Open Source Intelligence Techniques

Resources for Searching and Analyzing Online Information

Seventh Edition



Open Source Intelligence Techniques

Resources for Searching and Analyzing Online Information

Seventh Edition

Michael Bazzell

Open Source Intelligence Techniques:
Resources for Searching and Analyzing Online Information

Seventh Edition

Copyright © 2019 by Michael Bazzell

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means including information storage and retrieval systems without permission in writing from the author.

Seventh Edition First Published: October, 2019

Project Editors: Y. Varallo & A. Martin
Technical Editor: J. Edison

The information in this book is distributed on an "As Is" basis, without warranty. The author has taken great care in preparation of this book, but assumes no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

Rather than use a trademark symbol with every occurrence of a trademarked name, this book uses the names only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Due to the use of quotation marks to identify specific text to be used as search queries and data entry, the author has chosen to display the British rule of punctuation outside of quotes. This ensures that the quoted content is accurate for replication. To maintain consistency, this format is continued throughout the entire book.

Library of Congress Cataloging-in-Publication Data:
Application submitted

CONTENTS

About the Author **1**
Introduction **3**
SECTION I: OSINT Preparation **9**
CHAPTER 1: Computer Optimization **11**
Windows Reset 12
macOS Reset 12
Antivirus 13
Antimalware 15
Virtual Private Network 15
Password Manager 17
CHAPTER 2: Linux Virtual Machines **19**
VirtualBox 21
Ubuntu Linux 21
Snapshots 25
Exports and Clones 27
CHAPTER 3: Web Browsers **29**
Firefox Settings 29
Firefox Add-ons 32
Firefox Profiles 54
Chrome Settings 55
Chrome Extensions 55
Tor Browser 58
CHAPTER 4: Linux Applications **59**
Media Players 60
Media Download 61
Custom Scripts 63
Video Utilities 66
Instagram Utilities 71
Twitter Utilities 73
Documentation Utilities 79
Domain Utilities 82
Metadata Utilities 86
Linux Applications 87
Screen Capture Utilities 90
Linux Troubleshooting 92
CHAPTER 5: VM Maintenance & Preservation **95**
Master VM Creation 95
Application Updates 100
Custom Updates Script 101
Windows VM 102

CHAPTER 6: Android Emulation 105
Genymotion 106
Genymotion Configuration 106
Google Apps Installation 107
Android Apps 109
Contact Exploitation 115
Virtual Device Cloning 115
Virtual Device Export 116
Additional Android Emulation Options 117
CHAPTER 7: Custom Search Tools 119
HTML Tutorial 120
Element Inspection Tutorial 122
Tool Download & Modification 127
SECTION II: OSINT Resources & Techniques 129
CHAPTER 8: Search Engines 131
Google 131
Search Operators 132
Custom Search Engines 139
Alerts 143
Bing 143
Images 144
Archives 145
Translation 149
Groups 150
News & Newspapers 151
Search Engine Alternatives 156
Tor Search Engines 157
International Search Engines 159
Yandex 159
Yandex Operators 159
Private Search Engines 161
FTP Search 162
IntelTechniques Search Tool 165
CHAPTER 9: Social Networks: Facebook 167
Account Creation 168
Facebook Search: Official Options 170
Facebook Search: Profiles 172
Facebook Search: Base64 Encoding 174
IntelTechniques Facebook Search Tool 180
Facebook Search: Friends 182
Facebook Search: ID Creation Date 183
Facebook Search: Friends Extraction 183
Facebook Search: Email 184

CHAPTER 10: Social Networks: Twitter 187
Twitter Search 187
Twitter Search Operators 189
Deleted Twitter Posts 193
Twitter Biographies 196
TweetBeaver 197
Twitter Location Information 201
Tweet Deck 203
Twitter Analytics 205
IntelTechniques Twitter Tool 212
CHAPTER 11: Social Networks: Instagram 213
Instagram Search 213
Instagram Images 214
Instagram Metadata 216
Followers & Following 216
Likes & Comments 217
Third-Party Tools 220
Instagram Stories 222
IntelTechniques Instagram Tool 223
CHAPTER 12: Social Networks: General 225
LinkedIn 225
IntelTechniques LinkedIn Search Tool 228
Tumblr 230
Snapchat 230
Contact Exploitation 232
Account Export 233
International Networks 234
CHAPTER 13: Online Communities 237
Reddit 237
Reddit Deleted Content 238
Reddit Alternatives 244
Dating Websites 246
Forums 249
Online Prostitution 252
Craigslist 254
eBay 257
Amazon 259
IntelTechniques Communities Search Tool 260
CHAPTER 14: Email Addresses 263
Email Verification 263
Email Assumptions 264
Compromised Email Databases 265
Email Searching 271

IntelTechniques Email Search Tool 274
CHAPTER 15: Usernames 277
Username Search Engines 277
Compromised Accounts 280
Email Assumptions 281
IntelTechniques Username Search Tool 282
CHAPTER 16: People Search Engines 287
IntelTechniques Person Search Tool 293
Resumes 295
Gift Registries 297
Physical Addresses 299
Voter Registration 300
CHAPTER 17: Telephone Numbers 303
Carrier Identification 303
Caller ID Databases 304
Telephone Search Websites 310
Historical Search Websites 313
Search Engines 314
Loyalty Cards 316
IntelTechniques Telephone Search Tool 317
CHAPTER 18: Online Maps 319
Google Maps 319
Bing Maps 321
Additional Maps 321
Historic Imagery 324
Crowd-Sourced Street Views 326
IntelTechniques Maps Search Tool 328
Maps Manipulation 334
CHAPTER 19: Documents 335
Google Searching 335
Google Docs 336
Cloud Storage Providers 337
Presentation Repositories 339
Paste Sites 340
Document Metadata 342
Rental Vehicle Records 348
IntelTechniques Documents Search Tool 349
IntelTechniques Paste Sites Search Tool 349
CHAPTER 20: Images 353
Reverse Image Searches 353
Twitter Images 359
Flickr Images 359
Image Metadata 361

Image Manipulation 365
Image Forensics 365
IntelTechniques Images Search Tool 370
CHAPTER 21: Videos 373
YouTube Techniques 373
YouTube Comment Scraping 377
YouTube Deleted Videos 377
Reverse Video Searching 378
Video Search Options 381
Video Search Archives 383
Video Closed Captions 384
Live Video Streams 385
Periscope 385
TikTok 389
IntelTechniques Video Search Tool 391
CHAPTER 22: Domain Names 393
Domain Registration 393
Domain Search Tools 394
Historical Registration Data 396
Visual Depictions 399
Website Monitoring 403
Domain Analytics 404
Robots.txt 407
Search Engine Marketing Tools 409
Threat Data 412
Shortened URLs 415
IntelTechniques Domain Search Tool 418
CHAPTER 23: IP Addresses 419
IP Address Location 419
IP Address Search 420
Wigle 422
Shodan 423
IP Logging 425
IntelTechniques IP Address Search Tool 428
CHAPTER 24: Government & Business Records 431
General Government Records 431
General Court Records 431
General Business Records 434
Date of Birth Records 434
Social Security Records 435
Vehicle Identification Number Search 435
Vehicle Registration Search 436
Campaign Contributions 438

Criminal Information 439
Voter Registration Records 440
IntelTechniques Business & Government Search Tool 442
CHAPTER 25: Virtual Currencies 443
Blockchain 443
Virtual Currency Search 443
Virtual Currency Wallets 444
Virtual Currency Search APIs 444
IntelTechniques Virtual Currency Search Tool 446
CHAPTER 26: Advanced Linux Tools 449
Recon-ng 449
Skiptracer 459
Sherlock 460
Spiderfoot 461
CHAPTER 27: Data Breaches & Leaks 465
Voter Data 466
Ripgrep 467
Marketing Data 470
Snapchat Data 472
Combo List Collections 474
Hashes 478
Online Breach Data Resources 481
Data Leaks 484
Elasticsearch Databases 484
Elasticsearch Crawler 488
SQL Files 492
Usenet Data 494
IntelTechniques Breaches & Leaks Tool 499
SECTION III: OSINT Methodology 501
CHAPTER 28: Methodology & Workflow 503
OSINT Investigation Methodology 503
Case Workflow & Target Flowcharts 521
CHAPTER 29: Documentation & Reporting 531
Notetaking 531
OSINT Report Structure 534
Link Analysis & Timelines 539
CHAPTER 30: Policy & Ethics 555
Investigation Policies 556
Ethics 559
CONCLUSION: 563
INDEX: 564

About the Author

Michael Bazzell

Michael Bazzell investigated computer crimes on behalf of the government for over 20 years. During the majority of that time, he was assigned to the FBI's Cyber Crimes Task Force where he focused on various online investigations and open source intelligence (OSINT) collection. As an investigator and sworn federal officer through the U.S. Marshals Service, he was involved in numerous major criminal investigations including online child solicitation, child abduction, kidnapping, cold-case homicide, terrorist threats, and advanced computer intrusions. He has trained thousands of individuals in the use of his investigative techniques and privacy control strategies.

After leaving government work, he served as the technical advisor for the first season of the television hacker drama *Mr. Robot*. His books *Open Source Intelligence Techniques* and *Extreme Privacy* are used by several government agencies as mandatory training manuals for intelligence gathering and privacy hardening. He now hosts the *Privacy, Security, and OSINT Show*, and assists individual clients in achieving ultimate privacy, both proactively and as a response to an undesired situation.

INTRODUCTION

I taught my first Open Source Intelligence (OSINT) course in 1999 to a small group of local police chiefs in Illinois. I had not heard of the term OSINT at the time and I did not realize there was an official name for the methods I was teaching. I simply thought I was demonstrating a powerful way to use the internet as a part of everyday investigations. Most in the room had little interest in my session, as they had not experienced the internet much. Well, times sure have changed. OSINT is now quite the buzz word within several communities. Hacking conferences proudly host OSINT training sessions; cyber security groups include it within their strategies of hardening systems and customers; law enforcement agencies dedicate entire investigation units to online research into criminals; journalists rely on the methods every day; and even the social engineering crowd has adopted it as part of their playbook. I never knew OSINT would become a household name within most technology circles.

My company continues to provide OSINT training sessions, and the audiences seem to grow every year. It is no longer a course reserved for tech-savvy employees. We now see people with minimal online experience being thrown into investigation and analyst positions. We see crowds desperate for the latest investigation techniques, only to see those methods disappear without notice as social networks come and go, or change their search options. Search techniques seem to be more fickle than ever, and I am always concerned about losing an online resource.

Seventh Edition

The previous (sixth) edition of this book was originally released in January of 2018. Every time I publish a new version, I incorrectly assume it will be the final release. It always feels like a "complete" work, and I become naïve while thinking that nothing major will change in the near future. 2019 was a wake-up call. It presented the most drastic changes to my site and various online resources. Within a few months, many of the resources discussed in the previous edition disappeared, even my own search tools and forum. It all began at the beginning of the year.

In January, I noticed disturbing behavior within my public online forum. There were over 9,000 registered users, and it had always been a very active area for those in the online investigations and digital privacy communities. As with any forum, popularity brings trouble. Every day, a handful of users conducted automated attacks against the forum and the rest of my site. The attempts were malicious in nature, but there is no evidence any sensitive information was ever obtained. This varied from scraping usernames and messages to attempting various injections trying to gain access to other areas of the site. It was only a matter of time before a vulnerability surfaced that caused real trouble. This was a serious liability for me, and a chance I no longer wished to take. Online forums are notorious for being insecure and often used as an avenue to

unauthorized access to a website. Legal representation for my company had been pushing me to eliminate the public forum for over a year, but I always resisted.

Twice daily, my moderators needed to remove posts and sometimes remove members due to those who had refused to honor the rules. The number of hateful posts and threats had increased substantially. This nuisance distracted these employees from their other jobs, and they grew tired of the personal attacks from unhappy members when they were banned. This is common for all forums, and I was impressed with how long it took to attract the trolls and hateful people. I think that says a lot for the OSINT community. We had a great run, but in May of 2019, I closed the public forums permanently.

On June 6th, 2019, I woke up to find my website offline and my email inbox full of messages from confused members of my online video training. Overnight, my web host (Namecheap) had suspended my account and knocked my website completely offline. I had two emails from them waiting in my inbox. The first was from the abuse department and the second from their legal team. The abuse email notified me that my site was under attack and that I had exceeded my resources provided to my business plan. A severe Distributed Denial Of Service (DDOS) style of attack had been launched against my free online search tools. Simultaneously, a large company had threatened legal action against me (and my host) for encouraging my visitors to formulate search queries which revealed exposing (yet public) information. It appeared to be an orchestrated attack to shut down my online resources. Furthermore, and somewhat coincidentally, Facebook disabled their standard graph search on the same date. Graph search had allowed us to dig into user profiles for years and was a staple of my site and training. That was a tough day.

I was able to negotiate with Namecheap to get my site back online, but with restrictions. I was forced to place my free online search tools behind a login and out of access from the public. I was given strictly enforced resource limits and threatened with permanent suspension if I did not comply with the orders. I did not like this, but I had to think about the paid users of my online video training. I owed them unrestricted access and needed a solution fast. I also had a business to run, and my website was vital for that mission.

Word had spread throughout the community that the tools were gone. I received an equal mix of appreciation for the free usage over the past decade, and hateful messages from those who believed I had cheated them by removing the free resource. Several imposter sites appeared almost overnight, but none had the same functions as my tools which relied on PHP to fetch data and present it in a nice format. Today, there are at least ten websites attempting to recreate my previous public tools in order to sell their own training and services. One site is selling access to the previously free options. One major goal of this book is to teach you how to create, maintain, and possess your own set of search tools which can never be removed or censored. You will no longer need other sites to provide you with the tools you need. There is much more on this to come, but that was the motivation to start writing this book.

In August of 2019, I finally felt the urge to tackle the new world of OSINT and begin writing about online research methods again. Because so much had changed since the last edition, I knew I could easily meet my self-imposed rule in reference to my book revisions. The potential release must include at least 25% brand new material, 25% updated content, and 25% untouched stable and beneficial techniques. I believe that this seventh edition exceeds these criteria.

My primary goals with this edition are three-fold. First, we need to stop relying on third-party resources and tools. When my Facebook custom search tool was taken offline, I was bombarded by people pleading for access. They stated that they relied on it to conduct their investigations and were stuck until they had a replacement. I sympathized with their situation, but I was under severe constraints legally and from my web host. Many said I should simply move to a new host, but that did not solve the legal threats. Anywhere I moved would be closely monitored by those who wanted my site offline. I believe I have a better solution. This book will not rely on any of my previous online search tools. Instead, it will teach you how to make your own. We can no longer assume that our preferred search sites will be available forever.

This also applies to pre-built virtual machines. I have offered an OSINT virtual machine called Buscador for four years. While it served a great purpose, it became outdated. In this edition, we will build our own custom Linux virtual machine. This will help you understand every detail of the process and allow you to maintain it properly without reliance on an online repository. This will also allow you to properly explain how the tools work, and display much more confidence in your reports and testimony. Again, we need to be self-reliant.

Next, I have updated the entire content of the previous edition to be 100% functioning as of publication. I removed all old and outdated information, added many new options, and updated any resources which saw major changes. The "guts" of this book will be very similar to previous editions, but with a fresh explanation of each new technique.

Finally, I offer a section specifically tackling methodology. Knowing a bunch of websites which have data about your target is not very helpful if you do not know what to do with the information found. I see a new OSINT link collection appear online daily, sometimes with thousands of links. I do not believe these are very helpful unless you know where to start. The final section of this book will teach you the proper flow of investigating targets, documenting your findings, and explaining the valuable intelligence to those who receive your reports. I also include reporting templates to assist. I relied heavily on Jason Edison for this section, as he works full-time in this space and provides strategies which are beyond my capabilities. You will learn more about him in a later chapter.

Keeping a book up to date about ways to access information on the internet is a difficult task. Websites are constantly changing or disappearing, and the techniques for collecting all possible public information from them are affected. While the sixth edition of this book is still highly applicable, a lot has changed over the past two years. Much of this book contains new techniques that were previously not available. A growing number of Python tools has bombarded us with

new capabilities never available before. It is a very exciting time for internet investigations, and this book will explain how to harness these advanced technologies.

Fortunately, knowing methods for accessing data on one website often carries over nicely to other websites. This entire seventh edition was accurate as of October 2019. If, or more likely when, you find techniques which no longer work, use the overall lessons from the entire book to push through the changes and locate your content. Once you develop an understanding of the data, you will be ready to adapt with it.

What is OSINT?

Open Source Intelligence, often referred to as OSINT, can mean many things to many people. Officially, it is defined as any intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement. For the CIA, it may mean information obtained from foreign news broadcasts. For an attorney, it may mean data obtained from official government documents that are available to the public. For most people, it is publicly available content obtained from the internet.

Overall, this book includes several hundred sources of free information and software which could identify personal information about anyone you might be investigating. All of the resources are 100% free and open to the public, with a few minor exceptions. Each method is explained, and any creative search techniques involving the resources are detailed. When applicable, actual case examples are provided to demonstrate the possibilities within the methods. The book can be read in any order and referenced when a specific need arises. It is a guidebook of techniques that I have found successful in my investigations.

Locating this free online information is not the final step of OSINT analysis. Appropriate collection and reporting methods will be detailed and referenced. Whether the data you obtain is for an investigation, a background check, or identifying problem employees, you must document all of your findings. You cannot rely on the information being available online forever. A website may shut down or the data may be removed. You must preserve anything of interest when you find it. The free software solutions presented here will help you with that. OSINT search techniques do not apply only to websites. There are many free programs that automate the search and collection of data. These programs, as well as application programming interfaces, will be explained to assist the advanced investigator of open source intelligence. In summary, this book is to serve as a reference guide to assist you with conducting more accurate and efficient searches of open source intelligence.

What the book is not…

This is not a debate of the various opinions about online reconnaissance for personal information. It is not a historical look at OSINT or a discussion of your administrative policy.

Furthermore, it is not a how-to guide for criminals to steal your identity. Nothing in this book discusses illegal methods of obtaining information.

Book Audience

When I first considered documenting my OSINT techniques, the plan was to post them on my website in a private area for my co-workers. This documentation quickly turned into over 250 pages of content. It had grown too big to place on my site in a manner that was easy to digest. I changed course and began putting together this book as a manual to accompany my multiple-day training sessions. It has grown to a 600 page textbook which could never include every beneficial resource on the internet.

Many readers are in some form of law enforcement. Police officers can use these techniques to help locate missing children or investigate human trafficking. Intelligence analysts can apply these methods to a large part of their daily work as they tackle social media posts. Detectives can use the search techniques to re-investigate cases that have gone unsolved.

I also offer my live OSINT training to the private sector, especially security divisions of large corporations. This book can help these teams locate more concise and appropriate information relative to their companies. These methods have been proven successful for employees who monitor any type of threat to their company, from physical violence to counterfeit products. I encourage the use of these techniques to institutions that are responsible for finding and eliminating "bad apples". This may be the human resources department, applicant processing employees, or "head hunters" looking for the best people. The information about a subject found online can provide more intelligence than any interview or reference check.

Parents and teachers are encouraged to use this book as a guide to locating social media content posted by children. In many households, the children know more about the internet than the adults. The children use this to their advantage and often hide content online. They know that it will not be located by their parents and teachers, and often post inappropriate content. This book can empower the adults and assist with identifying important personal information which could pose a threat toward children.

A large portion of my intended audience is private investigators. They can use this book to find information without possessing a deep understanding of computers or the internet. Explicit descriptions and occasional screen captures will ensure that the techniques can be recreated on any computer. Several universities have adopted this book as required reading, and I am honored to play a small role in some amazing courses related to network security.

I realize that people who use these techniques for devious purposes will read this book as well. Colleagues have expressed their concern about this possibility. My decision to document these techniques came down to two thoughts. First, anyone that really wants to use this information in malicious ways will do so without this book. There is nothing in here that could not be replicated

with some serious searching and time. The second thought is that getting this information out to those who will use it appropriately is worth the risk of a few people using it for the wrong reasons. Please act responsibly with this information.

Finally, a parting thought before you begin your journey through OSINT analysis and collection. This book was written as a reference guide. It does not necessarily need to be read straight-through sequentially, but it was written as a chronological guide for most investigators. I encourage you to skip around when needed or if you feel overwhelmed. The second chapter about Linux may make you want to abandon the teachings before ever utilizing an online resource or website. When you encounter material that seems too technical or not applicable, please move on to the next topic and consider returning later. The book is suitable for all skill levels, and there is something here for everyone. You can always return to the advanced topics when more appropriate.

Digital Files

Throughout this book, I refer to several files which can be downloaded in order to simplify your usage of the various tools, techniques and scripts. These are hosted on my website, and available free to you under a creative commons license. Below is the link to the page on my site which offers all files referenced in the book in case you want to download them in advance.

https://inteltechniques.com/osintbook

I suspect many readers are concerned about the availability of these files if my site was suspended again or if I received another threat of lawsuits over my search tools. I have the same concern. Therefore, I have placed copies of these files on two different third-party file hosts, both outside of the United States, which makes them immune to many generic legal threats and take-down demands. I am sure this is overkill, but I hope it provides another layer of usability and sustainability for the content of the book.

Tools: mega.nz/#!7HwmyYrA!LE1RTmoReD7J5J1F3spxA6Cu5f1dMROkiPTOJtv3gMc
Files: mega.nz/#!mKwQjYBI!1fcu99HUjr3FW2CnxIHuKXA7KfpoWaqhC_WICOHpxxw
VM Steps: mega.nz/#!LS42hAiJ!8laKuFWZ88O4DHuM5XscXSXfyPWkD12m34ahjERru5k
Workflow: mega.nz/#!nOxkEaAQ!Fyn5a8VrklH0Da9JNXYSRcD9FcGdt6X3AorjHo_Opf4
Templates: mega.nz/#!PWgkxA5A!ljcm1y7JMgnSr2WdutVIOO_zkpDo4q1DfwdgZBkFqSg

Tools: app.nihaocloud.com/f/5df00ef7ecf442c9995e/?dl=1
Files: app.nihaocloud.com/f/1d142ce1628348e5b095/?dl=1
VM Steps: app.nihaocloud.com/f/ae8a8f624c814c428239/?dl=1
Workflow: app.nihaocloud.com/f/d1c7802a56224229a4d5/?dl=1
Templates: app.nihaocloud.com/f/5f664bbf1b2f41279d1a/?dl=1

SECTION I
OSINT PREPARATION

This entire section explains the essential steps which I believe any online investigator should complete before ever conducting a search online. We should never jump into an investigation without being digitally secure with a clean computer and software which has not been compromised from previous activity. We should jump into each investigation with confidence, knowing we are working from an environment without any contamination from previous investigations. It will take a lot of work to create our perfect playground, but replicating a pristine environment for each investigation will be easy.

The first four editions of this book began with search engine techniques. Right away, I offered my methods for collecting online information from various popular and lesser known search websites. This may have been due to my own impatience and desire to "dive in" and start finding information. This edition will begin much differently. Before you attempt any of the search methods within this book, I believe you should prepare your computing environment. I was motivated to begin with this topic after teaching a multiple-day OSINT class. On day two, several attendees brought laptop computers in order to attempt the techniques I was teaching during the course. During a break, I observed police officers searching Facebook on patrol vehicle laptops; private investigators using Windows XP while browsing suspects' blogs; and cyber security professionals looking at hacker websites without possessing any antivirus software, script blockers, or a VPN.

I have also been guilty of all of this. Early in my career of researching OSINT, I did not pay any attention to computer security or proper browsing habits. While I was aware of malicious software, I knew I could reinstall Windows if something really bad happened. This was reactive thinking. I believe that we must all proactively attack vulnerabilities in our privacy and security while conducting online research. This section is not meant to be a complete guide to computer security or a manual for total privacy. Instead, I hope to quickly and efficiently propose the most beneficial strategies which will protect you from the majority of problems. Applying the changes mentioned in this section will provide a valuable layer of security to your online investigations and overall computing habits.

In the following chapters, I will explain how to ensure your computer host is secure; configure virtual machines for each investigation; customize OSINT software which will be available at all times; create your own set of search tools to automate queries; prepare a virtual Android environment for mobile investigations; and easily clone all of your work for immediate replication if anything should become corrupt, damaged, or compromised. Your efforts now will pay off ten-fold in the future. Let's get started.

Chapter One
Computer Optimization

What data is on your computer? Is there a virus, malicious software, or spyware hanging around from casual browsing in questionable places? Does your internet cache include tracking cookies from Amazon, Facebook, Google, and others? Is there evidence of your last investigation stored within your bookmarks, documents, or download queue? If the answer is "maybe" or "I don't know" to any of these, you have a contaminated computer. If your investigation enters court testimony, you may have a lot of explaining to do once an expert witness who specializes in computer forensics takes the stand. If your screen captures display evidence unrelated to your investigation, you could compromise the entire case. You may think I am being overly cautious, but we can no longer take any chances when it comes to the purity of our online evidence.

First, we need to focus on the idea of a clean host. Your host machine is your traditional physical computer. It may be the laptop or desktop owned by your agency or purchased with personal funds. It is the device which you obtain your access to the internet, but not necessarily the direct avenue which will be used for your investigations. In the next chapter, I present my options for protection during online investigations by using a virtual machine (VM) on top of your host. Before we can consider building a VM, we must know we have a host without any contamination.

In a perfect world, you have just purchased a brand new computer and are ready to tackle your first investigation with it. There is no contamination because you have yet to turn it on. In the real world, you are stuck with used equipment, repurposed gear, and outdated operating systems. Regardless of whether you have a new machine or a hand-me-down, this chapter contains vital steps we must take before going online. Let's begin with the used machine.

Assume you possess a laptop computer which you have had for a year or two. You installed some traditional software such as Microsoft Office and maybe added a better browser such as Firefox. You have checked your email, logged in to your Amazon account, and browsed the web as anyone else would with their own machine. Each time you visit any website, you collect detailed temporary files which are personalized only for you. They contain session details and account profile data. We all constantly leave a digital trail within every device we touch. Now imagine that you begin an investigation into a target on this same machine. A Google search leads you to an Amazon Wishlist. Loading to that page connects your Amazon account to the query, and your name is visible within the screen capture. Even when you log out of services such as Amazon, persistent "cookies" linger and let companies know you are still online. They follow you to websites you visit, attempting to collect your interests in order to present targeted advertisements. Your fingerprint is now all over this investigation. I pick on Amazon, but Google, Facebook, and others are much worse.

If I have not convinced you that your machine is contaminated, consider the following scenario, which happened to me many years prior. While investigating your suspect, you check your email in another browser tab. You also take advantage of various instant messenger applications in order to communicate with colleagues. You finish your investigation and submit your evidence for discovery. A suppression hearing is scheduled because the opposing party in your case wants evidence thrown out. During arguments, the other side demands an exact clone of the computer used during the investigation be provided to their own digital examiner. The judge agrees, and orders you to allow the opposing side to make an identical clone of your machine. You begin thinking about the personal activity and online purchases which are now going to surface during this trial.

We can avoid all of this. I present a firm rule which will not sit well with all readers. You should possess a dedicated machine for the sole use of online investigations. No personal usage and no unnecessary activity. It should be a machine only used as part of your profession. Even if you only have a used computer, you can bring life back to the machine and start over. This requires much more than simply deleting files and running a computer cleaning application. These only remove the obvious data, and will never truly eliminate all contamination from the machine. To do things right, we must completely reformat and reinstall all software. **This will erase all data on your machine, so proceed with caution!** Let's attack this from the two most common operating systems.

Windows: First and foremost, backup any important data. Connect an external drive via USB and copy any documents, configuration files, and media which will be removed when you reformat the machine. Common locations include the Desktop, Downloads, and Documents folders within the home folder of the current user. Double check that you have everything you need, because the next step is to remove all data from the drive. Most modern Windows computers possess a hidden "restore" partition. To factory reset Windows 10, go to Start > Settings > Update & Security > Recovery and click the Get started button under Reset this PC. Select Remove everything, which results in the following two options:

Just remove my files
Remove files and clean the drive

Choose the "clean the drive" option and wait for completion. The result will be a new operating system free of any previous contamination. If you do not have this option, or possess an older Windows operating system, you will need the original installation media or a restore CD from the manufacturer. Upon boot, refuse any requests to create a Microsoft account, and only provide the necessary information to log in to Windows, such as a vague username and password.

Mac: Similar to Windows, make a backup of any valuable data. Common locations include the Desktop, Downloads, and Documents folders within the home folder of the current user. Restart the computer and immediately hold the Command and R keys until you see the Apple logo. Release the keys and proceed to the next step. While in Recovery Mode, you will see the "macOS

Utilities" window. Choose Disk Utility and click Continue, then select your startup disk and click Erase. Select Mac OS Extended (Journaled) as the format, click Erase, and wait until the process is finished. With your hard drive completely erased and free of any data, you can perform a clean installation of macOS. From the same macOS Utilities window, choose Reinstall macOS (Reinstall OS X in older versions). Allow the process to complete and reboot the machine. Create a generic login account and you have a brand new system.

You should now have a computer with no previous internet usage. This is our clean host. Now, we need to apply protection to the host, including antivirus and a solid VPN. It is likely that most readers already have an antivirus solution and are insulted at the mention of it in a book like this. I will keep my thoughts very brief. If you are using Microsoft Windows, you absolutely need antivirus software. If you are using an Apple computer, you might not. Antivirus applications only protect against known variants of viruses. They do not stop everything. A new virus can often bypass the best software detection solutions. A better defense is applying better browsing habits instead of relying on an application.

Antivirus (Windows)

There are a dozen popular antivirus companies that will provide a free solution. For most Windows users, I simply recommend to use Microsoft's products. Users of Windows 7 should use Microsoft Security Essentials, while Windows 8 and 10 users should use the default Windows Defender included with their installation. Privacy enthusiasts will disagree with this advice, and I understand their stance. Microsoft products tend to collect your computer usage history and analyze the data. Unfortunately, their core operating systems also do this, and it is difficult to disable long term. Therefore, I believe that Windows users are already disclosing sensitive information to Microsoft. Using their antivirus solutions will not likely enhance the data being collected.

Antivirus (Mac)

Mac users do not have any built-in antivirus protection, and most do not need any. The software architecture of Mac computers is much more secure, and viruses are rare (but they do still occur). I no longer recommend the free commercial products such as Avast, Kaspersky, and others. They tend to be more of an annoyance than helpful, and their business practices can be questionable. However, I do believe that it is irresponsible to have absolutely no protection whatsoever. When I conduct investigations from a Mac computer, I possess an open-source solution called ClamAV.

ClamAV (not to be confused with the unnecessary paid option of ClamXAV) is a community-driven antivirus database, which is freely available to anyone. It usually does not score very high on "Top 10 Antivirus" websites, which are commonly paid advertisements. However, it is completely free; does not run on your system non-stop; only executes when you desire; and can be completely removed easily. Unfortunately, there is no easy software installation process, and no point-and-click application. You will need to manually update the database through a Terminal

command, then scan your system from the same prompt. ClamAV does not remove any viruses by default, it only discloses the presence and location of suspicious files. In my use, ClamAV has never found a virus which impacted a Mac computer. Instead, it has identified numerous malicious files which target Windows machines, but were present on my system (mostly as email attachments). This notification allowed me to manually remove those files, which could prevent future infection of my Windows virtual machines. If you have concerns about having a "naked" Mac with no antivirus, the following instructions will configure your Mac to be better protected.

First, you must install a package manager called Brew. This program is very beneficial when there is a need to install programs that would usually already be present on a Linux computer. It also happens to have a pre-configured version of ClamAV ready to go. The easiest way to install Brew is to visit the website brew.sh and copy and paste the following command into the Terminal application (Applications > Utilities > Terminal).

/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"

After Brew is installed, type the following commands, hitting "Return" after each line, into the same Terminal application used previously.

- `brew analytics off`
- `brew install clamav`
- `sudo mkdir /usr/local/sbin`
- `sudo chown -R` `` `whoami` ``:admin /usr/local/sbin`
- `brew link clamav`
- `cd /usr/local/etc/clamav/`
- `cp freshclam.conf.sample freshclam.conf`
- `sed -ie 's/^Example/#Example/g' freshclam.conf`

These steps will install ClamAV; switch to the installation directory; make a copy of the configuration file; and then modify the configuration file to allow ClamAV to function. You are now ready to update your antivirus database and conduct a scan. Type the following commands into Terminal, striking return on your keyboard after each line.

- `freshclam -v`
- `clamscan -r -i /`

The first option will download all virus definition updates, and should be executed before each scan. The second option conducts a scan of the entire computer, and will only prompt you with details of found viruses. While it may appear to be dormant, it is working, and will notify you upon completion. All of these commands must be exact. In order to assist with this, I have created a web page with all of these commands at https://inteltechniques.com/clamav.

ClamAV may occasionally present a false-positive report of a virus. Do not panic. Research the file on the internet and identify the issues. If you receive reports of malicious files within email,

simply delete those messages. Note that the above scans only SEARCH for viruses, they do not REMOVE threats. If you would like to conduct a scan and automatically remove suspicious files, you must conduct a different command. Please note this could be dangerous, and could permanently remove necessary files. I always run a scan, research the threats found, and execute the following scan ONLY if I am confident the files should be removed.

- `clamscan -i -r --remove=yes /`

Antimalware

The use of ClamAV on Mac and Linux computers is more about preventing the spread of bad files to Windows users instead of protecting your own machine, but viruses do exist for non-Windows systems. I try to scan my entire drive at least once monthly on all machines. Whether on Windows or Mac computers, protection from malicious software, otherwise known as malware, is vital. Again, there are numerous free options from which to choose. I recommend Malwarebytes for both Windows and Apple users. It is completely free and thorough. I suggest executing, updating, and scanning at least once a week on every device that you use.

- Navigate to http://www.malwarebytes.org/ and select the "Free Download" option.
- Conduct a default installation.
- On a weekly basis, launch the program, update the database, and conduct a full scan.
- Malwarebytes will remove any issues it finds.

If prompted, decline any premium features. The free version is sufficient and preferred. Proper antivirus and antimalware protection will greatly enhance your overall computing experience. It will help your computer to run smoothly and may prevent malicious files from infecting your operating system. It will help protect the integrity of any online investigations. I refer to these steps as the "staples". They are the minimum requirements before proceeding and apply to any computer user. Ideally, you will never use your host operating system for any web browsing or investigations, and all of this will be overkill. However, it is better to be safe than sorry.

Virtual Private Network (VPN)

I believe that every OSINT researcher should possess and use a virtual private network (VPN) at all times. A VPN extends a private network across a public network, such as the internet. It enables users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network, thus benefiting from the functionality and security of the private network. A VPN masks your identity online. Two specific examples should help demonstrate the need for this resource.

If you are on your home computer, and connected to the internet, you are using a connection from your internet service provider (ISP). If you navigate to a website that is monitoring visitors, it knows your IP address, approximate location, and internet provider and type (cable, DSL, etc.).

However, if you are on that same computer, with the same internet connection, you can use a VPN to protect you. The VPN software connects your computer to one of their servers over the internet connection. This encrypted traffic cannot be deciphered by the ISP. When your traffic gets to the VPN server, it sends and receives your data, returning incoming packets to you. The websites that you visit believe that you have the IP address of the VPN server. They do not know what type of internet connection you possess nor your location.

Some readers may wonder why they cannot simply use the free Tor service for this scenario. While you could, it is not always advised. Tor connections can be too slow for constant use. Also, some websites will not let you access their services through a Tor proxy. Connecting to your bank through Tor will likely set off alarms, and may prevent you from access. I believe that Tor is great when you truly need to hide your entire connection, and I will discuss more on that later. I believe that every day browsing is better suited for a VPN.

If you work for a large corporation, you may already have access to a corporate VPN. Ask around and identify your options. These are great for security, but not so much for privacy. I never recommend a corporate VPN for online investigations. Instead, you need to purchase a VPN service. While there are a few providers that give away free VPNs, I never recommend them. They are extremely slow and often use your internet connection for other people's traffic. Instead, consider purchasing access from a reputable provider such as ProtonVPN or PIA.

VPNs can be launched in many ways. Some run through a firewall or router, which may be overkill for your needs. This is especially true if you conduct investigations from a laptop at multiple locations. Some use various web browser extensions which allow the VPN to intercept data. I do not recommend this as it would only protect your browser traffic instead of your entire host. My advice for you is to protect your entire host computer with a dedicated VPN application. This will also protect your virtual machines, which will be explained in the next chapter.

Both ProtonVPN and PIA provide a software application to all premium user accounts. I find this sufficient for our needs and installation is easy for both Windows and Mac. Each of these providers allow you to connect to your choice of dozens of servers worldwide. I can choose California when I want to appear on the west coast or New York when I want to appear in the east. I can choose London in order to bypass restrictions while watching the BBC online or Toronto when I need to appear as a Canadian user. Your yearly access can be used on up to ten devices simultaneously. My personal policy on VPNs is quite simple. I always use a VPN on any device that I connect to the internet. This includes desktops, laptops, and cell phones.

Lately, I have had better success with ProtonVPN than PIA in regard to online investigations. PIA is one of the largest VPN providers in the world, and many sites block them due to abuse and fraud. ProtonVPN is equally as secure, but not as popular. When a website blocks me because I possess an IP address from PIA, I can almost always connect to the site by switching over to ProtonVPN. ProtonVPN is a bit more expensive than PIA, but you may have less headaches. Any reputable VPN is better than no protection.

Password Manager

While you conduct your online investigations, you will likely create and maintain numerous accounts and profiles across various services. Documenting the profile details, including passwords, for these services can be a daunting task. A password manager provides a secure database to store all of the various settings in regard to these profiles. My choice is KeePassXC.

KeePassXC is an opensource password manager that does not synchronize content to the internet. There are many convenient online password managers which are secure and keep all of your devices ready for automated logins. Those are great for personal security, and millions of people are safely using them. However, it is not enough for our needs. Since you will be storing data connected to online investigations, you should protect it in an offline solution. KeePassXC is cross-platform and free. It will work identically on Mac, Windows, or Linux. Download the software from https://keepassxc.org, and conduct the following as an exercise.

- Launch KeePassXC and select Database > New Database.
- Provide a name to your new password database, such as Passwords.
- Move the encryptions settings slider completely to the right and click "Continue".
- Assign a secure password which you can remember but is not in use anywhere else.
- Click "Done" and select a safe location to store the database.
- Close the program and verify you can open the database with your password.

You now have a secure password manager and database ready for use. Assume you are ready to change the password to your covert Facebook profile. Navigate to the menu which allows change of password. Next, conduct the following within KeePassXC.

- Right-click within the right column and select "New Group".
- Name the group Facebook and click "OK".
- Select the Facebook group on the left menu.
- In the right panel, right-click and select "New Entry".
- Provide the name of your covert account, username, and URL of the site.
- Click the black dice icon to the right of the "Repeat" field.
- Click the eyeball logo underneath the black dice logo.
- Slide the password length slider to at least 40 characters.
- Copy the generated password and paste into the "Password" and "Repeat" fields.
- Change your Facebook password to this selection within your account.
- Click "OK" and save the database.

You successfully created a new, secure, randomly generated password for your covert profile. You will not remember it, but your password manager will. From this moment forward, you will change every password to any site that you access upon logging in. The next time you log in to

your secure sites, change any passwords. Allow your password manager to generate a new random password containing letters, numbers, and special characters. If the website you are using allows it, choose a password length of at least 50 characters. When you need to log in, you will copy and paste from the password manager. For each site which you change a password, your password manager will generate a new, unique string. This way, WHEN the site you are using gets breached, the passwords collected will not work anywhere else. More importantly, recycled passwords will not expose your true accounts after the breached data becomes public. There should be only a handful of passwords you memorize, which brings us to the next point.

The password to open your password manager should be unique. It should be something you have never used before. It should also contain letters, numbers, and special characters. It is vital that you never forget this password, as it gives you access to all of the credentials that you do not know. I encourage users to write it down in a safe place until memorized. It is vital to make a backup of your password database. When you created a new database, you chose a name and location for the file. As you update and save this database, make a copy of the file on an encrypted USB drive. Be sure to always have a copy somewhere safe, and not on the internet. If your computer would completely crash, and you lose all of your data, you would also lose all of the new passwords you have created. This would be a huge headache. Prepare for data loss now.

If you really want integrated browser support, KeePassXC has this option. You can install the browser extension and easily enter passwords into websites without leaving the browser. This will be explained later. I believe this is safe, and I have used it during investigations in the past. Today, I believe that copying passwords into websites should be a deliberate act that requires effort without automation. I don't want a machine doing this for me. The attraction to online password managers such as Lastpass and Dashlane is the ability to sync the password database to all devices over the internet. I understand the benefits of these features, but it also comes with risk. All reputable online password managers encrypt the passwords locally on the user's device before syncing with their own servers. Theoretically, no one at the password manager company would have the ability to see your individual passwords. However, nothing is hack-proof. It is only a matter of time before something goes wrong.

By keeping your passwords in an offline database, you eliminate this entire attack surface. By keeping your password manager ready in your host machine, you will have immediate access to it regardless of which virtual machine you are using during an investigation. That brings us to the next chapter. It is now time to create our investigation environment.

CHAPTER TWO
LINUX VIRTUAL MACHINES

Linux operating systems have been a part of my OSINT investigations and trainings for many years. They are lightweight, run on practically any hardware, cost nothing, and provide a level of security that cannot be obtained through traditional operating systems such as Microsoft Windows. During my training, I often demonstrate how I use Linux as a virtual machine (VM) or bootable USB device. In both scenarios, I can navigate to any malicious website, download every virus possible, and eliminate all traces of my activity by simply rebooting the system while reverting the VM. Upon reboot, there are no viruses and everything works exactly as intended when the system was created.

This chapter presents ways that you can harden your security by using a Linux operating system during your investigations. Many years ago, this may have been intimidating to non-technical users. Today, implementing Linux into your investigations is extremely easy. This chapter is intentionally at the beginning of the book in order to better protect you and your investigations right away. Once we start exploring the world of online search techniques, you will likely encounter malicious software or viruses at some point. If you investigate cyber criminals, this will be sooner rather than later. The malicious code will almost always target Windows machines. By choosing Linux as your investigations system, you greatly lessen the concern about infections. This chapter will vary from basic tutorials through advanced technologies. I present material on Linux before online searching because I want you to have a safe and secure environment for your research, without the fear of exposing your personal computer.

In 2015, I actively taught methods that would take a standard Linux system, such as Ubuntu or Mint, and install it to a USB device. This device could be inserted into a computer, booted, and a native Linux system would be present. When the machine was turned off, all history of that session was eliminated. This was a quick and easy way to conduct high-risk investigations while protecting the integrity of a personal computer. Unfortunately, this was slow, mostly due to the speed bottleneck of the USB device. It was a valid practice with good intentions, but not extremely applicable to most OSINT investigators today. Previous editions of this book had an entire chapter devoted to creating these devices.

In 2016, I was contacted by David Westcott. We knew each other through our OSINT work, and he asked if I was interested in creating a custom OSINT virtual machine. I had always considered this, but had concerns about my skills at hardening Linux systems and pushing out finished builds. David had worked on other public Linux releases, and was much more comfortable distributing custom systems. I began designing my dream OSINT build, sending him weekly requests, and he began taking the ideas and executing them within a test product. By 2017, the first public version of our new operating system was released and titled Buscador (*Seeker* in Spanish).

This concept is not new. Many readers are likely familiar with Linux systems such as Kali. This is a custom Linux build that includes hundreds of security testing tools for the cyber security community. It is considered an all-in-one operating system that has everything pre-configured upon installation. We wanted that same experience for the OSINT community. Buscador was designed from the ground up with considerations for OSINT investigations. The web browsers were pre-configured with custom settings and extensions, and numerous OSINT software applications are already set-up to accept your search queries.

An important part of Buscador was the applications. On many Linux builds, launching software is not similar to traditional operating systems. While the software is installed, you must still launch a Terminal window and type the specific commands required. This can be very difficult and unforgiving. There are seldom point-and-click icons that launch similar to Windows. This has always created a barrier between the geeks and the norms. Either you know how to issue Linux commands or you do not. If you don't, then you never get to take advantage of the power of Linux and Python.

We wanted to eliminate that barrier. We wanted to make powerful Linux programs easily accessible to everyone. My initial thought was to create Bash scripts similar to batch files in Windows, but David came up with a much easier and more appropriate way. Every tool inside Buscador had its own icon in the dock, executed by clicking with a mouse, which walked the user through the menus. After collecting the required data, each program executed the proper commands behind the scenes and delivered the content directly to the user. We believed this to be unique in our community. Every person, at any skill level, could use Buscador as a virtual machine.

Today, my views of pre-built virtual machines have changed slightly. While I am proud of our work with Buscador, and I still use it, I believe we should no longer rely on systems from third parties. Buscador is no longer updated with new versions, and there is no online repository to apply updates to the many applications within the virtual machine. When my web host suspended my account and took my site offline, I received hundreds of emails asking where Buscador could be downloaded. We should never rely on a single source for our tools. Many of the people using Buscador now receive errors due to outdated applications and most users do not have the training to apply their own updates in order to correct any issues. Creating a virtual machine which was user friendly had many benefits, but also some unintended consequences. My goal in this chapter is to help you create and maintain your own OSINT virtual machine.

Virtual Machines

Virtual machines (VMs) conduct emulation of a particular computer system. They are computer operating systems on top of computer operating systems. Most commonly, a software program is executed within an operating system, and individual operating systems can launch within that program. Each virtual machine is independent from the other and the host operating system. The environment of one virtual machine has no impact on any others. Quite simply, it is a way to

have numerous computers within your single computer. When finished, you can safely investigate a single target within a secure environment with no contamination from other investigations. You will be able to clone a master VM in minutes and will no longer need to worry about persistent viruses, tracking cookies, or leftover evidence.

Before creating a virtual machine, you must possess virtual machine software. There are several programs that allow you to create and execute virtual machines. Some of these are paid programs, such as VMWare. However, I will focus only on VirtualBox here. VirtualBox is completely free and easy to operate. All methods presented here for VirtualBox can be replicated on VMWare or any other virtualization software.

VirtualBox (virtualbox.org)

Volumes could be written about the features and abilities of VirtualBox. I will first explain how to install the application and then ways to configure a virtual machine. VirtualBox installation instructions can be found at virtualbox.org but are not usually straightforward. At the time of this writing, the following steps installed VirtualBox to my MacBook Pro.

- Navigate to virtualbox.org/wiki/Downloads.
- Click the first appropriate link for your OS, such as "OS X Hosts" or "Windows Hosts".
- If required, select the highest version NOT including "Beta", such as "6.0.14".
- Download the appropriate "dmg" file for Mac or "exe" file for Windows.
- Download the Extension Pack for your version, such as "6.0.12.vbox-extpack".
- Install the dmg or exe file with all default settings.
- Double-click the Extension Pack and add it to VirtualBox.

The only requirement for VirtualBox to function is a computer that supports virtualization. Any modern Apple product will work without any modification. Most mid-range and high-end Windows computers made within the past five years should have no problem, but may require you to enable virtualization support in the BIOS (Basic Input / Output System) during startup. Netbooks, older machines, and cheap low-end computers will likely give you problems. If you are in doubt about meeting this requirement, search for your model of computer followed by "virtualization" and you should find the answers. The rest of this section will assume that your computer meets this requirement.

Ubuntu Linux

This may be quite controversial to some Linux enthusiasts, but I recommend Ubuntu as an ideal Linux operating system for our OSINT machine. While I use a Debian build for my own personal use due to the privacy and security of that operating system, Ubuntu is more appropriate for a wider audience. Debian can present complications when trying to display the dock or enable sudo

features. If you are comfortable with Debian and insist on using it versus Ubuntu, go for it. For those who want an easier option, I will only explain the process using Ubuntu.

You may have heard of Ubuntu Linux. It is one of the most popular Linux distributions, and it is based on Debian as its backbone. Ubuntu is the product of a corporation called Canonical. Some criticize Ubuntu for including Amazon software in the default downloads, which provides affiliate funding to Ubuntu when you make an online purchase, but we will eliminate it once we are finished. The following could be replicated with other "flavors" of Linux if desired, with minimal differences within installation menus. As new versions of Ubuntu are released, you may see minor changes to the following tutorials. However, the general functions should remain the same. I urge readers to replicate these instructions with Ubuntu 18.04.

First, we need to download an ISO file containing the installation files for Ubuntu. The official website is very user-friendly, and can be reached at ubuntu.com/download/desktop. This presents a Long-Term Support (LTS) option, which is currently 18.04. Clicking the "Download" button should prompt you to download the ISO file required to install Ubuntu. This will be a large file with an extension of iso. During this writing, my file was titled "ubuntu-18.04.03-desktop-amd64.iso". This file behaves similar to a physical CD which would install Windows or any other operating system. The 64-bit version should apply to most readers and their computers. If you know you need a 32-bit version for an older computer, you will find it in the Alternative Downloads section of the site. Save the file to your Desktop or anywhere else to which you have easy access. Expect an approximate file size of 2GB. Next, open VirtualBox and click on the button labeled "New". The following steps should create a new VM appropriate for our needs.

- Provide a name of "OSINT Master".
- Choose your desired location to save the machine on your host (I chose my Documents).
- Select "Linux" as type, "Debian 64-bit" as version, and click "Continue" (or "Next").
- In the Memory size window, move the slider to select 50% of your system memory.
- Click "Continue" and then "Create".
- Leave the hard disk file type as "VDI" and click "Continue" (or "Next").
- Select the default option of "Dynamically allocated" and click "Continue" (or "Next").
- Choose the desired size of your virtual hard drive. If you have a large internal drive, 20GB should be sufficient. If you are limited, you may need to decrease that number.
- Click Create.

Your VM has been created, but it will do nothing upon launch. We need to tell it to boot from the ISO file which we downloaded previously. Select your new machine in the menu to the left and complete the following steps.

- Click the Settings icon.
- Click the Storage icon.
- Click the CD icon which displays "Empty" in the left menu.

- Click the small blue circle to the far right in the "Optical Drive" option.
- Select "Choose Virtual Optical Disk File".
- Select the Ubuntu ISO previously downloaded.
- Click "OK" and then "Start" in the main menu.

Your Ubuntu installation process should now start within a new window. You should be booting to the ISO file downloaded previously, which is behaving as if you had placed an Ubuntu install CD into the virtual computer. This is your first virtual machine running on top of your host operating system. We can now finish the installation with the following steps within the VirtualBox window of your Ubuntu installation.

- Select "Install Ubuntu".
- Select your desired language and location, then click "Continue".
- Select "Normal Installation", "Download Updates", and "Install third party…".
- Click "Continue".
- Select "Erase disk and install Ubuntu", then "Install Now". Confirm with "Continue".
- Choose your desired time zone and click "Continue".
- Enter a name, username, computer name, and password of "osint" for each. This is required as part of these tutorials. Since this is a virtual machine inside a secure computer, minimal security is acceptable. Choose to log in to Ubuntu automatically.
- Allow Ubuntu to complete the installation, and choose "Restart Now", then press Enter. If prompted, allow the installation media to be removed.

Your device should now boot to the login screen. In my example, it booted directly to the Ubuntu Desktop. The following will finish the default configuration.

- Press "Next" twice, select "No" and then "Next" when asked to help improve Ubuntu.
- Click "Done" to remove the welcome screen.
- If prompted to install updates, click "Remind me later".

You now have a functioning virtual machine which contains the basic programs we need to use the internet. By default, it is using your host computer's internet connection, and taking advantage of your host's VPN if you have it connected. Technically, we could start using this machine right away, but the experience would get frustrating. We need to take some additional steps to configure the device for optimum usage. The first step should be to install VirtualBox's Guest Additions software. This will allow us to take advantage of better screen resolution and other conveniences. Conduct the following steps.

- In the VirtualBox Menu, select Devices > "Insert Guest Additions CD Image".
- Click "Run" when the dialogue box pops up, provide your password when prompted.
- Allow the process to complete and restart the VM.

You should now have VirtualBox Guest Additions installed. You can test this by resizing the screen. If you make the Ubuntu VM full screen, you should see the overall screen resolution change with it. If this appears to be functioning, you can right-click the CD icon on the desktop and choose "Eject". If not, double-click the CD icon and choose "Run Software" in the upper right corner to repeat the process. Next, we should make some modifications within the VirtualBox program in order to experience better functionality. Shut down the Ubuntu VM by clicking on the down arrow in the upper right and choosing the power button, followed by "Shut down". In VirtualBox, select your Ubuntu VM and click the "Settings" icon. Next, conduct the following steps.

- In the "General" icon, click on the "Advanced" tab.
- Change "Shared clipboard" and "Drag n' Drop" to "Bidirectional".
- In the "Display" icon, change the Video Memory to the maximum.
- In the "Shared Folders" icon, click the green "+".
- Click the dropdown menu under "Folder Path" and select "Other".
- Choose a desired folder on your host to share data back and forth.
- Select the "Auto-mount" option and then "OK".
- Click "OK" to close the settings window.
- Restart your Ubuntu VM.
- Click the nine dots in the lower-left to open the "Applications" menu. Search "Terminal" and open the application. In it, type `sudo adduser osint vboxsf` and press enter.

You should now have a more robust display, copy and paste capabilities, and a new shared folder connecting your VM to your host on the Desktop. You can copy files from your VM directly to your host and vice versa. This has improved a lot of function, and now it is time to personalize the machine. Personally, I do not like the default colors and wallpaper of Ubuntu, and prefer something more professional. I conducted the following on my new VM.

- Right-click on the Desktop and select "Change Background".
- Click the "Background" image and then "Colors".
- Choose a solid color of preference.
- Repeat the process with the "Lock Screen" image.
- Back in the "Settings" menu, click "Notifications" and disable both options.
- Click the "Privacy" option, then click "Screen Lock" and disable the option.
- Close the window and click "Usage and History", then disable the option.
- Close the window and click "Problem Reporting", then disable the option.
- Close the window and click Power, changing "Blank Screen" to "Never".
- Click "Automatic Suspend" and disable the feature.
- Close all open windows.
- Right-click any undesired icons from the dock and select "Remove from Favorites".

These changes should create a more private and pleasing environment. Even if you removed the Amazon icon from the Dock, the program is still installed. Conduct the following to remove all traces of Amazon and the affiliate program attached.

- Click the nine small dots in the lower left to launch the Applications menu.
- Type Terminal in the search field and click on the application.
- Type `sudo apt purge ubuntu-web-launchers` and click "Y" when prompted.

It is important to keep the software on this master VM updated. There are different ways to do this, but I will focus on the easiest way within the operating system applications. While we do this, it may be a good time to add some commonly used applications to our Dock. Conduct the following steps.

- Click the nine dots to launch the Applications option.
- Type Terminal into the search field.
- Right-click on the application and select "Add to Favorites".
- Type Software into the search field and right-click on "Software Updater".
- Select "Add to Favorites".
- Press escape until all windows are gone.
- Launch the Software Updater icon from the Dock.
- Click "Install Now" and allow the updates to complete.

You now have the Terminal and Software Updater in your Dock for easy access. Check for updates weekly and keep your master copy ready for usage. This brings us to a conversation about the term "Master". Ideally, you will keep a copy of this VM clean and free of any internet usage or contamination. There are two ways to achieve this, and both have unique benefits. First, let's discuss Snapshots.

Snapshots

A great feature of virtual machines is the use of Snapshots. These "frozen" moments in time allow you to revert to an original configuration or preserve an optimal setup. Most users install the virtual machine as detailed previously, and then immediately create a snapshot of the unused environment. When your virtual machine eventually becomes contaminated with remnants of other investigations, or you accidentally remove or break a feature, you can simply revert to the previously created snapshot and eliminate the need to ever reinstall. Consider how you might use snapshots, as detailed in the following pages.

Upon creation of a new Ubuntu virtual machine, apply all updates as previously mentioned. Completely shut down the machine and open the Snapshots option within your virtual machine software. Create a new snapshot and title it "Master". Use this machine for a single investigation, and export all evidence to an external USB device, such as a flash drive. You can then "restore"

the Master snapshot, and it overwrites any changes made during the previous investigation. Upon reboot, all history and evidence is eliminated. This ensures that you never contaminate one investigation with another. When there are substantial updates available for Ubuntu, you can load the default configuration, and apply all updates. You can then shut the machine down completely and delete the Master snapshot, without saving it, and create a new snapshot titled Master. This new snapshot possesses all of the updates. If using this technique, I usually delete and create a new snapshot weekly. The use of snapshots is very similar between VirtualBox and VMWare, but let's take a look at the minor differences.

VirtualBox use of Snapshots

- Completely shut down the Virtual Machine.
- In the VirtualBox Menu, click on the Snapshots button in the upper right.
- Click on the blue camera icon to "take a snapshot".
- Create a name to remind you of the state of the machine, such as "New Install".
- Click OK.

You can now use your virtual machine as normal. If you ever want to revert to the exact state of the machine that existed at the time of the snapshot, follow these instructions:

- Completely shut down the Virtual Machine.
- In the VirtualBox Menu, click on the Snapshots button in the upper right.
- Select the desired snapshot to apply.
- Click on the blue camera icon with arrow to "restore snapshot".
- Deny the option to save the current data, and click Restore.

Optionally, if you ever want to remove a snapshot, simply use the icon with a red X. This will remove data files to eliminate wasted space, but you cannot restore to that image once removed. It will not impact the current machine state. Many users remove old, redundant snapshots after creating newer clean machines.

VMWare Use of Snapshots

- Completely shut down the Virtual Machine.
- In the VMWare Menu, click on the Snapshots button in the upper right.
- Click on the camera icon to "take" a snapshot.
- Create a name to remind you of the state of the machine, such as "New Install".
- Click Take.

You can now use your virtual machine as normal. If you ever want to revert to the exact state of the machine that existed at the time of the snapshot, follow these instructions:

- Completely shut down the Virtual Machine.
- In the VMWare Menu, click on the Snapshots button in the upper right.
- Select the desired snapshot to apply.
- Click on the camera icon with arrow to "restore" a snapshot.
- Click Restore.

Optionally, if you ever want to remove a snapshot, simply use the "delete" icon. Today, I rarely use snapshots, as I believe they are prone to user error. I much prefer Cloned machines. These require more disk space, but provide a greater level of usability between investigations. Examples are explained in greater detail in a later chapter.

VM Exports and Clones

If you ever want to preserve a specific state of Ubuntu, you can export an entire session. This may be important if you are preserving your work environment for court purposes. When I am conducting an investigation that may go to trial, or discovery of evidence will be required, I make an exact copy of the operating system used during the investigation. At the end of my work, I shut down the machine. I click on File and then Export within my virtual machine software and create a copy of the entire operating system exactly as it appeared at shutdown. This file can be imported later and examined. After a successful export, I restore my clean Master snapshot and I am ready for the next case. The exported file is added to my digital evidence on an external drive. I now know that I can defend any scrutiny by recreating the exact environment during the original examination.

As stated previously, I prefer Clones over Snapshots. I create an exact replica of my master VM for every investigation, and never use Snapshots within these unique VMs. For clarity, consider my routine for every OSINT investigation I conduct, which takes advantage of the "Clone" option within VirtualBox.

- Launch the Master VM weekly to apply updates or global changes, then close the VM.
- In the VirtualBox menu, right-click on the Master VM and select "Clone".
- Create a new name such as Case #19-87445.
- Click "Continue" (or "Next") then "Clone".

This creates an identical copy of the VM ready for your investigation. You have no worries of contaminating your master VM. You can keep this clone available for as long as needed while you continue to work within several investigations. You can export the clone in order to preserve evidence, or delete it when finished. Neither action touches the master copy. I commonly possess over a dozen cloned VMs, each attached to a specific investigation. It is similar to possessing a new computer for every case and having all of them at your disposal whenever you want to jump back into an investigation.

Errors

I wish I could say that every reader will be able to easily build virtual machines on any computer. This is simply not the case. While most computers are capable of virtual machine usage, many demand slight modifications in order to allow virtualization. Let's take a look at the most common errors presented by VirtualBox upon launch of a VM.

VT-x is disabled: Any version of this error is the most common reason your VM's will not start. This indicates that the processor of your computer either does not support virtualization or the feature is not enabled. The fix for this varies by brand of machine and processor. Immediately after the computer is turned on, before the operating system starts, enter the BIOS of the machine. This is usually accomplished by pressing delete, F2, F10, or another designated key right away until a BIOS menu appears. Once in the BIOS, you can navigate through the menu via keyboard. With many Intel processors, you can open the "Advanced" tab and set the "Virtualization (VT-x)" to "Enable". For AMD processors, open the "M.I.T." tab, "Advanced Frequency" Settings, "Advanced Core" settings, and the set the "SVM Mode" to "Enable". If none of these options appear, conduct an online search of the model of your computer followed by "virtualization" for instructions.

VT-x is not available: This is usually isolated to Windows 10 machines. Navigate to the Windows Control Panel and open "Programs and Features". Click "Turn Windows features on or off" and uncheck all "Hyper-V" features. Click "OK" and reboot. If the Hyper-V option is not enabled, enable Hyper-V, restart the computer, disable Hyper-V, and reboot again. Attempt to start your VM with these new settings. This may seem backwards, but it makes sense. Previous versions of VirtualBox cannot run if you are using "Hyper-V" in Windows. Basically, both systems try to get exclusive access to the virtualization capabilities of the processor. Hyper-V within Windows receives the access first and impedes VirtualBox from the capabilities. The latest version of VirtualBox attempts to correct this. If the previous setting did not help, try to re-enable all of the Hyper-V options within Windows, reboot, and try to boot your VM again.

If you are still experiencing problems, read the troubleshooting chapter of the VirtualBox manual at virtualbox.org/manual/ch12.html. Expand any errors received and search the provided error codes to identify further solutions.

Your master VM should only be used to install new software and apply updates. Throughout the next several chapters, I will reveal more about my usage protocols. Hopefully, you now have either VirtualBox or VMWare installed and an Ubuntu installation created as a virtual machine. You have chosen to either use Snapshots or Clones as part of your investigation (I prefer Clones). Now it is time to play with the many applications available for Linux. We should start with the most important application we have - web browsers.

CHAPTER THREE
WEB BROWSERS

If you are a Windows user, your default web browser is either Internet Explorer or Microsoft Edge. Apple users are presented Safari by default. I believe OSINT professionals should avoid these at all costs. All are inferior in my opinion, and you will encounter difficulties with some of the websites and services mentioned later. Therefore, we need a better browser in Linux.

Firefox (mozilla.org)

The most vital application in this chapter is the Firefox web browser. Most of the search methods that you will learn throughout this book must be conducted within a web browser. The Firefox browser has enhanced security and a feature called "add-ons" or "extensions". These are small applications which work within the browser that perform a specific function. They will make searching and documentation much easier. I also use the Chrome web browser when necessary, and will explain some customizations later. However, many of the extensions that I need are only compatible with Firefox. The following instructions apply to any recent version of Firefox, including builds for Windows, Mac, and Linux.

Downloading and installing Firefox is no different than any other application. Detailed directions are readily available on their website. The browser will not look much different from the browser you were previously using. When installing and executing, choose not to import any settings from other browsers. This will keep your browser clean from unwanted data. The next step is to ensure your browser is up-to-date. You can check your version of Firefox by clicking on the Menu button in the upper right (three horizontal lines), then the Help button (?), and finally the option labeled About Firefox. This will open a new window that will display the version of Firefox you are running, or a warning that the version you have is out-of-date.

For the purposes of this chapter, I will assume you are configuring the Firefox application included in your new Ubuntu VM. However, all of the methods explained here could be replicated within your host or any other computer. If you will be using your host computer for any web browsing, Firefox is highly recommended as the default browser. Regardless of where you will be conducting your online investigations, have a properly configured Firefox application.

Before identifying Firefox resources which will aid in our OSINT research, we must first secure our browser to the best of our ability. While the default Firefox installation is much more private and secure than most other browsers, we should still consider some modifications. I personally use Firefox for all of my OSINT investigations in my VMs, and as my default web browser on my personal laptop. I no longer possess multiple browsers for various tasks. I believe that Firefox is the most robust, secure, and appropriate option for almost any scenario. However, I recommend changing the following settings within Firefox.

- Click on the menu in the upper right and select "Options" or "Preferences".
- In the "General" options, uncheck "Recommend extensions as you browse" and "Recommend features as you browse". This prevents some internet usage information from being sent to Firefox.
- In the "Home" options, change "Homepage and new windows" and "New tabs" to "Blank page". This prevents Firefox from loading their services in new pages and tabs.
- In the Privacy & Security options, select the "Strict" option under "Content Blocking".
- Check the box titled "Delete cookies and site data when Firefox is closed".
- Uncheck the box titled "Ask to save logins and passwords for websites".
- Change the History setting to "Firefox will use custom settings for history".
- Uncheck the boxes "Remember browsing and download history" and "Remember search and form history".
- Check the box titled "Clear history when Firefox closes". Do not check the box titled "Always use private browsing mode", as this will break Firefox Containers.
- Uncheck "Browsing history" from the "Address Bar" menu.
- In the "Permissions" menu, click "Settings" next to Location, Camera, Microphone, and Notifications. Check the box titled "Block new requests..." for each of these options.
- Uncheck all options under "Firefox Data Collection and Use".
- Uncheck all options under "Deceptive Content and Dangerous Software Protection". This will prevent Firefox from sharing potential malicious site visits with third-party services. This leaves you more exposed to undesired software attacks, but protects your internet history.

Firefox allows users to modify many configuration settings, and some of these deal with privacy and security concerns. Though some of these changes can be made in the menu of Firefox's preferences, changes made through about:config tend to be more durable and granular. To access the list of configuration settings, open Firefox and type "about:config" into the Uniform Resource Locator (URL) field. This is the place where you would traditionally type the website you wish to visit. The terms URL, web address, and website will be used interchangeably throughout the book.

You will receive a warning about making changes within this area, but the modifications we make will be safe. Choose to accept the risks. Some of these about:config settings may already be on the correct setting, but most probably will not. To change most of these settings you can simply double-click the setting to toggle it between "True" and "False". Some may require additional input, such as a number. Because the list of about:config settings contains hundreds of entries, you should search for all of these through the search bar in the about:config interface.

- **geo.enabled**: **FALSE**: This disables Firefox from sharing your location.
- **browser.safebrowsing.phishing.enabled**: **FALSE**: This setting disables Google's "Safe Browsing" and phishing protection. If this setting is "true" Google will be able to

scan (and store) the sites that you visit for the presence of malware.

- **browser.safebrowsing.malware.enabled**: **FALSE**: Again, this disables Google's ability to monitor your web traffic for malware, storing the sites you visit.
- **media.navigator.enabled**: **FALSE**: Website operators will identify your computer as unique to enable tracking around the web. One such tactic is to track the status of your webcam and microphone (ON/OFF). This disables the ability to website operators to see this information.
- **dom.battery.enabled**: **FALSE**: Another technique used by website operators to track you is to view your exact battery levels. This setting blocks this information.
- **extensions.pocket.enabled**: **FALSE**: This disables the proprietary Pocket service.

WebRTC: These settings address a potential vulnerability of leaked IP addresses.

- **media.peerconnection.enabled: FALSE**
- **media.peerconnection.turn.disable: TRUE**
- **media.peerconnection.use_document_iceservers: FALSE**
- **media.peerconnection.video.enabled: FALSE**

It is not vital that all of these security settings be applied to your systems. Firefox natively respects your privacy and security more than other browsers. These recommendations are for those that want to tweak additional settings that may provide a new layer of protection, even if minimal. Next, I will discuss the biggest benefit of Firefox, which is the abundance of helpful browser extensions called add-ons.

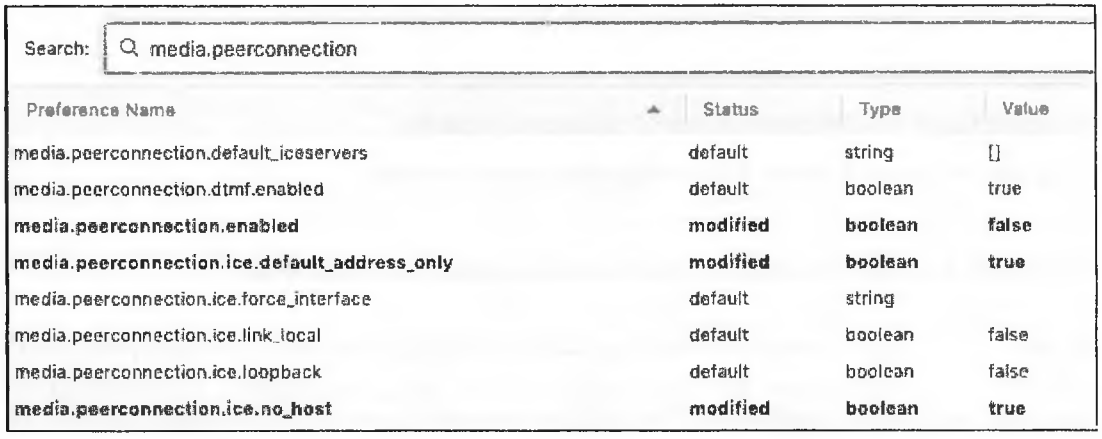


Figure 3.01: An "about:config" menu with customizations.

Firefox Add-ons (Extensions)

There are thousands of extensions available for Firefox. Some are helpful, some are worthless, and some are fun. This chapter will discuss several of them. The Firefox add-ons, sometimes called extensions, detailed here will include a website for each option. You can either visit the website and download the add-on, or search for it from within Firefox. The latter is usually the easiest way. While Firefox is open, click on the menu in the upper right and then "Add-ons". This will present a page with a search field in the upper right corner. Enter the name of the extension and install from there. The following are my recommendations, in order of importance.

Firefox Containers: Isolate specific sites within tabs which do not see settings from other sites.

uBlock Origin: Block undesired scripts from loading.

DownThemAll: Download bulk media automatically.

Bulk Media Downloader: Download bulk media automatically.

VideoDownloadHelper: Download media from a page with a click of a button.

FireShot: Generate screenshots of partial and entire web pages.

Nimbus: Alternative screen capture for large web pages.

HTTPS Everywhere: Ensure that you are accessing sites through a secure connection.

Exif Viewer: Identify Metadata embedded inside a photograph.

MJSONViewer: View API JSON and XML results properly in a browser.

User Agent Switcher: Emulate various browsers and devices.

Image Search Options: Conduct automatic reverse image searches.

Resurrect Pages: Enable historical search on deleted websites.

Copy Selected Links: Quickly copy all hyperlinks from a website.

OneTab: Collapse or expand tabs into a single resource.

KeePassXC Browser: Automatically enter stored usernames and passwords.

The following pages provide explicit instructions for installing and configuring each of these add-ons. At the end, I will explain how you can export your settings and replicate your work across practically any Firefox installation. This will preserve your work and allow you to receive an identical experience if conducting investigations across multiple computers. This will also benefit your virtual machines. Ideally, you would complete all browser configurations within your master VM before cloning, exporting, or use of snapshots.

Firefox Multi-Account Containers (addons.mozilla.org/addon/multi-account-containers/)

The first Firefox Add-on which I use daily is the Multi-Account Containers option from Mozilla. Multi-Account Containers allows you to separate your various types of browsing without needing to clear your history, log in and out, or use multiple browsers. These container tabs are like normal tabs, except the sites you visit will have access to a separate slice of the browser's storage. This means your site preferences, logged-in sessions, and advertising tracking data will not carry over to the new container. Likewise, any browsing you do within the new container will not affect your logged in sessions, or tracking data of your other containers. Below is an example.

On my personal laptop, I have a container tab open which I use to log in to my email provider. I have my inbox open in this tab. I want to order a product from Amazon, but I do not want them to see any cookies stored by my email provider. I also want to conduct a Google search, but do not want Google to see any data present from my Amazon search. I simply open a unique container tab for each of these events. Each sees the session as unique, and no data is shared from one service to another.

OSINT investigators can use this technique in many ways. With a traditional browser, you can only be logged in to one instance of a social network. If you are logged in to a covert Facebook account, then open a new tab and navigate to Facebook, you will be presented with the same logged-in account used in the previous tab. With containers, we can isolate this activity. You can log in to one Facebook account in one container, another Facebook account in a second container, and any additional accounts in their own containers. This applies to any service, such as Twitter, Reddit, or others. This allows us to simultaneously access multiple accounts within the same service without logging out or opening a different browser. This is a substantial update since the last book. Let's configure it for optimal use.

Once installed, you will see a new icon in the upper right in your Firefox browser which appears as three squares and a "+" character. Click on it and select the container you want to open. Default options include choices such as Personal and Shopping, but you can modify these any way you desire. You can create, delete, and edit containers from the main menu. When you click the "Edit Containers" or the "+" buttons, you can change the color or icon associated with a container or change the container name. The following tutorial replicates my configuration for OSINT investigations.

- Open the Multi-Account Containers menu and click the "Edit Containers" option.
- Delete all containers by clicking the trash can next to each.
- Click "Exit Edit Mode" to return to the menu.
- Click the + in the lower right.
- Enter the name of your new container, such as "Alias 01".
- Choose a desired color and icon.
- Repeat this process to create the number of containers desired.

On my machine, I have the following containers, which can be seen in Figure 3.02.

Alias 01	Alias 03	Leaks	Harmful
Alias 02	Alias 04	Google	

You can now either open a new container as a blank page or open links in a new or different container. The following are a few of my usage examples.

Multiple Logins: While in Firefox, I want to open Facebook inside a unique container. I click on the containers menu and select Alias 01. This opens a new blank tab within this container. I navigate to Facebook and log in to an alias account. I then want to log in to a second Facebook account, so I click on the containers menu and select Alias 02. This opens a new tab in that container. I then navigate to Facebook and receive a login prompt. I log in to my second account and can switch back and forth between tabs. You should note that Facebook can see you have the same IP address for each login, but they cannot see your cookies from one session to the other. You could replicate this process for any other social network or service. You could also have numerous Gmail accounts open within one browser.

Safety: While I am viewing my target's Twitter profile, I see a link to an external website from his page. I am logged in to a Twitter account within this container and I do not know what this linked website will try to load. I do not want to jeopardize my investigation. I right-click on the link and choose "Open link in New Container", and then select the desired container tab, such as "Harmful". The new tab will open within this container created for questionable websites. The page I open cannot see any cookies stored within my container associated with my Twitter login.

Dedicated Container: I assign specific websites to a container so they will always open in a that container. I use this for Google because I do not want my search history associated with other investigation activities. If I ever visit Google as part of an investigation, the site will open in a new container tab which I designated "Google". This is regardless of which tab I try to use. The following steps configure this option.

- Create a containers tab titled "Google".
- Click on the containers menu and open a new Google tab.
- Connect to google.com.
- Click the containers menu and select "Always open in Google".
- Close this tab and navigate to google.com from a standard tab.
- When prompted, select "Remember my decision…" and then "Open in…".

When complete, you have created a rule within Firefox. Any time you connect to google.com, regardless of the container you are in, or if you have Google set as your default search from within the URL field, Firefox will open a new "Google" tab to complete your connection. This isolates your Google traffic from any other tab, and applies to any Google sites, such as Google

Voice, Gmail, etc. If you regret making this type of rule, you can either delete the entire container or just the policy for that site. In this example, I can go to the Containers menu; click the Edit Containers option; then click the edit icon next to the Google container; then click the trash icon next to the websites I wish to remove.

When I first installed this add-on, I went a bit too far with customized containers. I wanted all Facebook pages to load in their own container, which prevented the ability to log in to multiple accounts. I removed this option and established the rule mentioned previously which allowed me to have multiple logins, but lost the isolation from Facebook to other websites. I created containers for most of the sites I visited, which was overkill. There is no perfect solution. Evaluate your needs and create the most appropriate set of containers vital to your investigation.

On my personal laptop, my containers are focused on privacy and isolate invasive services such as Amazon, Google, and online payment services. I also isolate financial websites and personal email tabs. I highly recommend applying these same strategies to your personal devices. I possess the following containers on my personal laptop.

- Communications: Personal email and calendar accounts
- Financial: Banking and credit card accounts
- Search: All Google queries
- Alias: Any social network accounts in another name

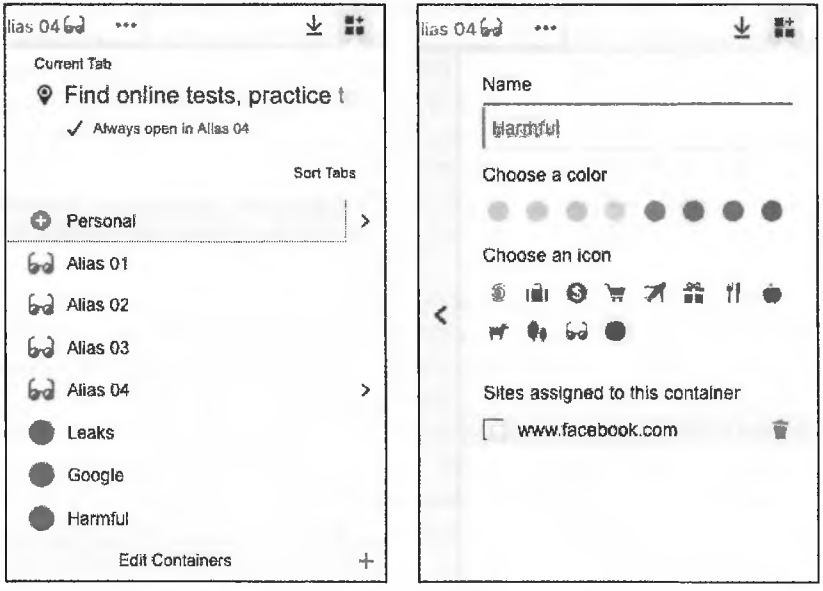


Figure 3.02: The Firefox Multi-Account Containers menus.

uBlock Origin (addons.mozilla.org/firefox/addon/ublock-origin)

In previous editions of this book, I recommended NoScript as my choice of script blocker. I no longer use NoScript at all. During the transition to Firefox 57, NoScript changed drastically. It became much more convenient to use, at a cost of functionality. We were no longer given granular control of the data that is passed through our browser, and lost several features required for private and secure browsing. I also previously recommended Adblock Plus and Disconnect as privacy add-ons that would help stop unwanted ads, tracking, and analytics. These are no longer present on my systems. I now only use uBlock Origin, as it replaces all three of the previous options. This section may seem a bit overwhelming, but experimenting with the advanced settings should help you understand the functionality. Let's start with the basics.

Install uBlock Origin from the Firefox add-ons page or directly by navigating to the application's website at https://addons.mozilla.org/en-US/firefox/addon/ublock-origin/. You are now protected on a basic level. By default, most known invasive advertisements, tracking code, and malicious content is blocked. This step alone would provide much needed protection from the internet. However, we can take it a step further.

Click on the uBlock Origin icon in the menu and select the "Dashboard" icon to the right, which appears as a settings option. This will open a new tab with the program's configuration page. On the Settings tab, click the option of "I am an advanced user". This will present an expanded menu from the uBlock Origin icon from now forward. Click on the "Filters" tab and consider enabling additional data sets that will protect your computer. I select all options within the Ads, Privacy, Malware Domains, and Annoyances categories. After you have made your selection, click the Update Now button at the top of the page. This will refresh all of the data and apply your new settings. You now have extended protection that will be applied to all visited websites without any interaction from you. When you encounter a web page with a lot of advertisements, such as a news media website, it should load much faster. It will block many of the pop-ups and auto-play media that can be quite annoying when conducting research. This protection will suffice for most users, but dedicated OSINT analysts may choose to take a more advanced approach.

After you have enabled the Advanced settings as explained above, clicking on the uBlock Origin icon should now present an expanded menu that will change as you visit different sites. In order to explain the function of this menu, I will conduct a demonstration using the website cnn.com. Figure 3.03 displays the default view. While this book is printed in black and white, your view will be in color, and likely all options will appear grey. Scrolling down this list of scripts that have either been loaded or blocked, you can see several questionable scripts such as Facebook, Sharethrough, and Turner. These scripts allow tracking across multiple websites and are the technology responsible for monitoring your interests, web history, and shopping habits.

This menu is split into three columns. The first simply identifies the type of code or domain name of the script. The second column is global settings. Anything changed here will apply to all website visits. The third column contains settings for the current website. A single plus sign (+)

indicates that less than ten scripts were allowed from that specific option. Two plus signs indicates that between ten and one hundred scripts were allowed. The single minus sign (-) indicates that between one and nine scripts were blocked from that domain, while the dual minus signs tell us that ten to one hundred scripts were blocked. In Figure 3.03, we know that over ten scripts were allowed to run from cdn.cnn.com, and at least one script was blocked by smetrics.cnn.com. This is all default behavior and provides a balance of functionality and security. uBlock Origin decides which content should be allowed and which should be blocked.

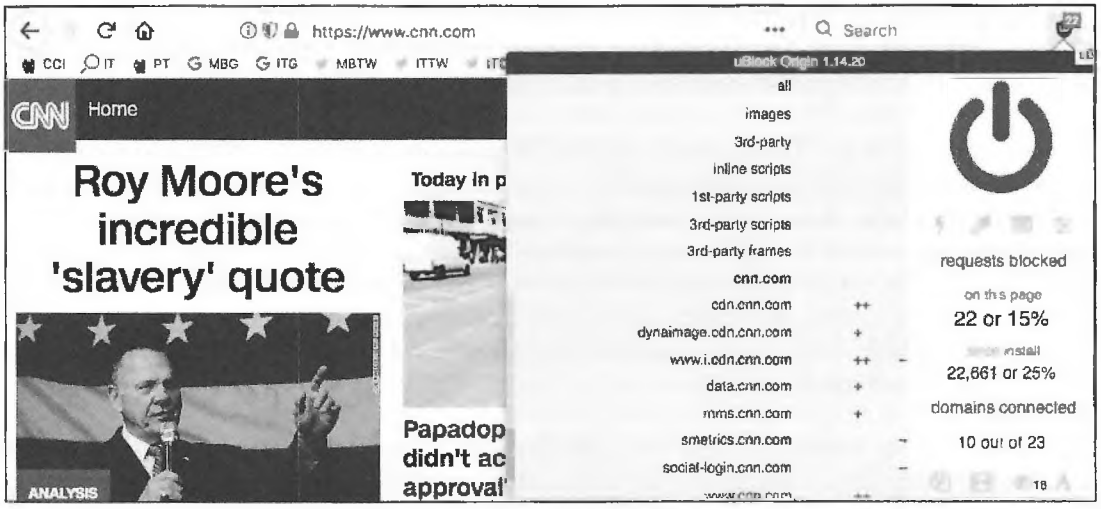


Figure 3.03: An advanced view of uBlock Origin.

Using this same page, let's modify the options. In Figure 3.04 (left), I have clicked on the far-right portion of the first cell in the third column. This turned the entire third column red in color. This action activated an option to refresh the page (middle arrows) and an option to save the change (upper left "padlock"). Clicking the padlock and then refreshing the page presented me with the example in Figure 3.04 (right). Since I blocked every script, the page would not fully execute. It could not load images, design scripts, or any JavaScript. This is not useful at all, so I disabled my actions by clicking on the middle section of the top cell in the third column, which turned the entire column back to grey in color. Saving these changes and refreshing the page brought me back to the example in Figure 3.03.

We can also take this to the opposite extreme. In Figure 3.05 (left), I clicked on the far-left portion of the top cell in the third column. This turned the entire column green in color, and allowed all scripts to load on cnn.com. This includes the dozens of intrusive scripts that could load advertisements on the page. You can also see that small plus signs confirm that scripts were allowed to run while the minus signs in Figure 3.04 (right) state the opposite. For most users, this allowance would seem irresponsible. However, there is a specific reason that we want the ability to allow all scripts. If you are collecting evidence, especially in criminal cases, you may

want to archive a page exactly as it was meant to be seen. When we block scripts, we are technically modifying the page (evidence). By intentionally allowing all scripts before the collection of the screen capture, we know that we are viewing the page in an unmodified format. This may be overkill for many investigators, but you should know your options.

Next, we will modify the second (middle) column, which will apply the settings globally. By default, all options are grey in color. This indicates that the default block list is applicable, and only invasive scripts will be blocked everywhere. I clicked on the far-right portion of the top cell in the second column. This turned the entire column red, and indicates that all scripts across all websites will be blocked. After I saved my changes, every website will only load the most basic text content. This will prohibit much of our research.

Loading a page such as a Twitter profile resulted in no usable content. By clicking on the uBlock Origin icon and clicking the middle sections of specific cells within the third column, I enabled those scripts without allowing everything on the page. While you cannot see the colors in Figure 3.05 (right), you can see the difference in shading. In this example, the entire second column is red. This indicates that all scripts are blocked globally. The third column is mostly red, but the options for twitter.com, twimg.com, and others are grey. Those scripts will be allowed, if approved by uBlock Origin's rules, only for that domain. If I load a blog that has scripts from Twitter, they would still be ignored.

These are extreme examples. Let's bring this back to some sanity. The following is how I recommend using uBlock Origin. Install, enable advanced options, and proceed with your work. When you arrive at a website that is blocking something you want to see, open the menu, and click on the far-left section of the top cell in the third column. That will allow everything to load on that page, and that page only. When you are about to navigate to a questionable site that may try to install malicious code on your machine, click on the far-right section of the top cell in the second column. That will block all scripts on all pages. Conduct your research and reverse the change when you are finished. Remember to click the save button (padlock) after each change and refresh the page.

Hopefully, you are practicing these settings and learning how this program functions. It is an amazing option that has protected me many times. If you are doing things right, you have likely completely messed-up your settings and are now blocking things you want while allowing things you do not. Don't worry, we can reverse all of our mistakes by first making the global (second column) settings back to grey (middle section of top cell). Next, return to the dashboard settings of the add-on, and click on the My Rules tab. In the second column (Temporary Rules), click Edit, highlight all of your customizations, and delete them. Click the Save button in this same column and then the Commit button to apply these settings everywhere.

The primary benefit of uBlock Origin over other options is the simple ability to block malicious scripts without customization, while having an option to allow or block any or all scripts at our disposal. This is a rarity in these types of add-ons. Another benefit is the ability to bypass website

restrictions, such as a news site blocking articles unless the visitor has a subscription service. Consider the following example with the Los Angeles Times. Visiting the page allows you to view three articles for free, but you must have a paid subscription in order to continue using the site. Figure 3.06 displays the results of my blocked access. If I click on the uBlock Origin menu while on this page, select the far-right (red) option on the right column under the setting for "3rd party scripts", then the lock icon, and reload the page. I am now allowed to see the article. An example of this is seen in Figure 3.07. This is because this website relies on a third-party script to identify whether a visitor is logged in to the service.

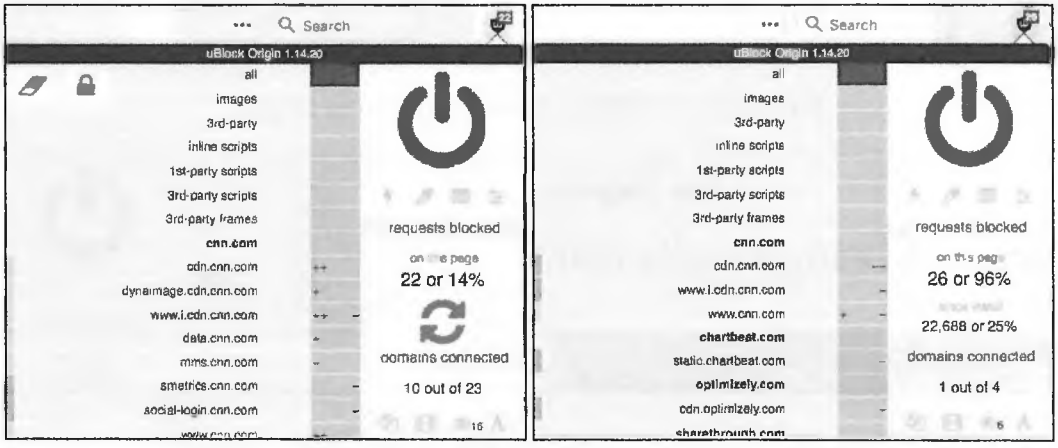


Figure 3.04: Disabled scripts within uBlock Origin.

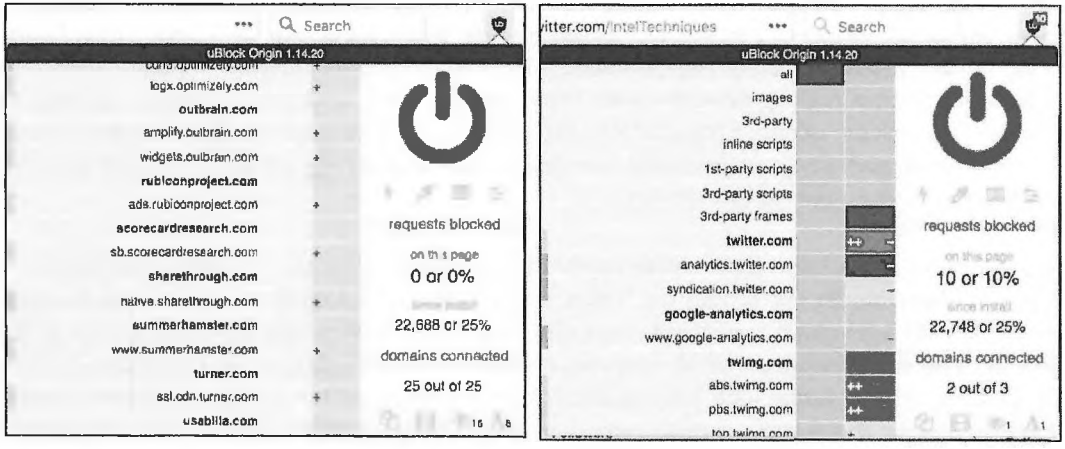


Figure 3.05: Fully and partially enabled scripts with uBlock Origin.



Figure 3.06: A website blocking access to an article.

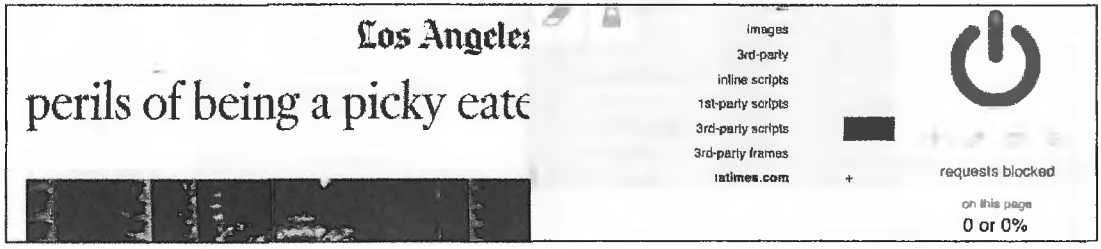


Figure 3.07: Bypassing a restriction with 3rd-party script blocking.

The final example of uBlock Origin within this chapter, which I rely on with my daily browsing, is the Inline Scripts blocker. For this demonstration, I will navigate to cnn.com. Clicking on almost any article presents a new page, including new annoyances. An undesired video begins playing while scrolling down the page, multiple images from unrelated articles populate over some of the article text, a pop-up advertisement interrupts my view, and over 56 scripts attempt to monitor your activity. uBlock Origin blocks the malicious scripts, but not all of the annoyances. We could block each script individually, but that is time consuming. Instead, consider a simple modification of the inline scripts setting.

Click on the uBlock Origin menu while on the desired page, select the far-right (red) option on the right column under the setting for "inline scripts", then the lock icon, and reload the page. The site should load much faster and block all of the inline scripts being pushed to you by the provider. You should notice that all pages on cnn.com load immediately, and without all of the undesired media interfering with your research. Clicking the grey area in this same box reverses the action. I apply this feature to practically every news website I visit. It blocks vital information if your desire is to obtain a screen capture as evidence, but provides a much more pleasing environment if you simply want to read the articles.

DownThemAll (addons.mozilla.org/firefox/addon/downthemall)

In the previous edition of this book, I complained about the loss of this extension when Firefox made a substantial change to their add-ons policies. Fortunately, we saw this staple return in late 2019. DownThemAll simplifies the process of extracting bulk data from a web page. It attempts to identify linked video, audio, images, documents, or any other type of media within a site. It then allows you to easily download everything at once. Consider the following example.

Archive.org possesses a copy of a website which once offered several gigabytes of marketing data containing millions of records on Americans, including full names, addresses, telephone numbers, and interests. The archive can be found at the following URL.

https://web.archive.org/web/20151110195654/http://www.updates4news.com:80/kyledata/

This URL presents hundreds of large CSV and TXT files. Later in this book, I discuss how to properly parse through this content and create your own searchable file with this example. For now, I simply need to download each file. While on the page, click on the DownThemAll toolbar menu and select "DownThemAll". In the new window, you should see all of the data links present on this site. Clicking the "All Files" box near the bottom selects each of the files. Clicking "Download" in the lower right begins the process of downloading all of the data from the page (which may take hours), and places each file in the default download location for your operating system. Please do not download this data set yet, as it will fill your disk space in your VM. We will discuss external storage methods later. Figure 3.08 displays the action window for this example. This add-on is a requirement for any browser I use. I find it to work better than the next option, but consider all alternatives for your own needs.

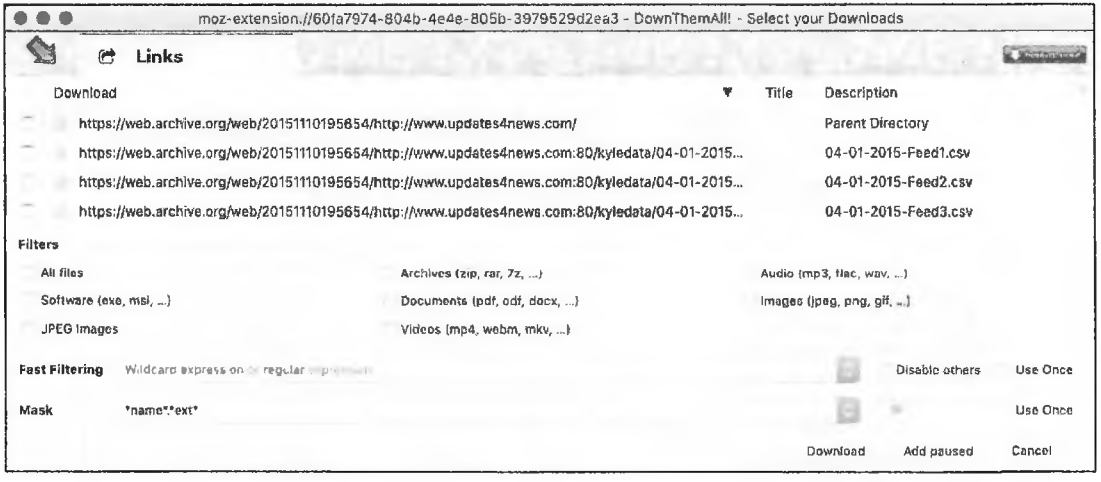


Figure 3.08: The DownThemAll download window.

Bulk Media Downloader (addons.mozilla.org/firefox/addon/bulk-media-downloader/)

Similar to DownThemAll, this add-on can make downloading a large amount of media files easy. It should serve as a backup in the event you find a page which DownThemAll will not function properly. If you locate a page of several audio or video files, it can be time consuming to save them all manually. Additionally, you run the risk of accidentally skipping a file. Bulk Media Downloader provides a solution. As an example, I navigated to Twitter and searched the word Video. This presented hundreds of embedded videos within a single page. I launched Bulk Media Downloader, which displayed a pop-up option over my browser. In this pop-up, I can select specific file types such as Video or Audio. I chose only the Video option and reloaded the Twitter page in the background. The Bulk Media Downloader tool began populating video links as I scrolled down the Twitter page. Figure 3.09 displays the result. Clicking the Download button retrieved all of the videos in MP4 format as seen in Figure 3.10. This utility works well on sites that have a large number of embedded audio or video files, as well as those that contain numerous documents. You can easily select or deselect entries individually, or select categories at the bottom that fit your needs.

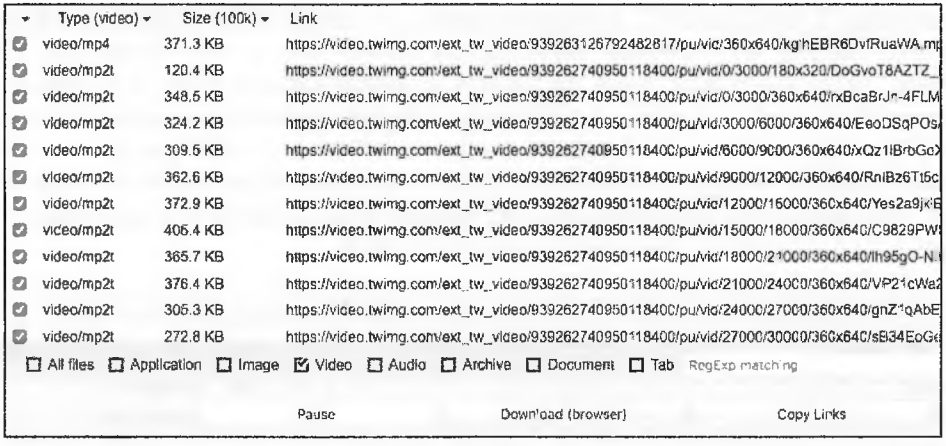


Figure 3.09: A Bulk Media Downloader window.

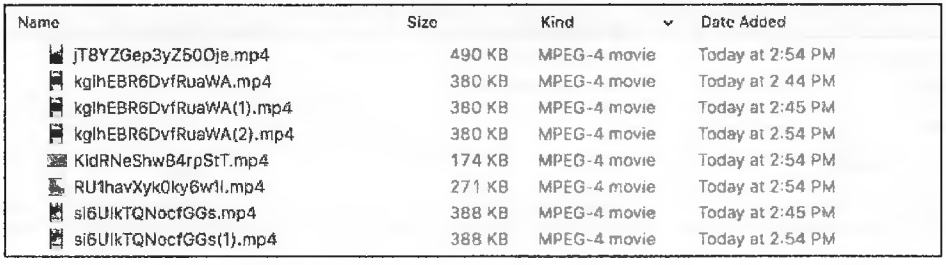


Figure 3.10: Files extracted from Twitter with Bulk Media Downloader.

Video Download Helper (downloadhelper.net)

This extension will assist with downloading media that is located during your search. It works well with videos such as those found on Vimeo and other hosts. It does not work well with YouTube, and we will use a better option in the next chapter. When this extension is enabled, an icon will appear within your browser that looks like three grey circles. Any time you open a website that includes media content, such as a video, these circles will turn to full color. This is an indication that the media on the page can be extracted. While this add-on will work immediately after installation, I have found specific configuration changes to be helpful to OSINT investigators.

- Click on the icon placed in your menu bar and select the icon for "Settings".
- Click the Behavior tab and change the Max concurrent downloads to 20.
- Change the Max Variants to 99.
- Select the Hide ADP Variants option and click "Save".

When downloading videos, especially from YouTube, the ADP format requires secondary conversion software to be installed. I do not like this option as it introduces unnecessary software to my machine. Furthermore, I never want to convert video evidence. I simply want to extract the options available directly from the source. Therefore, eliminating the ADP options from our view as explained above reduces the chance of downloading undesired content. In Figure 3.11 (left), the ADP options are present and would not be ideal download choices. In the example on the right, I have eliminated these choices and I am presented with more appropriate options.

You can now extract embedded media files from websites by clicking the icon and selecting the appropriate file. If your desired media is going to be used in court, I recommend downloading all sizes available. If you only want a personal archive, the largest size should be downloaded. You will now have a pure digital extraction of the target video. This is better than a screen capture or recording of the video because there is no loss of data or analog conversion. If downloading a large number of videos, consider the custom script that will be explained in the next chapter.

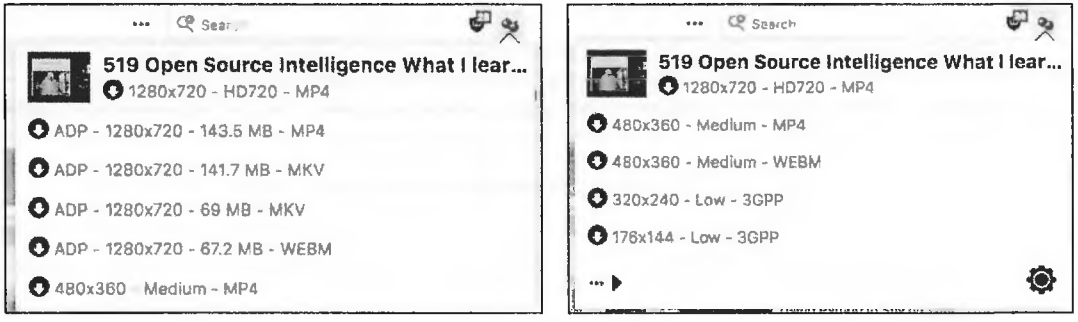


Figure 3.11: Menu options from Video Download Helper.

FireShot (addons.mozilla.org/en-us/firefox/addon/fireshot/)

Documenting and archiving your progress with an OSINT investigation is as important as the intelligence discovered. The general rule is that if you do not have proof of your findings, then they never existed. FireShot provides you with an easy solution to capturing all of your results. When enabled, this extension is a button in the upper right portion of your browser. It appears as a blue square containing the letter "S". Clicking the icon presents a menu with options. The best option is to select "Capture entire page" and then "Save to PDF". This will create a PDF document of the entire page exactly as it appears in your browser and then save it to anywhere you choose. The file can later be archived to a removable storage device. The title of the document will match the title of the web page and it will include the URL of the selected page.

This method is preferred over a standard screen capture for several reasons. A typical screen capture only captures the visible area and not the entire page. You must then open a program into which you "paste" the data and then save the file. The FireShot extension automates this and saves it in a format that is difficult to edit. This can be beneficial during testimony.

By accessing the "Options" area of the menu, you can assign customized naming features. Click "Show filename template settings" in the options page and change the default value to the following.

%n-%u-%t-%y-%m-%d-%H-%M-%S

This setting will change the default name of each page capture. Each file will be named a numerical value, followed by the website URL, followed by title, and followed by the date and time of capture. Changing the %n value to 0 and the Pad option to 3 will ensure that your captures always start with a numerical value of 0 and ascend chronologically. This can help determine the order of the evidence that you retrieved. Be sure to "Apply" and then "Save" after you have made your desired changes. Figure 3.12 displays a typical series of results. Notice that you can quickly see the order captured (first three digits), target website, description, and date & time.

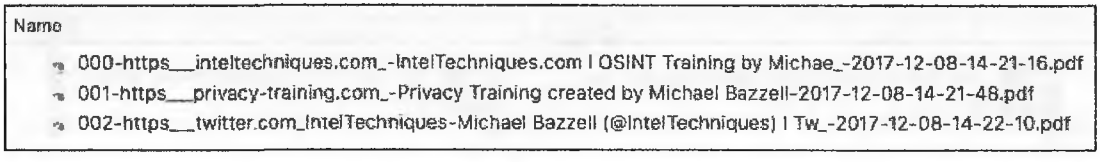


Figure 3.12: Results from FireShot screen captures.

Nimbus (addons.mozilla.org/en-US/firefox/addon/nimbus-screenshot)

While FireShot is my preferred screen capture utility within Firefox, there are some instances where it does not perform well. If you have a target's Facebook page that has a lot of activity present, this may create a screen capture too large for FireShot. The rendering process will likely expend all of the computer's video memory and fail to create the file. When this happens, I use Nimbus as my first backup. Nimbus allows you to specify whether you want to capture only the visible portion of the page, the entire page, or a custom selection from the page. The drop-down menu presents these choices and the result is saved as a PNG file. This is not optimal for online investigations, but is better than no capture at all. Another feature of Nimbus is the ability to manipulate captures. I believe that this is bad practice as we usually want to provide the most authentic and accurate evidence as possible. I do not want to manipulate any potential evidence. Therefore, I recommend the following configurations.

- Click on the Nimbus icon and choose the "gear" icon in the lower-right.
- In the Filename Template, insert {url}-{title}-{date}-{time}. This will name every capture with the URL and title of the target website along with date and time of capture.
- Check the Enable Quick Capture option and select the Entire Page option in the first row and Download option in the second row.

After these changes, clicking the Nimbus icon in the menu bar will no longer present a menu with options. Instead, it will automatically select the entire page, apply the proper file naming, and download the capture as a maximum quality PNG file to your Desktop. While a PDF file created with FireShot is the preferred file format, a PNG file has other advantages. The PNG file is more universal and does not require PDF viewing software such as Acrobat Reader. However, PNG files are easy to edit, and establishing the integrity of the file may be difficult. I believe that Nimbus should be used as a supplement to FireShot.

One common failure of both FireShot and Nimbus is the capture of extremely large Facebook and Twitter pages. While this is rare on computers that have ample resources such as processing power and RAM, it can be quite common on older machines with low specifications. Surprisingly, I have found FireShot to work better on large Twitter profiles and Nimbus to be best for large Facebook pages. I have no logic to offer for this discovery. Again, having both at our disposal will make us better prepared for online evidence collection. When both fail, consider Firefox's own solution on the next page.

Firefox Screenshot

You may not like either Fireshot or Nimbus. In general, you get what you pay for with these (they are free). When I have an extremely long Facebook or Twitter page, I find both of those options mediocre at best. Lately, I find myself using the embedded Firefox screenshot utility more than anything else. Consider the following example.

I loaded my own Twitter profile and scrolled back to posts from a year prior. This generated quite a long page and my computer fans increased speed due to the heat generated from my processor. I attempted a screen capture with both Fireshot and Nimbus, and each failed. I then executed the following with Firefox.

- Click on the three dots to the right of the address field.
- Click "Take a screenshot".
- Choose "Save full page".
- Click "Download".

The process ran for about three minutes and saved an image .png file to my default downloads directory. It was several megabytes in size. The default filename includes the word Screenshot, date, and title from the webpage. In this example, the file was titled as follows. Figure 3.13 displays a compressed and partial view of screenshots from three different sites.

Screenshot_2019-09-04 (1) Michael Bazzell (IntelTechniques) Twitter.png

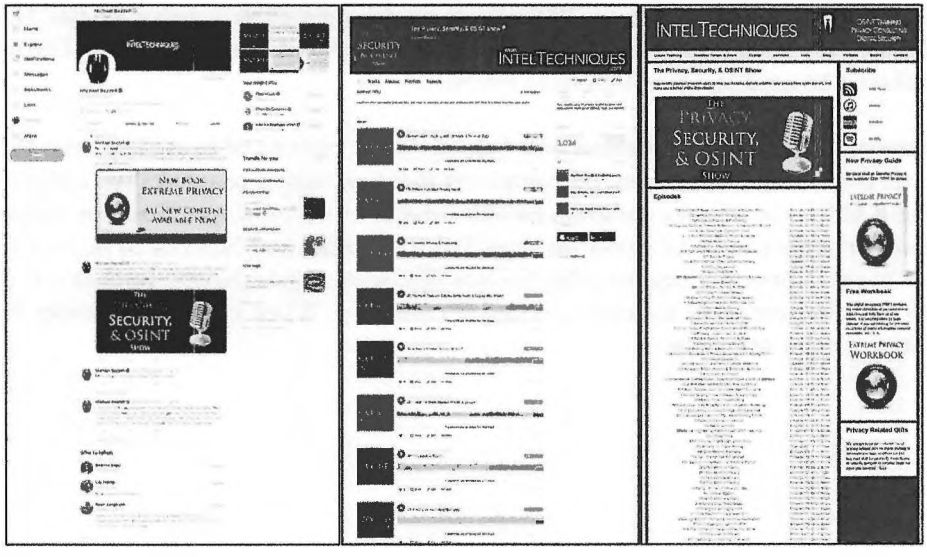


Figure 3.13: Partial screenshots from the embedded Firefox capture utility.

HTTPS Everywhere (addons.mozilla.org/en-us/firefox/addon/https-everywhere)

This extension encrypts your communications with many major websites, making your browsing more secure. It is produced as a collaboration between The Tor Project and the Electronic Frontier Foundation. Many sites on the web offer some limited support for encryption over HTTPS, but make it difficult to use. As examples, a site may default to unencrypted HTTP, or fill encrypted pages with links that go back to the unencrypted site. The HTTPS Everywhere extension fixes these problems by rewriting requests to these sites to HTTPS. This happens automatically after installing this add-on, and you do not need to take any additional actions. If the icon for this extension crowds your browser, you can remove it by clicking the Firefox menu, then Customize, then dragging the icon away from the menu bar. It does not need to be visually present to function.

Exif Viewer (addons.mozilla.org/en-us/firefox/addon/exif-viewer)

This extension provides right-click access to the Exif data embedded into images. Later chapters explain what Exif data is and how it can be useful. With this extension enabled, you can right-click on any full size image located on a web page. The menu option is "View Image Exif Data" and a new window will open when selected. This window will identify any available metadata about the image. Figure 3.14 (left) displays the right-click menu with a new option to View Image Exif Data. Figure 3.14 (right) displays partial results that identify the make and model of the camera used to capture an image.

Overall, most photos on social networks do not contain any metadata. They have been "scrubbed" in order to protect the privacy of users. However, many blogs and personal websites display images that still contain metadata. While Chapter Twenty will explain online websites that display this data, a browser add-on is much more efficient. In my experience, this extension will increase the amount of times that you will search for this hidden content.

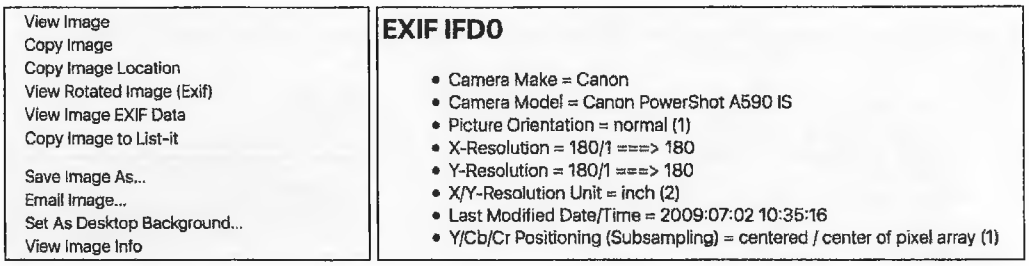


Figure 3.14: The right-click menu and result from an Exif Viewer search.

MJSONViewer (addons.mozilla.org/en-US/firefox/addon/mjsonviewer/)

This extension will probably go unnoticed, which is a good thing. MJSONViewer allows JSON and XML files to be opened and viewed within a web page instead of saving them and viewing the files in a text editor. These files are returned when querying specific types of data such as server content. This mostly applies to the Application Programming Interfaces (APIs) as discussed throughout the book. Many of these APIs deliver the results in a view that is not intended for web browsers, even Firefox. Your browser may try to download a file instead of displaying the content. With this extension installed, the API results should appear within the browser every time. Without this extension, some of the API searches may not function. There is nothing to configure with this add-on. If your results during later tutorials appear scrambled or completely missing, install this add-on.

Recent updates to Firefox's display of JSON and XML content may have made this add-on unnecessary, but I keep it activated for the rare instance it is needed. Figure 3.15 displays a result received in Firefox from a breached email service before installing this extension. Figure 3.16 displays the same result, parsed by this add-on.



Figure 3.15: Data from a JSON output without a JSON extension.

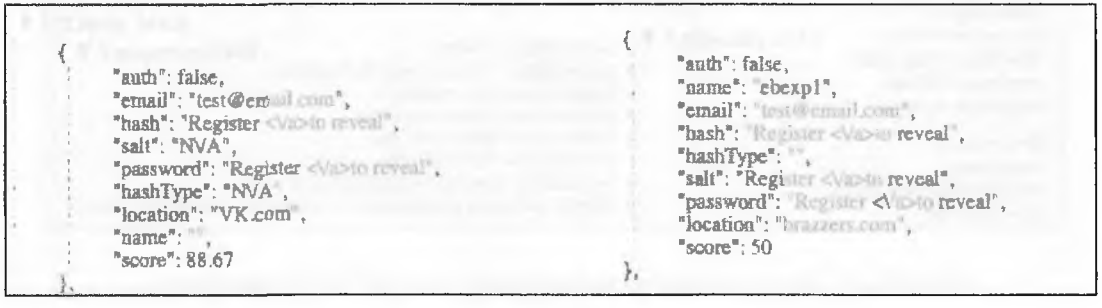


Figure 3.16: Data from a JSON output with a JSON extension.

User Agent Switcher (addons.mozilla.org/en-US/firefox/addon/user-agent-switcher-revived)

Occasionally, you may visit a website that does not want to cooperate with Firefox. Browsers notify websites of their identity and websites can alter or refuse content to certain products. One example is that some older websites require Microsoft's Internet Explorer to view the content. Even though Firefox is capable of displaying the information, a website can refuse the data to the browser. Another example is mobile websites that display different content if viewing from an iPhone instead of a computer. This can now all be controlled with User Agent Switcher.

When installed, you have a new option in your browser. The menu allows you to choose a mobile operating system, such as iOS or Android, or a desktop browser such as Internet Explorer or Chrome. It will also allow you to specify your operating system such as Mac or Windows. Whatever you choose, this data will be sent and confirmed to any site that you visit. If you visit a website of a tech-savvy target, he or she may know that you were looking around. You may also be revealing that you are using a specific browser, such as Firefox, and a Windows computer (common in government). You could now change your agent to that of a mobile device or Google Chromebook which may not look as suspicious. To do so, you would click on the menu bar icon, and simply select the desired configuration. To return to the default Firefox option in your native operating system, click on the checkmark icon in the lower left. Figure 3.17 displays an example where a mobile version of Yahoo was delivered to a desktop computer.

I have used this on several occasions to bypass poor security protocols. During one investigation, I had encountered a web forum of a hacking group that always appeared blank upon visit. Google had indexed it, but I could not see any content. By changing my default agent to Firefox on a Linux machine, I was allowed to see the content. The group had enabled a script that would only allow the page to be viewed on Linux computers. While still employed by the government, various mandated online training needed to be completed in order to maintain specific certifications. This government-hosted training was poorly designed and required users to access via Internet Explorer. Since I used an Apple computer, I could not connect until I changed my agent to Internet Explorer within my Firefox browser.

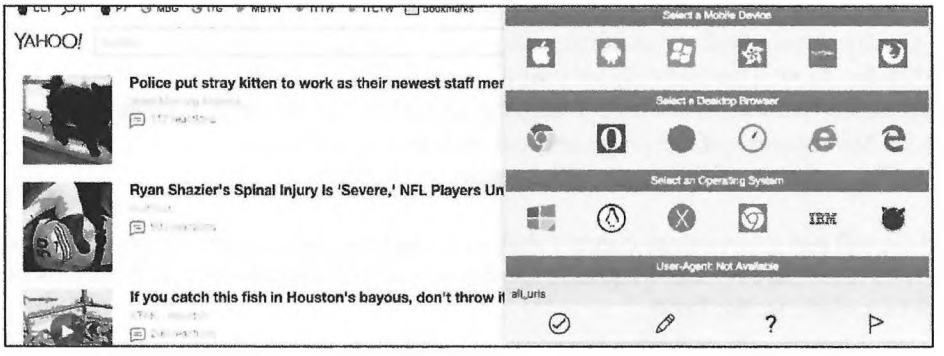


Figure 3.17: User Agent Switcher disguising a desktop system as a mobile device.

Image Search Options (addons.mozilla.org/firefox/addon/image-search-options/)

A later chapter explains reverse image search engines and how they can identify more target photographs. Popular options include Google Images and Tin Eye. This extension automates the reverse search when an image is right-clicked. When installed, "Image Search Options" is present when you right-click on an image. Highlighting this option presents several reverse image search services including Google, Bing, TinEye, Yandex, Baidu, and others. You will later learn how my online search tool will execute an image search across all of those services at once. However, this tool can be beneficial due to the convenience and obscure services such as Karma Decay, which looks for copies of images on Reddit. This add-on removes any excuse to not always check reverse images on target websites. With this add-on enabled, you will be ready to enhance your searches during that investigation. Figure 3.18 displays the options after right-clicking an image located on a target website.

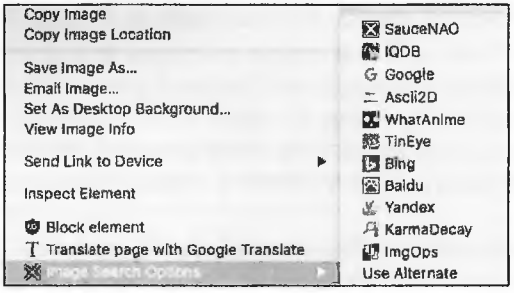


Figure 3.18: A reverse image search menu.

Resurrect Pages (addons.mozilla.org/en-US/firefox/addon/resurrect-pages/)

This extension provides a link to archived versions of websites whenever a page has been modified, is unavailable, or has been deleted. Right-clicking on any site will offer a new menu option of "Resurrect this page". That option will present the following archive services.

Google Cache: A standard cache of the target address from Google
Google Cache Text: The text-only view of a standard Google cache
The Wayback Machine: A link to the target page within The Internet Archive
Archive.is: Any captures of the target address domain on Archive.is
WebCite: Any captures of the target address domain on WebCite

This add-on will not give you any content that you could not locate manually from these sources. Instead, it serves as an easy way to quickly identify interesting content. You will learn more about online archives later in this book.

Copy Selected Links (addons.mozilla.org/en-US/firefox/addon/copy-selected-links/)

This simple add-on will identify any hyperlinks within the selected text of an individual web page. It will store the links within your operating system's clipboard, which allows you to paste them into any application of your choice. While only a small utility, it can quickly turn a large project into an easily completed task.

Using the utility is fairly straightforward. While on any website, select any or all text, right-click anywhere in the page, and select the "Copy selected links" option in the menu. The links will be stored in your clipboard and you can paste them into Notepad, Excel, or any other productivity application. There are unlimited uses for Copy Selected Links, and below are a few of my favorite.

Facebook: When I am on my target's list of Facebook friends, I will select all text and use Copy Selected Links to quickly record each hyperlink to an individual's profile. I will then paste these into Excel for later analysis. Comparison with previous captures identifies those that were "unfriended".

Twitter: When I am viewing a Twitter profile, I will use this utility to capture all links to external websites and photos.

YouTube: When viewing a person's YouTube videos page, Copy Selected Links allows me to paste the entire link collection of every linked video into a report.

eBay: While viewing results from a search for a specific fraudulent product, I can quickly copy the active hyperlinks to each auction and paste them directly into a report in seconds.

Human Trafficking: While viewing ad results for suspected human trafficking victims, I can copy all active hyperlinks and paste directly into a report, email, or memo for other investigators.

Documents: When I encounter a public FTP server or open web directory, this tool allows me to copy the native links to all files encountered. This is helpful for documentation after downloading all of the data.

Performing screenshots during my investigation in these examples would never identify the direct links to the visible content. Only hovering over the links would temporarily identify the source. A combination of screen captures and link collection with this add-on provides a much more comprehensive report.

OneTab (addons.mozilla.org/en-US/firefox/addon/onetab/)

This extension provides simple and effective management of browser tabs by allowing you to collapse any open pages into a list of bookmarks. Once installed, it is activated by clicking on the blue funnel icon on your toolbar. Doing so will close any open tabs and open the OneTab management page. The collapsed tabs are displayed on this page as a group of bookmarks. Each individual bookmark is made up of a page title, with the respective URL embedded as a link. Any previously saved tab groups are stored further down the page in reverse chronological order.

Interacting with the management page is straight forward. At the top of each tab set is an editable title which defaults to the number of tabs in that group. Left click on the title to change it to something logical for that set of links, such as "Username Subject X". To the right of the title is a date and time stamp indicating when the list was saved. You can drag and drop individual bookmarks to change the order in each group or to move links from one group to another. Clicking on "Restore all" will re-open each of the bookmarks into its own tab. The bookmarks are links, not saved pages, so you will be reloading any non-cached page content from the remote host. Selecting "Delete all" will destroy the bookmarks in that group. Selecting "More..." gives options to rename, lock, or star the tab group. "Star this tab group" pins that group to the top of your management page, independent of the date and time created.

Using "Share as web page" is not recommended as it creates an html page of your tab group on the OneTab servers. The preferred method for transferring OneTab bookmarks is to use the export feature, which is found in the upper right menu. This will allow you copy your links as plain text URLs and paste them into a more secure platform. OneTab defaults to deleting the bookmarks from a group if you select "Restore all". To change this behavior, select "Options" and select "Keep them in your OneTab list". Right clicking on a web page will bring up the OneTab context menu which allows for more granular tab selection, as seen in Figure 3.19 (right). OneTab collects no user data unless you intentionally click on the "Share as web page" feature. Barring that feature, all data is stored locally on your workstation.

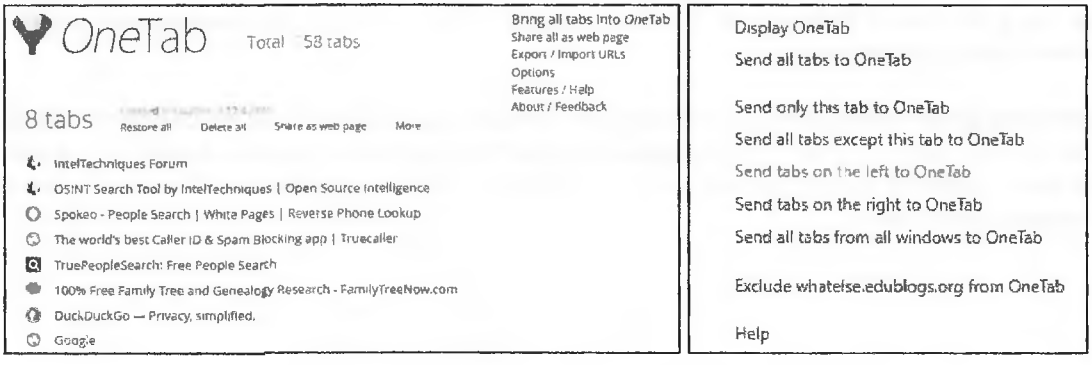


Figure 3.19: The OneTab management page (left) and context menu (right).

KeePassXC Browser (addons.mozilla.org/en-US/firefox/addon/keepassxc-browser/)

In a previous chapter, I explained the requirement to possess a reliable password manager within your investigative computer. I choose the application KeePassXC because it is secure and completely offline. I do not want to keep my usernames and passwords to my covert accounts within any online third-party database. While data exposure is unlikely, I cannot take the chance. Personally, I prefer to copy and paste my passwords directly from the password manager application into the browser. I realize I am more paranoid than the average investigator in matters related to privacy. I also respect that you may need a more convenient option to automatically enter your passwords into multiple covert accounts. My suggested solution offers the convenience of automatically populated passwords with the privacy of an offline database.

Installation of this extension is straightforward. However, it will not work until you make some modifications to your password manager application. This extension is simply a conduit to bring the program functionality into the browser. The first step is to make sure you have the KeePassXC program and database configured on the machine which you want access. If you are installing the browser extension within Firefox on your VM, but your program is on the host computer, they cannot see each other. In this scenario, you will possess KeePassXC and a password database on your VM (explained in the next chapter), and add this browser extension to Firefox within this same VM.

Once you have the application and extension installed, navigate to the "preferences", "options", or "tools" menu (varies by OS and version). Click on "Browser Integration" in the left menu and select the box to enable this feature. Specify which browsers should have access and save your changes. Back in Firefox, open the KeePassXC menu and click "Connect". You should be prompted within your KeePassXC application to allow this connection, and you will be asked to provide a name for the configuration. Once complete, you are ready to start using this extension.

It is important to provide the URL of each website within the entries in the KeePassXC application. When you store your username and password for each service, also include the URL in the appropriate field, such as https://twitter.com. It is vital to include the exact URL, including "https://". This allows the extension to match the URL on the current page to the entry in your password database. Consider the following typical scenario.

In my KeePassXC database, I have several entries for various covert Facebook profiles. When I visit facebook.com, I click on the username field at the login prompt. I immediately see a list of all of my covert accounts, which is retrieved from my KeePassXC database. I can also right-click within a login field and ask my browser to populate the credentials. When I select the desired username, that name and password are automatically populated into the browser. I can log in to that account without typing or pasting a password. All of this happens without relying on a connection to an online password manager. The data is transferred through your local machine, and your passwords are never sent via internet, within plain text or encrypted. This is an offline solution which provides a level of security an online password manager could never supply.

Exporting and Importing a Pre-Configured Profile

At this point, it may seem overwhelming when thinking about the abundance of add-ons and their proper configurations. I currently use several Windows, Apple, and Linux virtual machines and must keep my Firefox browser updated on all of them. I no longer manually update each browser. Instead, I maintain a single Firefox browser that includes all customizations that I desire. I then import these settings into any other browsers in order to replicate the experience across every computer that I use.

The following instructions allow you to export your final settings and import the same customizations into other investigative computers. The details will vary slightly based on your operating system and version. Only execute this tutorial on a new install of Firefox that has no saved settings, or within a version of Firefox that you want to be overwritten. Do not overwrite your current version if you have bookmarks, extensions, or other data that you want to keep. You should backup any settings if proceeding on an older install. As a final warning, the following steps will overwrite any custom options applied to your target Firefox installations.

- Open your CONFIGURED version of Firefox and click the menu button (three horizontal lines), click the help question mark, and then select Troubleshooting Information. The Troubleshooting Information tab will open.
- Under the Application Basics section, click on Open (or Show) Folder (or Directory). A window with your profile files will open. Close Firefox but leave this window open.
- Copy these files into a new folder on a removable drive.
- Open your NEW version of Firefox and click the menu button (three horizontal lines), click the help question mark, and then select Troubleshooting Information. The Troubleshooting Information tab will open.
- Under the Application Basics section, click on Open (or Show) Folder. A window with your profile files will open. Close Firefox but leave this window open.
- Paste the content of the new folder on your removable drive into this folder. Overwrite any files when prompted. Restart Firefox.

The result should be a copy of Firefox which contains every add-on you configured from this chapter. This profile could be copied to an unlimited number of computers, as long as the versions of Firefox were identical. I once visited a large fusion center which had created a custom Firefox profile for use on all computers. This made everyone's version of Firefox practically identical and ready for online investigations. When an employee was assigned to a different workstation, the look and feel of the browser was identical. This also included blocking of specific scripts globally throughout the organization with a custom version of uBlock Origin.

Chrome (google.com/chrome)

Chrome is an excellent browser that is known for being very fast and responsive. Chrome is also very secure by nature, but compromises privacy since Google receives a lot of data about your internet usage. Both Firefox and Chrome "sandbox" each tab. Sandboxing restricts the content in that tab to that tab only, preventing it from "touching" other tabs in the browser, or the computer's hardware. This is a very important feature in preventing malware from being installed when you visit a malicious website.

While I always prefer Firefox as my browser for investigations and daily usage, Chrome is my browser used during live training events. This is due to stability when loading dozens of tabs, and your system should have a lot of RAM if you want to take advantage of Chrome's power. For investigative purposes, Chrome can use several of the add-ons previously mentioned for Firefox. I highly recommend uBlock Origin as discussed previously on any browser that you use, including Firefox, Chrome, Safari, and Opera. The only time that I use Chrome during an investigation is when I am forced because of a Chrome-specific utility. There are a couple of extensions for Chrome which will not work on Firefox. I will focus on those here. Before discussing any investigative resources, I suggest you harden your Chrome security. Enter the Settings menu and consider the following changes.

Privacy: Beside the content settings button is a button labeled "Clear browsing data…" This button will open a dialogue that allows you to clear any or all of the data stored during your sessions. You may erase information for a period of time ranging from the last hour to "the beginning of time". You may wish to use this function to clear all of your browsing data daily.

Passwords and forms: I recommend disabling these features by unchecking both boxes: "Enable Autofill to fill out web forms in a single click", and "Offer to save your web passwords". If you have stored form-fill information or passwords in Chrome, I recommend removing any data before conducting investigations.

Chrome Extensions: To install add-ons in Chrome, navigate to the settings menu. Click "Extensions" on the upper left side of the Chrome interface. You will be presented with all the add-ons that are currently installed in Chrome. I recommend uninstalling any add-ons that you did not personally install or research for trustworthiness. Furthermore, most extensions previously explained for Firefox can be installed in the same manner in Chrome. The following Chrome-only extensions may provide additional benefit to your online research.

I offer a final note about Chrome. I believe it is very invasive into your usage and investigations. If you are concerned about the privacy issues surrounding Google's capture of internet traffic occurring within Chrome, consider switching to the Brave browser. It is based on Chrome, but has eliminated most of the intrusive behavior of Chrome. Of all of the privacy-related versions of Chrome, I believe Brave has the best execution. If you do not need Chrome, and can complete all of your investigations within Firefox, I believe you will possess more privacy.

Prophet (recruitingtools.com/prophet)

Prophet monitors the social networks that you visit and supplies additional details about the targets you are researching. It does not require an account. After installation, launch Prophet by clicking on the appropriate icon in the upper right of your browser. This works best when you are actively on the social network profile of your target. Figure 3.20 displays the view while launched from a person's Twitter page. The results identify her AboutMe, Facebook, FourSquare, Google+, LinkedIn, and Klout profiles. It also connects directly to her personal blog and Flickr page. The "Find Email Address" option reveals two verified email addresses that belong to the target.

This is a premium product with a free tier. I have found that a handful of daily searches keeps me within the free tier unless I search for email addresses. Those seem to deplete quickly, but I have never received an email address result which I could not replicate with free search options online. Overall, I like this tool. However, it will not find anything which you cannot replicate after learning the techniques throughout the rest of this book. Registration often requires a valid email address, and burners are usually forbidden. I must admit that I no longer possess this tool within my investigative machines, but I want you to know of the option. If this extension is the only reason you need Chrome on your VM, I think it is an easy "pass".

While I have successfully used this extension on numerous investigations in years past, I have seen very little benefit lately. While the methods that Prophet uses to obtain the data can be replicated with manual searching, it is a laborious process. This extension does save time if you plan to use it daily. In one investigation, I needed to quickly locate the Facebook pages connected to several Twitter profiles involved in a threat case. Clicking through each profile, with the Prophet sidebar expanded, immediately identified the majority of the accounts. A two-hour task was completed in less than fifteen minutes. This tool works best when executed from a Twitter, Facebook, or LinkedIn profile. It does not work well from blogs or personal websites.



Figure 3.20: A Prophet search result from a Twitter profile.

Hunchly (hunch.ly)

While FireShot and Nimbus were explained as free options that work with both Firefox and Chrome, they both have their limitations. Neither work extremely well with large social network profiles and both provide no type of file management solution. While I try to focus only on free resources, this book would not be complete without a discussion about Hunchly. Hunchly is a paid tool that is designed to optimize your data capture and analysis during an OSINT investigation. Hunchly takes full content captures of every page that you visit so that you don't lose information during the course of your investigation. Additionally, it automatically does the following:

- Creates a cryptographic signature for each page captured for verification purposes
- Automatically extracts EXIF metadata from every photo encountered
- Enables you to tag pages for easy organization of small or large cases
- Powerful full text search of all captured pages and EXIF data
- Flexible export and reporting options
- Automatic attachment of downloads including documents and video files
- API integration with tools such as Maltego

Hunchly is completely integrated with Google Chrome so you can stay in your browser while you are doing your investigative work. With Hunchly working in the background, you never have to worry about remembering to take screenshots or annotate with some tool. All of the pages are captured, timestamped and documented automatically.



Figure 3.21: The Hunchly capture tool in use.

Tor Browser (torproject.org)

Tor is an acronym for The Onion Router. Basically, it allows you to mask your IP address and appear to be browsing the internet from a false location. Normally, when you connect to the internet and browse to a website, that website can identify the IP address that was assigned to you from your internet service provider. This can often identify the city and state that you are in and possibly the business organization where you are currently located. In some instances, it can identify the building you are in if you are using public wireless internet access. The owner of the website can then analyze this information which may jeopardize your investigation. This is one of the many reasons that I recommend the uBlock Origin add-on for Firefox which was explained earlier. uBlock Origin will block most of the analytic code within websites that monitors your information, but it will not stop everything. Occasionally, you may want to change your IP address to make you appear to be someone else in a different country. This is where Tor excels.

The Tor bundle available for free download is completely portable and requires no installation. After download, unzip the file and extract all of the data. You are now ready to start the program by double clicking the "Start Tor Browser" icon. The first task that Tor will complete is to create a connection to a Tor server. This connects you to a server, usually in another country, and routes all of your internet traffic through that server. After the connection is successful, it will load a custom version of the Firefox browser. Now, every website that you visit through this browser will assume you are connecting through this new IP address instead of your own. This provides a layer of privacy to stay hidden from a suspect. This may be overkill for most investigations. If you are only searching and monitoring common services such as Facebook, Twitter, or YouTube, this service is not needed. If you are visiting personal websites and blogs of a tech savvy hacker, you should consider Tor. When using Tor, you may notice a drastic decrease in the speed of your internet. This is normal and unavoidable. This often improves the longer you are connected. To stop the service, simply close the browser. This will disconnect the Tor network and stop all services. Figure 3.22 displays the IP address assigned to me through the Tor Browser (top) and a browser not using Tor (bottom). Any activity conducted through the Tor browser is not associated with my real internet connection and appears to be originating in Canada.

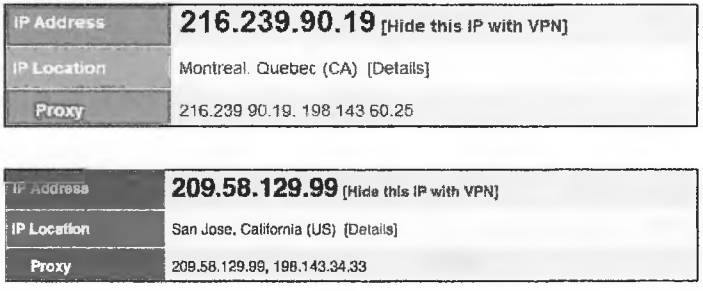


Figure 3.22: A Tor IP address and location (top) and actual data (bottom).

CHAPTER FOUR
LINUX APPLICATIONS

Hopefully, you now have a functioning Linux Ubuntu virtual machine and customized browsers. That alone provides a very secure environment for your online research. However, possessing customized software applications within your VM would greatly enhance your investigations. Ubuntu provides some basic applications in the default installation, but I want to take that to the next level. In this chapter, we are going to customize your Ubuntu VM with numerous OSINT applications which will automate many investigative tasks. First, let's discuss the approach.

In Buscador, David and I provided all applications pre-configured for use. This included icons to launch programs and scripts to help execute specific queries. As stated previously, we should not rely on third-parties to create and maintain these VM configurations (even my own public resources). The goal here is to teach you how to replicate that work and create your own custom VM. If conducting the following tutorials to your master VM, you will only need to take these steps once. Each clone you create will maintain all of your hard work.

Previous editions of this book included a chapter focused on Windows software applications beneficial to OSINT analysis. I now place the emphasis on Linux for three reasons. First, many of those applications have become outdated, or they are no longer maintained. Most do not function as originally intended. Second, I want to enforce better security within our investigations. As stated previously, I believe we should only conduct investigations within secure virtual machines which are free from any contamination of other cases. If I present Windows-based software solutions, I believe it encourages readers to simply install these applications directly to the host and conduct investigations inappropriately. Forcing you to consider Linux alternatives is safer. Finally, there are simply many more OSINT-related applications available for Linux than Windows. We can replicate practically every Windows application mentioned in the previous editions while protecting our investigation within a secure VM. There will be a learning curve if you are not familiar with Linux, but the data obtained during your investigations will be superior to the content retrieved from Windows application equivalents.

Some of this chapter may seem complicated at first. I promise everything becomes easier as you practice. I will explain the entire process of downloading, installing, configuring, and executing each program. Furthermore, I will demonstrate how to create and download scripts in order to easily automate your queries without the necessity of entering the terminal commands and switches. This will be a crash course in Linux, but the lessons learned now will pay off in future usage. Once we have our systems ready, then we can dive into various online search techniques. All scripts referenced throughout this entire chapter can be downloaded from my website at https://inteltechniques.com/osintbook.

Application Installation

Ubuntu possesses a software "store" in which you can point and click through various Linux applications and install them with ease. However, I discourage users from this method. Your choices are minimal and there are likely better alternatives available. Instead, we will use the Terminal for all of our application installations. If you followed the previous tutorials, you may already have the Terminal application in your software dock within your Ubuntu VM created earlier. If not, you can always find the Terminal application within the "Applications" area of Ubuntu by clicking the nine dots within the dock on the left of your screen. Open the Terminal application and leave it open while we install some required software. While I encourage readers to replicate this chapter manually, typing the commands directly, I will also maintain a text file with every required step at https://inteltechniques.com/osintbook. If anything should need updating, you will find modifications there. Let's ease into things slowly.

VLC Media Player

By default, your Windows and Mac operating systems include media players which allow execution of audio and video files. Your default Ubuntu virtual machine does not have this luxury. However, this is easy to correct. VLC is an application which can play practically any media files you throw at it. You could find VLC within the Ubuntu Software application, but I prefer to manually install it. This also provides our first explanation of installation commands in Linux. Within Terminal, type the following command, pressing return after.

```
sudo snap install vlc
```

Let's break down this command, as you will see similar instructions throughout this chapter.

sudo: This command executes any following text with elevated privileges. It is similar to running a program in Windows or Mac as the administrator. When using this command, you will be required to enter your password. Note that passwords entered within Terminal do not appear as you type them, but they are there. Simply press enter when finished typing. Any additional sudo commands in the same terminal session should not ask for the password again.

snap: Snappy is a software deployment and package management system developed by Canonical for the Linux operating system. The packages, called snaps, are easy to install and update.

install: This tells Ubuntu to install a specific software application. In this scenario, it instructed Ubuntu to install VLC. You may need to confirm installation when prompted by entering "y" for "yes".

After you have executed this command, you should see VLC installed within the "Applications" menu by clicking the nine dots icon in the dock to the left. You can launch the application within this menu, and it should be set as the default option for opening most downloaded media files.

FFmpeg

This is another set of media tools, but these only work within Terminal. We will need them when we start adding utilities to manipulate and download videos. Enter the following into Terminal and press enter after each.

```
sudo apt-get update
sudo apt-get install ffmpeg
```

apt-get update: This command updates the Ubuntu lists for upgrades to packages that need upgrading, as well as new packages that have just come to the repositories. It basically fetches information about updates from the repositories mentioned previously.

apt-get install: This tells Ubuntu to install a specific software application. In this scenario, it instructed Ubuntu to install FFmpeg. You may need to confirm installation when prompted by entering "y" for "yes".

Video Download

This may be my most-used utility in Linux. The Python script **YouTube-DL** is the backbone which will help us download bulk videos from YouTube and other sources. This is a Python script, and we will need to install "Preferred Installer Program", otherwise known as PIP. Enter the following into Terminal, which will provide the necessary files for both major Python versions, 2.x and 3.x.

```
sudo apt-get install python-pip
sudo apt-get install python3-pip
```

We can now install YouTube-DL by entering the following command.

```
sudo pip install --upgrade youtube_dl
```

We now have the necessary utilities installed for our media needs. This process is not too exciting, and you will not immediately see anything useful after these actions. However, you are laying the groundwork for the upcoming scripts. The YouTube-DL script is ready to use, but only through the Terminal application. Let's conduct an example in order to understand the features, and then discuss a way to automate queries.

Assume that we are looking for videos of Bob Ross teaching viewers how to paint with oils. After a search on YouTube of Bob Ross, I found the official Bob Ross channel located at https://www.youtube.com/user/BobRossInc. Clicking on the Videos option on that page navigated me to https://www.youtube.com/user/BobRossInc/videos. This page displays over 600 full episodes of his television program. Clicking one of these videos presented a URL of https://www.youtube.com/watch?v=lLWEXRAnQd0. I want to download the video from this

page in the best quality possible in order to attach it to my case. The default YouTube-DL command to do this is as follows.

```
youtube-dl https://www.youtube.com/watch?v=lLWEXRAnQd0
```

This will download the video to whichever directory you are in within Terminal. By default, you are likely in your home folder, but that is not a great place to save a video. Therefore, let's change the saved location to your desktop with the following commands.

```
cd ~/Desktop
youtube-dl https://www.youtube.com/watch?v=lLWEXRAnQd0
```

You should now see the video file on your Desktop in Ubuntu. Playback will likely appear poor within the VM, but you could copy that file out to your host for proper playback through the shared folder on your Desktop. When you retrieved this video, you may have noticed that two files actually downloaded (one video and one audio). This is because YouTube-DL has detected that you have FFmpeg installed, and it chose to download the highest quality option for you. If you had not possessed FFmpeg, you would have been given a lower quality version. This is because YouTube presents separate audio and video files to the viewer when a high quality output is chosen. YouTube-DL works in tandem with FFmpeg to download the best quality option and merge the result into a playable file. This is the best way to extract a single video from YouTube and other video sources. However, the true power is on bulk downloads.

Assume you downloaded the previous file in reference to your investigation. You then see that your target possesses hundreds of videos on his YouTube "videos" page. You want all of these videos, located at https://www.youtube.com/user/BobRossInc/videos. YouTube-DL can download these with the following single command.

```
youtube-dl https://www.youtube.com/user/BobRossInc/videos
```

This will take some time to download everything, especially if the videos are long. If you want to cancel this command, strike ctrl-c on your keyboard at any time. This terminates any process within Terminal, and may be beneficial throughout this chapter. Additional options to add to this command include the following.

`--all-subs` Downloads any closed captioning subtitles associated with the video(s)
`--all-formats` Downloads all versions of a video of any quality

Downloading all available formats of every video is likely overkill and could take several hours. However, obtaining the closed captioning for each video can be quite beneficial. The following command in Terminal will download the best quality video and audio tracks of the target video, merge them together, and download a text file of each of the embedded closed captioning files.

```
youtube-dl https://www.youtube.com/watch?v=lLWEXRAnQd0 --all-subs
```

The video appears on my Desktop, as that was the last place I had navigated within Terminal. Next to the video is two text files with a .vtt file extension. Below are the first two lines of each. The first represents the English subtitles, and the second are in Spanish.

```
00:00:29.988 --> 00:00:32.108
- Hello, I'm Bob Ross and I'd like to welcome

00:00:29.988 --> 00:00:32.108
- Hola, soy Bob Ross y me gustaría dar la bienvenida
```

Imagine the following scenario. You have a suspect on YouTube with hundreds of videos. You have been tasked to download every video and identify the exact files and time stamps of every time he said the word "kill". One command within YouTube-DL would provide all of the data you need to conduct a search within all text files, and the corresponding video evidence. In a later chapter, I will provide a single query which will search all of these text files at once.

This tutorial could be sufficient for your needs, as it will download the media you need. It works well on most video host websites. However, I would like to take things a step further. There are a few minor annoyances with this process. The first is that you must manually open Terminal in order to replicate this technique. Next, you must navigate to a desired folder for saving. Finally, you are at the mercy of YouTube-DL on the naming scheme for your saved file. The following may seem complicated at first, but it will simplify our future usage of this utility.

In both versions of the Buscador VM, David Westcott created scripts which would automate the various tools which we had installed. These were publicly available inside the VM, and were the best feature of the OS in my opinion. They simplified the Terminal tools into point-and-click utilities. The YouTube-DL script is as follows, with minor modifications due to changes in the command options. The following file is located within my online archive of Linux scripts at https://inteltechniques.com/osintbook. **In Chapter Five, we will incorporate all of the scripts explained throughout this entire chapter into our OSINT Master VM.** Before then, we should understand the code and function.

```
#!/usr/bin/env bash
timestamp=$(date +%Y-%m-%d:%H:%M)
url=$(zenity --entry --title "Video Downloader" --text "Enter target URL" -
--entry-text "" 2> >(grep -v 'GtkDialog' >&2))
if [ -n "$url" ]; then
       youtube-dl "$url" -o ~/Videos/Youtube-
DL/"$timestamp%(title)s.%(ext)s" -i -all-subs| zenity --progress --pulsate
--no-cancel --auto-close --title="Video Downloader" --text="Video being
saved to ~/Videos/Youtube-DL/" 2> >(grep -v 'GtkDialog' >&2)
       sleep 2
       nautilus ~/Videos/ >/dev/null 2>&1
else
       zenity --error --text "Missing URL, exiting"
       exit
fi
```

The first line (#) explains that this file is an executable script. The second line creates a current time stamp in order to prevent duplicate file names. The third line (url) creates a prompt for the user to input the URL of the target video. The fourth command (if) instructs the script to enter the proper Terminal command. This includes the YouTube-DL command, the entered URL, a defined folder for output, and a proper file name. The rest of the script launches the folder containing the final download and handles any errors. If you were to save this file to your Desktop as youtubedl.sh, you could launch it with the following Terminal commands.

`cd ~/Desktop` (Places Terminal in the Desktop directory)
`chmod +x youtubedl.sh` (Makes the script executable)
`./youtubedl.sh` (Launches the script)

The result would be the image seen in Figure 4.01 (left). If you entered either of the YouTube URLs used in the previous pages, the script would download the target data to your "Videos" folder. This solves many of the annoyances, but there is still one issue. You still need to open Terminal in order to launch the script. The solution to this is to create a ".desktop" file which we can launch from the dock. Conduct the following steps, with or without the downloaded file explained in a moment.

- Open the Applications menu (nine dots) and launch Text Editor.
- Enter the following text and save the file as youtubedl.desktop to your Desktop.

```
[Desktop Entry]
Name=Video Download
Comment=Launch Youtube-DL
Exec=/home/osint/Documents/scripts/youtubedl.sh
Terminal=false
Type=Application
Categories=Application;OSINT
Icon=/home/osint/Documents/icons/youtube-dl.png
```

- Close Text Editor.
- In Terminal, type: `cd Desktop`
- In Terminal, type: `sudo mv youtubedl.desktop /usr/share/applications/`
- Enter password if prompted.
- Open the Files program.
- Click on "Documents" in the left area.
- Right-click in the white area and select "New Folder".
- Create a folder titled "scripts" and enter this folder.
- Drag and drop the youtubedl.sh file from the Desktop to this folder.
- Click on "Documents" in the left area of the Files program.
- Right-click in the white area and select "New Folder".
- Create a folder titled "icons" and enter this folder.
- Download the compressed archive from https://inteltechniques.com/osintbook.
- Open the archive within your new VM.

- Copy all images from the "icons" folder in the archive into the "icons" folder in the VM.
- Open the Applications menu again and scroll down to "Video Download". If you do not see it, try searching "Video" within the top search field. This should present the icon permanently in this menu. You may need to repeat this step with upcoming tutorials.
- Right-click this program and select "Add to favorites".

You should now see a new icon in your Dock in Ubuntu. Clicking this should present the same dialogue as seen in Figure 4.01 (left). Entering your target URL from YouTube or other services should execute the appropriate options and save your evidence into your Videos folder, as seen in Figure 4.01 (right). The file name includes the date and time of download followed by the title of the video file. The subtitles are also included. The summary of these steps is that we created a desktop shortcut script; moved it to the proper system folder; created a new directory to store our custom scripts, moved our first script into it; downloaded the supplemental files; moved over data; and added our new launch icon into our dock. If you feel lost on any of this, do not worry. We will repeat the process in a moment with another program. As you practice entering these commands, the more comfortable you will become using Terminal for other configurations.

While this configuration took some time, it will save you much effort in the future. You can now download single videos, or entire video collections, without opening Terminal or typing a single command. We will replicate similar steps throughout this chapter, and copy over all shortcuts during Chapter Five. Every configuration, script, desktop shortcut, and icon is available within my archive of custom Linux files on my website at https://inteltechniques.com/osintbook. Alternatively, you can create your own with the details provided in these tutorials. Please note that these examples assume you chose "osint" as your username for your VM, as explained in a previous chapter. All of the demos in this chapter will assume this. If you deviated from this example, you will need to replace "osint" with your chosen username within the scripts.

While named YouTube-DL, this script works on most popular video websites. You should have no issues downloading individual or bulk videos from YouTube, Vimeo, LiveLeak, WSHH, and many others. The bulk download option has saved me numerous hours of manually downloading videos individually. I have yet to find any size or file number limitations. This utility is likely my most used program within Linux, aside from web browsers.

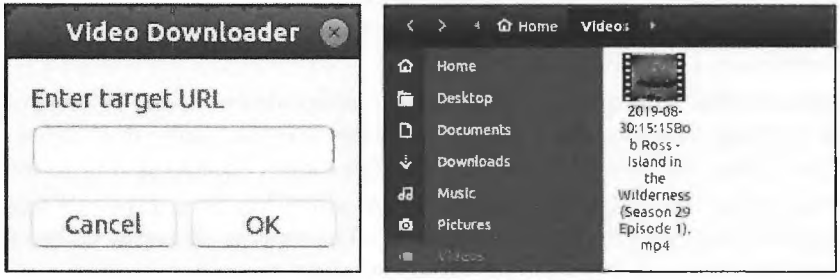


Figure 4.01: A video download prompt (left) and result (right).

I realize this is getting quite technical, and we are very early in the book. None of the steps in this chapter are required to complete the tutorials within the rest of the book. However, having your Linux VM customized and ready to go will be beneficial. In the previous example, you created a utility which will download videos from the internet. Once you see how this works and understand the benefits, you may not want to consider the online options presented later. Let's conduct an additional example of creating your own execution script from the beginning. Next, we will tackle various video utilities.

Video Utilities

The utilities in this section execute the previously installed tool called FFmpeg. This powerful Terminal-based utility can manipulate videos in order to assist with investigations. We will create scripts which will provide the following services.

- **Play a video:** This option will force FFmpeg to attempt to play any video file with multiple video codecs. This will often play videos that would not otherwise play using standard media players such as Windows Media Player and VLC. This will also play many surveillance videos without the need for third-party programs.
- **Convert a video to MP4:** This option simply converts any target video to a standard MP4 format. This is beneficial when the target video possesses a unique codec that prohibits playing universally. If the above option can play the video, this option can convert it so that any computer should be able to play it natively.
- **Extract video frames:** This is likely the most used utility within this set of applications. After supplying a target video, this tool will extract the still images from the video. The result is a folder of uncompressed bitmap (bmp) image files depicting practically every frame of the video. This is beneficial when close examination of frames is necessary.
- **Shorten a video (Low activity):** This version of the script removes single frames of a video that appear extremely similar to the frame preceding it. In other words, it takes out all of the frames which are the same (no action) and only leaves desired activity.
- **Shorten a video (High activity):** This version of the script is identical to the previous with one exception. It is a bit more forgiving for videos with high activity. This might be outdoor surveillance of people walking in the distance or a video with a time counter printed within the feed. If the previous option does not remove enough content, this version will be a bit more aggressive. Both of these work well with surveillance video recorded in real-time (no motion detection).
- **Extract audio:** This option extracts the raw audio file out of any video, converts it to a 320k MP3 file, and saves the file. I have used this to extract audio from video confessions and interviews, and it works well on any online videos downloaded from the internet.
- **Rotate video:** You may encounter a video which has been captured from a mobile device and is rotated 90 degrees clockwise. This option allows us to rotate the video counterclockwise to a traditional view, and flips the video vertically to generate a traditional view of a cellular video.

First, let's take a look at the manual commands which achieve the desired goals of each option. Assume that you possess a video titled evidence.mpg on your Desktop. After changing to your Desktop directory within Terminal (cd ~/Desktop), the following commands would be used.

Play a video:

```
ffplay evidence.mpg
```

This command simply plays the video inside a new window at full resolution.

Convert a video to MP4:

```
ffmpeg -i evidence.mpg -vcodec mpeg4 -strict -2 evidence.mp4
```

This command converts the video to a standard format, and saves it alongside the original.

Extract video frames:

```
ffmpeg -y -i evidence.mpg -an -r 10 img%03d.bmp
```

This command saves the still images from the video to a new folder alongside the original.

Shorten a video (Low activity):

```
ffmpeg -i evidence.mpg -strict -2 -vf
"select=gt(scene\,0.003),setpts=N/(25*TB)" evidence.mp4
```

This command converts the video to a version displaying activity, saving it alongside the original.

Shorten a video (High activity):

```
ffmpeg -i evidence.mpg -strict -2 -vf
"select=gt(scene\,0.005),setpts=N/(25*TB)" evidence.mp4
```

This command converts the video to a version displaying activity, saving it alongside the original.

Extract audio:

```
ffmpeg -i evidence.mpg  -vn -ac 2 -ar 44100 -ab 320k -f mp3 evidence.mp3
```

This command converts the video to a standard audio mp3, and saves it alongside the original.

Rotate video:

```
ffmpeg -i evidence.mpg  -vf transpose=0 evidence.mp4
```

This command rotates the video 90 degrees counter clockwise and saves it alongside the original.

As you can see, these commands can be lengthy and technical. We want to create one single custom script that will allow us to choose which service we want to execute. This is a bit different than the previous example, as it will combine several options into one dialogue. The following two pages display the script we used on our VM. Again, please download all scripts from the archive mentioned previously instead of trying to type these manually.

```
#!/usr/bin/env bash
#Script for ffmpeg interaction

zenity --info --text="The next window will prompt for a target media file"
--title="Video Utilities"

sleep 1

#prompt for file selection
ffmpeg_file=$(zenity --file-selection --title "Video Utilities")
timestamp=$(date +%Y-%m-%d:%H:%M)

#define choices
opt1="Play a video"
opt2="Convert a video to mp4"
opt3="Extract video frames"
opt4="Shorten a video (Low Activity)"
opt5="Shorten a video (High Activity)"
opt6="Extract Audio"
opt7="Rotate Video"

#make sure file is selected
if [ -n "$ffmpeg_file" ]; then

        #prompt for user choice selection
        ffmpeg=$(zenity  --list  --title "Video Utilities" --text "What do
you want to do?" --width=400 --height=400 --radiolist  --column "Choose" --
column "Option" TRUE "$opt1" FALSE "$opt2" FALSE "$opt3" FALSE "$opt4"
FALSE "$opt5" FALSE "$opt6" FALSE "$opt7")

        #perform actions based on selection
        case $ffmpeg in

        $opt1 )
          ffplay "$ffmpeg_file"
          ;;
        $opt2 )
          ffmpeg -i "$ffmpeg_file" -vcodec mpeg4 -strict -2
"/home/osint/Videos/$timestamp.mp4" | zenity --progress --pulsate --no-
cancel --auto-close --title="ffmpeg" --text="Converting Video to mp4"
           nautilus "/home/osint/Videos/" >/dev/null 2>&1
          ;;
        $opt3 )
          mkdir "/home/osint/Videos/$timestamp-frames"
          ffmpeg -y -i "$ffmpeg_file" -an -r 10
"/home/osint/Videos/$timestamp-frames/img%03d.bmp" | zenity --progress --
pulsate --no-cancel --auto-close --title="ffmpeg" --text="Extracting
Frames"
          nautilus "/home/osint/Videos/" >/dev/null 2>&1
          ;;
        $opt4 )
```

```
          ffmpeg -i "$ffmpeg_file" -strict -2 -vf
"select=gt(scene\,0.003),setpts=N/(25*TB)" "/home/osint/Videos/$timestamp-
low.mp4" | zenity --progress --pulsate --no-cancel --auto-close --
title="ffmpeg" --text="Shortening video (Low Activity)"
          nautilus "/home/osint/Videos/" >/dev/null 2>&1
          ;;
        $opt5 )
          ffmpeg -i "$ffmpeg_file" -strict -2 -vf
"select=gt(scene\,0.005),setpts=N/(25*TB)" "/home/osint/Videos/$timestamp-
high.mp4" | zenity --progress --pulsate --no-cancel --auto-close --
title="ffmpeg" --text="Shortening video (High Activity)"
          nautilus "/home/osint/Videos/" >/dev/null 2>&1
          ;;
        $opt6 )
          ffmpeg -i "$ffmpeg_file" -vn -ac 2 -ar 44100 -ab 320k -f mp3
"/home/osint/Videos/$timestamp.mp3" | zenity --progress --pulsate --no-
cancel --auto-close --title="ffmpeg" --text="Extracting Audio"
           nautilus "/home/osint/Videos/" >/dev/null 2>&1
         ;;
        $opt7 )
          ffmpeg -i "$ffmpeg_file" -vf transpose=0
"/home/osint/Videos/$timestamp.mp4" | zenity --progress --pulsate --no-
cancel --auto-close --title="ffmpeg" --text="Rotating Video"
           nautilus "/home/osint/Videos/" >/dev/null 2>&1
        esac

else
    zenity --error --text "No file selected, exiting"
    exit
fi
```

Much of this script is similar to the previous example. However, there are some unique differences. In the YouTube-DL script, we only gave it one task. In this script, we have multiple options. Let's walk through each section.

The first two lines (#) announce that this is an executable script and provide an identifier for us. The (zenity) line prompts the user with a dialogue telling them to pick a file in the next window. This could be eliminated if desired, but a good reminder for new users. The (sleep) option places one second in between actions in order to prevent potential errors. The (#prompt) section instructs the script to open a Files window in order to allow the user to select a target video from the VM. It then defines that video within the script and obtains a time stamp for unique naming. The (#define) section identifies the options for the user, which match our previous manual entries. The (#make) section confirms that the user chose a file, and the script would fail if the user did not. The next (#prompt) section displays the utility selection window to users which allows them to choose the appropriate action to take against the target video.

Each ($opt) line designates the specific action for that option. As an example, ($opt2) instructs the script to run the ffmpeg command within Terminal; add the -i switch; load the target video; insert specific ffmpeg operations, save the file to the Videos folder; add the timestamp to the file

name; present a status window; and finally open the Videos folder when complete, in order to display the result.

If desired, you could add new options to this script when you identified additional ffmpeg features you would like to implement. This applies to every script presented within this chapter. Once you have the basic templates, you can modify these any way that is beneficial to your investigations.

Find this script within the Linux archive file you previously downloaded from my site. Save this script into the "scripts" folder in your "Documents" folder. This should be the same location where you saved the YouTube-DL script. Next, we need to make a shortcut to open this script, just as we did previously. In the Linux archive download I presented earlier, there is a file titled "ffmpeg.desktop" Place this file on the Desktop within your VM, and conduct the following tasks within Terminal.

- `cd Desktop`
- `sudo mv ffmpeg.desktop /usr/share/applications/`
- Enter password if prompted.

This will move the shortcut file to the appropriate location. The reason we cannot use the Files program to do this is because the destination is a special system folder, and it needs elevated privileges in order to place data of inside it. Using terminal with the sudo command authorizes this task. Below is the content of the ffmpeg.desktop file.

```
[Desktop Entry]
Name=Video Utilities
Comment=Launch ffmpeg
Exec=/home/osint/Documents/scripts/ffmpeg_interact.sh
Terminal=false
Type=Application
Categories=Application;OSINT
Icon=/home/osint/Documents/icons/ffmpeg.png
```

This should look very similar to the previous shortcut. Again, this file assumes you have applied the same paths and usernames as my demonstration throughout this chapter. Next, let's add a shortcut icon to our Dock in order to have easy access to this new utility.

- Open the Applications menu again and scroll down to "Video Utilities".
- Right-click this program and select "Add to favorites".

You should now have a new icon in your Dock directly below the previous YouTube-DL option. You can drag-and-drop these in a different order if desired. The new icon launches a dialogue that first notifies you that "The next window will prompt for a target media file". After clicking OK, you are presented with a file manager window that prompts you to choose a target video file. This could be a downloaded YouTube video, an unplayable surveillance video, or any other downloaded video content retrieved from any source. After you select the target video, you are

presented with a menu of the options mentioned previously. After choosing the action, the script processes the video and creates the new evidence. All of this happens without opening Terminal or typing any commands.

The previous two examples explained the details of how these scripts and shortcuts function. The remaining utilities we need to add are quite similar and do not need the same attention to detail. Instead, I present a description of each utility, the content of each script and shortcut, the exact commands to include in your Ubuntu VM, and an example of usage. Let's start with Instagram.

Instagram Utilities

There are two independent programs which assist with bulk-download of Instagram data. The first is **Instalooter**, which has been a staple within Buscador since the beginning. Since it occasionally stops working, pending an updated release, we also want **Instaloader** available at all times. The results of each are very similar, but it is good to have redundant options. First let's install each program into our VM.

```
sudo pip install instalooter
sudo pip3 install instaloader
```

The pip3 option was used because Instaloader requires a newer version of Python (3.x). Now that you have both programs installed, we can test them with the following two commands within Terminal. Please note this is not my account, but serves as a good demonstration with minimal content.

```
cd Desktop
instalooter user mikeb
instaloader user mikeb
```

The Instalooter should have downloaded several individual images to your Desktop, while Instaloader should have created a new folder titled mikeb on your Desktop containing the same images. Since these two programs save data in a unique way, our script will need unique options for each. The entire script is displayed on the following page, and is titled instagram.sh within the Linux files archive. Notice that the Instalooter options are nearly identical to our previous scripts. However, the Instaloader option has unique lines. Under ($opt2), I had to instruct the application to create a new directory and enter that folder. This is because Instaloader does not allow you to specify the saved location of the data. Therefore, we will cheat and make the script enter that folder for us. In both of these options, all data will be saved to folders containing the target username within the folder titles. These will be in the "instalooter" and "instaloader" folders in your "Documents" folder.

```
#!/usr/bin/env bash
##Instagram Menu Script

#define choices
opt1="instalooter"
opt2="instaloader"

timestamp=$(date +%Y-%m-%d:%H:%M)

socialmenu=$(zenity  --list  --title "Instagram:Choose Tool" --text "What
do you want to do?" --width=400 --height=200 --radiolist  --column "Choose"
--column "Option" TRUE "$opt1" FALSE "$opt2"  2> >(grep -v 'GtkDialog'
>&2))

case $socialmenu in

      $opt1 ) #instalooter
      handle=$(zenity --entry --title "Instalooter" --text "Enter
Instagram User ID" 2> >(grep -v 'GtkDialog' >&2))

        if [ -n "$handle" ]; then

            if [ ! -d "/home/osint/Documents/instalooter/$timestamp-
$handle/" ]; then

                mkdir /home/osint/Documents/instalooter/$timestamp-
$handle/
            fi

        instalooter user $handle
/home/osint/Documents/instalooter/$timestamp-$handle/ -v -m -d -e
2>/dev/null | zenity --progress --pulsate --no-cancel --auto-close --
title="Instalooter" --text="Grabbing all photos for: $handle" 2> >(grep -v
'GtkDialog' >&2)

        nautilus /home/osint/Documents/instalooter/$timestamp-$handle/
>/dev/null 2>&1

        else
            zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
            exit
        fi
      ;;
      $opt2) #Instaloader
      mkdir ~/Documents/instaloader
      cd ~/Documents/instaloader
      handle=$(zenity --entry --title "Instaloader" --text "Enter
Instagram User ID" 2> >(grep -v 'GtkDialog' >&2))
      instaloader $handle
      nautilus /home/osint/Documents/instaloader/$handle/ >/dev/null 2>&1
      esac
```

The following is the content from our "instagram.desktop" shortcut file included in the Linux files archive.

```
[Desktop Entry]
Name=Instagram
Comment=Launch Instalooter
Exec="/home/osint/Documents/scripts/instagram.sh"
Type=Application
Categories=Application;OSINT
Icon=/home/osint/Documents/icons/instagram.png
Terminal=false
```

Place the "instagram.sh" script in the "scripts" folder in your "Documents" folder next to the others. Place the "instagram.desktop" file directly onto your Desktop and type the following commands into Terminal.

- `cd Desktop`
- `sudo mv instagram.desktop /usr/share/applications/`

Next, let's add a shortcut icon to our Dock in order to have easy access to this new utility.

- Open the Applications menu again and scroll down to "Instagram".
- Right-click this program and select "Add to favorites".

You now have yet another application at your fingertips, and no requirement to use Terminal or typed commands. This utility attempts to extract media from a target's Instagram profile. While screen captures of an account may suffice for initial investigations, you should consider downloading any images directly from the source when possible. Launching this script from the dock prompts you to choose either Instalooter or Instaloader. You are then prompted to supply an Instagram handle, such as mikeb. This will begin the extraction of all images associated with the username. After providing mikeb as my target, Instalooter and Instaloader both quickly downloaded all 10 photos from his account and placed them in the Documents folder. It also opens the target folder upon completion to immediately display the evidence. In the past, I have used these utilities to download thousands of images from a single account.

Twitter Utilities

In previous editions of this book, I presented two options for automated downloading of Twitter data via a Linux script. The first was a custom solution which required access to Twitter's Application Programming Interface (API). This was problematic and API keys can be abused and then terminated. The second option relied on a Terminal version of Tinfoleak, and the results were not reliable. This time, we are fortunate to have a new option called **Twint**. Open Terminal and install all necessary software with the following command.

- `pip3 install twint`

Depending on when you read this, that single command may be all you need to use this program. In Terminal, type `twint -h` and press enter. If you see several usage options, then you are all set. If you receive a "not found" response, Twint is not properly installed. While writing this chapter, the previous command did not work for me (or for many others). Instead, I had to enter the following into a new Terminal window. Note that this will also install **Git**, which will be beneficial for future program installations.

- `cd Downloads`
- `mkdir Programs`
- `cd Programs`
- `sudo apt install git`
- `git clone https://github.com/twintproject/twint.git`
- `cd twint`
- `sudo python3 setup.py install`

This will create a folder in your Downloads directory titled Programs, and a new folder there titled "twint". If you have any issues with the installation or usage of Twint, visit the official website for Twint at https://github.com/twintproject/twint. It is possible there have been substantial changes since publication. Hopefully, you now have everything you need to conduct manual searches within Terminal. The queries which I have found most beneficial to my investigations are below. In each of these examples, the target Twitter account is my own profile (@IntelTechniques). You can type each of these directly into Terminal to achieve the same results which I have summarized below.

- `twint -u inteltechniques`: Display all of the Tweets from @inteltechniques' timeline on the screen. This serves as a good test to see if your installation is functioning.
- `twint -u inteltechniques -o inteltechniques.csv --csv`: Acquire all of the Tweets from @inteltechniques and save them to a spreadsheet titled inteltechniques.csv.
- `twint -u inteltechniques --followers -o inteltechniques-followers.csv --csv`: Save all followers of @inteltechniques to a spreadsheet.
- `twint -u inteltechniques --following -o inteltechniques-following.csv --csv`: Save all accounts followed by @inteltechniques to a spreadsheet.
- `twint -u inteltechniques --favorites -o inteltechniques-favorites.csv --csv`: Save all favorites of @inteltechniques to a spreadsheet.
- `twint -s osint -o osint-keyword.csv --csv`: Save every Tweet containing "osint" from all of Twitter to a spreadsheet titled osint.csv. This can take a long time.

There are many additional options within Twint, but these are the basics which are valuable for any investigation. I have added each of these into my automated script for easy access. The entire script can be seen on the following two pages, and is titled "twitter.sh" in the Linux file archive.

```
#!/usr/bin/env bash
##Twitter Menu Script

#define choices
opt1="Download User's Tweets"
opt2="Download User's Followers"
opt3="Download User's Following"
opt4="Download User's Favorites"
opt5="Search a Keyword"

timestamp=$(date +%Y-%m-%d:%H:%M)

socialmenu=$(zenity  --list  --title "Twitter Tool" --text "What do you want
to do?" --width=400 --height=300 --radiolist  --column "Choose" --column
"Option" TRUE "$opt1" FALSE "$opt2" FALSE "$opt3" FALSE "$opt4" FALSE "$opt5"
2> >(grep -v 'GtkDialog' >&2))

case $socialmenu in
      $opt1 )
      handle=$(zenity --entry --title "Download Tweets from User" --text
"Enter Twitter Username" 2> >(grep -v 'GtkDialog' >&2))
        if [ -n "$handle" ]; then
            if [ ! -d "/home/osint/Documents/$timestamp-$handle/" ]; then
                mkdir /home/osint/Documents/$timestamp-$handle/
            fi
        twint    -u    $handle    -o    /home/osint/Documents/$timestamp-
$handle/$handle.csv --csv 2>/dev/null | zenity --progress --pulsate --no-
cancel --auto-close --title="Twint" --text="Downloading Tweets from: $handle"
2> >(grep -v 'GtkDialog' >&2)
        nautilus /home/osint/Documents/$timestamp-$handle/ >/dev/null 2>&1
        else
            zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
            exit
        fi;;
      $opt2 )
      handle=$(zenity --entry --title "Download Followers for User" --text
"Enter Twitter Username" 2> >(grep -v 'GtkDialog' >&2))
        if [ -n "$handle" ]; then
            if [ ! -d "/home/osint/Documents/$timestamp-$handle/" ]; then
                mkdir /home/osint/Documents/$timestamp-$handle/
            fi
        twint -u $handle --followers -o /home/osint/Documents/$timestamp-
$handle/$handle-followers.csv --csv 2>/dev/null | zenity --progress --pulsate
--no-cancel --auto-close --title="Twint" --text="Downloading Followers of:
$handle" 2> >(grep -v 'GtkDialog' >&2)
        nautilus /home/osint/Documents/$timestamp-$handle/ >/dev/null 2>&1
        else
            zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
            exit
        fi;;
```

```
        $opt3 )
        handle=$(zenity --entry --title "Download Profiles Followed by User"
--text "Enter Twitter Username" 2> >(grep -v 'GtkDialog' >&2))
          if [ -n "$handle" ]; then
               if [ ! -d "/home/osint/Documents/$timestamp-$handle/" ]; then
                    mkdir /home/osint/Documents/$timestamp-$handle/
               fi
          twint -u $handle --following -o /home/osint/Documents/$timestamp-
$handle/$handle-following.csv --csv 2>/dev/null | zenity --progress --pulsate
--no-cancel --auto-close --title="Twint" --text="Downloading Follows of:
$handle" 2> >(grep -v 'GtkDialog' >&2)
          nautilus /home/osint/Documents/$timestamp-$handle/ >/dev/null 2>&1
          else
               zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
               exit
          fi;;
        $opt4 )
        handle=$(zenity --entry --title "Download Favorites from User" --text
"Enter Twitter Username" 2> >(grep -v 'GtkDialog' >&2))
          if [ -n "$handle" ]; then
               if [ ! -d "/home/osint/Documents/$timestamp-$handle/" ]; then
                    mkdir /home/osint/Documents/$timestamp-$handle/
               fi
          twint -u $handle --favorites -o /home/osint/Documents/$timestamp-
$handle/$handle-favorites.csv --csv 2>/dev/null | zenity --progress --pulsate
--no-cancel --auto-close --title="Twint" --text="Downloading Favorites from:
$handle" 2> >(grep -v 'GtkDialog' >&2)
          nautilus /home/osint/Documents/$timestamp-$handle/ >/dev/null 2>&1
          else
               zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
               exit
          fi;;
        $opt5 )
        handle=$(zenity --entry --title "Search a Keyword" --text "Enter
Keyword" 2> >(grep -v 'GtkDialog' >&2))
          if [ -n "$handle" ]; then
               if [ ! -d "/home/osint/Documents/$timestamp-$handle/" ]; then
                    mkdir /home/osint/Documents/$timestamp-$handle/
               fi
          twint    -s    $handle    -o    /home/osint/Documents/$timestamp-
$handle/$handle-keyword.csv --csv 2>/dev/null | zenity --progress --pulsate
--no-cancel --auto-close --title="Twint" --text="Downloading Search Results
of: $handle" 2> >(grep -v 'GtkDialog' >&2)
          nautilus /home/osint/Documents/$timestamp-$handle/ >/dev/null 2>&1
          else
               zenity --error --text "No handle entered, exiting" 2> >(grep -
v 'GtkDialog' >&2)
               exit
          fi
esac
```

After saving this script to the Documents/Scripts folder, we now need a shortcut to launch it. Below is the exact text included within the file titled twitter.desktop within the Linux files archive.

```
[Desktop Entry]
Type=Application
Name=Twitter
Categories=Network;OSINT
Exec=/home/osint/Documents/scripts/twitter.sh
Icon=/home/osint/Documents/icons/twitter.png
Terminal=true
```

Place this "twitter.desktop" file directly onto your Desktop and type the following commands into Terminal.

- `cd Desktop`
- `sudo mv twitter.desktop /usr/share/applications/`

Next, let's add a shortcut icon to our Dock in order to have easy access to this new utility.

- Open the Applications menu again and scroll down to "Twitter".
- Right-click this program and select "Add to favorites".

Finally, let's discuss the results of this new program. After I clicked the new Twitter icon in my Dock, which can be seen in the lower left of Figure 4.02 (left), I received a menu, visible in Figure 4.02 (left). I entered "inteltechniques" as the target in the menu, as seen in Figure 4.02 (right), and allowed the program to execute. I was presented with a new Documents window which contained a single file titled inteltechniques.csv. Opening this file within any spreadsheet program, such as the included LibreOffice, reveals the data obtained. A very partial view of this spreadsheet is visible in Figure 4.03. I use this for every Twitter target I investigate because the text files are in standard CSV (comma separated value) format and can be easily imported into other software. Twint is my chosen Twitter utility because it conducts true scraping and does not require a Twitter API key.

Let's pause for a moment and catch our breath. You have accomplished a lot in a short amount of time. You have a secure Linux virtual machine installed and configured. You also have four new utilities ready in your Application Dock. You are ready to download videos with the Video Downloads script; manipulate the new video evidence with the Video Utilities menu; extract all Instagram images from your target's profile; and acquire many different categories of Twitter data during your next investigation. Once we lock in your "Master VM", you will never need to replicate this process in the future. Now that you understand the concept of installing a program; creating or downloading an automated script; and creating or downloading a desktop shortcut to launch our new option, we can now pick up the pace. We have more to add, but I will not provide explicit breakdowns of each. Let's begin with some documentation utilities.

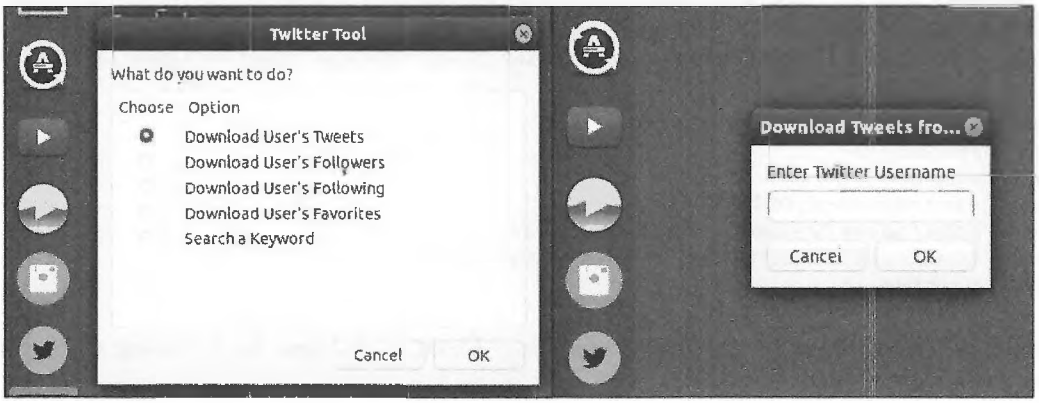


Figure 4.02: A Twitter tool menu (left) and username prompt (right).

C	D	E	F
created_at	date	time	tweet
1567183450000	2019-08-30	10:44:10	This week I announce a new OSINT book being published in 2020. options for credit ca
1567177120000	2019-08-30	08:58:40	Yes. I have a very lengthy chapter outlined for data sets, breaches, leaks, etc.
1567176123000	2019-08-30	08:42:03	More details on today's show in a few hours
1567175888000	2019-08-30	08:38:08	The new OSINT professional must be self-sustaining and possess their own tools and
1567175859000	2019-08-30	08:37:39	Writing has begun...The 7th edition of my #OSINT book will be released in early 2020.

Figure 4.03: Results from a Twitter username search.

LibreOffice

LibreOffice is a free and open source office suite, comprised of programs for word processing (Writer), the creation and editing of spreadsheets (Math), and PowerPoint (Impress). The Writer program can be used within your Linux VM for creating reports while the Math program can be useful with opening CSV files created by the previous utilities. Your installation of Ubuntu should already have this productivity suite installed. If not, you can add the package with the following Terminal command.

```
snap install libreoffice
```

Personally, I do not like typing my reports from within a VM. I prefer to launch my desired word processor within my host operating system. This creates a small layer of protection between the OS which you are using during your "covert" investigation (VM), and the machine which is documenting the written report (host). This is all personal preference, and you may choose another route. Much later in the book, you will learn several documentation strategies for your next online investigation.

EyeWitness

This Python script automates the collection of screen captures from websites. Imagine the following scenario. You are investigating a long list of website addresses that was provided to you as a lead. Maybe they were websites visited from a suspect's computer, a list of social network profiles discovered during your previous Twitter scrapes, or just a self-created list of URLs associated to your investigation. Manually visiting each site and using a screen capture tool can be overwhelming. Instead, let's automate the task. Install EyeWitness to your VM with the following commands within a new instance of Terminal.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/ChrisTruncer/EyeWitness.git`
- `cd EyeWitness/setup`
- `sudo ./setup.sh`
- `sudo apt-get install xvfb`

If you received an error because you do not have Git installed, reopen Terminal and repeat the process, but enter `sudo apt install git` before the first command. In these steps, we entered the new folder you created during the installation of Twint (Programs). This will be where we store applications which do not traditionally install into your operating system. You can now execute EyeWitness, but you must first navigate to the folder where the Python script is located. We will correct this with a custom script in a moment, but let's go ahead and test the application within Terminal. Conduct the following steps.

- Open your Applications menu and launch Text Editor.
- Type or paste some URLs and save the file to your Desktop as sites.txt.
- Open Terminal and enter the following commands.
- `cd ~/Downloads/Programs/EyeWitness`
- `./EyeWitness.py -f ~/Desktop/sites.txt --web -d ~/Documents/EyeWitness/`

When finished, you should have a new file within your Documents/EyeWitness folder titled Report.html. Double-clicking this file opens it within Firefox. Figure 4.04 displays the first two results from my sites.txt file which contained exactly the following data.

```
https://inteltechniques.com
https://computercrimeinfo.com
https://instagram.com/mikeb
```

The results include screen captures of each target website and detailed information including the server IP address, page title, modification date, and full source code of the page. This is very beneficial in two ways. First, it automates the task of researching multiple target websites. Second, it saves each page as it existed at a specific moment in time. If your suspect deletes or modifies a website after you have captured it, you possess evidence of a previous state. The annoyances of

this method include the requirement to create a text file and need to articulate the location of the data within Terminal. We can make this easier with a custom script.

The script on the following page will prompt you to choose between a single URL or multiple URLs. If you choose single, it will prompt you for the URL and then it will execute EyeWitness. The report will be saved into a new folder titled with a current timestamp within the Documents/EyeWitness folder. At the end of the process you will be asked if you would like to open the report. If you choose the option for multiple URLs, you will be prompted to choose the text file containing all of the target addresses. EyeWitness will then conduct the queries and generate the report as described in the single URL option. The following script and shortcut file are included within the Linux files archive mentioned previously, titled eyewitness.sh and eyewitness.desktop.

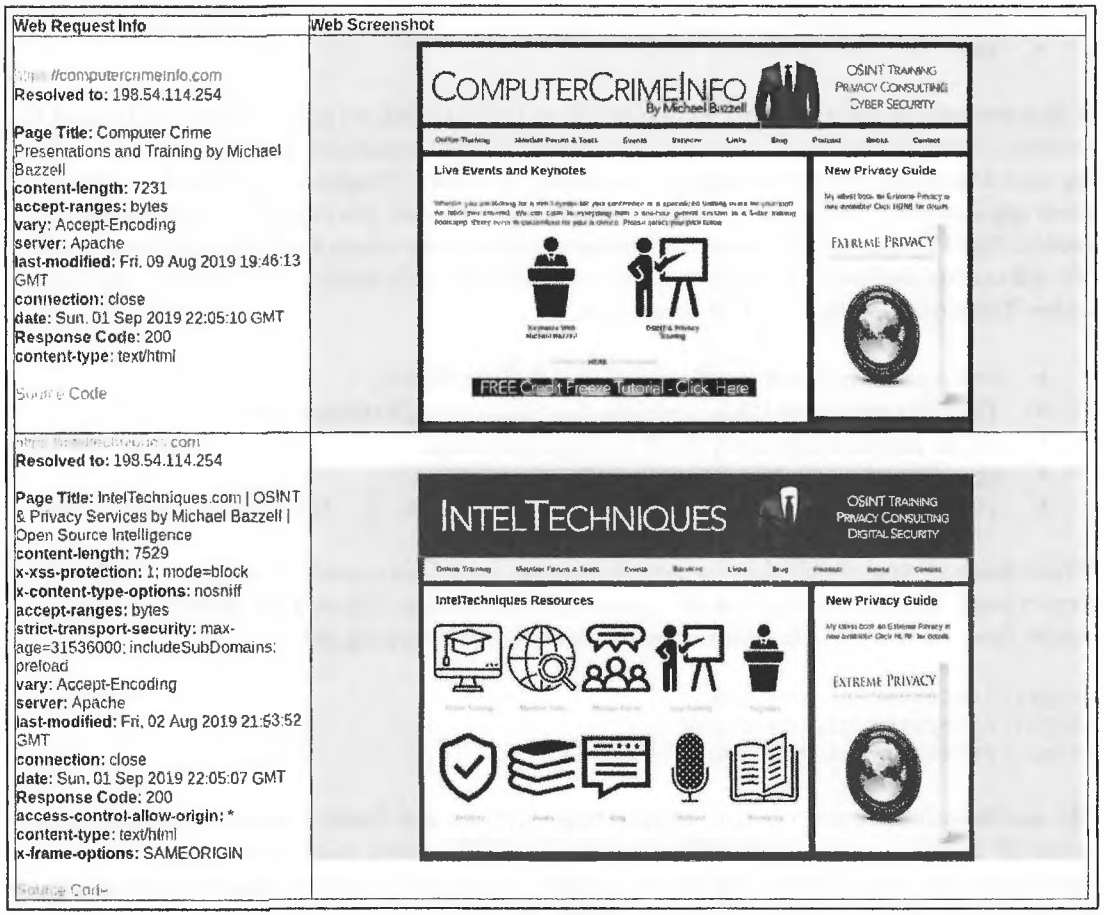


Figure 4.04: An EyeWitness report.

```
#!/usr/bin/env bash
##EyeWitness Script

      #define choices
      opt1="Single URL"
      opt2="Multiple URLs (File)"
      eyewitness=$(zenity  --list  --title "EyeWitness" --text "Do you have
Single or multiple URLs?" --width=400 --height=200 --radiolist  --column
"Choose" --column "Option" TRUE "$opt1" FALSE "$opt2" 2> >(grep -v 'GtkDialog'
>&2))
      case $eyewitness in
      $opt1 ) #Single
            domain=$(zenity --entry --title "EyeWitness" --text "Enter
target URL (ex: https://www.google.com)" --entry-text "" 2> >(grep -v
'GtkDialog' >&2))
            if [ -n "$domain" ]; then
            cd /home/osint/Downloads/EyeWitness
            ./EyeWitness.py --web --single "$domain"
            else
            zenity --error --text "No URL entered, exiting" 2> >(grep -v
'GtkDialog' >&2)
            exit
            fi;;
      $opt2 ) #Multiple
            eyewitness_file=$(zenity --file-selection --title "URL List" -
-text "Select File of URLs" 2> >(grep -v 'GtkDialog' >&2))
            if [ -n "$eyewitness_file" ]; then
            cd /home/osint/Downloads/EyeWitness
            ./EyeWitness.py --web -f "$eyewitness_file"
            else
            zenity --error --text "No file found, exiting" 2> >(grep -v
'GtkDialog' >&2)
            exit
            fi
      esac
```

```
[Desktop Entry]
Type=Application
Name=EyeWitness
Categories=Network;OSINT
Exec=/home/osint/Documents/scripts/eyewitness.sh
Icon=/home/osint/Documents/icons/eyewitness.png
Terminal=true
```

Place the "eyewitness.sh" file into the "scripts" folder in your "Documents" folder next to the other scripts. Place the "eyewitness.desktop" file directly onto your Desktop and type the following commands into Terminal.

- `cd Desktop`
- `sudo mv eyewitness.desktop /usr/share/applications/`

Next, let's add a shortcut icon to our Dock in order to have easy access to this new utility.

- Open the Applications menu again and scroll down to "EyeWitness".

This utility can be very beneficial when you have dozens, hundreds, or even thousands of domains of interest. I once converted a list of over 300 suspect Twitter accounts into a folder of screen captures of each. This allowed me to quickly identify which accounts were truly valuable to my investigation by simply viewing the evidence similar to photos in an album. The report was several hundred pages, but was generated in only a few minutes.

Domain Utilities

Let's create another complicated script, and then we can finish with some standard applications. In the previous utility, we created a new folder called Programs in your Downloads folder. We will continue to use this within the following script. These application files are usually stored within other areas of your operating system, but I want them easily accessible to you, which will make it easier to keep them updated. First, let's gather all of the files needed with the following commands within a new instance of Terminal. Note that "&&" combines two commands.

```
cd ~/Downloads/Programs
sudo snap install amass
git clone https://github.com/aboul3la/Sublist3r.git
cd Sublist3r && sudo pip install -r requirements.txt
cd ~/Downloads/Programs
git clone https://github.com/s0md3v/Photon.git
cd Photon && sudo pip3 install -r requirements.txt
cd ~/Downloads/Programs
git clone https://github.com/laramies/theHarvester.git
cd theHarvester && git checkout 8b88a66
sudo pip3 install -r requirements.txt
```

You should now have three new folders within a folder titled "Programs" inside your "Downloads" folder. These folders contain the Python scripts necessary to launch the programs. Next, let's take a look at the manual commands required to execute each program. I will use my own website during the demonstrations. You would need to execute each script from within the proper folder containing the Python files. Our custom script allows us to use a menu without the need for Terminal commands. This is included in the Ubuntu Files archive as "domains.sh".

```
amass enum -src -ip -brute -d inteltechniques.com
python sublist3r.py -d inteltechniques.com
python3 photon.py -u inteltechniques.com -l 3 -t 100
python3.6 theHarvester.py -d inteltechniques.com -b bing,google
```

```
#!/usr/bin/env bash
##Domain Interact Menu Script

#define choices
opt1="Amass"
opt2="Sublist3r"
opt3="Photon"
opt4="TheHarvester"

timestamp=$(date +%Y-%m-%d:%H:%M)
fqdnregex="\b((xn--)?[a-z0-9]+(-[a-z0-9]+)*\.)+[a-z]{2,}\b"

domainmenu=$(zenity  --list  --title "Domain Tool" --text "What do you want
to do?" --width=400 --height=300 --radiolist  --column "Choose" --column
"Option" TRUE "$opt1" FALSE "$opt2" FALSE "$opt3" FALSE "$opt4" 2> >(grep -
v 'GtkDialog' >&2))

case $domainmenu in
      $opt1 ) #Amass
            domain=$(zenity --entry --title "Amass" --text "Enter target
domain name" --entry-text "" 2> >(grep -v 'GtkDialog' >&2))
            if [ ! -z "$domain" ]; then
                  #verify legit domain
                  if [[ $domain =~ $fqdnregex ]]; then
                        #Run Tool
                        mkdir /home/osint/Documents/Amass/
                        amass intel -whois -ip -src -d $domain    -o
/home/osint/Documents/Amass/$timestamp-$domain.1.txt
                        amass   enum   -src   -brute   -d   $domain   -o
/home/osint/Documents/Amass/$timestamp-$domain.2.txt -d $domain
                        sleep 3
                        nautilus          "/home/osint/Documents/Amass/"
>/dev/null 2>&1
                        exit
                  else
                        zenity --error --text "Doesn't appear to be a
legitimate domain, exiting!" 2> >(grep -v 'GtkDialog' >&2)
                        exit
                  fi
            fi;;
                  $opt2 ) #Sublist3r
            domain=$(zenity  --entry  --title  "Sublist3r"  --text  "Enter
target domain name" --entry-text "" 2> >(grep -v 'GtkDialog' >&2))
            if [ ! -z "$domain" ]; then
                  #verify legit domain
                  if [[ $domain =~ $fqdnregex ]]; then
                        #Run Tool
                        mkdir /home/osint/Documents/Sublist3r/
                        cd /home/osint/Downloads/Programs/Sublist3r
                        python     sublist3r.py     -d     $domain     -o
/home/osint/Documents/Sublist3r/sublist3r_$domain.txt
                  else
```

```
                        zenity --error --text "Doesn't appear to be a
legitimate domain, exiting!" 2> >(grep -v 'GtkDialog' >&2)
                        exit
                fi
        fi;;
                $opt3 ) #Photon
        domain=$(zenity --entry --title "Photon" --text "Enter target
base URL (ex: https://www.inteltechniques.com)" --entry-text "" 2> >(grep -
v 'GtkDialog' >&2))
        if [ ! -z "$domain" ]; then
                #verify legit domain
                if [[ $domain =~ $fqdnregex ]]; then
                        #Run Tool
                        mkdir /home/osint/Documents/Photon/
                        cd /home/osint/Downloads/Programs/Photon/
                        python3 photon.py -u $domain -l 3 -t 100 -o
/home/osint/Documents/Photon/$timestamp-$domain
                        sleep 3
                        nautilus
"/home/osint/Documents/Photon/$timestamp-$domain" >/dev/null 2>&1
                        exit
                else
                        zenity --error --text "Doesn't appear to be a
legitimate domain, exiting!" 2> >(grep -v 'GtkDialog' >&2)
                        exit
                fi
        fi;;
                $opt4 ) #TheHarvester
        domain=$(zenity --entry --title "TheHarvester" --text "Enter
target domain name" --entry-text "" 2> >(grep -v 'GtkDialog' >&2))
        if [ ! -z "$domain" ]; then
                #verify legit domain
                if [[ $domain =~ $fqdnregex ]]; then
                        #Run Tool
                        mkdir /home/osint/Documents/theHarvester/
                        python3.6
/home/osint/Downloads/Programs/theHarvester/theHarvester.py -d $domain -b
baidu,bing,google,yahoo,censys,virustotal                               -f
/home/osint/Documents/theHarvester/$timestamp-$domain.html
                        if                              [               -f
/home/osint/Documents/theHarvester/$timestamp-$domain.html  ]; then
                                firefox
/home/osint/Documents/theHarvester/$timestamp-$domain.html
                        fi
                else
                        zenity --error --text "Doesn't appear to be a
legitimate domain, exiting!" 2> >(grep -v 'GtkDialog' >&2)
                        exit
                fi
        fi;;
esac
```

Below is the desktop shortcut for this utility, titled domains.desktop within the Ubuntu Files archive.

```
[Desktop Entry]
Type=Application
Name=Domains
Categories=Network;OSINT
Exec=/home/osint/Documents/scripts/domains.sh
Icon=/home/osint/Documents/icons/domains.png
Terminal=true
```

Place the "domains.sh" file into the "scripts" folder in your "Documents" folder next to the other scripts. Place the "domains.desktop" file directly onto your Desktop and type the following commands into Terminal.

- `cd Desktop`
- `sudo mv domains.desktop /usr/share/applications/`

Next, let's add a shortcut icon to our Dock in order to have easy access to this new utility.

- Open the Applications menu again and scroll down to "Domains".
- Right-click this program and select "Add to favorites".

In a later chapter, I explain investigative techniques for domain names, and the importance of searching for subdomains. A subdomain is a dedicated address that is part of a larger domain, such as pics.inteltechniques.com. While inteltechniques.com might be a main landing page, pics.inteltechniques.com could possess hidden content not available within the main website. Programs within this script such as **Amass** and **Sublist3r** request a domain name and then fetch subdomains associated with the target. This is a vital step for any investigation into a domain name. The following explains each utility in this new menu, including the results for a test query of cnn.com.

Amass: This option takes the longest to run, but it is the most thorough. It uses a brute force option which attempts to determine any possible subdomains. It creates two reports, both of which will be located in the Documents/Amass folder. In my test with cnn.com, it found 282 subdomains, such as the following.

footballclub.cnn.com
edition.cnn.com
fave.edition.cnn.com

Sublist3r: This program scans much faster, but will only find common subdomains. This may be sufficient for most tasks. It creates a report located in the Documents/Sublist3r folder. In my test, it found 808 subdomains of cnn.com such as the following.

trends.cnn.com
tours.cnn.com
coupons.cnn.com.

Photon: This option does not attempt to find subdomains. Instead, it searches for internal pages within a target website. It creates a report located in the Documents/Photon folder. Examples include the following.

https://cnn.com/webview/
https://cnn.com/NOKIA
https://cnn.com/Quickcast/

TheHarvester: This program searches a supplied domain with the intent of providing email addresses associated to the target. It creates a report located in the Documents/theHarvester folder. During my search of cnn.com, it located 72 hosts and 16 email addresses. Further investigation into these results may reveal new sources of information.

In each of these scenarios, you are greeted with text files containing numerous URLs. If desired, you could submit each to EyeWitness. This would create a new report with screen captures from any valid addresses. This allows for a quick review of the evidence to see if anything warrants further investigation.

We have now made it through the difficult section of this chapter. The remaining programs have standard launch options and do not require custom scripts to make them user friendly.

MediaInfo

This is a utility for displaying hidden metadata within a media file. The most common example is the metadata within a video obtained directly from the original source. First, open Terminal and type the following to install the application within the Applications menu of Ubuntu.

- `sudo apt-get install mediainfo-gui`

You can now click on the MediaInfo icon in the Applications menu to launch the program. Click on the "File" option in the menu and then open either a file or folder. The default view offers little data, so click on "View" and then "Text" from the menu. This presents all metadata available within the file. In my test, I received the (partial) output visible in Figure 4.05 from a video sent to me via email directly from a mobile device. It identified the make, model, and operating system version of the device, along with GPS information about the location during capture. Note that this type of data is usually not available from videos downloaded from the internet. The ideal scenario is that you possess a video file sent directly from a device.

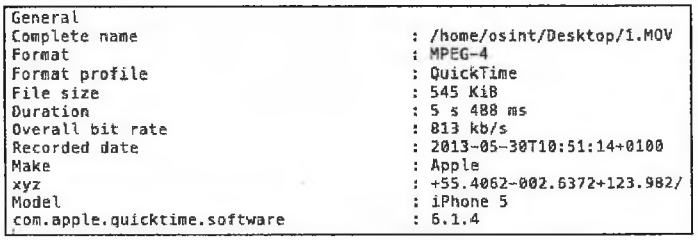


Figure 4.05: A MediaInfo result.

Exif Data

Image metadata, also called Exif data, will be explained later in more detail. It can include the make, model, and serial number of a camera or location data of the capture. Third-party applications to view this data are no longer necessary because Ubuntu has this feature embedded in the default image viewer. Simply open any image, click the three horizonal lines in the upper right menu, and select Properties. The Metadata and Details tabs display all available details such as camera information, location, and dates. However, this does not help much in regard to bulk images. We can easily install a Terminal-based utility which allows for automated export called **ExifTool**. The following command in Terminal will complete the process.

```
sudo apt install libimage-exiftool-perl
```

You can now export the metadata from multiple images into a spreadsheet. Consider the following example where I possess a folder of images on my Desktop titled "Evidence". I want a spreadsheet of all available Exif data in reference to these images. The following commands in Terminal will create a report on my Desktop titled Report.csv.

```
cd ~/Desktop/Evidence
exiftool * -csv > ~/Desktop/Report.csv
```

The second command launches ExifTool (exiftool), reads all files in the current folder (*), specifies creation of a CSV file (-csv), and saves the file output to the Desktop titled as Report.csv (> ~/Desktop/Report.csv). I do not use this feature often, but it has saved me a lot of work when needed. I once possessed a folder of 200 images retrieved from a directory on my target's blog. This command created a spreadsheet with all metadata from the images, which identified GPS locations valuable to the investigation.

HTTrack (httrack.com)

There are several ways to make an exact copy of a static website and this software will walk you through the process. You may want to do this when you locate a target website and are concerned the site could be taken down. Any time that you find any content that will be used in court, you should archive the entire site. This application automates the process. The result is a locally stored

copy in which you can navigate as if it were live. This is beneficial in court when internet access is not appropriate or a website has been taken offline. The following command within a new Terminal session will download the software and configure it within your operating system.

```
sudo apt-get install webhttrack
```

You can now type `webhttrack` at any Terminal prompt and the software will execute. The following is the desktop shortcut for this utility, titled httrack.desktop within the Linux files archive.

```
[Desktop Entry]
Type=Application
Name=HTTrack
Categories=Network;OSINT
Exec=webhttrack
Icon=/home/osint/Documents/icons/httrack.png
Terminal=true
```

Place the "httrack.desktop" file directly onto your Desktop and type the following commands into Terminal.

- `cd Desktop`
- `sudo mv httrack.desktop /usr/share/applications/`

The program is now available within the Applications menu. When the application loads, clicking the "next" button will bring you to the project screen. Give your project a title, preferably the website name, and choose a location to save all of the data. Click next and then "add URL". Enter the exact website that you want to archive. Do not enter any login or password. Click "next", then "finished", and you are done. The application will begin extracting all public information from the site. This can take a while, depending on the size of the site. When complete, you will have an exact copy in which you can navigate and search offline. Archiving to an external drive creates a replica that can be held for future analysis. This program only works on smaller static websites, and cannot archive large dynamic sites such as social networks. It is a very old program, and was never intended to function within today's latest technology. However, it still works surprisingly well on small business websites.

Google Earth Pro (earth.google.com)

Google maps is an online website that is discussed later. Google Earth Pro is a standalone application that takes the Google Map data to another level. With this application, we have access to many mapping tools. These tools can import data from spreadsheets and help you visualize the content. In order to maintain the scope of open source intelligence, I will focus on only a few specific tools. First, we need to enter the following commands within Terminal to install the software. Note the first three lines are one command.

- `wget -O google-earth64.deb http://dl.google.com/dl/earth/client/current/google-earth-stable_current_amd64.deb`
- `sudo dpkg -i google-earth64.deb`
- `sudo apt-get -f install; rm google-earth64.deb`

Within the application, the first step is to display your location of interest. This can be accomplished by typing the address or GPS coordinates in the upper left search field. When you see your target location and have set the zoom to an appropriate level, you are ready to start adding layers. By default, you will only see the satellite imagery of the location. The menu on the left possesses options for adding new content to this view. The last box in this menu is titled "Layers". Inside of this menu are several data sets that can be enabled and disabled by the checkbox next to each. I recommend disabling all layers and then enabling one at a time to analyze the data that is added to your map view. The following details will explain the layers of interest.

Photos - Digital images uploaded through social networking sites
Roads - Text layer of road names
3D Building - Alternative 3D view of some locations
Gallery - User submitted content including YouTube videos

Another Google Earth tool available that is often overlooked is the Historical Imagery option. This can be activated by selecting the "clock" icon in the upper menu bar of the application. This will open a slider menu directly below the icon. This slider can be moved and the result will be various satellite images of the target location taken at different times. Figure 4.06 displays a target area with the Historical Imagery option enabled. The view has been changed to the satellite image obtained on 05/30/2008. Usually, the quality of the images will decline as you navigate further back in time, as seen in Figure 4.07, through a satellite image obtained in 2000. This can be useful in identifying changes in the target location such as building modifications, additional vehicles, and land changes. Drug enforcement agents often use this tool to monitor suspected drug growth at a target location.

KeePassXC

I previously mentioned KeePassXC in previous chapters. It is a solid offline password manager which possesses an optional browser plugin. If you choose to use this within your VM, you will need to install the software with the following command in Terminal.

```
sudo snap install keepassxc
```

This creates a shortcut in the Applications menu, and this software behaves identical to the examples previously provided. This is only required if you want to take advantage of the browser plugin for automated entry of credentials into websites.

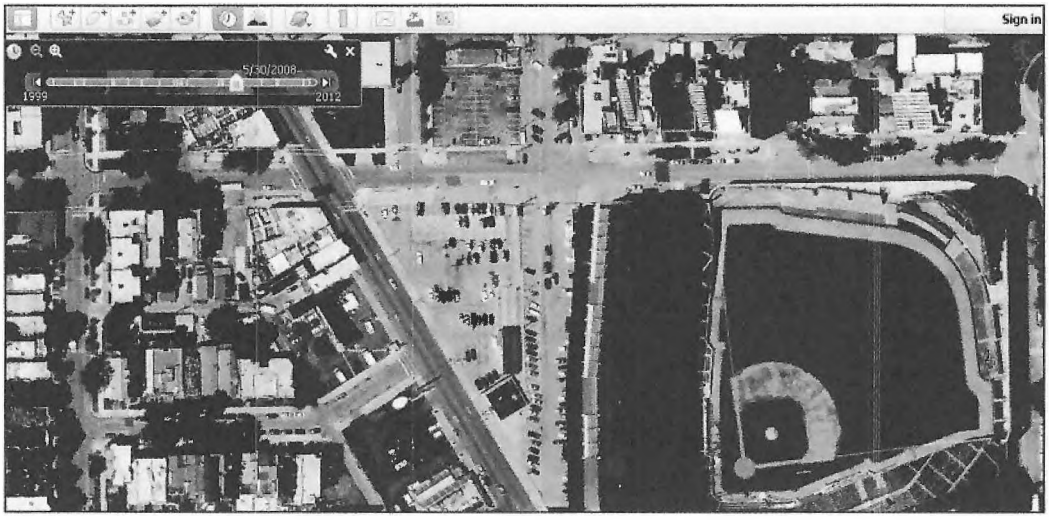


Figure 4.06: A Google Earth view of historic imagery from 2008.



Figure 4.07: A Google Earth view of historic imagery from 2000.

Screen Capture

I previously mentioned four unique methods of capturing website evidence within your browser. However, you may need a more robust solution for capturing your entire screen. This can be approached from two specific avenues. First, you could record either a still or video capture of your entire computer from your host. In other words, you could use the default capturing software within your Windows or Mac machine to record video or save a screenshot of your Linux VM. I do not recommend this. Recording from your host displays evidence of your other

operating system. While it would likely never be an issue, it exposes you unnecessarily. Consider the following.

You are a detective investigating a homicide. You find evidence online implicating your suspect. You are on a Windows computer, but all of your investigation was conducted within a Linux VM. You launch recording software on the Windows computer and record a video of your work. This video was submitted to the defense. For a brief moment, the video captured a file on your Windows desktop titled Accounts.xlsx. It is a spreadsheet containing all of your covert online accounts, has no connection to the investigation, and was not intended to be exposed. The defense makes a motion to analyze this file, as they believe it could be associated with the investigation. The judge approves, and you must share all of your covert accounts with the other side. Does this sound far-fetched? It happened to a colleague of mine in 2014.

I have become more paranoid of digital mistakes than necessary, but I believe we should never take chances. Therefore, I recommend that all of your screen captures be executed within your VM. Fortunately, Linux has many options. First, let's consider the default Ubuntu screenshot utilities. The following keyboard keys and key combinations create high-resolution exact screenshots of your VM.

- PrtSc: Save a screenshot of the entire screen to the "Pictures" directory.
- Shift + PrtSc: Select and save a screenshot of a specific region to the "Pictures" directory.
- Alt + PrtSc: Save a screenshot of the active window to the "Pictures" directory.
- Ctrl + PrtSc: Copy the screenshot of the entire screen to the clipboard.
- Shift + Ctrl + PrtSc: Select and copy the screenshot of a specific region to the clipboard.
- Ctrl + Alt + PrtSc: Copy the screenshot of the current window to the clipboard.

If you are using a Mac keyboard, the F13 key replicates "PrtSc". Some Windows keyboards may have a key labeled Print Scrn, Prnt Scrn, Prt Scrn, Prt Scn, Prt Scr, Prt Sc or Pr Sc. If you simply want a screenshot of your Desktop, such as evidence of your output within a Linux application, this is your best option. If you need video of your activities, we will need to install a third-party application.

Kazam

Kazam is a minimal tool for screen recording. It also includes screenshot support, but I find the native Ubuntu option easier and faster. Kazam is most suitable for getting the task done quickly without providing many options. The following steps will install and execute a Kazam capture.

- In Terminal, enter `sudo apt install kazam`.
- Launch Kazam from the Applications menu (make a shortcut if desired).
- Click "Capture" within the application.
- After the countdown, your entire screen is being captured.

- When finished, click the icon in the upper right and choose "Finish Recording".
- Choose "Save for later" and click "Continue".
- Choose the location for your recording, and click "Save" in the upper right.

You now possess a high-resolution video of your entire session. This can be very beneficial in many scenarios. I have used this to document a specific portion of my investigation when a simple still capture would not suffice. It could also be used to explain a search technique through video, which can be easily replicated by the person receiving a copy. I appreciate the simplicity of this application, and the lack of any branding or logos within the videos.

Troubleshooting

I do not expect everything presented here to work flawlessly for every reader. I wish it were that simple. You may encounter issues. My best advice is to restart the process with which you are having the trouble, and follow all instructions exactly as written. Even then, you may experience frustration. The following are some tips which have helped members of my online video training.

Scripts and Shortcuts

Throughout this chapter, I have explained ways to copy my digital files into your own VM. If you have done that, you are likely in good shape. If you are missing any, you will experience problems. My advice is to download a fresh copy of my Linux file archive from my website at https://inteltechniques.com/osintbook. Place this file on your Desktop within the VM. Right-click the file and choose "Extract Here". This will place respective folders titled scripts, icons, and shortcuts on your Desktop. Open Terminal and enter the following commands.

```
cd Desktop
cd shortcuts
sudo mv * /usr/share/applications/
enter password
cd ..
cd scripts
sudo mv * ~/Documents/scripts/
cd ..
cd icons
sudo mv * ~/Documents/icons/
```

You should now have all of the custom files within the appropriate folders. As a reminder, all of the commands required to install and configure all applications from this chapter are online at https://inteltechniques.com/osintbook. This allows for easier copy and paste functions.

Executable Files

If you downloaded my scripts and copied them to your VM, they should launch without any further steps. However, if you typed your own, they likely do not possess the authority to execute.

If you have ".sh" scripts in your "scripts" folder in your "Documents" folder, the following commands in Terminal will apply the appropriate execution rights to each.

```
cd ~/Documents/scripts
chmod +x *.sh
```

Hidden Files

By default, Ubuntu hides system files which should not normally be manipulated. You may need to see these within the Files application. The following steps allow you to see all files on the system, and the change is persistent between reboots.

- Open the "Files" program in either your Dock or Applications menu.
- Click on the three horizontal lines in the upper right area.
- Select "Show Hidden Files".
- Close the Files program.

Program Updates

After your initial installation of a Linux program, it may stop functioning. This is usually due to a change within the online service which is queried. Updating the program will usually resolve any issues, and this is explained in the next chapter titled VM Maintenance & Preservation.

Large File Transfer

We have configured our VM to possess a shared folder on the Desktop. This is an avenue to transfer files from within the VM to the host and vice versa. This can be used to copy evidence from within the VM to your host for preservation to external media. I have experienced limitations with this method. If I have hundreds of large videos, I often receive errors when I try to copy all of them to my host through the shared folder. This is much more prominent in VirtualBox versus the premium option VMWare. The best solution I have found is to transfer evidence directly to a USB drive within the VM. If desired, you could choose this method and eliminate any shared folder on the VM.

When you insert a USB drive into your host computer, it is immediately read by your primary operating system. Your running VM does not see the device. Within the VirtualBox menu, you can choose "Devices", then "USB", and try to click on the USB drive if found. The issue you may face is that your host has the drive locked and will not allow the VM to take over. You can usually resolve this by ejecting the USB drive within your host. On both Windows and Mac, you should see an eject icon next to the drive or by right-clicking the device. After ejecting, attempt to load the USB device again through the VirtualBox menu. This may present another issue. If your USB device is formatted specifically for Mac or Windows, it may not be readable by Linux.

I suggest specifying a USB device solely for use within this VM, and formatting it in a way so that it can be read universally by any operating system.

- Insert the USB device into your computer while the VM is running.
- Eject the device from your host OS if necessary.
- In the VirtualBox VM window, click on "Devices", "USB", and the device name.
- In Ubuntu, launch the Applications menu.
- Type "Disks" into the search field.
- Click on the USB drive listed in the left column.
- Click on any boxes within the "Volumes" section.
- Click the minus (-) icon and "Delete" when prompted.
- Repeat until there are no volumes present.
- Click the "+" icon, click "Next", provide a name, choose "Fat", and click "Create".

This will create a USB drive on which files can be read and written within any file system. You cannot possess files larger than 4GB in size, but I have never had an issue with that during investigations. When you insert the USB device into your computer, you can choose to allow your primary OS see it, or load it into your VM. I have found this method to be much more reliable than transferring large files through the shared folder. You should practice both techniques and pursue the most appropriate option for your investigations.

New Programs

We have only scratched the surface of the possibilities within a Linux virtual machine. New OSINT programs and scripts emerge weekly. Hopefully, this chapter has given you the foundation required to understand the installation and configuration of any future applications. The methods here can be replicated to create your own custom Linux build which works best for your investigative needs.

Linux Tips

- When typing in Terminal, you can hit the tab key and it will suggest all the possible options that start with string you have typed so far. It will also auto-complete your command if only one option exists. For example, if you are trying to navigate to your Downloads/Programs folder, you can just type cd ~/Dow [tab], then "Pro" [tab] to complete the command. Typing cd ~/D [tab][tab] would list all folders starting with D.
- Use the "up" arrow key to navigate through previous commands.
- "ctrl-c" kills any running process and "ls" displays the contents of a directory.
- You can copy and paste to/from Terminal, but only with right-click. "ctrl-v" is unstable.
- Keyboard arrows will move you through a Terminal command, but mouse clicks do not.
- You can modify the size of your Dock icons in Settings > Dock.

CHAPTER FIVE
VM MAINTENANCE & PRESERVATION

You now likely possess a secure computer, a custom Linux VM, several Linux applications, and shortcuts to easily launch each with user prompts for data. You could begin online investigations now, but I have a request. Delete it all. That's right, let's start over, but cheat a bit. Think about the VM you built. Did you have any issues installing software? Did you make any mistakes? Did you test the new applications with your own user data? It is likely that you slightly contaminated your perfect virtual machine while learning the process. I do not expect you to start over at the beginning of the book, but I do ask that you now repeat the steps with a new machine. Consider the previous chapters an education, and now we will generate a new clean machine.

In the following pages, I display only the abbreviated steps taken throughout the previous instruction. There are no explanations or demonstrations. Consider this the cheat sheet to replicate our previous work. `All Terminal commands are in this font.` Afterwards, I will explain many considerations for updates, backups, and locking in your "Master VM". After this, we can finally start to dig into some investigation techniques. I promise this will be much easier now that you have worked through the process. The entire list here is available digitally on my website at https://inteltechniques.com/osintbook/linux.txt. It should be easier to copy and paste commands versus typing them directly. **These steps will change as programs are updated! I will update this file as necessary on my website.** The linux.txt file overrides anything here.

- Open VirtualBox and delete any previous virtual machines created from this book.
- Within VirtualBox, click on the button labeled "New".
- Provide a name of "OSINT Master".
- Choose your desired location to save the machine on your host.
- Select "Linux" as the type, and "Debian 64-bit" as the version.
- Click Continue.
- In the Memory size window, move the slider to select 50% of your system memory.
- Click Continue.
- Click Create.
- Leave the hard disk file type as VDI and click Continue.
- Select the default option of "Dynamically allocated" and click Continue.
- Choose the desired size of your virtual hard drive (20GB+).
- Click Create.
- Click the Settings icon.
- Click the Storage icon.
- Click the CD icon which displays "Empty" in the left menu.
- Click the small blue circle to the far right in the "Optical Drive" option.
- Select "Choose Virtual Optical Disk File".

- Select the Ubuntu ISO downloaded previously.
- Click "OK".
- Click "Start" in the main menu.
- Select "Install Ubuntu".
- Select your desired language and location, then click "Continue".
- Select "Normal Installation", "Download Updates", and "Install third party..."
- Click "Continue".
- Select "Erase disk and install Ubuntu", then "Install Now". Confirm with "Continue".
- Choose your desired time zone and click "Continue".
- Choose a name, username, computer name, and password of "osint" for each.
- Select "Log in automatically".
- Allow Ubuntu to complete the installation, and choose "Restart Now".
- Press "Enter" to reboot.
- Press "Next" twice.
- Select "No" and then "Next" when asked to help improve Ubuntu.
- Click "Done" to remove the welcome screen.
- If prompted to install updates, click "Remind me later".
- In the VirtualBox Menu, select Devices > "Insert Guest Additions CD Image".
- Click "Run" when the dialogue box pops up, provide your password when prompted.
- Allow the process to complete, press enter, and shut down the VM (Upper right).
- In VirtualBox, select your VM and click "Settings".
- In the "General" icon, click on the "Advanced" tab.
- Change "Shared clipboard" and "Drag n' Drop" to "Bidirectional".
- In the "Display" icon, change the Video Memory to the maximum.
- In the "Shared Folders" icon, click the green "+".
- Click the dropdown menu under "Folder Path".
- Select "Other".
- Choose a desired folder on your host to share data back and forth.
- Select the "Auto-mount" option and then "OK".
- Click "OK" to close the settings window.
- Restart your Ubuntu VM. Right-click and eject the "CD" on your Desktop.
- In Terminal, type `sudo adduser osint vboxsf`
- Right-click on the Desktop and select "Change Background".
- Click the "Background" image and then "Colors".
- Choose a solid color of preference.
- Repeat the process with the "Lock Screen" image.
- Back in the "Settings" menu, click "Notifications" and disable both options.
- Click the "Privacy" option, then click "Screen Lock" and disable the option.
- Close the window and click "Usage and History", then disable the option.
- Close the window and click "Problem Reporting", then disable the option.
- Close the window and click Power, changing "Blank Screen" to "Never".

- Click "Automatic Suspend" and disable the feature.
- Close all open windows.
- Right-click any undesired icons from the dock and select "Remove from Favorites".
- Click the nine small dots in the lower left to launch the Applications menu.
- Type Terminal in the search field and click on the application.
- Type: `sudo apt purge ubuntu-web-launchers`, click "Y" when prompted.
- Click the nine dots to launch the Applications option.
- Type Terminal into the search field.
- Right-click on the application and select "Add to Favorites".
- Type Software into the search field and right-click on "Software Updater".
- Select "Add to Favorites".
- Press escape until all windows are gone.
- Launch the Software Updater icon from the Dock.
- Click "Install Now" and allow the updates to complete. Reboot.
- Open Firefox.
- Click on the menu in the upper right and select "Preferences".
- In the General options, uncheck "Recommend extensions as you browse".
- In the General options, uncheck "Recommend features as you browse".
- In the Home options, change "Homepage and…" and "New tabs" to "Blank page".
- In the Privacy & Security options, select the Strict option under Content Blocking.
- Check the box titled "Delete cookies and site data when Firefox is closed".
- Uncheck the box titled "Ask to save logins and passwords for websites".
- Change the History setting to "Firefox will use custom settings for history".
- Uncheck the boxes "Remember browsing…" and "Remember search…".
- Check the box titled "Clear history when Firefox closes".
- Uncheck "Browsing history" from the "Address Bar" menu.
- In the Permissions menu, click "Settings" next to "Location, Camera, Microphone…".
- Check the box titled "Block new requests..." on each of these options.
- Uncheck all options under "Firefox Data Collection and Use".
- Uncheck all options under "Deceptive Content and Dangerous Software Protection".
- Close all menus and return to Firefox.
- Enter "about:config" into the address field and hit enter.
- Search and change geo.enabled to FALSE.
- Search and change browser.safebrowsing.phishing.enabled to FALSE.
- Search and change browser.safebrowsing.malware.enabled to FALSE.
- Search and change media.navigator.enabled to FALSE.
- Search and change dom.battery.enabled to FALSE.
- Search and change extensions.pocket.enabled to FALSE.
- Search and change media.peerconnection.enabled to FALSE.
- Search and change media.peerconnection.turn.disable to TRUE.
- Search and change media.peerconnection.use_document_iceservers to FALSE.

- Search and change media.peerconnection.video.enabled to FALSE.
- Click the Firefox menu and choose "Addons".
- Search and Install Add-on titled Firefox Containers.
- Search and Install Add-on titled uBlock Origin.
- Search and Install Add-on titled Bulk Media Downloader.
- Search and Install Add-on titled VideoDownloadHelper.
- Search and Install Add-on titled FireShot.
- Search and Install Add-on titled Nimbus.
- Search and Install Add-on titled HTTPS Everywhere.
- Search and Install Add-on titled MJSONViewer.
- Search and Install Add-on titled User Agent Switcher.
- Search and Install Add-on titled Image Search Options.
- Search and Install Add-on titled OneTab.
- Search and Install Add-on titled Resurrect Pages.
- Search and Install Add-on titled Copy Selected Links.
- Search and Install Add-on titled KeePassXC Browser.
- `sudo snap install vlc`
- `sudo apt-get install ffmpeg` (press Y)
- `sudo apt-get install python-pip` (press Y)
- `sudo apt-get install python3-pip` (press Y)
- `sudo pip install --upgrade youtube_dl` (press Y)
- Download https://inteltechniques.com/osintbook/vm-files.zip to VM Desktop
- Right-click file and choose "Extract here".
- `mkdir ~/Documents/scripts`
- `mkdir ~/Documents/icons`
- `cd ~/Desktop/vm-files/scripts`
- `sudo mv * ~/Documents/scripts`
- `cd ~/Desktop/vm-files/icons`
- `sudo mv * ~/Documents/icons`
- `cd ~/Desktop/vm-files/shortcuts`
- `sudo mv * /usr/share/applications/`
- Delete vm-files file and folder from Desktop.
- Open Applications menu, and click "All" if necessary.
- Right-click Domains, EyeWitness, Instagram, Twitter, and Video Utilities/Download.
- Choose "Add to Favorites" for each.
- `cd ~/Downloads`
- `mkdir Programs`
- `cd Programs`
- `sudo pip install Instalooter`
- `sudo pip3 install Instaloader`
- `sudo apt install git` (press Y)

- `git clone https://github.com/twintproject/twint.git`
- `cd twint`
- `sudo python3 setup.py install`
- `cd ~/Downloads/Programs`
- `git clone https://github.com/ChrisTruncer/EyeWitness.git`
- `cd EyeWitness/setup`
- `sudo ./setup.sh`
- `sudo apt-get install xvfb`
- `cd ~/Downloads/Programs`
- `sudo snap install amass`
- `git clone https://github.com/aboul3la/Sublist3r.git`
- `cd Sublist3r && sudo pip install -r requirements.txt`
- `cd ~/Downloads/Programs`
- `git clone https://github.com/s0md3v/Photon.git`
- `cd Photon && sudo pip3 install -r requirements.txt`
- `cd ~/Downloads/Programs`
- `git clone https://github.com/laramies/theHarvester.git`
- `cd theHarvester && git checkout 8b88a66`
- `sudo pip3 install -r requirements.txt`
- `sudo apt-get install mediainfo-gui` (press Y)
- `sudo apt install libimage-exiftool-perl` (press Y)
- `sudo apt-get install webhttrack` (press Y)
- `wget -O google-earth64.deb http://dl.google.com/dl/earth/client/current/google-earth-stable_current_amd64.deb`
- `sudo dpkg -i google-earth64.deb`
- `sudo apt-get -f install; rm google-earth64.deb`
- `sudo snap install keepassxc`
- `sudo apt install kazam` (press Y)
- Shut down the VM from the menu in the upper right of the Desktop.

You should now have a new virtual machine titled OSINT Master. This is your clean machine with no contamination from any investigation or testing. It has all of the software we want, and it is ready to be used. Next, let's consider an "OSINT Test" machine. This is the VM which you can practice Linux commands, test new programs, or create new scripts. It is a VM which will never be used for any investigations. Its sole purpose is to give you a safe playground to experiment. Complete the following tasks within VirtualBox.

- Right-Click the VM titled "OSINT Master", click "Clone", and title it "OSINT Test".
- Supply the desired storage location and click "Continue".
- Select "Full Clone" and click the "Clone button".

You now have a fully functional cloned "test VM". Any activity within that machine will not change anything in any other VMs. Repeat this cloning process any time you wish to conduct an investigation. In this scenario, assume I created a new clone titled Case #2019-143. I can open this new VM which appears identical to my master. I can conduct my investigation and close the machine. All evidence is stored within the machine and can be extracted to my shared folder or USB drive if desired. All three of my VMs are visible in Figure 5.01. The Android VM visible at top is explained in the next chapter.

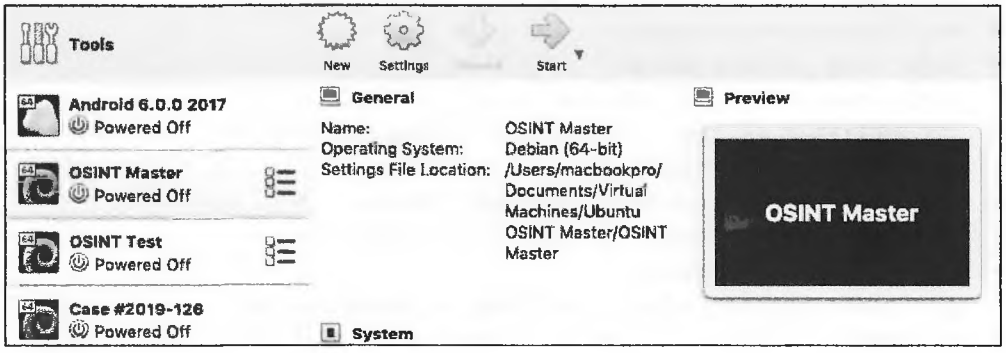


Figure 5.01: Cloned machines within VirtualBox.

Updates

Assume that you have not touched your virtual machines in some time, and you are ready to launch a new investigation. You likely have software updates which need to be applied. Instead, launch your OSINT Master VM and conduct the following within Terminal. These commands will update your operating system, installed applications, and custom programs created during the previous chapter.

- `sudo apt-get update`
- `sudo apt-get upgrade`
- `sudo -H pip install --upgrade youtube-dl`
- `sudo pip install instalooter -U`
- `sudo pip3 install Instaloader -U`
- `sudo pip3 install Twint -U`
- `cd ~/Downloads/Programs/EyeWitness`
- `git pull https://github.com/ChrisTruncer/EyeWitness.git`
- `cd ~/Downloads/Programs/Sublist3r`
- `git pull https://github.com/aboul3la/Sublist3r.git`
- `cd ~/Downloads/Programs/Photon`
- `git pull https://github.com/s0md3v/Photon.git`
- `sudo snap refresh`

I believe these commands should be executed as often as possible. You always want your Master VM to have the latest software. Manually entering each of these commands on a daily or weekly basis is exhausting. This is why I have created a script to update everything we have installed in our OSINT Master VM. This file is included within the scripts folder of the Linux files archive, and the exact content is below.

```
#!/usr/bin/env bash
##Updates Script

sudo apt-get update
sudo apt-get upgrade
sudo -H pip install --upgrade youtube-dl
sudo pip install instalooter -U
sudo pip3 install Instaloader -U
sudo pip3 install Twint -U
cd ~/Downloads/Programs/EyeWitness
git pull https://github.com/ChrisTruncer/EyeWitness.git
cd ~/Downloads/Programs/Sublist3r
git pull https://github.com/aboul3la/Sublist3r.git
cd ~/Downloads/Programs/Photon
git pull https://github.com/s0md3v/Photon.git
sudo snap refresh
```

You can execute this script with the following Terminal command.

- `sudo ~/Documents/scripts/./updates.sh`

Once you have your Master VM created, configured, and updated, it is best to export a copy as a backup. You have put a lot of time into this, and I would hate to see you lose the hard work. If your computer or storage device would crash beyond repair, you would need to start over. If your Master VM would become corrupted, restoring the data would be quite difficult. I keep an exported copy of my Master VM on a USB drive in case of emergency. In VirtualBox, conduct the following steps.

- Shut down the VM you want to export.
- Single-click on the VM within the VirtualBox menu.
- Click on "File" in the menu and then "Export Appliance".
- Confirm the selection and click "Continue".
- Choose the "File" export location on your drive and click "Continue".
- Click "Export".

This produces a single large file which contains your entire OSINT Master VM. You could import this file into any instance of VirtualBox with the "File" and then "Import Appliance" menu options. Let's take a look at my procedures for all of my VM's.

- I launch my OSINT Master VM weekly and apply all updates.
- I export my OSINT Master VM monthly as a backup.
- I conduct all software auditing and install new apps for testing on the OSINT Test VM.
- I create a new clone of the OSINT Master for every new investigation.
- At the end of the investigation, I export all evidence to an external drive.
- If necessary, I create an exported copy of investigation VMs for later review.
- I delete investigation VMs when no longer needed.

This plan ensures that every investigation is completed on a clean VM with absolutely no contamination from previous investigations. My exported investigation VMs can be provided to other investigators or as part of discovery in litigation. I am prepared to testify with confidence, if required.

Windows VM

While I prefer to conduct all investigations solely inside Linux, I respect that there may be a need for a Windows VM. In fact, I am using one now. I write all of my books within an offline copy of Microsoft Word. I create protected press-ready PDFs with Adobe Acrobat Pro. Neither of these applications run reliably on a Linux machine, and my personal laptop possesses Debian as the host operating system. Therefore, I keep a Windows VM for all writing. Installing Windows inside VirtualBox is not difficult, but licensing may be an issue. Therefore, we will rely on the official Microsoft Windows 10 VM available directly from their website.

- Navigate to https://developer.microsoft.com/en-us/microsoft-edge/tools/vms/.
- Choose "MSEdge on Win10" as the Virtual Machine.
- Choose "VirtualBox" as the Platform, and click "Download Zip".
- Unzip the downloaded file and keep a copy with your other VM backups.
- In VirtualBox, click "File" and "Import Appliance" in the menu.
- Choose the "ovf" file which you extracted from the zip file and click "Continue".
- Make any desired modifications as explained previously and click "Import".
- In the VirtualBox menu, click "Settings" then "Storage".
- Click the first "+" to add an optical drive, click "Leave empty", and click "OK".
- Before launching, create a snapshot of the VM as explained previously.
- Double-click the new Windows 10 machine to launch.
- Enter the password of "Passw0rd!" to enter Windows.
- In the VirtualBox menu, click "Devices" and "Insert Guest Additions CD".
- Through the Files explorer, double-click the mounted CD and choose "Yes".
- Click "Next", "Next", and "Install" to configure the default options.
- Reboot when prompted.

You now have a fully functioning and legal Windows 10 VM at your disposal. You can resize the window as desired and install any Windows applications. You can configure the copy and paste options, shared folder, or any other customizations as demonstrated previously. This is a 90-day trial, at which time the VM will no longer boot. You can revert to the original snapshot you created at any time to restart the 90-day trial. Surprisingly, this is allowed and encouraged from Microsoft. You will notice branding within the desktop advising it is a trial. If that bothers you, you must acquire a license and install from traditional media. This method is the easiest (and cheapest) option to possess a legal copy of Windows within a VM at no cost.

If you prefer to possess a valid license on a budget, you can find Windows 10 Long-Term Servicing Branch (LTSB) licenses on eBay for $5.00 each. LTSB is a specialized edition of Windows 10 Enterprise that promises the longest intervals between feature upgrades of any version of the operating system. While other Windows 10 servicing models push feature upgrades to customers every six months, LTSB does so only every two or three years. That means fewer changes during a set timeline, a less-involved upgrade effort, and fewer disruptions as well as fewer possibilities for applications breaking because of a modification of the OS. If you want a long-term Windows 10 solution, I recommend this option over all other licenses. It does not include the Edge browser, any Microsoft Store apps, or the Cortana voice-activated digital assistant. I view this as a benefit.

Hopefully, you now have a virtual machine in pristine condition ready for your investigations. Keep your Master clean and only use it for updates. Make clones of it for each investigation. Make sure you have a safe backup copy stored outside of your primary device. Most importantly, understand the steps you have taken. It is very likely that you will need to modify some of these commands as things change from the software developers. If you should receive an error on any specific step in these tutorials, search that exact error online and you should be presented many solutions. I send a final and sincere "Thank you" to David Westcott for opening my eyes to the many ways in which Linux can be customized for our needs as online investigators.

Chapter Six
Android Emulation

For several years, online researchers have been navigating through various social networking websites for information about individuals. Whether it was older sites such as Friendster and Myspace, or newer networks such as Twitter and Facebook, we have always flocked to our web browsers to begin extracting data. Times have changed. Today, an entire generation of social network users rarely touch a traditional computer. They operate completely from a cellular telephone or tablet. Many of the networks through which individuals engage will only operate on a mobile device. Services such as Snapchat, Tinder, and Kik do not allow a user to access content from a traditional web browser. As this shift occurs, investigators must transition with it. Our preparation is not complete until we have disposable Android environments in place.

This chapter will focus on the huge amount of information available through mobile platforms that is not accessible through a web browser. I will explain a method of emulating a portable device within a traditional computer. Before we dive into the nuts and bolts of making things work, we should discuss why emulation is the way to go. In my investigations, documentation is my primary reason for launching a simulated mobile device within my computer operating system. If I conducted my investigation on an actual smartphone, documenting my findings can be difficult. Mobile screen captures only cover a small amount of visible content. Extracting any captured images can be a hassle. Referencing my findings within a final report can become very tedious. When using Android emulation within my traditional computer, I can easily create numerous screen captures, record a video of my entire investigation, and paste my results directly into the report.

Privacy and security are also important reasons to consider emulation versus directly investigating from a portable device. I have seen many law enforcement investigators conduct a search or use an app directly from their personal or work phones. This opens that device to scrutiny and discovery. An attorney could rightfully request a copy of the investigator's phone in order to conduct an independent forensic analysis. That would make most people nervous. Additionally, if I encounter malicious software or a virus from my portable device, it could affect all future investigations using that hardware. Emulation will remedy both of these situations.

The idea of Android emulation is to recreate the mobile operating experience within an application on your computer. This application will execute in the same manner that your web browser, word processor, or email client would open. It will have the exact same appearance as if you were staring at a telephone or tablet. Any actions that you take within this emulated device will not affect anything else on your computer. Think of it as an encapsulated box, and nothing comes in or gets out, very similar to our Linux VM explained previously. A great feature of emulation is that you can create unlimited virtual devices. You could have one for every investigation in order to prevent any contamination.

Some readers will question why I chose to explain Android emulation instead of iPhone. The most obvious reason is the number of options. I will explain software solutions for recreating the Android environment on your computer. The iPhone simulator will only function on Apple computers and has very limited features. The Android techniques will work on any major operating system. Additionally, we can create Android virtual machines that possess all original functionality. The iPhone simulator will not connect to most applications and features. There are more options for Android emulation than what I present in this chapter. My goal is to focus on the most user-friendly and feature rich solutions that are available without cost. My overall emulator of choice is Genymotion.

Genymotion (genymotion.com/fun-zone)

This application-based solution is extremely easy to use. It works with Windows, Mac, and Linux operating systems. I will provide details for the Windows installation, but the principles apply across all platforms. The operation of virtual devices after installation is identical on all operating systems. First, you will need to install the application.

First, you must create a free account online at genymotion.com. This can be all alias information, and the login will be required in order to fully use the application. After you have created the account and successfully logged in to the site, navigate to genymotion.com/fun-zone and click on the "Download Genymotion Personal Edition" link. This presents the standard download page for Windows, Mac, and Linux. If prompted, choose the version without VirtualBox, as you should already have that program installed. Executing the download and accepting all default installation options will install all of the required files. When the setup process has completed, you will have a new icon on your desktop titled Genymotion. This entire process should occur on your HOST operating system, and not within a virtual machine.

Execute this application and note that an Android virtual machine may already be pre-installed and ready for launch. Instead of accepting this default option, consider creating your own machine in order to learn the process for future investigations. I recommend deleting this machine by clicking the menu icon to the right of the device and choosing "Delete". Perform the following instructions in order to create your first custom Android devices.

- In the left menu, expand the "Android API" menu and select the highest number. My option was 9.0 at the time of this writing. On the right, choose the device. I chose "Google Pixel XL" since I have a high-resolution screen, and clicked "Add custom device". You may want to choose a device with a smaller screen for your hardware.
- Rename this device similar to Android 9.0 Master. Change the "Android Version" to the highest option and click "Install". This will download and configure the device for immediate use, and can take several minutes.
- Launch the new device by double-clicking the new machine present in the Genymotion software. The machine will load in a new window which should appear similar to the

screen of an Android telephone. Click "OK" to any feature notifications. Figure 6.01 (left) displays the default view of the home screen of version 9.0.0.

- Navigate within the Android emulator by single-clicking on icons and using the "Back" icon in the lower left that appears similar to a left facing arrow.
- Consider the following customizations to improve the look and feel of the device. Figure 6.01 (right) displays the view of the home screen after these configurations.
- Drag any app icons up and drop them in the "Remove" option.
- Click and hold the bottom of the screen and drag up to view installed applications.
- Drag the Settings icon to your home screen and open the app.
- Choose "Display", then "Sleep", and select "30 Minutes".
- Choose "Security", then "Screen Lock", and choose "None".
- Press and hold the main window, select Wallpaper, and change if desired.
- Shut down the device and open VirtualBox.
- Similar to the VM settings, change the Video Memory to the maximum.
- Change the Memory size to half of the system resources.
- Relaunch your device from within the Genymotion application.

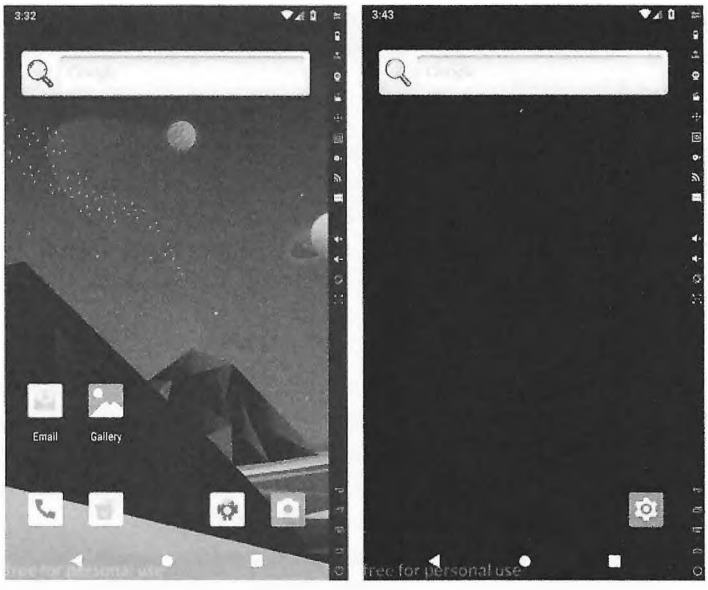


Figure 6.01: A default Android screen (left) and the custom version free of clutter (right).

You should now have a functioning replica of a standard Android device. However, you are missing several features. The biggest void is the absence of key applications such as Google Play and Gmail. Without core Google services, you cannot download apps to your device as part of your investigation tools. This has been the biggest hurdle with emulation. Consequently, there is

finally an official fix, and an alternative option for advanced users. First, let's try the easy way by using the Genymotion built-in Google features.

- While inside our virtual Android device, click the "Open GAPPS" icon in the upper right corner. Accept the agreement and allow Google Apps to install. Select the option to restart the devices.
- Your browser should open to https://opengapps.org/?source=genymotion. Select "ARM64", the version of the device that you created (9.0.0), and "Stock". Click the red download option in the lower right and save the large file to your Desktop. Do NOT open the downloaded zip file.
- Drag-and-Drop the downloaded zip file into your running Android device. Accept any warnings. You may receive errors. When complete, close and restart the device.

You should now have the Google Play Store in your applications menu. Launching it should prompt you to connect to an existing or new Google account. Consider using an anonymous account that is not used for anything else. I do not recommend creating a new account from within this virtual machine because Google will likely demand a cellular telephone number for verification. I prefer to create Google accounts from a traditional computer before connecting to the virtual Android device. After syncing with an active Google account on your new device, you should now be able to enter the Google Play Store. You should also now see all core Google services in your applications menu.

Many users have found this official way to fail them on occasion. I have found that many popular apps will not load with the most current versions of these utilities. Therefore, I offer an alternative option. Consider creating a second virtual device to have in case of future problems. The following instructions will restore the Play Store, emulate a more appropriate ARM driver (which will make some apps work better), and execute a patch that will eliminate those annoying Google crashes that have plagued this method for years.

- Click only the "6.0 API" filter in the Genymotion menu and double-click "Custom Phone 6.0.0 – API 23". Rename and allow the device to be created. Launch it.
- Download the "ARM Driver" at inteltechniques.com/android. Drag and drop the zip file directly into your running virtual Android device. Agree to the warning, and acknowledge the completion. Close the device and restart.
- Download the "Google Play version 6.x" file at inteltechniques.com/android. Drag and drop the zip file directly into your running virtual Android device. Agree to the warning, and acknowledge the completion. Close the device and restart.
- Log in to a Google account that you will use to download apps to the device. Close any errors that appear. Close the device and restart.
- Download the Benzo Patch file at inteltechniques.com/android. Drag and drop the zip file directly into your running virtual Android device. Agree to the warning, and acknowledge the completion. Close the device and restart.

You should now have a fully-functioning Android 6 device with Google Play and no errors. Please consider your experience with each installation. Android 9 is much larger in size, flashier, and overall polished. Android 6 appears a bit more dated, but is more responsive to interaction. Personally, I prefer Android 6 over 9 for my virtual machine investigations. It simply feels better, but your mileage may vary. My best advice is to have both ready to go at all times.

You can now install any apps within the Play Store. If any apps refuse to install because of an incompatible device, download the desired app from **APK Pure** (apkpure.com) and drag and drop it into the machine. The addition of Google Play will allow you to natively install Android applications as if you were holding a real telephone or tablet. Launch Google Play and you will be able to search, install, and execute most apps to your new virtual device. After you install a new program, click on the applications menu. Click and hold the new app and you will be able to drag it to your home screen. Figure 6.02 displays the screen of my default investigation emulator. I will later explain how these programs can be used for intelligence collection. First, you should understand the features embedded into the Genymotion software.

When you launch an Android virtual machine, you will see a column on the right side of the window and a row of icons horizontally on the bottom. The bottom icons are part of the emulated Android system. Clicking the first icon will navigate you backward one screen from your current location. If you are within an app, this would take you back one step each time that you press it. The second icon represents the "Home" option and will always return you to the home screen. The third button is the "Recent Apps" option and it will load a view of recently opened applications. The icons on the right of the emulator are features of Genymotion and allow you to control aspects of the Android machine from outside of the emulator. The following page displays this column of options, which should help explain each of these features. Note that many features are not available in the free version, but I have never found that to be a hindrance to my investigations. Genymotion is quite clear that if you plan on making money by designing an app through their product, you should pay for a license. Non-commercial usage allows unlimited use of the free personal version.

The GPS option within Genymotion is the most beneficial feature of their toolset. Clicking this icon and clicking the Off/On switch will execute the location spoofing service. You can either supply the exact coordinates directly or click on the "Map" button to select a location via an interactive Google map. Figure 6.03 (left) displays the default GPS menu in the disabled state. Figure 6.03 (right) displays coordinates entered. I recommend changing the altitude, accuracy, and bearing settings to "0". Close this window and you will see a green check mark in the GPS button to confirm that your location settings are enabled.



GAPPS Indicator: Confirms Google Services are installed.

Battery Indicator: It does not have any impact on your virtual machine.

GPS: Enable and configure the current location reported to the device.

Webcam: Use your computer's webcam for live video within an app.

Screen Capture: Not available in the free version.

Remote Control: Not available in the free version.

Identifiers: Not available in the free version.

Disk I/O: Not available in the free version.

Network Configuration: Not available in the free version.

Phone: Not available in the free version.

App Sharing: Not available in the free version.

Volume Up

Volume Down

Screen Rotate: Flip your view into horizontal mode similar to a tablet.

Pixel Configuration: Not available in the free version.

Back Button: Moves back one screen from current app location.

Recent Apps: View recently opened applications.

Menu: Simulates the "Menu open" option within an application.

Home: Returns to the Home screen.

Power: Shuts down the device.

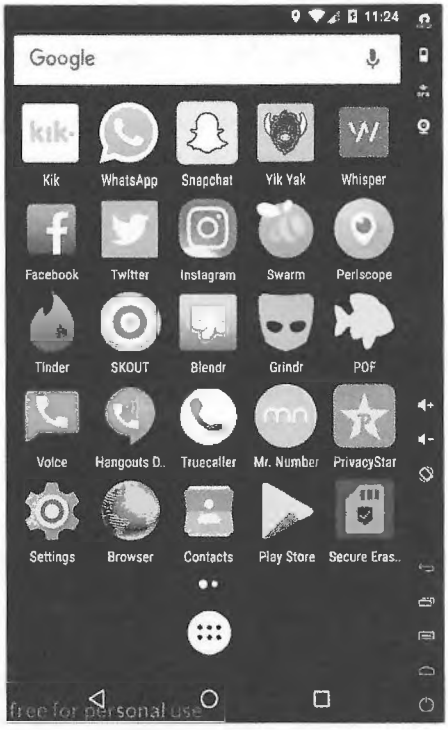


Figure 6.02: A custom Android emulator home screen with several apps installed into groups.

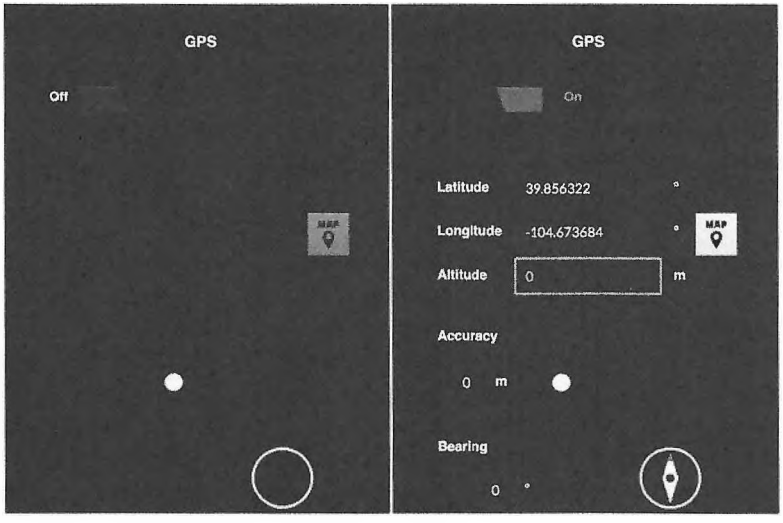


Figure 6.03: A disabled Genymotion GPS menu (left) and spoofed GPS (right).

Now that you have selected a location to broadcast through your device, you should test this configuration. My preferred way of doing this is to open Google Maps within the Android emulator and click the small blue target inside the white circle in the lower left. This will zoom Google Maps into the location where it believes you are located. You could also load the Bing Maps web page within a browser and ask it to center to your location. With both of these options, you may be prompted to "Allow" or "Deny" the device from obtaining your location details. You must choose "Allow" for this to function. After you have your desired location configured and you have confirmed accuracy, you can start to put this feature to work. The following tutorials will explain how location-aware applications could be used in investigations.

Facebook: The Facebook app on Android will appear similar to a compressed view of a standard profile page. The benefit of the mobile app is the ability to check into places. When you click the "Check In" option, Facebook will present businesses near your current spoofed location. With my configuration, Facebook presented the terminals and airlines at the Denver airport. If you choose a location, and create a post on your timeline, Facebook will verify that you were there. I have used this when I need to portray that I am somewhere I am not. This method can help you establish credibility within your pseudo profile. You could easily create the illusion that you were working at a business all day or out clubbing all night.

Real World Application: I once helped a domestic violence victim confuse her ex-husband with this technique. I posted from her Facebook account "accidentally" leaving my spoofed location enabled. He stalked her every move online. After wasting his time going to random places trying to find her, and always finding the location to be closed, he began doubting the information that he uncovered about her whereabouts.

Twitter: The first time that you use Twitter within your Android environment, you will be asked if you want to share your location. While I usually discourage this type of activity, sharing your spoofed location can have many benefits. Similar to Facebook, you can make yourself appear to be somewhere which you are not. You may want to confuse your target. If you know that he or she will be monitoring your social networks using the techniques in later chapters, this method should throw them off and be misleading.

Tinder: This dating app relies on your location in order to recommend people in your area that want to "hook up". It will use your Facebook account in use on your device for the login credentials. The preferences menu will allow you to specify the gender, age range, and distance of the targeted individuals. Most people use this to identify members of their sexual preference within one mile of their current location. The users can then chat within the app. I have used this to identify whether a target was at home or another location. I have found that the most recent version of Tinder will not load on an emulated device. However, older versions from APKPure often work correctly. You may need to test those with both older and newer versions of Android.

Real World Application: During one investigation, I discovered that my target was a Tinder user. I set my GPS in my Android emulator to his residence. I could then search for men his age

within one mile and identify if he was at home. If I did not get his profile as a result, I could change my GPS to his work address or favorite bar. When I received his profile in the results, I knew that he was near the spoofed location. I could do all of this from anywhere in the world.

Bumble/Blendr/Badoo/Skout/Down: These dating apps use various databases of user profiles. They are similar to Tinder, but do not require a Facebook account. This could be an additional option for locating a target who uses dating apps. The same method applied to Tinder would work on these networks as well.

Real World Application: I once used these during a cheating spouse investigation. I connected with a covert female Facebook profile who was recently accepted as a "friend" with the suspected cheating spouse. Launching the Down app confirmed that he had an account. Swiping "Down" on his profile alerted him that I wanted to "get down" with him. This quickly resulted in a very incriminating chat that was later used in litigation. In addition to identifying the location of targeted individuals, these apps could be used to identify people who are currently at a crime scene or gathering. I once used this technique to simply document people who were present near a state capitol during a credible bomb threat. When these people denied their presence during interviews, I had data that disagreed with their statements. Those who were lying quickly recanted their false statements and saved investigators a large amount of time.

Snapchat: As of 2019, I am no longer able to execute Snapchat on Genymotion. The latest versions block this behavior, and older versions require an update before launch. The only limited success I have had is with Android Studio which exceeds the scope of this book. Even when I was able to force the program to function, constant errors prohibited actual use.

Kik Messenger: Kik is an instant messaging application for mobile devices. It is modeled after BlackBerry's Messenger and uses a smartphone's data plan or Wi-Fi to transmit and receive messages. It also allows users to share photos, sketches, mobile webpages, and other content. You must create a free account within the app and you can then search any username or Kik number. Many users do not share personal details, but you can still use the app during your investigation for covert communication with a target.

Real World Application: Child exploitation is prominent on Kik Messenger. Pedophiles have been quoted in news sources stating "I could go on it now and probably within 20 minutes have videos, pictures, everything else in between off the app. That's where all the child porn is coming off of" and "I can get anybody I want. I can achieve my sexual desires through this app". In 2014, a parent confiscated her 15-year-old daughter's cellular telephone after it was discovered that the minor was sending nude photos of herself to an older man at his request. I was able to use my Android emulator to log in as the child; continue conversations with the pedophile; and develop evidence to be used during prosecution. Documentation was easy with screen captures and screen recording.

WhatsApp: WhatsApp Messenger is an instant messaging app for smartphones that operates under a subscription business model. The proprietary, cross-platform app enables users of select feature phones to use the internet to communicate. In addition to text messaging, WhatsApp can be used to send images, videos, and audio media messages. Locations can also be shared through the use of integrated mapping features. It is the most globally popular messaging app with more than 800 million active users. You will need to create an account and provide a telephone number for verification. This number can be a cellular, landline, or VOIP number. I have had success using free Google Voice numbers. After you have an account, you can communicate directly with any target using the service. I have found that several of my targets refuse to converse over traditional text messaging, but freely text over WhatsApp. If you conduct any online covert operations, you should have this set up ahead of time.

Text Messaging: If you conduct online investigations and communicate with a suspect, it is very possible that you may be asked to send or receive a standard SMS text message. Since your virtual device does not possess a cellular connection, and it is not assigned a telephone number, there are no native opportunities for this activity. However, you can install Google Voice and Hangouts Dialer within your virtual device. Voice will allow you to send and receive SMS text messages, and Hangouts Dialer will allow you to make and receive voice calls at the number issued by Google Voice. With this setup, you can conduct all of your communications over the virtual device, and preserve all of the evidence within a single archive.

Caller ID Apps: A later chapter explains reverse caller ID services and how they can identify subscriber information associated with telephone numbers. There are several additional services that only support mobile use. Privacy Star is a powerful service that previously supported web search but now mandates that you install the app. After installation, you can search unlimited cellular and landline numbers in order to identify the owners. Other options include Mr. Number and True Caller.

Secure Eraser: As time passes, the size of your Android virtual devices will grow. System and app updates alone will increase the size of your files quickly. Much of this size is unnecessary. When these virtual machines download new data and update the files, the old files remain, and are not usable. Basically, your virtual devices start to take up a lot of space for no reason. Secure Eraser helps with this. On your master copy, after you have updated all of your software, launch Secure Eraser and change Random to 0000-0000. Click the start button and allow the process to complete. This will remove all of the deleted files. Restart your machine and then clone or export the device. The new copy will reflect the reduction of file size, but the master will still be large.

There are many other similar apps. Now that you have an idea of how to integrate mobile applications into your investigations, you can apply the same techniques to the next future wave of popular apps. Many social network apps have no association with location. This content can still have value to an investigation. Some apps, such as Kik, only function within a portable device. You cannot load a web browser on a traditional computer and participate with these networks.

However, you can access them from within your Android virtual machine. The following tutorials may help you find new uses for these popular apps.

Contact Exploitation

Mobile apps often urge users to invite friends into the app environment. When you first join Twitter, the app requests access to your contacts in order to connect you with "friends" who are also Twitter users. This is one of the most reliable ways which apps can keep you within their ecosystem. As investigators, we can use this to our advantage. I have found that adding my unknown target's cellular telephone number to the Android phone's address book will often obtain the following information relative to the target.

- Associated Facebook accounts (name) from the "Find Friends" feature.
- Google Play purchases and reviews (interests) from the Google Play Store.
- Associated Twitter accounts (name) from the "Find Friends" feature.
- WhatsApp usernames and numbers (contact) registered to the cell number.

Basically, entering a target's phone numbers and email addresses into your address book on an Android emulator forces many apps to believe that you are friends with the person. It overrides many authority protocols that would otherwise block you from seeing the connection from the real details to the connected profiles. Let's consider another example using the popular secure messaging program Signal. When I downloaded the Signal app, it wanted me to register a telephone number. I chose a Google Voice number and configured the app. I then added my target's cellular number into my Android contact list and asked Signal to search for friends. Signal immediately confirmed that my target was active on Signal. This alone is valuable in regard to behavior, but not very helpful to establish identity. If I launch a new window to send a message to the number, even if I do not send the data, I may see a name associated with the account. This would need to be a deliberate act by the target, but this behavior is common.

Virtual Device Cloning

There are several beneficial features that are disabled in the free version of Genymotion. Options such as cloning, resetting, sharing, and renaming are restricted unless you purchase a premium license. The following tutorials replicate these missing features, and do so legally. Genymotion relies heavily of VirtualBox to function. We can access our Android virtual devices within VirtualBox for better control of our investigations. Similar to the earlier instruction about using a clean virtual machine for every investigation, you should consider a new Android virtual device every time to research a target. The steps taken previously may seem too complicated to execute every day, so you may want to maintain a master copy and clone it.

The paid version of Genymotion will allow you to clone any machine. This is very beneficial when you have a custom emulator that contains a lot of configuration. You can instantly make a copy of that machine and use a new version for each investigation. However, the free edition of

this software has disabled this feature. Instead, you will need to either manually create each machine that you want to use or clone the machine through VirtualBox. The benefit of manually installing and configuring your virtual devices is that you will keep your skills sharp. It should only take you ten minutes to create a new machine and incorporate the Google core services. However, that may be too time consuming if you use a new device for every investigation. The following instructions will clone the exact state of any virtual Android device within Genymotion.

- Create and customize an Android virtual device as desired. Configure Google Play and any other apps that you want present in all cloned copies. Optionally, execute the app "Secure Eraser" to eliminate unnecessary hard drive space. Exit the machine and close Genymotion completely.
- Open VirtualBox from your Applications folder (Mac) or Start menu (Windows). You should see the identical names of your Android machines visible within Genymotion. Right-click the machine that you want to duplicate and select "Clone". Figure 6.04 displays this program with a right-click menu option from an active machine.
- Provide a name for your new machine. This could be "Investigation Master Copy" or "2018-1234". Choose the options of Full Clone and Current machine state and click the Clone button. VirtualBox will create an exact duplicate of the chosen machine in the default folder for VirtualBox machines. You can identify this folder by right-clicking your new machine and choosing "Show in Finder" (Mac) or "Show on disk" (Windows).

You can now use this cloned device to conduct your investigation. Any changes made within it will have no impact on the master device. In fact, I title my master investigation devices "Master 9.0" and "Master 6.0". This way, I know to only open these to apply updates, and never for active investigations. Every time I need to use a device to research a target, I quickly clone the master and keep all of my cases isolated.

The presence of Android versions 6.0.0 and 9.0.0 may seem redundant. These numbers refer to the software version of the Android operating system within them. Possessing multiple versions is vital due to current and outdated apps. Most relevant apps will work fine on either version; however, some apps will not. I have encountered apps in the Play Store that would not allow installation under 6.0.0, but installed fine under 9.0.0. I have also witnessed the exact opposite. Having both versions available makes us prepared for either situation. You may be reading this book long after this writing and have more options present. Ultimately, I would have devices created for the most recent version and at least one previous version. If it is the year 2025, this may all be irrelevant.

Virtual Device Export

You may be asked to provide all digital evidence from your investigation as a matter of discovery. This could happen to a forensic examiner hired in a civil case or law enforcement prosecuting a criminal case. This is the precise reason that I create a new virtual device for all of my investigations. Not only is it a clean and fair environment, it is easy to archive and distribute when

complete. The following instructions will generate a large single file that contains the entire virtual operating system and apps from your investigation.

- Exit Genymotion and open VirtualBox in the same manner as mentioned previously.
- Select the target virtual device, click on "File" in the menu bar, and select Export Appliance. Select the device again and provide the save location and name of the file.
- Click Export and allow the process to complete. The final result will consist of a single file that can be archived to DVD or flash media.
- This file can be imported into VirtualBox by choosing the Import Appliance option in the File menu. This would allow another investigator to view the exact investigation environment as you.

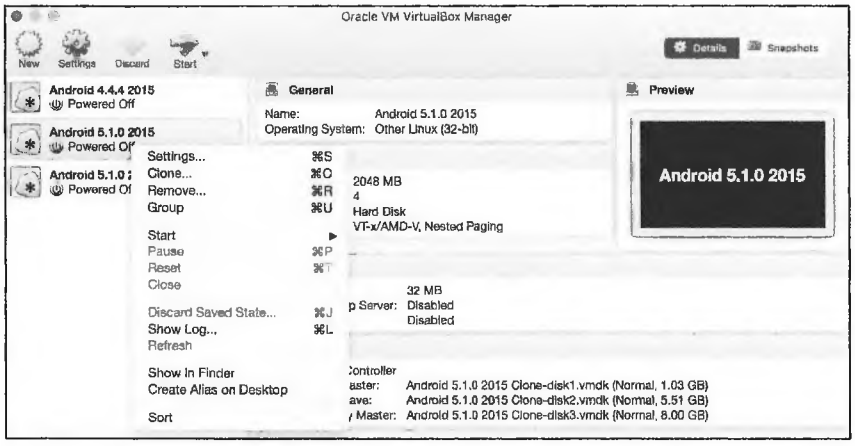


Figure 6.04: A VirtualBox menu with a clone option in the menu.

Genymotion is not your only option. **BlueStacks** (bluestacks.com), **Andy** (andyroid.net), and **NoxPlayer** (bignox.com) all offer the same basic functionality. After installation, most of these programs work the same way as Genymotion. I choose Genymotion over these because of the ability to import and export evidence as a virtual machine. While the others have their own backup and export options, I find the options presented here to be more transparent and acceptable in court. I encourage you to experiment with all of the options, and choose any that work best for you. Overall, I believe the future of OSINT collection will become more focused on mobile apps that have no website search option. In order to conduct thorough online investigations, mobile environment emulation is required. I highly recommend practicing these techniques with non-essential apps and data. This will better prepare you for an actual investigation with proper evidence control.

Chapter Seven
Custom Search Tools

I mentioned in the introduction that my online search tools were forced offline when my website was suspended in June of 2019. Part of my agreement with my web host, after receiving complaints from their abuse and legal departments, was that I would remove the interactive tools from public view. However, I never agreed to prevent others from creating their own sets or offering downloadable copies which can be run locally from any computer. The goal of this chapter is to help you create and maintain your own custom search tools which can automate queries for any investigation. First, let's talk about why this is so important.

I assumed my search tools would be around as long as I maintained my site. I learned the hard way that nothing lasts forever. We can no longer rely on third-party tools, a theme which I have overly-emphasized throughout this entire book. Any online search tool outside of your control could disappear any day. That is not the worst possible outcome. We never truly know what information various online search tools are storing about our searches and activity. Many aggregated search sites possess numerous tracking cookies and practically all "link collections" force embedded analytics capturing data about every visitor. Creating and hosting your own tools eliminate these issues. We still must query sensitive data to a final repository of information, but let's eliminate the middle-man. All of the tools presented in this chapter, and referenced throughout the book, do not need to be placed online within a website. You can store them on your computer and launch them without fear of questionable connections. Let's get started.

First, download a copy of all search tool templates used within the entire book. This can be found at https://inteltechniques.com/osintbook/tools.zip. Unzip this archive to a destination of your choice. If using a Linux virtual machine, I suggest saving them to the Desktop for easy access. If using your host computer, anywhere should suffice. Be sure to extract all of the files within the zip file. This collection reveals a folder titled "Tools" consisting of multiple files within it. Technically, you have everything you need to replicate my once public search tools. However, it is up to you to modify these as needed. You will eventually want to remove dead sources, add new features, and modify the structures due to changes at third-party websites. I will use my Email Search Tool as a demonstration. Figure 7.01 displays the current view of the Email tool.

As you can see, there are several individual search options and a "Submit All" feature at the bottom. Inserting an email address into any of these fields will query that address through the designated option, or the final field executes a search through all of them. Let's pick apart one of these queries from within the code. By default, double-clicking on any of the files within the search tool folder opens the selected option within your default web browser. This is required for any of them to function. In order to edit the files, we must open them within an HTML editing tool or any text processing application. If you are on a Mac, that could be TextEdit, Windows users have Notepad, and Linux users have Text Edit. All work fine for our needs.



Figure 7.01: The Email Search Tool available online.

If you open the file titled email.search.html within a text editor (File > Open), you will see the code which makes this document function within a web browser. Below are several lines.

```
<!DOCTYPE html>
<head>
<title>Email Search Tool</title>
</head>
<body>
<script type="text/javascript">
function doSearch01(Search01)
{window.open('https://haveibeenpwned.com/unifiedsearch/' + Search01,'Search01window');}
</script>
<form onsubmit="doSearch01(this.Search01.value); return false;">
<input type="text" name="Search01"id="Search01" size="30" placeholder="Email Address"/>
<input type="submit" style="width:120px" value="HIBP Breaches" /><br /></form>
```

Let's dissect each line in order to understand these commands.

```
<!DOCTYPE html>
```
-This informs your web browser that this is indeed a web page, even if offline.

```
<head>
```
-This informs your browser that the "head" portion of the page begins now.

```
<title>Email Search Tool</title>
```
-This represents the title of the page, visible in the browser tab.

```
</head>
```
-This discloses the end of the "head" section.

```
<body>
```
-This informs your browser that the "body" portion of the page begins now.

```
<script type="text/javascript">
```
-This identifies the following text as a JavaScript command.

```
function doSearch01(Search01)
```
-This tells the page we want it to "do" something, and the task is called Search01.

```
{window.open('https://haveibeenpwned.com/unifiedsearch/' + Search01,'Search01window');}
```
-This instructs the page to build a URL, add a piece of data, and open the result in a tab.

```
</script>
```
-This identifies the end of the script.

```
<form onsubmit="doSearch01(this.Search01.value); return false;">
```
-This creates a form to generate the URL, looking for a specific value.

```
<input type="text" name="Search01" " id="Search01" size="30" placeholder="Email Address"/>
```
-This creates a form input identified as Search01 with "Email Address" populated in the field.

```
<input type="submit" style="width:120px" value="HIBP Breaches" /><br /></form>
```
-This creates the Submit button with specific text inside, inserts a new line, and closes the form.

```
</body>
```
-This identifies the end of the "body" section.

```
</html>
```
-This is the closing command ending the page.

This only represents the first search option within this tool, but it is quite powerful. This collects a target email address and queries the website HaveIBeenPwned to identify known data breaches containing this account. This technique will be explained in more detail later in the Email chapter. This also demonstrates the need for a search tool versus simply visiting the search site. If you go to haveibeenpwned.com, you can enter an email address to conduct your search. The new page presented does not include a static URL for that search. The page with your results still possesses a simplified address of haveibeenpwned.com, and not something static and similar to haveibeenpwned.com/test@email.com. Bookmarking this page would not present the search results which you have just achieved. This is why I am using a different static address of https://haveibeenpwned.com/unifiedsearch/test@email.com. It presents the same content, but is a text-only format. Below is another example to explain this.

Conducting the search on the official site presents a graphical output similar to that seen in Figure 7.02. However, the static address I just mentioned presents a view from the HaveIBeenPwned API, as seen in Figure 7.03. The same data can be found in each offering, but the text view can be copied and pasted more easily. It also possesses a static URL which can be referenced in your report and recreated later. You may be wondering where this URL came from. It is not advertised on the site, and is not an official option within the API (which is now a paid service, but this URL is free). That is our next tutorial.

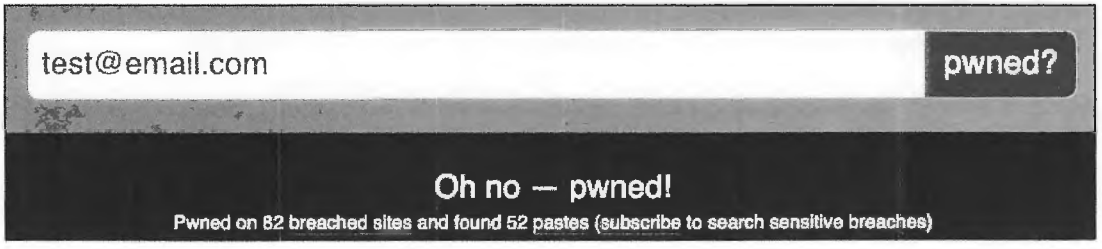


Figure 7.02: A result from the HaveIBeenPwned website.


```
Name:          "BitcoinTalk"
Title:         "Bitcoin Talk"
Domain:        "bitcointalk.org"
BreachDate:    "2015-05-22"
AddedDate:     "2017-03-27T23:45:41Z"
ModifiedDate:  "2017-03-27T23:45:41Z"
PwnCount:      581407
Description:   "In May 2015, the Bitcoin forum <a href=\"https://www.cryptocoinsnews.com/bitcoin-exchange-btc-e-bitcointalk-forum-
               genders, birth dates, security questions and MD5 hashes of their answers plus hashes of the passwords themselves."
LogoPath:      "https://haveibeenpwned.com/Content/Images/PwnedLogos/BitcoinTalk.png"
```


Figure 7.03: A result from the HaveIBeenPwned API.

Navigate to haveibeenpwned.com within Firefox and allow the page to load. Conduct the following steps to identify the exact URL which submits a query to return data about your target.

- Right-click on the page and choose "Inspect Element".
- Click the "Network" tab in the new window at the bottom of the page.
- Type an email address into the website and execute the search.
- Scroll through the new text in the Inspector window at the bottom of the page.
- Click on the result displaying the target email address with "xhr" in the "Cause" column.
- Copy the URL in the window to the right under "Headers" as seen in Figure 7.04.

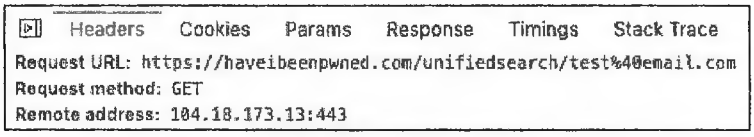


Figure 7.04: The static URL from a query as seen in Inspector.

With this method, we can identify the URL structures for our tool. In the example displayed previously, our tool presented a line of code which included the URL required for the search.

```
{window.open('https://haveibeenpwned.com/unifiedsearch/' + Search01,'Search01window');}
```

This line instructs the tool to open a new browser window, navigate to the website https://haveibeenpwned.com/unifiedsearch/, followed by whatever text was entered into the search tool, and define that new window (or tab) with a unique name in order to prevent another search within our tool from overwriting the page. This results in our tool opening a new tab with our desired results (https://haveibeenpwned.com/unifiedsearch/test@email.com). Let's look at another search option within this same tool with a slightly different structure.

The Dehashed search option is unique in that it requires quotation marks surrounding the input. In other words, you must enter "test@email.com" and not simply test@email.com. This requires us to add an additional character after the target input has been provided. Below is the example for this query. Note that double quotes (") are inside single quotes ('), which appears quite messy in print. Always rely on the digital files to play with the actual content.

```
{window.open('https://dehashed.com/search?query="' + Search05 + '"', 'Search05window');}
```

This line instructs the tool to open a new browser window, navigate to the website https://dehashed.com/search?query=", followed by whatever text was entered into the search tool, plus another quotation mark (a single quote, double quote, and another single quote), and define that new window (or tab) with a unique name in order to prevent another search within our tool from overwriting the page. The lesson here is that you can add as many parameters as necessary by using the plus (+) character. You will see many examples of this within the files that you have downloaded. Remember, every search tool presented in this book is included in digital format. You only need to modify the options as things change over time.

Now let's assume that Dehashed made a change to their search structure. Your tool currently possesses the following instruction.

```
{window.open('https://dehashed.com/search?query="' + Search05 + '"', 'Search05window');}
```

This is because the URL structure of the search is as follows:

https://dehashed.com/search?query="test@email.com"

Assume that Dehashed changed the structure to the following:

https://dehashed.com/?query="test@email.com"&trial

Your new line within the tool would need to be manipulated as follows:

```
{window.open('https://dehashed.com/?query="' + Search05 + '"&trial', 'Search05window');}
```

Submit All

Many of the online search tools offer a "Submit All" button at the bottom of the options. This executes each of the queries referenced above the button and can be a huge time saver. If you open one of the search tools with this option in a text editor, you will see the code for this at the bottom. It appears very similar to the other search options, but there are multiple "window.open" elements such as those listed below.

```
window.open('https://haveibeenpwned.com/unifiedsearch/' + all, 'Search01window');
window.open('https://dehashed.com/search?query=' + all, 'Search05window');
```

In each of the tools, I have simply replicated the individual search options within one single "Submit All" feature. If you modify a search tool within the code next to the manual search, you should also update it under the final option to execute all queries. If you feel overwhelmed with all of this, do not panic. None of this is required at this point. Your own custom offline search tools are already configured and functioning. If a specific desired tool stops functioning, you can use this chapter to change your scripts.

Now, let's assume that you found a brand new search service which was not included in the downloadable search tools. You will need to modify the tools to include this new option. Again, we will use the email tool as an example. Open the "Email.html" file within a text editor. Look through the text and notice that each search script possesses an identifier similar to Search01, Search02, Search03, etc. These must each be unique in order to function. You will notice that the final option (before the submit all feature) is Search 22. We now know that our next option should be Search 23. The final option (Search 22) is exactly as follows.

```
<script type="text/javascript">
function doSearch22(Search22)
{window.open('https://en.gravatar.com/site/check/' + Search22, 'Search22window');}
</script>
<form onsubmit="doSearch22(this.Search22.value); return false;">
<input type="text" name="Search22" "id="Search22" size="30" placeholder="Email Address"/>
<input type="submit" style="width:120px" value="Gravatar" /><br /></form>
```

Assume that you found the website leakprobe.net (which is a real site somewhat valuable for email investigations, as explained in the later Email chapter). You want to add it to the tools, but a search does not reveal a static URL. Repeat the process discussed earlier in order to identify the static address. Load leakprobe.net and conduct the following steps.

- Right-click on the page and choose "Inspect Element".
- Click the "Network" tab in the new window at the bottom of the page.
- Type an email address into the website and execute the search.
- Scroll through the new text in the Inspector window at the bottom of the page.
- Click on the result displaying the target email address with "xhr" in the "Cause" column.
- Copy the URL in the window to the right under "Headers".

In this scenario, the URL was the following:

https://leakprobe.net/ajax.php?query=searchLeaks&username=&email=test%40email.com

Since we are not searching a username, we can shorten this to:

https://leakprobe.net/ajax.php?query=searchLeaks&email=test%40email.com

You would next copy the "Search22" script and paste it at the end of the tool (before the Submit All feature). You can then edit the script, which should look like the following, using Search23 and our new URL.

```
<script type="text/javascript">
function doSearch23(Search23)
{window.open('https://leakprobe.net/ajax.php?query=searchLeaks&email=' + Search23,
'Search23window');}
</script>
<form onsubmit="doSearch23(this.Search23.value); return false;">
<input type="text" name="Search23" "id="Search23" size="30" placeholder="Email Address"/>
<input type="submit" style="width:120px" value="LeakProbe" /><br /></form>
```

All we changed within this copy and paste job was the target URL, the Search22 identifiers, and the descriptor. You can place this section anywhere within the tools, as it does not need to be at

the end. Spoiler alert: I have already added this script to the Email Search Tool. I placed it directly below the WeLeakInfo option. Take a look at my final code within the tool and see if you can identify the changes. Now that you know how to add a manual search entry, we must also include it within the Submit All feature. I added the following line into the tool for you to review.

```
window.open('https://leakprobe.net/ajax.php?query=searchLeaks&email=' + all, 'Search23window');
```

I encourage you to delete this new entry from the tool and see if you can replicate the work. Note it is titled Search23, so any new options added would need to start with Search24. These numbers do not need to be sequential throughout the tool, but they must be unique. If you decide to add a new feature to one of the tools, this lesson will assist. Finally, let's focus on using the tools.

You may have noticed that there are several files within the Tools folder. Launching any of these opens that specific tool, such as Email.html, but a menu exists within each of the pages in order to navigate within the tool to the desired page. The file titled "_Start_Here.html" is the "Main menu", and might be appropriate to set as your browser's home page. Clicking on the desired search option within the left side of the menu opens that specific tool. As an example, clicking on "Twitter" presents numerous Twitter search options. These will each be explained at the end of each corresponding chapter.

Populate All

You may have noticed that most of the tools have an option to populate all of the fields from a single entry. This is beneficial as it prevents us from copying and pasting target data within multiple fields. Below is the code within the first portion of the Email search tool. This tells your browser to populate anything you place into the first field within every field on that page which has an ID of "Search" plus any numbers. In other words, it would populate both examples on the previous page because they have "id=Search22" and id="Search23". Test this within the Email search tool. Make sure each "id" field is unique, as no two can be the same on one page.

```
<script type="text/javascript">
function doPopAll(PopAll)
{var pop = document.getElementById("PopAll");
for (j = 1; j <= 100; j++){
if (j < 10){
j = "0" + j.toString();}
else {j = j.toString();}
console.log(j)
item = document.getElementById("Search" + j);
if (item != null){
item.value = pop.value;}}}</script>
<form onsubmit="doPopAll(this.PopAll.value); return false;">
<input type="text" id="PopAll" name="PopAll" size="30" placeholder="Email Address"/>
<input type="submit" style="width:120px" value="Populate All" /><br /></form><br>
```

When I need to search a specific target, I do not copy the data into each search field and press the corresponding button for each service. I place the input directly into the "Populate All" option and then execute any individual searches desired. Alternatively, I place my target data into the "Submit All" option and let it go. If using Firefox, this will fail on the first attempt. This is because you have pop-ups blocked by default, and Firefox is trying to protect you from multiple new pages loading automatically. The following steps will prevent this.

- Open the Email.html search tool included in your downloaded offline search tools.
- Place any email in the last option and press the Submit All button.
- A new tab will open, but close it.
- Back in the Email search tool, you should see a yellow banner at the top.
- Click the Preferences button within the banner.
- Click the first option to "Allow pop-ups for file…"
- Confirm multiple tabs opened.

This will prevent your pop-up blocker from blocking that specific page. You would need to repeat the process for each of the other tools, such as Twitter, Facebook, etc., which can be quite a burden. If desired, you can disable the pop-up blocker completely, but that carries risks. You may visit a malicious website which begins loading new tabs. I do not see this as much as in years past, but the threat does still exist. If conducting your research within a VM, I do not see a huge risk in disabling this blocker. If you do, all of the tools will function without specific modifications to the blocker. Make this decision carefully.

- Click the Firefox menu in the upper right and choose Preferences or Options.
- Click on Privacy & Security and scroll to Permissions.
- If desired, uncheck the "Block pop-up windows" option.

While I would never do this on my primary browser used for personal activity on my main computer, I have disabled the pop-up blocker within my OSINT Master VM (and therefore all clones). It simply saves me headaches when trying to use automated tools. If only using the single queries within the tool, your pop-up blocker will not interfere. Regardless of whether you manually allow each tool to bypass the pop-up blocker, disable blocking completely, or leave it all untouched, I highly recommend that you become familiar with these search tools before you rely on them. Practice with various types of information. Experience how the URLs are formed, and understand how to modify them if needed. Each of these tools will be explained in the following chapters as we learn all of the functions.

Simplified Modification

I am sure some readers are frustrated at the technology presented here. Some may look at this code and cite numerous ways it could be made better. I agree this is amateur hour, as I am not a strong HTML coder. Other readers may be confused at all of this. For those, there are two

options which simplify things. First, ignore this entire chapter and simply use the free tools without any modification. Some options will break eventually as sites come and go, but that should not impact the other fields. Second, don't worry too much about adding new features. Instead, simply replace any searches that stop functioning. If the Dehashed shuts down tomorrow, simply wait for a replacement. When that happens, modify only the URL and name, leaving the structure as-is. You have a strong start with the current tools template. Very minimal modifications as things break will keep you in good shape.

License & Warranty

These tools are released to you for free under a Creative Commons Attribution-NonCommercial-ShareAlike license. Full details of any allowances and restrictions can be found online at https://creativecommons.org/licenses/by-nc-sa/4.0/. You are free to share and modify the content any way you desire, as long as you respect the following conditions.

Attribution: You must identify and link to IntelTechniques.com as the source, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. (The lawyers demand this one)

NonCommercial: You may not use the material for commercial purposes. (Do not sell these or offer them as part of a commercial product/website/training)

ShareAlike: If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. (This keeps the work free)

The software is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

As a final reminder, all tools can be downloaded from these primary and backup locations:

https://inteltechniques.com/osintbook/tools.zip
https://mega.nz/#!7HwmyYrA!LEIRTmoReD7J5J1F3spxA6Cu5f1dMROkiPTOJtv3gMc
https://app.nihaocloud.com/f/5df00ef7ecf442c9995e/?dl=1

The current version of these tools is 1.0. I have placed a text file disclosing the version within the download. If I make any changes, a log will be kept within this text file. Please follow by blog or Twitter account for any updates. I send a huge thanks to "Jesse" and Justin Seitz for their continuous help with my website education. Ultimately, it is now your responsibility to update your tools as desired. The torch has been passed. We have finished the tutorials to prepare our machines for our investigations. It is now time to jump into various search techniques.

Section II
OSINT Resources & Techniques

Some may consider this section to be the "guts" of the book. It contains all of the OSINT tips, tricks, and techniques which I have taught over the past twenty years. Each chapter was rewritten and confirmed accurate in October 2019. All outdated content was removed, many techniques were updated, and over 75 new resources were added.

The first four editions of this book only consisted of this section. Only recently have I adopted the preceding preparation section and the methodology topics toward the end. OSINT seems to have become a much more complex industry over the years. It is exciting to watch the community grow and I am honored to play an extremely small role.

This section is split into several chapters, and each explains a common type of target investigation. I have isolated specific topics such as email addresses, usernames, social networks, and telephone numbers. Each chapter provides every valuable resource and technique which I have found beneficial toward my own investigations. No book could ever include every possible resource, as many tools become redundant after a superior version has been identified. I do my best to limit the "noise" and simply present the most robust options for each scenario. This section should serve as a reference when you encounter a specific need within your own investigations.

CHAPTER EIGHT
SEARCH ENGINES

The first stop for many researchers will be a popular search engine. The two big players in the United States are Google and Bing. This chapter will go into great detail about the advanced ways to use both and others. Most of these techniques can apply to any search engine, but many examples will be specific for these two. Much of this chapter is unchanged from the 6th edition.

Google (google.com)

There are entire books dedicated to Google searching and Google hacking. Most of these focus on penetration testing and securing computer networks. These are full of great information, but are often overkill for the investigator looking for quick personal information. A few simple rules can help locate more accurate data. No book in existence will replace practicing these techniques in a live web browser. When searching, you cannot break anything. Play around and get familiar with the advanced options.

Quotation Marks

Placing a target name inside of quotation marks will make a huge difference in a quick first look for information. If I conducted a search for my name without quotes, the result is 176,000 pages that include the words "Michael" and "Bazzell". These pages do not necessarily have these words right next to each other. The word "Michael" could be next to another person's name, while "Bazzell" could be next to yet another person's name. These results can provide inaccurate information. They may include a reference to "Michael Santo" and "Barry Bazzell", but not my name. Since technically the words "Michael" and "Bazzell" appear on the page, you are stuck with the result in your list. In order to prevent this, you should always use quotes around the name of your target. Searching for the term "Michael Bazzell", including the quotes, reduces the search results to 34,000.

Each of these pages will contain the words "Michael" and "Bazzell" right next to each other. While Google and other search engines have technology in place to search related names, this is not always perfect, and does not apply to searches with quotes. For example, the search for "Michael Bazzell", without quotes, located pages that reference Mike Bazzell (instead of Michael). This same search with quotes did not locate these results. Placing quotes around any search terms tells Google to search exactly what you tell it to search. If your target's name is "Michael", you may want to consider an additional search for "Mike". If a quoted search returns nothing, or few results, you should remove the quotes and search again.

This search technique can be vital when searching email addresses or usernames. When searching the email address of "Michael@inteltechniques.com", without quotes, I receive 6,400 results.

When I search "Michael@inteltechniques.com" with quotes, I receive only 8 results that actually contain that email address. When your quoted search, such as "Michael Bazzell", returns too many results, you should add to your search. When I add the term "FBI" after my name, the results reduce from 34,000 to 12,000. These results all contain pages that have the words "Michael" and "Bazzell" next to each other, and include the term "FBI" somewhere on the page. While all of these results may not be about me, the majority will be and can easily be digested. Adding the occupation, residence city, general interest, or college of the target may help eliminate unrelated results.

Search Operators

Most search engines allow the use of commands within the search field. These commands are not actually part of the search terms and are referred to as operators. There are two parts to most operator searches, and each are separated by a colon (:). To the left of the colon is the type of operator, such as "site" (website) or "ext" (file extension). To the right is the rule for the operator, such as the target domain or file type. The following will explain each operator and the most appropriate uses.

Site Operator

Google, and other search engines, allow the use of operators within the search string. An operator is text that is added to the search, which performs a function. My favorite operator is the "site:" function. This operator provides two benefits to the search results. First, it will only provide results of pages located on a specific domain. Second, it will provide all of the results containing the search terms on that domain. I will use my name again for a demonstration. I conducted a search of "Michael Bazzell" on Google. One of the results is a link to the website forbes.com. This search result is one of multiple pages on that domain that includes a reference to me. However, this search only displayed one of the many pages on that domain that possessed my name within them. If you want to view every page on a specific domain that includes your target of interest, the site operator is required. Next, I conducted the following exact search.

site:forbes.com "Michael Bazzell"

The result was all eight pages on forbes.com that include my name within the content. This technique can be applied to any domain. This includes social networks, blogs, and any other website that is indexed by search engines.

Another simple way to use this technique is to locate every page that is part of a specific domain. A search query of site:inteltechniques.com displays all 677 pages that are publicly available on my personal website. This can be a great way to review all of the content of a target's personal website without attempting to navigate the actual site. It is very easy to miss content by clicking around within a website. With this technique, you should see all of the pages in a format that is easy to digest. Also, some of the pages on a website that the author may consider "private" may actually

be public if he or she ever linked to them from a public page. Once Google has indexed the page, we can view the content using the "site" operator.

Real World Application: While conducting private background checks, I consistently use the site operator. A search such as "site:amazon.com" and the target name can reveal interesting information. A recent background check of an applicant that signed an affidavit declaring no previous drug or alcohol dependencies produced some damaging results. The search provided user submitted reviews that he had left on Amazon in reference to books that he had purchased that assisted him with his continual addiction to controlled substances. Again, this result may have appeared somewhere in the numerous general search results of the target; however, the site operator directed me exactly where I needed to look.

File Type Operator

Another operator that works with both Google and Bing is the file type filter. It allows you to filter any search results by a single file type extension. While Google allows this operator to be shortened to "ext", Bing does not. Therefore, I will use the original "filetype" operator in my search examples. Consider the following search attempting to locate PowerPoint presentation files associated with the company Cisco.

"Cisco" "PowerPoint"

The result is 7,909,000 websites that include the words Cisco and PowerPoint in the content. However, these are not all actual PowerPoint documents. The following search refines our example for accuracy.

"Cisco" filetype:ppt

The result is 22,300 Microsoft PowerPoint presentations that contain Cisco within the content. This search only located the older PowerPoint format of PPT, but not newer files that may have the PPTX extension. Therefore, the following two searches would be more thorough.

"Cisco" filetype:ppt
"Cisco" filetype:pptx

The second search provided an additional 12,000 files. This brings our total to over 40,000 PowerPoint files, which is overwhelming. I will begin to further filter my results in order to focus on the most relevant content for my research. The following search will display only newer PowerPoint files that contain the exact phrase Cisco Confidential within the content of the slides.
"Cisco Confidential" filetype:pptx

The result is exactly 17,976 PowerPoint files of interest. There are many uses for this technique. A search of filetype:doc "resume" "target name" often provides resumes created by the target

which can include cellular telephone numbers, personal addresses, work history, education information, references, and other personal information that would never be intentionally posted to the internet. The "filetype" operator can identify any file by the file type within any website. This can be combined with the "site" operator to find all files of any type on a single domain. By conducting the following searches, I was able to find several documents stored on the website irongeek.com.

site:irongeek.com filetype:pdf
site:irongeek.com filetype:ppt
site:irongeek.com filetype:pptx

Previously, Google and Bing indexed media files by type, such as MP3, MP4, AVI, and others. Due to abuse of pirated content, this no longer works. I have found the following extensions to be indexed and provide valuable results.

7Z: Compressed File
BMP: Bitmap Image
DOC: Microsoft Word
DOCX: Microsoft Word
DWF: Autodesk
GIF: Animated Image
HTM: Web Page
HTML: Web Page
JPG: Image
JPEG: Image
KML: Google Earth
KMZ: Google Earth
ODP: OpenOffice Presentation
ODS: OpenOffice Spreadsheet
ODT: OpenOffice Text
PDF: Adobe Acrobat
PNG: Image
PPT: Microsoft PowerPoint
PPTX: Microsoft PowerPoint
RAR: Compressed File
RTF: Rich Text Format
TXT: Text File
XLS: Microsoft Excel
XLSX: Microsoft Excel
ZIP: Compressed File

Hyphen (-)

The search operators mentioned previously are filters to include specific data. Instead, you may want to exclude some content from appearing within results. The hyphen (-) tells most search engines and social networks to exclude the text immediately following from any results. It is important to never include a space between the hyphen and filtered text. The following searches were conducted on my own name with the inclusion of excluded text. Following each search is the number of results returned by Google.

"Michael Bazzell" 34,300
"Michael Bazzell" -police 27,670
"Michael Bazzell" -police -FBI 27,000
"Michael Bazzell" -police -FBI -osint 6,010
"Michael Bazzell" -police -FBI -osint -books 4,320
"Michael Bazzell" -police -FBI -osint -books -open -source 404
"Michael Bazzell" -police -FBI -osint -books -open -source -"mr. robot" 9

The final search eliminated any results that included any of the restricted words. The nine pages that were remaining referenced other people with my name. My goal in search filters is to dwindle the total results to a manageable amount. When you are overwhelmed with search results, slowly add exclusions to make an impact on the amount of data to analyze.

InURL Operator

We can also specify operators that will focus only on the data within the URL or address of the website. Previously, the operators discussed applied to the content within the web page. My favorite search using this technique is to find File Transfer Protocol (FTP) servers that allow anonymous connections. The following search would identify any FTP servers that possess PDF files that contain the term OSINT within the file.

inurl:ftp -inurl:(http|https) filetype:pdf "osint"

The following will dissect how and why this search worked.

inurl:ftp – Instructs Google to only display addresses that contain "ftp" in the URL.

-inurl:(http|https) – Instructs Google to ignore any addresses that contain either http or https in the URL. The separator is the pipe symbol (|) located above the backslash key. It tells Google "OR". This would make sure that we excluded any standard web pages.

filetype:pdf – Instructs Google to only display PDF documents.

"osint" – Instructs Google to mandate that the exact term osint is within the content of results.

Obviously, this operator could also be used to locate standard web pages, documents, and files. The following search displays only blog posts from inteltechniques.com that exist within a folder titled "wp" (WordPress).

inurl:/wp/ site:inteltechniques.com

InTitle Operator

Similar to InURL, the "InTitle" operator will filter web pages by details other than the actual content of the page. This filter will only present web pages that have specific content within the title of the page. Practically every web page on the internet has an official title for the page. This is often included within the source code of the page and may not appear anywhere within the content. Most webmasters carefully create a title that will be best indexed by search engines. If you conduct a search for "osint video training" on Google, you will receive 115,000 results. However, the following search will filter those to 863. These only include web pages that had the search terms within the limited space of a page title.

intitle:"osint video training"

Note that the use of quotation marks prevents the query from searching "video training" within websites titled "osint". The quotes force the search of pages specifically titled "osint video training". You can add "all" to this search to force all listed words to appear in any order. The following would find any sites that have the words osint, video, and training within the title, regardless of the order.

allintitle:training osint video

An interesting way to use this search technique is while searching for online folders. We often focus on finding websites or files of interest, but we tend to ignore the presence of online folders full of content related to our search. As an example, I conducted the following search on Google.

intitle:index.of OSINT

The results contain online folders that usually do not have typical website files within the folders. The first three results of this search identified the following publicly available online data folders. Each possess dozens of documents and other files related to our search term of OSINT. One provides a folder structure that allows access to an entire web server of content. Notice that none of these results points to a specific page, but all open a folder view of the data present.

http://cyberwar.nl/d/
http://bitsavers.trailing-edge.com/pdf/
http://conference.hitb.org/hitbseccconf2013kul/materials/

OR Operator

You may have search terms that are not definitive. You may have a target that has a unique last name that is often misspelled. The "OR" (uppercase) operator returns pages that have just A, just B, or both A and B. Consider the following examples which include the number of results each.

"Michael Bazzell" OSINT 6,390
"Mike Bazzell" OSINT 413
"Michael Bazzell" OR "Mike Bazzell" OSINT 19,800
"Michael Bazell" OR "Mike Bazell" OSINT 107
"Michael Bazzel" OR "Mike Bazzel" OSINT 575

Asterisk Operator (*)

The asterisk (*) represents one or more words to Google and is considered a wild card. Google treats the * as a placeholder for a word or words within a search string. For example, "osint * training" tells Google to find pages containing a phrase that starts with "osint" followed by one

or more words, followed by "training". Phrases that fit this search include: "osint video training" and "osint live classroom training".

Range Operator (..)

The "Range Operator" tells Google to search between two identifiers. These could be sequential numbers or years. As an example, OSINT Training 2015..2018 would result in pages that include the terms OSINT and training, and also include any number between 2015 and 2018. I have used this to filter results for online news articles that include a commenting system where readers can express their views. The following search identifies websites that contain information about Bonnie Woodward, a missing person, and between 1 and 999 comments within the page.

"bonnie woodward" "1..999 comments"

Related Operator

This option has been proven very useful over the past year. It collects a domain, and attempts to provide online content related to that address. As an example, I conducted a search on Google with the following syntax.

related:inteltechniques.com

The results included no references to that domain, but did associate it with my other websites, my Twitter page, my Blackhat courses, and my book on Amazon. In my investigations, this has translated a person's personal website into several social networks and friends' websites.

Google Search Tools

There is a text bar at the top of every Google search result page. This allows for searching the current search terms within other Google services such as Images, Maps, Shopping, Videos, and others. The last option on this bar is the "Tools" link. Clicking this link will present a new row of options directly below. This provides new filters to help you focus only on the desired results. The filters will vary for each type of Google search. Figure 8.01 displays the standard search tools with the time menu expanded.

The "Any time" drop-down menu will allow you to choose the time range of visible search results. The default is set to "Any time" which will not filter any results. Selecting "Past hour" will only display results that have been indexed within the hour. The other options for day, week, month, and year work the same way. The last option is "Custom range". This will present a pop-up window that will allow you to specify the exact range of dates that you want searched. This can be helpful when you want to analyze online content posted within a known time.

Real World Application: Whenever I was assigned a missing person case, I immediately searched the internet. By the time that the case is assigned, many media websites had reported on the incident and social networks were full of sympathetic comments toward the family. In order to avoid this traffic, I set the search tools to only show results up to the date of disappearance. I could then focus on the online content posted about the victim before the disappearance was public. This often led to more relevant suspect leads.

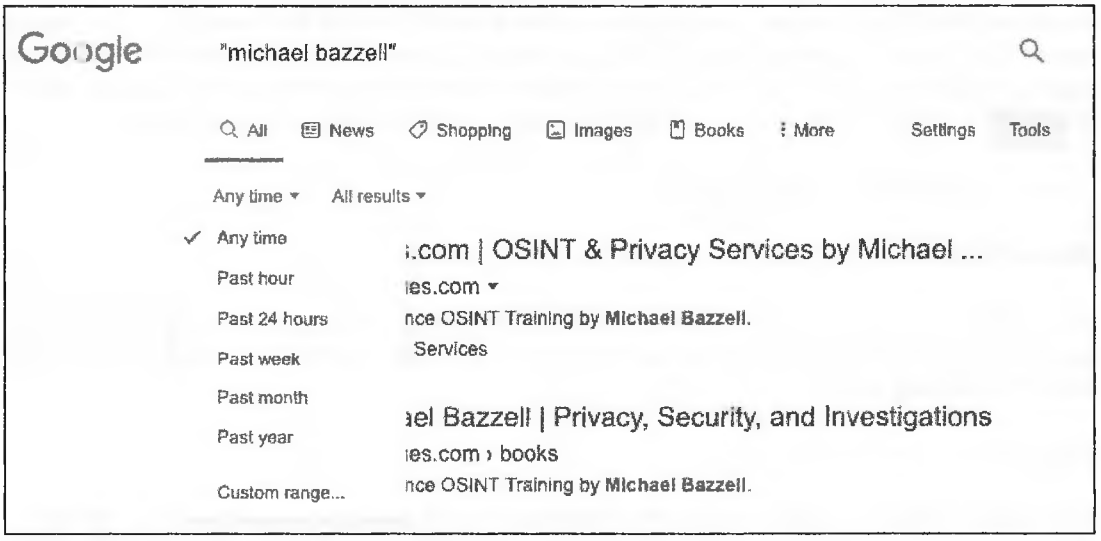


Figure 8.01: A Google Search Tools menu.

Dated Results

Google can be very sporadic when it comes to supplying date information within search results. Sometimes you will see the date that a search result was added to the Google index and sometimes you will not. This can be frustrating when you desire this information in order to identify relevant results. There is a fairly unknown technique that will force Google to always show you the date of each search result.

When you modify the "Any Time" option under the Search Tools menu, you will always see a date next to each result. If you are only searching for recent information, this solves the issue. However, if you are conducting a standard search without a specific date reference, the dates next to each result are missing. To remedy this, you can conduct a specific search that includes any results indexed between January 1, 1 BC and "today". The appropriate way to do this is to add "&tbs=cdr:1,cd_min:1/1/0" at the end of any standard Google search. Figure 8.02 (top) displays the results of a standard search for the terms OSINT Tools. The exact URL of the search was "google.com/?#q=osint+tools". Notice that the result does not include a date next to the item.

Figure 8.02 (bottom) displays the results of this same search with the specific data added at the end. The exact URL of this search was the following address.

google.com/?#q=osint+tools&tbs=cdr:1,cd_min:1/1/0

Notice that the result now has the date when the content was first indexed by Google. You can also now sort these results by date in order to locate the most recent information. The search tools menu also offers an "All results" menu that will allow you to choose to see "all results" or "Verbatim". The All Results will conduct a standard Google search. The Verbatim option searches exactly what you typed. One benefit of the Verbatim option is that Google will often present more results than the standard search. It digs a little deeper and gives additional results based on the exact terms you provided.

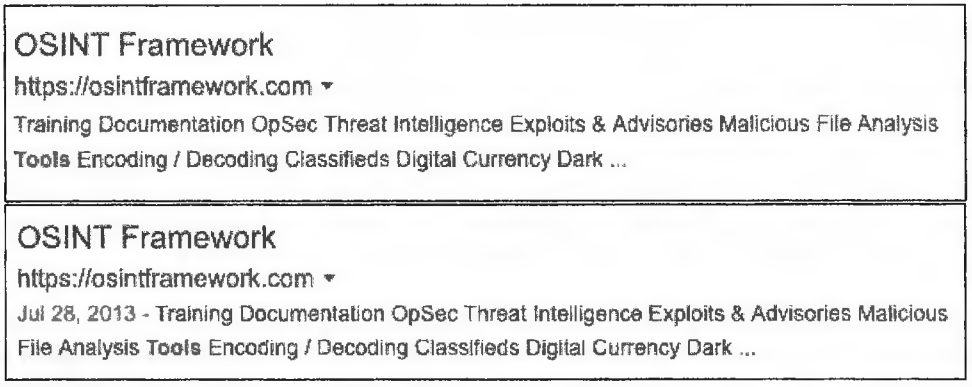


Figure 8.02: Results without (top) and with (bottom) date injection.

Google Custom Search Engines (google.com/cse)

Now that you are ready to unleash the power of Google, you may want to consider creating your own custom search engines. Google allows you to specify the exact type of searches that you want to conduct, and then create an individual search engine just for your needs. Many specialty websites that claim to search only social network content are simply using a custom engine from Google. For our first example, we will create a basic custom search engine that only searches two specific websites.

After you log in to a Google account, navigate to the website listed above. If you have never created an engine, you will be prompted to create your first. Enter the first website that you want to search. In my example, I will search inteltechniques.com. As you enter any website to search, Google will automatically create another field to enter an additional website. The second website that I will search is computercrimeinfo.com. Provide a name for your custom engine and select "Create". You now have a custom search engine. You can either embed this search engine into a website or view the public URL to access it from within any web browser.

This basic functionality can be quite powerful. It is the method behind my custom Pastebin search engine discussed in a later chapter. In that example, I created a custom search engine that scoured dozens of specific websites in order to retrieve complete information about specific topics. This is only the first layer of a Google custom search engine. Google offers an additional element to its custom engines. This new layer, labeled Refinements, allows you to specify multiple actions within one custom search engine. The best way to explain this is to offer two unique examples.

For the first example, I want to create a custom search engine that will allow us to search several social networks. Additionally, we will isolate the results from each network across several tabs at the top of our search results. The first step will be to create a new custom search engine in the same way that we did previously. Instead of specifying the two websites mentioned earlier, we will identify the websites to be searched as the following.

Facebook.com
Twitter.com
Instagram.com
LinkedIn.com
YouTube.com
Tumblr.com

While this is not a complete list of active social networks, it represents the most popular social networks at the time of this writing. At this point, our custom search engine would search only these websites and provide all results integrated into one search result page. We now want to add refinements that will allow us to isolate the results from each social network.

After you have added these websites, navigate to the control panel option in order to view the configuration of this custom search engine. On the left menu, you should see an option called "Search Features". This will present a new option at the top of the page labeled "Refinements". Click the "add" button to add a new refinement for each of the websites in this example. You should create these in the same order that you want them to appear within the search results. For this demonstration, I created the following refinements in order.

Facebook
Twitter
Instagram
LinkedIn
YouTube
Tumblr

When each refinement is created, you will have two options of how the search will be refined. The option of "Give priority to the sites with this label" will place emphasis on matching rules, but will also reach outside of the rule if minimal results are present. The second option of "Search only the sites with this label" will force Google to remain within the search request and not disclose other sites. I recommend using the second option for each refinement.

Now that you have the refinements made, you must assign them each to a website. Back on the "Setup" menu option, select each social network website to open the configuration menu. Select the dropdown menu titled "Label" and select the appropriate refinement. Repeat this process for each website and save your progress. You should now have a custom search engine that will not

only search several specific social network websites, but it should also allow you to isolate the results for each network. Navigate back to the control panel view and select the Public URL button to see the exact address of your new engine. Go to that address and you should see a very plain search engine. You can now search any term or terms that you want and receive results for only the social networks that you specified. Additionally, you can choose to view all of the results combined or only the results of a specific network. Figure 8.03 displays the results when I searched the term osint. In this example, I have selected the Twitter refinement in order to only display results from twitter.com.

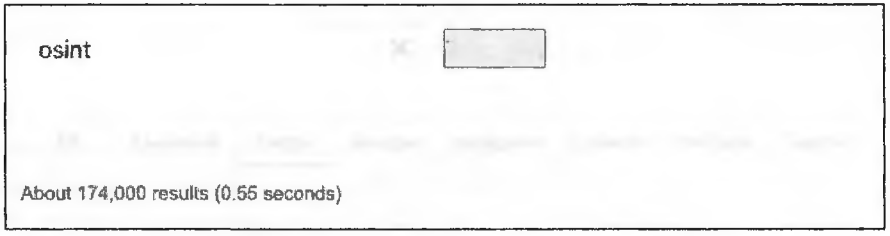


Figure 8.03: A Twitter refinement in a Google Custom Search.

You can now bookmark this new search engine that you created and visit it whenever you have a target to search. You can take your custom search engines to another level by adding refinements that are not website specific. In the next example, we will make a search engine that will search the entire internet and allow us to filter by file type.

Create a new custom search engine and title it "Documents". Add only "google.com" as the website to be searched. We do not actually want to search google.com, but a website is required to get to the control panel. Save your engine and open the control panel to configure the options. In the "Sites to search" portion, choose the "Search only included sites" option and change it to "Search the entire web but emphasize included sites" option. Delete google.com from the sites to be searched. You now basically have a custom search engine that will search everything. It will essentially do the same thing as Google's home page. You can now add refinements to filter your search results. Navigate to the search features menu and add a new refinement. Title the new refinement "PDF"; leave the default setting of "Give priority to the sites with this label"; and enter the following in the "Optional word(s)" field.

ext:pdf

This will create a refinement that will allow you to isolate only PDF documents within any search that you conduct. Save this setting and create a new refinement. Title it DOC; leave the default search setting; and place the following in the "Optional word(s)" field.

ext:doc OR ext:docx

This will create a new tab during your search results that will allow you to isolate Microsoft Word documents. By entering both the doc and docx formats, you will be sure to get older and newer documents. The word "OR" tells Google to search either format. Repeat this process for each of the following document types with the following language for each type.

XLS (Excel Spreadsheets) – ext:xls OR ext:xlsx OR ext:csv
PPT (PowerPoint Files) – ext:ppt OR ext:pptx
TXT (Text Docs) – ext:txt OR ext:rtf
WPD (Word Perfect Docs) – ext:wpd
ODT (Open Office Docs) – ext:odt OR ext:ods OR ext:odp
ZIP (Compressed Files) – ext:zip OR ext:rar OR ext:7z

Figure 8.04 displays the results of a search for the term osint within this new engine. The All tab is selected which reveals 717,000 results. Clicking the PowerPoint presentations option (PPT) reveals 45 files which contain the term. There are endless possibilities with this technique. You could make an engine that searched for audio and video files with extensions such as mp3, mp4, mpeg, avi, mkv, etc. You could make an engine that isolated images with extensions such as jpg, jpeg, png, bmp, gif, etc. You could also replicate all of this into a custom engine that only searched a specific website. If you were monitoring threats against your company, you could isolate only these files that appear on one or more of your company's domains.

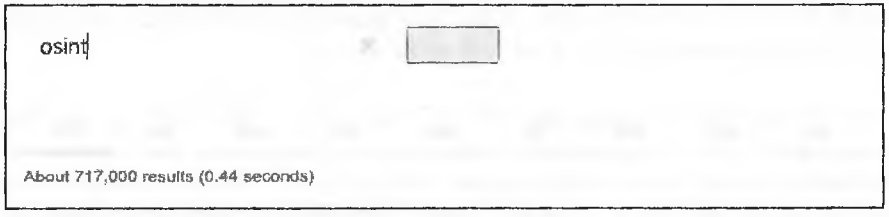


Figure 8.04: A PowerPoint file refinement within a Google Custom Search.

As a service to readers, I have publicly shared these two custom search engines at the following URL-shortened addresses. I present a third option in the Documents chapter.

Social Networks: https://bit.ly/2ma3ANe
Document Formats: https://bit.ly/2nKAqEB

One negative aspect to custom Google search engines is that they only display the most relevant 100 results. These are presented in ten pages of ten results per page. If you are searching very specific terms, this may not be an issue. However, standard searches can be limiting. Overall, I do not rely heavily on custom search engines for large websites. They simply cannot replicate a properly structured manual search. However, they can be beneficial for smaller sites which you often query.

Google Alerts (google.com/alerts)

When you have exhausted the search options on search engines looking for a target, you will want to know if new content is posted. Checking Google results every week on the same target to see if anything new is out there will get mundane. Utilizing Google Alerts will put Google to work on locating new information. While logged in to any Google service, such as Gmail, create a new Google Alert and specify the search term, delivery options, and email address to which to send the alert. In one of my alerts, Google will send an email daily as it finds new websites that mention "Open Source Intelligence Techniques" anywhere in the site. Another one of my alerts is for my personal website. I now receive an email when another site mentions or links to my website. Parents can use this to be notified if their child is mentioned in a website or blog. Investigators that are continuously seeking information about a target will find this beneficial.

Real World Application: A police detective was assigned a runaway case where a 15-year-old had decided to leave home and stay with friends at an unknown location. After several extensive internet searches, a Google Alert was set up using the runaway's name and city of residence. Within three days, one of the alerts was for a blog identifying the runaway and where she was currently staying. Within 30 minutes, the unhappy child was back home.

Bing (bing.com)

Google is not the only great search engine. While Google is the overwhelming choice of search engines used today, other sites should not be ignored, especially when having trouble locating any information on a subject. Bing is Microsoft's competition to Google and provides a great search experience. In 2009, Yahoo search (yahoo.com) began using the Bing search engine to produce search results. This makes a Yahoo search redundant if a Bing search has already been conducted. The same tactics described previously, and in the numerous Google books, can be applied to any search engine. The site operator and the use of quotes both work with Bing exactly as they do with Google. Bing also introduced time filtered searching that will allow you to only show results from the last 24 hours, week, or month. There are a couple of additional operators that are important that only apply to Bing. Bing offers an option that will list every website to which a target website links, and is the only search engine that offers this service.

Bing LinkFromDomain

I conducted a search on Bing of "LinkFromDomain:inteltechniques.com". Note that there are no spaces in the entire search string and omit the quotation marks. This operator creates a result that includes every website to which I have a link, located on any of the pages within my website. This can be useful to an investigator. When a target's website is discovered, this site can be large and contain hundreds of pages, blog entries, etc. While clicking through all of these is possible, sometimes links are hidden and cannot be seen by visually looking at the pages. This operator allows Bing to quickly pull links out of the actual code of the website.

Bing Contains

Earlier, I discussed searching for files with specific file extensions on Google. The "filetype" and "ext" operators that were explained both work on Bing the same way. However, Bing offers one more option to the mix. The "contains" operator allows you to expand the parameters of the file type search. As an example, a Bing search of "filetype:ppt site:cisco.com" returns 6,200 results. These include PowerPoint files stored on the domain of cisco.com. However, these results do not necessarily include links on the cisco.com website to PowerPoint files stored on other websites. A search on Bing for "contains:ppt site:cisco.com" returns 131,200 results. These include PowerPoint files that are linked from pages on the domain of cisco.com, even if they are stored on other domains. This could include a page on cisco.com that links to a PowerPoint file on hp.com. In most cases, this search eliminates the need to conduct a filetype search, but both should be attempted.

Google Images (images.google.com)

Google Images scours the web for graphical images based on a search term. Google obtains these images based on keywords for the image. These keywords are taken from the filename of the image, the link text pointing to the image, and text adjacent to the image. This is never a complete listing of all images associated with a subject, and will almost always find images completely unrelated to a target. In the case of common names, one should enclose the name in quotes and follow it with the city where the subject resides, place of employment, home town, or personal interests. This will help filter the results to those more likely to be related to the subject. When results are displayed, clicking the "Tools" button will present five new filter menus. This menu will allow you to filter results to only include images of a specific size, color, time range, image type, or license type. The most beneficial feature of Google Images is the reverse image search option. This will be explained in great detail later in the book.

Bing Images (bing.com/images)

Similar to Google, Bing offers an excellent image search. Both sites autoload more images as you get toward the end of the current results. This eliminates the need to continue to load an additional page, and leads to faster browsing. Bing also offers the advanced options available on Google, and adds the ability to filter only files with a specified layout such as square or wide. Bing provides a "filter" option in the far right of results that provides extended functionality. The People tab offers restriction for images of "Just faces" and "Head & shoulders". It also provides suggested filters with every image search. Clicking image search links may provide additional photographs of the specific target based on the listed criteria. This intelligence can lead to additional searches of previously unknown affiliations.

Web Archives

Occasionally, you will try to access a site and the information you are looking for is no longer there. Maybe something was removed, amended, or maybe the whole page was permanently removed. Web archives, or "caches" can remedy this. I believe that these historical copies of websites are one of the most vital resources when conducting any type of online research. This section will explain the current options in order from most effective to least.

Google Cache (google.com)

When conducting a Google search, notice the result address directly below the link to the website. You will see a green down arrow that will present a menu when clicked. This menu will include a link titled "Cached". Clicking it will load a version of the page of interest from a previous date. Figure 3.05 (first image) displays a search for phonelosers.org which returns a result that includes a cached version of the page. This version was taken four days prior to the current date, and displays information different from the current version. The second option visible within this menu, titled "Similar", identifies web pages that contain content similar to the listed result.

If you have a specific page within a website that you want to view as a cached version, type the exact website into Google to link to the cached page. For example, if I wanted to see a previous view of the podcast for The Phone Show, an audio archive about telephone pranks, I would conduct a Google search for the site "www.phonelosers.org/snowplowshow". This will return the main landing page as well as sub-pages that will each have a cached view. If any of these pages were to go offline completely, Google would hold the last obtained version for viewing. I could have also typed the following directly into any Google search page to be navigated directly to the cached page.

cache:www.phonelosers.org/snowplowshow

Bing Cache (bing.com)

Similar to Google, Bing offers a cached view of many websites. Searching for a domain name, such as phonelosers.org, will present many results. The first result should link to the actual website. Directly next to the website name is a small green down arrow. Clicking it will present the option of "Cached page". Clicking this link will display a previous version of the target website as collected by Bing. Figure 8.05 (second image) displays their menu option.

Yandex Cache (yandex.com)

The Russian search engine Yandex will be explained in great detail later, but it is important to note now that it also possesses a cache option. Very similar to Google and Bing, Yandex presents a green drop-down menu directly under the title of the search result. Figure 3.05 (third image) displays their cache menu option. Selecting the Cached page option opens a new tab displaying

the most recent Yandex archive of the page. The top banner displays the date and time of capture, the original website address, and a search option to highlight selected keywords within the result. The biggest strength of the Yandex cache is the lack of updates. While this may sound counterintuitive, an older cache can be very helpful in an investigation. Assume that the Phone Losers website was your target. At the time of this demonstration, September 7, 2019, the Google, Bing, and Yandex caches of this page were dated as follows.

Google: September 6, 2019
Bing: September 7, 2019
Yandex: September 1, 2019

Google and Bing tend to have very recent results which often appear identical to the live view. However, the Yandex option from a week prior is more likely to contain modified content. You can often locate a cached version of a page that is older than the Yandex version on Baidu.

Baidu Cache (baidu.com)

This Chinese search engine is the least productive as far as cached copies of websites are concerned, but it should not be ignored. It will be explained further during a later discussion about international engines. The results of a search on Baidu are mostly in Chinese, but can still be valuable to those that cannot read the text. At the bottom of each search result is a green link to the website that hosts the content of the result. While this also includes a drop-down menu, the cache option is not there. Instead, look for a word in Chinese directly to the right of this link. In Figure 3.05 (fourth image) it is displayed as 百度快照. Clicking this link will open a new tab with the cache result, which Baidu refers to as a snapshot. In my experience, the presence of this linked option does not always mean that a cached version exists.

The Wayback Machine (archive.org/web/web.php)

The Wayback Machine will provide a much more extensive list of options for viewing a website historically. Searching for phonelosers.org displayed a total of 1,280 captures of the site dating from 12/21/1997 through 6/10/2019 (Figure 8.06). Clicking the links presents quite a display of how the site has changed. Graphics are archived as well, proving that we should always think twice about which photos we post to the internet. Each view of the archived page will allow the user to click through the links as if it were a live page on the original web server. Clicking through the timeline at the top of each page will load the viewed page as it appeared on the date selected.

Wayback Search

Until 2016, you could not search keywords across Wayback Machine data. You had to know the exact URL of a target website, or at least the domain name. Today, we can search any terms desired and connect directly to archived data. At the time of this writing, a search bar was present at the top of every Wayback Machine page. If that should change, you can also conduct a search

via a direct URL. The following address searched "Michael Bazzell" throughout the entire archive of information.

https://web.archive.org/web/*/Michael Bazzell

The results identify over twenty websites that include these terms. Within those sites are dozens of archived copies of each. This data represents decades of content at your fingertips. Much of it is offline and unavailable on the current public internet. Many domains have completely shut down. Furthermore, websites that I own appear within the results, even though I have specifically blocked archiving them through a configuration file on my server. You would not find these by searching the domains directly through the Wayback Machine. This is a reminder that we should check all available resources before completing our investigations.

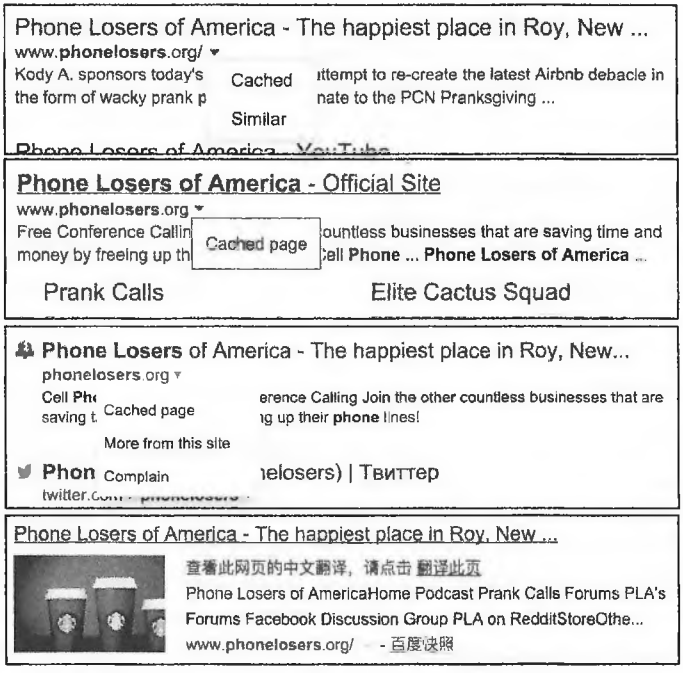


Figure 8.05: Cache menu options on Google, Bing, Yandex, and Baidu.

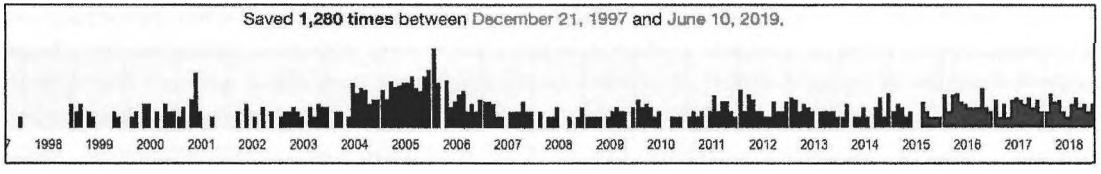


Figure 8.06: Wayback Machine results for an archived website.

Archive.Today (archive.today)

There is another option for historic archives of web pages. However, it is not as powerful as the Wayback Machine and my successes have been minimal. Archive.Today will allow you to search a domain and display any captured archives of the home page. They also offer a wildcard service that will search an entire domain for any previously captured pages. If I wanted to search inteltechniques.com for any archived pages, I would enter the following two queries into the search field.

https://inteltechniques.com/*
*.inteltechniques.com

The first would look for any archived pages within the chosen domain. The second would look for any sub-domains such as mail.inteltechniques.com or ftp.inteltechniques.com. I have rarely received results from the sub-domain search through this service. The strength of Archive.Today is their stance on ignoring the noarchive rule of the robots.txt file. This file will be fully explained in a later chapter. In brief summary, it is a set of rules included on a web server that instructs search engines on allowable ways to index the website. The noarchive function tells a search engine to never archive anything on the site. As an example, my website inteltechniques.com has this enabled. Therefore, Google, Bing, Yandex, and Baidu do not have a cache of any pages on that domain. However, Archive.Today does, and they have multiple versions from the past three years. Consequently, this service should always be checked when dealing with a tech savvy target that blocks traditional engines from caching a page.

Searching All Resources

Occasionally, there are websites that surface claiming to be able to extract and rebuild entire websites from online caches. In my experience, none of these have ever provided a complete historical view versus a manual approach. Engines such as Bing and Yandex generate a unique code when a cache is displayed. This action prevents most automated search tools from collecting archived information. I do not believe any option, other than navigating to each resource, will present you with the content that you need. I bookmark each of these services in an individual folder titled Archives and open each tab when I have a domain as a target. I have also created an online tool that will collect your target domain and forward you to the appropriate archive page. This will be explained later when discussing domain searches.

Finally, it is important to acknowledge that these resources can be beneficial when everything on a website appears to be present and unaltered. While caches work well on websites that have been removed and are completely empty, they also can tell a different story about websites that appear normal. Any time that I find a website, profile, or blog of interest, I immediately look at caches hoping to identify changes in content. These minor alterations can be very important. They highlight information that was meant to be deleted forever. These details can be the vital piece of your investigation puzzle. Most people have no idea that this technique exists.

Non-English Google Results

Not every piece of information that will be useful to you will be obtained by standard searches within English websites. Your target may either be from another country or have associates and affiliations in another country. While Google and Bing try to pick up on this, the technology is not perfect. Google has a search site and algorithm that change by location. For example, google.fr presents the French search page for Google. While this may produce the same overall results, they are usually in a different order than on google.com. Google no longer maintains a page with links to each international version of its search, but I have a preferred method.

2Lingual (2lingual.com)

This page will allow you to conduct one search across two country sites on Google. The Google search will display a plain search box and choices of two countries. The results will display in single columns next to each other. Additionally, the foreign results will be automatically translated to English. This feature can be disabled, if desired. The first few sponsored results (ads) will be similar, but the official results following should differ. This site can also be helpful when demonstrating to someone the importance of searching targets through multiple countries.

Google Translator (translate.google.com)

Many websites exist in non-English languages. As internet enthusiasts, we tend to focus on sites within our home area. There is a wealth of information out there on sites hosted in other countries which are presented in other languages. Google Translator will take text from any site or document and translate the text to a variety of languages. Usually, the service will automatically identify the language of the copied and pasted text. Selecting the desired output will provide the translation. Alternatively, you can translate an entire website in one click which will give a native view of the layout of the site. Instead of copying individual text to the search box, type or paste in the exact URL (address) of the website you want translated. Clicking the "Translate" button will load a new page of the site, which will be translated to English. This translation is rarely, if ever, perfect. However, it should give you an idea of the content presented on the page. This will also work on social network sites such as Twitter and Instagram.

Bing Translator (bing.com/translator)

A few years after Google introduced free translation services, Bing created their own product. At first glance, it looks like a replica of Google's offering. However, Bing's translations are usually slightly different than Google's results. Similar to Google, you can also type or paste an entire foreign website to conduct a translation of everything on the target page.

DeepL (deepl.com/translator)

While smaller than Google or Bing, this may be the most accurate translator service I have found.

PROMT Online Translator (online-translator.com)

There are dozens of additional online translation tools available. Almost all of them allow translation of a small amount of text at a time. Some use either the Google or Bing translation service. One last online translation tool worth mentioning is PROMT Online Translator. It is unique from the dozens of other options in that it allows translation of entire websites similar to Google and Bing. This service provides an independent translation and can be considered a third source.

I am often asked during training which of the three services I use during investigations. My answer is all four. This is important for two reasons. The obvious benefit is that you will receive four unique translations that will be very similar. The minor variations may be important, especially when translating Tweets and other shortened messages that may not be grammatically correct in any language. The second reason is to show due diligence during my investigation. I always want to go above and beyond what is required. Translating a foreign web page through four different services emphasizes my desire to conduct an unbiased investigation.

Google Input Tools (google.com/inputtools/try)

There is one last feature regarding foreign language searching that I have found useful. Google's Input Tools allow you to type in any language you choose. Upon navigating to the above website, choose the language of your target search. In Figure 8.07, I have chosen Arabic as the language and typed "Online Investigation" on a standard English keyboard. The result is how that text might appear in traditional Arabic letters. I have had the most success with this technique on Twitter. When supplying any search term on Twitter, the results are filtered by the presence of the keywords entered and only in the language provided. Searching "Online Investigation" on Twitter only provides results that have that exact spelling in English characters. However, searching the Arabic output provides Tweets that include the Arabic spelling of the selected words. This technique is extremely important when you have located a username in a foreign language. As with all computer-generated translation services, the results are never absolutely accurate. I expect this technology to continue to improve.



Figure 8.07: A Google Input Tools translation from English to Arabic.

Google Groups (groups.google.com)

Google Groups provides access to both Usenet groups and Non-Usenet Google groups. Usenet groups are similar to mailing lists. The Usenet archive is complete and dates back to 1981. Since

many people posted to these groups using their real name or email address, identifying their opinions on controversial topics is effortless. Additionally, searching a real name will often provide previous email addresses that may not be known to the searcher. This provides new intelligence for future searches. While none of this is usually damaging to the submitter, it helps provide an overall view of the target of interest. Many of the newer groups used are created through Google and conform to practically any interest imaginable. Most users continue to use a real name, screen name, email address, or a combination of all three. Searching these posts is similar to any Google search.

Yahoo Groups (groups.yahoo.com)

While Google Groups will search many of the non-Google online groups, it will not pick up all of the Yahoo Groups. The content of most Yahoo Groups is public and will allow viewing without membership. Searching by real name or screen name will often produce results. Any time I have a target with a Yahoo email address, I search the username before @yahoo.com through Yahoo Groups. Many people who possess a Yahoo email address have had it for many years. Yahoo Groups were very popular in the early days of the internet. In my experience, you are more likely to find content from Yahoo users in Yahoo Groups than Gmail users in Google Groups.

I have had multiple successes with searches in Google and Yahoo Groups. Most have been associated with pedophiles or background checks. In pedophile cases, I have identified new evidence based on historic conversations in various Yahoo Groups. This never identified new victims or generated new cases, but the details strengthened the current charges by showing a pattern of inappropriate interest in children. With background checks, this content has been extremely valuable. While applicants can easily clean their blogs and social profiles, they cannot easily purge their history within these groups. Many people simply have forgotten about the content, which was often posted a decade earlier. In one example, an applicant had signed an affidavit claiming to have never abused any type of narcotics. However, a post to a newsgroup indexed by Google revealed that he had battled cocaine addiction for several years.

Google News Archive (news.google.com)

This can be an amazing resource of information about a target. In the past, if someone relocated to a new geographical area, he or she could leave the past behind and start over. Today, that is difficult. Google's News Archive is continually adding content from both online archives and digitized content from their News Archive Partner Program. Sources include newspapers from large cities, small towns, and anything in between. The link referenced above will allow for a detailed search of a target's name with filters including dates, language, and specific publication. In order to display this menu, click on the down arrow to the right of the search box. This can quickly identify some history of a target such as previous living locations, family members through obituaries, and associates through events, awards, or organizations.

Google Newspaper Archive (news.google.com/newspapers)

The previous option focused solely on digital content, such as your local newspaper website. Google's Newspaper archive possesses content from printed newspapers. All results on this site consist of high-resolution scanned newspaper pages. In my experience, this collection is not as extensive as the next option discussed. However, it is definitely worth a look, and will likely continue to grow.

Newspaper Archive (newspaperarchive.com)

This paid service provides the world's largest collection of newspaper archives. The high resolution PDF scans of entire daily newspapers range in date from the 1800's until present. The first four editions of this book explained a method of using the Google Site operator and cached results to obtain practically any page of this newspaper collection without paying or subscribing. These vulnerabilities have all been patched and none of those techniques work today. Fortunately, Newspaper Archive still offers a 14-day free trial with unlimited access to every archive. While multiple trials can be obtained, each require a unique credit card number and email address. Many libraries have asked this service to scan their entire microfilm archives and make them freely available online. You will not find any mention of this free alternative on their home page, but a bit of searching will guide you to the right place. The following search on Google identifies hundreds of public libraries that pay for your access to their archives.

site:newspaperarchive.com "This archive is hosted by" "create free account"

The first part of the search tells Google to only look at the website newspaperarchive.com. The second part mandates that the exact phrase "This archive is hosted by" appears in the result. The final piece isolates only the newspaper collections that are available for free, and without a credit card. This identifies the landing pages of the various libraries that have made their collections freely available. While you will still be required to register through the service, payment is not required for these collections. Consider the following usage that will likely present you with free views of Newspaper Archive whenever you need them.

On 12/13/2017, I navigated to newspaperarchive.com/advancedsearch/ and conducted an advanced search for anyone named Michael Williams from Cedar Rapids, Iowa. Newspaper Archive presented several results from the Cedar Rapids Gazette. Clicking on any of these results prompted me to create an account and forced me to enter a valid credit card number to proceed. I could not create an account from any of the pages without providing payment. Instead, I conducted the following Google search.

site:newspaperarchive.com "This archive is hosted by" "cedar rapids gazette"

The first result was a direct connection to crpubliclibrary.newspaperarchive.com. Clicking this link presented a page dedicated to searching over 40 newspapers within the Cedar Rapids and

Des Moines areas. In the upper right corner was a link titled "Create Free Account". I clicked this link and provided generic details and a throwaway email address. The membership choices now include a completely free option, which will only allow access to the Iowa newspapers. After creating my free account, I returned to crpubliclibrary.newspaperarchive.com and repeated the search of my target. Every link allowed me full unrestricted access to the high-resolution images.

While still logged in to this account, I navigated to delawarecolib.newspaperarchive.com, the direct page associated with the Delaware County Library (which I found through the original Google search in this section). I was not authorized to view this newspaper collection. However, after clicking "Create Free Account" on this page, I entered the same data as previously provided to the Iowa newspaper previously. After verifying my email address, I was allowed immediate access to this series of newspapers.

This technique will not obtain access to every collection on Newspaper Archive. However, it will provide a surprising amount of free access to huge collections internationally. During an hour of downtime, I created a free account on every library collection I could locate, using the same credentials on each. I can now log in to my single Newspaper Archive account and navigate the site from any page. When I reach a newspaper of interest after a search, I will be given full access if it is within a free collection. This is all thanks to the local libraries that have paid this site to give free access to the public. If the free trial of Newspaper Archive or the free library collections do not offer enough content, consider the following options.

Old Fulton (fultonhistory.com/Fulton.html): 34,000,000 scanned newspapers from the United States and Canada.

Library of Congress US News Directory (chroniclingamerica.loc.gov):
Scanned newspapers from the United States dated 1836-1922.

Library of Congress US News Directory (chroniclingamerica.loc.gov/search/titles):
Scanned newspapers from the United States dated 1690-Present.

Google Advanced Search (google.com/advanced_search)

If the search operators discussed in this chapter seem too technical, Google offers an advanced search page that simplifies the process. Navigating to the above website will present the same options in a web page that are possible by typing out the operators. This will help you get familiar with the options, but it will be beneficial to understand the operators for later use. The Advanced Search page will allow you to specify a specific phrase for which you are searching, just like the quotes in a search will allow. The site and filetype operators used earlier can be achieved by entering the desired filters on this page. It should be noted that the file type option on this page is limited to popular file types, where the filetype operator can handle many other file extensions.

Bing Advanced Search (search.yahoo.com/web/advanced)

Bing does not technically provide an advanced search page similar to Google's. However, since Yahoo uses Bing's search, you can use Yahoo's advanced search page as a replacement. This page will allow you to easily create a search that filters by individual terms, exact phrases, omitted terms, specific domains, file formats, and languages.

Additional Google Engines

Google isolates some search results into specialized smaller search engines. Each of these focuses on a unique type of internet search. The following engines will likely give you results that you will not find during a standard Google or Bing search. While some results from these unique searches will appear within standard Google results, the majority will be hidden from the main page.

Google Blogs (google.com)

Google removed its original blog search in 2014. It was quite helpful and focused mostly on personal websites, especially those with a blogging platform. Today, this is nothing more than a subsection of Google News. You can load the Blogs option under the "All News" menu within the Tools option on any Google News results page. Alternatively, you can navigate to the following address, replacing TEST with your search terms.

google.com/search?q=TEST&tbm=nws&tbs=nrt:b

The website above displays a standard Google search option, but the results appear much differently. A standard Google search of my name reveals my website, Twitter, and Amazon pages in the first results. The Google Blogs option reveals several personal and professional (media) blogs that mention my name. These results are likely buried within the standard Google search.

Google Patents (google.com/?tbm=pts)

Google probably has the best patent search option on the internet. It allows you to search the entire patent database within any field of a patent. This can be useful for searching names associated with patents or any details within the patent itself. If you need further help, Google offers an advanced patent search at google.com/advanced_patent_search.

Google Scholar (scholar.google.com)

Google Scholar is a freely accessible web search engine that indexes the full text of scholarly literature across an array of publishing formats. It includes most peer-reviewed online journals of Europe's and America's largest scholarly publishers, plus many books and other non-peer reviewed journals. My favorite feature of this utility is the case law and court records search. I

have located many court records through this free website that would have cost money to obtain from private services.

Advangle (advangle.com)

This is a simple and convenient builder of complex web search queries for both Google and Bing. The service allows you to quickly build a query with multiple parameters, such as the domain, language, or date published. Immediately you will see the result of this query in Google or Bing search engines. You can save your queries in an Advangle account if you want to restore a search to identify new results. Any condition in a query can be temporarily disabled without removing it completely. This allows you to quickly try several combinations of different conditions and choose the most suitable for your needs. Figure 8.08 displays the search page with filters for my exact name, on my website, within the past month, and only PDF files. The Google and Bing "Open" options will launch a new tab with the exact terms required for these options.

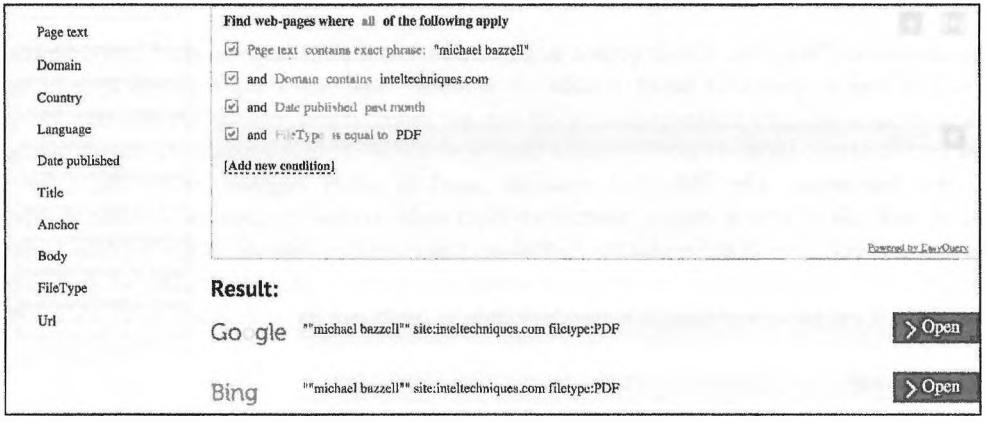


Figure 8.08: An Advangle search menu in use.

Keyword Tool (keywordtool.io)

Keyword Tool displays autocomplete data from Google, Bing, YouTube, and the App Store. You have likely noticed that Google quickly offers suggestions as you type in your search. This is called autocomplete. If I were to type "macb" into Google, it would prompt me to choose from the most popular searches when people typed those letters. This information may lead you to new terms to search in reference to your investigation. The advantage of Keyword Tool over Google is that Google only provides the five most popular entries. Keyword Tool provides the ten most popular entries. Additionally, you can choose different countries to isolate popular terms. You can also see results from similar searches that Google does not display.

Real World Application: I have successfully used this technique during the investigation of many businesses. I was once asked by a medium-sized business to investigate reports of a faulty

product that they had recently recalled. They wanted to see customer complaints. After searching the typical review websites, I conducted a search with Keyword Tool. I discovered that the 9th most popular search involving this specific product name included a term that was a misspelling of the product name. It was different enough in spelling that my searches were missing this content. Knowing this information, I was able to locate more relevant data for the client.

Other Alternatives

Google and Bing are great, but they do not do it all. There will always be a need for specialized search engines. These engines usually excel in one particular search method which justifies the lack of search power in other areas. The sites listed in this next section represent the extreme minority when it comes to search traffic. It is often sites like these that implement the technologies that we later take for granted in more popular engines.

Exalead (exalead.com)

Headquartered in Paris, this search engine has gained a lot of popularity in the United States. The main search engine provides many results on popular searches. I have found that individual targets without a strong internet presence do not get many, if any, results on this site. However, this site excels in two areas. It works well in finding documents that include the target mentioned within the document. The "filetype" operator used in other engines works the same here. Voxalead, an Exalead search engine, searches within audio and video files for specific words. This is thanks to speech to text technologies. Voxalead will search within all of the spoken audio of a file for references to the text searched. The results are presented in a timeline view. Currently, the majority of the results of this new product link to news media and public news video files.

Searx (searx.me)

This is considered a meta-crawler, as it presents results from both Google and Bing. It often gets dismissed as another comparison search site, but there are many other advantages to using this service. First, conducting a search will provide results from the two main search engines, but will remove duplicate entries. This alone is a quick way to conduct your due-diligence by checking Google and Bing. Next, the top row of options will allow you to repeat this redundancy-reducing option by checking results on Google's and Bing's Images, News, and Videos sections. Next to each result on any search page is a "cached" link. Instead of opening the Google or Bing cache, clicking this will open the cached page of the target website through the Wayback Machine. Finally, a "proxied" option next to each result will connect you to the target website through a proxy service provided by Searx. This is basically a layer of privacy preventing the website owner from collecting data about you, such as your IP address. Technically, Searx.me opened the target site, and their data would be tracked instead of yours. There are ways for adversaries to bypass this "anonymity", but it is decent protection for most sites.

The final benefit of this service over all others is the easy ability to export search results as a file. The "Links" section to the right of all search pages displays options to download a csv, json, or rss file of the results. The csv option is a simple spreadsheet that possesses all of the search results with descriptions and direct links. I find this helpful when I have many searches to conduct in a short amount of time, and I do not have the ability to analyze the results until later.

Million Short (millionshort.com)

This website offers a unique function that is not found on any other search engine. You can choose to remove results that link to the most popular one million websites. This will eliminate popular results and focus on lesser known websites. You can select to remove the top 100,000, 10,000, 1,000, or 100 results.

Tor Search Engines

Tor is free software for enabling anonymous communication. The name is an acronym derived from the original software project name The Onion Router. Tor directs internet traffic through a free, worldwide, volunteer network consisting of more than six thousand relays to conceal a user's location and usage from anyone conducting network surveillance or traffic analysis. Using Tor makes it more difficult for internet activity to be traced back to the user. This also applies to a website that is hosted on the Tor network. Usually, these sites include illegal drug shops, child pornography swaps, and weapon sales. Because these sites are not hosted on publicly viewable networks, they are hard to locate and connect. Two search engines and a proxy aid this process.

Ahmia (ahmia.fi)

This is a very powerful Tor search engine. While no engine can index and locate every Tor website, this is the most thorough option that I have seen. It should be the first engine used when searching Tor related sites. The links in the results will not load if searching through a standard browser and connection. Using the Tor Browser discussed previously is the ideal way to use this service.

Dark Search (darksearch.io)

This engine appeared in 2019 and appears quite promising. When I conducted a search of "OSINT" within Ahmia, I received 5 results. The same search on Dark Engine revealed 51 results. It appears to index Tor sites well, and I received many results when querying email addresses of targets. This has replaced Ahmia for most of my Tor-based investigations.

Onion Link (onion.link)

Similar to Ahmia, Onion Link attempts to identify websites within the Tor network. It uses a Google Custom Search Engine (CSE) and appends ".link" to each search result. This allows you to open these links through Onion Link's own Tor connection which appears to pull the data from their own cached sources. This makes the viewing of these pages much quicker with faster page loads, and eliminates the need to be on the Tor Browser. While relying on Google to index these pages is a bit amateur, the minimal results here are often different than other Tor engines.

Tor2Web (tor2web.org)

Whenever you see a URL like libertygb2nyeyay.onion, it is a Tor Onion website. As mentioned earlier, you cannot connect directly to these links without being connected to the Tor network. However, you can replace ".onion" within the address to ".onion.to" in order to view the content. In the above example, navigating to the website libertygb2nyeyay.onion.to will display the content using the Tor2Web proxies. This connects you with Tor2web, which then talks to the onion service via Tor, and relays the response back. This is helpful when locating Tor links on Twitter.

Tor Search Sites

I believe some of the strongest Tor search engines exist only on the Tor network. You cannot access them from a standard internet connection, and the Tor Browser is required for native use. My favorite is "Not Evil", which can be found at the following address if connected to Tor.

hss3uro2hsxfogfq.onion

Since Tor2Web allows us to use their proxy, we can connect to "Not Evil" by navigating directly to the following Tor2Web proxy address, without being on the Tor Browser.

hss3uro2hsxfogfq.onion.to

This presents the home page of the search site, and allows for a keyword search. However, searching through this portal while being connected through the Tor2Web proxy can present difficulties. Instead, consider conducting a search within a URL submission. In the following web address, I am connecting to Tor2Web's proxy of the search engine and requesting a search results page for the term OSINT.

hss3uro2hsxfogfq.onion.to/index.php?q=OSINT

This type of submission will be much more reliable than counting on the proxy to conduct your search and return an additional proxy-delivered page.

International Search Engines

Search engines based in the U.S. are not the primary search sites for all countries. Visiting search sites outside of the U.S. can provide results that will not appear on Google or Bing. In Russia, Yandex is the chosen search engine. Yandex offers an English version at yandex.com. These results are often similar to Google's; however, they are usually prioritized differently. In the past, I have found unique intelligence from this site when Google let me down. In China, most people use Baidu. It does not offer an English version; however, the site is still usable. Striking the "enter" key on the keyboard after typing a search will conduct the search without the ability to understand the Chinese text. New results not visible on Google or Bing may be rare, but an occasional look on these sites is warranted.

Yandex (yandex.com)

In a previous edition of this book, I only made a brief reference to Yandex and quickly moved on. In the past few years, I have discovered many advanced features of Yandex which justify an expanded section. Visually, the Yandex home page and search results pages do not possess additional search operators. These options are only available by issuing a direct command within your search. While this can be more cumbersome than a Google search, the results can include much new data. Some of these searches can be overkill for daily use, but those who conduct brand reputation monitoring or extensive background checks may take advantage of this.

Exact terms: Similar to Google and Bing, quotation marks will search for exact terms. Searching "Michael Bazzell" inside of quotes would search those terms, and would avoid "Mike" or "Bazel".

Missing word: You can search an exact phrase without knowing every word of the phrase. A search for "Open Source * Techniques" inside of quotation marks will identify any results that include that phrase with any word where the asterisk (*) is located. This identified not only results with the title of this book, but also results for "Open Source Development Techniques" and "Open Source Responsive Techniques". This search can be very useful for identifying a person's middle name. "Michael * Bazzell" produced some interesting results.

Words within the same sentence: The ampersand (&) is used in this query to indicate that you want to search for multiple terms. "Hedgehog & Flamingo", without the quotation marks, would identify any websites that contained both of those words within one sentence. If you want the results to only include sentences that have the two words near each other, you can search "Hedgehog /2 Flamingo". This will identify websites that have a sentence that includes the words Hedgehog and Flamingo within two words of each other.

Words within the same website: Similar to the previous method, this search identifies the searched terms within an entire website. "Hedgehog && Flamingo", without quotation marks, would identify pages that have both of those words within the same page, but not necessarily the same sentence. You can also control the search to only include results that have those two words

within a set number of sentences from each other. A search of "Hedgehog && /3 Flamingo", without the quotation marks, would identify websites that have those two words within three sentences of each other.

Include a specific word: In Google and Bing, you would place quotation marks around a word to identify pages that contain that word in them. In Yandex, this is gained with a plus sign (+). Michael +Bazzell would mandate that the page has the word Bazzell, but not necessarily Michael.

Search any word: In Google and Bing, you can use "OR" within a search to obtain results on any of the terms searched. In Yandex, this is achieved with the pipe symbol (|). This is found above the backslash (\) on your keyboard. A search of "+Bazzell Michael|Mike|M", without quotation marks, would return results for Michael Bazzell, Mike Bazzell, and M Bazzell.

Exclude a word: Google and Bing allow you to use a hyphen (-) to exclude a word in a search. Yandex does not technically support this, but it seems to work fine. The official Yandex operator is the tilde (~). A typical search would look like "Michael Bazzell ~ Mike", without the quotation marks. This would identify websites that contained Michael Bazzell, but not Mike Bazzell. I prefer to stick with the hyphen (-) until it no longer works.

Multiple identical words: This is a technique that I have needed several times in the past before I learned of Yandex's options. You may want to search for websites that contain a specific word more than once. An example might be if you are searching for someone that has two identical words in his or her full name. "Carina Abad Abad" would fit in this scenario. You could use quotation marks to identify the majority of the results, but you would filter out anything that was not exact such as Abad,Abad, Abad-Abad, or AbadAbad. This is where the exclamation point (!) comes in. A search of "!Carina !Abad !Abad", without quotation marks, would identify any results that included those three words regardless of spacing or punctuation.

Date specific searches: While Google provides a menu to filter your searches by date, Yandex makes you work harder for it. You must specify the date range within the search. The following queries should explain the options.

date:20111201..20111231 OSINT – Websites mentioning OSINT between December 1-31, 2011
date:2011* OSINT – Websites mentioning OSINT in the year 2011
date:201112* OSINT – Websites mentioning OSINT in December of 2011
date:>20111201 OSINT – Websites mentioning OSINT after December 1, 2011

Standard operators: Most of the operators explained earlier for Google and Bing should also work in Yandex. The commands for Site, Domain, Inurl, and Intitle should work the same way. Yandex maintains a list of operators at https://yandex.com/support/search/how-to-search/search-operators.html. All Yandex operators work together and multiple operators can be used to form very specific searches. Figure 8.09 displays the results for a search of any websites from 2013 with the phrase Michael Bazzell and the word OSINT while excluding the word Mike.

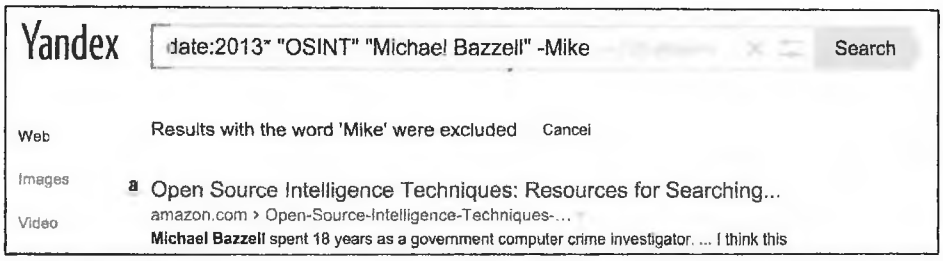


Figure 8.09: A custom Yandex search.

There are hundreds of additional international search engines. Of those, most are extremely specialized and do not offer great general search. The following have been most beneficial to my international investigations, in order of usefulness.

Yandex (yandex.com): Russia
Baidu (baidu.com): Asia
Goo (search.goo.ne.jp): Japan
Daum (search.daum.net): Korea
Parseek (parseek.com): Iran

I Search From (isearchfrom.com)

If you want to search Google within a version specified for another country, this site simplifies the process. Choose the country and language, and the tool will do the rest. While testing this service, I entered Japan as my country, English as my language, an iPad as my device, and OSINT as my search term. I was presented with google.co.jp search page in a tablet view. Many results were similar to the U.S. version, but all were in a unique order. I find this useful when searching for international targets when I do not want bias toward a U.S. user.

Search Engine Colossus (searchenginecolossus.com)

This website is an index of practically every search engine in every country. The main page offers a list of countries alphabetically. Each of these links connects to a list of active search engines in that country. I stay away from this service when searching American-based subjects. However, if my target has strong ties to a specific country, I always research the engines that are used in that area through this website.

DuckDuckGo (duckduckgo.com)

This search engine with a clean interface offers two unique services. It has gained a lot of popularity because it does not track anything from users. Engines, such as Google, record and maintain all of your search history and sites visited. This can be a concern to privacy advocates and those with sensitive investigations. Additionally, it uses information from crowd-sourced

websites such as Wikipedia and Wolfram Alpha to augment traditional results and improve relevance. You will receive fewer results here than at more popular search engines, but the accuracy of the results will improve.

Start Page (startpage.com)

Similar to DuckDuckGo, Start Page is a privacy-focused search engine that does not reveal your connection information to traditional search engines. The difference here is that Start Page only includes Google results versus DuckDuckGo's collaboration of multiple sources. The benefit to this is the ability to use Google's advanced search options while still protecting your identity. This includes filtering by date, images, and videos. Another benefit is the ability to open any result through a "proxy" link. This option, labeled "Proxy" next to each search result, opens the linked page through Start Page's servers and displays the content within their site. This protects your IP address from anyone monitoring connections at the target website. While this technique is not foolproof, it provides a valid layer of protection. My search strategy involves Start Page whenever I have a sensitive search that I do not want to associate with my computer or internet connection. This might include investigations that involve highly sensitive topics such as tech-savvy stalker suspects.

FTP Search

I believe that the searching of File Transfer Protocol (FTP) servers is one of the biggest areas of the internet that is missed by most online researchers. FTP servers are computers with a public IP address used to store files. While these can be secured with mandated access credentials, this is rarely the case. Most are public and can be accessed from within a web browser. The overall use of FTP to transfer files is minimal compared to a decade ago, but the servers still exist in abundance. I prefer the manual method of searching Google for FTP information. As mentioned earlier, Google and Bing index most publicly available data on FTP servers. A custom search string will be required in order to filter unwanted information. If I were looking for any files including the term "confidential" in the title, I would conduct the following search on Google and Bing.

inurl:ftp -inurl:(http|https) "confidential"

The result will include only files from ftp servers (inurl:ftp); will exclude any web pages (-inurl:(http|https); and mandate that the term "confidential" is present (""). I have located many sensitive documents from target companies with this query. The above search yielded 107,000 FTP results. However, these specific hits are not the only valuable data to pursue. Consider the following example. I want to locate PDF documents stored on FTP servers that contain "cisco" within the title or content, and I conduct the following search on Google.

inurl:ftp -inurl:(http|https) "cisco" filetype:pdf

This results in 20,000 options within multiple FTP servers hosted on numerous domains. The first result is hosted on the Southwest Cyberport FTP server and connects to a PDF document at the following address. It appears to be a chapter of a textbook.

ftp://ftp.swcp.com/pub/cisco/03chap01.pdf

Manually changing the last "01" to "02" loads the second chapter of the book. However, it is easier to eliminate the document name altogether and browse the directory titled "cisco". The first of the following addresses displays the contents of that folder, while the second displays the content of the "pub" folder. Copy these directly into a web browser to see the results.

ftp://ftp.swcp.com/pub/cisco/
ftp://ftp.swcp.com/pub/

This type of manual navigation will often reveal numerous publicly available documents that traditional searches withhold. I have located extremely sensitive files hosted by companies, government agencies, and the military. Most File Transfer Protocol (FTP) servers have been indexed by Google, but there are other third party options that are worth exploring. At the end of each description, I identify the number of results included for the search "Cisco" "PDF".

Global File Search (globalfilesearch.com)

Global File Search provides one of the few web-based engines for searching files on these public servers. At the time of this writing, the site claims to have indexed 243 terabytes of files in public FTP servers. Anyone searching for intelligence on any business should take a look at this site. The results are usually minimal, but very reliable.

"Cisco" "PDF": 121

Napalm FTP (searchftps.org)

This FTP search engine often provides content that is very recent. After each result, it displays the date that the data was last confirmed at the disclosed location. This can help locate relevant information that is still present on a server. While it generated the most results of all four services, many of them were no longer available on the target FTP servers. Some could be reconstructed with cached copies, but not all.

"Cisco" "PDF": 3,384

Mamont (mmnt.ru)

This Russian FTP server allows you to isolate search results by the country that is hosting the content. This is likely determined by IP address. While most of the filtered results will be accurate,

I recommend searching through the global results before dismissing any foreign options. My favorite feature of this engine is the "Search within results" option. After conducting my search, I checked this option and my search field was cleared. I entered "router" and clicked search again. I was prompted with the 436 results within my original hits that also included the word router. While this could have been replicated manually, I appreciate the option.

"Cisco" "PDF": 789

For comparison, Google found 19,600 results for inurl:ftp -inurl:(http|https) "Cisco" "PDF".

Nerdy Data (nerdydata.com/search)

Google, Bing, and other search engines search the content of websites. They focus on the data that is visually present within a web page. Nerdy Data searches the programming code of a website. This code is often not visible to the end user and exists within the HTML code, JavaScript, and CSS files with which most users are not familiar. This code can be extremely valuable to research in some scenarios. Viewing the source code of a website can be done by right-clicking on the background of a page and selecting "View Source". The following two examples should explain a small portion of the possibilities with this service.

In later chapters, you will learn about free services that try to identify additional websites that may be associated with your target website. The backbone of these services relies on the indexing of programming data of websites. Nerdy Data may be the purest way of searching for this data. If you were to look at the source code of one of my previous websites (no longer online), you would have seen at the bottom that I used a service called Google Analytics. This service identifies the number of visitors to a website and the general area where they are located. The following is the actual code that was present.

```
<script type="text/javascript">
try {var pageTracker = _gat._getTracker("UA-8231004-3");
pageTracker._trackPageview();
} catch(err) {}</script>
```

The important data here is the "UA-8231004-3". That was my unique number for Google Analytics. Any website with which I used the service with would have needed to have that number within the source code of the page. If you searched that number on Nerdy Data a few years prior, you would have received interesting results. Nerdy Data previously identified three websites that were using that number, including computercrimeinfo.com and two additional sites that I maintain for a law firm. You can often find valuable information within the source code of your target's website.

Many web designers and programmers steal code from other websites. In the past, this would be very difficult to identify without already knowing the suspect website. With Nerdy Data, you can

perform a search of the code of concern and identify websites that possess the data within their own source code. In 2013, I located a custom search website at the YGN Ethical Hacker Group that inspired me to create my own similar search service. I was curious if there were any other search websites that possessed this basic code that might give me more ideas. I looked at the source code of the website and located a small piece of code that appeared fairly unique to that service. I conducted a search on Nerdy Data for the following code.

http://yehg.net/q?[keyword]&c=[category] (q?yehg.net&c=Recon)

This code was within the JavaScript programming of the search website. The search results identified 13 websites that also possessed the same code. Two of these results were hosted on the creator's website, and offered no additional information. Three of the results linked to pages that were no longer available. Three of the results linked to pages that were only discussing the code within the target website and how to improve the functionality. However, four of the results identified similar search services that were also using the programming code searched. This revealed new search services that were related to the website in which I was interested. This same technique could be used to identify websites that are stealing proprietary code; locate pages that were created to attempt to fool a victim into using a cloned site; or validate the popularity of a specific programming function being used on hacking websites globally.

Qwant (qwant.com)

Qwant attempts to combine the results of several types of search engines into one page. It was launched in 2013 after two years of research. It has an easily digestible interface that displays results in columns titled Web, News, Knowledge, Social, and Shopping. There is a Google "feel" to it and the layout can be changed to your own preferences. A default search of my own name provided the expected results similar to Google and Bing. Clicking on the "People" tab at the top introduced new results not found on the other engines. The results included recent posts from Twitter, Facebook, LinkedIn, and Myspace from and about people with my name.

IntelTechniques Search Tool

At this point, you may be overwhelmed with the abundance of search options. I can relate to that, and I do not take advantage of every option during every investigation. During my initial search of a target, I like to rely on the basics. I first search Google, Bing, Yandex, and the smaller search engines. In order to assist with this initial search, I created a custom tool that will allow you to quickly get to the basics. Figure 8.10 displays the current state of this option, which is included in the search tools archive mentioned previously.

The search options will allow you to individually search directly through Google, Bing, Yahoo, Searx, Yandex, Baidu, Exalead, DuckDuckGo, Start Page, Google Newsgroups, Google Blogs, FTP Servers, data folders, Google Scholar, Google Patents, Google News, Google Newspapers, The Wayback Machine, and others. Across all options, each search that you conduct will open

within a new tab within your browser. The search all takes place on your computer within your browser, directly to the sources. The "Submit All" option will allow you to provide any search term that will be searched across all of the services listed. Each service will populate the results within a new tab in your internet browser. Regardless of the browser that you use, you must allow pop-ups in order for the tool to work. You can also use any of the search operators discussed previously within this tool, including quotation marks.

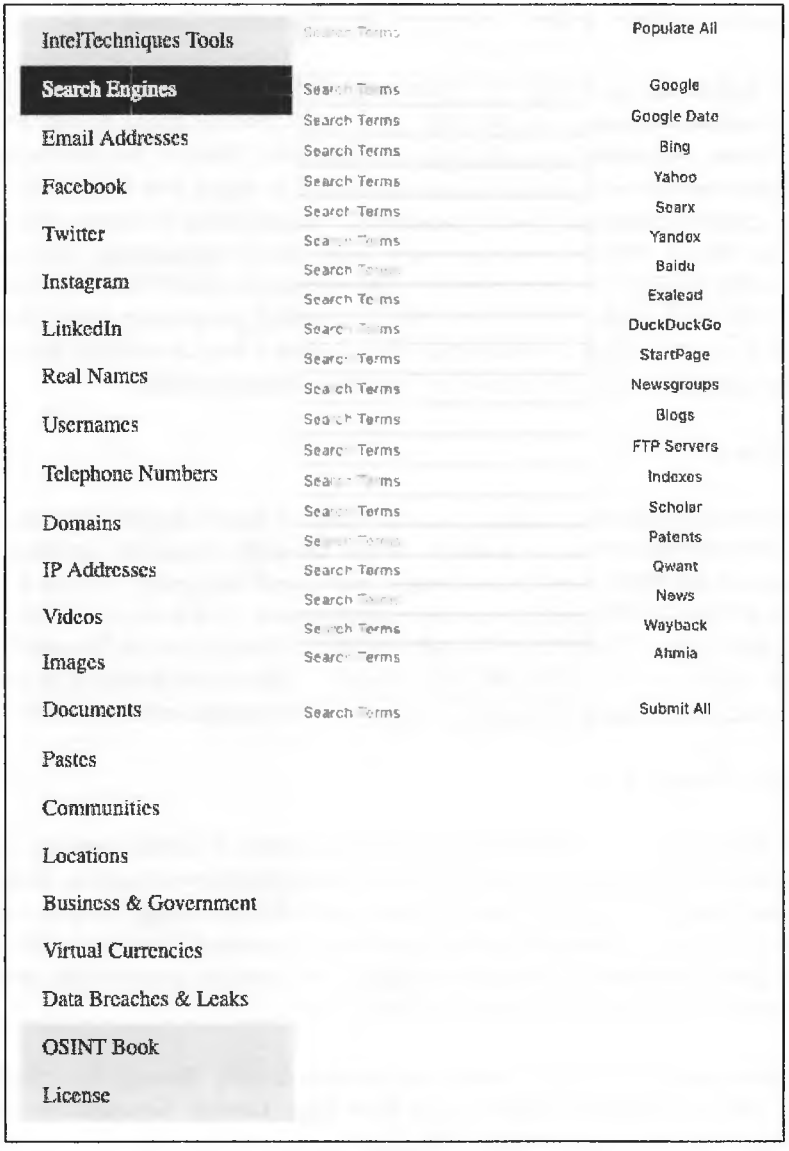


Figure 8.10: The IntelTechniques Custom Search Engines Tool.

CHAPTER NINE
SOCIAL NETWORKS: FACEBOOK

I hesitate writing anything within this chapter. It seems that most valuable Facebook search techniques disappeared in 2019. There are still many methods we can apply, but the future outlook for targeted queries is dim. I worry this chapter will become outdated before anything else, but there are some stable searches which should have longevity. Before we dive in, let's take a look at the 2019 Facebook timeline.

June 6th, 2019: This was the big date which began the decline of the Facebook graph. Previous profile queries all failed, and everything seemed doomed.

June 17th, 2019: Various researchers developed online search tools and browser extensions which brought back most of the Facebook Graph functionality. Hundreds of OSINT researchers flocked to these and restored our missing techniques.

August 1st, 2019: On this date, all browser-based extensions which leveraged the Facebook Graph stopped working. Facebook implemented new encryption which terminated all functionality.

August 2nd, 2019: The Facebook username to user ID conversion tool on my website stopped working. It appeared that my web server was being blocked by Facebook.

August 3rd, 2019, 12:33 pm ET: I switched the user ID conversion tool to a new web server, and all appeared to be working again.

August 3rd, 2019, 12:38 pm ET: Facebook began blocking my new web server.

September 8th, 2019: On this date, we saw methods such as using the Facebook Messenger application to search telephone numbers disappear, as well as most email search options. This appeared deliberate, and more evidence of Facebook's desire to lock down the platform.

October 1, 2019: This is when I wrote this chapter. You have been warned that many of the techniques here may no longer work by the time you read this. Here we go…

There are hundreds of social networks that act as storage for details of a person's life. Information that was once held privately within a small group of friends or family is now broadcasted to the world via public websites. Searching these websites has always been a high priority for intelligence gathering. This chapter should identify some new techniques that can be applied on any target. Before proceeding with any of the methods here, it is important to discuss covert accounts.

Covert Accounts

Some networks' search options are severely limited without being logged in to an account. In fact, I do not recommend trying any searches on Facebook without having a clean account in place. Covert accounts on all of the social networks mentioned here are free and can be completed using fictitious information. However, some networks will make this task more difficult than others. Google, Facebook, Twitter, Instagram, and Yahoo are known to make you jump through hoops before you are granted access. We begin this chapter discussing ways around this.

Email: It is vital that you possess a "clean" email address for your covert accounts. Every social network requires an email address as a part of account registration, and you should never use an already established address. Later chapters explain methods for researching the owners behind email addresses, and those techniques can be applied to you and your own accounts. Therefore, consider starting fresh with a brand-new email account dedicated toward use for covert profiles.

The choice of email provider is key here. I do not recommend GMX, ProtonMail, Yahoo, Gmail, MSN, or any other extremely popular providers. These are heavily used by spammers and scammers, and are therefore more scrutinized than smaller providers. My preference is to create a free email account at **Fastmail** (fastmail.com). This established mail provider is unique in two ways. First, they are one of the only remaining providers which do not require an established email address in order to obtain a new address. This means that there will be no connection from your new covert account to any personal accounts. Second, they are fairly "off-radar" from big services such as Facebook, and are not scrutinized for malicious activity.

Fastmail will provide anyone unlimited free accounts on a 30-day trial. I suggest choosing an email address that ends in fastmail.us instead of fastmail.com, as that domain is less used than their official address. This is a choice during account creation. Once you have your new email address activated, you are ready to create covert profiles. Note that the free trial terminates your access to this email account in 30 days, so this may not be best for long-term investigations. Personally, I possess a paid account which allows me 250 permanent alias email addresses.

Facebook: This is by far the most difficult in terms of new account creation. For most new users, Facebook will require you to provide a cellular telephone number where a verification text can be sent and confirmed. Providing VOIP numbers such as a Google Voice account will not work anymore. I have found only one solution. Turn off any VPN, Tor Browser, or other IP address masking service and connect from a residential or business internet connection. Make sure that you have cleared out all of your internet cache and logged out of any accounts. Instead of creating a new account on facebook.com, navigate directly to m.facebook.com. This is the mobile version of their site which is more forgiving on new accounts. During account creation, provide the Fastmail email address that you created previously. In most situations, you should bypass the requirement to provide a cellular number. If this method failed, there is something about your

computer or connection that is making Facebook unhappy. Persistence will always equal success eventually. I find public library Wi-Fi our best internet option during account creation.

Twitter: Many of the Twitter techniques presented later will not require an account. However, the third-party solutions will mandate that you be logged in to Twitter when using them. I highly recommend possessing a covert account before proceeding. As long as you provide a legitimate Fastmail email address from a residential or business internet connection, you should have no issues. You may get away with using a VPN to create an account, but not always.

Instagram: Instagram is similar to Twitter. Unless you are creating multiple accounts per day from the same internet connection, you should receive no resistance in creating an anonymous account.

Google/Gmail/Voice: While Google has become more aggressive at refusing suspicious account registrations, they are still very achievable. As with the previous methods, Google will likely block any new accounts that are created over Tor or a VPN. Providing your Fastmail address as an alternative form of contact during the account creation process usually satisfies their need to validate your request. I have also found that they seem more accommodating during account creation if you are connected through a standard Chrome browser versus a privacy-customized Firefox browser. While that seems a bit shady, it makes sense (Google owns Chrome).

Some readers may assume that they can simply use their personal and accurate social network account to search for information. While this is indeed possible, it is risky. Some services, such as Instagram may never indicate to the target that your specific profile was used for searching. Others, such as Facebook, will indeed eventually notify the target that you have an interest in him or her. This is usually in the form of friend recommendations. On any service, you are always one accidental click away from sending a friend request from your real account to the suspect. For these reasons, I never use a real social network profile during any investigation. I like to maintain multiple accounts at all times in case one is suspended or deleted by the social network.

The topic of undercover operations quickly exceeds the scope of this book about search techniques. Volumes could be written about proper photo use and the psychology of posts in order to create an assumption that the person is real. For our purposes, we only need a covert account. We will not add any personal details or photos. We will not post any messages privately or publicly. We simply need to be logged in to real accounts in order to pacify the social networks. I will assume that you now have social network accounts created that are not in your real name and possess no personal identification about you. It is time to dig into social network profiles and extract data.

Facebook users tend to keep their information a little more secure than users of other social networking sites. By default, a new Facebook user must specify the privacy settings to their account during the creation of their profile. This is mostly thanks to privacy advocates who

continuously protest Facebook's privacy policies. Many of these user settings simply do not promote privacy and leave the user's information exposed for anyone to see. This section will explain numerous ways to obtain user information that is not visible on the public profile.

Official Facebook Options

Once logged in, a simple search field will be present at the top of any Facebook page. Typing in the target's real name should lead to some results. Unlike Twitter, Facebook users usually use their real name when creating a profile. This profile is also usually linked to an employer, graduating high school class, or college alumni. Once a user's profile is located, the default view is the "timeline" tab. This will include basic information such as gender, location, family members, friends, relationship status, interests, education, and work background. This page will also commonly have a photo of the user and any recent posts on their page. With billions of active users, it will be likely that you will locate several user profiles under the same name as your target. There are a few things that you can do to find the right person.

If your target's name is Tom Johnson, you have your work cut out for you. This does not mean that you will never find his Facebook page, but you will need to take additional steps to get to your target. When searching the name, several possibilities may appear in a drop-down menu. This is obviously not the complete list of Tom Johnsons that are present on Facebook. At the bottom of this list is an option to see all of the profiles with your target name. This is also not the complete list. Clicking the "people" tab in the upper menu will take you to a scrolling list of all profiles matching your search. You can look through these and hope to identify your target based on the photo, location, or additional information displayed in this view.

Next, consider using the traditional filter options available on most Facebook pages. Figure 9.01 displays the main filter bar at the top of every Facebook results page. This will seek Posts, Photos, Videos, Places, Groups, Events, and other options based on the provided search terms. Within each section is a secondary filter on the left. Figure 9.01 demonstrates the "Photos" option after searching "Chicago". The filter on the left can further filter results based on location and date. These filters can also be applied using direct URLs. After a search of the term "OSINT", the following addresses replicate each of the filters on a standard Facebook results page.

All: https://www.facebook.com/search/top/?q=osint
Posts: https://www.facebook.com/search/posts/?q=osint
People: https://www.facebook.com/search/people/?q=osint
Photos: https://www.facebook.com/search/photos/?q=osint
Videos: https://www.facebook.com/search/videos/?q=osint
Marketplace: https://www.facebook.com/search/marketplace/?q=osint
Pages: https://www.facebook.com/search/pages/?q=osint
Places: https://www.facebook.com/search/places/?q=osint
Groups: https://www.facebook.com/search/groups/?q=osint
Apps: https://www.facebook.com/search/apps/?q=osint

Events: https://www.facebook.com/search/events/?q=osint
Links: https://www.facebook.com/search/links/?q=osint

If these filters still provide too many results, we will filter further in a moment.

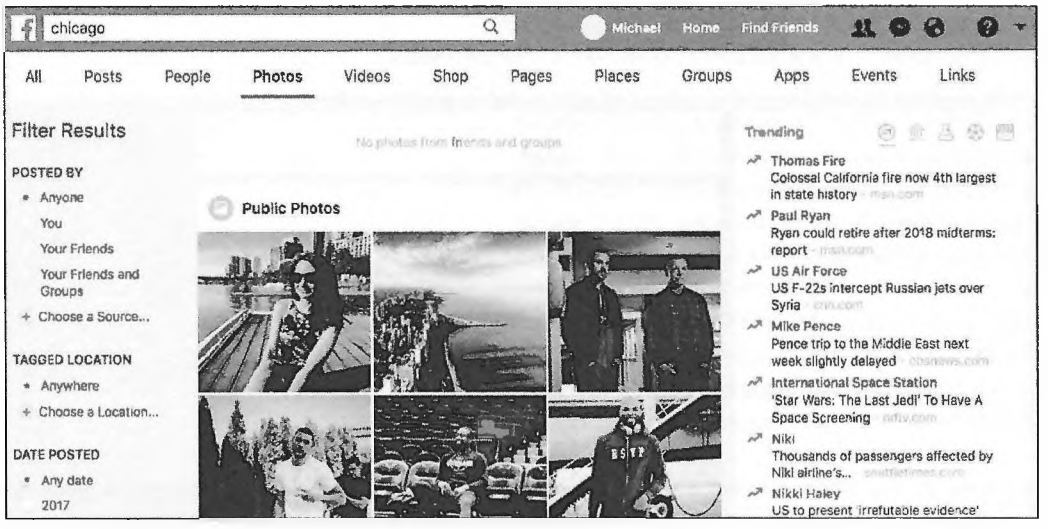


Figure 9.01: The Facebook filter options.

If you have located a Facebook profile of interest, you may be able to click the various sections in order to reveal all publicly available content. My preference is to query via direct URL so that I know I did not miss anything. Assume that your target is Mark Zuckerberg. His profile is available at the following URL.

https://www.facebook.com/zuck

This indicates his Facebook username is "zuck". We can apply this to the following URLs, each which connect directly to the associated public information page. These direct URLs will be beneficial to our Facebook tool presented at the end of the chapter.

Timeline: https://www.facebook.com/zuck
About: https://www.facebook.com/zuck/about
Employment: https://www.facebook.com/zuck/about?section=work
Education: https://www.facebook.com/zuck/about?section=education
Locations: https://www.facebook.com/zuck/about?section=living
Contact Info: https://www.facebook.com/zuck/about?section=contact-info
Basic Info: https://www.facebook.com/zuck/about?section=basic-info
Relationships: https://www.facebook.com/zuck/about?section=relationship
Family Members: https://www.facebook.com/zuck/about?section=family

Bio: https://www.facebook.com/zuck/about?section=bio
Life Events: https://www.facebook.com/zuck/about?section=year-overviews
Friends: https://www.facebook.com/zuck/friends
Profile Photos: https://www.facebook.com/zuck/photos
Photo Albums: https://www.facebook.com/zuck/photos_albums
Videos: https://www.facebook.com/zuck/videos
Check-Ins: https://www.facebook.com/zuck/places_visited
Recent Check-Ins: https://www.facebook.com/zuck/places_recent
Sports: https://www.facebook.com/zuck/sports
Music: https://www.facebook.com/zuck/music
Movies: https://www.facebook.com/zuck/movies
TV Shows: https://www.facebook.com/zuck/tv
Books: https://www.facebook.com/zuck/books
Likes: https://www.facebook.com/zuck/likes
Events: https://www.facebook.com/zuck/events
Facts: https://www.facebook.com/zuck/did_you_know
Reviews: https://www.facebook.com/zuck/revies
Notes: https://www.facebook.com/zuck/notes

This presents the end of basic Facebook search techniques. Everything presented until now should apply for a while, and these URL structures should not change soon. Analyzing all of the evidence identified through these URLs should present substantial information. Many of these pages, such as a person's timeline, will load continuously. Pressing the space bar on your keyboard should load everything until the end. I have found taping down the space bar helpful for long pages. From here, we must dig deeper into profile data and apply some fairly technical methods in order to get to the next level. This will not be too difficult thanks to the automated tools presented later. For now, let's focus on an individual profile.

Profile Information

At this point, you should be able to locate a target's profile by name; analyze the publicly available content; and search by topic. That is just the tip of the iceberg. Facebook collects a lot of additional information from everyone's activity on the social network. Every time someone "Likes" something or is tagged in a photo, Facebook stores that information. Extracting these details can be difficult.

In order to conduct the following detailed searches, you must know the user number of your target. This number is a unique identifier that will allow us to search otherwise hidden information from Facebook. Prior to mid-2015, the easiest way to identify the user number of any Facebook user was through the Graph API. While you were on a user's main profile, you could replace "www" in the address with "graph" and receive the profile ID number of that user. This no longer works because Facebook removed the ability to search their graph API by username. However, we can still obtain this powerful number through two search options.

The first option involves viewing the source code of any user's Facebook profile. The process for this will vary by browser. In Firefox and Chrome, simply right-click on a Facebook profile page and select "View Page Source". Be sure not to hover on any hyperlinks during the right-click. A new tab should open with the text-only view of the source code of that individual profile. Within the browser, conduct a search on this page for "entity_id". This will identify a portion of the code within this page that contains that specific term. As an example, the following is the source code immediately before and after the search result.

"entity_id":"651620441","profile_session_id"

In this example, the user ID of this profile is 651620441. We will use this number for numerous searches within the next instruction. Some users prefer to look at the URLs of a target's photos in order to identify the user ID, but I believe this is bad practice. If a user has no photos, this will not work. Also, Facebook's photo displays often hide this information from plain site. I prefer to rely on the source code view or my Facebook tools for this identification. This number will allow us to obtain many more details about the account. Until July of 2019, there were dozens of online search tools which would identify the user ID number (4) when supplied a username (zuck). Almost all of these stopped functioning, including my own, when Facebook began aggressively blocking these search tools. While you may find an online option which still functions, we should not rely on these.

If you find yourself searching for the user ID number often, you might consider a "bookmarklet" for your browser. This is a small amount of JavaScript which loads as a bookmark in your browser. Pressing it displays the user ID number of the Facebook profile currently displayed. The following steps will place a button within your Firefox browser for this task.

- In Firefox, click on View > Toolbars > Bookmarks Toolbar to make it visible.
- Navigate to https://pastebin.com/zMdwsWps.
- Copy all of the text within the RAW Paste Data window.
- Within the Bookmarks Toolbar in Firefox (directly below the address field), right-click and select "New Bookmark".
- Paste the content from Pastebin in the "Location" field.
- Provide a name of FBUID in the "Name" section and click "Add".

Figure 9.02 displays this technique. The bookmarklet is visible in the upper left bookmarks toolbar (FBUID), I am viewing the page of "zuck", and his user number is "4". I rely on this tool heavily, but I also know how to replicate the results myself in case the script should stop functioning. If Facebook were to ever change the "entity_id" label, you could search the source code on the "zuck" page for his number (4). This should identify any label changes, and you can modify your search if needed. Now that we can locate a user number of our target, let's do something with it.



Figure 9.02: A Facebook bookmarklet which obtains user ID numbers.

Facebook Base64 Encoding

Prior to June of 2019, a simple URL would display content posted by an individual. As an example, the following URL would display all photos posted by a specific user (4).

https://facebook.com/search/4/photos-by

This technique no longer works, and the replacement method is much more difficult. Instead of "facebook.com/search", our base URL is as follows.

https://facebook.com/search/photos/?q=*&epa=FILTERS&filters=

This is followed by the structure of the following.

{"rp_author":"{\"name\":\"author\",\"args\":\"[USERID]\"}"}

However, it must be presented in Base64 format, which would be the following.

eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

Therefore, the entire URL would be the following.

https://www.facebook.com/search/photos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

Confused? I sure was. It took me a while to process what was going on here. Let's start over, and approach the creation of a custom URL in three phases. First, let's tackle the Facebook search URL. In the previous example, I displayed the following.

https://facebook.com/search/photos/?q=*&epa=FILTERS&filters=

Let's break this down. The following explains each section of this URL

https://facebook.com/	The Facebook domain
search/	Instructs Facebook to conduct a search
photos/	Specifies the type of information desired
?q=*	Searches everything (*)
&epa=FILTERS&filters=	Finishes the URL with a filter demand

Next, we must formulate our target data and convert it to a specific type of encoding called Base64. This is likely used because it is extremely common and can be generated by any browser. Previously, I presented the following data.

{"rp_author":"{\"name\":\"author\",\"args\":\"[USERID]\"}"}

This tells Facebook that we are searching for information from a specific profile (author), and the [USERID] should contain your targets user ID number as determined earlier. If we were targeting "zuck", and knew his user number was "4", we would have the following data.

{"rp_author":"{\"name\":\"author\",\"args\":\"4\"}"}

Notice the position of "4" as the user number. Now, we must convert this data into Base64. I prefer the website https://codebeautify.org/base64-encode and we can check our work at the decoding version at https://codebeautify.org/base64-decode. When I copy and paste the previous data into this website, I receive the following.

eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ==

The final two "==" are optional, and not necessary. When I paste this value into the decoder on this website, I receive the exact data originally entered. Let's take a look at the screen captures. Figure 9.03 (top) displays the desire text data, including my target's user ID number. Figure 9.03 (bottom) displays the result coded in Base64. Figure 9.04 (top) displays the opposite technique by converting the previous Base64 result back to the original desired data in Figure 9.04 (bottom). I realize this is confusing, but our search tools will simplify all of this in a moment.

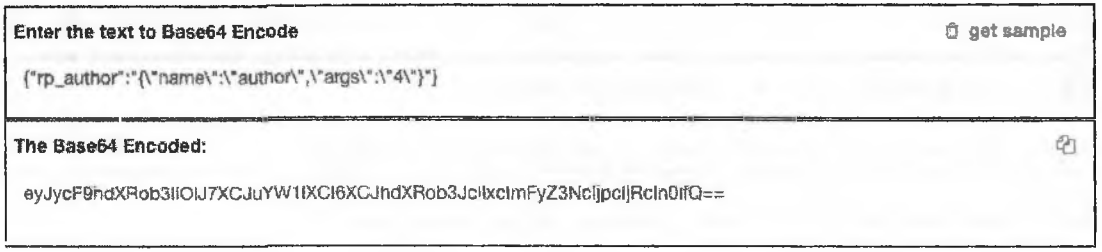


Figure 9.03: Encoding text to Base64 on https://codebeautify.org/base64-encode.

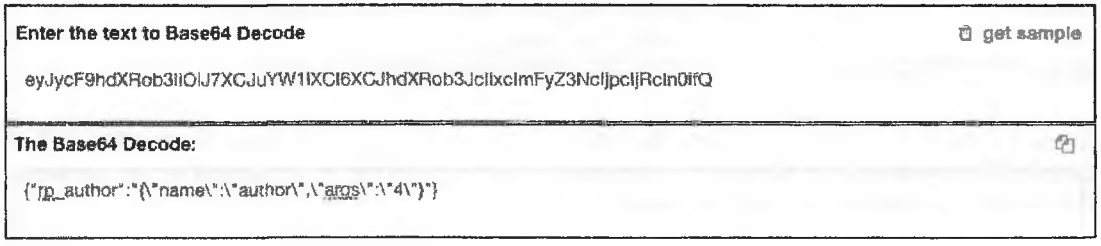


Figure 9.04: Decoding text to Base64 on https://codebeautify.org/base64-decode.

Let's take a look at the entire URL as follows. Below it is a breakdown.

https://www.facebook.com/search/photos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

https://facebook.com/	The Facebook domain
search/	Instructs Facebook to conduct a search
photos/	Specifies the type of information desired
?q=*	Searches everything (*)
&epa=FILTERS&filters=	Finishes the URL with a filter demand
eyJycF9hdXRob3IiOiJ7X…	{"rp_author":"{\"name\":\"author\",\"args\":\"4\"}"}

Figure 9.05 displays the results of this URL. Facebook has provided images posted by our target, some of which do not appear on his profile. Clicking "See All" opens even more images. Let's start over with another example, this time focusing on public posts.

Assume we want to identify posts created by "zuck". We know his user number is "4", so we would want to combine https://facebook.com/search/posts/?q=*&epa=FILTERS&filters= with the Base64 encoding of {"rp_author":"{\"name\":\"author\",\"args\":\"4\"}"}. This creates the following URL.

https://www.facebook.com/search/posts/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

A partial result can be seen in Figure 9.06. These are public posts, but some may not be from his profile. In one investigation, I had a suspect with absolutely no posts within his timeline. However, this method identified hundreds of posts he made on other people's profiles. These included comments within posts and public messages to others.



Figure 9.05: Image results from a Base64 search.

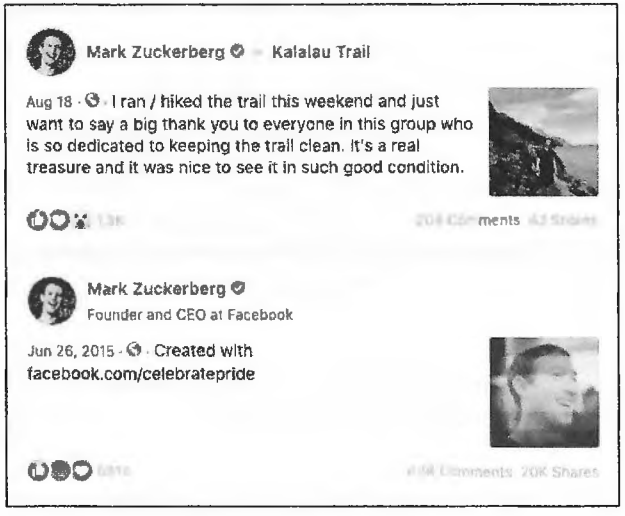


Figure 9.06: Post results from a Base64 search.

Let's make this simpler. In a moment, I will present the most common URL queries which I have found valuable to my investigations. Then, I will present the portion of my custom Facebook tools which automates the entire process. First, below is a summary of each search by section.

Posts by User:	Public posts by a specific user on any page
Photos by User:	Public photos by a specific user on any page
Videos by User:	Public videos by a specific user on any page
Posts by Location:	Public posts submitted from a specific city
Photos by Location:	Public photos submitted from a specific city
Videos by Location:	Public videos submitted from a specific city
Profiles by Employer:	Profiles associated with employees of a specific business
Profiles by City:	Profiles associated with residents of a specific city
Profiles by School:	Profiles associated with students of a specific school
Common Friends:	Common profiles which are friends with two targets
Posts by Date:	Posts by keyword filtered by a date range

If your target was the Facebook user number "4", the URLs would be as follows. Notice that the Base64 encoding is identical on each.

Posts by User:
https://www.facebook.com/search/posts/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3Ii
OiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

Photos by User:
https://www.facebook.com/search/photos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3Ii
OiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

Videos by User:
https://www.facebook.com/search/videos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3Ii
OiJ7XCJuYW1lXCI6XCJhdXRob3JcIixcImFyZ3NcIjpcIjRcIn0ifQ

If your target was Chicago, user number "108659242498155", the URLs would be as follows.

Posts by Location:
https://www.facebook.com/search/posts/?q=*&epa=FILTERS&filters=eyJycF9sb2NhdGlv
biI6IntcIm5hbWVcIjpcImxvY2F0aW9uXCIsXCJhcmdzXCI6XCIxMDg2NTkyNDI0OTgxN
TVcIn0ifQ

Photos by Location:
https://www.facebook.com/search/photos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJsb2NhdGlvblwiLFwiYXJnc1wiOlwiMTA4NjU5MjQyNDk4MTU1XCJ9In0

Videos by Location:
https://www.facebook.com/search/videos/?q=*&epa=FILTERS&filters=eyJycF9hdXRob3IiOiJ7XCJuYW1lXCI6XCJsb2NhdGlvblwiLFwiYXJnc1wiOlwiMTA4NjU5MjQyNDk4MTU1XCJ9In0

Profiles by City (with a name of Tom):
https://www.facebook.com/search/people/?q=tom&epa=FILTERS&filters=eyJjaXR5Ijoie1wibmFtZVwiOlwidXNlcnNfbG9jYXRpb25cIixcImFyZ3NcIjpcIjEwODY1OTI0MjQ5ODE1NVwifSJ9

If your target was Harvard, user number "105930651606", the URLs would be as follows.

Profiles by Employer (with a name of Tom):
https://www.facebook.com/search/people/?q=tom&epa=FILTERS&filters=eyJlbXBsb3llciI6IntcIm5hbWVcIjpcInVzZXJzX2VtcGxveWVyXCIsXCJhcmdzXCI6XCIxMDU5MzA2NTE2MDZcIn0ifQ

Profiles by School (with a name of Tom):
https://www.facebook.com/search/people/?q=tom&epa=FILTERS&filters=eyJzY2hvb2wiOiJ7XCJuYW1lXCI6XCJ1c2Vyc19zY2hvb2xcIixcImFyZ3NcIjpcIjEwNTkzMDY1MTYwNlwifSJ9

As a reminder, I obtained the user ID numbers for the profiles of "zuck" and "harvard" by searching the source code of each profile for "entity_id", or with the bookmarklet explained previously. However, city profiles do not possess this data. Fortunately, the user ID for cities is within the URL. When I searched for Chicago on Facebook, I clicked the official City of Chicago profile which possessed the following URL.

https://www.facebook.com/places/Things-to-do-in-Chicago-Illinois/108659242498155/

The last group of numbers (108659242498155) is the user ID number.

It is extremely important to note that I did not discover this Base64 conversion technique. The online researcher NEMEC (https://twitter.com/djnemec) was the first to post about this method. Practically every online search tool generating these types of URLs, including mine, is due to the work of this person. I send many "thanks" on behalf of the entire OSINT community for this work.

IntelTechniques Facebook Search Tool

If you have not already skipped to the next chapter after all of this confusion, it is now time to simplify the process. Finally, let's use the custom Facebook search tool to replicate all of this work. Figure 9.07 displays the current version of the tool. The following pages walk you through each section with real examples.

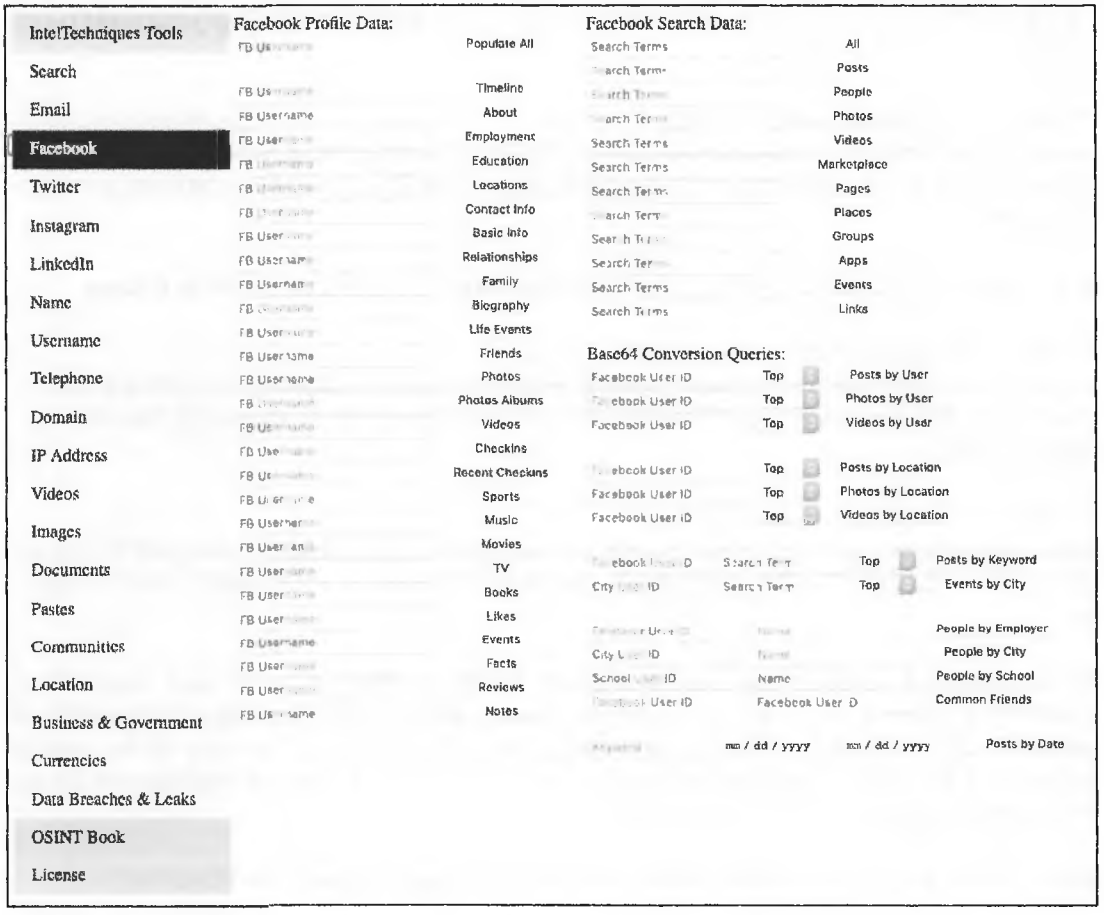


Figure 9.07: The IntelTechniques Facebook Tool.

The first input box in the "Facebook Profile Data" section allows you to enter a Facebook username, such as zuck, and the tool will populate that data within all of the search options in that section. This makes it easy to begin your queries. You can then use the submit buttons to search each entry, such as Timeline or Photos. Note there is no "Submit All" option here. This is due to Facebook's scrutiny into accounts which automate any queries. Submitting too many requests simultaneously will result in a suspended account.

The next section titled "Facebook Search Data" allows you to replicate the instruction on keyword searches throughout all twelve sections of Facebook, such as Posts, Photos, and Pages. This presents immediate results without relying on Facebook to present your desired data from a generic keyword search.

The "Base64 Conversion Queries" section presents the true power behind this tool. All of the technical search techniques presented within the previous pages are replicated here without the need to do any of the work. Instead of summarizing this section, let's run through a demonstration of each option.

Posts/Photos/Videos by User

I entered the user number of "4" (zuck) in the first portion, as seen in Figure 9.08 (left). This allows me to query posts, photos, and images posted by that user to any public pages. The dropdown menu allows me to filter by year, if desired, which is helpful if the target posted a lot of content. This option is available for all three queries. This often identifies content unavailable within the target's own profile.

Posts/Photos/Videos by Location

I entered the user number of "108659242498155" (Chicago) in the first portion, as seen in Figure 9.08 (right). This allows me to query posts, photos, and images posted from that location to any public pages. The dropdown menu allows me to specify a year, if desired. This often identifies content unavailable through traditional search techniques.



Figure 9.08: IntelTechniques Search Tool options with date filtering.

People by Employer, City, and School

I entered the user number of "105930651606" (Harvard) in the first and third boxes, as seen in Figure 9.09. This allows me to query people's profiles by their employer or school affiliation. These search options require at least one additional piece of information. In this example, I searched the name "Tom". Notice the difference in the results within Figure 9.10. The left result identifies people named Tom who WORK at Harvard, while the right displays people named

Tom who ATTENDED Harvard. This search option can quickly identify your target when you only know a small amount of information, or if they are using a false last name.

Common Friends

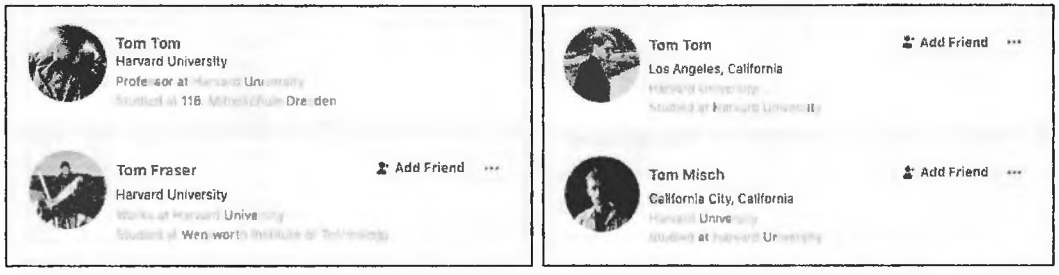
This query requires one of your targets to display a public friends list. Enter the user ID number of two subjects and identify any mutual friends they have in common. This is beneficial when trying to find associates of two suspects while avoiding family, friends, and colleagues of only one of them.

Posts by Date

This option allows you to filter to a specific set of dates and conduct a keyword search. Figure 9.09 displays a query for any mention of the term "osint" during the month of September 2019. Clicking these fields presents an interactive calendar for easy selection.



Figure 9.09: IntelTechniques Search Tool options for profile search.



9.10: Facebook results from an IntelTechniques tool query.

I suspect that these queries will change over time. I do not present the source code of these specific queries within the book because they are quite complicated and lengthy. If this tool should be updated, I will make an announcement at https://inteltechniques.com/osintbook.

Facebook ID Creation Date

Digital forensics enthusiast and private investigator Josh Huff at LearnAllTheThings.net has conducted a lot of research into the assignment of user ID numbers to Facebook profiles. We know that these numbers are assigned in chronological order, but the intelligence goes much further beyond that. His research could fill several pages, and in some situations, he can identify the month when a specific user account was created. For the sake of space and most useful details, he has provided the following to me for publication.

Facebook transitioned from 32-bit numbers, such as 554432, to 64-bit numbers that begin with 100000 between April and December of 2009. Therefore, if your target's number is less than 100000000000000, the account was likely created prior to April 2009. An account with a 15-digit ID number would have been created after December 2009. We can break down the numbers by year. The following are rough estimates that should only be used as a general guideline.

2006: Numbers less than 600400000
2007: 600400000 - 1000000000
2008: 1000000000 - 1140000000
2009: 1140000000- 100000628000000
2010: 100000629000000 - 100001610000000
2011: 100001611000000 - 100003302000000
2012: 100003303000000 - 100004977000000
2013: 100004978000000 - 100007376000000
2014: 100007377000000 - 100008760000000
2015: 100008761000000 - 100010925000000
2016: 100010926000000 - 100014946000000
2017: 100014947000000 - 100023810000000
2018: 100023811000000 -

Facebook Friends Extraction

I was recently presented a Facebook scenario without any obvious solution. Assume that you find the Facebook profile of your target, and there is a large list of "Friends". You want to document this list, and a screenshot is not good enough. You want to capture the hyperlinks to each account and have a data set that can be manipulated or imported elsewhere. There are several outdated online tools that claim to "scrape" this data, but none of them work. In 2016, Facebook changed their Terms of Service (TOS) which now blocks any type of scraping of Friend data. Furthermore, the Facebook API no longer allows the display of this data either. There is a solution, and it will involve minimal geek work.

First, identify the page of your target. For this example, I will use the following public profile:

https://www.facebook.com/darya.pino/friends

She has several friends, so I will hold down the space bar on my keyboard to load the entire page. I will now highlight her entire friends list and use ctrl-c or right-click>copy to copy the data. I find it best to click directly above the left side of the first friend and hold until the lower right area of the last friend. The friends list should highlight. Now, open Microsoft Excel. Click on the "B" in column B to highlight the entire column. Paste the content with either ctrl-v or right-click>paste. This will appear disorganized, but the data is there. The images will be on top of the user data, which will not work for a final report.

Use F5 to launch the "Go To" menu and select "Special" in the lower left. Select "Objects" and click OK. This will select all of those images. Hit the delete key to remove them. You will now see only the text data (with hyperlinks). Now click on the "A" in column A and paste the friend content again with either ctrl-v or right-click>paste. Right-click any cell in this column and choose "Clear Contents". This will remove any text, but keep the images.

Place your mouse in between columns A and B and resize column A to be a bit larger than one of the images. Do the same with Column B to fit in all of the text. Use the "Find and Replace" feature to find every instance of "Add Friend" and replace it with nothing. This will remove those unnecessary entries. In the "Home" menu, choose "Format" and then "Auto Fit Row Height". This will eliminate unnecessary spacing. Select Column B and Left Justify the text. Your final result will be a clean spreadsheet with all of the images, names, and active links from your target's Facebook "friends" page. This is not the cleanest way of doing things, but it will work.

Email Search

As stated previously, we lost all standard email address search options within Facebook in 2019. However, there is one remaining technique which allows submission of an email address and provides identification of an associated profile. However, it is not simple or straightforward. It will take some work to set up, but will then be available as needed. This is often referred to as the "Page Role" trick. The premise is that you create a Facebook business page and then assign another profile to possess management rights. When you enter the email address of the target, Facebook confirms the name and profile to make sure you truly want to give away authorization to the person. We can then cancel the request without any notification to the target. The following steps replicate this technique.

- While logged in to any covert Facebook profile, click "Create" and then "Page" in the upper-right corner. Click "Get Started" under the "Business or Brand" option.
- Assign a random name to your profile, select any category, and click "Continue".
- Skip all optional steps.
- Once you see the new profile, click the "Settings" button in the upper-right.
- On the new menu, click "Page Roles" in the left column.
- In the "Assign a New Page Role" section, enter the target email address.

This should present any Facebook profiles associated with the address entered. Figure 9.11 displays a result. I entered an email address in this field and was presented the only Facebook profile which was associated with the account. I can now search this full name within Facebook, look for the image previously displayed, and scour the target profile for valuable information.

Facebook continuously makes both minor and major changes to their search functions. Some of these instructions may not work one day and work fine the next. Your mileage will vary as Facebook scrutinizes your covert profiles, VPN protected networks, and overall "vibe" as a fake user. Hopefully, this chapter has given you new ideas on ways to completely analyze your next target on Facebook and ideas to circumvent the next roadblocks.

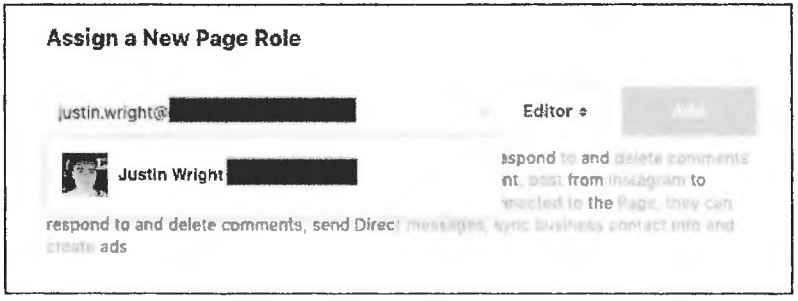


Figure 9.11: An email search through Facebook under the Page Role menu.

Chapter Ten
Social Networks: Twitter

Twitter is a social network and microblogging service that limits most posts to 280 characters. In 2019, Twitter reported that there were over 500 million Twitter posts, or "Tweets", posted every day. Basically, users create a profile and post Tweets announcing their thoughts on a topic, current location, plans for the evening, or maybe a link to something that they feel is important. A user can "follow" other users and constantly see what others are posting. Likewise, a user's "followers" can constantly see what that user is up to on a constant basis. The premise is simply sharing small details of your life for all of your friends to see, as well as the rest of the world. Most users utilize the service through a cellular phone, which can capture the user's location and broadcast the information if the location feature is enabled and the user approves sharing. Obtaining information from Twitter can be conducted through various procedures.

Twitter Search (twitter.com/explore)

This is the official site's search interface, but it is nothing different than the search field at the top of any Twitter profile or search result. I only present this because navigating to twitter.com often offers a signup page, but no option to search.

Twitter Advanced Search (twitter.com/search-advanced)

This page will allow for the search of specific people, keywords, and locations. The problem here is that the search of a topic is often limited to the previous seven to ten days. Individual profiles should display Tweets as far back as you are willing to scroll. This can be a good place to search for recent data, but complete archives of a topic will not be displayed. The following explains each section.

All of these words: The order of wording is ignored here, and only the inclusion of each of the words entered is enforced.

This exact phrase: Every Twitter search takes advantage of quotes to identify exact word placement. Optionally, you can conduct the search here to get precise results without quotes.

Any of these words: You can provide multiple unique terms here, and Twitter will supply results that include any of them. This search alone is usually too generic.

None of these words: This box will filter out any posts that include the chosen word or words.

These Hashtags: This option will locate specific posts that mention a topic as defined by a Twitter hashtag. This is a single word preceded by a pound sign (#) that identifies a topic of

interest. This allows users to follow certain topics without knowing usernames of the user submitting the messages.

From these accounts: This section allows you to search for Tweets from a specific user. This can also be accomplished by typing the username into the address bar after the Twitter domain, such as twitter.com/JohnDoe92. This will display the user's profile including recent Tweets.

To these accounts: This field allows you to enter a specific Twitter username. The results will only include Tweets that were sent to the attention of the user. This can help identify associates of the target and information intended for the target to read.

Mentioning these accounts: While these messages might not be in response to a specific user, the target was mentioned. This is usually in the form of using "@". Anyone mentioning me within a Tweet may start it with @inteltechniques.

Dates: The final option allows you to limit a search to a specific date range. We will do this manually in just a moment.

Overall, I do not ever use the Twitter Advanced Search page. In the following pages, we are going to replicate these searches within our own custom tool options, which will be much more powerful than these standard solutions. The results of any of these searches can provide surprisingly personal information about a target, or generic content that includes too much data to be useful. This data can be used for many types of investigations. Law enforcement may use this data to verify or disprove an alibi of a suspect. When a suspect states in an interview that he was in Chicago the entire weekend, but his Twitter feed displays his Tweet about a restaurant in St. Louis, he has some explaining to do. Private investigators may use this content as documentation of an affair or negative character. Occasionally, a citizen will contact the authorities when evidence of illegal activity is found within a person's Tweets. The possibilities are endless. First, let's find your target's Twitter profile.

Twitter Person Search

Locating your target's Twitter profile may not be easy. Unlike Facebook, many Twitter users do not use their real name as their profile name. You need a place to search by real name. I recommend Twitter's "People" search page. Search any target name and click the "People" tab in the top horizontal menu. Scrolling down the list I can look through the photo icons and brief descriptions to identify my target. Clicking on the username will open the user's Twitter profile with more information. This does not require a login.

Followerwonk Bios (followerwonk.com/bio)

Twitter's People option can be great when you know the exact name that a target used when creating an account. If you are unsure of the real name, or if the target has a very common name,

Followerwonk can help you identify the profile that you are seeking. This service allows you to search Twitter profiles for any keyword that may help you locate a profile of interest. You can choose the default "Twitter Bios Only" option or switch to "Search Twitter Profiles". The "More Options" under the main search box will display numerous fields including Location, Name, and Follower details. A search of "John Smith" reveals 21,156 Twitter profiles. However, a search of "John Smith" from "New York City" reveals only 125 profiles. Filtering out profiles that have not posted at least 100 Tweets reveals only 44 profiles. This is a manageable number of profiles that can be viewed to identify the target.

Twitter Directory (twitter.com/i/directory/profiles)

If you still cannot locate your target's profile, you may need to resort to the Twitter Directory. This awkward and difficult monstrosity tries to allow you to browse through the millions of Twitter profiles alphabetically. First, choose the first letter of the first name of your target. This will present a range of options. You then need to select the range in which your target would be listed, and that selection would open a new window with hundreds of name range options such as "Mike Hall – Mike Hirsch". You will need to keep using this drill-down method until you reach a list of actual profiles that meet your criteria. I do not enjoy this method, but sometimes it is all that I have. I once found my target this way after he used a misspelled version of his real name.

If searching by real name through the previous three methods does not produce your target, your best option is to focus on the potential username. I will discuss username searches at length in a later chapter (Usernames). Overall, any associated usernames from other networks such as Instagram, Snapchat, and YouTube, should be tried on Twitter in the format of twitter.com/username.

Search Operators

Similar to the way that search engines use operators as mentioned in previous chapters, Twitter has its own set of search operators that will greatly improve your ability to effectively search Twitter. Two of the most powerful options are the "to" and "from" operators. I use these daily in order to filter results. Consider the following examples using our target Twitter username of twitter.com/sultryasian. We can obviously navigate directly to her page, but it is full of promoted Tweets, ReTweets, and whatever content that she wants us to see. Instead, the following search within any Twitter window will limit our results only to her outgoing Tweets from her account. Clicking the Latest option in the Twitter menu will place these in reverse-chronological order.

from:SultryAsian

This provides a better view into her thoughts. Both her Twitter profile and this results page give us insight to the messages that she is sending out, but what about the incoming content? With most traditional Twitter investigations, we tend to focus on one side of the conversation. There is often a plethora of associated messages being sent to the attention of the target that go

unchecked. In order to see all of the posts being publicly sent to her, we would search the following.

to:SultryAsian

We now see all of those incoming messages that she cannot control. While she can prohibit them from being seen on her profile, she cannot block us from this search. When I have a missing person or homicide victim, I would much rather see the incoming messages versus the outgoing. We can also combine these options to create an extremely targeted query. At first glance, I did not see many incoming messages to SultryAsian from HydeNS33k. However, the following Twitter search tells a different story. It isolates only these posts.

to:SultryAsian from:HydeNS33k

Search by Location

If you are investigating an incident that occurred at a specific location and you have no known people involved, Twitter will allow you to search by GPS location alone. The Twitter Advanced Search allowed us to search by zip code, but that can be too broad. The following specific search on any Twitter page will display Tweets known to have been posted from within one kilometer of the GPS coordinates of 43.430242,-89.736459.

geocode:43.430242,-89.736459,1km

There are no spaces in this search. This will be a list without any map view. They will be in order chronologically with the most recent at top. The "1km" indicates a search radius of one kilometer. This can be changed to 5, 10, or 25 reliably. Any other numbers tend to provide inaccurate results. You can also change "km" to "mi" to switch to miles instead of kilometers. If you want to view this search from the address bar of the browser, the following page would load the same results.

https://twitter.com/search?q=geocode:43.430242,-89.736459,1km

You can add search parameters to either of these searches if the results are overwhelming. The following search would only display Tweets posted at the listed GPS coordinates that also mention the term "fight". Notice that the only space in the above search is between "km" and "fight".

geocode:43.430242,-89.736459,1km "fight"

It would be inappropriate to finish this section without a discussion about the lack of geo-enabled Tweets. Several years prior, this search would have been highly productive, as an alarming number of Twitter users were unknowingly sharing their locations with every post. Today, it is the opposite. The default option for Twitter is NOT to share location. A user must enable this option

in order to appear within these search results. In my experience, catching a criminal from a location-enabled Tweet is extremely rare. However, we should be aware of the possibility.

Mandatory and Optional Search Terms

You may have a scenario that requires a specific search of both mandatory and optional terms. Twitter does not provide a published solution for this. However, it does support this type of search. Assume that you are investigating threats against your target named Michael Parker. You believe that people may be tweeting about him with reference to violence. Searching his name alone produces too many results. Since you only want posts that include violent terms, the following search on any Twitter page may be appropriate.

"Michael Parker" kill OR stab OR fight OR beat OR punch OR death OR die

The name within quotes forces Twitter to only give you results on those exact terms. That is your mandatory portion. The words kill, stab, fight, beat, punch, death, and die are all optional because the term "OR" is between each. This term must be uppercase, and will only require one of the optional words be present within the search results.

Date Range Search

If you are searching vague terms, you may want to filter by date. This option is now available on the advanced search page, but I believe it is important to understand how Twitter performs this task. Assume that you are investigating a bomb threat that occurred several weeks or months ago. A search on Twitter of the terms "bomb threat" will likely apply only to recent posts. Instead, consider a date specific search. The following query on any Twitter page would provide any posts that mention "bomb threat" between January 1, 2015 and January 5, 2015.

since:2015-01-01 until:2015-01-05 "bomb threat"

My favorite use of this search technique is to combine it with the "to" operator or a name search (or both). This allows you to go further back in time than standard profile and Twitter feed searches typically allow. Consider an example where Twitter user humanhacker is your target. You can visit his live Twitter page and navigate back through several thousand Tweets. However, you will reach an end before obtaining all Tweets. This could be due to Twitter restrictions or browser and computer limitations. He currently has 13,000 Tweets. Even if you could make it through all of his Tweets, you are not seeing posts where he is mentioned or messages sent publicly to him. I recommend splitting this search by year and including mentions and messages directed toward him. The following search within Twitter displays all Tweets from the Twitter name humanhacker between January 1, 2012 and December 31, 2012.

from:humanhacker since:2012-01-01 until:2012-12-31

This may create a more digestible collection of Tweets that can be collected and archived appropriately. There may be no other way of identifying these messages since you cannot likely scroll back that far. In my investigations involving targets with several thousand posts, I conduct multiple searches within Twitter that span several years. The following would collect yearly sets of Tweets posted by humanhacker since 2006.

from:humanhacker since:2006-01-01 until:2006-12-31
from:humanhacker since:2007-01-01 until:2007-12-31
from:humanhacker since:2008-01-01 until:2008-12-31
from:humanhacker since:2009-01-01 until:2009-12-31
from:humanhacker since:2010-01-01 until:2010-12-31
from:humanhacker since:2011-01-01 until:2011-12-31
from:humanhacker since:2012-01-01 until:2012-12-31
from:humanhacker since:2013-01-01 until:2013-12-31
from:humanhacker since:2014-01-01 until:2014-12-31
from:humanhacker since:2015-01-01 until:2015-12-31
from:humanhacker since:2016-01-01 until:2016-12-31
from:humanhacker since:2017-01-01 until:2017-12-31
from:humanhacker since:2018-01-01 until:2018-12-31
from:humanhacker since:2019-01-01 until:2019-12-31

This same technique can be modified to display only incoming Tweets to humanhacker for these years. Replace "from" with "to" to obtain these results. The 2008 posts would appear as follows.

to:humanhacker since:2008-01-01 until:2008-12-31

You can combine all of these options into a single result, but I only recommend this after you have attempted the more precise options mentioned previously. While the next search should theoretically display all of his outgoing Tweets, incoming Tweets, and mentions, it is not always complete. The following would include this data for 2008.

"humanhacker" since:2008-01-01 until:2008-12-31

There are many uses for a date range search. Any supported Twitter search should work combined with dates. This might include a location search for a specific date related to an investigation. As a test of the possibilities, consider that you want to identify an email address for this target. His live Twitter page will not reveal this, as he no longer posts his email, likely to prevent spam. However, the following search is quite productive.

from:humanhacker email since:2006-01-01 until:2009-12-31

This isolates only his posts from the beginning of Twitter until the end of 2009. Only four results are present, including the Tweet as seen in Figure 10.01. We will use this data during our automation process as discussed later (sorry Chris).

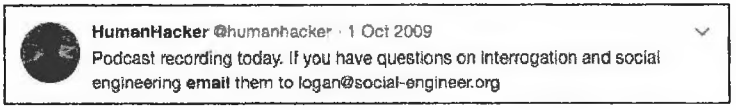


Figure 10.01: An old Twitter post including an email address of the target.

Media and Likes

You may want to filter all results from a target Twitter profile and only see those which have some type of media embedded. There is not a search operator to force this, but the following direct URL will display only these posts.

https://twitter.com/humanhacker/media/

A trend that has seen rapid adoption over the past few years is the "Liking" of posts. When a user wants to acknowledge something said by another user, but does not necessarily want to respond or ReTweet, clicking the small heart icon indicates that the post was "liked". The following direct URL displays all of the posts that our target liked. For unknown reasons, this query requires you to be logged in to an account.

https://twitter.com/humanhacker/likes/

Deleted, Suspended, and Missing Tweets

Twitter users may delete their own accounts if there is suspicion that an investigation is under way. If this happens, searching on Twitter will not display any of the posts. Furthermore, a person might only delete individual Twitter posts that are incriminating, but leave non-interesting posts on the profile to prevent raising suspicion associated with deleting an entire account. Some users may find their accounts suspended for violating Twitter's terms of service. In any of these scenarios, it is still possible to retrieve some missing posts using various techniques.

If I encounter a Twitter user that has recently deleted some or all of their messages, I conduct a cache search of their profile. There are various ways to do this, and I will demonstrate the most common. In this example, I conducted a search on Twitter for "deleted all my Tweets" on December 15, 2017. This provided many users that recently posted that they had just deleted all of their content. This helped me identify a good target for this type of demonstration. The first user I located was "WestCornfield". He had one Tweet and it referenced deleting all of his posts, and it is seen in Figure 10.02.

I attempted a search on Twitter of from:WestCornfield, which provided no results. I conducted a search of to:WestCornfield, which provided dozens of incoming messages from his friends. This was a good start. I then went to Google and conducted a search for "Twitter WestCornfield". The first search result was a link to the user's live Twitter page. Instead of clicking on the link, I chose the Google Cache view of his profile by clicking the small green "down arrow" next to the URL and selecting "Cached". This view identified twenty deleted Tweets from this account. Two of these posts can be seen in Figure 10.03. Google identified this capture as taken on December 12, 2017.

This may be enough for your investigation. Occasionally, I need to identify content that was deleted weeks or months before my investigation. The previous technique will likely not provide much assistance because the Google Cache is probably a recent copy of their live page. The cache may be missing Tweets you want to see. I next replicated this process on Bing and Yandex. Bing's cached view was taken on December 7, 2017 while Yandex's cached view was collected on November 3, 2017. Each of these possessed unique posts and images. Figure 10.04 displays a recovered post from Bing. Next, I returned to Google to obtain further data. I searched the following, which provided only results that possess a URL that begins with twitter.com, then my target's username, then "status". This will force Google to present direct links to actual posts.

site:twitter.com/westcornfield/status

The result was 56 posts. When I clicked on each of these, Twitter informed me that the post had been deleted. However, opening the cached version from the Google result displayed each of the posts. In Figure 10.05, you can see that Google is now identifying deleted posts as far back as October 2017. This process should also be repeated using the cached view options of Bing and Yandex. Next, we should check the Wayback Machine as mentioned in Chapter Eight. If you recall, you can search their archives by keywords or direct URL. The following address connects us directly to their archive of his account.

http://web.archive.org/web/*/twitter.com/WestCornfield

This identified a capture of his profile on December 6, 2017. Opening this archive displayed his Twitter profile dating back to November 8, 2017. Figure 10.06 displays this deleted Tweet.

While our target removed his content from his profile, he did not remove his history. In order to see the Twitter posts that he had previously liked before wiping out his page, we can navigate to the following URL. In this example, we see the hundreds of messages that identify his interests.

https://twitter.com/WestCornfield/likes/

While every investigation is unique, I wanted to demonstrate the importance of checking every source. These searches took less than three minutes using my custom Twitter search tool

discussed in the next section. While you will likely never rebuild an entire deleted account, the posts obtained with this technique are something you did not have before.



Figure 10.02: A live Twitter post announcing Tweet deletion.

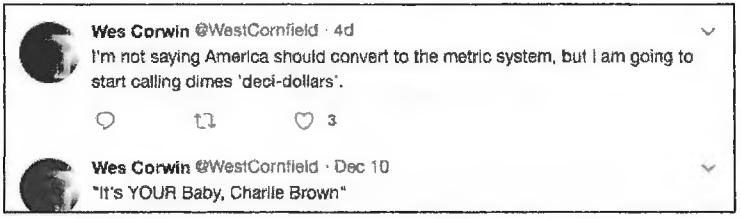


Figure 10.03: Google cached Twitter posts recovered after deletion.

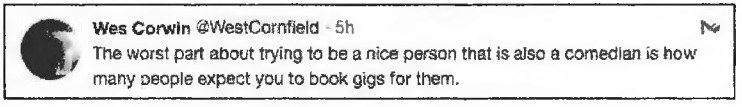


Figure 10.04: A Bing cached Twitter post recovered after deletion.

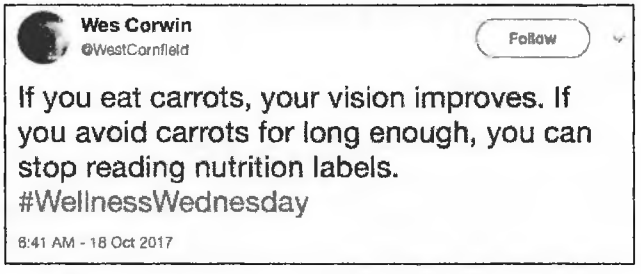


Figure 10.05: A Google cached Twitter message URL of a deleted post.

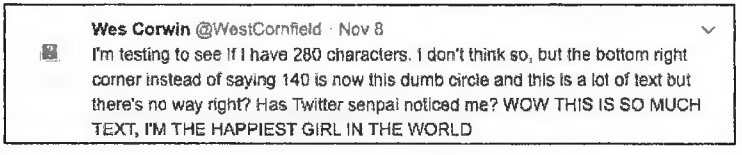


Figure 10.06: A recovered deleted Tweet from the Wayback Machine.

Twitter Post Details

Assume that you have identified an individual Tweet of interest. The URL of the message will appear similar to https://twitter.com/IntelTechniques/status/1167446287228100609. You have the tools to create a screen capture, but you may want a few more details. If there is an image embedded into the Tweet, you can click on it to see a larger version. However, this is sometimes not the original image size. In order to see the original full-size version, right-click the image and choose "View image". This will load a new URL such as the following.

https://pbs.twimg.com/media/EDOaNaJVUAIOqM3?format=png&name=900x900

You can save this image by right-clicking and choosing "Save image as". The only vital part of this URL is the first portion, which will visually appear the same, and can be used as the static URL for that image as follows.

https://pbs.twimg.com/media/EDOaNaJVUAIOqM3?format=png

If we return to the message, we can see the date and time of the post directly below the content. In this scenario, it was posted on August 30, 2019 at 8:37 AM. This is likely local time to the account owner, but we should never assume this. If we right-click on the Tweet and choose "View Source", we can search within the source code of that Tweet. While in your browser, strike Control-F (Windows/Linux) or Command-F (Mac) and search for "time_zone". This should result in one entry. In this case, it reads "Eastern Time (US & Canada),utc_offset:-14400,tzinfo_name:America/New_York". We now know this Tweet was sent at 8:37 AM EST. While this may seem minor, it could be a huge deal if your investigation makes its way to a courtroom. Knowing exact dates and times is mandatory.

Twitter Bio Changes (spoonbill.io)

Similar to the way that users delete Tweets and comments, they also modify the information within their Twitter Bio on their profile page. Several sites have come and gone which attempt to record these modifications, and my current favorite is Spoon Bill. Use of this free service from the home page requires you to log in to your Twitter account. However, a direct URL query will display any stored results. If you were researching my Twitter handle, the following address would bypass the account login.

https://spoonbill.io/data/inteltechniques/

This page displays several changes I made to my account on 9/8/16, including changing my name and location. Searching more active people will usually reveal many pages of changes, and will always display the previous content as stricken, and highlight any changes in green. I do not recommend creating an account to use this service, it will demand access to the people you follow, which could jeopardize your investigation.

Real World Application: In 2017, I assisted a law enforcement agency with a missing juvenile case. Authorities had suspicion initially that she may have run away, but became more concerned when all of her social networks became dormant and her cell phone stopped being used. Her Twitter profile possessed no valuable Tweets, and I could not find any deleted content to recover. I supplied her name to Spoon Bill and immediately received a log of changes to her Twitter profile. Two months prior to her disappearance, she listed her Snapchat username on her bio. This led me to her Snapchat profile photo. This photo and username had not been present in any of my previous investigations, and revealed a whole new world of leads. The username was associated with an email address, which was used by the missing person to create another Facebook profile. The username from Facebook revealed her user ID number, which was used in the Facebook search tools mentioned previously. The "hidden" images connected to this account provided many interesting suspects. Within an hour, a single suspect had emerged from the new discoveries, and the juvenile was located safely nearby. I cannot overstate the necessity to retrieve modified and deleted content during every Twitter investigation.

First Tweet – Keyword (ctrlq.org/first)

If you are ever following a trending hashtag or unique keyword, you may want to know the user that first posted the topic. This is not available natively within Twitter, but you can use this website to perform the search. A result for the first time a person used the hashtag of OSINT (#OSINT) reveals the message and the date of October 27, 2008. I once used this site to determine the person responsible for posting an inappropriate photo of a minor that included very specific text within the post. It had been ReTweeted many times, and this tool quickly located the first post.

First Follower (socialrank.com/firstfollower)

It may be beneficial to know the first follower of your target on Twitter. This will likely be the person that introduced the target to Twitter and can often identify a former associate. If you want only this one piece of information, First Follower will usually give you the result. However, I have found that you must often click the search option several times before the result is displayed.

TweetBeaver (tweetbeaver.com)

This is currently my absolute favorite third-party Twitter reporting tool. It is the most robust online option available to us for exporting content from an account or researching associations. There are currently fourteen unique options within this site, and I will explain each of them with usage scenarios. Note that you must be logged in to a Twitter account for any of these to work, as they all leverage the Twitter API in order to function. Please only use covert accounts, and never your own personal Twitter login information.

Convert Name to ID: This is the most reliable way to take a Twitter username, such as jms_dot_py, and convert it to the corresponding user number, which is 817668451. This can be

vital for investigations. Users can always change their username at any time, but the user number cannot be modified.

Convert ID to Name: This is the opposite of the above technique. If you had previously identified jms_dot_py as your target Twitter account only to find later that he had changed his username, you could easily locate his profile by providing the assigned user number (817668451).

Check if two accounts follow each other: As the title implies, this option quickly sorts out whether two users follow each other. An actual output appears below.

@inteltechniques does not follow @jms_dot_py
@jms_dot_py follows @inteltechniques

Download a user's favorites: This is the first tool where we can choose to either display the results on the screen or download them as a CSV spreadsheet. This option simply extracts a user's favorites (or likes) as discussed earlier. The results include the original author, date and time, text of the message, a direct URL to the post, and the author's bio, as seen in Figure 10.07.

Tweet author	Date posted	Text text	URL	Tweet author's biography
@TerrorFanatics	Wed Dec 06 02:20:06 +0000 2017	https://t.co /S4MDS86jLf	www.twitter.com/TerrorFanatics/statuses /938231545235636225	Tweeting (and retweeting) the best #Horror articles, videos, memes, writers, podcasts, filmmakers, etc.

Figure 10.07: A TweetBeaver result for a user's favorites.

Search within a user's favorites: If the previous technique produces too many results, this option allows you to filter by keyword. Since you could search within the file you downloaded or on the screen of results, I find this feature to be of little use.

Download a user's timeline: This may be the most useful of all of these options. Provide a target Twitter name and TweetBeaver will extract the most recent 3,200 posts from the account. Furthermore, it will include the date and time of each post and the direct URL to the message. When I have a Twitter target of interest, I run this tool on the account daily. It has helped me obtain the new posts every day, and identify previous posts deleted after my initial acquisition. Figure 10.08 displays the first line of a result.

Tweet author	Date posted	Text text	URL
@jms_dot_py	Fri Dec 15 20:56:36 +0000 2017	@_prasket @TeriRadichel thanks so much you guys!	www.twitter.com/jms_dot_py /statuses /941774513942945792

Figure 10.08 A TweetBeaver user timeline result.

Search within a user's timeline: Similar to the favorites search tool, I find this one redundant.

Get a user's account data: This utility provides a great summary of the account information publicly available from any Twitter account. The benefit of this method of obtaining the data is that it is quick and presented in a standard view. I can collect this information about many users, and all results will have the same format. This can aid in presentation during prosecution. Figure 10.09 displays the actual result from this target.

Screen name	Twitter ID	Name	Biography	Account created date
jms_dot_py	817668451	Justin Seitz	Creator of @Hunchly. Blogging & training #OSINT techniques. Wrote a couple of @nostarch books. @Bellingcat contributor. @C4ads fellow.	Tue Sep 11 15:44:20 +0000 2012
Location	**URL**	**Time zone**	**Geo enabled**	**Language**
Saskatoon, Saskatchewan	http://automatingosint.com/blog	Central Time (US & Canada)	not set	en
Verified	**Tweets**	**Followers**	**Friends**	
not verified	9041	7095	2112	

Figure 10.09: A TweetBeaver user account export.

Bulk lookup user account data: Similar to the previous, but allowing bulk submissions, as demonstrated in a moment.

Download a user's friends list: This option collects a target's list of accounts that he or she follows. This is similar to a typical friends list on Facebook, but approvals on either end are not required.

Download a user's followers list: This is a list of the people who follow the target on Twitter. This is less likely to contain actual friends, but all associated accounts should be investigated.

Find common followers of two accounts: This is a newer feature which can quickly identify people of interest based on co-targets. Consider the following. Twitter user jms_dot_py has been identified as a suspect in your embezzlement investigation. He seems to be very friendly with Twitter user humanhacker, and he once discussed tax evasion within public posts with this new second person of interest. Humanhacker has 33,000 followers and jms_dot_py has 12,500. There is no great way to look through these in order to find other suspects. However, TweetBeaver quickly identified the 371 people who follow both targets. This can be much more manageable, but is still a lot of data.

Find common friends of two accounts: Similar to the above, this only looks at the "friends" of each target. In other words, the people each target follows on Twitter. At the time of this writing, TweetBeaver identified the 57 people of interest. The following were two of the accounts.

The Innocent Lives	@InnocentOrg	ID 854674794482216960
Michael Bazzell	@IntelTechniques	ID 257644794

Find conversations between two users: Once you have identified two people of interest, you can focus on the conversations between them. You could replicate this with the previous Twitter operators, but this route presents a nice spreadsheet of the results. In my scenario, I located six conversations between me and jms_dot_py. Figure 10.10 displays one result.

@jms_dot_py	Fri Nov 23 16:24:48 +0000 2018	@IntelTechniques @replyall @AGoldmund @frankahearn Not gonna lie, I am fanboying a bit over @AGoldmund	https://twitter.com /jms_dot_py/statuses /1066004849721954304	Twitter Web Client

Figure 10.10: A message filtered by TweetBeaver.

Bulk Account Data Download

In early 2018, TweetBeaver introduced a bulk lookup feature which can accept up to 15,000 Twitter usernames. This is quite impressive, and I have put it through many tests. The bulk feature on TweetBeaver allows you to input numerous accounts within one query. As a demonstration, I added every Twitter account mentioned in this chapter up to this point into the TweetBeaver bulk lookup utility. The entry appeared as follows.

inteltechniques	SultryAsian	humanhacker
jms_dot_py	HydeNS33k	WestCornfield

This immediately created a CSV spreadsheet which was downloaded to my computer. The following are screen captures of the data. This gives me an immediate view into the accounts. If I had hundreds or thousands of Twitter usernames, this would allow me to sort by location or popularity. I could also sort by creation date in order to identify newly created accounts. This is possibly the most useful third-party tool when you have numerous accounts of interest.

Screen Name	Twitter ID	Name	Description	Account created date
@screen_name	ID 257644794	Michael Bazzell	Open Source Intelligence (OSINT) Training and Tools. International Privacy Consultant.	Fri Feb 25 21:46:04 +0000 2011
@screen_name	ID 817668451	Justin Seitz	Creator of @Hunchly. Blogging & training #OSINT techniques. Wrote a couple of @nos	Tue Sep 11 15:44:20 +0000 2012
@screen_name	ID 3308415728	SultryAsian	Known to most as M. Not as serious as I look.	Fri Aug 07 03:49:08 +0000 2015
@screen_name	ID 7711241370	Jek Hyde	Infosec Auntie ðŸ‡¨ðŸ‡¦ Red Team Analyst at Fortune 1 ðŸ•µðŸ¿â€â™€ï¸ Thief ðŸ—¡ Security Quality	Wed Aug 31 23:14:55 +0000 2016
@screen_name	ID 46998400	HumanHacker	This is the official Twitter account of all things SEORG - The SEVillage, SEPodcast, and tl	Sun Jun 14 00:47:39 +0000 2009
@screen_name	ID 368971022	Wes Corwin	The Dean Malenko of Stand-Up Comedy and The Kazuyuki Fujita of Roast Comedy. See	Tue Sep 06 14:58:56 +0000 2011

Location	URL	Time Zone	Geo-enabled	Language	Verified	Tweets	Followers	Following
Washington D.C.	http://inteltechniques.com	Eastern Time (US & Canada)	not set	en	verified	266	5456	0
Saskatoon, Saskatchewan	http://automatingosint.com/blog	Central Time (US & Canada)	not set	en	not verified	9163	7316	2123
not set	not set	Pacific Time (US & Canada)	not set	en	not verified	2248	1894	622
Dallas, TX	not set	not set	not set	en	not verified	3820	12817	404
USA	http://www.social-engineer.org	not set	not set	en	not verified	11498	25274	273
Dallas, TX	http://wescorwincomedy.wordpress.com	Pacific Time (US & Canada)	enabled	en	not verified	42	2725	2193

Real World Application: The day this feature was released, I was investigating a suspicious Twitter account associated with violent threats toward a celebrity. The suspect had sanitized the account which prohibited obtaining valid location data. However, the suspect had numerous followers and people he was following. Using TweetBeaver, I could extract these accounts easily and supply them to the bulk lookup utility. I then sorted his friends by location which revealed a strong presence in a specific Midwest city. Of those target accounts of interest, I could see that a few were Geo-enabled. This provided home addresses for two people. The person search tools described later identified full names and telephone numbers of these two friends. Group photos found online of these targets identified one person always nearby, but without a Twitter account associated with this group. The same photos on Facebook identified my new target by name. This process led to a positive identification of my suspect. Bulk search tools help tremendously. If you only use one third-party Twitter analysis tool, I recommend TweetBeaver.

Location Information by User

While privacy-aware individuals have disabled the location feature of their accounts, many users enjoy broadcasting their location at all times. Identifying a user's location during a Twitter post is sometimes possible through various methods. Prior to 2014, identifying the GPS details of every post of many users was simple. Today, most of these identification techniques are no longer working. A manual Twitter search method for identifying posts by location was explained earlier. That technique is best for current or live information, and is limited to only recent posts. You may have a need for historical details from previous posts from a specific location. I have had better success with historical data than current content in regard to geolocation. I believe this is because most people unknowingly shared their location while Tweeting for many years. When Twitter changed the default location option to "Disabled", most users never intentionally re-enabled the feature. Fortunately, there are third-party websites that collected this historical data, and can assist with easy searching. The following options will work best when you are investigating events that occurred several years prior. While you may get lucky and receive some recent posts, the majority could be quite old.

Tweet Mapper (keitharm.me/projects/tweet)

If your target does possess recent Tweets with location data, this map should reveal them.

Omnisci (omnisci.com/demos/tweetmap/)

Omnisci is a massive database platform developed through collaboration between MIT and Harvard. Historically, each college had their own interface into this data, which supplied Twitter post locations from past Tweets. Each interface provided new ways of searching information. Both websites have been disabled, and the entire project has warped into Omnisci. This website can search by topic, username, or location. It can also combine all three options to conduct a detailed search. Results appear as blue dots on a dark map. Each dot represents a Tweet which possesses location data to the chosen area.

One Million Tweet Map (onemilliontweetmap.com)

This service only displays the most recent one million Tweets on an international map. They do not have access to every Tweet available, often referred to as the "firehose", but they offer new Tweets every second. I would never rely on this map for complete data about a location. However, monitoring a large event can provide live intelligence in an easily viewed format. I recommend using a mouse scroll wheel to zoom into your location of interest. Once you are at a level that you can see single Tweets, you can click any of them to see the content. The page will automatically refresh as new information is posted.

Search by Email Address

Technically, Twitter does not allow the search of a Twitter user by providing an email address. If you attempt this type of search on Twitter, you will receive no results, even if a profile exists for a user with that email address. To bypass this limitation, you can use a feature offered by online email providers that will communicate with Twitter.

This technique will require a Twitter account and Yahoo email account. Through a web browser, connect to an alias Twitter account in one tab and connect to an alias email account on Yahoo Mail in another tab. If you do not have a Yahoo email account, the creation process is easy at mail.yahoo.com. On the "Contacts" page of your Yahoo account, select "Add a New Contact". This will present a form that will allow you to enter the email address of the target. The "Name" fields can contain any data, and I usually use numbers beginning at 001. Save this entry and go to your Twitter page.

Click on the "Privacy and safety" option in the settings menu, then "Discoverability and contacts". You can then manage your contacts and allow Twitter to sync to your Yahoo account. You will need to confirm that you want to give Twitter access to your Yahoo contacts by clicking "I agree" on the pop-up window. Twitter will then identify any user profiles that are associated with the target email addresses within Yahoo. Twitter will encourage you to add these friends, but do not choose that option. If you do, your target will be sent a notification from your account.

All My Tweets (allmytweets.net)

This website provides a clean display of all of a user's Twitter posts on one screen. It will start with the most recent post and list previous posts on one line each. This view will display up to 3200 messages on one scrollable screen. This provides two analysis methods for digesting large amounts of data. Holding CTRL and F on the keyboard will present a search box. Any search terms can be entered to navigate directly to associated messages. The page can also be printed for archiving or distribution to assisting analysts. This is my preferred way of reading through the Tweets of an active user. This also prevents you from constantly loading more Tweets at the end of every page throughout a profile. While I prefer TweetBeaver for this task, it is always good to have options.

Sleeping Time (sleepingtime.org)

This site allows for a search of an exact Twitter profile name, and provides the average time period that this user sleeps. The historical Tweets are analyzed according to the times of posts. Data is then presented that suggests when the user is usually sleeping due to lack of posting during a specific time period. A query of Kevin Mitnick revealed that he is likely to be sleeping between 12am and 7am according to his Tweets. Although the idea was probably executed as a fun site, it can be quite useful.

Real World Application: Police often want the element of surprise on their side when contacting suspects. Whether this is to execute a search warrant or simply contact a subject when he or she is most likely to be home, knowing the habits of the individual can be beneficial. Locating a possible sleep pattern of the individual will decrease the chances of showing up at an empty house, only to discover that the subject works a strange night shift somewhere. Sleeping Time may have alerted an investigator that the average sleep time is 2pm to 10pm, creating an opportunity to catch the subject at home. This also works for process servers, private investigators, bill collectors, and even salesmen.

Tweet Deck (tweetdeck.twitter.com)

Tweet Deck is owned by Twitter, and it can take advantage of the Twitter "Firehose". This huge stream of data contains every public post available on Twitter. Many Twitter services do not have access to this stream, and the results are limited. Tweet Deck requires you to create and log in to an account to use the service. This user account is not the same as a Twitter account. The "Create Account" button on the website will walk you through the process. Alias information is acceptable and preferred. The plus symbol (+) in the upper left area will add a new column to your view. There are several options presented, but the most common will be "Search" and "User". The "Search" option will create a column that will allow you to search for any keywords on Twitter. The following is a list of search examples and how they may benefit the investigator.

"Victim Name": A homicide investigator can monitor people mentioning a homicide victim.
"School Name": A school can monitor anyone mentioning the school for suspicious activity.
"Subject Name": An investigator can monitor a missing person's name for relevant information.
"Event": Officials can monitor anyone discussing a special event such as a festival or concert.

The "User" option will allow you to enter a Twitter username and monitor all incoming and outgoing public messages associated with the user. If several subjects of an investigation are identified as Twitter users, each of the profiles can be loaded in a separate column and monitored. Occasionally, this will result in two of the profiles communicating with each other while being monitored. You can also use the Geo search mentioned earlier within Tweet Deck. A column that searches "geocode:43.430242,-89.736459,1km" will display a live feed of Tweets posted within the specified range. A more precise search of "geocode:43.430242,-89.736459,1km fight"

would add the keyword to filter the results. Figure 10.11 displays Tweet Deck with several searches.

The columns of Tweet Deck are consistently sized. If more columns are created than can fit in the display, the "Columns" option with left and right arrows will provide navigation. This allows for numerous search columns regardless of screen resolution. This is an advantage of Tweet Deck over the other services discussed. Tweet Deck is one of my Twitter staples. I use it at some point during every investigation. I recommend familiarizing yourself with all of the features before needing to rely on it during your searches.

Real World Application: In 2019, I was investigating a death threat toward a celebrity. I launched Tweet Deck and began monitoring. First, I created a search column with "to:myclient". This started a stream of numerous people mentioning my client, which was too much to monitor. Next, I created a column of "to:myclient kill OR die OR shoot OR death". This presented very few tweets being sent to the celebrity including specific hateful words. However, it did identify a suspect. A tweet was sent to the celebrity stating "I hope you die a fiery death tonight". I then created another column of "from:suspect to:myclient". This identified every tweet he was sending to the celebrity. Since I had to move on to other resources, I set one final Tweet Deck column of "from:suspect to:myclient kill OR die OR shoot OR death" and added an alert. This instructed Tweet Deck to play an audible sound if any new messages met the criteria. The same suspect was arrested a week later while attempting to burglarize her apartment.

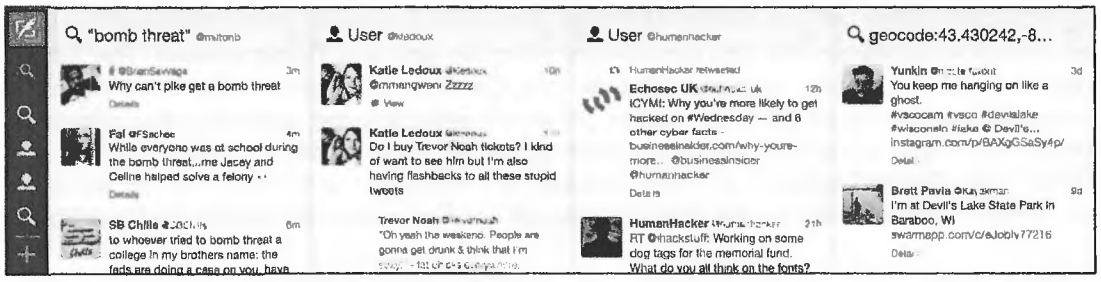


Figure 10.11: A Tweet Deck search screen.

Hootsuite Feed (hootsuite.com)

While Tweet Deck is my preferred viewer of live Twitter information, it does not display well for large audiences. If I am broadcasting my screen to a digital projector for a room full of people to see, the text is usually too small to accurately view from a distance. Hootsuite offers a solution to this predicament. If you wanted to display a live feed of anyone mentioning "OSINT" on Twitter, you can navigate to the following website in your web browser, after connecting to a Twitter account.

https://hootsuite.com/feed/OSINT+Search

You can replace "OSINT" in the previous address with any term or terms of interest. The result will be a live stream with a very large font that could be viewed from a long distance away. This view can be beneficial for situations where Twitter streams are monitored by a group of people in an operations center.

Twiangulate (twiangulate.com)

Earlier, I explained how I use TweetBeaver to filter most of my Twitter friend and follower data. If it should become unavailable, there are two additional websites which can assist. Twiangulate identifies mutual friends on two specific accounts. In one example, 521 people were friends with one of my subjects. However, only 15 were friends with both targets of my investigation. This can quickly identify key users associated within an inner circle of subjects. All 15 subjects were listed within the results including full name, photo, bio, and location. While Twiangulate has assisted me in the past, I now recommend Followerwonk as a superior solution when you have multiple suspects.

Followerwonk (followerwonk.com)

This service offers more options than Twiangulate and will let you compare up to three users. The second tab at the top of the page, titled "Compare Users", will allow you a more thorough search. Figure 10.12 displays the analysis of three subjects. You can see that the first and second subject do not have any people in common that they follow on Twitter. This can indicate that they may not know each other in real life, or that they simply have different tastes in the people whom they find interesting. However, the first and third subjects have 79 people in common that they follow on Twitter. This is a strong indication that they know each other in real life and have friends in common. Clicking on the link next to this result will display the identities of these people as seen in Figure 10.13.

This default search on Followerwonk is a good start. A more valuable search is to analyze the people who follow these users. The previous example identified people whom our targets followed. This will often include celebrities, businesses, and profiles that probably have no impact on your investigation. However, the people who follow your targets are more likely to be real people who may have involvement in your investigation. Figure 10.14 displays the results of the same targets when the search criteria was changed to "Compare their followers" in the dropdown menu next to the search button. We now see that the first and second subject still have no one in common. The first and third subject have 200 people who follow both of their Twitter feeds. You can click the result link to identify these 200 people.

Followerwonk possesses other search capabilities for user analysis. The first tab at the top of the screen will search any term or terms to identify any Twitter bios that contain those words. This may identify profiles that were missed during the search on twitter.com for messages. The third tab, titled "Analyze Followers", allows you to enter a single Twitter handle and either analyze the

people the user follows or the people who follow that user. The second option usually provides more relevant results.

The information provided during this search will display numerous pie charts and graphs about the user. The most useful is the map that identifies the approximate location of the people connected to the person's Twitter account. Figure 10.15 displays a map for one of the users searched previously. This provides a quick indication of the regions of interest to the target. Figure 10.16 displays a detailed level of an area near Denver. Each small dot identifies an individual Twitter account of a person that follows the target and lives or works in the general area. This location data is very vague and does not usually correctly correlate with the address on the map. This should only be used to identify the general area, such as the town or city, of the people who are friends with the target on Twitter. In the past, I have used this data to focus only on people in the same area as my homicide victim. I temporarily eliminated people who lived in other states and countries. This helped me focus on subjects that could be contacted quickly and interviewed.

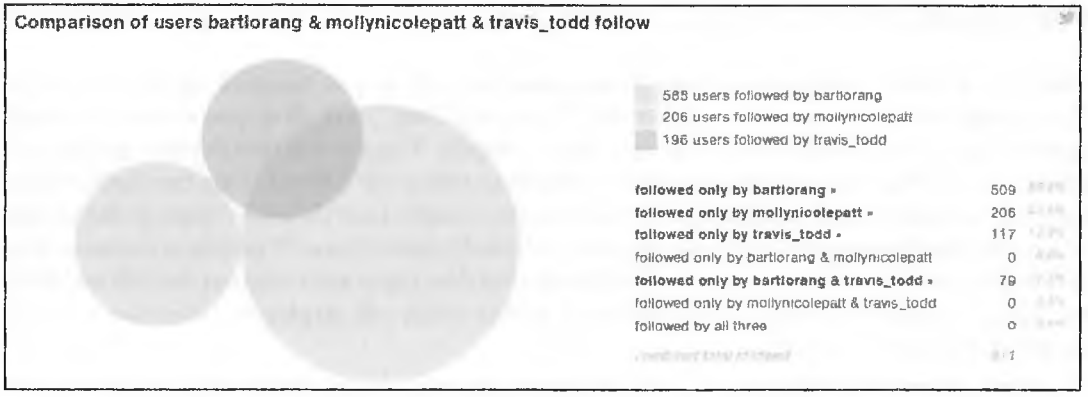


Figure 10.12: A Followerwonk user comparison.

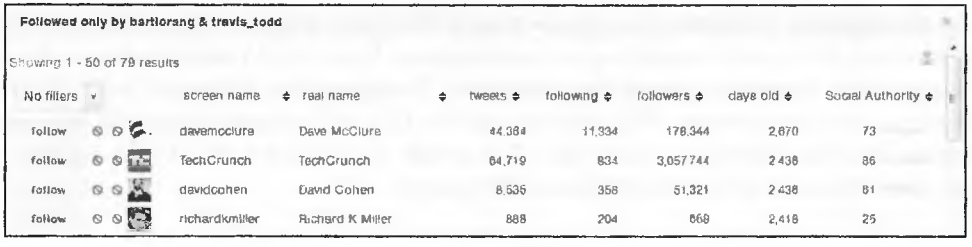


Figure 10.13: A Followerwonk list of users.

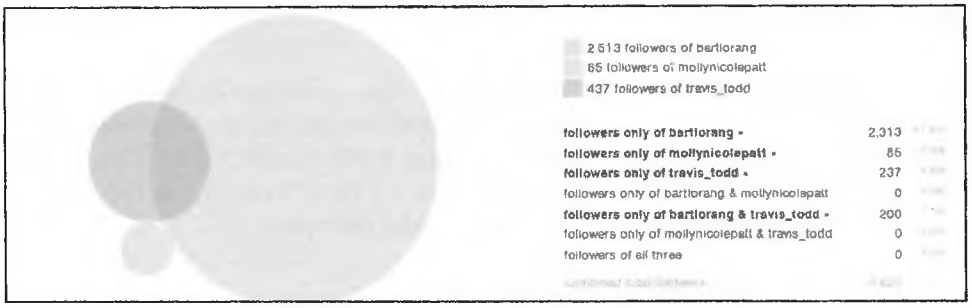


Figure 10.14: A Followerwonk user comparison.



Figure 10.15: A Followerwonk map of users connected to a target.

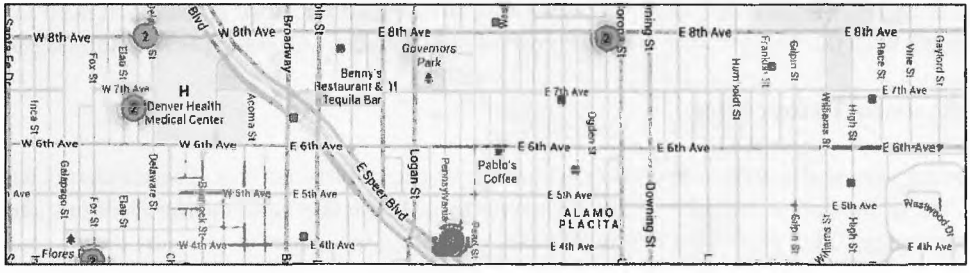


Figure 10.16: A detailed view of a Followerwonk map of connected users.

Fake Followers

There are a surprising number of Twitter accounts that are completely fake. These are bought and sold daily by shady people who want to make their profiles appear more popular than they really are. I have seen a lot of target profiles that have been padded with these fake followers. There are three websites that will assist in distinguishing the authentic profiles from the fraudulent. They all require you to be logged in to a Twitter account, and I will compare the results of each.

SparkToro (sparktoro.com)

The most robust option was SparkToro. Analyzing my own account, it declared that 25% of my followers were "accounts that are unreachable and will not see the account's tweets". It also provided some metrics and a full explanation as to how it achieves its results. This is something the others do not disclose. Once you have logged in to a Twitter account, you can query further users at the following static URL (using jms_dot_py as an example).

https://sparktoro.com/fake-followers/jms_dot_py

Twitter Audit (twitteraudit.com)

This option identified 9% of my followers as "fake" and provided very few details. It also allows for submission through a static URL, but you still need to request an audit once you get to the page. The following would display results for jms_dot_py.

https://www.twitteraudit.com/jms_dot_py

Status People (fakers.statuspeople.com)

This service provided a report stating that 28% of my followers were fake, 39% inactive, and 33% "good". There is no option to query via a static URL within the site. Overall, I do not put much faith into any of these, especially when the results appear so random. If all three services present a high score of over 75% fake followers, that would be more revealing. These services can be beneficial during elections. They can reveal political accounts which falsely appear popular.

Miscellaneous Twitter Sites

Every week, a new site arrives that takes advantage of the public data that Twitter shares with the world. These sites offer unique ways of searching for information that Twitter does not allow on their main page. This partial list is a good start to finding information relevant to your target. I encourage readers to follow my blog, podcast, and Twitter account for any updates which surface after this book has been published.

Twitonomy (twitonomy.com)

One Twitter analytics website that stands out from the rest is Twitonomy. This is the most complete analytics service that I have found for a single Twitter handle. A typical user search would fill four pages of screenshots. A search of the user "humanhacker" immediately revealed the following details.

He has posted 8,689 Tweets	He has 17,079 followers
He is following 170 people	He joined Twitter on June 14, 2009

He averages 5 Tweets per day
He has mentioned 4,175 other Twitter users
He replies to 34% of posts to him
He has ReTweeted 621 posts (15%)

The remaining sections of this page identify current posts, followers, people following, favorites, and lists. The main analytics portion identifies the average number of posts by day of the week and by hour of the day. It also displays from which platforms the user Tweets. Figure 10.17 displays this portion. This data discloses that the target has an Android and an iPhone, and that the majority of his Twitter time is spent on a Mac computer. This also identifies his preferred web browser, check-in utility, photo sharing service, and video sharing service. Other information includes his Tweets that are most "favorited" and "ReTweeted"; the users to whom he replies most often; and the users whom he mentions more than others. If you have a Twitter name of interest, I highly recommend searching it through Twitonomy.



Figure 10.17: A portion of a Twitonomy search result identifying user platforms.

Trendsmap (trendsmap.com)

Monitoring trends on Twitter can provide intelligence on a global scale. The keywords that are currently being posted more than any other terms in a specific area could be of interest. This can identify issues about to surface that may need attention. This type of analysis is common during large events such as protests and celebrations. Several websites offer this service, but I choose Trendsmap. You can search either topics or a location. Searching a location will provide the top keywords being posted as well as a heat map to identify peak usage.

Tinfoleak (tinfoleak.com)

This Twitter analytics tool provides a simple yet thorough report. The website requires that you log in to your Twitter account through this service, and the report includes the following information relevant to the target.

Twitter ID #	Location	Hashtags
Account Creation Date	Time Zone	User Mentions
# of Followers	Number of Tweets	Metadata from Images
# Following	Twitter Clients Used	Geo-Location Data

FollerMe (foller.me)

This service is very similar to the previous Twitter analytics options. Providing a Twitter username presents the typical bio, statistics, topics, hashtags, and mentions analysis that you can find other places. I find the following option of most interest to my investigations. I previously explained Sleeping Time as a resource to learn a target's usual sleep pattern based on posting times. FollerMe provides a bit more detail including posting patterns per hour. Note that the results are displayed in Universal Time, so you will need to convert as appropriate for your suspect. Figure 10.18 displays the results for my account. Since I am on the east coast (UTC -5), this example indicates that I tend to never post before 8:00 a.m. or after 11:00 p.m. My peak Tweeting time is at 11:00 a.m. There is a very obvious sleep pattern present within this result.

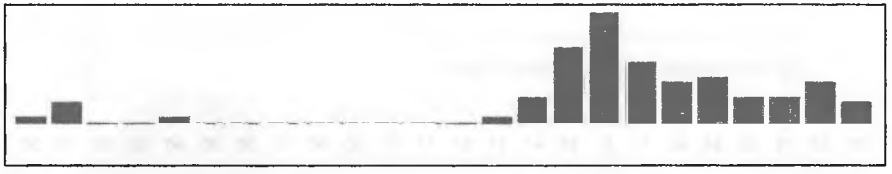


Figure 10.18: A FollerMe posting history by time.

TweetTopic (tweettopicexplorer.neoformix.com)

This simple tool provides one feature that I have found helpful in my investigations. Once you supply the target Twitter username, it collects the most recent 3,200 Tweets and creates a word cloud. This identifies the most common words used within posts by the target. There are several sites that do this, but this service takes it a vital step further. Clicking on any word within the result displays only the Tweets that include the selected term. In Figure 10.19, you can see that I tend to Tweet about OSINT more than anything else. Clicking on any of the "ios" or "Facebook" circles would immediately identify posts related to those terms. I have used this when I have a target with too many posts to read quickly. TweetTopic allows me to quickly learn what my target posts about and immediately delve into any topics of interest to my investigation.

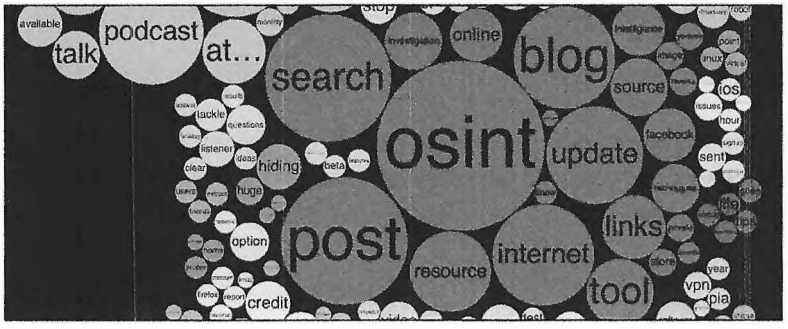


Figure 10.19: A TweetTopic interactive word cloud.

Social Bearing (socialbearing.com)

This robust solution combines the features of many resources previously listed into one search. It relies on the Twitter API, so it will only analyze the most recent 3,200 Tweets. While searching my own account, I received a summary which included the following. Investigative benefits are next to each.

- Total audience reach: This tells me whether the target has true followers or "fakes".
- Total impressions: This tells me an accurate size of the target's audience.
- Total ReTweets: This discloses if the target's audience engages with the content.
- Total audience favorites: This confirms engagement from the target's audience.
- Tweet sentiment: This indicates positive or negative tone within comments.
- Tweet types: This identifies new content versus Re-Tweets of others.
- Tweet sources: This discloses the source of submission, such as mobile, desktop, or API.
- Domains shared: This summarizes the web links posted within Tweets.
- Word cloud: This provides a summary of the most common words posted.

The graphical output of this resource is impressive, but the CSV export is more useful. Figure 10.20 displays the web view, and the CSV export option can be seen in the upper middle. Additionally, we can query an account directly via static URL. This allows us to include this resource in our automated tools. My account can be seen at the following URL.

https://socialbearing.com/search/user/inteltechniques

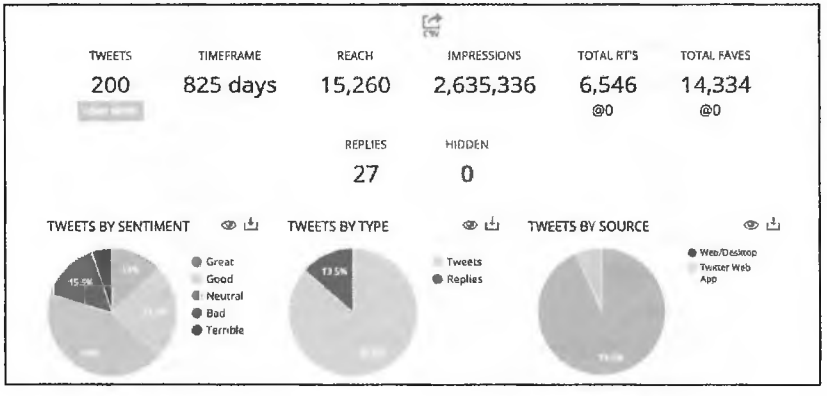


Figure 10.20: A Social Bearing report.

On a final note about Twitter, I believe it is under-utilized by most investigators. We tend to find our target's profile, scroll a bit, and quickly dismiss it. In my experience, the bulk of relative information, and the most valuable to my case, is never present in this live view.

IntelTechniques Twitter Search Tool

I found myself using many of these manual Twitter techniques daily. In order to prevent repetitive typing of the same addresses and searches, I created a custom tool with an all-in-one solution. Download the archive mentioned previously to access this resource. This page includes embedded JavaScript that will structure and execute web addresses based on your provided information. Figure 10.21 displays the current state of this tool. Any utilities that do not look familiar will be described in the remaining pages of this book. This tool will replicate many of the Twitter searches that you have read about here. The first option on the left side populates all of the search fields with your target's Twitter name when entered there. This should save time and avoid redundant pasting of data. Clicking "Go" next to each option executes the query in a new tab in your browser. As a reminder, this page collects absolutely no information from your searches. They are all conducted within your own browser, and no data is captured by my server.



Figure 10.21: The IntelTechniques Custom Twitter Search Tool.

Chapter Eleven
Social Networks: Instagram

This edition presents the first time I have included a chapter dedicated to Instagram. While Facebook and Twitter are likely going to offer the most bang for your buck in terms of web-based social network presence, Instagram has captured a large share of the market over the past two years. Instagram is a photo-sharing service that is now owned by Facebook. With well over 1 billion active monthly users, the amount of content available here is overwhelming. This application works alone or in correlation with Facebook and Twitter to distribute the photos. This service is very popular with most photo sharing Twitter users and should not be ignored. Searching for Instagram content through Twitter's website will not provide much of the content. Surprisingly, there is no search feature on the Instagram home page. If you want to use Instagram's search database, you must connect directly to an account such as Instagram.com/mikeb. However, this search field only identifies users and hashtags related to the search terms. It does not provide a true keyword search. We will use search engines for this in a moment.

In previous editions of this book, I detailed several third-party Instagram search options that unlocked a lot of hidden content within user accounts. On June 1, 2016, Instagram tightened their API, and this killed most of the useful websites. My own Instagram tools page suffered drastically from the new restrictions, and I had to start over with new options. Fortunately, you still have many search options for your next Instagram investigation. Let's start with keyword searching. I have found greater success with a custom Google search instead of an Instagram search field. The following query on Google will produce 261 results that display Instagram posts that mention "OSINT" within the post title or comments.

site:instagram.com "OSINT"

This same term searched on an Instagram page only displayed users that have the term within the username. When searching "#OSINT" on Instagram, I was provided a list of hashtags that include the keyword. Each of these hashtags are associated with multiple Instagram posts. Consider the following example in order to identify the benefits of searching away from the Instagram website. While on an Instagram page, I searched for my target "Darren Kitchen" from Hak5. This presented several profiles associated with people having that name, but none of them were my target. Instead, I went to Google and conducted the following search.

site:instagram.com darren kitchen hak5

This produced three results, but none of them connected to my target. Since people seldomly use last names on Instagram, I modified my search to the following.

site:instagram.com darren hak5

The first result was my target's profile (@hak5darren). Similar to Facebook and Twitter, a person's Instagram profile only tells a small portion of the story. Conducting the following search on Google revealed 165 results.

site:instagram.com "hak5darren"

These are the various pages and posts that contain the text of my target's username. Many of these are posts from the target, which we have already seen by looking at his profile. A few are from associates of the target and may otherwise be missed. I now only want posts that contain "@hak5darren" within the content. This is more likely to find other people mentioning my target.

site:instagram.com "@hak5darren"

Similar to the previous search, it forces Google to ignore the target's profile, and only focus on people that are posting "to" the target with @hak5darren instead of just his username alone. These searches can be modified to include or exclude keywords, usernames, real names, or any other terms with which you have interest. You could also repeat these on Bing and Yandex for better coverage. Hopefully, you now have a target profile of interest. Now we can dig into our target within Instagram and through third-party resources.

Instagram Images

There is one obstacle with Instagram searches that needs to be addressed. When you identify a profile of interest, there will always be a profile image at the top of the page. This is usually a 150x150 pixel profile image and a full page of small-resolution post images, all of which are small and heavily compressed. In 2015 and 2018, Instagram changed the way these thumbnails and original images are stored. The goal of this technique is to locate these low-resolution images and identify the full high-resolution images that were originally uploaded. For an example, I will pick on the user "hak5darren". His profile is at instagram.com/hak5darren. When I right-click on his profile picture, I am blocked from opening the image alone in a new tab. In fact, there is no right-click menu whatsoever. If I right-click on the images, I also receive no option to view the image alone, which previously would grant us a larger version. Instagram is actively blocking us, so we will need to dig into the source code of the page.

When I right-click anywhere else on the profile, I have the option to "View source". This opens a new tab which displays the HTML source code (text) behind the profile page. I can also navigate directly to view-source:https://www.instagram.com/hak5darren/ to see this result. Within this text, I can search for "og:image" and receive only one result. The extremely long URL within this line should appear similar to the following.

https://scontent-dfw5-1.cdninstagram.com/vp/c73900a418f897522ae8062791e18723/5DF569DF/t51.2885-19/s150x150/13534173_1620017484977625_1916767107_a.jpg?_nc_ht=scontent-dfw5-1.cdninstagram.com

Notice the section of "s150x150". This tells Instagram to only display a 150 pixel square icon of the target. This is a highly compressed thumbnail image. In previous years, we could remove that section and display the full-sized image, but that no longer functions. Instead, we can only enlarge the profile image slightly. In the previous URL, notice the number directly before ".jpg". The last set of numbers is "1916767107". Back in the page of source code for the profile, conduct a search for these numbers. You should receive three results, the third is a URL similar to the following.

https://scontent-dfw5-1.cdninstagram.com/vp/fe834028ba5d34e8f935f88a14eb0127/5E0356C7/t51.2885-19/s320x320/13534173_1620017484977625_1916767107_a.jpg?_nc_ht=scontent-dfw5-1.cdninstagram.com

This URL presents a profile image which is now twice as large as the original image provided within the profile. This is not the full-size image, but it possesses a higher resolution. Fortunately, we can do much better in regard to the images from his posts. When clicking on the first image on his profile, we are given a new URL of https://www.instagram.com/p/B0Hze9xJ8bM/ which displays the image and comments. The image is highly compressed. Clicking on the image does nothing, and right-clicking gives us no options associated with photos. Instead, right-click on this post image and select "View source" again. This opens a new tab similar to the following.

view-source:https://www.instagram.com/p/B0Hze9xJ8bM/

Conduct a search within this text for "1080". The URL immediately following the first hit should appear similar to the following. You should eliminate any text following the second instance of "cdninstagram.com", such as "\u0026_nc_cat=109".

https://scontent-dfw5-1.cdninstagram.com/vp/300f5c8b017ffa319029985d0ad047de/5DF349E8/t51.2885-15/e35/66279715_634911447015657_3017525331735088197_n.jpg?_nc_ht=scontent-dfw5-1.cdninstagram.com

This image has much higher resolution, and is our best option for viewing and archiving. Once you load this URL in the browser, you can right-click on the image and easily save it. Does this seem like too much trouble for every image of interest on Instagram? No argument here. If you prefer to allow a third-party service to replicate this for you, the following sites will download the best quality image available within a post. Regardless of your method, you should understand the functions allowing for the capture of these high-resolution images.

downloadgram.com
dinsta.com/photos
instadp.org/instagram-downloader
gramsave.com

Metadata Details

You should also consider digging into the source code of your evidence in order to identify further details that could be valuable. First, I like to identify the user number of an account. Similar to Facebook and Twitter, people can change their username on Instagram, but not their user number. Right-click on your target profile page and select the option to view the source code. This will open a new tab with a lot of pure text. With either ctrl-f or command-f, conduct a search for exactly the following. The numbers directly after this data will be the user number of your target. In our previous example, his user number is 340416780.

"owner":{"id"

While we are looking at the source code of pages, we can use this technique to identify the exact time that a post was published. Instagram only identifies the date of a post, and not the time. This detail can be crucial to a legal investigation. For this example, I located an image posted by our target at the following address.

https://www.instagram.com/p/Btbtlu1HIQ3/

Viewing the source code of this page, search for "uploadDate". The following is the entire line, and we now know the exact time, in universal time (GMT) format.

"2019-02-03T20:26:51"

I realize these details are very minor. However, documenting this content can announce the difference between an average OSINT investigator and an expert. If you plan to serve any legal process later, knowing the user number and exact time of activity will be crucial.

Followers & Following

Instagram now requires users to be logged in to an account in order to view the followers of a target or the profiles that a target follows (friends). Viewing these lists is not a challenge, but proper documentation can be tricky. The following explains how I choose to view and document every Instagram follower and friend of my target account, using hak5darren as an example.

After logging in to your account and navigating to the target profile, you will be able to simply click on "Followers" or "Following". In this demonstration, I chose the people he is "following", often referred to as friends. This opened a new window on my screen with the first twenty people displayed. Since hak5darren follows 199 people, I can scroll down the entire list to load all of the accounts. You will not be able to see all of the accounts at once because the window only displays ten within its boundaries. This causes screen capture tools to be useless in this scenario. If you are using either the Firefox browser we created within Linux or your own version of Firefox with the recommended add-ons, you can easily copy this list. After scrolling down to load all of the

users, press either ctrl-a (Windows & Linux) or command-a (Mac) on your keyboard. This will highlight or "select" the entire list. Next, right-click within this window and choose "Copy selected links". This copies all of the Instagram account hyperlinks within the target's following list into your computer clipboard. Open your desired documentation software, such as Word, Excel, or any text editor, and paste the contents into the application. I prefer Excel (or Calc within the free LibreOffice suite) for this task. Pasting the results confirmed 199 accounts, four of which are shown in Figure 11.01. Repeat the process with "Followers".

196	https://www.instagram.com/dragorn/
197	https://www.instagram.com/smacktwin/
198	https://www.instagram.com/serafinamoto/
199	https://www.instagram.com/moon_spot/

Figure 11.01: A partial spreadsheet of people being followed by the target.

In my experience, Instagram limits the number of accounts within this window to 1,000. This should be sufficient for most friends of the target, but may be limited when looking at the followers of famous people.

Likes & Comments

Similar to Twitter, users can "Like" or comment on each post. Unique from Twitter, this is only seen when you load the target post in your browser. In other words, if a user posts a comment to a photo published by our target, we do not see that comment within the account of the person who made it. We only see it within the specific post directly from the target. We also cannot search within Instagram by keyword in order to identify comments, but we can with Google as explained momentarily. First, let's focus on Likes.

Viewing an individual Instagram post reveals a heart icon below any comments. Clicking that icon "likes" the post from your account, so this should be avoided. Directly below the heart is a summary such as "557 Likes". Clicking on this opens a new window, which identifies each account that liked the post. Slowly scrolling through this list expands it until all accounts are present. Similar to the previous method of capturing friends and followers of a target, a manual approach is best here. Pressing ctrl-a or command-a on the keyboard selects all of the content, and ctrl-c or command-c copies it to your computer memory, which can be pasted into a spreadsheet or word processor. You can repeat this process for the comments within the post.

If you locate a comment which is valuable to your investigation, you may want to document the date and time of the post. This is not available within the standard view. As an example, look at the post at https://www.instagram.com/p/Bqi6r9uA0Y3. Inside the comments includes a response of "German Beer" from a friend. At the time of this writing, Instagram displayed "41W" under the post, indicating it was posted 41 weeks previous to the current date. This is not sufficient for our needs. While on the page, right-click and choose the option to view the source

code. Next, search for the term (German Beer) and notice the text immediately after the result, as displayed below.

"created_at":1543045570

1543045570 is a Unix time stamp. This is a way to track time as a running total of seconds. This count starts at the Unix Epoch on January 1st, 1970 at UTC. Therefore, the Unix time stamp is the number of seconds between a particular date and the Unix Epoch. In our example, 1543045570 represents 08/19/2018 @ 4:05pm (UTC). I used unixtimestamp.com for this translation.

Complete Post Analysis

Assume that you have located your suspect (https://www.instagram.com/hak5darren) and found an incriminating post on his account (https://www.instagram.com/p/BK1KWEthQkb). This is a high-priority investigation, and extensive manual documentation is justified. You have already attacked his account with Instaloader and Instalooter as explained in a previous chapter. You possess all of the images from the account. The following outlines the next steps I would take toward this target.

- View the source code of https://www.instagram.com/p/BK1KWEthQkb.
- Search "1080" within the source page.
- Copy the entire URL directly before the third instance of "1080".
- Paste the URL into a browser and download the full version of the suspect image.
- Return to https://www.instagram.com/p/BK1KWEthQkb.
- Scroll through the comments, and expand any if necessary, by clicking "+".
- Click the summary of likes below the heart icon.
- Scroll through them until all are loaded, select all with ctrl-a/command-a on keyboard.
- Open a new spreadsheet through Microsoft Office or LibreOffice and paste the results.
- Rename this tab "BK1KWEthQkb" and open a new tab.
- Repeat the process with any other posts of interest.
- Return to the account (https://www.instagram.com/hak5darren) and click "Followers".
- Load the data, copy it, and paste it into the new sheet titled "Followers".
- Repeat the process for "Following".
- On the target profile, scroll down until all posts are loaded and select/highlight all.
- Right-click and choose "Copy selected links". Paste results into a new sheet titled Posts.

You now have a spreadsheet with multiple pages inside it. The first provides the comments, details, and likes of a specific post. The second displays all of the target's followers. The third shows his friends, and the fourth provides the direct URL of every post for later analysis or extraction. Figure 11.02 displays a partial view of my example, highlighting the final sheet created.

This was a lot of work, and I do not replicate these steps for every Instagram investigation. You must decide when the extra work is warranted. Unfortunately, there is no reliable automated solution which extracts all data.



23	https://www.instagram.com/p/BEnKp-4G9Sg/			
24	https://www.instagram.com/p/BEnKgNCm9SP/			
25	https://www.instagram.com/p/BEnJbyBm9RF/			
26	https://www.instagram.com/p/BEnJDgWm9QJ/			
	BK1KWEthQkb	Followers	Following	Posts

Figure 11.02: A spreadsheet created from a target Instagram post.

Hashtags

A hashtag is a word or phrase preceded by a hash sign (#) used on social media websites and applications, especially Twitter and Instagram, in order to identify messages on a specific topic. If I wanted to view all posts tagged with "osint", the following direct URL would be appropriate.

https://www.instagram.com/explore/tags/osint/

I do not pay much attention to hashtags within my investigations. I prefer to focus more on search terms. Many people do not properly tag a post, and you can miss a lot by limiting your search to a hashtag. If I search for hashtags of #shooting, I may see valuable data. If a potential target only posts the term "shooting" without the "#", I would miss the post. This is why I rely much more heavily on searching within Google, as explained next.

Google Instagram Search

We have exhausted all useful aspects of collecting data from within Instagram.com. The search functions are extremely limited there, so we must utilize a third-party search engine to find further data and connections. In my experience, Google is our best option, and you should be familiar with the queries conducted at the beginning of the chapter. Let's conduct a demonstration.

Earlier, we searched site:instagram.com "hak5darren" and received 165 results. Browsing through these three pages reveals no evidence of a post including the word "Pager". However, I know he posted a photo with that term in the title. Searching the following reveals the target post.

site:instagram.com "hak5darren" "pager"

This confirms that Google indexed the post, but did not present it when I searched the suspect's Instagram username. This means that we should also query search terms when we conduct searches associated with a target. The search tools presented in a moment simplifies this, but let's conduct one more manual attempt. I see that "hak5darren" posts comments to "snubs" within

his posts. I would like to search for any Instagram posts, within any accounts, which contain either a mention of or comment by both "hak5darren" and "snubs". The following is appropriate.

site:instagram.com "hak5darren" "snubs"

Twitter Instagram Search

Many people who post to Instagram broadcast the links on Twitter. Following this data can reveal unique comments from Twitter users. In rare scenarios, it can reveal Instagram posts from private accounts. Posts themselves are not private, only ACCOUNTS are private. If I know the exact URL of your "private" post, I should be able to see the content. If you post an Instagram link on your public Twitter account, that post is not private. If you post images to Instagram and have no followers, you may post the same link to Twitter where you have an audience. Because of these reasons, we should always check Twitter for Instagram posts. The user "hak5darren" on Instagram is also "hak5darren" on Twitter. The following Google search identifies posts on Twitter (site:twitter.com) which mention "hack5darren", and possess an Instagram image URL (instagram.com/p).

site:twitter.com "hak5darren" "instagram.com/p"

Third-Party Tools

Overall, the various Instagram search websites do nothing more than what we replicated in the previous instruction. However, these sites could identify that one piece of vital evidence that we missed. Therefore, it is always best to have as many resources as possible. I have found the following third-party resources beneficial to my investigations.

Search My Bio (searchmy.bio)

This site was mentioned previously in the Twitter tools. It indexes Instagram account biographies (bios) and makes them searchable. The bios are the brief description of the account, as provided by the user. These often include additional usernames, locations, interest, and contacts. A search of the term "gmail.com" within this site revealed 312,314 profiles which include an email address within the bio. I often use this service to search Snapchat usernames with hopes of connecting an account to an Instagram profile.

StalkFest (stalkfest.com)

With such a negative name, this resource is likely to disappear sooner than later. The benefit here is that it allows you to display friends and followers within a static URL. The following options apply to our target, which display posts, friends, and followers.

https://stalkfest.com/account/hak5darren/
https://stalkfest.com/account/hak5darren/friends/
https://stalkfest.com/account/hak5darren/followers/

The standard search feature in the upper portion of this site allows any keyword. You can then choose if this is a hashtag or username search. The benefit here is a partial search. On Instagram, you must search an entire hashtag to get any results. On StalkFest, you can supply any portion of the topic of interest. A search of "osint" reveals numerous hashtags including osinti, osintrend, and osintxu. A static URL for this query is as follows.

https://stalkfest.com/search/tag/osint/

This technique becomes much more powerful when searching a partial username. Assume I want to know everyone who possess an Instagram username which includes "hak5", but not EXACTLY "hak5". I can search within the website or supply the following direct URL.

https://stalkfest.com/search/user/hak5/

The results identify dozens of accounts, and Figure 11.03 displays partial results, including my original target.

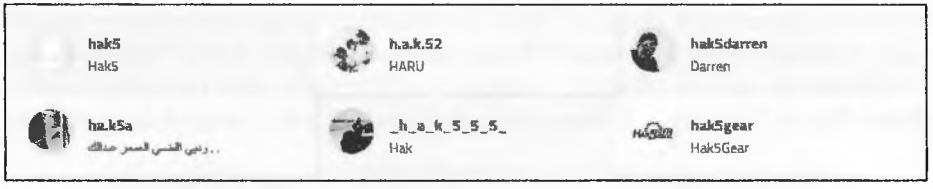


Figure 11.03: A StalkFest partial username result.

IG Audit (igaudit.io)

Similar to Twitter auditing tools, this service aims to identify fake followers. Searching our previous target reveals he has a 73% legitimate follower rate, which is high.

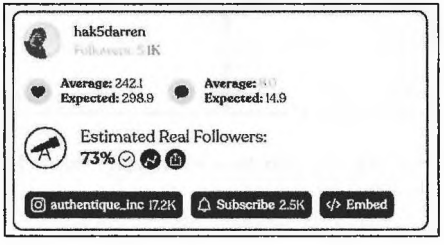


Figure 11.04: An IG Audit report.

Instagram Stories

Instagram Stories are vertical photos or videos, up to 15 seconds in length, which disappear 24 hours after posting. Instead of being displayed in the user's feed, they are displayed at the top of an active user's app when they are logged in. Since Stories disappear quickly, they are often used much more casually than the public Instagram feed. I have found the following utilities helpful for displaying and downloading this temporary content. Below the list, I demonstrate my favorite.

Storiesig (storiesig.com)
IGStorie (igstorie.com)
Instaview (instaview.me)
Story Saver (storysaver.net)

Of these, IGStorie has been the most beneficial. First, it allows query via static URL, which allows me to add it to our custom search tools. Second, it was the only option which displayed "Stories Highlights" of previously expired content. Finally, it provides an option to download all highlights of a specific post as a single archive. This includes images and videos embedded into the content. Consider the following example.

Using the search tool, I navigated directly to the following URL of my target.

https://igstorie.com/profile?username=ambermac

I clicked "Watch/Download Highlights" under the content of interest and was presented multiple short videos. I clicked on the link titled "Download All Highlights As Zip" and was presented a new link to obtain the content. This file contained multiple mp4 video files which could now be played offline. Figure 11.05 displays the results, which include a very helpful date and time stamp.



Figure 11.05: Instagram Stories results from IGStorie.

IntelTechniques Instagram Search Tool

If this seems like a lot of work for minimal information, consider using my custom Instagram Search Tool. Figure 11.06 displays the current version of this page, which includes some of the search options previously discussed. It cannot replicate the source code techniques, but may save you time with username and keyword searching.

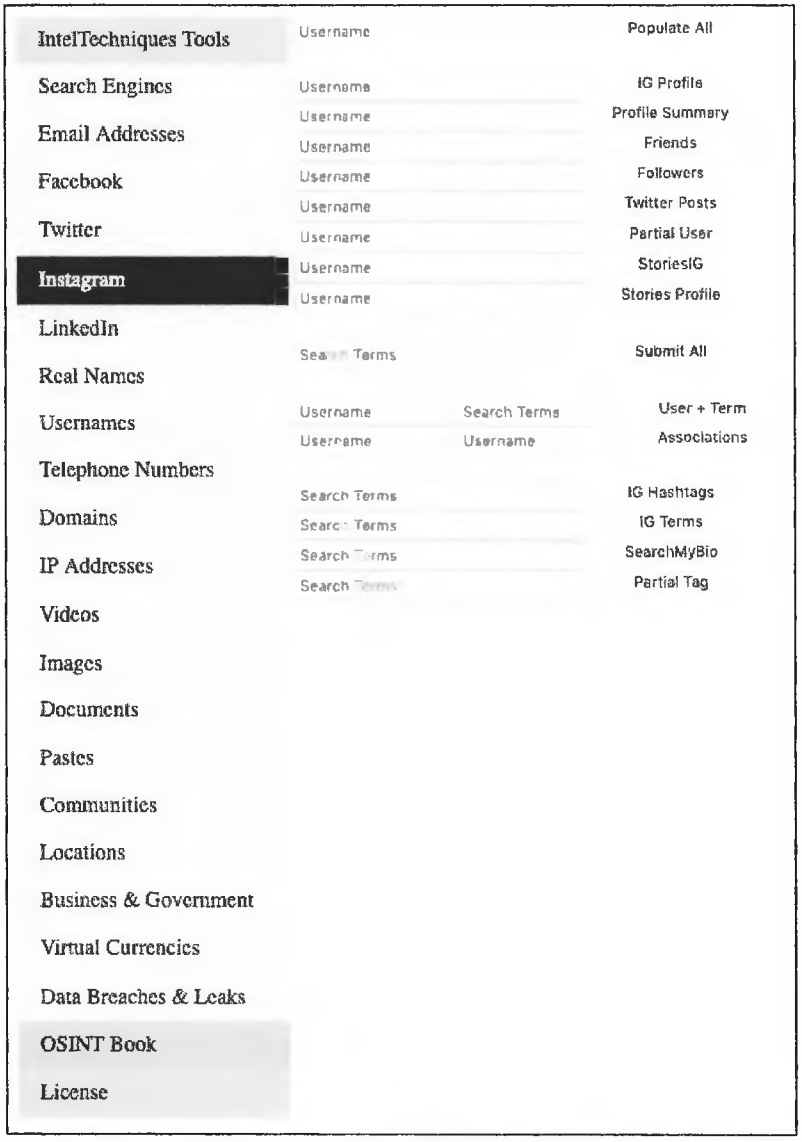


Figure 11.06: The IntelTechniques Custom Instagram Search Tool.

CHAPTER TWELVE
SOCIAL NETWORKS: GENERAL

Facebook, Twitter, and Instagram will each provide more data than all other social networks combined. However, there are many other options, and all should be researched. While we see high usage on the most popular sites, we may not see much incriminating evidence. Instead of finding your suspect confessing to your investigation on Facebook or Twitter, you may be more likely to find his grandmother testifying to his innocence. Smaller networks often provide more intimate details. Your suspect may feel exposed on the bigger networks, but a bit more private on smaller websites. Let's start with the most popular options and work our way down to the niche sites.

LinkedIn (linkedin.com)

When it comes to business-related social networking sites, LinkedIn is the most popular. It is owned by Microsoft and currently has more than 610 million subscribers internationally. The site requires searchers to create a free profile before accessing any data. As with any social network, I recommend creating a basic account with minimal details. The search field of any page offers a search using a real name, company, location, or title. These searches will often lead to multiple results which identify several subjects. The upper center portion of this results page will offer some basic refinements to the search to filter by name, title, company, school, location, and others. Knowing the real name will be most beneficial. The results page should include the target's employer, location, industry, and possibly a photo. After identifying the appropriate target, clicking the name will open that user's profile. If searching a common name, the filters will help limit options.

The profiles on LinkedIn often contain an abundance of information. Since this network is used primarily for business networking, an accelerated level of trust is usually present. Many of the people on this network use it to make business connections. Some of the profiles will contain full contact information including cellular telephone numbers. This site should be one of the first stops when conducting a background check on a target for employment purposes. The target profile often contains previous employment information, alumni details, and work associates. Aside from searching names and businesses, you can search any keywords that may appear within someone's profile. Since many people include their phone numbers or email addresses in their profile, this can be an easy way to identify the user of that specific data. Visiting this profile identifies further information as well as confirmation of the target number.

Searching by Company

If you are searching for employees of a specific company, searching the company name often provides numerous profiles. Unfortunately, clicking on any of these profiles presents a very

limited view with the name and details redacted. The name of the employee is not available, but the photo and job description are usually visible. You are now required to upgrade to a premium account, or be in the same circles as the target, in order to get further information. Instead, consider the following technique.

Search for the business name of your target company, or the employer of your target individual. I typed "Uber" into the search bar and received the official business page on LinkedIn. Clicking the "See all 77,788 employees on LinkedIn" link presented me with numerous employee profiles such as those visible in Figure 12.02. Notice that the names are redacted and only "LinkedIn Member" is available. Clicking this first result prompts me with "Profiles out of your network have limited visibility. To see more profiles, build your network with valuable connections". We struck out, but there are ways that you can proceed in order to unmask these details.

First, copy the entire job description under the "LinkedIn Member" title. In this example, it is "Account Executive at Uber". Use this in a custom Google search similar to the following.

site:linkedin.com "Account Executive at Uber"

The results listed will vary from personal profiles to useless directories. Since Uber is such a large company, I had to view many pages of results until I identified my target. When I opened the 24th search result, the LinkedIn page loaded, and her photo confirmed it was the correct target. The easier way would have been to search the images presented by Google. After the above search is conducted, click on the Images option within Google and view the results. Figure 12.03 (left) displays a section, which easily identifies the same image as the LinkedIn target. Clicking this will load the profile page with full name and details.

Another way to accomplish this is to navigate through the profiles in the "People also viewed" column. These pages include other profiles viewed that are associated with whichever person you are currently analyzing. These people may not be friends or co-workers with your target, but there is a connection through the visitors of their pages. As an example, I returned to the Google search at the top of this page. I clicked on the first search result, which was not my target. However, in the "People also viewed" area to the right, I saw my target, including her full name and a link to her complete profile. Figure 12.03 (right) displays this result.

Finally, the last option is to conduct a reverse image search on the photo associated with the target's profile. Full details of this type of search will be presented later. For this demonstration, I will right-click on her photo and choose "Copy image location". On the Google Images page, I can click the camera icon and submit this URL. While the first result is not the target, clicking the page does present a link to the target's unmasked page. I will later explain many detailed ways to fully query reverse image search options within the upcoming Images chapter. For now, we need to focus on finding our individual targets.

Searching by Personal Details

You might get lucky and find your target with a simple name search. Unfortunately, this is rarely the case. With 610 million profiles, LinkedIn must make some assumptions when choosing the profiles to display after a search. This is especially true with common names. Let's conduct several queries as part of a demonstration of LinkedIn's URL structure. Searching "John Smith" produces practically useless results at the following URL.

https://www.linkedin.com/search/results/content/?**keywords**=john%20smith

Instead of the default "keywords" parameter, let's force a change with the following URL.

https://www.linkedin.com/search/results/content/?**firstName**=john

Changing to "firstName" displays content associated with people named John. This is still fairly unhelpful and includes a lot of content which is not applicable to our target. Now, let's specify the full name of our target in the following URL.

https://www.linkedin.com/search/results/people/?**firstName**=john&**lastName**=smith

The results now only include links to profiles of people with the real name of John Smith. This is much more useful and may be all you need to identify your target. With a name such as John Smith, we need to go a few steps further. The following URL adds his employer (Microsoft).

https://www.linkedin.com/search/results/people/?**firstName**=john&**lastName**=smith &**company**=microsoft

If you wanted to go further, we could specify his title (Manager) and school (Oklahoma) in the following URL. If you know these details about your target, you could start with this, but I find that providing too many details can work against you. Figure 12.01 displays the result of this URL, which identified the only person on LinkedIn which fit the criteria.

https://www.linkedin.com/search/results/people/?**firstName**=john&**lastName**=smith &**company**=microsoft&**title**=manager &**school**=Oklahoma



Figure 12.01: A LinkedIn result via direct URL.

The previous URL query is the most precise option we have. This has been the most beneficial structure I have found to navigate directly to my target. However, it can fail. If your suspect did not provide the school attended or current employer to his profile, you will not receive any leniency from LinkedIn within this search. However, we can rely on Google to help us. Your target may have mentioned an employer somewhere else within the profile or listed a school within the "Interests" area. The following search reveals many profiles possibly associated with the current target example.

site:www.linkedin.com john smith microsoft manager Oklahoma

IntelTechniques LinkedIn Search Tool

These search techniques may seem complicated, but we can simplify them with a custom search tool. The "LinkedIn" section of the tool which you previously downloaded possesses several advanced search features, as seen in Figure 12.04. Practically every "Recruiter Search" style of online website uses these exact methods.

Searching by Country

While LinkedIn is an American company, it is a global social network. If you know that your target is in a specific country, you can filter your search accordingly. This can be done manually by navigating to a foreign subdirectory such as uk.linkedin.com (UK), ca.linkedin.com (Canada), or br.linkedin.com (Brazil). Each of these pages query the entire LinkedIn network, but each places emphasis on local individuals.

PDF Profile View

You may want a quick way to collect the publicly available details from the profiles that you find. One option is to have LinkedIn generate a PDF of the information. While on any profile, click the "More..." button and choose "Save to PDF". This will not extract any private details, but will make data collection fast.

Posts

In years prior, LinkedIn was a place to create a profile and communicate directly with another person. Today, it is a true social network with Posts, Likes, and Comments. Conducting a keyword search through any LinkedIn page or my own custom tool will present anything applicable. Clicking the three dots within a post allows you to copy a static URL to the content, which is beneficial during documentation. Replicating the "friends" and "likes" capture methods described in the Instagram chapter will allow you to document your findings.

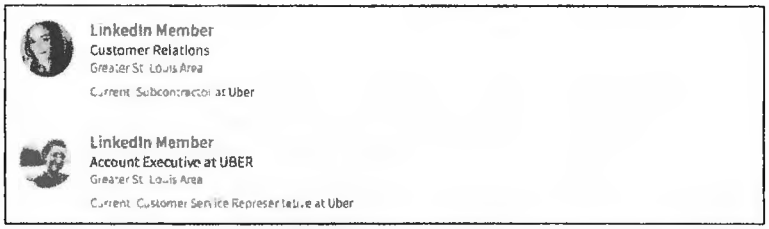


Figure 12.02: Redacted employee results from a business search.



Figure 12.03: Google Images results (left) and un-redacted LinkedIn results (right).



Figure 12.04: The IntelTechniques LinkedIn Custom Search Tool

Tumblr (tumblr.com)

Tumblr was purchased by Yahoo in 2013, Verizon became the owner when it acquired Yahoo in 2017, but then sold Tumblr to WordPress owner Automatic in 2019. While I believe Yahoo/Verizon never took advantage of the reach of this network, I suspect we will see a Tumblr resurgence over the next couple of years. Tumblr is half social network and half blog service. At the time of this writing, there were 471 million blogs and hundreds of billions of posts. These posts can include text, photos, videos, and links to other networks. The search method and layout is no longer user-friendly. The search feature will only identify blogs that were specifically tagged by the creator with the provided search terms. I suggest using a custom Google search. As an example, I conducted a search of "bazzell" within the official Tumblr search and received three results. I then conducted the following search in Google and received the appropriate 244 results.

site:tumblr.com "bazzell"

Snapchat (snapchat.com)

Overall, Snapchat is a difficult OSINT resource. It is only available officially as a mobile application, and there is not any sort of native web search. The majority of the content is set to auto-destruct after a specific amount of time, and is privately delivered from one user to another. In 2018, we started to see much more public content. While there are some extended search options within the app itself, I will only focus here on three web resources.

Snap Map (map.snapchat.com)

This interactive map allows you to query public "Stories" by location. In Figure 12.05, I have zoomed into LAX and I can click on any of the heat maps to load a post. You will see the approximate time in the upper left, but not much additional information. Viewing the source code of any story page will display a static URL to the post, but no additional details. I only find this service valuable when monitoring a high-profile event at a specific location. The search feature only allows entry of a location in order to quickly zoom to an area of interest.



Figure 12.05: A Snap Map result from an airport.

Snapdex (snapdex.com)

The most powerful third-party engine I have located is Snapdex. Enter a username within the search option or navigate directly to snapdex.com/mike and view the bio, a profile picture, location data, and public "snaps". This will only be useful if you know your target's Snapchat name, as any additional search features appear missing. Note that creating a Snapchat username is optional, and many profiles cannot be searched this way.

Telegram (telegram.org)

Most content on Telegram is encrypted private communication between individuals. However, "Channels" were added in 2015 and have become quite popular. These are publicly visible and often include content replicated from other social networks and websites. There is no official search option on telegram.org, but options exist based on the public data. I have found site searches on Google to be the most helpful. The following examples may help explain. I searched "osint" within different official Telegram domains and received very unique results from each.

site:telegram.me "osint" (71 results)
site:t.me "osint" (598 results)
site:telegra.ph "osint" (47 results)

I have also found **Telegram.im** (telegram.im/tools/search.php) helpful when searching channels, users, and groups by name. It provides a graphical interface and robust search feature. Searching "osint" through the standard search option produced thirty results possessing usernames and bio details which included the term osint. Filtering users by the term "osint" revealed seven accounts of interest. I can also filter Channels and Groups by keyword.

Access to private profiles

There have been several "hacks" in the past that would allow a type of "back door entry" into a profile that is marked as private. By the time these methods become known, the vulnerability is usually corrected by the social network host. Websites or applications that publicly claim to be able to access this secured data are most often scams or attempts to steal your own passwords. In my experience, it is best to avoid these traps and focus on finding all available public information. At the time of this writing, an application had recently surfaced that claimed to be able to obtain all information from within private Facebook accounts. The program did not work, but installed a malicious virus instead. If it seems too good to be true, it probably is.

Contact Exploitation

Previously, I explained how to add cell phone numbers and email addresses as contacts to an Android virtual machine in order to supply them to various apps. When programs received the numbers directly from the contacts list, it believed that the contacts were "friends"; therefore, they often identify the names and accounts associated with each number or email. I refer to this technique as contact exploitation, and the Android technique is not the only option for this type of activity. This technique works throughout several social networking environments. I keep covert Yahoo and Gmail accounts solely for adding my target's contact information and asking networks to find friends based on this data. I am often presented with profiles in my target's true name as well as alias accounts. Many people are tricky enough to create an alias profile, but too lazy to connect it to a covert email address or Google Voice number.

We now know that locating someone's social network profile can reveal quite a lot about them. Just knowing a target name can make use of the people search engines that will identify places to seek more information. Unfortunately, sometimes the investigator does not know the target's name. A common scenario is an investigation about some type of event. This could be a specific violent crime, a bomb threat on a campus, or inappropriate chatter about a particular business. All of these scenarios require search engines that monitor social network traffic. There is an abundance of these types of services. Some will work better on reactive investigations after an incident while others show their strength during proactive investigations while monitoring conversations. Many of the sites mentioned here will find the same results as each other.

Overall, some of the strongest methods of searching social network traffic have already been discussed in the Facebook and Twitter chapters. Searching for traffic at the source, such as on twitter.com, will usually provide more accurate and updated content than on an aggregated website of multiple sources. Furthermore, searching specific services through Google or Bing may sometimes quickly locate results that would be difficult to obtain anywhere else. The use of the site operator explained previously will take you far. Aside from direct searches on social networks and targeted search engine queries, there are other options. The accuracy of the services mentioned in the rest of this chapter varies monthly. Hopefully, you will find some of these websites to have value in your investigations. I believe that the options in this chapter should be used to supplement, not replace, the results obtained from previous methods.

Custom Search Engines

The previous Search Engines chapter discussed the creation of custom search engines on Google. Two of the final products created were the Social Networks Search Engine and the Smaller Networks Search Engine. Both of these offer a quick and thorough search of both popular and lesser known social networks. They can often identify communication involving your target. I have found these to be the best initial searches for general content about a specific topic, username, real name, or event. Both of these engines can be accessed at the URLs below.

https://cse.google.com/cse?cx=001580308195336108602:oyrkxatrfyq
https://cse.google.com/cse?cx=001580308195336108602:fdcl5hqdbge

The first queries larger networks such a Facebook, Twitter, Instagram, and LinkedIn. The second focuses on smaller, and outdated, networks such as TheHoodUp, Black Planet, Asian Ave, and Migente. If you do not want to type out the entire URLs, both options can also be found at https://tinyurl.com/y5oscmhp and https://tinyurl.com/y2c9kpej.

Social Searcher (social-searcher.com)

I had previously discouraged users from attempting searches on the first version of Social Searcher. Since then, I have begun to rely on their free service to digest data located on the main social networks. You can provide any keywords, usernames, or terms and receive the most recent results from Facebook, Twitter, Instagram, and the overall web. It allows email alerts to be created for notification of new content matching your query. One of the unique features of this website is the free ability to export search results into CSV format. This output contains the Twitter username, date & time, and entire message among other information. Having this in a spreadsheet format can be incredibly beneficial. This document also included dozens of Reddit and other network posts. The document could be imported into any other collection system.

Account Export Options

The following utilities can be very useful during investigations. Consider a scenario where you have cooperation from a target and consent to view social network accounts. Many suspects will allow you to peek into their online activity with hopes that it will stop your suspicion about them as a suspect. If you have explicit consent, consider collecting all available content from within the target's profiles. Up to this point, I have explained how to properly collect data from the open and public internet. This does not include a person's email content, calendar entries, truly private photos, or personal communication. I believe that any time you have permission from the target to view their accounts with their volunteered credentials, you should also request to archive the contents. This can be very difficult if done manually. The following techniques identify the easiest and most automated solution for the most popular environments.

Google Takeout (takeout.google.com/settings/takeout)

If your target has a Google account, there is an abundance of data available in several different areas. This can include Gmail messages, YouTube channels, Blogs, Calendars, Contacts, and many others. While logged in as the target, navigate to the above website. By default, every option should be selected. Clicking "Next" will forward you to a download page. Accept the default settings and click "Create Archive". Google will package every possible piece of data from the user's account and present it in very large compressed zip files. These can be opened or stored for later view.

Facebook (facebook.com/help/1701730696756992)

Facebook does not offer a specific page for archiving data, but the feature is embedded into the user settings for every account. While logged in as the target, click the top right menu of any Facebook page and select "Settings". Click "Download a copy of your Facebook data" below the General Account Settings tab. Click "Start My Archive". Facebook will send an email to the address on file for the target. A download link in that message will present a compressed file of the entire contents of the target's Facebook profile. It is important to have consent on the target's email account in addition to Facebook for this method to work.

Twitter (help.twitter.com/en/managing-your-account/how-to-download-your-twitter-archive)

Similar to Facebook, Twitter allows you to export an entire account when authenticated as the target user. Click on the profile icon in the upper right area and choose "Settings". On the left menu, choose "Your Twitter Data" at the bottom. Scroll to the bottom of the page and click "Twitter Archive". Click the button labeled "Request Your Archive" and the process will begin. A link will be sent to the email address on file for the target, and it will connect to a compressed file containing the entire account.

Instagram (help.instagram.com/181231772500920)

Log in to the target Instagram account, select the "Settings" option, click the "Privacy & Security" link, and select "Data Download".

During my criminal investigations, I always asked for written consent to view and collect data within my suspect's accounts. When allowed, I would ask them for his or her password and advise that I would go start the process of viewing and collecting the data. I would then execute the archive collection process and return to the interview. This way, downloading the data was consensual, and the target can withdraw consent at any time. Practically every popular service allows users to export their own data. Searching the provider and "export my data" on Google should present you with a tutorial.

International Social Networks

While this book is heavily focused on social networks popular in the United States, they tend to be fairly global with an international presence. This is especially true for Facebook and Twitter. However, there are many social networks which are not popular within the United States that are the primary networks to local residents abroad. This section attempts to identify and explain the most popular foreign networks that may be used by your international targets.

Russia: VK (vk.com)

VK is basically a Russian version of Facebook. You can create a new free account or log in using your existing Facebook credentials. Most search options function without logging in to an account. The page at vk.com/people offers advanced search options which allow filtering by location, school, age, gender, and interests. Most profiles publicly display a user's birthday, location, and full photo collection.

Russia: Odnoklassniki (ok.ru)

Odnoklassniki works similar to most other social media platforms. It is intended to be a way to communicate with friends, as well as an opportunity to network with other people with similar interests. The service is concentrated on classmates and old friends, and translates to "Classmates" in Russian. The official search page is located at ok.ru/search, but you will need to create an account to take full advantage of the options. I have found a targeted site search on Google to be most effective. Searching for Michael Smith would be conducted as follows.

site:ok.ru "michael smith"

The links connect directly to profiles, which can be browsed as normal. These will appear very similar to Facebook profiles. The upper-right portion of a profile will announce the date of the user's last login. Most of the profile details are public, and do not require any type of URL trickery in order to expose the details.

China: QZone (qq.com)

QZone is typically used as a blogging and diary platform, much like LiveJournal. Most of the loyalty to the platform is due to the popularity of the instant messaging tool "QQ" provided to all users. Unfortunately, this is a one-to-one messaging platform, so opportunities for public search are not present. The search options on qq.com pages provide results similar to Chinese search engines such as Baidu. The searches are not restricted to the social network profiles. I have found the following search on Google or Baidu to work best for English queries. Replace "Michael Smith" with your target's real name or username.

site:user.qzone.qq.com "michael smith"

China: Renren (renren.com)

Literally translated as "Everyone's Website", Renren is a Chinese remake of Facebook. It is one of the most popular Chinese social networks. Users earn points for activities such as logging in, posting status messages, commenting, and receiving comments. As users earn points, their level on the site increases, which unlocks special emoticons, profile skins, and the ability to go "invisible" and view other users' profiles without their knowledge. The home page does not allow

profile search, but browse.renren.com does. Clicking on any profiles from this query will prompt you to sign in to an account. However, a targeted site search should eliminate this. The following Google search identified several various pages that contained Michael Smith.

site:renren.com "michael smith"

If the results are too overwhelming, you can use the following structure to filter the content.

site:blog.renren.com "michael smith" (Filter for blog results only)
site:page.renren.com "michael smith" (Filter for profile results only)
site:zhan.renren.com "michael smith" (Filter for news results only)

Latin America: Taringa (taringa.net)

Taringa has a presence in every country in the Spanish-speaking world. Its main markets are Argentina, Spain, Colombia, Chile, Peru, and the U.S. Latino community. The search functionality is fairly straightforward, but the following URLs may produce more efficient results.

taringa.net/buscar/posts/?q=OSINT (Searches OSINT within post comments)
taringa.net/buscar/comunidades/?q=OSINT (Searches OSINT within community posts)
taringa.net/buscar/shouts/?q=OSINT (Searches OSINT within "Shout" posts)
taringa.net/buscar/imagenes/?q=OSINT (Searches OSINT within images)

These social networks represent only a portion of the available options. If you encounter your target within any of these social networks, you should research additional options for that region. Overall, resort to custom Google searches when the foreign language barriers become an issue.

Everything Else

The details of this chapter are often the first to become outdated after a new release. Instead of over-explaining all of the new and upcoming social media monitoring options, I leave a list of resources here for you to research. They all have strengths and (many) weaknesses. You may find a perfect solution here, but you are more likely to see more of the same.

Board Reader (boardreader.com)
HashAtIt (hashatit.com)
UVRX (uvrx.com/social.html)

CHAPTER THIRTEEN
ONLINE COMMUNITIES

Online communities are very similar to social networks. The thin line which has separated the two is slowly disappearing. While social networks cater to a broad audience with many interests, these communities usually relate to a specific service or lifestyle. Some online communities do not get indexed by search engines; therefore, the presence of a target's participation will not always be found through Google or Bing. Any time that a target's interests or hobbies are located through the previous search techniques, you should also seek the online communities that cater to that topic. This can often identify information that is very personal and private to the target. Many people post to these communities without any regard to privacy. Some communities require registration to see the content, which is usually free. Occasionally, a cached version of the pages on the site is available without registering. This chapter will provide methods of infiltrating these communities to maximize the intelligence obtained.

Reddit (reddit.com)

Contrary to some previous editions, I now start this chapter with Reddit. This social news aggregation, web content rating, and discussion website went from a place for those "in-the-know" to a resource often cited on mainstream media. More users than ever post, reply to, and read the user-submitted content in the form of either a link or text, each submitted to a specific category known as a Subreddit. Other users then vote the submission up or down, which is used to rank the post and determine its position on the website's pages. The submissions are then discussed on the "comments" page of every entry. The Subreddits cover practically any topic you can imagine. If your target has the slightest interest in the internet, he or she has probably been to Reddit. As of 2018, there were over 900,000 Subreddits and 250 million registered users.

Reddit Search

The official search option on Reddit has been plagued with problems since inception. The search field in the upper right of every page will allow you to query any terms you desire, but the results have not always been optimal. In 2016, I saw the search function improve drastically, and even with some added new features. When typing terms into the search field on any Reddit page, the results will be from all pages, including thousands of Subreddits. While you should understand this option, and even execute target searches from the home page on occasion, we should consider some advanced search options.

We can replicate that standard keyword search within a URL. This is beneficial for bookmarking searches of interest that need checked often. The format for a search about OSINT is as follows.

https://www.reddit.com/search?q=OSINT

The results from such a generic search can be quite overwhelming. With the following URL, we can force Reddit to only deliver results if our search term is within the title of a post, and not simply present anywhere within the comments.

https://www.reddit.com/search?q=title:OSINT

If you know the name of the Subreddit, you can navigate directly with the following structure.

https://www.reddit.com/r/OSINT/

If you locate a username of interest while searching Reddit, you can load all of that user's posts and comments by clicking on the name. Alternatively, the following URL can be used.

https://www.reddit.com/user/inteltechniques

If you have a target website, and you want to know if the URL has ever been posted as a submission link, the following URL will display all results.

https://www.reddit.com/search?q=site:inteltechniques.com

You can obtain much further analysis from a third-party website called Redstatz. The direct URL below displays the top Subreddits, recent activity, date range, top authors, linked content, and recent posts associated with a target domain.

http://redstatz.com/report/inteltechniques.com

If Reddit is not providing the results you think you should be receiving, you should return to our previous instruction on Google searching. The following query would identify any posts, categories, or users that included the word "surveillance".

site:reddit.com "surveillance"

If you wanted to force Google to restrict its searching to a specific Subreddit, such as OSINT, you would add "/r/osint" after the first portion. If you wanted to restrict the searching to a specific user, you would add "/user/inteltechniques" to the end. The following are two examples.

site:reddit.com/r/osint "surveillance"
site:reddit.com/user/inteltechniques "surveillance"

Deleted Content

If you have identified any Reddit content of interest, you should consider checking any online third-party archives. These historic representations of an account will often disclose previously

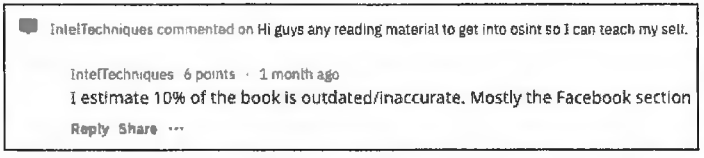
deleted or modified content. It is extremely common for Reddit users to edit or delete a comment entirely, especially if it was controversial. I have investigated numerous Reddit accounts where the evidence I expected to find was not present. First, I always search the standard archive options that were explained previously. The following three direct URLs would attempt to display historic versions of a Reddit user's profile. You could replace the Reddit user URL within each of these with a Subreddit address or Reddit post URL.

webcache.googleusercontent.com/search?q=cache:reddit.com/user/inteltechniques
web.archive.org/web/*/https://www.reddit.com/user/inteltechniques
archive.fo/https://www.reddit.com/user/inteltechniques

You may get lucky with these queries, but the results are often only the tip of the iceberg. These will display the historic view of a Reddit user account at a specific moment in time. While this may provide enough evidence for your investigation, you should also take the next step and identify any further deleted content. In order to dig much deeper, we will rely on Pushshift.

Pushshift (files.pushshift.io)

This huge archive contains over 300GB of data including the most publicly posted content on Reddit since 2005. This provides us an amazing collection of most deleted posts. The next time your target wipes out various Reddit posts before you can collect the evidence, Pushshift may reveal the deleted content. This site allows you to download the entire archive, but that may be overkill for most users. Instead, we can take advantage of their robust application programming interface (API). First, let's assume that you are only interested in a specific user that has deleted all content. I will use my own account as a demonstration. Below is a partial post which I created on Friday, July 19, 2019 at 5:47:32 PM GMT.



On September 1, 2019, I deleted all of my posts, including the above content. Below is the result you should see when you load https://www.reddit.com/user/inteltechniques.



As I wrote this on September 10, 2019, the following direct URL queried the entire data set for any posts that have been archived by Pushshift for user inteltechniques.

https://api.pushshift.io/reddit/search/comment/?author=inteltechniques

This URL presented a lot of information about my deleted posts, but I was most interested in the following. It displayed the entire deleted comment and the subreddit location. Further, it provided a Unix time of 1563580052, which converts to Friday, July 19, 2019 5:47:32 PM GMT.

body:	"I estimate 10% of the book is outdated/inaccurate. Mostly
created_utc:	1563580052

This URL will display the most recent 25 posts, regardless of whether they are still on Reddit or have been removed. This is a great start, but our target may have thousands of posts. The following URL adds two options at the end to force sorting in ascending format and display 1000 comments within a single page.

https://api.pushshift.io/reddit/search/comment/?author=inteltechniques&sort=asc&size=1000

If you are seeking a specific post with unique wording, you can accomplish this with the following URL. This example would identify public and deleted posts mentioning my username.

https://api.pushshift.io/reddit/search/comment/?q=inteltechniques

Each of these searches may present too much content and may not be easy to digest. We can filter unwanted content in order to produce less results. The following would repeat our previous search, but only display content from the Subreddit Privacy.

https://api.pushshift.io/reddit/search/comment/?q=inteltechniques&subreddit=privacy

If you wanted to limit results to a single user with a timeframe between 5 days prior to your search and 30 days prior to your search, you would navigate directly to the following URL.

https://api.pushshift.io/reddit/search/comment/?author=inteltechniques&after=30d&before=5d

In order to demonstrate the value of this, consider the following real example. On December 29, 2017, Reddit user iPhoNewsRO posted on Twitter "great so my reddit account was hacked and used for scams and now it got deleted. lost 5 years of saved posts". Figure 13.01 displays the current view of the account. On December 30, 2017, I navigated to the following URL, which displayed 100 of the most recently deleted comments. His earliest post was "HMU = hit me up aka send me a DM/PM" to Reddit user "Obey_Kush". Evidence of this interaction is not present anywhere on the live view of Reddit.

https://api.pushshift.io/reddit/search/comment/?author=iPhoNewsRO&sort=asc&size=100

Note that all of these searches only identify results which are comments and not user submissions. A submission is a new topic, and a comment is a post within a specific submission. In order to

replicate all of these queries for user submissions, simply replace "comment" in each example to "submission". The following URL only displays my deleted submissions.

https://api.pushshift.io/reddit/search/submission/?author=inteltechniques

In a moment, I will demonstrate my offline search tools which you can use to simplify this entire process. You can also experiment with an online tool at https://redditsearch.io. The ability to extract deleted content from a community as large as Reddit is a substantial OSINT technique. I encourage you to monitor Reddit for new features and changes in order to update your own tools as needed.



Figure 13.01: A deleted Reddit account that was recovered with Pushshift.

Images

Reddit is well-known for hosting entertaining images and memes. The majority of linked images on Reddit are either self-hosted with Reddit or hosted on a photo-sharing site called **Imgur** (imgur.com). This can be very beneficial when you are investigating an image post that has been removed from Reddit. If a user posts a photo to Imgur, then links it to a Reddit post, and then deletes the post, the image is still online. You will no longer have a link to the image, and randomly searching Imgur will be unproductive. Instead, we can browse all of the Reddit images on Imgur with a direct URL. The following address will display the current images, in reverse chronological order, associated with the Subreddit titled NetSec. Scrolling will continuously load older images.

https://imgur.com/r/netsec

If you find an image of interest, you should consider searching the name of the image within Pushshift. Let's run through an example. Assume that you suspected your target was posting images of his antique vehicle on the Subreddit /r/projectcar, but he deleted the posts before you could find them. You should first navigate to the following page of related images on Imgur.

https://imgur.com/r/projectcar

Assume you then located a potential suspected vehicle image at the following address.

https://imgur.com/J0C7Mi9

You should right-click on the image to open view the full-size version in a new tab.

https://i.imgur.com/J0C7Mi9.jpg

The filename of the image is J0C7Mi9. The following two URLs search this image within both submissions and comments within all of Reddit, including deleted posts.

https://api.pushshift.io/reddit/search/submission/?q=J0C7Mi9
https://api.pushshift.io/reddit/search/comment/?q=J0C7Mi9

The second URL provides the following data within the result.

```
author: mrmoto1998
body: [Obligatory pic of the moldsmobile](https://imgur.com/J0C7Mi9.jpg)
created_utc: 1568129504
link: /r/projectcar/comments/d296fl/tore_up_some_carpet_in_the_moldsmobile/eztg3jk/
```

You now know the author, date, and original link of the post on Reddit. The previous example may be an extreme case scenario, but the possibilities are endless. The important message is to search any keyword data through Pushshift when your investigation is associated with Reddit.

Next, you should consider a reverse image search. This will be explained in detail later, but you should know now that you have a Reddit-specific reverse image search option called Karma Decay. Assume that you located an image on Imgur at the following URL.

https://imgur.com/r/funny/0DnE1aB

You can navigate to karmadecay.com, supply this address, and immediately see if that image has been posted to any other locations within Reddit. If you wanted to bookmark a direct URL for future checking, you could use the following to obtain the same result.

http://karmadecay.com/imgur.com/r/funny/0DnE1aB

Note that Karma Decay blocks pornographic images. If your investigation involves any adult photos, you will need to use a different service called NSFW Reddit Reverse Image Search (i.rarchives.com). Enter any image into that website and it will search for other copies within Reddit. Optionally, you can submit directly via URL as follows.

http://i.rarchives.com/?url=http://i.imgur.com/mhvSa.jpg

Investigation Subreddits

There are many Subreddits that can provide a unique benefit to an investigator, three of which are outlined here. There are several versions of each of these, but those that I present here have the most history of being helpful. You will find additional options with a bit of searching.

Reddit Bureau of Investigation (reddit.com/r/rbi)

This active community helps other Reddit users solve crimes and other problems. Internet gurus will help find deadbeat parents; computer specialists will aid in tracking stolen devices; and private investigators will assist with investigation techniques. I have used this option several times during my career. The most successful cases involved hit and run traffic crashes. In 2013, I assisted a northern Illinois police department with the investigation of a fatal car crash. The offender fled the area and an elderly woman died. Three small pieces of the offending vehicle were left at the scene. After posting this information to local media outlets, I submitted it to RBI. Within minutes, several vehicle body shop employees were tracking down the parts and eventually tied them to a specific year and model of a 10-year-old vehicle. This information led to the arrest of the subject. Another victim of an unrelated hit and run traffic crash posted a blurry photo of the suspect vehicle and asked for assistance. Within hours, a Reddit user identified the license plate through digital correction techniques.

Pic Requests (reddit.com/r/picrequests)

A constant frustration in my work is blurry, out of focus, or grainy digital images. Commonly, I will receive surveillance photos that are too dark or light to identify anything of value in the image. Occasionally, I will find images on social networks that could be beneficial if they were just a touch clearer. Pic Requests saves the day. This Subreddit consists of digital photo experts that can perform Photoshop magic on practically any image. Many Reddit users will request old photos colorized, torn photos digitally repaired, or unwanted subjects removed from an image. I have uploaded several surveillance images to this group with a request for assistance. The users have been incredibly helpful by identifying digits in blurred license plates and turning dark surveillance footage into useful evidence.

What Is This Thing? (reddit.com/r/whatisthisthing)

I am consistently amazed at the results from this Subreddit. What Is This Thing is a place where you can post a digital photo of practically anything, and someone will know exactly what it is while providing detailed and cited additional information. Many users post images of old antiques and intricate items hoping to identify something valuable in their collection. I use it to identify tattoo meanings, graffiti, suspicious items mailed to politicians, vehicle parts, and just about anything else that is presented to me during my investigations.

Real World Application: In 2012, I was asked to assist with a death investigation of a "Jane Doe". I submitted a sanitized version of a tattoo on her back that appeared to be Chinese symbols. Within five minutes, a Reddit user identified the symbols, their meaning, and references to the region of China that would probably be related to my investigation. A reverse image search of his examples led to information about a human trafficking ring with which the victim was associated. This all occurred over a period of one hour.

If you plan to use these techniques on Reddit, please consider a few things. You should create a free account now and hold on to it. Creating a new account and asking for help minutes later can be viewed as rude. I like to use accounts that appear to have been established a long time ago. If you are visible as an active member of Reddit with a history of comments, this might encourage other active members to assist you. You should never be demanding in your requests. Remember, these people are volunteering to help you. Many of them possess a skill set that cannot be found elsewhere. I also never upload any content that is not already publicly available. If digital images were released to the press, I have no problem releasing them to Reddit. If my target image is already on a public social network, I see no reason it cannot be linked through Reddit.

While Reddit seems to get most of the attention in this type of community, there are alternative options that are growing rapidly. These include Voat, 4chan, Hacker News, and others. I will briefly discuss the most common search options, which can be replicated with my Custom Communities Search Tool that is explained later.

Voat (voat.co)

When launched in 2014, Voat appeared to be one of many Reddit clones that was surfacing. In June of 2015, it gained a lot of steam as a solid Reddit competitor due to some backlash in the community. At that time, Reddit had just banned several Subreddits that were known to contain harassing content. Many of the hundreds-of-thousands of subscribers that felt betrayed by Reddit fled to Voat as a replacement. It is now known as the home for former Reddit early adopters. While it is very similar to Reddit, you will immediately notice that there is absolutely no search option on the home page, which also requires viewers to be logged in. We can bypass these restrictions with direct URL access. This also helps those who wish to save specific searches for future use. These are very similar to Reddit. Notice the inclusion of "&b=on" (searches body of posts) and "&nsfw=on" (searches adult content).

Text Search: https://searchvoat.co/?t=OSINT&b=on&nsfw=on
User Search: https://searchvoat.co/?u=inteltechniques&b=on&nsfw=on
Domain Search: https://searchvoat.co/?d=inteltechniques.com&b=on&nsfw=on
Subverse (Subreddit) Search: https://searchvoat.co/?s=OSINT&b=on&nsfw=on
Subverse (Subreddit) Page: https://voat.co/v/OSINT
Google Search: https://www.google.com/search?q=site:voat.co+OSINT

4chan (4chan.org)

4chan is a mess. It is an image-board website and users generally post anonymously, with the most recent posts appearing above the rest. 4chan is split into various boards with their own specific content and guidelines, modeled from Japanese image-boards. The site has been linked to internet subcultures and activism groups, most notably Anonymous. The site's "Random" board, also known as "/b/", was the first board to be created, and is the one that receives the most traffic. This site is full of bullying, pornography, threats, and general illicit behavior. It has also been the focus of numerous investigations. Similar to Voat, there is no search feature. In this scenario, we will use 4chansearch.com. The following examples are direct URLs that take advantage of this third-party search option, each using "OSINT" as a search term.

Active Search: http://4chansearch.com/?q=OSINT&s=4
Archives Search: http://4chansearch.com/?q=OSINT&s=7
Archives Alternative: https://archive.4plebs.org/_/search/text/OSINT/order/asc/
Google Search: https://www.google.com/search?q=site:4chan.org%20OSINT

Hacker News (news.ycombinator.com)

While this site is targeted toward a tech-savvy community, general discussion topics are followed by millions of users daily. Fortunately, we have a lot of control with searching specific posts, keywords, users, and favorites. The following searches locate data based on a keyword (OSINT) and user (inteltechniques).

Text Search: https://hn.algolia.com/?query=OSINT&type=all
Username Search: https://news.ycombinator.com/user?id=inteltechniques
User Posts: https://news.ycombinator.com/submitted?id=inteltechniques
User Comments: https://news.ycombinator.com/threads?id=inteltechniques
User Favorites: https://news.ycombinator.com/favorites?id=inteltechniques
Google Search: https://www.google.com/search?q=site:news.ycombinator.com+OSINT

Meetup (meetup.com)

Meetup consists of users and groups, with all communication related to events where people actually meet in real life. Each user creates a profile that includes the person's interests, photos, and username. A group is created by one or more users and is focused on a general interest in a specific location. An example would be the "Houston Dog Park Lovers", which is a Houston-based group of dog owners that meet at dog parks. Each group will post events that the members can attend. The majority of the events posted on Meetup are visible to the public and can be attended by anyone. Some groups choose to mark the details of the event as private and you must be a member to see the location. Membership is free and personal information is not required.

You can search Meetup by interest or location on practically any page. Once you locate a group

page, you can browse the members of the group. This group page will also identify any future and past events sponsored by the group. These past events will identify the users that attended the event as well as feedback about the event. This site no longer offers the option to search by username. In order to do this, you will need to use a search engine as described in a moment. A user profile will often include links to social networks and messages from associates on the website. Additionally, these profiles will identify any future Meetup events that the user plans on attending. Because of this, the site has been used in the past by civil process servers, detectives, and the news media to locate people that had been avoiding them. The following Google search structures have been most helpful in my experience.

Name Search (John Morrison): site:meetup.com inurl:member john morrison
Event Search (Protest): site:meetup.com inurl:events Protest
Post Search: site:meetup.com inurl:discussions Protest
Location Search (Zip-62025): meetup.com/find/events/?radius=5&userFreeform=62025
Google Keyword Search (OSINT): site:meetup.com OSINT

Dating Websites

When investigating cheating spouses, background information, personal character, or applicant details, dating sites can lead to interesting evidence. The presence of a dating profile does not mean anything by itself. Millions of people successfully use these services to find mates. When a target's profile is located, it will usually lead to personal information that cannot be found anywhere else. While many people may restrict personal details on social networks such as Facebook, they tend to let down their guard on these intimate dating websites. In my experience, the following will apply to practically every dating website.

- You must have an account to browse profiles, which is usually free.
- You must have a premium (paid) account to contact anyone.
- If a target uses one dating service, he or she likely uses others.

Instead of explaining each of the dating services, I will focus on methodology of searching all of them. While each website is unique and possesses a specific way of searching, they are all very similar. Overall, there are three standard search techniques that I have found useful, and they are each identified below.

Username: Every dating website requires a username to be associated with the profile, and this data is searchable. Surprisingly, most users choose a username that has been used somewhere else. I have seen many dating profiles that hide a person's real name and location, but possess the same username as a Twitter account. The Twitter account then identifies name, location, and friends. Additional username search tools will be presented later.

Text Search: This is a technique that is often overlooked. Most dating network profiles include an area where the users can describe themselves in their own words. This freeform area often

includes misspellings and obvious grammatical errors. These can be searched to identify additional dating networks since many users simply copy and paste their biography from one site to another. In 2013, I was teaching an OSINT course in Canada. During a break, one of the attendees asked for assistance with a sexual assault case that involved the dating website Plenty Of Fish. The unknown suspect would meet women through the online service and assault them. All of the information on his profile was fake, and the photos were of poor quality and unhelpful. Together, we copied and pasted each sentence that he had written in his bio for the profile. Eventually, we found one that was very unique and grammatically worded poorly. A quoted Google search of this sentence provided only one result. It was the real profile of the suspect on Match.com, under his real name, that contained the same sentence describing himself. The high-quality photos on this legitimate page were used to verify that he was the suspect. An arrest was made within 24 hours.

Photo Search: In later chapters, I explain how to conduct reverse-image searching across multiple websites. This technique can compare an image you find on a dating network with images across all social networks, identifying any matches. This will often convert an "anonymous" dating profile into a fully-identifiable social network page. This applies to any dating networks, and photos will be your most reliable way of identifying your target.

The list of popular dating websites grows monthly. The following are the current most popular services.

Match (match.com)
Plenty of Fish (plentyoffish.com)
eHarmony (eharmony.com)
OK Cupid (okcupid.com)
Christian Mingle (christianmingle.com)
Ashley Madison (ashleymadison.com)
Adult Friend Finder (adultfriendfinder.com)
Farmers Only (farmersonly.com)
Elite Singles (elitesingles.com)
Zoosk (zoosk.com)
Friendfinder-X (friendfinder-x.com)
Badoo (badoo.com)

Tinder (tinder.com)

A section about online dating would not be complete without a reference to Tinder. The simplest explanation of Tinder is that it connects you with people in your immediate geographical area who are also using the service. Some call it a dating app, some refer to it as a "hook-up" app. Either way, it is probably the most popular dating service available today. While this service was once natively available only through a mobile app, they have recently begun allowing account access via their website. However, I find this to be full of frustration. In order to access Tinder via their web interface, you must provide either Facebook account information or a mobile telephone number. I never advise connecting any third-party service to any covert Facebook account, so that option is out. Instead, you can provide a Google Voice number, which is a bit more acceptable. Supply a covert Google Voice telephone number and prepare for the issues.

First, Tinder will send you a text message with a code to verify your number. Supplying this code to Tinder passes the first hurdle. If you are using a VPN, you will immediately be sent to a series of tests to verify you are human. If you have been annoyed by Google Captcha pop-ups in the past, Tinder's options take frustration to a new level. In most scenarios, you will be asked to complete a series of 20 small tasks. In my experience, completing them perfectly results in a failure report 99% of the time. Tinder simply does not like a combination of a VOIP number and a VPN. Providing a real cellular number seems to pacify Tinder somewhat. Providing a true number and internet connection without a VPN seems to make it happy. However, you sacrifice privacy and security in order to do so. If you accept these risks, you can proceed with the web-based access to Tinder. After I explain the process, I will present my chosen solution.

In order to access Tinder from your web browser, several things must be perfectly aligned in order to prevent account blocking. Tinder gets bombarded with fraudulent accounts, and their radar for investigative use is very sensitive. The following instructions assume that you do not have an existing Tinder account.

- Connect to a public Wi-Fi location, without a VPN, near your target.
- Click the Login button at tinder.com and choose "Log In With Phone Number".
- Supply a Google Voice number.
- Confirm the text message received on Google Voice number.
- Complete registration with alias name and photo.

These instructions may seem simple, and too good to be true. They are. Tinder has begun blocking any type of GPS spoofing, even if done manually through the browser inspector. They focus much more on the networks through which you are connected. The previous edition explained ways to spoof GPS and pretend to be at another location. In my experience, these tricks simply do not work anymore. If you are able to connect through the web version of Tinder, it is unlikely to be of any use. Furthermore, your "matches" will likely be far away from your actual location. Personally, I no longer try to make this work.

Instead, I keep an old Android phone ready for any Tinder investigations. I have the Tinder app installed along with the "Fake GPS Location" app. I keep the Tinder app logged in using a covert Google Voice number. Before I open Tinder, I set my desired location through the Fake GPS Location application. Upon loading Tinder, I can control the search settings through the app. I usually choose to limit the search to a few miles away from my target's location. In my experience, this will not work on an iPhone due to GPS spoofing restrictions. Since Tinder actively blocks emulators, connecting through Genymotion does not work. This will simply require a dedicated Android device.

Real World Application: I have two recent experiences with covert Tinder accounts to share. The first is a human trafficking investigation with which I was asked to assist. The investigation unit received information that a well-known pimp had recruited several children to enter his

world of prostitution. He was not promoting them on websites, as he believed it was too risky. Instead, he posted profiles to Tinder while located at a run-down motel. From several states away, I spoofed my GPS on my Android device to the motel of interest. I set my search settings for females within five miles aged 18-19. I immediately observed two images of what appeared to be young girls in a shady motel room. I "swiped right" to indicate interest, and was immediately contacted by the pimp pretending to be one of the girls. We agreed on a price and he disclosed the room number. The local agency working the case began immediate surveillance while a search warrant was obtained. While waiting, they arrested two men before the "date" could begin, and also arrested the pimp after a search warrant was obtained for the room.

The other example I have presents a much different view of Tinder usage. An attorney reached out requesting assistance with a cheating spouse investigation. He was looking for evidence which confirmed the husband was involved with other women. He provided several images of the man and common locations which he was known to frequent. After many failures, I had Tinder launched with the GPS spoofed to the suspect's office. I claimed to be a woman looking for a man his age, and I was eventually presented an image of my target. I swiped right, as did he, and we began a conversation. The evidence piled up immediately.

What photo should you use? First, uploading one photo and vague information to your account looks suspicious. Providing images of other people without their consent is wrong. Using stock photography from the internet will quickly get you banned from Tinder. I rely heavily on **Fiverr** (fiverr.com). I search for people willing to send unique photos of themselves for a few dollars. I once paid a 21-year-old woman $30 for five "selfies" while dressed in different outfits. I received a signed consent form allowing me to upload them to Tinder with the intent of luring cheating spouses. At first, she assumed I was a pervert with a unique fetish. After my explanation, she loved the idea and was eager to assist.

Tinder Profiles

Tinder users can optionally create a username within the network. Instead of being limited to the identity of "Tom, 30, NYC", a user can claim a specific username such as Tom911. This generates a web profile which can be seen publicly (often unknowingly to the user). The following format would display a user with that username. In order to see an active profile, try the second link.

https://www.gotinder.com/@Tom911
https://www.gotinder.com/@MikeB

Discord (discordapp.com)

Discord is a free voice, video and text chat application and digital distribution platform which was originally designed for the video gaming community. Today, it is heavily used within various hacking, doxing, and other communities associated with cybercrime. Some call it a modern-day IRC replacement. Discord allows users to create virtual servers which further divide into text and

voice channels. Discord stores unlimited logs of all channels within every server. Anybody who joins a Discord server has full access to all server history. Access to a Discord server is granted through invites in the form of a URL. Discord is classified as a "deep web" resource, as Discord servers are unable to be indexed by conventional search engines such as Google. I present two initial thoughts on Discord investigative techniques.

- You must receive some type of invite in order to join a server. Once you have an invite, joining is easy and covert details are accepted. You can often find generic invite links within related forums or simply by asking a member of the server. Administrators will know that you joined, but will only see the details you provided during registration.
- Once you are in the server, you should have access to complete chat history since inception. In 2018, I was asked to assist the podcast Reply All with an investigation into the OG Users Discord. Members of this group stole social network accounts from random people and sold them to each other and the public. The primary avenue for communication was through a designated Discord server. The episode is titled 130-The Snapchat Thief if you want to hear more.

Let's conduct a demonstration of finding, joining, and archiving a Discord server. First, I navigated to **Disboard** (disboard.org). This free service indexes numerous Discord servers which have open invitations. I conducted a search for the term "osint" and received one result of "Team Omega Cybersecurity and Analysis". The static Disboard link was the following.

https://disboard.org/server/join/605819996546924544

This immediately forwarded to the Discord link at the following address. This is the official invitation link which could be shared by members of the channel. I found it through Disboard because someone from the group likely posted the details with the intent of increasing usage. If you do not find any servers of interest on Disboard, try **Discord Me** (discord.me).

https://discordapp.com/invite/DBtGker

I was greeted with a login window asking for the name I wished to use in the channel. I provided OSINTAuthor and completed a Captcha. I was immediately given an error which demanded a cellular telephone number in order to join this server. This is common when using a VPN, hardened browser, and guest login. Therefore, I never recommend this route. Instead, register for a Discord account at https://discordapp.com/register, but take a few precautions. In my experience, creating the account from within Chrome appears less "suspicious" than Firefox. Connecting from a network without a VPN seems to allow registration while an IP address from a VPN results in another telephone demand. Therefore, I create a handful of accounts any time I am at a hotel or library. I create them from within my Windows VM using Chrome on the public Wi-Fi without a VPN. These accounts can be stored until needed.

I prefer to conduct all Discord investigations within their official Windows application while inside a virtual machine. Conduct the following steps to replicate my Discord machine.

- Clone your Master Windows 10 machine previously created.
- Title your new VM "Discord" and conduct the following inside the Windows VM.
- Download and install the Discord app from https://discordapp.com/download.
- Download the first file titled "DiscordChatExporter.zip" from the website located at https://github.com/Tyrrrz/DiscordChatExporter/releases.
- Extract the contents of the zip file to your Windows VM Desktop.
- Launch DiscordChatExporter.exe from within the new folder.
- Launch the Discord app, provide your account credentials, and connect to the target Discord server (example: https://discordapp.com/invite/DBtGker).
- Press ctrl-shift-I on the keyboard to launch the Discord developer options.
- Click the arrows in the upper right and select "Application".
- Double-click "Local Storage" and select "https://discordapp.com".
- Press ctrl-r on the keyboard and look for "Token" in the right menu.
- Select and copy the entire token key (without the quotes).
- Paste the token into the DiscordChatExporter program and click the arrow.
- Select the desired target server on the left and the target channel on the right.
- Choose the "CSV" export format and leave the dates empty.
- Choose the save location and click "Export".

The result is a text-only CSV file with the entire contents of the exported server and channel. You should repeat this process for each channel of your target server. In 2019, I located a Discord server used to share file-sharing links to recent data breaches. I exported all of the content, which included thousands of links to mega.nz, Google Drive, Dropbox, and other hosts. The day after I created the export, the Discord server was removed due to policy violations. However, I already had the full export and could investigate the links at my own pace. If desired, you could export an HTML report which would be much more graphically pleasing. I prefer the CSV option because I can import and manipulate the text easier than an HTML web page.

Discord is not the only platform for this type of communication, but I find it to be the most popular with amateur cyber criminals. **Slack** (slack.com/) appears very similar, but it is targeted more toward professionals. **Riot** (about.riot.im) and **Tox** (tox.chat) each possess encrypted communications and better overall security, but adoption is lower than Discord. I believe you should be familiar with all of these environments, and be ready for an investigation within any of them. I keep an investigation VM with all four applications installed and configured. This can be a huge time-saver when the need arises.

Board Reader (boardreader.com)

Online forums provide a unique place for discussion about any topic. If you can think of the subject, an entire site full of people is probably hosting a discussion about the topic. These are usually referred to as user forums. Sometimes, these sites are excluded from being indexed by search engines. This can make locating them difficult. A new wave of forum search sites fills this void. Board Reader queries many forum communities, message boards, discussion threads, and other general interest groups which post messages back and forth. It also offers an advanced search which allows you to choose keywords, language, date range, and specific domain. If you have trouble filtering results on other forum search sites, this can be useful.

Craigslist Forums (craigslist.org)

This forum is categorized by topic instead of location, but location filtering is supported. These forums are not indexed by most search engines, so a manual search is the only way to see the content. In order to search these areas, you must create a free user account. As usual, you can use fictitious information in your profile. After logging in, you can search by keyword on this main page, but not by screen name. This option will identify posts matching the search terms within any of the topics.

The "Handle" option will search by username but can only be seen by clicking on any topic. I entered the "Yoga" room which displays the additional "Handle" search option. This will identify the posts of an individual user. I have found this "handle" option useful to an investigator. As a general rule, most people will use the same username across several sites. Craigslist is no exception. If you have identified a username of a target, a search on the Craigslist forums is worth a look. Although you will not get a result every time you search, the commentary is usually colorful when you do. When you locate a username of a target on the Craigslist forums, searching that username will provide an abundance of information within the user's profile page. These often display a forwarding email address, date joined, photograph, and up to 50 posts made during the past 31 days. These can provide great intelligence on the target.

Online Prostitution

Craigslist was once used by many prostitutes nationwide as an avenue to meeting "Johns". Likewise, many people used the site to locate a prostitute. In 2009, Craigslist was forced to remove the "Erotic Services" section that hosted these posts announcing this activity. In 2018, Backpage was forced offline and seized by various government agencies. Today, it is difficult to find a post offering prostitution on Craigslist and impossible on Backpage. This does not mean that the prostitutes and their clients simply stopped the potentially illegal behavior. Instead, they found new resources. There are many sites online that aid in prostitution and human trafficking. A few of the big players are listed here. I encourage you to investigate which services are applicable to your cities of interest.

https://5escorts.com/
https://www.bedpage.com/
https://cityoflove.com/
https://cityxguide.com/
http://craigslistgirls.com/
https://www.eros.com/
https://www.escort-ads.com/
https://escortfish.ch/
https://www.humaniplex.com/
http://ibackpage.com/
https://escortindex.com/
https://onebackpage.com/
https://openadultdirectory.com/escorts/
https://preferred411.com/
https://sipsap.com/
http://skipthegames.com/
https://www.slixa.com/
https://www.stripperweb.com/
http://theotherboard.com/
https://www.tsescorts.com/
https://www.tnaboard.com/

Escort Review Websites

These types of services may be difficult for some readers to understand. I was also surprised when I first found them. This is where prostitution clients communicate with each other and leave reviews of their experiences with the prostitutes. These "Johns" document the slightest details of the experience including price, cleanliness, and accuracy of the photograph in the ad. Furthermore, this is the first location that will announce an undercover operation by the police. This is important for law enforcement, as this could create an officer safety issue. It is also how the police can determine when a new name or photo should be used in future covert ads. Another purpose for this data is to create documentation of the reviews of an arrested prostitute. This can prove valuable in court for the prosecution of offenses. There are several of these services, and every metropolitan area will have a preferred website by the customers. A Google search of "Escort reviews Anaheim" will get you to the popular options. Of course, replace Anaheim with your city of interest.

The Erotic Review (theeroticreview.com)

If you do not know of any individual services that prostitution clients are using in your area, The Erotic Review is a safe bet. Practically every metropolitan area has a presence here. Much of this site will not be available unless you join as a premium member. However, there should be plenty of visible free content for basic investigations. Most of the posts are unsuitable for this book. At the time of this writing, this site was blocking U.S. IP addresses. Switching my VPN to Canada or any other country bypassed the restriction.

Real World Application: While participating in an FBI Operation, I focused on locating juvenile prostitutes and women forced into the sex industry by pimps. One easy way to determine if a sex worker was traveling extensively was to search her number through various escort websites. If it returned numerous cities with postings, that was a strong indication that she was a full-time sex worker and was likely not traveling alone. Every contact that we made with traveling prostitutes resulted in the identification of the pimps that transported them to the stings.

Online Newspaper Comments

Practically every newspaper now has some sort of online presence. Most digital editions allow readers to leave comments about individual articles. These comments can usually be seen at the bottom of each web page. While the value that these comments add to the newsworthiness of each piece of news is debatable, the content can be important to an investigation. In years past, most newspapers hosted their own digital comment delivery system within their website. This often resulted in a large headache while trying to maintain order, prevent feuds between readers, and delete direct threats. Today, most news websites use third party services to host these comments. The most popular are Facebook and Disqus. When Facebook is utilized, most people use their real names and behave better than when using only a username on Disqus. Any complaints about the comment activity can be referred to Facebook since they technically store the content. Searching Facebook comments can be conducted through the technique explained previously.

In order to search for content within the Disqus comment system, you can conduct a custom Google search. First, it is important to understand how the Disqus system is recognized by Google. There is an option to log in to a Disqus account and you can "upvote" or "downvote" each comment to show your approval. The words visible on this page that were provided by Disqus are important for the search. The word "comments" will be visible on every Disqus provided environment and there will also be a link to disqus.com. Therefore, the following search on Google should provide any websites that have the Disqus comment delivery system and also have a reference to OSINT.

"osint" "disqus" "comments"

This may produce some non-Disqus results that happen to possess all three words, but those should be rare. This will also identify many pages that do not contain any comments whatsoever. In order to only receive results that actually have comments, alter your search to the following.

"osint" "disqus" "1..999 comments"

This instructs Google to only display results that contain the keywords "OSINT" and "Disqus" and also contain the exact phrase of any number between 1 and 999 followed immediately by the term "comments". This would provide results that contain any number of comments with the exception of "0" or over "1000". The "1..999" portion is the Google range operator that will display any number within the specified range.

Craigslist Auctions (craigslist.org)

Craigslist is one big online classified ad for every area of the world. The site can ease the pain of finding an apartment; provide numerous options for buying a vehicle locally; or assist in locating just about any item or service that you can imagine that is within driving distance of your home.

It is also a landing spot for stolen goods, illegal services, and illicit affairs. While Craigslist offers a search option, the results are limited to active posts only. You can also only search within one category at a time. You can browse through the posts individually, but this will be overwhelming.

Government and private investigators have found much success in locating stolen goods within this site. To start, you must find the Craigslist site for your area. Often, simply visiting craigslist.org will direct you to the landing page for your geographical area. If this does not happen, navigate through your country, then your state, then your metropolitan area to see listings around you. If the theft occurred recently, a live search in the "for sale" section may produce results. I do not recommend searching from the main page, as there are no advanced options. Instead, click on any section title. For example, clicking on the "for sale" section will take us to that area. The top of the page will have a search field that will search all of the categories in this section. Additionally, we can filter by price range, posts that contain images, or terms that only appear in the title of the post.

Craigslist also has features that allow you to view results by list view, gallery view, or map view. These locations will only refer to the city of the item, and not exact GPS location. The gallery view can be used as a "photo lineup" to identify a stolen item. The map view can be beneficial when only looking for items within surrounding areas. Four new options on the upper right of every result page allow you to sort the items by newest listings (default), relevance, lowest price, and highest price. Most pages with items for sale will also allow you to filter the results so that only items being sold by individuals are listed. This would eliminate businesses and dealers. The default is to show both, and I recommend leaving that unless you are overwhelmed by the number of results.

If a thief sells the item on Craigslist, he or she will usually delete the post after the transaction is complete. If the post is deleted, it will not be listed in the results of a search on Craigslist. This is where Google and Bing come into play. Both Google and Bing collect information from Craigslist posts to include in their search results. This collection can never be complete, but a large archive of posts is available. Searching Google or Bing with "site:craigslist.org" (without quotes) will search through archived posts on Craigslist that are both active and removed. Similar to the previous example, you can search "site:craigslist.org laptop Edwardsville" (without the quotes). This search produced 572 results that match these criteria on Google. These include the current posts that were available with the live search on craigslist.org as well as posts that have been recently deleted from Craigslist. If you wanted to focus only on a specific regional area of Craigslist, changing the search to "site:stlouis.craigslist.org laptop Edwardsville" would filter results. This example would only show listings from the St. Louis section of Craigslist. You can use any region in your custom searches.

The results that are still current will link to the actual post and display all content of the post. If a search result links to a post that has been deleted from Craigslist, a standard "page not found" error will be returned. You can still get additional information from this deleted post by looking through the text supplied on this search page. The brief description will often disclose an email

address or telephone number. Some listings may have a cached view, but lately this has been rare. In a scenario where thousands of search results are presented by Google or Bing, you can add search terms to filter to a more manageable amount of posts. Adding the make or model number of the product may quickly identify the stolen property.

You can also search by terms other than the product of interest. Many people that use Craigslist do not want to communicate through email sent from the website. Most users will include a telephone number in the post as a preferred method of communication. The overwhelming majority of these telephone numbers belong to the cellular telephone of the user submitting the post. This can be a huge piece of intelligence for an investigator attempting to identify a person associated with a telephone number. It is common that a criminal will purchase a cellular telephone with cash and add minutes to it as needed. This makes it difficult for someone to identify the criminal from the phone number. Ironically, the same criminal will post the telephone number as well as a name on a public internet site for the world to see. Sometimes, a person will post both a cellular and a landline telephone number on the same post. This allows an investigator to associate these two numbers, and a quick internet search should identify the owner of the landline telephone number.

Another way to search Craigslist posts is to identify screen names. Craigslist discourages inserting a screen name or email address within a post; however, most people have figured out how to bypass this limitation. Instead of someone typing their email address within their posts, they will insert spaces between the first portion of the email address (username) and the second portion of the email address (domain name). For example, instead of the user typing their email address as JohnDoe911@gmail.com, he or she may identify the account as "JohnDoe911 at gmail com". This would be enough to prevent Craigslist's servers from identifying the text as an email address and prohibiting the post. Fortunately for the investigator, this information is indexed by Craigslist and other search engines to be retrieved.

You can search any keyword in either the official Craigslist site or on Google and Bing using the "site" operator. In my experience, Bing offers more results of archived Craigslist posts than Google. If you do not have success with Bing, Google should be searched as well. Many private investigators find the "personals" section of interest. The "Casual encounters" area is well known for extramarital affairs. If you want to only search all live Craigslist posts, regardless of which geographical area it exists, you can use sites such as totalcraigsearch.com, adhuntr.com, and searchalljunk.com.

Craigslist has a few advanced search operators that may be of interest. It supports a phrase search with quotation marks such as "low miles". It accepts the hyphen (-) operator to exclude terms such as honda black -red. This search finds postings that have 'honda' and 'black' but not 'red'. A pipe symbol (|) provides "OR" searches such as honda | toyota. This search finds postings that have 'honda' or 'toyota' (or both). You can group terms together in parentheses when queries are complicated. A search of red (toyota | honda) -2000 -2001 finds listings that have 'red' and either 'honda' or 'toyota' (or both) but do not have 2000 or 2001. Wildcards are accepted as follows.

hond* civ* (match "honda civic", "honda civil", etc.)
wood floo* (matches "wood floors", "wood flooring", etc.)
iphone* (matches "iphone", "iphones", "iphone5", etc.)

Craigslist's email alert feature has made third party tools for this purpose unnecessary. After logging in to your account, you can customize alerts to send an email to you when specific search terms are located.

Real World Application: Many thieves will turn to the internet to unload stolen items. While eBay requires banking information or a credit card to use their services, most thieves prefer Craigslist's offer of anonymity. My local police department successfully located a valuable stolen instrument this way and set up a sting to arrest the thief. Often, the thief will be willing to bring the item to you in order to get some quick cash. Another tip that has helped me during investigations is to look for similar backgrounds. When I had a group of gang members stealing iPhones from vehicles and pockets, they would sell them right away on Craigslist. Since there were hundreds of legitimate iPhones listed, identifying the stolen units can be difficult. By looking for similarities in the backgrounds, I could filter the list into interesting candidates. Finding unique backgrounds, such as tables or flooring, within several posts can be suspicious. Additionally, I have found posts that include "hurry", "must sell today", and "I will come to you" to be indicators of illegal activity.

eBay Auctions (ebay.com)

eBay is an online auction site. Since the site requires a user's financial information or valid credit card to post items for sale, many thieves have moved to Craigslist to unload stolen goods. eBay offers an advanced search that will allow filters that limit to auctions from a specific location, or specified distance from the location. On any search page, there is an "Advanced" button that will display new options. Of these options, there is a category titled "show results". The last option in this category is titled "items near me". Here, you can select a zip code and filter results to a minimum of 10 miles from the zip code selected. This will now allow you to search for any item and the results will all be from sellers near a specific zip code. This location option will remain active as you search for different keywords. These searches will only search current auctions that have not expired. In order to search past auctions, select the "Completed listings" option under the category of "Search including". If you want to conduct your searches directly from a URL, or if you want to bookmark queries that will be repeated often, use the following structure. Replace TERMS with your search keywords and USER with your target's username.

Keyword: ebay.com/dsc/i.html?&LH_TitleDesc=1&_nkw=TERMS
Sold: ebay.com/sch/i.html?_from=R40&_nkw=TERMS&LH_Sold=1&LH_Complete=1
Complete: https://www.ebay.com/sch/i.html?_from=R40&_nkw=TERMS&LH_Complete=1
Username: https://www.ebay.com/usr/USER
User Feedback: https://feedback.ebay.com/ws/eBayISAPI.dll?ViewFeedback2&userid=USER
User Items: https://www.ebay.com/sch/USER/m.html

User: http://www.ebay.com/sch/ebayadvsearch/?_ec=104&_sofindtype=25&_userid=USER
User Followers: https://www.ebay.com/usr/USER/followers#followers
User Following: https://www.ebay.com/usr/USER/all-follows?priflwtype=people#people

Flippity (flippity.com)

An alternative to the location feature on the official eBay site is Flippity. This site performs the same function as mentioned previously, but with less work on the user's part. The results of your search will appear on a map with the ability to minimize and expand the radius as desired. This is a quick way to monitor any type of items being sold in a specific community.

GoofBid (goofbid.com)

Not everyone uses spellcheck. Some people, especially criminals, will rush to list an item to sell without ensuring that the spelling and grammar are correct. You could conduct numerous searches using various misspelled words, or you can use GoofBid. This site will take your correctly spelled keyword search and attempt the same search with the most commonly misspelled variations of the search terms. Once this helped me identify a thief selling a "saxaphone". Another alternative to this service is **Fat Fingers** (fatfingers.com).

Search Tempest (searchtempest.com)

If you find yourself searching multiple geographical areas of Craigslist and eBay, you may desire an automated solution. Search Tempest will allow you to specify the location and perimeter for your search. It will fetch items from Craigslist, eBay, and Amazon. You can specify keywords in order to narrow your search to a specific area. Advanced features allow search of items listed within the previous 24 hours, reduction of duplicates, and filtering by categories. While I encourage the use of these types of services, I always warn people about becoming too reliant on them. These tools could disappear. It is good to understand the manual way of obtaining data.

OfferUp (offerupnow.com)

This service is steadily stealing the audience currently dominated by Craigslist. OfferUp claims to be the simplest way to buy and sell products locally. A search on their main page allows you to specify a keyword and location. The results identify the usual information including item description and approximate location. OfferUp follows the eBay model of including the seller's username and rating. The unique option with OfferUp is the ability to locate the actual GPS coordinates associated with a post instead of a vague city and state. This information is not obvious, but can be quickly obtained. While on any post, right-click and choose to view the page source. Inside this new tab of text should be two properties titled offerup:location:latitude and offerup:location:longitude. You can search for these in your browser by pressing ctrl-f (Windows) or command-f (Mac). Next to these fields should display GPS coordinates. In my experience, these precise identifiers will either identify the exact location of the target, or a location in the

neighborhood of the suspect. I would never rely on this all the time, but I have had great success getting close to my targets through this technique.

Amazon (amazon.com)

Amazon is the largest online retailer. Users flock to the site to make purchases of anything imaginable. After the receipt of the items ordered, Amazon often generates an email requesting the user to rate the items. This review can only be created if the user is logged in to an account. This review is now associated with the user in the user profile. An overwhelming number of users create these product reviews and provide their real information on the profile for their Amazon account. While Amazon does not have an area to search for this information by username, you can do it with a search engine. A search on Google of site:amazon.com followed by any target name may link to an Amazon profile and several item reviews. The first link displays the user profile including photo, location, and the user's review of products purchased.

This technique of using Google or Bing to search for profiles on websites that do not allow such a search can be applied practically everywhere. Many sites discourage the searching of profiles, but a search on Google such as "site:targetwebsite.com John Doe" would provide links to content matching the criteria. The difficulty arises in locating all of the sites where a person may have a profile. By now, you can search the major communities, but it is difficult to keep up with all of the lesser known networks.

FakeSpot (fakespot.com)

There is an abundance of fake reviews on Amazon, which can make it difficult to determine which reviews accurately describe a product and which are provided by employees associated with the seller. FakeSpot attempts to identify products that are likely misrepresented by the review community. During a search for a remote-controlled drone, I found that the Amazon "Best Seller" possesses over 53% fake reviews, and top reviewers "tri nguyen" and "EUN SUN LEE" appear to be automated reviewers based on other products. This service also supports analysis of reviewers on Yelp and Trip Advisor.

Pinterest (pinterest.com)

Pinterest is an online "pinboard" where users can share photos, links, and content located anywhere on the internet. It is a way to rebroadcast items of interest to a user. People that follow that user on Pinterest can keep updated on things that the user is searching and reading. The search feature on the main website is useful for keyword searches only. It will search for any term and identify posts that include those words within the description. A search of my last name displayed several photos of people. Clicking each of these links will present the full-page view of the photo and any associated comments. This page will also identify the full name of the person that uploaded the content and the original online source. Clicking on the full name of the user will open the user profile which should include all "pinned" content. Unfortunately, you cannot

search a person's full name or username on Pinterest and receive a link to their profile page. To do this, you must use Google. The following direct search URLs will identify the usernames (BILL) and keywords (CRAFTS) present on Pinterest.

Username: https://www.pinterest.com/BILL/
User Pins: https://www.pinterest.com/BILL/pins
User Boards: https://www.pinterest.com/BILL/boards
User Followers: https://www.pinterest.com/BILL/followers/
User Following: https://www.pinterest.com/BILL/following
Pins Search: https://www.pinterest.com/search/pins/?q=CRAFTS
Boards Search: https://www.pinterest.com/search/boards/?q=CRAFTS
Google Search: https://www.google.com/search?q=site:pinterest.com+CRAFTS

IntelTechniques Communities Search Tool

Similar to the previous search tools, this option attempts to simplify the various search techniques presented within this chapter. Figure 13.02 displays the current view. This tool should replicate all of the specific URLs cited within this topic. While the chances of your target appearing here are lower than large social networks, this resource should not be ignored. In my experience, the details obtained about a target from online communities are usually much more intimate and personal than the public blasts on the larger sites.

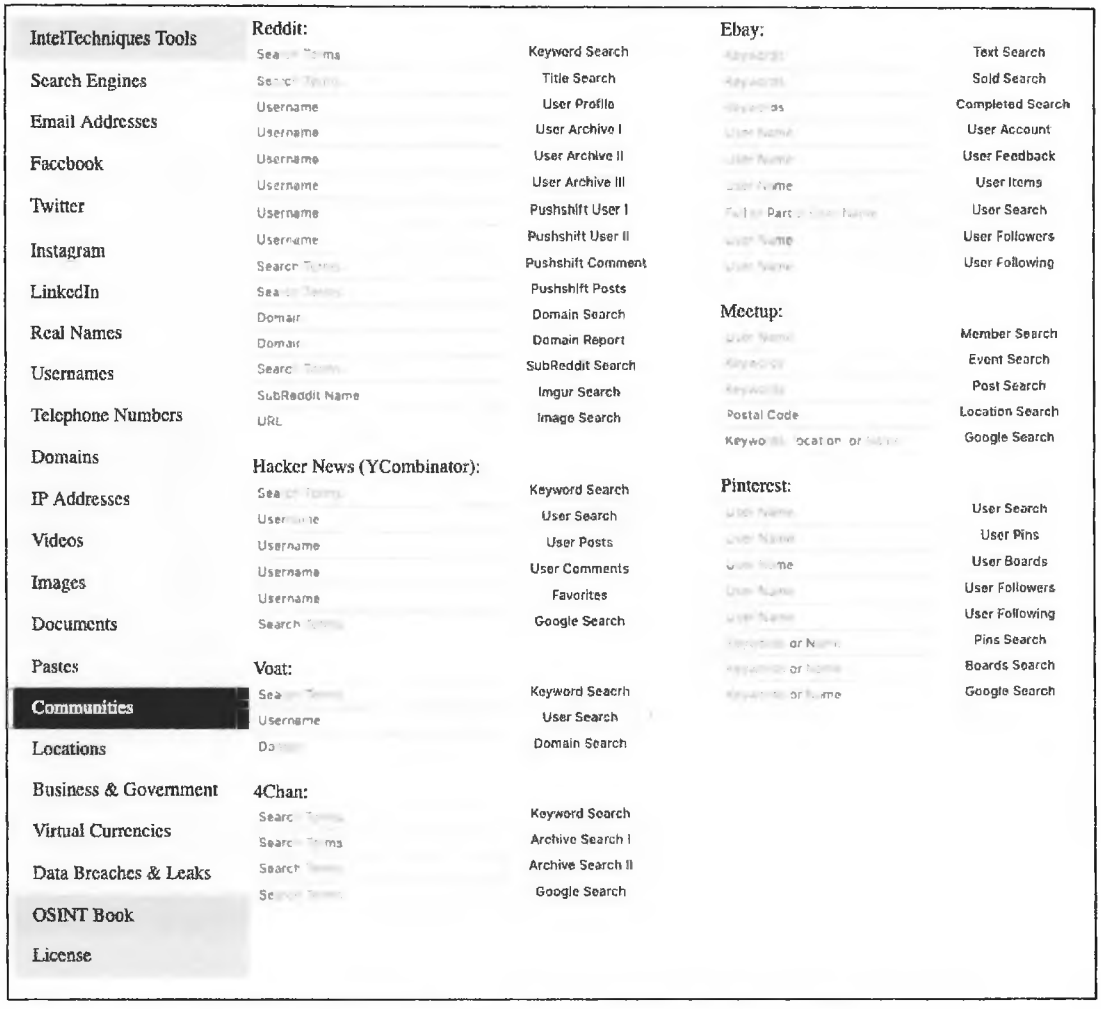


Figure 13.02: The IntelTechniques Communities Search Tool.

CHAPTER FOURTEEN
EMAIL ADDRESSES

Searching by a person's real name can be frustrating. If your target has a common name, it is easy to get lost in the results. Even a fairly unique name like mine produces almost 20 people's addresses, profiles, and telephone numbers. If your target is named John Smith, you have a problem. This is why I always prefer to search by email address when available. If you have your target's email address, you will achieve much better results at a faster pace. There may be thousands of John Wilsons, but there would be only one john.wilson.77089@yahoo.com. Searching this address within quotation marks on the major search engines is my first preference. If you receive absolutely no results in these searches, you should next validate the email address.

TruMail (trumail.io)

When searching for a target by email address, you may find yourself receiving absolutely no results. If this happens, you need to consider whether the email address that you are searching is valid. It is possible that the address was copied incorrectly or is missing a character. There are several websites online that claim to be able to verify the validity of an email address. Most of these do not work with many of the free web-based email providers. One service that stands out from this crowd is TruMail. The sole purpose of the service is to identify if an email address is active and currently being used. After entering an email address, you will be presented with immediate results which will identify if the address is valid or invalid. Further information will identify potential issues with the address. As an example, I searched book@inteltechniques.com, and received the results displayed below. This indicates that the domain provided (inteltechniques.com) is configured for email and the server is active. Otherwise, the account is valid and everything else checks out. It also identifies whether an address is a "catchAll", which may indicate a burner account from that domain. The results confirm it is not a free webmail account nor a disposable temporary account. I find this tool to be more reliable than all the others when searching email addresses, but we should always consider alternatives.

"address": "book@inteltechniques.com",
"username": "book",
"domain": "inteltechniques.com",
"validFormat": true,
"deliverable": true,
"fullInbox": false,
"hostExists": true,
"catchAll": true,
"gravatar": false,
"disposable": false,
"free": false

Verify Email (verify-email.org)

This service provides minimal data, but functions with free webmail hosts. When I searched each of my Gmail accounts, I received "OK" as a response. When I changed one character, I immediately received "BAD" as a result. You are limited to five queries per day (per IP address).

Email Hippo (tools.verifyemailaddress.io)

Email Hippo searches both corporate and personal email accounts, and it provides an additional feature not present in the others. As you validate your target addresses, the responses appear at the bottom as a collection. Choosing the Export option allows you to download all results as a PDF document, Word document, or Excel spreadsheet. Note that you are only allowed 20 free searches per day. In my experience, this has been sufficient for most investigations.

Email Assumptions

You may know about one address, but not others. It can be productive to make assumptions of possible email addresses and use the verifiers to see if they exist. For example, if your target's name is Jay Stewart and he has an email address of jay112003@yahoo.com, you should conduct additional searches for the addresses of jay112003@gmail.com, jay112003@hotmail.com, jay112003@live.com, and others. If you already know your target's username, such as a Twitter handle, you should create a list of potential email addresses. If I had no email address or username for my target (Jay Stewart), but I knew that he worked at the Illinois Medical District Commission (medicaldistrict.org), I may start searching for the following email addresses.

jstewart@medicaldistrict.org
j.stewart@medicaldistrict.org
jay.stewart@medicaldistrict.org
stewartj@medicaldistrict.org

These are merely assumptions of potential addresses. Most, if not all, of them do not exist and will provide nothing for me. However, if I do identify an existing address, I now have a new piece of the puzzle to search. These addresses can be verified against the previous three tools.

Email Format (email-format.com)

If the previous email assumption techniques were unproductive or overkill for your needs, you may want to consider Email Format. This website searches a provided domain name and attempts to identify the email structure of employee addresses. When searching medicaldistrict.org, it provided several confirmed email accounts under that domain and made the assumption that employee emails are formatted as first initial then last name. Our target would have an email address of jstewart@medicaldistrict.org according to the rules. I use this service to help create potential email lists from names collected from Facebook, Twitter, and other social networks. I can then verify my list with the services previously mentioned.

Online Email Databases

If you have a confirmed email address, there are numerous online tools which query multiple databases for any information associated with the account. Many of these are paid premium services that promise results and rarely deliver after payment has been received. I have found almost all paid "email background check" options to be a waste of time and money. Instead, consider the free alternatives, which are more likely to present relative results. In order from most lucrative to least, you should search your target email address through the following options.

Compromised Accounts

Email addresses are compromised regularly. Hacker groups often publicly post databases of email addresses and corresponding passwords on websites such as Pastebin. Manually searching for evidence of compromised accounts can get complicated, but we will tackle this later in the book. Several online services now aid this type of investigation. These services provide one minimal piece of information about any email address entered. They disclose whether that address appears within any publicly known hacked email databases. While they will never disclose the owner, any email content, or passwords, they will confirm that the target's email account has been compromised at some point. They will also identify the service which was targeted during the breach.

This helps us in two ways. First, it confirms an email address as valid. If your suspect account is todd007@gmail.com, and that address was compromised in a breach in 2015, you know the address is valid, it was active, and is at least a few years of age. Second, you know the services which need to be investigated. If todd007@gmail.com was included in the Dropbox and LinkedIn breaches, you should attempt to locate those profiles. Compromised account data is absolutely the most beneficial technique I have used within my investigations in the past three years. We will dive much deeper into data sets later, but let's focus on online services first. These are also good websites to check your own address in order to identify vulnerabilities.

Have I Been Pwned (haveibeenpwned.com)

This is a staple in the breach community. This site allows entry of either a username or email address, but I find only the email option to be reliable. The result is a list and description of any public breaches which contain the provided email address. The descriptions are very helpful, as they explain the type of service associated and any details about the number of users compromised. In a test, I searched an old email of mine that was used to create several covert accounts many years prior. The result was notification that the email address had been compromised on six websites, including Bitly, Dropbox, LinkedIn, Myspace, River City Media, and Tumblr. The data present here is voluntarily contributed to the owner of the site. Cyber thieves steal various credential databases and allow Have I Been Pwned (HIBP) to confirm the legitimacy. HIBP then makes the content searchable and credits the thief by name. The thief can then charge more money for the stolen goods as HIBP has vetted the content. It's a weird game.

Dehashed (dehashed.com)

While Have I Been Pwned is often considered the gold standard in regard to breached account details, we cannot ignore Dehashed. It takes a more aggressive approach and seeks breached databases for their own collection. Using the same email address provided during the previous test, I received two additional breach notifications. These included lesser-known breaches which had not yet publicly surfaced on HIBP. When combining results from both of these services, you would now know that this target email address was likely a real account (you received results); it had been used online since 2012 (the date of the oldest breach according to HIBP); it is connected to an employed individual (LinkedIn); and it exists in spam databases as a U.S. consumer (River City Media). Figure 14.01 displays a typical result.

I believe that Have I Been Pwned and Dehashed complement each other, and one should never be searched without the other. This search alone often tells me more about a target email address, even if it never identifies the owner. Dehashed allows unlimited search for free, but will not disclose passwords without a premium account. They advertise the ability to see all associated passwords for a small fee. I have tried this service in the past, and it worked great. However, as I write this, several people have complained that Dehashed is no longer responding to sales or emails. My attempts to contact customer support went unanswered, so use caution. If they resurface offering paid results, I have found them to be the most affordable option.



Figure 14.01: A Dehashed breached account result.

We Leak Info (weleakinfo.com/)

This service is very similar to Dehashed. You are allowed a free unlimited search tier, but receive no password details. The paid tier unlocks the password field and displays any passwords associated with the target email address as reported by the various breaches. I have never tested the paid services, as We Leak Info has had a poor reputation in the past as a criminal enterprise. Overall, I do not recommend paying any of these services unless absolutely necessary. I will explain self-hosting options later which can replicate many of these breached databases. Figure 14.02 displays a typical result from a free search. We Leak Info allows search of many types of target data, including phone numbers, IP addresses, and names. The static URL for an email search is as follows.

https://weleakinfo.com/search?type=email&query=test@email.com



Figure 14.02: A We Leak Info breached account result.

Leaked Source (leakedsource.ru)

Data from this service will very likely be redundant to the information received from the previous options. However, it is always good to have choices, and to confirm the data received. When searching an email address, Leaked Source displays the company which was breached and the approximate date when searching an email address. Below is a typical example.

WarFrame.com has: 1 result(s) found. This data was hacked on approximately 2016-04-09
NeoPets.com has: 4 result(s) found. This data was hacked on approximately 2013-10-08

Leak Probe (leakprobe.net)

One unique piece of information included within Leak Probe results is the username associated with an email address. In the first example below, I supplied test@email.com as my target account. Leak Probe displayed potential associated usernames of ebexp1 and acton66. This data is retrieved directly from the breaches, which are identified in the results. A URL is displayed below the examples. This is one of the providers which does not sanitize pornography breaches.

ebexp1 test@email.com brazzers.com
acton66 test@email.com xHamster.com

https://leakprobe.net/ajax.php?query=searchLeaks&username=&email=test@email.com

Spycloud (spycloud.com)

This service is a bit different. They are extremely aggressive in regard to obtaining fresh database breaches. They possess many data sets which are not present in HIBP. However, they do not display details about accounts which you do not own. Our only option is through their free API. The following URL submits a query for test@email.com.

https://portal.spycloud.com/endpoint/enriched-stats/test@email.com

The results are in JSON format, which may appear like pure text within your browser. The following result was presented to me when I searched an email address within my own domain.

They basically tell you that the email address queried is present within multiple database breaches, but the identity of each is not available. I use this service to simply verify an email address.


```
"you": {
  "discovered": 7,
  "records": 8,
  "discovered_unit": "Months"

"company": {
  "discovered": 1,
  "records": 11,
  "discovered_unit": "Month",
  "name": "inteltechniques.com"
```


Ghost Project (ghostproject.fr)

The benefit of this service is that it displays a partial view of passwords associated with email addresses within a breach. The weakness is that it possesses a relatively small data set of 1.4 billion credentials. This consists of a "Combo List", which will be acquired later. The following examples identify the masking protocol used to block portions of each password.

test@email.com:55J********
test@email.com:Big*****

Cynic (ashley.cynic.al)

Note that many of these services will not display sensitive results, such as whether your target had an account on the adultery website Ashley Madison. For this type of search, use Cynic. The following result after an email search identifies an account within the Ashley Madison breach.

test@test.com was found. However, someone else could have signed up using your email address.

PSBDMP (psbdmp.ws)

Up to this point, all of these services display content received directly from known data breaches. PSDMP takes a different approach. It monitors Pastebin for any posts including email addresses and/or passwords. They use the Pastebin API to index every post in nearly real-time. Figure 14.03 displays an actual search result of an email address. The hyperlinks on the left identify the randomly assigned name to the "paste". The center column describes the data within the breach, and the final column displays the date of acquisition. Clicking the first link takes you to the post on Pastebin. Searching the target email address on that page should reveal the entry. Typical results appear similar to "test@email.com:filbert" and "test@email.com:simone". The data after each colon is usually a password obtained from within a breach.



Figure 14.03: A PSBDMP breached account result.

IntelligenceX (intelx.io)

A common hurdle with PSBDMP is that posts on Pastebin get removed often. If a complaint is filed about content which violates their terms of service, Pastebin removes the entire post. This creates a dead hyperlink and no data. IntelligenceX attempts to solve this by collecting their own copy of all data indexed. You can search for free and receive partial results, or create a free trial to see everything. In my experience, you are only limited to the number of burner "trial" email addresses with which you have access. Figure 14.04 displays an actual result of a paste containing Bitcoin account credentials. The redaction was applied here after retrieval. The raw results are not redacted and display millions of true passwords.

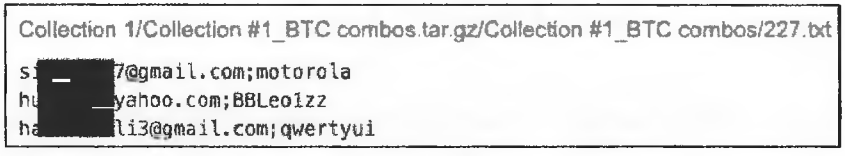


Figure 14.04: An IntelligenceX breached account result.

Combining All Results

I search all email addresses connected to any investigation through every resource listed here. You never know when one service may present a unique response. The results can be very telling. I have much more confidence that an email address is "real" when it appears within a breach. When I search an account which has never been seen within any breach, I assume it is either a brand new account or a "burner" address created for a single purpose. Note that any password details obtained could be accurate or completely wrong. Never attempt to log in to someone's account, as that would be considered computer intrusion. We should use these techniques only for passive intelligence gathering and never as credentials for access.

The following page contains a fictitious example of a breached data summary which would be included with my full OSINT report. I present it only to give you an idea of one way to present your own findings. Always modify any templates and examples toward your own reporting strategies.

Breach Data Report

Target: test@company.com

Service:	Date:	Source:	Details:
LinkedIn	2012	HIBP, Dehashed, WeLeakInfo	Password available
Dubsmash	2018	HIBP, Dehashed	Password available
XKCD	2019	HIBP, Dehashed	Password available
Exploit.in Combo	2016	HIBP, Dehashed, WeLeakInfo	Password available
Anti-Public Combo	2016	HIBP, Dehashed, WeLeakInfo	Password available
Pastebin	2012	PSBDMP, IntelX, Pastebin	Password Found
Pastebin	2015	PSBDMP, IntelX, Pastebin	Password Found
Ashley Madison	2012	Cynic	Credit Card Available

First date present:	2012
Latest date present:	2019
Potential profiles:	LinkedIn, Dubsmash, XKCD, Ashley Madison
Potential passwords:	Pa$$w0rd3, Pa$$w0rd123, Kelly

This is an extremely simple example. In my experience, I have created summaries with three full pages of breached accounts and known passwords within a few email accounts. Once I explain the concept of collecting and querying your own databases near the end of the book, your reports could become quite lengthy.

LinkedIn (linkedin.com)

If my target is likely on LinkedIn, my favorite query is a direct URL submission which attempts to identify the LinkedIn profile associated with the account. Note that you must be logged in to LinkedIn for this to work, and that LinkedIn is likely recording the search in relation to your profile. Always conduct this search with an alias account. The following URL will either present the name and LinkedIn profile connected to the email address, or nothing at all.

https://www.linkedin.com/sales/gmail/profile/viewByEmail/email@here.com

Gravatar (gravatar.com)

This service is responsible for many of the small image icons that you see next to a contact in your email client. You may notice that some incoming emails display a square image of the sender. This is configured by the sender, and this image is associated with any email address connected. You do not need to wait for an incoming message in order to see this image. While the Gravatar home page does not offer an email address search option, we can conduct a query directly from

the following URL, replacing the email address with your target information. This image can later be searched with a reverse image query as explained later.

https://en.gravatar.com/site/check/test@gmail.com

Emailrep.io (emailrep.io)

This is similar to a verification service, but with added features. Below is an actual search result, and I provide an explanation after each detail [within brackets]. As you will see, this is an impressive free search, and at the top of my list for every investigation.

"email": "redacted@gmail.com",	[Email address provided]
"reputation": "high",	[Likelihood to be a real email address]
"suspicious": false,	[Indications of spam or malicious use]
"references": 20,	[Number of online references]
"blacklisted": false,	[Blocked by spam lists]
"malicious_activity": false,	[Known phishing activity]
"malicious_activity_recent": false,	[Known recent phishing activity]
"credentials_leaked": true,	[Present within breaches]
"credentials_leaked_recent": false,	[Present within recent breaches]
"first_seen": "07/01/2008",	[Date first seen online]
"last_seen": "02/25/2019",	[Date last seen online]
"domain_exists": true,	[Whether domain is valid]
"domain_reputation": "n/a",	[Reputation of domain]
"new_domain": false,	[Domain recently registered]
"days_since_domain_creation": 8795,	[Days since the domain was registered]
"spam": false,	[Marked as spam]
"free_provider": true,	[Free provider such as Gmail]
"disposable": false,	[Disposable provider such as Mailinator]
"deliverable": true,	[Inbox able to receive mail]
"accept_all": false,	[Address is a catch-all]
"valid_mx": true,	[Domain possesses an email server]
"spf_strict": true,	[Domain email security enabled]
"dmarc_enforced": false,	[Dmarc email security enabled]
"profiles": "youtube","google","github","twitter","linkedin",	[Profiles associated with email]

Domain Connections

Every domain name registration includes an email address associated with the site. While many people will use a privacy service to hide personal details, this was not always the case. Fortunately, many free services have been collecting domain registration details and offer queries of current and archived domain registration data. They will identify domain names that were registered with the target email address entered. This is beneficial when you have a tech-savvy target that may

have registered websites of which you are not aware. This also works on domains that no longer exist. As a test, I provided an email address of brad@notla.com. The following results identify the details found by each service. The use of affiliate IDs obtained by services such as Analyze ID will be explained during the domain instruction presented later.

Whoxy (whoxy.com/reverse-whois)
Full Name
Home Address
Telephone Number
11 Domain Names
Registrars and Hosts

Domain Big Data (domainbigdata.com)
Full Name
Home Address
Telephone Number
8 Domain Names

DNS Trails (dnstrails.com)
7 Domain Names

Whoismind (whoismind.com)
Gravatar image
3 Domain Names

Analyze ID (analyzeid.com)
Amazon Affiliate ID
AdSense Affiliate ID
5 Domain Names

Spytox (spytox.com)

The first general email search service I use is Spytox. This is a "White Pages" style of query and you will potentially see a name, city, and telephone number associated with the account. The paid options are never worth the money.

That's Them (thatsthem.com)

The majority of the email addresses and usernames I have searched through this service returned no results. However, on occasion I received detailed results such as full name, address, phone number, and vehicle information. Although this is rare, I believe That's Them should be on your list of email and username search resources.

Public Mail Records (publicemailrecords.com)

The data here is believed to be populated by the River City Media data leak. Results include full name and home address. This site has been scraped by many malicious actors and the data is commonly used to generate spam email lists.

Search People Free (searchpeoplefree.com)

This data set is suspected to be generated from various marketing database leaks. Because of this, we usually receive much more complete results. Most results include full name, age, current home address, previous addresses, telephone numbers, family members, and business associations.

Many Contacts (manycontacts.com/en/mail-check)

This premium service offers a free individual email lookup utility. It provides links to any social networks associated with the target email address. During a test search of a target's personal Gmail account, Many Contacts immediately identified the subject's LinkedIn, Twitter, Facebook, FourSquare, Instagram, and Flickr accounts. Hyperlinks connect you directly to each profile page. This is one of my first searches when I know I possess a target's primary email account. I have experienced "no results" when using a common VPN provider, so expect some blockage.

Imitation

The assumptions of email addresses can be valuable in discovering potential valid accounts, as mentioned earlier. Imitation of any target email addresses can reveal more details, and confirm association with online activities. Consider the following example. Your target email address is bill@microsoft.com, and you want to know if he is a Mac or Windows user. You could first navigate to appleid.apple.com/account and attempt to make an Apple account with that address. If allowed to proceed past the first screen, then that user does not already have an account associated with the target address. If you are informed that the email address is already in use, then you know that your target is a Mac user, and that specific address controls the account. You could then navigate to signup.live.com and attempt to create an account with the target address. If denied, then you know that your target is already a Windows user, and the supplied address controls the account.

This method can be replicated across practically all websites and services. I have used this technique to confirm that target email addresses are associated with services from Yahoo, Gmail, Facebook, Twitter, and many others. I have also identified real-world services in use by my target by attempting to create accounts with local cable, power, and water companies, supplying the target email account and being notified that the address was "already in use". Knowing the local cable provider of a suspect can seriously limit the geographical area where he or she could be residing. Be sure not to fully execute any account creations with your target email address, as he

or she will receive a notification. This method should only be used on an initial account creation screen, and never submitted.

Email Provider

If your target's email address ends in gmail.com or yahoo.com, the identity of the email provider is quite obvious. However, business addresses and those with custom domain names do not notify you of the service that hosts the email. A domain's email provider is the company listed in the domain's MX record. The email provider may be the same as the domain's hosting company, but could also be a separate company. You may need to know the email provider in order to issue a court order for content or subscriber data. You may simply want to document this information within your investigation for potential future use. Regardless of your needs, the following will obtain the email provider from almost any address.

Navigate to **MX Toolbox** (mxtoolbox.com) and enter the domain of the email address, such as phonelosers.org. The result should include a hostname and IP address. These identify the email provider for the target domain. In this example, the host is mx1.sub4.homie.mail.dreamhost.com. This tells me that Dreamhost is likely the email provider. You could have also replicated this process in one click using the email search tool at **Records Finder** (recordsfinder.com/email).

These techniques help me identify the email providers or hosts behind business email accounts. If I am conducting a consensual penetration test, this information may lead me toward a social engineering attack against the host. If I am trying to connect an individual to a shell company, this may associate the same small provider with each target. I believe every thorough OSINT report should include a brief mention about the domain email provider. This should be checked as the investigation continues. Changing providers could be a sign of paranoia or intent to conceal evidence. Law enforcement can use this information in order to secure a proper search warrant.

IntelTechniques Email Address Search Tool

Similar to the previous IntelTechniques search tools, I have created a custom email address search tool as explained previously in Section I. This is my most used tool, and the option which requires the most updates. As you find new or updated resources, you will likely want to keep this tool functioning 100 percent. Figure 14.05 displays the version provided at the time of this writing.



Figure 14.05: The IntelTechniques Custom Email Search Tool.

CHAPTER FIFTEEN
USERNAMES

Once you have identified a username for an online service, this information may lead to much more data. Active internet users often use the same username across many sites. For example, the user "amanda62002" on Myspace may be the same "amanda62002" on Twitter and an unknown number of other sites. When you identify an email address, you may now have the username of the target. If a subject uses mpulido007@gmail.com as an email address, there is a good chance that he or she may use mpulido007 as a screen name on a number of sites. If the target has been an internet user for several years, this Gmail account was probably not the first email address used by the target. Searches for mpulido007@yahoo.com, mpulido007@hotmail.com, and mpulido007@aol.com may discover new information. Manual searching of this new username information is a good start. Keeping up with the hundreds of social websites available is impossible. Visiting the following services will allow you to search usernames across several websites, and will report links to profiles that you may have missed. After the details of each service, I provide a comparison chart of features.

KnowEm (knowem.com)

KnowEm is one of the most comprehensive search websites for usernames. The main page provides a single search field which will immediately check for the presence of the supplied username on the most popular social network sites. A search for the username "inteltechniques" provides information about the availability of that username on the top 25 networks. If the network name is slightly transparent and the word "available" is stricken, that means that there is a subject with a profile on that website using the supplied username. When the website is not transparent and the word "available" is orange and underlined, there is not a user profile on that site with the supplied username. For an online researcher, these "unavailable" indications suggest a visit to the site to locate that user's profile. The results could indicate that the target username is being used on Delicious and Twitter, but not Flickr or Tumblr. The link in the lower left corner of the result will open a new page that will search over 500 social networks for the presence of the supplied username. These searches are completed by category, and the "blogging" category is searched automatically. Scrolling down this page will present 14 additional categories with a button next to each category title stating "check this category". This search can take some time. In a scenario involving a unique username, the search is well worth the time. Fortunately for us, and our tools, the following direct URL can be used to save two steps.

https://knowem.com/checksocialnames.php?u=inteltechniques

Check User Names (checkusernames.com)

This site searches approximately one third of the sites on KnowEm, but it links directly to the

target's profile when one is identified. Because this service relies on KnowEm, and not its own internal structure, it could experience outages. Unfortunately, it does not provide a direct search URL since it queries KnowEm through numerous requests.

Name Checkr (namecheckr.com)

This service appeared in late 2014 and conducts the same type of search as the previous competitors. The only slight advantage here is that the search is conducted faster than other sites. Additionally, you have a live hyperlink that will navigate to any identified accounts of the target.

User Search (usersearch.org)

This service stands out a bit from the others in that it only provides actual profile results. It searches the supplied username for a presence on 45 of the most popular websites (basic option) or 115 total websites (advanced option) and returns a list of identified profiles matching the target. While this service is the slowest of all options, this could be an indication of account verification for more accurate results. I have also found their email address search occasionally valuable. The advanced username search can be executed via URL as follows.

https://usersearch.org/results_advanced.php?URL_username=inteltechniques

NameVine (namevine.com)

This username search service provides a unique feature missing in the rest. It allows you to begin typing any partial username and it will immediately identify registered accounts within the top ten social networks. This could be beneficial when you are not sure of the exact name that your target is using. If your target has a Twitter username of "Bazzell", the previous services will easily identify additional accounts that also possess this name. If you think that your target may be adding a number at the end of the username, it could take some time to search all of the possibilities. With NameVine, you can quickly change the number at the end of the username and get immediate results. It will search Twitter, Facebook, Pinterest, YouTube, Instagram, Tumblr, Wordpress, Blogger, and Github. It will also check for any ".com" domains that match. The benefit of this service is the speed of multiple searches. URL submission is as follows.

https://namevine.com/#/inteltechniques

Lullar (com.lullar.com)

Lullar will search by email address, username, or real name. The search excels with an email address or screen name, but I do not recommend the real name option for reliable results. Lullar takes a different approach with the search results. When conducting the search, the results page appears almost immediately. This is because Lullar is not actually conducting any real analysis of user profiles. The results displayed are only the links that would open the page of the target's

profile based on the address, or URL, of that profile. For example, if I search for the username JohnDoe911, Lullar does not check any sites to see if this user has a profile. Instead, it generates the appropriate links that would function for that username. In the case of Twitter, it creates a link to twitter.com/JohnDoe911. This link will be presented whether there is a profile at this address or not. On the down side, you will often be presented with links that do not function. On a positive note, you may get links that do function and are so new that other engines have not indexed them yet.

User Sherlock (usersherlock.com)

While I NEVER recommend using this site's email search feature (it discloses search attempts and your approximate location to the email target), the username option is relatively safe. The only benefit here is the ability to see extended information about each positive result. Real names, locations, and direct links are presented within the minimal results offered. Most of this will be redundant details from the previous queries. The static search URL is as follows.

http://www.usersherlock.com/usersearch/inteltechniques

Social Searcher (social-searcher.com)

This option provides a unique response. It is not a traditional username search site which displays accounts OWNED by the target. Instead, it queries for social network posts which MENTION the target. While searching my own username, I did not locate any profiles associated with me. Instead, I received numerous LinkedIn posts and profiles which mentioned my username. The static URL is as follows.

https://www.social-searcher.com/search-users/?ntw=&q6=inteltechniques

Profilr (profilr.social)

This service does not have a functioning search option, but the direct URL displayed below navigates directly to your target. It only queries Facebook, Twitter, and Github, but it tends to find profiles which do not exactly match the username.

https://www.profilr.social/search/inteltechniques

Skype Username (web.skype.com)

Identifying a Skype username can be an important lead. It could direct you toward additional searches of the newly found data. Unfortunately, usernames are not obviously available when researching a real name on Skype. However, a quick method will reveal any Skype username when searching by name or email address. While logged in to a Skype account within the website or application, navigate to the search area. This section allows you to enter a real name or email

address, and then conducts a search of the Skype user directory. Any results will appear immediately below. Clicking on these results displays the user's basic profile details including a photo. If the user chose not to include a photo, a silhouette graphic appears. Right-click on either image format and choose to "Open image in a new tab" (Chrome) or "View image" (Firefox). The new tab will contain the image that was already available. However, the address in the URL will reveal the Skype username of the target. The following URL was created from the profile image of a Skype user with an email address of lorangb@gmail.com. It identifies the target's Skype username as bart.lorang. The username will always be between two forward slashes (/) after the word "users" within the URL.

https://api.skype.com/users/bart.lorang/profile/avatar?cacheHeaders=1

Compromised Accounts

In the previous chapter, I explained how I use Have I Been Pwned (haveibeenpwned.com) and Dehashed (dehashed.com) in order to identify online accounts associated with a target email address. While HIBP does not always work well with usernames, both Dehashed and We Leak Info (weleakinfo.com) perform well. Similar to the previous instruction, enter your target's username with the intent to discover any data breaches which include these details. The custom search tool begins with multiple breach queries, as I find that information to be the most valuable at the beginning of an investigation.

We can also make assumptions in order to identify target accounts. Assume that your suspect is IntelTechniques on Twitter. A search of the email addresses of IntelTechniques@gmail.com, IntelTechniques@hotmail.com, and IntelTechniques@yahoo.com at HIBP or Dehashed might reveal an active account that appears within a compromised database. Unfortunately, this manual search is time consuming and takes a lot of redundant effort. Therefore, consider using this option within the custom search tools.

Figure 15.01 displays three search fields near the bottom of the page. Each queries a provided username through HIBP, Dehashed, and We Leak Info. The code appends the most popular email domains after the supplied username, and conducts a search within a new tab for each. If you provide IntelTechniques as the target username, the following email addresses are queried within independent browser tabs.

IntelTechniques@gmail.com
IntelTechniques@yahoo.com
IntelTechniques@hotmail.com
IntelTechniques@protonmail.com
IntelTechniques@live.com
IntelTechniques@icloud.com
IntelTechniques@yandex.com
IntelTechniques@gmx.com
IntelTechniques@mail.com
IntelTechniques@mac.com
IntelTechniques@me.com

Any positive results indicate that an address exists matching the search criteria, and that address was present within a breach. If you look at the code behind the tools, it appears as follows.

setTimeout(function() {window.open('https://dehashed.com/search?query=%22' + all3 + '@gmail.com%22', '1leak3window');},1000);

The Timeout function places a short pause in between searches. This ensures that we do not upset the server by conducting too many automated queries at once. If desired, you could replace "gmail.com", or any other options, with an email provider more appropriate for your investigations. If you are outside the U.S., you may want to replace several of these. You could also add more at the end. If I were to add "qq.com" after the last option, it would appear as follows. The 35000 is the next Timeout option to ensure that we have 3000 milliseconds in between each search execution.

setTimeout(function() {window.open('https://dehashed.com/search?query=%22' + all3 + '@qq.com%22', '13leak3window');},35000);

Similar to how compromised database searches are the most powerful email search option that I use, querying usernames in this manner can be equally important. This search tool eliminates any laborious process and removes any excuses not to conduct this type of search every time.

Email Assumptions

The previous option made assumptions about usernames within email addresses which may appear on breach search sites. This option replicates that method within a standard search engine. This option populates a search on Google of the target username with the addition of the most popular email address domains, including quotation marks and the OR operator, as previously discussed. If your target username is IntelTechniques, you could manually conduct the following Google search.

"IntelTechniques@gmail.com"OR"IntelTechniques@yahoo.com"OR"IntelTechniques@hotmail.com"OR"IntelTechniques@protonmail.com"OR"IntelTechniques@live.com"OR "IntelTechniques@icloud.com"OR"IntelTechniques@yandex.com"OR"IntelTechniques@gmx.com"OR"IntelTechniques@mail.com"OR"IntelTechniques@mac.com"OR"IntelTechniques@me.com"

If desired, you could copy this query and paste it into Bing and Yandex, but I have found the results to be very unreliable. Your custom search tool makes this much easier. Enter your username in the "Email Search" option at the end in order to replicate this search. It will usually generate the most results within the quickest amount of time.

IntelTechniques Username Search Tool

Similar to the custom email search tool, this page assists with an automated search of some of the techniques mentioned previously. This page, seen in Figure 15.01, allows you to execute manual queries or use the option to attempt all search possibilities. As a reminder, you must allow pop-ups for these pages if you want to use the "Submit All" option.

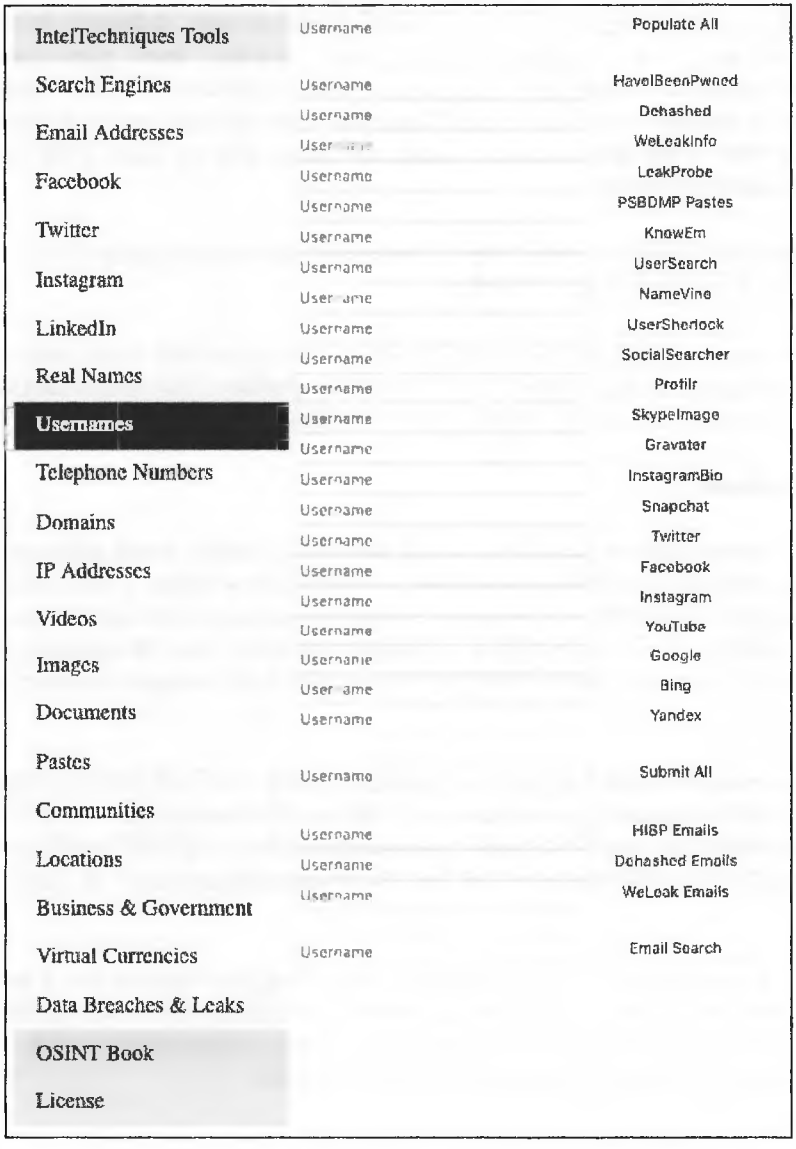


Figure 15.01: The IntelTechniques Custom Usernames Search Tool.

University Homepages

These automated searches for usernames can be very productive. However, they will not locate accounts within all online communities. One large untapped resource is the massive presence of university personal web pages and usernames. Most universities issue each student a university email address. These usually follow a standard naming convention such as the following.

lastname.firstname@university.edu

If you can identify the convention that the school uses and know the full name of the target, you can often determine the email address of the student. This address can be searched in order to identify any social networks that may have been missed. Furthermore, the first part of that address is usually the username that would have been issued to the student for a homepage. The target may have never used this email address online and a search result may appear empty. That does not mean that there is not data available. The chance that the target created some type of homepage while attending the university is high. Finding the content is easy.

Hopefully, the searches explained earlier have helped in identifying a university that the target attended. A search for the university's website will reveal the domain name that the university uses. For example, Southern Illinois University at Edwardsville's website is siue.edu. We can now take that information and conduct a specific search for any personal pages on their domain. The search should look like:

site:siue.edu laura

I picked the name of Laura at random just to identify any student or employee personal website on the SIUE domain. One of the results was a link to a personal website belonging to "Laura Swanson". The link was:

www.siue.edu/~lswanso/

This indicates that the naming convention for personal websites is a tilde (~) followed by the first initial and then the first six letters of the last name. If the target of interest was "Scott Golike", the personal website would probably be at:

www.siue.edu/~sgolike/

We can also assume the possibility of his school issued email account to be sgolike@siue.edu. A few searches using previously discussed techniques should confirm if this address belongs to the target. A search using the email to Facebook profile technique may identify an abandoned profile.

We can now navigate to this personal school page and see if there is any content. If there is, we can collect the data and conduct an analysis for intelligence and further research leads. If there is no page at this address, it does not mean that there has never been data there. This only indicates that there is no current content on this website. When students graduate, universities will usually remove all of the personal content from the servers. As discussed previously, this is never an excuse to stop looking. You can now take the URL of a target and conduct a search on The Wayback Machine (wayback.archive.org).

As an example, I can navigate to the first personal link for "Laura Swanson". Figure 15.02 displays a portion of the live page at www.siue.edu/~lswanso/. If this page did not exist and the site contained no content, you could check on The Wayback Machine. Figure 15.03 shows the search results for this personal page and identifies numerous archives dating back to 1997 for this site. Checking all of these options presents the many different versions of the site including one from 2005 (Figure 15.04) and the first capture from 1997 (Figure 15.05). This presents new data that would not have been uncovered with conventional searches. When a personal website is located, earlier versions should be archived.

Real World Application: While assisting another agency, information had developed in regard to a suspect in a priority investigation. After all online search attempts revealed nothing of value in locating the subject, a deleted student personal page was located using this method. It contained a list of friends, roommates, family members, and interests that were not previously known. This information helped locate the individual within hours.

It should be noted that some institutions will not follow a standard naming convention for all students and faculty. Additionally, there will be occasions when two or more students will have a name similar enough to create the same username. Usually, there is a plan in place to thwart these duplications. Sometimes it is as simple as adding the number "2" after the username.

Universities are not the only places that create personal web pages based on a member name. Several internet service providers allow each subscriber a personal space online as part of the provider's main website. Comcast provides 25MB of storage in a folder with the title of the subscriber's username. For example, if the email address of the customer was crazycheetah70@comcast.net, the username for the service would be crazycheetah70. The URL to view the personal web page would be:

home.comcast.net/crazycheetah70

The following is a sample list of personal web page addresses from additional internet providers, using "crazycheetah70" as a username example. You should also search for internet providers in the target's area and attempt to find deleted pages on The Wayback Machine, Google Cache, Bing Cache, Yandex Cache, and the other services discussed in Chapter Eight.

360.yahoo.com/crazycheetah70
crazycheetah70.webs.com
crazycheetah70.weebly.com
webpages.charter.net/crazycheetah70
sites.google.com/ crazycheetah70
about.me/ crazycheetah70
angelfire.com/ crazycheetah70
geocities.com/ crazycheetah70
reocities.com/crazycheetah70
crazycheetah70.tripod.com
home.earthlink.net/~crazycheetah70
home.comcast.net/~crazycheetah70

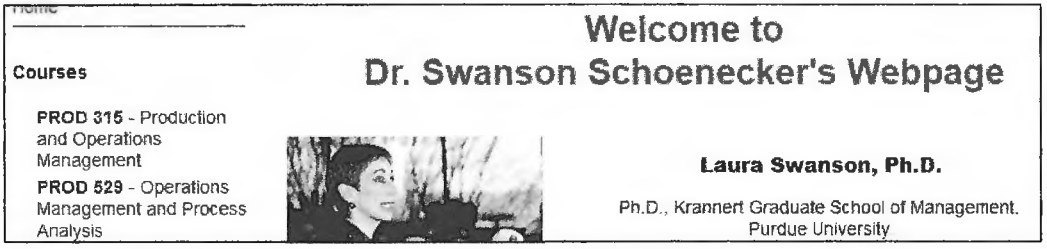


Figure 15.02: A current personal page on a university website.

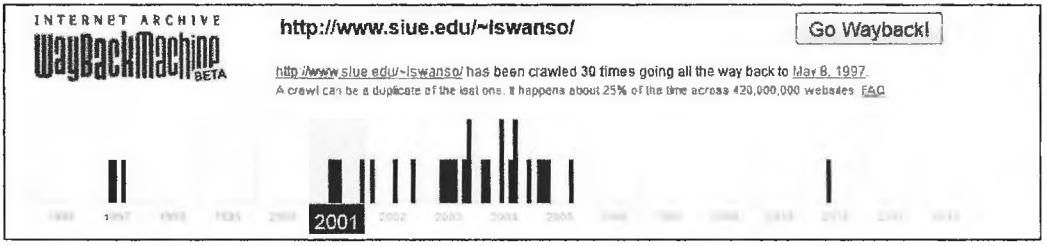


Figure 15.03: A Wayback Machine timeline of available versions of a website.

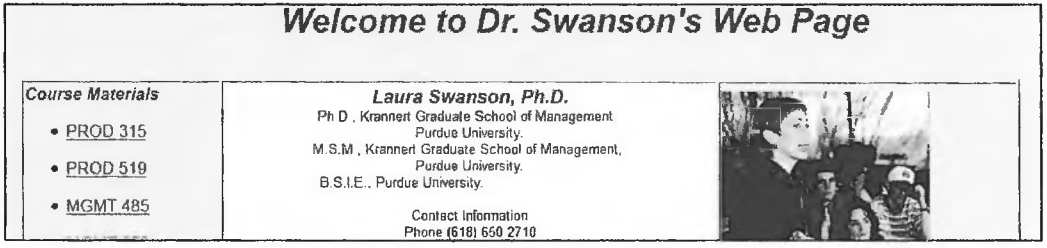


Figure 15.04: A previous version of the website.

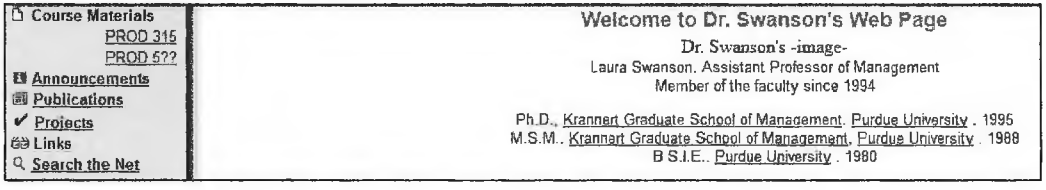


Figure 15.05: A previous version of the website.

CHAPTER SIXTEEN
PEOPLE SEARCH ENGINES

Just as Google and Bing specialize in searching content on the internet, people search engines specialize only in finding content about a particular person. Many of these sites utilize search engines such as Google and Bing to help compile the data, and then present a summary style interface that is easy to consume. The sites listed here each have their own strengths and weaknesses. Standard searches are free on all of them; however, each site generates revenue in some form. Usually, this is by displaying advertisements that often appear to be a report function within the site. I do not recommend purchasing any of the premium paid services until all free options have been exhausted. These details are often focused on targets in the United States, but many services are starting to reach past North America. Searching a target's real name often leads to the discovery of home addresses, telephone numbers, email accounts, and usernames. The following resources are presented in order of most beneficial within my investigations to least.

True People Search (truepeoplesearch.com)

This service has provided the most comprehensive and consistent free report based on real name. The results usually include current address, previous addresses, telephone numbers (including cellular), email addresses, relatives, and associates. There are no blatant ads, but the "Background Reports" links forward you to a paid service which will likely return the same details. This is my first stop when trying to locate an individual in the U.S. The direct search URL is as follows.

https://www.truepeoplesearch.com/results?name=michael%20bazzell

Fast People Search (fastpeoplesearch.com)

This service appears to rely on the exact same database as True People Search. However, it possesses one huge advantage. True People Search gained a lot of media attention as a "stalking" assistant, which resulted in numerous articles instructing individuals to remove their profiles with an opt-out link. Many people have removed their information, but practically none of them replicated the procedure on Fast People Search. If your target removed his or her profile from the previous option, it might still be present here. The direct search URL is as follows.

https://www.fastpeoplesearch.com/name/michael-bazzell

Nuwber (nuwber.com)

A newcomer in 2017 was Nuwber. I first learned about this service from various members of the privacy forum at my website. They were discussing the importance of removing their own

personal details from this site through various opt-out procedures available at the time. I found my own information to be quite accurate. Therefore, this makes for a great OSINT resource. The default landing page allows search of a first and last name. The results are presented by location, and each profile often includes full name, age range, home address, telephone number, and neighbors. The direct search URL is as follows.

https://nuwber.com/search?name=michael%20bazzell

XLEK (xlek.com)

This was another service that surprised many privacy-conscious people. A typical entry contains a target's full name, current home address, current home telephone number, email addresses, previous home addresses, additional telephone numbers, possible relatives, and age. I recently located a target's email address from this service, which was not present on any other site. Using the techniques mentioned in previous chapters, I was able to create a full dossier about him. The default page does not present a direct URL, but using the browser inspector, we find the following.

https://xlek.com/search_results.php?fname=michael&lname=bazzell&locations:all

Family Tree Now (familytreenow.com)

In 2016, this website emerged and launched an uproar online. This site is targeted toward those that want to conduct family history research, and its specialty is connecting a person to his or her relatives. The results do not display home addresses, but simply the age of the target and a large list of family members. After gaining a lot of online popularity, many people started complaining about the availability of this sensitive information. While this type of service is nothing new, people were outraged at this violation of their privacy. Since Family Tree Now sources all of their data from public databases, they defended their product, which is still available today. This attention may have been the inspiration for this company to take things to another level with True People Search. The direct search URL is as follows.

https://www.familytreenow.com/search/genealogy/results?first=Michael&last=Bazzell

Intelius (intelius.com)

Intelius is a premium service that provides reports about people for a fee. Most of the information is from public sources, but some of it appears to come from private databases. Searching for any information on the main website will always link you to a menu of pricing options. The information will never be displayed for free. However, the page that lists the report options does possess some interesting information. This free preview identifies an exact age, possible aliases, cities lived in, previous employers, universities attended, and relatives. If the subject is married, this will usually identify the spouse. In most situations, it will identify the maiden name of the

person's wife. Anything that you do not see on this main screen, you must pay a fee. I never recommend purchasing any of this data. Users are usually disappointed with the results. The direct search URL is as follows.

https://www.intelius.com/people-search/Michael-Bazzell

Radaris (radaris.com)

This service has many similarities to Intelius. However, the data set is unique. The business model is to entice you into purchasing an entire profile. I only use the service for the limited free content available in the preview. After searching a name, select the most appropriate target and choose "Full Profile" in the lower right of the result. This will open the full view of any free information. This will often include the target's middle name, age, current address, previous address, landline telephone number, and links to social networks. The Background Check options will forward you to a third-party premium access website which I do not recommend. The direct search URL is as follows.

https://radaris.com/p/Micheal/Bazzell/

UFind (ufind.name)

This service starts to creep into the "grey" areas of people search websites. The previous options all rely on extremely public data sets such as property tax records. Aside from the traditional address and telephone results, UFind taps into additional sources such as social networks, voter registration data, donation reports, vehicle titles, domains, and even visitors of the White House. The results are extremely comprehensive and have identified targets which were invisible everywhere else. The direct search URL is as follows.

http://ufind.name/Michael+Bazzell

Spytox (spytox.com)

This service provides limited data during a free search, but the records are often updated from premium resources. In my experience, the free option will only provide current city, state, and telephone number. The direct search URL is as follows.

https://www.spytox.com/michael-bazzell

Search People Free (searchpeoplefree.com)

This database looks awfully familiar to True People Search and Fast People Search, but I occasionally locate an individual present here and nowhere else. One benefit here is that the email

addresses identified within the results are hyperlinks. They connect to additional names on occasion. The direct search URL is as follows.

https://www.searchpeoplefree.com/find/michael-bazzell

John Doe (johndoe.com)

I also suspect this database originated from the same source as the previous, but all options should be queried. The strength here is the abundance of "associates". These often include relatives and former roommates. The direct search URL is as follows.

https://johndoe.com/people/michael-bazzell

That's Them (thatsthem.com)

In late 2014, a new website quietly entered the crowded scene of people search services. On the surface, it was just another service that aggregated publicly available information. Consequently, a closer examination revealed That's Them to contain information that is not available anywhere else for free. This service has many options, and most will be discussed in this book. For the purposes of this chapter, I will focus on the "Name and Address" search option in the top menu of the website. Entering a full name with city and state is preferred, but not required. Results often display the person's age range, cell phone number, landline number, full address, religion, financial details, home IP address and any associated email addresses. I searched my own name to test the accuracy of the results. My profile correctly identified similar information as well as the exact VIN number of a previous vehicle. This type of data is impressive without any fees. I have found their details of religion and financial information to be unreliable. Note that the options to purchase additional information are advertisements from third-party companies, and should be avoided. The direct search URL is as follows.

https://thatsthem.com/name/Michael-Bazzell

Spokeo (spokeo.com)

Spokeo is probably the most well-known of all of the people search engines. There are two very distinct versions of this service, free and premium. The premium service will provide a large amount of accurate data, but at a cost. The free version provides an interface that is easy to navigate. The results from a target name search will be presented after choosing a state and city. Only the states and cities where the target name has a presence will be shown. Choosing this target will display a profile with various information. Within this data will be several attempts to encourage you to purchase a premium account. Basically, anything that you do not see within this profile will cost you money. Any links from the profile will present a membership plan with pricing. The profile will often display full name, gender, age, and previous cities and states of residency. However, it no longer presents the actual current address. Well, kind of...

As an example, I searched for people with my name and located an exact match. The profile page only identified a city in West Virginia as a home address. However, the small Google Map in the upper left corner places the marker on top of the address. Zooming out should identify the nearest intersection which can be replicated within a full online mapping service to identify the address. Spokeo is one of many sites which present an animate gif file while you wait on the results. This is unnecessary and is in place to make you believe data is being queried in real time. The following static search URL bypasses this annoyance.

https://www.spokeo.com/Michael-Bazzell?loaded=1

Advanced Background Checks (advancedbackgroundchecks.com)

This is another service that includes advertisements for premium options. It appears to present data very similar to Intelius. Surprisingly, the majority of their data archive is free without any payment. The main search results appear redacted and entire addresses and telephone numbers are masked. With many other services, clicking these details prompt the user for payment. Instead, this service opens a new page revealing the entire record. This often includes home address, home landline telephone number, age, and relatives. Clicking the "See full info" button reveals previous addresses, additional telephone numbers, and aliases. Overall, this service is extremely useful for U.S. targets. The direct search URL is as follows.

https://www.advancedbackgroundchecks.com/names/michael-bazzell

Yasni (yasni.com)

On the surface, Yasni appears to be another standard people search engine. Much of the content received will be duplicate data, but there are a few areas where Yasni works differently. The home page will give three search options. For most OSINT purposes, the last option is the desired search. It will accept a real name or username and forward you to a results page. Real name search will present a large number of links associated with your target's name. As with other engines, many of these results will be about a person other than your target. The first box on the results page will include a "lives/works in" option that will display the cities of the users identified with the search. Clicking on a location that looks appropriate for your target will load a new results page that will provide all search results about your specific target. These links could all be found using search engines and operators, but this will take the hassle out of that technique. Though to obtain complete results on a target, you should still visit a standard search engine. This Yasni page will identify news articles, websites, and social networks related to your target. By default, the search is conducted internationally. Yasni is a German site and searching outside of the United States is one of the strengths of the service. The search bar includes an option to filter the results by specific countries, but the United States is not listed as an option. If you have a target that lives in another country, Yasni is a great tool. The direct search URL is as follows.

http://www.yasni.com/michael+bazzell/check+people?sh

Zaba Search (zabasearch.com)

This site appears to have several search options at first glance. Unfortunately, all but one will forward to an Intelius site, which will require a fee. There is one very specific free option on this page. Providing any real name and state of residence will provide a results page with full name, date of birth, address and phone number. In my experience, this often includes unlisted telephone numbers and addresses. Clicking on practically anything else on this results page will take you to a sponsored link. When I use this resource, I only rely on the information obtained on the first result page. The direct search URL is as follows.

https://www.zabasearch.com/people/michael+bazzell/

People Search Now (peoplesearchnow.com)

This database appears to have the same parent company as True People Search, and possesses the same data. However, this search should be included in the event that a target has removed details from other related websites. Due to increased awareness of exposed personal information, I am seeing many people request removal of personal online data. The direct URL is as follows.

https://www.peoplesearchnow.com/person/michael-bazzell

WebMii (webmii.com)

This service emphasizes information associated with social networks. I have never located any home addresses or telephone numbers, but I have found online images that were not available on any other search engine. This is not the most productive option, but one to consider when desperate for details. The direct search URL is as follows.

http://webmii.com/people?n=%22Michael%20Bazzell%22

Social Searcher (social-searcher.com)

I explained this service within the username chapter, but it also applies here. It focuses on social media presence, including mentions within posts. The direct search URL is as follows.

https://www.social-searcher.com/search-users/?q6=Michael+Bazzell

Truth Finder (truthfinder.com)

There is nothing very special about this service, as it will likely have data similar to the other sites already mentioned. However, there is one major annoyance. When you search this site, you are bombarded by fake progress meters that insinuate that a huge report is being prepared about your target. On average, a real name search takes over 14 minutes due to these constant "please

be patient while we find more details" notifications. The solution to this is to either conduct the query from their information removal page at truthfinder.com/opt-out or submit a direct URL as follows.

https://www.truthfinder.com/results/?firstName=Michael&lastName=Bazzell&state=ALL

People By Name (peoplebyname.com)

The title of this site is misleading. It only allows the query of a telephone number, which is not beneficial when searching by name. However, the following URL structure displays all profiles associated with the provided name.

http://www.peoplebyname.com/people/Bazzell/Michael

White Pages (whitepages.com)

This is the official White Pages website that will conduct a reverse name or address search. The results will include known residents and neighbors. Often, these neighbors listed will include current and previous residents. This data is pulled from public information and is rarely complete. The direct search URL is as follows.

https://www.whitepages.com/name/Michael-Bazzell

Replicating explanations of additional people search databases will only display more redundancy than already present. I leave you with a few others with which I have had success with in the past.

Find People Search (findpeoplesearch.com)
Public Records (publicrecords.directory)
Public Mail Records (publicemailrecords.com/name_search)
Whoodle (whoodle.com)

IntelTechniques Person Search Tool

The abundance of free person search tools can get overwhelming. They each have strengths and weaknesses, and none of them are consistent on accurate results. In years past, I would manually visit each site and enter my target information. This would usually result in small, yet valuable, pieces of information from each service. Today, I use a custom search tool that I created to search all possible sites simultaneously. This free tool will not locate any information that you could not find manually. Instead, it attempts to save you time by automating the process. Figure 16.01 displays the current state of the tool, which is in your downloaded archive from earlier. You can either enter the first and last name of your target within the search fields of each service, or enter this information only once at the final set of search fields. This latter option will launch several new tabs and conduct your search across each service.

During my live training courses, I am often questioned about two specific absences from this tool. The first is the inability to enter a middle initial or name. While I could make this an option, I find that a lot of people search websites do not always possess a middle name. Therefore, entering this data could harm an investigation by omitting valuable results. Some ask why I excluded some resources mentioned in the book. This is due to the way that those sites accept or refuse preset queries via a URL.



Figure 16.01: The IntelTechniques Custom Person Search Tool.

How Many Of Me (howmanyofme.com)

This minimalistic site provides a simple interface to find out how many people exist that have a specific name. In my case, there are 16 people in the United States with my name. This is obtained from census data and can help determine how effective a targeted search will be. For example, if your target has a very unique name, you may still get numerous results to links of social network sites. In order to determine the likelihood that all of these profiles apply to the target, How Many Of Me can tell you whether your target is the only person that has that name. This site provides no intelligence about someone with a common name. I have used this in the past to determine appropriate covert names to use online.

Classmates (classmates.com)

Classmates is a very underrated resource for the internet searcher. Unfortunately, you must create a free account to take advantage of the worthy information inside the site. This free account can contain fictitious information and it is necessary to complete a profile on the site to access the premium features. After you are logged in, you can search by first and last name. If you know the school that was attended, the results will be much more accurate. This should provide the school attended as well as the years the target attended the school. My new interest in this site is due to the availability of scanned yearbooks. The collection is far from complete, but there are a surprising number of complete yearbooks available to browse. Analyzing this content can be very time consuming, as the yearbook must be manually browsed one page at a time. The information obtained should be unique from any internet search previously conducted.

Resumes

Resume searching was mentioned earlier. Those methods will identify many documents, especially if the word "resume" is inside the file or file name. These results are only a portion of available content that could be extremely valuable to your investigation. I believe that resumes are an ideal target since they usually contain sensitive information that is not posted anywhere else. Many people will include a cellular number and personal email address on a resume, but would never consider placing these details on a social network. If the resume is publicly available, regardless of whether the target realizes this, we can gather good intelligence. The following techniques aim to assist in locating this valuable data. Detailed searches within Google or Bing will identify many resumes hosted publicly on websites and cloud-based document storage services. If my target's name is Michael Bazzell, I have found the following exact searches valuable on Google, Bing, and Yandex.

"Michael Bazzell" "Resume"
"Michael Bazzell" "Curriculum Vitae"
"Michael Bazzell" "CV"
"Michael Bazzell" "Resume" filetype:doc
"Michael Bazzell" "Curriculum Vitae" filetype:doc

"Michael Bazzell" "CV" filetype:doc
"Michael Bazzell" "Resume" filetype:pdf
"Michael Bazzell" "Curriculum Vitae" filetype:pdf
"Michael Bazzell" "CV" filetype:pdf
"Michael Bazzell" "Resume" site:docs.google.com
"Michael Bazzell" "Curriculum Vitae" site:docs.google.com
"Michael Bazzell" "CV" site:docs.google.com

While these queries will likely locate any resumes with text, they will fail on many resume images. Numerous resume hosting websites have realized that various data scraping engines scour their resume collection and "steal" their content. This has encouraged some services to store images of resumes that do not contain text that can be easily searched. While this is a decent layer of protection, it is not enough to keep out of Google results. Since Google scans images for Optical Character Recognition (OCR), it knows what words are within an image. After conducting the above searches within traditional engines, attempt them within Google Images (images.google.com). A search of "Mary Johnson" "Resume" on Google Images revealed hundreds of images of resumes. A manual inspection of each identified many pieces of sensitive information.

CV Maker (cvmkr.com)

This website allows users to create free professional resumes and CVs. Currently, over 5 million have been created and are stored within the service. The home page does not offer a search option, as this service is not intended to be used as a people finder. However, we can rely on a Google search to get us the content we want. The following identifies the resume of our target.

site:cvmkr.com "john pratt"

The search result opens a PDF. Within that file is the un-redacted content, which identifies his full email address, telephone number, and home address. On rare occasions, I have found this PDF option to be missing from my target profile. When this happens, we can create a direct link to the full details. In this example, our target's page is at cvmkr.com/7J0N. The following URL presents the entire PDF with the visible details. Basically, adding "?pdf=1" at the end of the URL should always present the full resume view. Since Google indexes all of the PDF files that are located, you can also perform searches for telephone numbers and email addresses using the site operator previously mentioned.

https://cvmkr.com/7J0N?pdf=1

Indeed (indeed.com)

Indeed has a powerful collection of resume data. Because the term "resume" is not present in any of the content pages, you will likely not obtain this data during your standard searches. Entering

your target name on Indeed under the "Find Resumes" option may present new results. Contact information is usually redacted. However, detailed work experience, education, and location are commonly present.

Ripoff Report (ripoffreport.com)

If your target conducts any type of business with the public, he or she will likely upset someone at some point. If your target regularly provides bad service or intentionally commits fraud within the business, there are likely many upset victims. Ripoff Report is a user-submitted collection of complaints about businesses and individuals. I have had numerous investigations into shady people and businesses where these reviews by previously unknown victims were beneficial.

Gift Registries

Decades ago, people were surprised at the gifts presented to them after a wedding or birth. Today, we create online registries identifying the exact products desired, and within moments someone can purchase and ship the "thoughtful" gift with very little effort. As an investigator, I have always enjoyed the plethora of personal details within these registries, which tend to stay online long after the related event. Before identifying the best resources, let's take a look at the types of details we can acquire from some random targets.

Partner Name: When I am investigating someone, that person usually knows that they are under a microscope. He or she tends to stop posting to social media and starts scrubbing any online details. However, their partner tends to ignore the threat of investigation and continues to upload sensitive information applicable to the target. Therefore, online wedding and baby registries help me identify the most lucrative target aside from the original suspect. In an example from the wedding registry website theknot.com, I received over 200 results for Michael Wilson, which also includes the name of the future spouse.

Maiden Name: In the example above, the results only identified future weddings. However, modifying the year in the search menu allows us to view past weddings. This will divulge a woman's maiden name. This can be beneficial in order to better locate a Facebook page or other family members that may be off my radar. I can also use this to search old yearbooks, criminal details, and previous addresses.

Date / State: Many counties will only share marriage certificates if the requestor knows the exact names of each party and the exact date of the event. We have everything we need in order to file a request. Marriage certificates often include full details of all parents, witnesses, and the officiant. Furthermore, I now have their anniversary date which can be helpful during a phishing attack or social engineering attempt. You might be surprised at the number of people that use their anniversary as a security question to an online account.

Ceremony Details: The Knot and other wedding registry sites offer the couple a free website to announce details about the upcoming (or past) event. This usually includes an embellished story about how they met, fell in love, and he proposed. While this could be good knowledge for social engineering, I am usually more interested in the wedding party. This will usually include the closest friends of my target, which will be next on my investigation list.

Items: While it may be fun to look at the items desired by a couple, there is much we can learn about their lives based on these details. In an example from Figure 16.02, I now know that a random Michael Wilson, who is getting married in San Antonio in November 2017, will be going to his honeymoon in Maui (#2), snorkeling (#3), at the airport carrying a Lowepro backpack (#4), checking red/black suitcases (#5), capturing everything on a Canon HD camcorder (#6), dining at the Lahaina Grill (#7), and staying at a fancy nearby hotel (#8).

Other recent examples associated with actual targets identify the types of phones used, vehicles driven, and subjects of interest. While The Knot requires both a first name and last name to conduct a search, providing two asterisks (**) as the first name will present every entry online including the provided last name.

Children: The items within a baby registry will usually provide little to no value. Knowing the brand of diapers preferred or favorite crib style has never helped me in the past. However, knowing a due date and location of the target can be beneficial for future searching. Unfortunately, The Bump only allows searching of upcoming births, and not any past profiles. Fortunately, Google has our backs. The following Google search revealed multiple baby registries from the past few years associated with Michael Wilson.

site:registry.thebump.com "michael wilson"

Gifts: The most fruitful registries in regard to identifying personal preferences of a target are the various gift registries. Of all these, Amazon is the most popular. The following are the most common wedding, baby, and gift registries, with direct links to the most appropriate search pages. I highly encourage you to conduct a detailed Google "Site" search after attempting the proper method.

The Knot: https://www.theknot.com/registry/couplesearch
The Bump: https://registry.thebump.com/babyregistrysearch
Amazon Baby: https://www.amazon.com/baby-reg/homepage/
Amazon Wedding: https://www.amazon.com/wedding/
Target Wedding: https://www.target.com/gift-registry/
Target Baby: https://www.target.com/gift-registry/baby-registry
Kohl's Wedding: https://www.myregistry.com/kohls-wedding-registry.aspx
Registry Finder: https://www.registryfinder.com
My Registry: https://www.myregistry.com



Figure 16.02: A search result from a gift registry website.

Find a Grave (findagrave.com)

While I assume that your goal is to find living targets, you should also have a resource for locating proof of deceased individuals. I have used this website numerous times to locate the graves of recently deceased people. While not necessarily "proof" of death, it provides a great lead toward locating a death certificate and living family members.

Addresses

The target of your investigation may be an address of your suspect. You may want to know who else lives at a residence. There are dozens of websites that possess databases of address information. I have outlined a few here that are unique from those already discussed. Additionally, the following websites which were previously discussed all allow reverse search of a residential address.

https://www.fastpeoplesearch.com/
https://www.whitepages.com
https://www.peoplefinder.com/reverse-address-lookup
https://www.peoplesearchnow.com/
https://www.truepeoplesearch.com/

https://radaris.com
https://www.intelius.com/property-records
https://www.advancedbackgroundchecks.com/address.aspx
https://www.spokeo.com/reverse-address-search
https://thatsthem.com/reverse-address-lookup
https://homemetry.com

Voter Registration (www.blackbookonline.info/USA-Voter-Records.aspx)

Many people will have their address and telephone number unlisted in public telephone books. This prevents their information from appearing on some websites. If any of these people are registered voters, their address may still be public. In order to locate this data, you will need to connect to the county clerk of the county of residence. The link here will display a list of all fifty states. Clicking the state of the target will present all of the counties with known online databases of voter registration content. These will often display the full name of the voter and full address. This can be sorted by name or address depending on what information you have about the target. Later chapters present additional voter registration search options.

Zillow (zillow.com)

This is a popular real estate information site. Entering an address will identify data such as purchase price history, sale status, satellite view map, estimated value, and surrounding real estate information. If the house is for sale or was recently for sale, and the home was listed on a real estate site, you will probably see the sale information here. If this includes interior photographs, which most do, you can view the interior of the house.

Google (google.com)

If all else fails, or you believe you are missing something, check Google. Searching the address should identify any leftover information about your target address. When searching, place the street address in quotes excluding the city. An example search may appear similar to "1234 Main" "Bethalto" IL. This will mandate that any results include the exact address and the exact city name, but they do not necessarily need to be right next to each other on the page. If you place the entire address including the city inside a single set of quotes, you would miss any hits that did not have this exact data. This search should reveal numerous home sale websites that may have unique interior photos.

Spokeo (spokeo.com)

Spokeo was explained earlier as a tool to search a target's real name. A reverse address search will also provide interesting information. Choosing the "Address" option and supplying a full address will identify the last names of any occupants. This is obtained from sources such as utility bills and shipments. The first names of the subjects will be masked and only the first initial is visible.

Paying a monthly fee will eliminate this masking, but is usually unnecessary. Instead, a custom search on Google will usually identify the target names. A site-limited Google search of your target's last name and address details usually identifies the full name of the target. The following search on Google would identify the full names of every person with the last name of Bazzell living at 121 Main Street in Houston.

site:spokeo.com "Bazzell" "121 Main" "Houston"

Non-U.S. Targets

This chapter is heavily focused on U.S. targets. If you have a subject of interest outside of America, I have found the following to be beneficial.

Australia: https://www.peoplesearch.com.au/People_Search_by_Location/
Canada: https://www.canadapages.com
Canada: https://www.infobel.com/en/Canada
France: https://www.infobel.com/en/France
Germany: https://www.infobel.com/en/Germany
Spain: https://www.infobel.com/en/Spain
UK: https://www.192.com/
UK: https://www.peopletraceuk.com/
UK: https://www.gov.uk/government/organisations/land-registry

CHAPTER SEVENTEEN
TELEPHONE NUMBERS

There are hundreds of websites that claim the ability to search for information on telephone numbers and addresses. These vary from amazingly accurate results to sites that only include advertisements. If I have a target telephone number, there are three phases of my search. First, I want to identify the type of number and provider. The type could be landline, cellular or internet, and the provider could be the company supplying the service. Next, I want to identify any subscriber information such as the name and address associated with the account. Finally, I want to locate any online web content with a connection to the target number. This can all lead to more intelligence and additional searches. The majority of cellular numbers can now be identified if they are registered in someone's name. If you have an address, you will want to identify the people associated with the address and any telephone numbers the subjects use. This chapter will highlight the sites that can assist you with these tasks.

Carrier Identification

Ten years ago, I often queried telephone number porting websites to identify the provider of my target's telephone number. This would identify the cellular company that supplied service to my suspect. I would use that information within my court order to demand subscriber data about the target. Knowing the provider was essential as to not waste time requesting records from companies that had no data to provide. The websites used back then have either disappeared, or now charge a substantial fee for access. Five years ago, I noticed that an overwhelming amount of my target telephone numbers were connected to Voice Over Internet Protocol (VOIP) services such as Google Voice and Twilio. This would often be indicated by a result of "Broadband" or "Internet" instead of something obvious such as "Verizon". Until recently, the absence of a specific provider was a hurdle during investigations. Today, we have more sophisticated services that can identify exact provider details on practically any number.

Free Carrier Lookup (freecarrierlookup.com)

I begin with this option because it has been the most stable and does not limit searching as others do. This site requests any domestic or international telephone number and produces a report which includes the country, type of service, and provider associated with the number. While many online services can identify the provider of a cellular or landline number, this option provides the best details about VOIP numbers. I submitted several telephone numbers associated with various providers that I could personally confirm. The following identifies the results of these searches. The first column represents the provider as I knew it, the second column is the type, and the third is the provider displayed by Free Carrier Lookup. This is far from complete, but I wanted to demonstrate the ability to convert internet-based numbers into identifiable companies.

Verizon	Mobile	Verizon Wireless
AT&T	Mobile	AT&T Wireless
Google Voice	VOIP	Google/Level 3
Sudo (US)	VOIP	Twilio/Level 3 (SMS-Sybase) (MMS-SVR)
Sudo (Canada)	VOIP	Iristel Inc.
Sudo (UK)	VOIP	aql Wholesale
Sudo (FR)	VOIP	TRANSATEL
Blur	Landline	Twilio/Level 3 Communications
TextNow	VOIP	Enflick/Bandwidth.com (SVR)
On/Off	VOIP	Peerless Network

If this service cannot provide the data you need, or if you want another source to provide more confidence in the result, you should also consider **Carrier Lookup** (carrierlookup.com). Unfortunately, it restricts users to one free search per day. If both of these services disappear by the time you need them, a Google search of "carrier lookup" should provide new alternatives. Once you have identified the provider, you should focus on the subscriber data associated with the number. We will tackle this in several different ways, as many things changed in 2019. First, we should focus on the most reliable options, which often require registration for a free trial. This chapter becomes quite technical quickly, but the tools at the end will simplify everything.

Caller ID Databases

In 2013, I began experimenting with reverse caller ID data. These are the same databases that identify a telephone number on your landline caller ID display. Often, this will include the name associated with the number. Until recently, this was something that only appeared on landline numbers, but that has changed. Now many cellular telephone numbers have name information associated with them. This name information is usually extracted from the cellular telephone service provider. I was immediately shocked at the accuracy of these results while searching cellular telephone numbers that were otherwise untraceable. On many of my investigations, this technique has eliminated the need to obtain court subpoenas to discover subscriber information.

The reason we can access this data is because it is necessary for telephone systems that do not already possess caller ID options. New systems that operate over the internet, referred to as VOIP systems (Voice Over Internet Protocol), do not receive caller ID data natively. This is something that we have taken for granted while it was provided by our telephone companies. Today, many businesses must purchase access to this data from resellers. This presents us with an opportunity.

I scoured the internet for every business that provides bulk caller ID data to private companies. Some offer a free website for testing; some require you to submit queries through their servers; and others make you register for a free trial. I have tested all of them and identified those that are easy to access and give the best results. First, I will focus only on easy and reliable ways to search an individual number through specific web addresses.

Twilio (twilio.com/lookup)

This company provides VOIP services to many apps, companies, and individuals. An extended feature of their internet-based phone service is the ability to identify incoming calls through caller ID. Fortunately for us, they provide a page on their site that allows queries against their database, but it requires you to register for a free trial account. Look for the "Sign up and try Lookup" button and provide alias information including a burner email address. Click the confirmation link in the email and you should be all set. If you are required to verify a telephone number, Google Voice numbers are accepted. You should be presented with a $15 credit within Twilio. This should provide numerous searches.

Click the three dots within a circle on the left menu in the dashboard and select the "Lookup" option. This is the screen which allows querying of numbers, and you may want to bookmark this page. I usually only select the "Include caller name" unless I want confirmation of the carrier. These searches are $0.01 each, which means you should be able to conduct 1500 queries on this single trial. While a bit of an annoyance to set up the first time, the work is justified. Make sure you have a valid email and Google Voice account ready before registering.

On the Lookup search page, insert the target number and allow the page to generate a result, which will appear directly below. The result will include the type of provider and any name associated with the billing. If I only had one website to search, this would be it. Below is an actual redacted result.

PHONE NUMBER	+16180000000
NATIONAL FORMAT	(618) 000-0000
COUNTRY CODE	US
CALLER NAME	M Bazel
CALLER TYPE	null
CARRIER NAME	Google (Grand Central) - Level3 - SVR
CARRIER TYPE	voip

Open CNAM (opencnam.com)

If I could only search two websites, this would be my second pick. Similar to Twilio, you must create a free trial account. You must provide an email address and telephone number, and I have used temporary addresses and Google Voice to bypass this restriction. On occasion, the service will call with a verification code instead of sending a text message. Once your account is activated, you should see a $0.50 credit in the balance. This allows for approximately 150 free queries.

Once inside the dashboard, select "CNAM Delivery" and then "CNAM Query Tool" in the left panel. Enter your target phone number and retrieve the results. You will usually receive the carrier and name associated with cellular and landline numbers. Open CNAM is one of the leading caller

ID providers, and this page is designated for companies wanting to test the accuracy of their data. The following is an actual redacted result.

"name": "Bazel Michael",
"number": "+16180000000",
"price": 0.0033,
"uri": "/v3/phone/+16180000000"

Caller ID Service (calleridservice.com)

The previous websites provided the easiest search options, but they often demand a telephone number in order to complete registration. You might want to consider a more sophisticated approach which could be incorporated into your search tools. This will require a bit more work on your part, but it will be justified. Caller ID Service has provided good accuracy with cellular telephone searches. You must register for the service to gain access to a free trial, and the process is very easy. Navigate to this website and register for a free account. Upon completion, you will receive an email with an API license key that is valid for approximately 20 free successful searches. You will not be charged for empty results. You must validate this address by clicking the link included in the message. This is their way of verifying that you provided an accurate and real email address. The following information was sent to my account.

Username: jwilson555
Password: mb555555
Auth KEY: 0b253c059b9f26e588ab101f4c2332b496e5bf95
Balance : 0.12

You are now ready to submit requests for caller ID information. To do this, you must formulate an API request in your browser that includes your username, authentication key, and target telephone number to search. This is easier than it sounds. All we need is a number to search.

https://cnam.calleridservice.com/query?u=jwilson555&k=c2332b496e5bf95&n=6187271200

This queries the domain (calleridservice.com), our username (jwilson555), our authentication key (c2332b496e5bf95), and our target number (6187271200). The service confirmed that this cellular number belongs to John Williams. I recommend saving the address of your first query as a bookmark or favorite. However, you should leave off the target telephone number at the end of the address. This will prevent the service from charging you for a credit every time you load this template. You can then add the new target number at the end of the bookmark and conduct your searches easily. Caller ID Services grants you $0.12 in free searches, which will allow you up to 25 queries. Obtaining an additional free trial will only require a different email address. We will add this URL to our search tools later.

Bulk Solutions (bulkcnam.com)

Bulk CNAM works very similar to Caller ID Services. You must register for a free account, and you will be granted limited free searches. You must provide a valid email address during the registration and will be required to validate that address after you receive an email with your license key. The custom address (URL) that we need to create only requires your license key and the target number. A username is not necessary. The format of my free trial key is as follows.

https://cnam.bulkcnam.com/?id=b03c6513f688f89ee3f&did=6187271200

This queries the domain (bulkcnam.com), my free trial license (b03c6513f688f89ee3f), and my target telephone number (6187271200). A premium subscription to Bulk Solutions costs $0.009 per successful query. Below is an actual result after searching a VOIP number associated to me.

M BAZZELL

Open CNAM (opencnam.com)

Open CNAM was the second search option that was discussed earlier in this chapter. We used their dashboard to conduct queries within a traditional web page. We can also conduct our search within a direct URL. In your dashboard, you can view your "Account Authentication Details". This will display an account ID and access token. The following address contains the structure of a proper query.

http://api.opencnam.com/v2/phone/+16187271200?account_sid=f10&auth_token=AU5c43d8

This queries the domain (opencnam.com), the target number (16187271200), our account ID (f10), and our license number (AU5c43d8). Note that Open CNAM requires a "1" before the ten-digit number.

Everyone API (everyoneapi.com)

This service is owned by the same company as Open CNAM (Telephone Research LLC). The difference with this option is that it will display the current address, gender, carrier information, previous carrier, and subscriber name of the owner. This is a substantial upgrade from the previous options. It will also search the number in a social network database, but successes are limited. This premium service is a bit more expensive than the previous options, but a complete telephone number search can be obtained during a free trial.

Once you have created your free trial, you can use the dashboard to submit queries similar to Twilio and others. This is the simplest approach and allows you to monitor your balance in the upper right corner. The results are impressive. The following was displayed when searching my previously issued government cellular telephone number.

"address": "REDACTED HOME ADDRESS",
"carrier name": "AT&T Illinois"
"carrier name_o": "Verizon"
"cnam": "BAZZELL,MICHAEL",
"gender": "M",
"name": "Michael P Bazzell"
"number": "+16184620000",
"type": "person"

We can also submit this request within a URL in our browser. The format of the URL for the request is as follows. You would replace "6184620000" with your target number, "xxx" with your account SID available in your dashboard, and "yyy" with your license key. Note that Everyone API requires a "1" before the ten-digit number.

https://api.everyoneapi.com/v1/phone/+16184620000?account_sid=xxx&auth_token=yyy&data=name,address,location,cnam,carrier,carrier_o,gender,linetype,image,line_provider,profile

Service Objects (serviceobjects.com/products/phone/reverse-phone-lookup-service)

In 2014, Service Objects removed their free online telephone lookup demo titled GeoPhone Plus 2. However, they will still allow you to generate a free API license which will allow you 500 free searches. Navigate to this website and complete the "Free API Trial Key" offer. You will receive an email similar to the following.

This is your DOTS GeoPhone Plus 2 API Trial License Key: WS77-OAZ3-xXxX

You can now use this key within a custom URL to search the registered owners of landline and cellular telephone numbers. The exact format is the following. Note that you would change "8475551212" to the target telephone number and "WS77-OAZ3-xXxX" to your license key.

http://trial.serviceobjects.com/gppl2/api.svc/GetPhoneInfo?PhoneNumber=8475551212&TestType=full&LicenseKey=WS77-OAZ3-xXxX

The response will be in XML data format. However, it will be easy to read. Below is an example.

<Provider><Name>NEW CINGULAR WIRELESS PCS, LLC - IL</Name>
<LineType>WIRELESS</LineType>
<Name>JOHN ADORJAN</Name>
<Address>12142 S. 22nd<Address/>
<City>Chicago<City/>
<State>IL <State/>
<DateFirstSeen>2014-06-20</DateFirstSeen>

This entry identifies the target number as a wireless service provided by Cingular since June of 2014. The registered owner of the number is John Adorjan residing at 12142 S. 22nd in Chicago. Before reverse caller ID lookups, this information would have required a subpoena. In one recent scenario, a cellular telephone number searched on Service Objects revealed the name "Jennifer S" in the result. During an interview with this subject, she disclosed that "Jennifer S" is how she identifies her account on the telephone bill that she shares with other family members. She was unaware that this data was sent to the receiving number. On many searches, the full name will be present. This should explain why you may be noticing a caller's name on your caller ID display when he or she is calling from a cellular number.

Reverse Caller ID Summary

There are other caller ID options available on the internet and I encourage you to investigate any companies which have surfaced since this research. Most services offer a free trial, even if it is not advertised. Adding a few dollars to each service may provide more queries than you will ever need. Take advantage of the free trials to determine which services work best in your investigations. In 2019, I ranked the following in terms of most useful to least.

#1) **Twilio**: This was the most reliable, but contained minimal information (name only).
#2) **Everyone API**: This displayed the most information, but some queries returned nothing.
#3) **Open CNAM**: The results here were very similar to Twilio.
#4) **Bulk Solutions**: I received outdated information on several queries.
#5) **Caller ID Service**: This service experienced numerous outages and empty results.
#6) **Service Objects**: These results appeared to be from a White Pages database.

In 2019, I purchased premium memberships through EveryoneAPI and Open CNAM. A $10 purchase provided over 1,000 queries at each provider. In a moment, I will explain how you can modify the search tools to query all of these services at once when you provide a single telephone number. The results will all display immediately for you. I use my own custom page every day. Overall, reverse caller ID services can tell us more about a target telephone number than the standard people search engines. In many cases, you can immediately obtain data that would have required a subpoena just a few years prior. Always utilize all of the services in order to gauge the confidence in the results. If this is overkill for your needs, there are other web-based search engines that are easier to use.

Caller ID Test (calleridtest.com)

This site was designed to input a telephone number and test the Caller ID display feature. It is nothing more than a standard lookup service, but I have found the data to be unique from other sources on some occasions. Unfortunately, I have also found the availability of this service to be completely unreliable. While the site is usually present, the results don't always populate. However, this resource should be checked as a last resort when the other processes have failed.

Telephone Search Websites

In previous editions of this book, I summarized a handful of people search websites which allowed the query of a telephone number. These sites all possess unique data sets and each should be searched. Most of the results originate from sources such as property tax data, marketing leaks, phonebooks, and various data breaches. Instead of explaining each site, which becomes quite redundant, I will display the static URL of a search submission. Many of these links avoid unnecessary loading screens and advertisements. This will help us with the automated tool at the end. Overall, we cannot control the results, and telephone search is mostly "what you see is what you get". Replace the demo number (618-462-0000) with your target number.

411 https://www.411.com/phone/1-618-462-0000
800 Notes https://800notes.com/Phone.aspx/1-618-462-0000
Advanced Background Checks https://www.advancedbackgroundchecks.com/618-462-0000
America Phonebook http://www.americaphonebook.com/reverse.php?number=6184620000
Caller Smart https://www.callersmart.com/phone-number/618-462-0000
Dehashed https://dehashed.com/search?query=6184620000
Fast People Search https://www.fastpeoplesearch.com/618-462-0000
Info Tracer https://infotracer.com/phone-lookup/results/?phone=6184620000
John Doe https://johndoe.com/phones/6184620000
Numpi https://numpi.com/phone-info/6184620000
Nuwber https://nuwber.com/search/phone?phone=6184620000
OK Caller https://www.okcaller.com/6184620000
People Search Now https://www.peoplesearchnow.com/phone/618-462-0000
Phone Owner https://phoneowner.com/phone/6184620000
Reverse Lookup https://www.reverse-lookup.co/618-462-0000
Search Bug searchbug.com/tools/reverse-phone-lookup.aspx?TYPE=phonerev&FULLPHONE=6184620000
Search People Free https://www.searchpeoplefree.com/phone-lookup/618-462-0000
Spytox https://www.spytox.com/reverse-phone-lookup/618-462-0000
Sync.me https://sync.me/search/?number=16184620000
That's Them https://thatsthem.com/phone/618-462-0000
True People Search https://www.truepeoplesearch.com/results?phoneno=(618)462-0000
US Phonebook https://www.usphonebook.com/618-462-0000
WeLeakInfo https://weleakinfo.com/search?type=phone&query=6184620000
White Pages https://www.whitepages.com/phone/1-618-462-0000
WhoCalld https://whocalld.com/+16184620000
Yellow Pages https://people.yellowpages.com/whitepages/phone-lookup?phone=6184620000
Zabasearch https://www.zabasearch.com/phone/6184620000

Google https://www.google.com/search?q=618-462-0000
Bing https://www.bing.com/search?q=618-462-0000
Yandex https://yandex.com/search/?text=618-462-0000

Search Engines

Google and Bing were once a great place to find basic information about a target phone number. These sites can still provide valuable information, but the amount of spam that will display in the results is overwhelming. Many of the links presented will link to sites that will charge a fee for any information associated. This information is usually the same content that could have been located with an appropriate free search. I do not recommend giving in to these traps. While we can't ignore a traditional search of telephone numbers, we can customize the queries in order to achieve the best results. Before explaining advanced telephone owner identification, we should take a look at appropriate search engine structure.

Most people use traditional search engines as a first step toward identifying the owner of a telephone number. The number is usually provided in a standard format such as 202-555-1212. This can confuse some search engines because a hyphen (-) is often recognized as an operator to exclude data. Some engines might view that query as a search for 202 but not 555 or 1212. Additionally, this search might identify a website that possesses 202-555-1212 within the content but not one that contains (202) 555.1212. If this is your target number, all of the following should be searched in order to exhaust all possibilities. The quotation marks are important to prevent the hyphen from being seen as an operator.

"2025551212"
"202-555-1212"
"202.555.1212"
"202 555 1212"
"(202) 5551212"
"(202) 555-1212"
"(202) 555.1212"
"(202)5551212"
"(202)555-1212"
"(202)555.1212"

This may seem ridiculous, but I am not done. Many websites forbid users to post a telephone number, such as many auction sites, but people try to trick this restriction. They will type out a portion of their number to disclose contact information. While not a complete list of options, the following should also be searched.

"two zero two five five five one two one two"
"two zero two five five five 1212"
"two zero two 555 one two one two"
"two zero two 555 1212"
"202 five five five one two one two"
"202 555 one two one two"
"202 five five five 1212"

This list would not capture a post that included (202) 555 twelve twelve, but you get the point. After submitting these through Google, you should attempt each through Bing. In my effort to always provide search tools which automate and simplify these techniques, I have added this feature to your telephone tools presented at the end of the chapter. The lower portion of this page, displayed later in Figure 17.03, allows you to enter a numerical and written target telephone number. Clicking the submit button launches a series of JavaScript commands that launch eight new tabs within your browser. The first four are custom Google searches with the target data and

the last four repeat the process on Bing. The following four searches are conducted on both services, using the example data entered previously.

"2025551212"OR"202-555-1212"OR"202.555.1212"OR"202 555 1212"

"(202) 5551212"OR"(202) 555-1212"OR"(202) 555.1212"OR"(202)5551212"OR"(202)555-1212"OR"(202)555.1212"

"two zero two five five five one two one two"OR"two zero two five five five 1212"OR"two zero two 555 one two one two"OR"two zero two 555 1212"

"202 five five five one two one two"OR"202 five five five one two one two"OR"202 five five five 1212"

Notice that these queries use quotation marks to obtain exact results and the OR operator to search multiple options independently from each other. You will likely receive many false positives with this method, but you are less likely to miss any relevant results. While this is a great starting point for number searches, it is much less reliable than the next method.

True Caller (truecaller.com)

This service stands alone as the most creative telephone number lookup service. True Caller is an app for smart devices that displays caller ID information of incoming calls. If you receive a call on your phone, and the number is not in your contacts, True Caller searches its database and provides any results on your screen. You can then choose to accept or deny the call. This is fairly standard and is not the interesting aspect of this service. The fascinating part to me is the source of their caller database. It is completely crowd-sourced.

When you install the app, you give it permission to collect all of your contacts and upload them to the master database. Basically, millions of users have uploaded their contact lists for the world to see. The next amazing thing to me is the ability to search within this data on the True Caller website. You must connect to the service via a covert Facebook or Google account, but that is not difficult. When I first found this service, I was skeptical. I entered the cellular number of my government issued cellular telephone expecting to see no results. The response was "Mike Bazell". My jaw dropped. My super-secret number was visible to the world. This means that someone in my circle, likely another government employee, installed True Caller on his or her phone and had my information in their contacts. Until someone else populates data for this number, it will always be present in the database.

Real World Application: During my final year of government investigations, I queried a target number associated with a homicide through this service. The result was "Drug Dealer Matt". One of the target's customers must have installed True Caller. One of three potential suspects was named Matt, earned our spotlight, and later an arrest.

Old Phone Book (oldphonebook.com)

I first noticed this niche service in late 2018. It provides historical White Pages landline listings from 1994-2014. The sources are official databases collected from many years of telephone CD-ROMs. These were purchased by various companies throughout several decades as a more convenient option than traditional phone books. The data is quite impressive, and the following direct URLs allow us to add these to our tools. Results include historic addresses attached to each year. Figure 17.01 displays an actual redacted result from the official website. This provides an old address, and assumes that the target moved to a new address between 1998 and 2001. This search is vital for background checks.

http://www.oldphonebook.com/searchphone2.php?syear=1994&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=1995&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=1996&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=1997&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=1998&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2001&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2002&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2003&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2007&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2008&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2013&sphone=6184620000
http://www.oldphonebook.com/searchphone2.php?syear=2014&sphone=6184620000

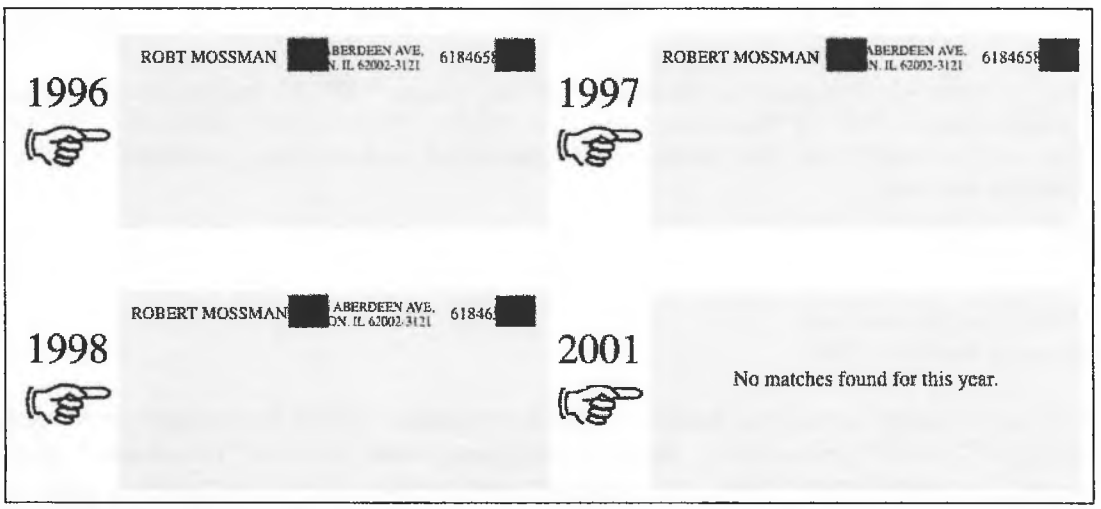


Figure 17.01: A partial redacted result from Old Phone Book.

Craigslist (craigslist.org)

Craigslist has already been discussed in earlier chapters, but the phone search options should be further detailed. Many people use Craigslist to sell items or services. The posts that announce the item or service available will often include a telephone number. These numbers will belong to a landline or cellular provider. This can be a great way to identify unknown telephone numbers. Some posts on Craigslist will not allow a telephone number to be displayed on a post. It is a violation of the rules on certain types of posts. Some people choose not to list a number because of automated "scrapers" that will grab the number and add it to databases to receive spam via text messages. Either way, the solution that most users apply to bypass this hindrance is to spell out the phone number. Instead of typing "314-555-1212", the user may enter "three one four five five five one two one two". Some will get creative and post "314 five five five 1212". This is enough to confuse both Craigslist's servers as well as the spammers. This can make searching difficult for an analyst. The hard way to do this is to conduct several searches similar to the following.

site:craigslist.org "314-555-1212"
site:craigslist.org "314" "555" "1212"
site:craigslist.org "three one four" "five five five" "one two one two"
site:craigslist.org "314" "five five five" "1212

This list can get quite long if you try to search every possible search format. One search that will cover most of these searches in a single search attempt would look like the following.

site:craigslist.org "314"|"three one four" "555"|"five five five" "1212"|"one two one two"

The "|" symbol in this search is the same as telling Google "OR". In essence, we are telling Google to search "314" or "three one four", then "555" or "five five five", and then "1212" or "one two one two". With this search, you would receive a result if any combination of the following was used.

314-555-1212
3145551212
314 555 one two one two
three one four 555-1212

This search will not catch every possible way to post a phone number. For example, if the user had typed "314 555 twelve twelve", the above technique would not work. The researcher must consider the alternative ways that a target will post a number on a website. It may help to imagine how you would post the target number creatively on a site, and then search for that method. Additionally, searching for only a portion of the number may provide results. You may want to try searching only the last four digits of the number. This may produce many unwanted results, but your target may be within the haystack. An automated option is included in the search tools.

Escort Index (escortindex.com)

If you have any suspicion that the target of your investigation is involved in prostitution, drugs, or any related activity, Escort Index should be checked against the telephone number of your subject. This website aggregates all of the prostitution classifieds and review sites into one search. It extracts the telephone numbers from all online classified pages and allows you to search by the target telephone number. One of my training examples identified 37 online photos, 20 escort ads, several reviews by "Johns", ages used by the target, the last known location, and locations visited based on postings in online classifieds. Any time I have a target telephone number that is likely involved in criminal activity, I conduct a brief search on this site.

https://escortindex.com/search?search=754-240-1522

Spy Dialer (spydialer.com)

This service takes a new approach on identifying the user of a cellular telephone number. While it offers a typical telephone number search tool, which appears to extract data from crowd-sourced databases, the real power lies in the voicemail retrieval. Most cellular users have an outgoing voicemail message that identifies them by name. Others create a custom message with their own voice. Any of this can help determine the user of the number. Spy Dialer attempts to connect to the service provider of the cellular number; extract the outgoing voicemail message; and present it to you in mp3 format for listening and downloading. All of this is normally completed without ringing the target's telephone. You should note that on some occasions, the target telephone number rang during testing. It did not identify the caller, but the unusual single ring may raise suspicion with a paranoid target. In my experience, this happens about 10 percent of the time. If the target dials the Nevada-based number of the missed call, he or she will be notified that a Spy Dialer call was placed. My successes with this method outweigh the risk. I have had several investigations that involved "anonymous" cash cellular telephones that announced the owner's name on the outgoing message. When a successful result is displayed, you can click the link below the player to download the audio file to your computer.

Sly Dial (slydial.com)

This service conducts an inquiry into a cellular telephone number the same way as Spy Dialer's attempt. It contacts the cellular provider of the telephone number and sends you straight to the outgoing voicemail message of the target. However, there are two big differences.

Sly Dial does not work through a website. Instead, you must call a general Sly Dial telephone number and follow the automated prompts. You must listen to a brief advertisement before your call is placed. Finally, the service will play the target's outgoing voicemail message through this audible telephone call. Since a website is not involved, there is no option to download an audio file of the call. We can obtain an audio copy of this message by placing the call through Google Voice and recording the session by pressing the "4" button on the dial pad.

Sly Dial does not usually ring the suspect's telephone. It will likely not show "missed call" or any other indicator that a call occurred. In my testing, less that 5 percent of the attempts actually cause the target telephone to ring only one time. Calling the missed call back reveals nothing about the identity of the number. Ultimately, there is a very small chance that the target will know that someone attempted a call. In the rare occurrence that the telephone rings, the target will never know the identity of the person making the calls. To use the Sly Dial service, call 267-759-3425 (267-SLYDIAL) from any telephone service including landlines, cellular lines, or VOIP. Follow the directions during the call. If this number does not work, visit slydial.com for updates.

I want to stress the following one additional time. Use these services at your own risk. If accidentally notifying your target that you are conducting these types of activities could compromise your investigation, avoid these two techniques.

Grocery Reward Cards / Loyalty Cards

Most grocery chains have adopted a reward/loyalty card system that mandates the participant to enroll in their program. The consumer completes an application and receives a plastic card to use during checkout for discounts. Many of these stores only offer a sale price if you are a member in the program. Most consumers provide a cellular telephone number to the program and use that number during checkout. This eliminates the need of possessing a physical card in order to receive the discount. Instead, they type their cell number into the card swiping machine to associate the purchase with their membership. These programs contain a huge database of telephone numbers and the registered users. There is no online database to access this data. However, you can obtain this data if you are creative.

Assume that your target telephone number is 847-867-5309. If you have tried every technique mentioned at this point to identify the owner and failed, you may consider a query with a local grocery chain. The easiest method is to enter the store, purchase a pack of gum, and enter the target telephone number as the reward/loyalty program number. You will likely receive a receipt with the target's name on the bottom. Figure 17.02 (left) displays a portion of the actual receipt that I received when using this number. If you prefer to avoid entering a store, drive to the company's gas station outside of the store. Figure 17.02 (right) displays the notification I received when entering this same number at the pump. Note that this number is fictional. However, it has been registered at practically every grocery store in the United States. Try to use it the next time you make a purchase.

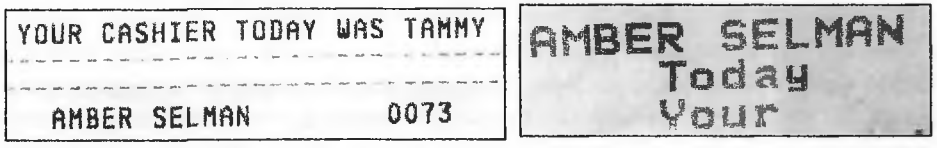


Figure 17.02: A receipt (left) and gas pump (right) identifying the owner of a cell number.

Contact Exploitation

I mentioned this technique previously during the Android emulation chapter. If I receive no valuable information about a target telephone number with the previous options, I proceed to my Android emulator. I add the number as a friend named "John" in the contacts of the device. I then open any applicable application, such as Twitter, Instagram, and dating apps, and ask them to "find my friends". More often than not, an app links me to a social network profile which is associated with the only number in my contact list.

Real World Application: In 2019, I was asked to assist an attorney in a child custody and divorce case. He provided the cellular telephone numbers of the suspected cheating husband and an unknown number of a potential mistress identified from telephone records. I added both numbers to the contacts list in my Android emulator and checked every app daily. This method identified a dating profile which included the husband's photo, but with a different name. On the same day which he was served divorce papers, I observed both numbers show up in the encrypted communications app Signal. While it did not identify the person on the other end of the communication, the coincidence of the two numbers joining Signal on the same day is revealing.

IntelTechniques Telephone Search Tool

Of all the automated search tools, this one may save the most time. You can submit an area code, prefix, and number to each of the options and execute a single search. This can become quite time consuming, so I encourage users to take advantage of the "Submit All" option near the end. If you have enabled pop-ups for this page, it will execute all queries above the button in new tabs. The Caller ID area automates the execution of the services mentioned at the beginning of the chapter. In order for these to function, you will need to request a free trial or paid account from each service. You will then need to open the Telephone.html file within a text editor and make a few modifications. I will demonstrate the EveryoneAPI option as an example. Inside the file, you will see the following line within the Caller ID section.

```
{window.open('https://api.everyoneapi.com/v1/phone/+1' + CID01 +
'?account_sid=USERNAME&auth_token=TOKEN&data=name,address,location,cnam,carrie
r,carrier_o,gender,linetype,image,line_provider,profile' + CID01, 'CID01window');}
```

Replace the USERNAME option with the ID provided by the service and replace TOKEN with the token issued to your account. Doing so eliminates the need to log in to these services each time you wish to conduct a query.

The Number & Words area on the search tool replicates the techniques mentioned previously which attempt to search both the numbers and words of a target telephone number within Google, Bing, and Craigslist. After entering the numbers, the words automatically populate. Figure 17.03 displays the current state of the tool. The right column offers a complete replication of the Old Phone Book service. This takes advantage of the direct URLs and Iframes, which

allow us to see the results within the tool. This tool possesses the most detailed (and lengthy) code behind the HTML file. This is due to a requirement for several brackets of numbers. Since each site possesses a unique URL structure, we need the isolation between area code, prefix, etc. However, the underlying principles are the same, and you can edit this tool as needed in the same way as the previous search tool files.

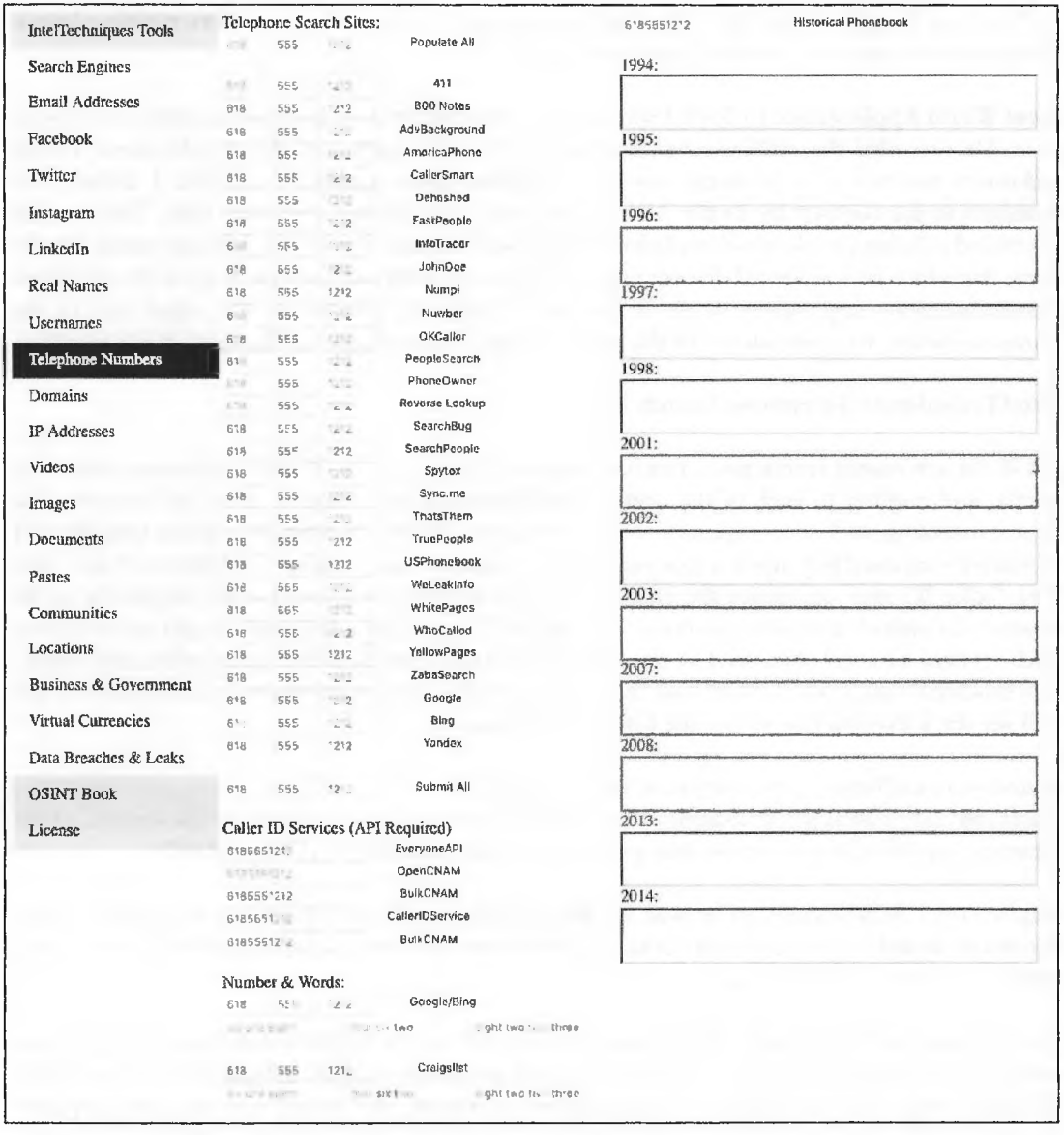


Figure 17.03: The IntelTechniques Custom Telephone Search Tool.

CHAPTER EIGHTEEN
ONLINE MAPS

The presence of online satellite images is not news anymore. Most of you have already "Googled" your own address and viewed your home from the sky. This view can get surprisingly detailed when using the zoom feature. Alleys, sheds, and extended driveways that are hidden from the street are now visible thanks to this free service. Many tactical units will examine this data before executing a search warrant at a residence. Aerial maps are helpful for viewing the location of exiting doors, escape routes, stairs, and various obstructions. The rapidly growing availability of the Street View option now gives us more data. This chapter explains detailed use of Google, Bing, and other mapping services. At the end, I present my custom maps tool which provides an automated solution to collecting every possible view associated with your target of interest.

Google Maps (maps.google.com)

In 2014, Google made several changes to their online maps service. They introduced a new feature with Street View that made the default view full screen. This eliminates the Google search bar, side menu, browser menus, and any other items from blocking a larger view. Additionally, Google streamlined the entire Maps experience to make everything easier to use. Unfortunately, they also eliminated many of the features that were beneficial to researchers and investigators. Fortunately, they have re-enabled some of these missing features. The following basics of Google Maps are now default for all users.

Search Bar: The Google Maps search bar can now accept practically any type of input. A full address, partial address, or GPS coordinates will immediately present you with a mapped view. Company names and types of businesses, such as "café" will highlight locations that may be of interest. This search field is the first stop. Attempt any search relevant to your investigation and you may be surprised at how accurate Google is. You can collapse this entire menu by clicking the left arrow next to the search field. This will return you to a full screen view of the map.

Satellite/Earth View: The lower left area of any map will offer a satellite view. The satellite view is a direct view from the sky looking almost straight down. The 3D view in the lower right is similar, but offers the tilt option. While in the 3D view, click on the small "rotation" icon to the right of the map. This will shift the view 45 degrees and a second click will shift an additional 45 degrees. While satellite views of maps are now well-known, we see continuous enhancements that are not advertised. A satellite view of your target location is always vital to every investigation.

Street View: If the Street View option is available, Google has been in the area and captured a photo of the location from the street. Dragging and dropping the small orange man in the lower right menu will open a street view from ground level at the area specified. You can navigate

through this view by clicking forward, clicking and dragging, or scrolling and zooming. This view can be zoomed in by double-clicking and panned left and right by dragging the mouse while holding a left click. Double-clicking an area of the street will refresh the window to the view from that location. Clicking "Back to map" in the lower left will return you to the standard map view.

Historic Street View: In late 2014, Google began offering the ability to view all stored street view images for any single location. This option is available within the standard street view layout within the search area of the upper left corner. Click on the small clock in order to launch a pop-up window. This new view will allow you to move a slider bar which will present different views. The month and year of image capture will also appear for documentation. Figure 18.01 displays this method which presents an additional view of a parking lot from a few years prior. Additional options include views from several years prior. This can often reveal additional vehicles or missing structures associated with an investigation.

Distance Measurement: Google Maps reintroduced the distance measurement tool after completely disabling the classic maps interface in 2015. While in map or satellite view, right-click on your starting point and choose "Measure distance". Click anywhere on the map to create a path you want to measure. Further clicks add additional measuring points. You can also drag a point to move it, or click a point to remove it. The total distance in both miles or kilometers will appear under the search box. When finished, right-click on the map and select clear measurement.

GPS Coordinates: Clicking on any point will load a small window in the bottom center that identifies the exact GPS coordinates of the chosen location. If this is not visible, right-click any point and select "What's here". For the purposes of our search tool, a static URL displaying a satellite view is as follows. In this example, 41.9483333 is the latitude coordinate and -87.6595718 is the longitude.

https://maps.google.com/maps?t=k&q=loc:41.9483333+-87.6595718



Figure 18.01: Historic Street View options from Google Maps.

Bing Maps (bing.com/maps)

Similar to Google Maps, Bing offers a map view, satellite view, and street view. Bing does offer something that is not always available in Google. Bing possesses a "Bird's Eye View" which displays four distinct angled views of a location. This may display signs, advertisements, pedestrians, and other objects with clear visibility. While Google is rolling out their own 45-degree satellite view, the areas covered are minimal at the time of this writing. In my experience, the imagery provided by Bing is often superior in quality compared to Google Maps. A side by side comparison can be seen in a few pages with the custom maps tool. By default, this view will always be of the south side of a location, looking north. The curved arrows in the upper right corner allow you to navigate to three additional views which display the west, north, and east sides of the location. The Bing static URL with satellite view is as follows. The final four URLs represent the 3D satellite view from each direction (N,E,S,W), as determined by the rotation number at the end of each URL (0,90,180,270). We will load all of these with the custom tool at the end of the chapter.

https://www.bing.com/maps?&cp=41.9423~-87.6533&lvl=20&sty=a
https://www.bing.com/maps?&cp=41.9423~-87.6533&lvl=20&sty=o&w=100%&dir=0
https://www.bing.com/maps?&cp=41.9423~-87.6533&lvl=20&sty=o&w=100%&dir=90
https://www.bing.com/maps?&cp=41.9423~-87.6533&lvl=20&sty=o&w=100%&dir=180
https://www.bing.com/maps?&cp=41.9423~-87.6533&lvl=20&sty=o&w=100%&dir=270

Zoom Earth (zoomearth.com)

This multiple satellite imagery website presents views from NASA, Bing, and ArcGIS. Occasionally, the ArcGIS data is more recent than Google or Bing. The smooth interface will easily provide a comparison of the available images for any location. One advantage of Zoom Earth is the ability to view satellite images in true full-screen mode. This allows creation of full-screen captures without branding, menus, or borders. This could be more appropriate for live demonstration instead of a standard Google or Bing window. The static URL with satellite view is as follows. The "20z" is the maximum zoom level.

https://zoom.earth/#view=41.9423,-87.6533,20z

Here (wego.here.com)

This service has experienced a lot of transformation over the past ten years. Originally a Nokia company with a unique set of satellite imagery, it now mostly contains identical images to Bing Maps. However, I find unique images on occasion, and we should always consider every resource. The static URL with satellite view is as follows.

https://wego.here.com/?map=41.9423,-87.6533,19,satellite&x=ep

Yandex (yandex.com/maps)

Satellite views from Yandex are often poor quality and low resolution when focused on the U.S. However, international locations tend to possess clearer results. The static URL with satellite view is as follows.

https://yandex.com/maps/?l=sat&ll=-115.222540%2C36.233776&z=10

Descartes Labs (maps.descarteslabs.com)

This service offers a very unique search option that I have not found present on any other site. After locating a target of interest, it displays a satellite view sourced from the National Agriculture Imagery Program (NAIP). The unique part is the ability to search based on image. In other words, you can select a monument, park, building, or any other view and request any images that match. As an example, I selected the baseball grounds at Wrigley Field and I was immediately presented hundreds of baseball fields all over the world. I have yet to determine how I would execute this strategy within an investigation, but this feature has potential.

https://search.descarteslabs.com/?layer=naip_v2_rgb_2014-2015#lat=41.9423000&lng=-87.6533000&skipTut=true&zoom=17

Land Viewer (eos.com/landviewer)

This resource will not present detailed views of your target's home. The images here are often generated from weather satellites, and restrict zoom levels to a city view. Most locations offer four active satellites that constantly retrieve images and five inoperative satellites that store historic imagery dating back to 1982. I have only used this resource to document potential weather at a crime scene (clouds, rain, snow, or clear). The static URL is as follows.

https://eos.com/landviewer/?lat=40.38735&lng=-107.18468&z=11

Landsat Look (landsatlook.usgs.gov)

This is very similar to Land Viewer, but with additional historic images. There are three unique maps which can be viewed within a browser. The following URLs present each option.

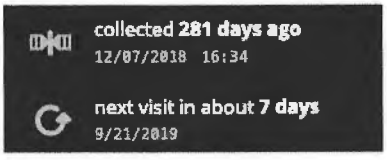
https://landlook.usgs.gov/viewer.html
https://landsatlook.usgs.gov/viewer.html
https://landsatlook.usgs.gov/sentinel2/viewer.html

Each map possesses unique satellite views ranging from 1972 to present. The menu in the upper right of each map displays the image options. The instructions for accessing this data is not straightforward, so I will provide a typical scenario.

- Drag and zoom to your target location.
- Select your desired satellite in the upper-right menu.
- Choose desired date range.
- Click the "Show Images" button.

Map Box (labs.mapbox.com/labs/terra)

This service also displays historic land images which do not provide many details. The benefit of Map Box is the notification of the next capture date. While researching a target area, the service displayed the image below. This identifies the date of the current image and an announcement of the next time the satellite will be over the target area. The static URL is below the image.



https://labs.mapbox.com/bites/00145/#12/41.9423/-87.6533

Wikimapia (wikimapia.org)

This may be my favorite option. It combines the various satellite options for Google, Bing, and others. It also possesses crowd-sourced data and images. The menu in the upper right allows you to choose the satellite service while clicking on any location or building presents a report of images, websites, and details about the target. The static URL is as follows.

http://wikimapia.org/#lang=en&lat=41.940894&lon=-87.649441&z=18&m=w

Zillow (zillow.com)

This website displays information about homes which are currently available for sale or have recently sold. I usually search this option in hopes of locating interior images of a suspect home. This can often identify personal belongings, interests, and family information. It is a bit of a long-shot, but has been vital to a few of my investigations. The static URL is as follows.

https://www.zillow.com/homes/for_sale/globalrelevanceex_sort/41.9423,,-87.6533

Snapchat (map.snapchat.com)

This was mentioned previously while discussing social networks, but should also be present here. This map allows you to zoom into a location and view any public stories posted through

Snapchat. I find this does not work well when investigating a home address, but can be incredibly useful when researching a specific event, school, or public area. The static URL is as follows.

https://map.snapchat.com/@40.088928,-81.754226,15z

Dual Maps (data.mashedworld.com/dualmaps/map.htm)

This website provides a satellite view of a location on both Google Maps and Bing Maps simultaneously. The Google view on the left will also contain a search field in the lower left corner. Searching an address in this field will center both maps on the same address. This will provide a comparison of the satellite images stored on each service. This can quickly identify the service that has the better imagery. While this can provide a quick side-by-side comparison, upcoming automated solutions are preferred.

Historic Imagery

Researching different satellite views of a single location can have many benefits. These views are all of the current content stored within each service. However, these mapping services continuously update their offerings and usually present the most recent option. You may want to view the previous content that was available before an image was updated.

Historic Aerials (historicaerials.com)

If you need satellite imagery from several years prior, you can visit Historic Aerials. The quality will often be poor, especially as you view imagery from previous decades. After you enter an address, you will be presented all available options on the left side of the page. Figure 18.02 displays several results of the same location over a twenty-year period. These views will be unique from all of the previously mentioned services.

World Imagery Wayback (livingatlas.arcgis.com/wayback)

I began using this site in 2018 as an alternative to Historic Aerials. The quality is superior, the number of satellite views is higher, and the functionality is smoother. Figure 18.03 displays four images acquired from the service of a specific location. Each provides a unique view from 2014, 2015, 2016, and 2017.

Real World Application: Combining several satellite views can provide much information about a target's residence. Before the execution of a search warrant, it is beneficial for police to collect as much map information as possible. This will give updated views of a drawn map, satellite imagery directly above the house, four angled views from the sky, and a complete view of the house and neighboring houses from the street, including vehicles. This can be used to identify potential threats such as physical barriers, escape routes, and video surveillance systems in place.

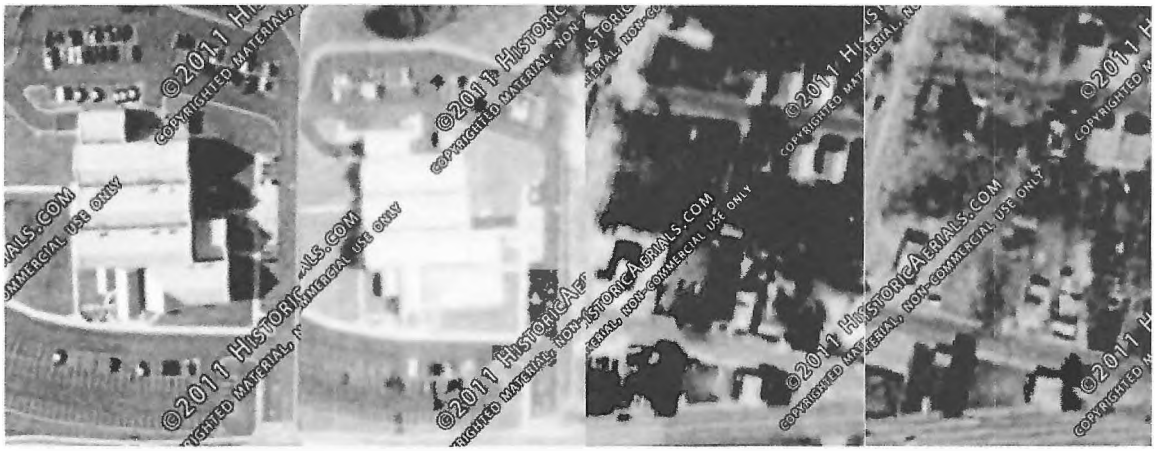


Figure 18.02: Multiple views of a location through Historic Aerials.

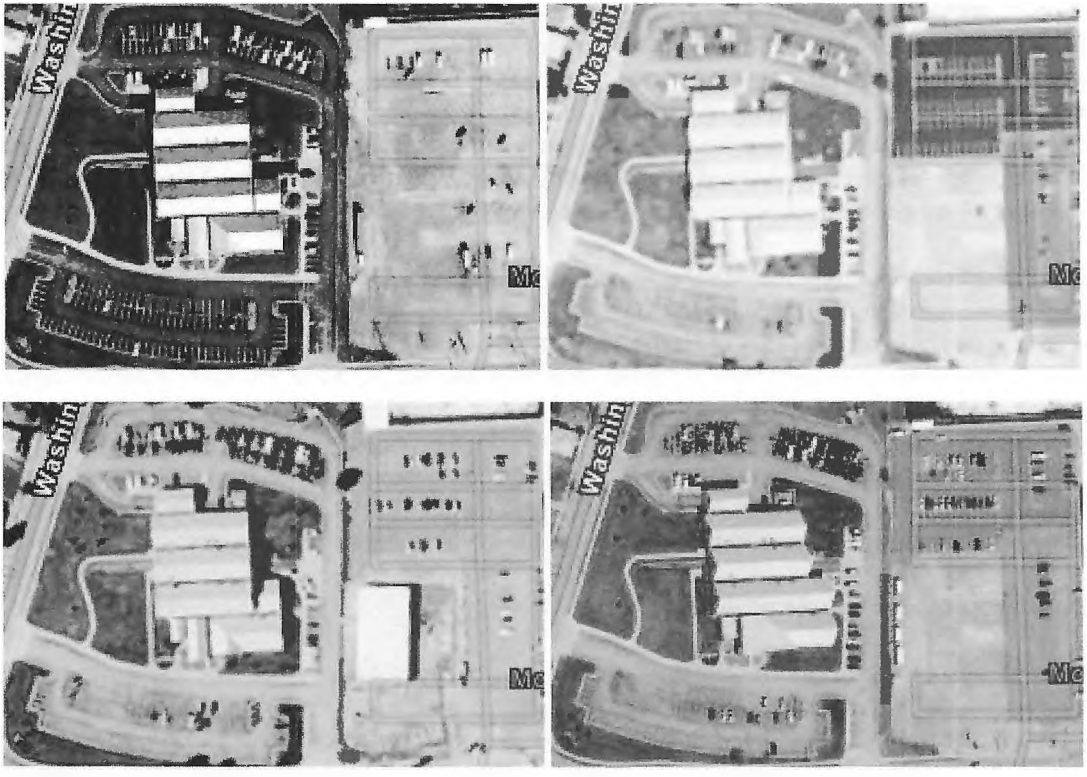


Figure 18.03: Historic satellite images from World Imagery Wayback.

Crowd-Sourced Street Views

Satellite and Street View maps from services such as Google and Bing are nothing new. Most of you can view the top and front of your home from multiple online websites. With street view options, these services are fairly responsible and block most faces and license plates. This makes it difficult for investigators trying to identify a suspect vehicle parked at a home or present at a crime scene prior to the incident. We have two services that offer unique street-level views that may remove these limitations.

Mapillary (mapillary.com)

This service appears similar to other mapping websites when first loaded. You see a typical map view identifying streets, landmarks, and buildings. Enabling the satellite view layer displays images from the Open Street Map project. The real power is within the crowd-sourced street view images. While in any map view, colored lines indicate that an individual has provided street-level images to Mapillary, usually from a GPS-enabled smart phone. This is actually quite common, as many people record video of their driving trips which could be used in case of an accident. These services make it easy and automated to upload these images. The sites then embed these images within their own mapping layers for the public to see. Clicking these colored lines reveals the street view images in the lower portion of the screen. Expanding these images allows you to navigate through that individual's images similar to the Google Street View experience. Figure 18.04 displays a street view layered over a satellite view. The username of the Mapillary member and date of image capture appears in the lower left.

In some of these images, the services appear to be redacting license plates with a typical "Blur Box" as seen in Figure 18.05 (left). A few feet later, the box disappears and a partially legible license plate is revealed, as seen in Figure 18.05 (right). It seems like these services are attempting to determine when a license plate is legible, and then blurring it. When the plate is farther away and more difficult to read, it is ignored. This can work in our favor. In Figure 18.06 we can use selective cropping and photo manipulation to obtain the registration. The left image appears unaltered as it is difficult to read. The right image was cropped; inverted with Photoshop; brightness turned down to 0%; and contrast heightened to 100%. The result is a legible license plate. I believe most registration plates can be made visible within these services.

Open Street Cam (openstreetcam.org)

After exhausting your options on Mapillary, continue to Open Street Cam. The site functions in the same way, but possesses unique images. These sites allow you to identify the uploader's username, mapping history, number of posted images, and profile image. You can also select to watch all of the captured images from a specific user as he or she travels daily. I can't begin to imagine the amount information available about a user's travel habits if he or she were to become a target of an investigation. While there is not coverage of every area like we see with Google Maps, the databases are growing rapidly, and should be included when using other mapping tools.



Figure 18.04: A crowd-sourced street view from Mapillary.



Figure 18.05: A vehicle with a blurred registration plate (left) and clear (right).



Figure 18.06: An illegible registration plate (left) and manipulated view (right).

IntelTechniques Maps Search Tool

This custom search tool has two independent portions. The first option allows entry of a physical address in traditional format which can execute several searches previously mentioned. The first field presents the Google Maps API page that includes the GPS coordinates of your target. These will be used in the second portion of the page. The Zillow Homes, Rehold Homes, and Google Homes options conduct searches attempting to identify any interior images from house sale websites. These have been very valuable during briefings before a search warrant execution.

The second portion allows entry of GPS coordinates for several mapping options. The "Populate All" simplifies entry into the query options. Each search will open results in a new tab, or the final "Submit All" will display all satellite imagery from multiple providers within new tabs. It currently fetches images from options mentioned here which allow a static URL with GPS for submission. Figure 18.07 displays the current view of the tool. Figures 18.08 through 18.23 on the following pages display the results from these providers when searching Wrigley Field in Chicago, using the search tool to generate each result.

I use this tool during practically every investigation. Whether I have a residential address of a target or a business address of the scene of an incident, these satellite views have relevance. Having fourteen unique views of an individual location would have seemed unimaginable in decades past. Today, we take it for granted. I believe that capturing these images while present is important. These services disappear for unknown reasons or images are overwritten with new content. Archiving these views during your research preserves the evidence.

Classic Maps (gokml.net/maps#)

In previous editions of this book, I explained the ways that users could access the original Google Maps, often referred to as the Classic Maps. There were a handful of URLs that still presented Google's classic view with additional tools. Google has officially disabled all of these hidden sites, but there is still one option left. Gokml.net hosts an interactive map tool that appears very similar to the classic version of Google Maps. It uses Google's live stream of mapping data, but displays options no longer available in the current Google Maps. Most of these are related to the layout of the page and allow you to customize the view. It allows you to split the view horizontally and vertically, with Street View on one side and Satellite View on the other. While this does not present any new content, it does provide additional view options.

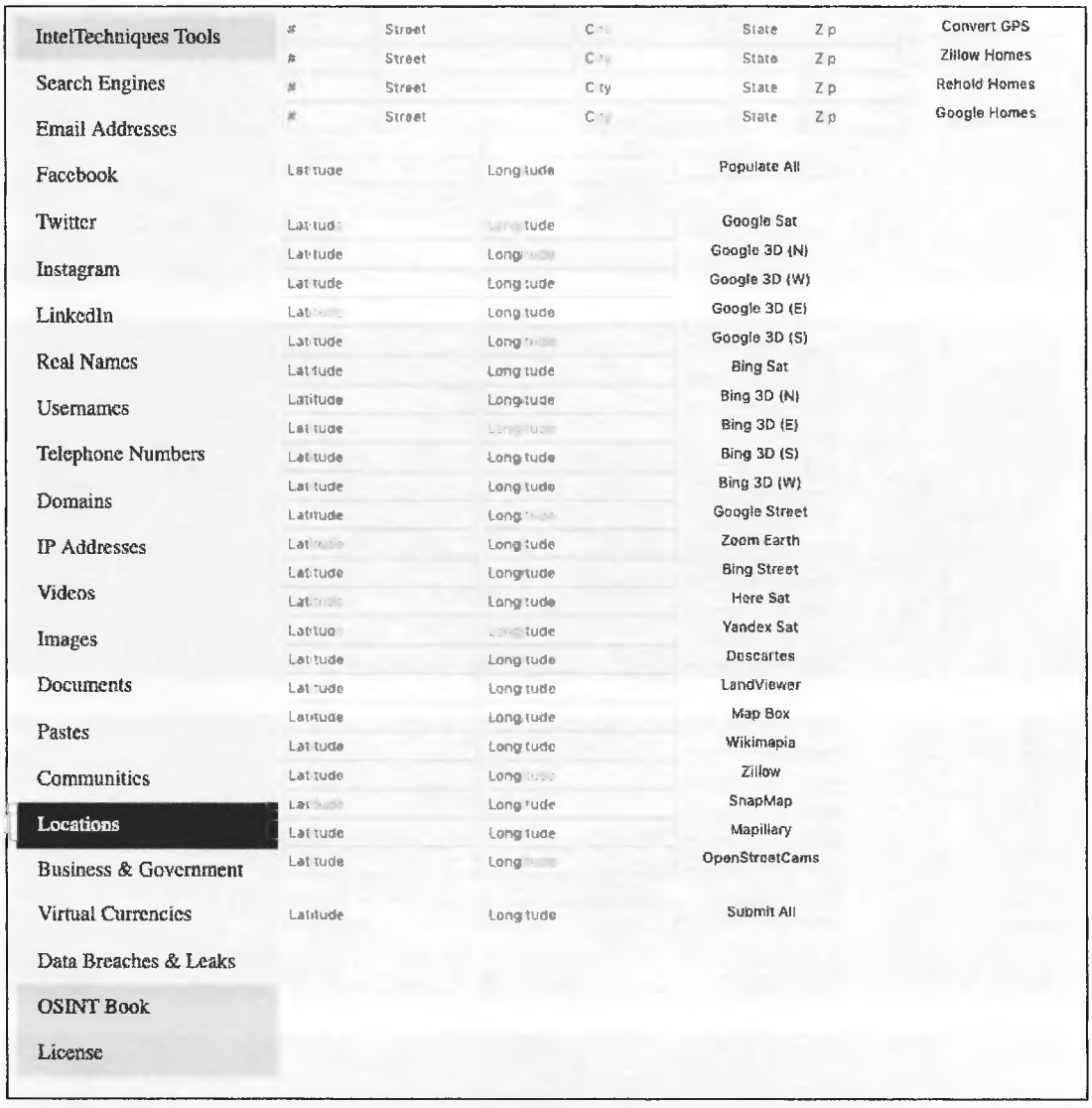
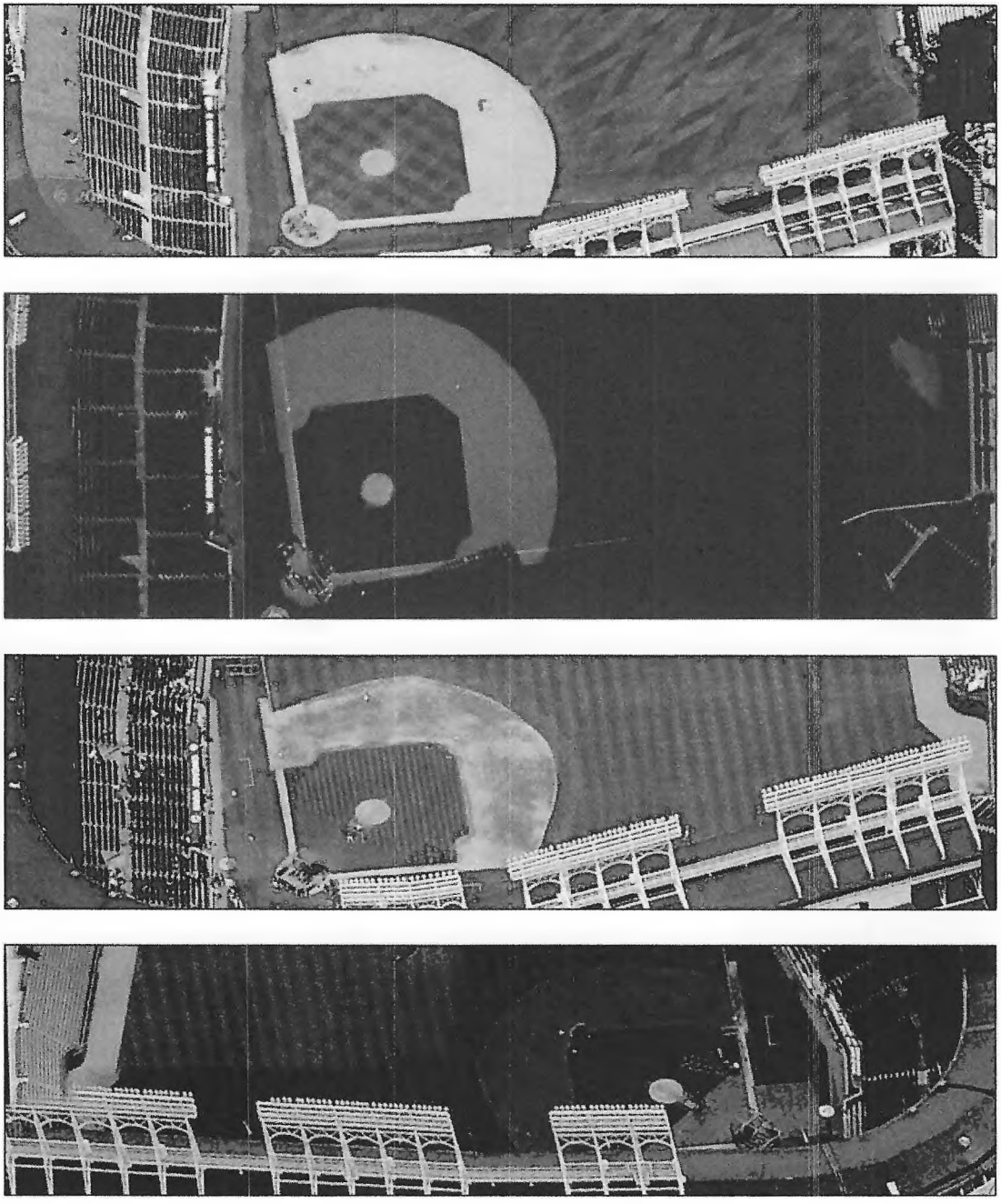
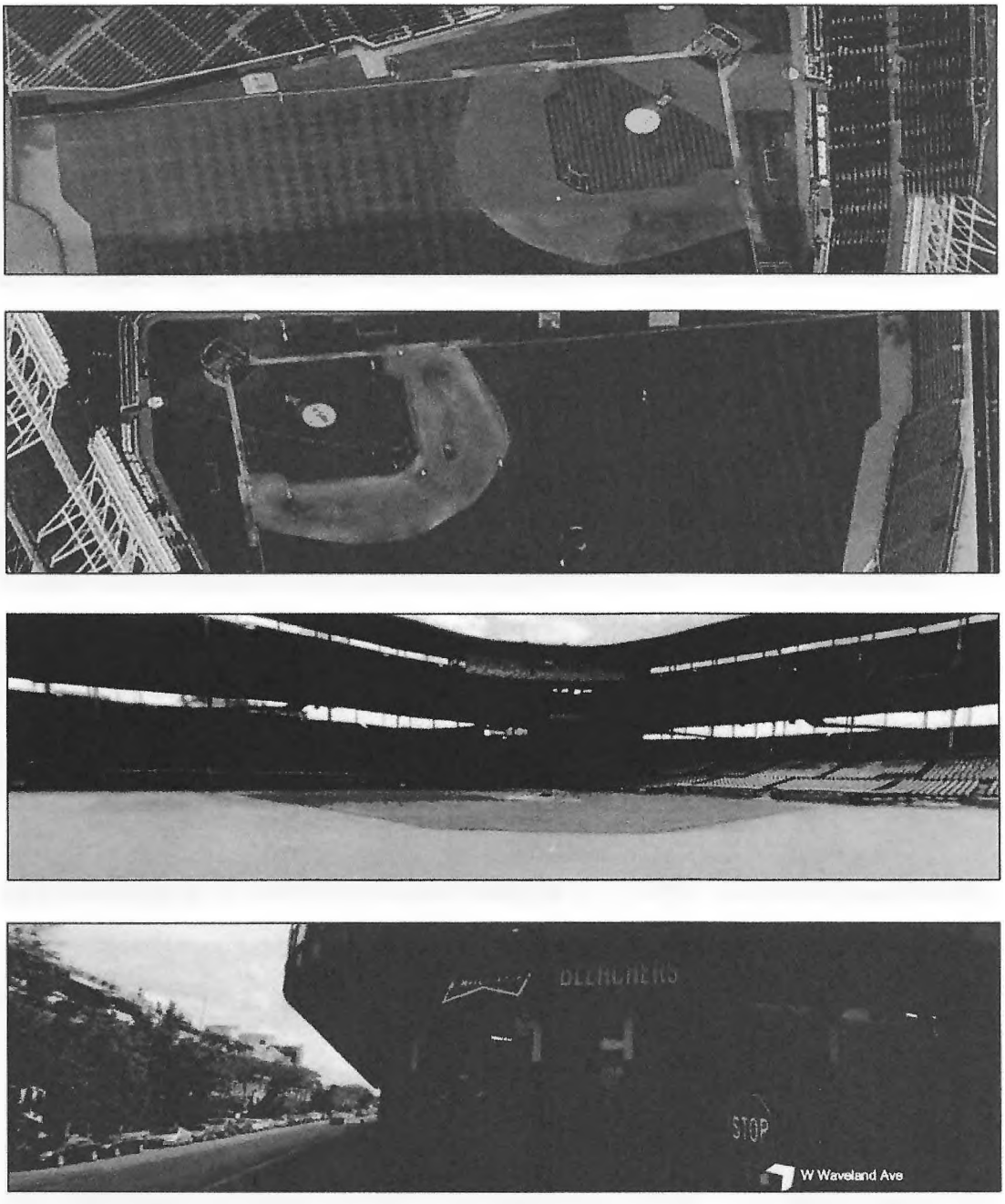


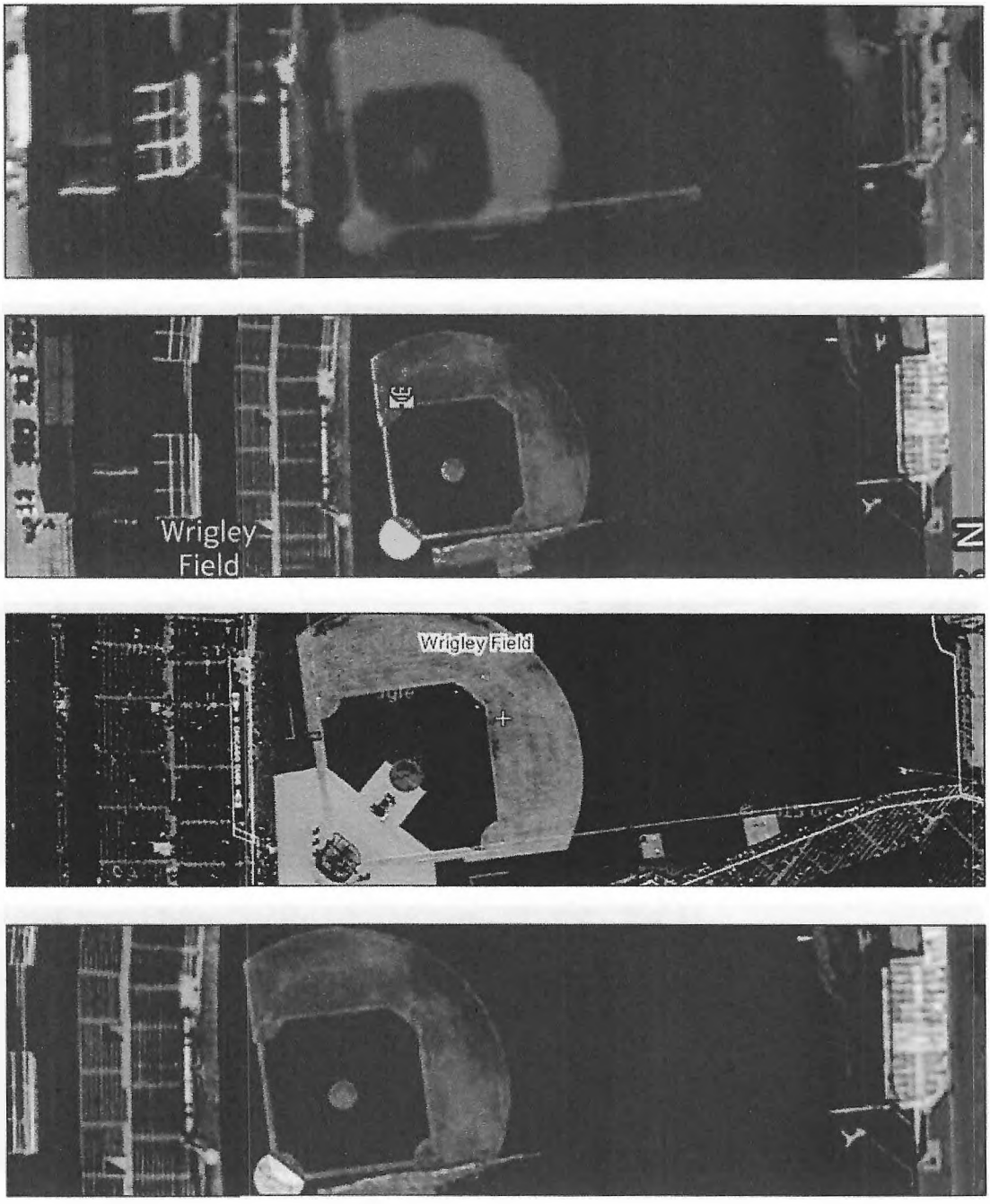
Figure 18.07: The IntelTechniques Custom Maps Search Tool.



Figures 18.08 through 18.11: Satellite views from Google, Bing, Bing Bird's Eye (north), and Bing Bird's Eye (east).



Figures 18.12 through 18.15: Satellite views from Bing Bird's Eye (south), Bing Bird's Eye (west), Google Street View, and Bing Street View.



Figures 18.16 through 18.19: Satellite views from Terra Server, Here, WikiMapia and ArcGIS.



Figures 18.20 through 18.23: Unique views from Yandex, Descartes, Snapchat, and Mapillary.

Scribble Maps (scribblemaps.com)

The default view of mapping services such as Google and Bing may be enough for your situation. Occasionally, you may want to modify or customize a map for your needs. Law enforcement may want to create a map to be used in a court case; a private investigator may want to customize a map to present to a client; or a security director may want to use this service to document the inappropriate Tweets that were found during the previous instructions. Scribble Maps offers one of the easiest ways to create your own map and add any type of visual aids to the final product.

The default view of your new map at Scribble Maps will display the entire world and a menu of basic options. I close this menu by clicking the small "x" in the upper right corner. You can then manually zoom into an area of interest or type in an address in the location bar at the top of the map. This will present you with a manageable area of the map. The lower right corner will allow you to switch from a traditional map view to a satellite or hybrid view.

The menu at the top of the map will allow you to add shapes, lines, text, and images to your map. Practicing on this map can never be replaced with any instruction printed here. Mastering the basics of this application will make occasional use of it easy. Figure 18.24 displays a quick sample map that shows a title, a line, a marker, and graphics. The menu can be seen in the upper left portion. When finished, the "Menu" button will present many options to print, save, or export your map. I also highly recommend **Free Map Tools** (freemaptools.com). This service provides multiple advanced options such as mapping a radius around a point of interest.

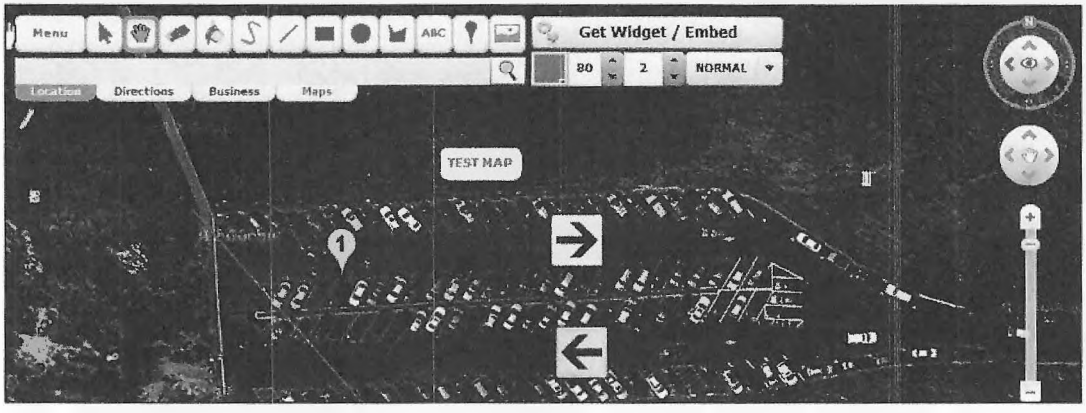


Figure 18.24: A basic custom map created with Scribble Maps.

CHAPTER NINETEEN
DOCUMENTS

The open source intelligence discussed up to this point has focused on websites that include valuable information about a target. A category of intelligence that is often missed during OSINT research is documents. This type of data usually falls into one of three classes. The first is documents that include information about the target within the contents of the file. These can include online PDF files that the target may not know exist. The second class is documents that were actually created by the target. These files can make their way into public view unintentionally. Finally, the third class includes the metadata stored within a document that can include vital information about the true source of the document. The following techniques explain manual searching and retrieving of these documents and automated software solutions for analysis of found content.

Google Searching (google.com)

A very basic way of locating documents that are publicly available on a specific website, or related to a specific topic, is to use Google. The "filetype:" (or "ext:") search operators explained previously can be used for this task. An example of a search query for all Microsoft Word documents stored on the domain of inteltechniques.com would be the following.

site:inteltechniques.com filetype:doc or site:inteltechniques.com filetype:docx

If you wanted to locate all documents that reference a specific topic, you can use the filetype operator without a specific website listed. An example of a search query for all Excel spreadsheets that contain the acronym OSINT would be the following.

filetype:xls "OSINT"

This search yielded 82 results for Excel documents. If you wanted to search for a specific person's name within any spreadsheets, such as John Doe, you would type the following query.

filetype:xls "John Doe"

The following table includes the most common document file types and the associated file extensions. As explained previously, both Google and Bing are capable of searching any file type regardless of the file association. Please note that this is a partial list, and I identify new possibilities constantly.

Microsoft Word	DOC, DOCX
Microsoft Excel	XLS, XLSX, CSV
Microsoft PowerPoint	PPT, PPTX
Adobe Acrobat	PDF
Text File	TXT, RTF
Open Office	ODT, ODS, ODG, ODP
Word Perfect	WPD

If you wanted to search all of these file types at once, the following string in Google or Bing would find most documents on the topic of OSINT. You could change that term to anything else of interest.

"OSINT" filetype:pdf OR filetype:doc OR filetype:xls OR filetype:xlsx OR filetype:docx OR filetype:ppt OR filetype:pptx OR filetype:wpd OR filetype:txt

This query basically tells the search engine to look for any reference to the term OSINT inside of a PDF file, Microsoft Word file, et cetera, and display all of the results. The Google Custom Search Engine described in Chapter Eight is a great resource for this exact type of search. However, I highly recommend having an understanding of the manual search process. It will give you much more control than any automated solution. The first three editions of this book contained several third-party document search services. Most of them either disappeared or now rely solely on a Google custom engine. Therefore, I no longer recommend any of them. They simply cannot compete with a properly structured document search on Google or Bing.

Google Docs (docs.google.com)

The idea of storing user created documents on the internet is gaining a lot of popularity. Keeping these files "in the cloud" eliminates the need for personal storage on a device such as a CD or flash drive. In addition, storing files on the internet allows the author to access and edit them from any computer with an internet connection. A common use of these document-hosting sites is to store them only during the editing phase. Once the document is finished and no longer needed, the user may forget to remove it from public view. Google is one of the most popular document storage websites. It allows users to embed the stored documents into their own websites if desired. Searching the site is relatively easy.

Many Google Mail (Gmail) users take advantage of Google's free service for document storage called Google Docs or Google Drive. When a document is created, it is private by default and not visible to the public. However, when people want to share documents with friends or coworkers, the sharing properties must be changed. While it is possible to privately share files with individual Google users, many people find it easier to make the documents public. Most of these users probably assume that the files will not be seen by anyone other than the intended recipients. After all, who would go out searching for other people's documents? We will.

The Google Docs and Google Drive websites do not offer the option to search these public files, but you can do this using Google search. Now that Google Docs allows search engines to index most of the public files, you should be able to find them with some specific search methods. The following search examples will explain a few of the options that would be conducted on google.com. The exact search is listed with the expected result. These should be used as a starting point for the many possibilities of document searching.

site:docs.google.com "resume" - 22,700 online resumes
site:docs.google.com "resume" "Williams" - 2,491 resumes with the name Williams
site:docs.google.com "Corey Trager" - 1 document (resume) belonging to the target
site:docs.google.com 865-274-2074 - 1 document containing the target number

Google categorizes the documents that are created by the user. The examples below identify searches that would display documents by type.

site:docs.google.com/presentation/d – 865,6000 PowerPoint presentations
site:docs.google.com/drawings/d – 68,600 Google flowchart drawings
site:docs.google.com/file/d – 6,945,000 images, videos, PDF files, and documents
site:docs.google.com/folder/d – 94,000 collections of files inside folders
site:docs.google.com/open – 1,400,000 external documents, folders, and files

In 2013, Google began placing some user generated documents on the "drive.google.com" domain. Therefore, any search that you conduct with the method described previously should be repeated with "drive" in place of "docs". The previous search for the telephone number would be the following.

site:drive.google.com 865-274-2074

Microsoft Docs (docs.microsoft.com)

Similar to Google Drive, Microsoft Docs offers that ability to store and share documents. The service is not as popular as Google Drive. However, there are thousands of publicly visible documents waiting to be found. The shared files are stored on the docs.microsoft.com domain. A query for resumes would be as follows. This search could be conducted on Google or Bing. The result on Google was 63,400 resume files with personal information.

site:docs.microsoft.com "resume"

Amazon Web Services (amazonaws.com)

Amazon Web Services (AWS) is a large collection of servers that supply storage and internet application hosting in "the cloud". Instead of purchasing expensive hardware, many companies and individuals rent these servers. There are numerous documents available for download from

these servers when searched appropriately. I cannot overstate the value of searching Amazon's servers. This is where most of the voter data that was heavily discussed during the 2016 election originated. I have personally located extremely sensitive documents from this source on numerous occasions. The following structure will identify files indexed on google.com.

site:amazonaws.com

The following search examples will explain a couple of the options. The exact search is listed with the expected result. These should be used as a starting point for the many possibilities of document searching.

site:amazonaws.com ext:xls "password" – 114 Excel spreadsheets containing "password"
site:amazonaws.com ext:pdf "osint" – 182 PDF files containing "osint"

Another option is the Amazon Cloudfront servers. CloudFront is a content delivery network (CDN) offered by Amazon Web Services. Content delivery networks provide a globally-distributed network of proxy servers which cache content, such as web videos or other bulky media. These are provided more locally to consumers, thus improving access speed for downloading the content. We can apply the same previous search techniques on this domain. The following search on Google yielded 129 results of pages on various Cloudfront servers containing the acronym "OSINT".

site:cloudfront.net OSINT

Gray Hat Warfare (buckets.grayhatwarfare.com)

AWS hosts more than simple documents. Many companies host large databases and data sets within "buckets" on AWS servers. Custom Google searches may locate some of this content, but never a full index of all data. This is where Gray Hat Warfare excels. It has created a searchable database of over one billion files, all publicly stored with AWS servers. Free users are limited to the first 350 million of these files, which is substantial. A search of "OSINT" revealed only 11 results while "password" displayed over 7,000. The results are often large files which must be opened carefully. A search of "password xls" provided three documents with active credentials. A direct search URL is as follows.

https://buckets.grayhatwarfare.com/results/password

Google Cloud Storage (cloud.google.com)

This is Google's response to Amazon's AWS. It is a premium file storage web service for storing and accessing data on the Google Cloud Platform infrastructure. It is heavily used by all types of businesses and tech-savvy individuals. The number of publicly available sensitive documents is growing at an alarming rate. Below are a few examples.

site:storage.googleapis.com ext:xlsx OR ext:xls - 2,310 Spreadsheets
site:storage.googleapis.com "confidential" - 9,502 Documents
site:storage.googleapis.com "confidential" ext:pptx - 11 PowerPoint files marked as confidential

Presentation Repositories

With unprecedented online storage space at all of our fingertips, many people choose to store PowerPoint and other types of presentations in the cloud. Several free services have appeared to fill this demand. Of those, the following have the majority of publicly available documents.

Slide Share (slideshare.net)
ISSUU (issuu.com)
Slidebean (slidebean.com)
Prezi (prezi.com)

Slide Share and ISSUU allow native searching within their websites. However, Prezi does not have this option. For all three, I recommend a custom Google search with the site operator. If I want to locate presentations including the term OSINT from Slide Share, I would use the following query.

site:slideshare.net "OSINT"

Scribd (scribd.com)

Scribd was a leading cloud storage document service for several years. Since 2014, it has shifted its focus toward ebook sales. However, the plethora of stored documents is still accessible. This can be valuable for historical content posted, and likely forgotten, by the target. A search field is at the top of every page on the site within their collapsible menu. Searching for your target name should produce any public books stored through this service that includes the target name on any page of the publication. Clicking "Documents" in the menu will present more relevant information. Most of these documents are intentionally stored on the site and any groundbreaking evidence of criminal activity will not be included. Instead, the primary use of the site for OSINT investigations is the large number of documents related to businesses. Entering any large corporation name should display several pages of viewable documents related to the company. Often, these include documents that the company's security personnel would not authorize to be online. Searching for "FOUO", an acronym for "for official use only", produced hundreds of results. While none of these appeared to be officially classified, they were not intended to be posted to a public website. If you are presented with an unmanageable amount of results, the filter options appear directly above the first document result. These will allow you to search by language, size, file type, and date uploaded. Identifying the user that uploaded a document is as easy as locating the document. In the upper-center of any page containing a document, there is an area that will identify the subject that uploaded the file. This also acts as a link to this user's profile on the website. The profile will display any information that the user supplied as well as a feed of recent activity of that user on the site. This can help identify other documents uploaded by a specific user.

PDF Drive (pdfdrive.com)

This service scans the internet for new PDF files and archives them on its own site. This can be helpful when the original source removes the content. Fortunately for us and our tools, the following static URL conducts a query across the entire domain.

https://www.pdfdrive.com/search?q=osint

WikiLeaks (search.wikileaks.org)

Some websites are created for the sole purpose of leaking sensitive and classified documents to the public. Wikileaks is such a site. When an Army soldier named Bradley Manning was arrested in 2010 for uploading classified government information to the site, Wikileaks became a household name. People then began to flock to the site to catch a glimpse of these controversial documents and videos. The official Wikileaks site finally provides a working search option. It will allow you to enter any search terms and will provide results of any leaked documents that contain these terms. Both the government and the private sector should be familiar with this site and the information that is identified with their respective agency.

Cryptome (cryptome.org)

Another site that strives to release sensitive and classified information to the public is Cryptome. Most of the information is related to freedom of speech, cryptography, spying, and surveillance. Much of the content could be considered conspiracy theories, but several official documents get released daily. Cryptome does not provide a search for their site and there are no third-party providers that cater to this service. Therefore, we must rely on Google or Bing to find the documents. A structured query of "osint" should function well. This technique using the search terms of "bradley manning" linked to 77 documents surrounding his investigation.

site:cryptome.org "osint"

Paste Sites

Paste Sites are not technically storage services for documents. They are websites that allow users to upload text for public viewing. These were originally designed for software programmers that needed a place to store large amounts of text. A link would be created to the text and the user could share the link with other programmers to review the code. This is still a common practice, but other users have found ways to abuse this technology. Many hacking groups will use this area of the internet to store compromised account information, user passwords, credit card numbers, and other sensitive content. There are dozens of sites that cater to this need, and very few of them have a search feature.

Pastebin (pastebin.com)

Pastebin is the most popular paste site in the United States. Criminal hacker groups often use this site to release illegally obtained data to the public. A previous release included the home addresses and personal information of many police officers near Ferguson, Missouri during protests and riots. More recently, stolen bank records and credentials from Venezuela were posted with encouragement to infiltrate the company. This is one of the sites that will allow for a search from within the site. This function performs a search through Google in the same way we could with the "site" operator. Typing in a target name, email address, or business name may reveal private information not intended for the public. For law enforcement, typing in the last four digits of a stolen credit card number may identify a link to the thief. If successful, the target is most likely outside of the country. Regardless, this is a valuable piece to the case and an impressive explanation to the victim. Unfortunately, most of the users leave a default username of "Guest". Pastebin allows a direct URL search, which relies on Google for indexing, as follows.

https://pastebin.com/search?q=osint

New paste sites come and go monthly. There is no way to present a current and complete list. However, I will focus on the most stable and prominent options which allow search. In a moment, I present a custom search tool which queries all 51 of these at once. The following sites can be individually queried with the site operator, such as site:doxbin.org "osint".

batman.gyptis.org/zerobin
cl1p.net
codepad.org
codepaste.net
controlc.com
doxbin.org
dpaste.com
dpaste.org
dumpz.org
friendpaste.com
gist.github.com
hastebin.com
heypasteit.com
hpaste.org
ideone.com
ivpaste.com
jsbin.com
justpaste.it
mysticpaste.com
nopaste.info
0bin.net
paste.debian.net
paste.ee
paste.fedoraproject.org
paste.frubar.net
paste.kde.org
paste.lisp.org
paste.pound-python.org
paste.opensuse.org
paste.org
paste.org.ru
paste.ubuntu.com
paste2.org
pastebin.ca
pastebin.com
pastebin.fr
pastebin.gr
pastebinz.com
pastefs.com
pastehtml.com
pastelink.net
pastie.org
p.ip.fi
privatebin.net
slexy.org
snipplr.com
snipt.net
sprunge.us
textsnip.com
tidypub.org
wordle.net

Document Metadata

When an original document is found online, it is obviously important to analyze the visible content of the file. This includes the file name, written text, and an original location of the document. Digging deeper will expose more information. There is data embedded inside the document that cannot be seen by simply looking at the content of the file. This data is called metadata and can be very valuable to any type of investigation. It can often include the computer name on which the document was created, the username of the computer or the network, the software version used, and information about the network to which the computer is connected. The best way to view all of this information is to use software-based metadata viewers, but you can also view this "hidden" information online through a web browser.

Several online sites will allow you to upload documents for analysis. To do this, click the "browse" button on the pages detailed below. This will enable a file explorer that will allow you to select the document that you want analyzed. The result often identifies a created and modified date, the original title, three applications used to create the document, and a username. A further search of this username through the previously discussed techniques could produce a wealth of information about the author of the document. The following websites allow you to upload a locally stored document or submit a URL of a file for analysis. Please use caution with this technique. If the document is already posted online, there is very little risk of allowing a URL analysis. However, a locally stored file that has never been on the internet may require a second thought. If the content is sensitive, you may not want to upload to any service. If the file contains classified information, you could be jeopardizing your clearance. In these situations, use the methods discussed in a moment. If this is not a concern the following websites work well.

Extract Metadata (extractmetadata.com)
Jeffrey's Viewer (exif.regex.info/exif.cgi)
ExifInfo (exifinfo.org)
Get Metadata (get-metadata.com)

Document Metadata Applications

If I need to analyze the metadata stored within documents, I prefer to do so locally on my machine. I do not want to share any potential investigation details with an online resource. This is especially true if the documents are not already online. You may possess a folder of files which were retrieved from a suspect computer or emailed directly to you. In these scenarios, we should be cautious as to not distribute any evidence electronically to any websites. I present two solutions, with a third unstable option.

ExifTool (sno.phy.queensu.ca/~phil/exiftool)

You already possess a document metadata viewer within your custom Linux virtual machine. It is called ExifTool and we installed it during the previous chapters. This is a terminal-based solution, but the function is quite simple. Assume that you have used the previous techniques to download several Word documents in .docx file format onto your Desktop in a folder titled

Evidence. The following steps within Terminal would navigate to the proper folder, generate a spreadsheet with the metadata included within the documents, and title it Report.csv on your Desktop in the Evidence folder.

```
cd ~/Desktop/Evidence
exiftool * -csv > ~/Desktop/Evidence/Report.csv
```

Let's conduct an example and take a look at the results. I performed the following Google search:

ext:docx "osint"

This provided 371 results, all of which were Microsoft Word documents. I downloaded the first four into the Evidence folder on my Desktop. After executing the commands above, I launched the spreadsheet. Figure 19.01 displays a small portion of the interesting results. This tells me the names of the individuals who created and last modified the documents; the companies involved; the software which was used; and even the amount of time they spent editing the content. This is extremely valuable information which should be collected during every investigation in which documents are obtained.

Application	AppVersion	Company	CreateDate	Creator	LastMo	By	TotalEditTime
Microsoft Office Word	12		2011:03:18 20:29:00Z	marko.prij	marko.J	ac	1.2 hours
Microsoft Macintosh Word	15		2016:04:25 14:56:00Z	Kirk Haye	Kirk Ha		26 minutes
Microsoft Office Word	12		2017:08:03 09:53:00Z	Hakon201	Hakon2		4 minutes
Microsoft Office Word	14	United States Army	2016:06:13 08:31:00Z	john.t.rich	IMO-P0	lin. Jonas A	3 hours

Figure 19.01: Document metadata results from ExifTool.

FOCA (www.elevenpaths.com/labstools/foca)

You may desire a Windows-based solution which possesses a user-friendly interface. FOCA was once the premier document metadata collection and extraction tool. It was created to search, download, and analyze documents and their metadata in one execution. Unfortunately, Google and other search engines began blocking the search and download behavior of the software. Fortunately, the analysis portion still works perfectly. The following steps will download and install FOCA to your Windows VM or any other Windows device.

- Navigate to https://www.elevenpaths.com/labstools/foca.
- Click the hyperlink at the bottom for the "previous version of FOCA".
- Accept the agreement and click "Download".
- Double-click the .zip file and extract the contents.
- Launch FOCA.exe from the "bin" folder within the "FOCAPro" folder.
- If prompted, select "Download and install this feature" to install the .net framework.

FOCA should launch and present a window with many options. Today, the vast majority of the features no longer work, but the document analysis is helpful. Assume you possess the same four documents mentioned previously on the Desktop of your Windows VM.

- Open FOCA and click the Metadata folder in the left menu.
- Drag and drop the documents into the FOCA window.
- Right-click any of the documents and choose "Extract all metadata".
- Right-click any of the documents and choose "Analyze metadata".

You can now click through the menus on the left to view any metadata details such as email addresses, names, and computers associated with these files. Figure 19.02 displays the result with the files mentioned previously. I have highlighted the Users section and redacted the screenshot. The benefit of this method is the user interface, but you sacrifice a reporting option. The previous ExifTool method is not as pretty, but the spreadsheet result is helpful.

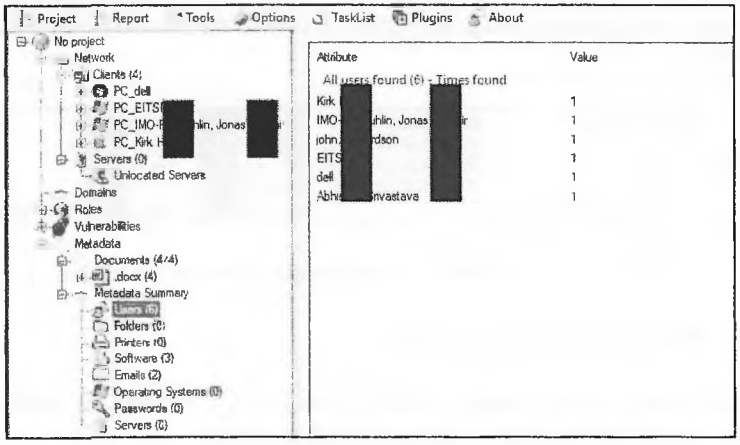


Figure 19.02: A FOCA analysis.

Metagoofil

I mentioned that FOCA was originally designed to locate and download documents, but no longer functioned. Metagoofil is a Linux option for replicating FOCA's original features. It does not always work well, as you are at the mercy of search engines to provide the data. When Google decides to block Metagoofil, its usefulness ceases. If desired, launch Terminal, and enter the following commands to add Metagoofil to your Linux VM.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/opsdisk/metagoofil.git`
- `cd Metagoofil`
- `pip3 install -r requirements.txt`

Metagoofil is now installed on your virtual machine. The following command launches Python 3 (python3); loads the script (metagoofil.py); sets the target domain (-d cisco.com); sets the file type as pdf (-t pdf); and saves the output to a new folder called "cisco" on the Desktop (-o ~/Desktop/cisco/). The second command asks for docx and xlsx files.

- `python3 metagoofil.py -d cisco.com -t pdf -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t docx,xlsx -o ~/Desktop/cisco/`

This automatically downloads any found documents. We could now analyze these documents with ExifTool as mentioned previously with the following commands within Terminal. The following commands navigate you to the directory with the files and then creates a report.

- `cd ~/Desktop/cisco`
- `exiftool * -csv > ~/Desktop/cisco/Report.csv`

As you may suspect, I prefer to download and analyze document metadata within Linux. While the command to launch ExifTool is easy, the commands for Metagoofil can become complicated, especially when we start querying for multiple types of documents. Therefore, let's create a script and shortcut in order to automate the process. If you completed the steps in previous chapters, you already have these installed. However, let's walk through them manually.

First, we need a script which offers two options. The first will be to execute Metagoofil and try to download any documents from a domain. The second will repeat that process and then create a report with all of the metadata. Finally, we need a Desktop shortcut file which allows us to launch the script without Terminal. This should all seem familiar from the first section of the book. These two files are in your Linux files archive, and titled metagoofil.sh and metagoofil.desktop. The following pages display the entire content of each. If you want to update Metagoofil at any time, enter the following commands, or add these to your update script as explained previously.

- `cd ~/Downloads/Programs/metagoofil`
- `git pull https://github.com/opsdisk/metagoofil.git`

You may wonder why I did not include this script within the Linux VM creation chapter. The answer is that this program often functions improperly. It is not a fault of the application, but due to restrictions by Google. When you execute any type of automated query against Google, it quickly blocks your IP address from this behavior. If you are connected to the internet through a VPN, this blockage is even more common. In my experience, this utility works fairly well when NOT using a VPN. Your mileage may vary, but I want you to possess all options possible.

```
#!/usr/bin/env bash
##Metagoofil Menu Script
#define choices
opt1="Metagoofil-Only"
opt2="Metagoofil+Exiftool"
timestamp=$(date +%Y-%m-%d:%H:%M)
fqdnregex="\b((xn--)?[a-z0-9]+(-[a-z0-9]+)*\.)+[a-z]{2,}\b"
domainmenu=$(zenity  --list  --title "Choose Tool" --text "What do you want
to do?" --width=400 --height=400 --radiolist  --column "Choose" --column
"Option" TRUE "$opt1" FALSE "$opt2" 2> >(grep -v 'GtkDialog' >&2))
case $domainmenu in
       $opt1 ) #Metagoofil-Only
       domain=$(zenity --entry --title "Metagoofil-Only" --text "Enter
target domain name" --entry-text "" 2> >(grep -v 'GtkDialog' >&2))
       if [ ! -z "$domain" ]; then
       #verify legit domain
       if [[ $domain =~ $fqdnregex ]]; then
       #Run Tool
       mkdir /home/osint/Documents/Metagoofil
       mkdir /home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain"
       mkdir
/home/osint/Documents/Metagoofil/"$timestamp"_results_"$domain"
       sleep 1
       python3 /home/osint/Downloads/Programs/metagoofil/metagoofil.py -d
$domain -w -t pdf,doc,xls,ppt,docx,xlsx,pptx -o
"/home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain""
       sleep 1
       nautilus "/home/osint/Documents/Metagoofil/" >/dev/null 2>&1
       exit
       fi
       else
       zenity --error --text "Doesn't appear to be a legitimate domain,
exiting!" 2> >(grep -v 'GtkDialog' >&2)
       exit
       fi
;;
       $opt2 ) #Metagoofil+Exiftool
       domain=$(zenity --entry --title "Metagoofil+Exiftool" --text "Enter
target domain name" --entry-text "" 2> >(grep -v 'GtkDialog' >&2))
       if [ ! -z "$domain" ]; then
       #verify legit domain
       if [[ $domain =~ $fqdnregex ]]; then
       #Run Tool
       mkdir /home/osint/Documents/Metagoofil
       mkdir /home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain"
       mkdir
/home/osint/Documents/Metagoofil/"$timestamp"_results_"$domain"
       sleep 1
       python3 /home/osint/Downloads/Programs/metagoofil/metagoofil.py -d
$domain -w -t pdf,doc,xls,ppt,docx,xlsx,pptx -o
"/home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain""
       sleep 1
```

```
        if [[ $(find
"/home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain"" -maxdepth 1
-type f) ]]; then
        find "/home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain""
-maxdepth 1 -type f | exiftool
/home/osint/Documents/Metagoofil/"$timestamp"_docs_"$domain"/* -csv >
~/Documents/Metagoofil/"$timestamp"_results_"$domain"/Report.csv
        nautilus "/home/osint/Documents/Metagoofil" >/dev/null 2>&1
        exit
        else
        echo "No files found for parsing, exiting"
        exit
        fi
        else
        zenity --error --text "Doesn't appear to be a legitimate domain,
exiting!" 2> >(grep -v 'GtkDialog' >&2)
        exit
        fi
        fi
;;
esac
```

The desktop shortcut file is as follows.

```
[Desktop Entry]
Type=Application
Name=Metagoofil
Categories=Network;OSINT
Exec=/home/osint/Documents/scripts/metagoofil.sh
Icon=/home/osint/Documents/icons/domains.png
Terminal=true
```

The metagoofil.sh file should be placed in your Documents/scripts folder if it is not already there and the metagoofil.desktop file should be placed in the /usr/share/applications folder as explained in the first section of this book. Again, if you followed the instructions in previous chapters, this has already been done for you. This script, available within the Applications window in your Linux VM, launches a window which prompts you to enter a target domain name. It then tries to extract any pdf, doc, xls, ppt, docx, xlsx, or pptx file types from the provided domain. The second menu option takes this further by creating a metadata report for the files downloaded. Personally, my protocol in regard to document metadata collection and analysis is as follows.

- If I already possess numerous documents on my computer, I create a metadata CSV spreadsheet using ExifTool. I then analyze this document.
- If my target website possesses few documents, I download them manually through Google or Bing within a web browser.
- If my target website possesses hundreds of documents, I use Metagoofil, but only download one file type at a time. If my target were cisco.com, I would execute the following commands in Terminal.

- `python3 metagoofil.py -d cisco.com -t pdf -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t doc -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t xls -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t ppt -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t docx -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t xlsx -o ~/Desktop/cisco/`
- `python3 metagoofil.py -d cisco.com -t pptx -o ~/Desktop/cisco/`

Real World Application: Dennis Lynn Rader, also known as the BTK killer, sent a floppy disk to the Wichita Police Department containing a Microsoft Word document in reference to his killings. The police examined the metadata of this document and determined that it was made by a subject named "Dennis". Links to a Lutheran church were also located within this data. Conducting OSINT searches on these two pieces of information helped to identify the suspect and make an arrest.

Free OCR (free-ocr.com)

You may occasionally locate a PDF file that has not been indexed for the text content. These types of PDF files will not allow you to copy and paste any of the text. This could be due to poor scanning techniques or to purposely prohibit outside use of the content. You may desire to capture this text for a summary report. These files can be uploaded to Free OCR and converted to text documents. OCR is an acronym for optical character recognition. Basically, a computer "reads" the document and determines what the text is inside the content. The result is a new document with copy and paste capability.

Rental Vehicle Records

The details of rental vehicles are not technically documents, but the data seemed to fit this category the best. The following options have been controversially received during training and may not be appropriate for everyone. I present these methods to you as theories, and you should evaluate if the techniques are suitable for your research. Several vehicle rental companies offer an option to access your receipts online. This is probably designed for customers that leave a vehicle at the business after hours and later need a receipt. While the processes to retrieve these documents are designed to only obtain your own records, it is very easy to view others.

Enterprise (enterprise.com)

At the bottom of every Enterprise web page is an option to "Get a receipt". Clicking this will present a form that must be completed before display of any details. Enterprise will need the driver's license number and last name. Providing this information will display the user's entire rental history for the past six months to three years. Testing with my own data provided two years' worth of results. Each document will link to the entire receipt from that rental. These receipts include the start date and time, end date and time, vehicle make and model, pick up

location, total mileage, lease name, and form of payment. This information could be very beneficial to any drug case or private investigation.

Hertz (hertz.com)

Similar to Enterprise, Hertz has a link at the bottom of every page titled "Get a receipt". You can search by driver's license number or credit card number and will need a matching last name. The receipt will be very similar to the example in the Enterprise demonstration.

Alamo (alamo.com)

Alamo titles their receipt retrieval link "Find a Receipt" and it is located in the lower right portion of every page. The process is identical to the previous two examples. The only difference is that you must choose a date range. I usually select a start date of one year prior to the current date and the end date of the current date.

IntelTechniques Documents & Pastes Search Tools

If these operators seem overwhelming, consider the Google Custom Search Engines (CSE) that were explained in Chapter Eight. They apply to most of the methods discussed here. These engines will present a simple search field ready for any keyword desired. I have found an all-in-one Google CSE to be a bit unreliable and sporadic with results. Therefore, I created my own options, which can be seen in Figures 19.03 and 19.04. The first tool, titled Documents.html in your download from previous chapters has two sections, both of which are explained below.

The first section (Figure 19.03) queries documents by file types. It allows entry of any terms or operators in order to locate PDF, DOC, DOCX, XLS, XLSX, CSV, PPT, PPTX, KEYNOTE, TXT, RTF, XML, ODT, ODS, ODP, ODG, ZIP, RAR, 7Z, JPG, JPEG, PNG, MPG, MP4, MP3, and WAV documents. The "Submit All" option will execute each search in its own tab.

The second section (Figure 19.03) allows entry of any terms or operators in order to locate files stored within Google Docs, Google Drive, Microsoft Docs, Amazon AWS, Cloudfront, SlideShare, Prezi, ISSUU, Scribd, PDF Drive, and others. The "Submit All" option will execute each search in its own tab.

The Pastes.html search tool (Figure 19.04) presents a Google custom search engine (CSE) which queries all 51 paste sites mentioned previously. In this example, I have searched the term "osint" and received 676 results within the search tool.

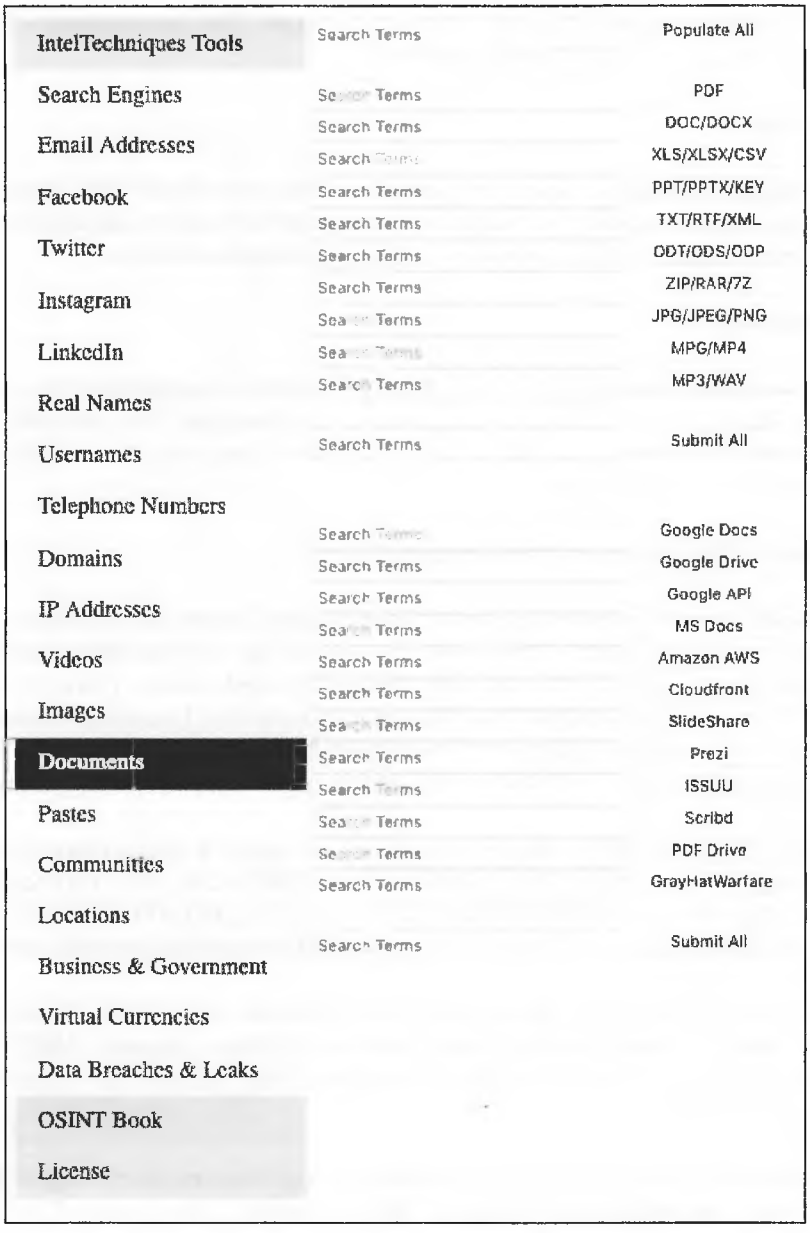


Figure 19.03: The IntelTechniques Documents Search Tool.



Figure 19.04: The IntelTechniques Pastes Search Tool.

Chapter Twenty
Images

Thanks to cameras on every data cellular phone, digital photograph uploads are extremely common among social network users. These images can create a whole new element to the art of open source intelligence analysis. This chapter will identify various photo sharing websites as well as specific search techniques. Later, photo metadata will be explained that can uncover a new level of information including the location where the picture was taken; the make, model, and serial number of the camera; original uncropped views of the photos; and even a collection of other photos online taken with the same camera. After reading this information, you should question if your online photos should stay online.

Google Images (images.google.com)

During my live training sessions, I always encourage attendees to avoid individual searches on various photo sharing websites such as Flickr or Picasa. This is because most of these searchable sites have already been indexed by Google and other search engines. Conducting a search for "Oakland Protest" on Flickr will only identify images on that specific service that match. However, conducting the same search on Google Images will identify photos that match the terms on Flickr and hundreds of additional services. Similar to Google's standard search results, you can use the Search Tools to filter by date. Additionally, you can further isolate target images by size, color, and type, such as photographs versus line drawings. I no longer conduct manual searches across the numerous photo sharing sites. Instead, I start with Google Images.

Bing Images (bing.com/images)

Similar to Google, Bing also offers an image search. While it is not as beneficial as the Google option, it should never be overlooked. On several occasions, I have located valuable pictorial evidence on Bing that was missing from Google results. The function is identical, and you can filter search results by date, size, color, type, and license type. When searching for relevant data about a target, I try to avoid any filters unless absolutely necessary. In general, we always want more data, not less. The search techniques explained in Chapter Eight all apply to queries on Google Images and Bing Images.

Reverse Image Searches

Advancements in computer processing power and image analysis software have made reverse image searching possible on several sites. While a standard search online involves entering text into a search engine for related results, a reverse image search provides an image to a search engine for analysis. The results will vary depending on the site used. Some will identify identical

images that appear on other websites. This can be used to identify other websites on which the target used the same image. If you have a photo of a target on a social network, a reverse analysis of that photo may provide other websites on which the target used the same image. These may be results that were not identified through a standard search engine. Occasionally, a target may create a website as an alias, but use an actual photo of himself. Unless you knew the alias name, you would never find the site. Searching for the site by the image may be the only way to locate the profile of the alias. Some reverse image sites go further and try to identify other photos of the target that are similar enough to be matched. Some will even try to determine the sex and age of the subject in the photo based on the analysis of the image. This type of analysis was once limited to expensive private solutions. Now, these services are free to the public.

Google Reverse Image Search (images.google.com)

One of the more powerful reverse image search services is through Google. Rolled out in 2011, this service is often overlooked. On any Google Images page, there is a search field. Inside of this field on the far right is a light grey camera icon that appears slightly transparent. Figure 20.01 (first) displays this search field. Clicking on this icon will open a new search window that will allow for either a web address of an online image, or an upload of an image file on your computer. In order to take advantage of the online search, you must have the exact link to the actual photo online. Locating an image within a website is not enough. You will want to see the image in a web browser by itself, and then copy the address of the image. If I want to view the image from the actual location, I must right-click on the image and select "view image" with my Firefox browser. Chrome users will see "open image in new tab" and Internet Explorer users will see "properties" which will identify the URL of the image. This link is what you want in order to conduct a reverse image analysis. If you paste this link in the Google Images reverse online search, the result will be other similar images, or exact duplicate images, on other sites. Visiting these sites provides more information on the target.

Note that adding context to the reverse-search field after submission can improve accuracy. As an example, a reverse-search of a photo from LinkedIn might produce many inaccurate results, but including the name or employer of your target will often display only applicable evidence. Another way to use this service is to search for a target within the Google Images search page. The images in the results will present additional options when clicked. A larger version of the image will load inside a black box. The three options to the right of the image will allow you to visit the page where the image is stored, view the image in full size, or "Search by image". Clicking the "Search by image" link will present a new search results page with other images similar to the target image. These connect to different websites which may contain more intelligence about the subject.

Bing Reverse Image Match (bing.com/images)

In 2014, Bing launched its own reverse image search option titled "Image Match". This feature can be launched from within any page on Bing Images by clicking the Image Match icon to the

right of the search field. Figure 20.01 (second) displays this option. This service does not seem to be as robust as Google's. In my experience, I often receive either much fewer results, although they do match. On a few occasions, I have received matched images that Google did not locate.

TinEye (tineye.com)

TinEye is another site that will perform a reverse image analysis. These results tend to focus on exact duplicate images. The results here are usually fewer than those found with Google. Since each service often finds images the others do not, all should be searched when using this technique. Figure 20.01 (third) displays the search menu. The icon on the left prompts the user to provide a location on the hard drive for image upload while the search field will accept a URL.

Yandex Images (images.yandex.com)

Russian search site Yandex has an image search option that can conduct a reverse image search. Similar to the other methods, enter the full address of the online image of interest and search for duplicate images on additional websites. In 2015, Yandex began allowing users to upload an image from their computers. Overall, these results will be limited. However, this option is vital for any international investigations. Figure 20.01 (fourth) displays the reverse image search icon in the far-right portion.

Baidu Images (graph.baidu.com)

Similar to Yandex, the Chinese search engine Baidu offers a reverse image search. Baidu currently offers no English version of their website and only presents Chinese text. Navigating to the above website offers a search box that contains a small camera icon to the right. Clicking this presents options for uploading an image (button to left) or providing the URL of an online image within the search field itself. The results will identify similar images on websites indexed by Baidu. Figure 20.01 (fifth) displays the search page only available in Chinese. In my experience, this reverse search option fails more than it functions.

Regardless of the services that you are executing, I urge you to use caution with sensitive images. Similar to my view of analyzing online documents for metadata, I believe that submitting online photos within these engines is harmless. If the photo is already publicly online, there is very little risk exposing it a second time. My concern involves child pornography and classified photos. As a former child pornography investigator and forensic examiner, there were several times that I wanted to look for additional copies of evidence online. However, I could not. Even though no one would know, and the photos would never appear any place they should not, conducting reverse image searches of contraband is illegal. It is technically distributing child pornography (to Google). While working with a large FBI terrorism investigation, I had possession of ten photos on which I wanted to conduct a reverse image search. The photos were part of a classified case, so I could not. Overall, never submit these types of photos from your hard drive. It will always come back to haunt you.

Whenever I have any public images that call for reverse image searching, I always check all five of these services. While I rarely ever get a unique result on Baidu, it only takes a few seconds to check every time. This diligence has paid off in the past. These manual searches do not need to be as time consuming as one may think. We can automate much of this process to save time and encourage thorough investigations. First, we should take a look at direct URL submission. For the following examples, assume that your target image is the cover of this book from the web page at inteltechniques.com/book1.html. The actual target image is stored online at the URL of https://inteltechniques.com/img/book1.png. The following direct addresses would conduct a reverse image search at each service listed.

Google: https://www.google.com/searchbyimage?site=search&sa=X&image_url=https://inteltechniques.com/img/book1.png

Bing: http://www.bing.com/images/searchbyimage?FORM=IRSBIQ&cbir=sbi&imgurl=https://inteltechniques.com/img/book1.png

TinEye: http://www.tineye.com/search/?url=https://inteltechniques.com/img/book1.png

Yandex: https://www.yandex.com/images/search?img_url=https://inteltechniques.com/img/book1.png &rpt=imageview

Baidu: https://graph.baidu.com/details?isfromtusoupc=1&tn=pc&carousel=0&image=&image= https://inteltechniques.com/img/book1.png

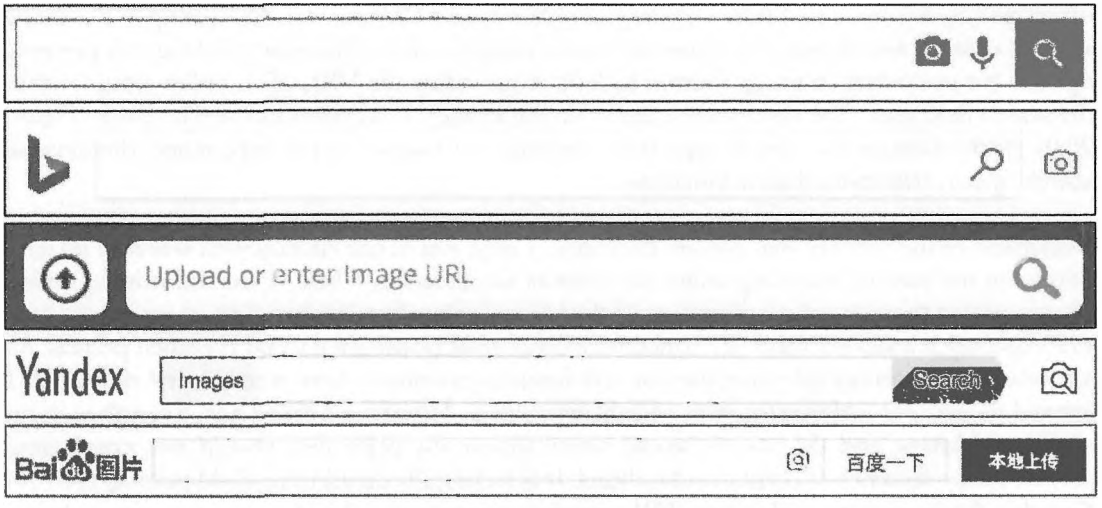


Figure 20.01: Reverse image search options from Google, Bing, TinEye, Yandex, and Baidu.

Cropped Reverse Image Searching

Beginning in 2018, I noticed that both Google Images and Bing Images were returning fewer results than in previous years. It seemed as though each were trying to limit the number of matching photos, possibly with the intent to present only relevant images based on previous search history or whatever they "think" we want from them. In 2019, I had an investigation focused around an online image. When I conducted a reverse image search, I received one result, which was a copy I already knew existed. When I cropped the image to only display my target, I received more search results. I find this technique applicable to Google and Bing, but I believe it works best with Yandex. The following is a demonstration of this method, using a public image recently posted to Twitter.

Figure 20.02 (left) is the original image obtained from Twitter. A reverse image search through Google, Bing, and Yandex revealed numerous results, but none of them contained my target displayed on the far left. I cropped the image, as seen in Figure 20.02 (right), to display only the target and resubmitted to Yandex. This immediately identified numerous images of the target. Figure 20.03 displays one of these images. Both Google and Bing displayed no results from this cropped image. I cannot stress enough the importance of reverse searching images through Yandex. I find their service superior to all others.



Figure: 20.02 An original image (left) and cropped version (right).

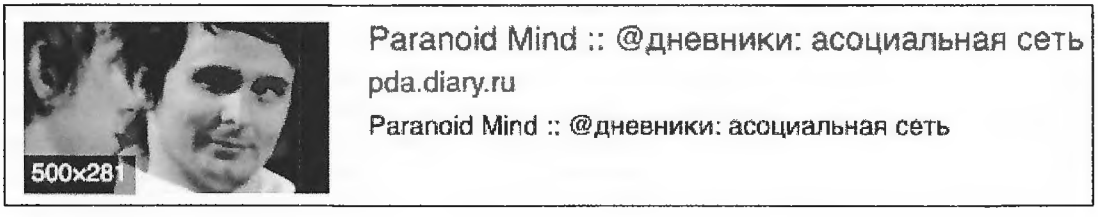


Figure 20.03: A reverse image search result from Yandex.

Karma Decay (karmadecay.com)

This service was mentioned in Chapter Thirteen and has a very specific specialty which can be beneficial to an internet researcher. It is a reverse image search engine that only provides positive results that appear on the website Reddit. It was originally launched as a way for users to identify when someone reposted a photo that had previously been posted on the website. The user could then "down-vote" the submission and have it removed from the front page. We can use this in investigations to locate every copy of an individual photo on Reddit. You can either provide a link to an image or upload an image from your computer. The following static URL submits our target image for reverse analysis on Reddit.

https://karmadecay.com/search?q=https://inteltechniques.com/img/book1.png

Root About (rootabout.com)

This is another specialized reverse image search utility which only queries against images stored on the Internet Archive and within Open Library (also an Internet Archive product). Any results will likely contain public images such as published photos and book covers. I have yet to receive any benefit to my investigations with this service, but it should still be available in your arsenal of tools. Root About does not support search via a direct URL.

Wolfram Image Identification Project (imageidentify.com)

While this is not a traditional reverse image search, it does provide value. The goal of this service is to identify the content of an image. If you upload a photo of a car, it will likely tell you the make, year, and model. An upload of an image containing an unknown Chinese word may display a translation and history details. The site prompts you to upload a digital file from your computer, but you can also drag and drop an image from within a web page in another tab.

Real World Application: These reverse image search sites can have many uses to the investigator. In 2011, I searched a photo of damage to a popular historic cemetery that was vandalized. The results included a similar photo of the suspect showing off the damage on a blog. An arrest and community service soon followed. Later, while working with a private investigator, I was asked to locate any hotels that were using the client's hotel images on websites. A reverse image search identified dozens of companies using licensed photos without authorization. This likely led to civil litigation. More recently, a federal agent asked me to assist with a human trafficking case. He had a woman in custody that spoke little English. She was arrested during a prostitution sting and was suspected of being a victim of trafficking. A reverse image search from one online prostitution ad located all of her other ads which identified the regional areas that she had recently been working, a cellular telephone number connected to her pimp, and approximate dates of all activity.

Pictriev (pictriev.com)

Pictriev is a service that will analyze a photo including a human face and try to locate additional images of the person. The results are best when the image is of a public figure with a large internet presence, but it will work on lesser-known subjects as well. An additional feature is a prediction of the sex of the target as well as age.

Twitter Images

For the first several years of Twitter's existence, it did not host any photos on its servers. If a user wanted to attach a photo to his or her post, a third-party photo host was required. These have always been free and plentiful. Often, a shortened link was added to the message, which forwarded to the location of the photo. Twitter now hosts photos used in Twitter posts, but third-party hosts are still widely used. The majority of the images will be hosted on Instagram, which was explained previously. If you have already identified your target's Twitter page, you will probably have the links you need to see the photos uploaded with his or her posts.

Many Twitter messages have embedded images directly within the post. Twitter now allows you to search keywords for photo results. After you conduct any search within the native Twitter search field, your results will include a filter menu on the top. The "Photos" results will only include images which have a reference to the searched keyword within the message or hashtag. You can also filter this search for people, videos, or news.

Photo-Sharing Sites

In order to find a photo related to a target, the image must be stored on a website. The most common type of storage for online digital photos is on a photo-sharing site. These sites allow a user to upload photographs to an account or profile. These images can then be searched by anyone with an internet connection. Almost all of these hosts are free for the user and the files will remain on the site until a user removes them. There are dozens of these services, many allowing several gigabytes worth of storage. While I mentioned earlier that a Google Images or Bing Images search was most appropriate for all photo sharing hosts, Flickr deserves a mention.

Flickr (flickr.com)

Flickr, purchased by Yahoo and now owned by SmugMug, was one of the most popular photo-sharing sites on the internet. Many have abandoned it for Twitter and Instagram, but the mass amount of images cannot be ignored. The majority of these images are uploaded by amateur photographers and contain little intelligence to an investigator. Yet there are still many images in this "haystack" that will prove to be beneficial to the online researcher. The main website allows for a general search by topic, location, username, real name, or keyword. This search term should be as specific as possible to avoid numerous results. An online username will often take you to that user's Flickr photo album.

After you have found either an individual photo, user's photo album, or group of photos by interest, you can begin to analyze the profile data of your target. This may include a username, camera information, and interests. Clicking through the various photos may produce user comments, responses by other users, and location data about the photo. Dissecting and documenting this data can assist with future searches. The actual image of these photos may give all of the intelligence desired, but the data does not stop there. A search on Flickr for photographs related to the Occupy Wall Street protesters returned over 157,000 results.

Flickr Map (flickr.com/map)

Flickr attempts to geo locate all of the photos that it can. It attempts to identify the location where the photo was taken. It will usually obtain this information from the Exif data, which will be discussed in a moment. It can also tag these photos based on user provided information. Flickr provides a mapping feature that will attempt to populate a map based on your search parameters. I believe this service is only helpful when you are investigating a past incident at a large event or researching the physical layout of a popular attraction.

Flickr API

There are three specific uses of the Flickr Application Programming Interface (API) that I have found helpful during many online investigations. The first queries an email address and identifies any Flickr accounts associated with it. The second queries a username, and identifies the Flickr user number of the connected account. The final option queries a Flickr user number and identifies the attached username. Unfortunately, all of these features require a Flickr API key. I have included a throwaway key within the search tools explained at the end of the chapter. However, it may not function for long after the book is published. If the key should be terminated by Flickr, simply request your own free key at https://www.flickr.com/services/api/. Once issued, replace my test key (27c196593dad58382fc4912b00cf1194) within the code of the tools to your own. A demonstration may help to explain the features. First, I submitted the following URL to Flickr in order to query my target email address of test@test.com.

https://api.flickr.com/services/rest/?method=flickr.people.findByEmail&api_key=27c196593dad58382fc4912b00cf1194&find_email=test@test.com

I immediately received the following result.

user id="8104823@N02"
username>intellectarsenal

I now know that my target possesses a Flickr account associated with the email address, the username for the account, and the unique user number which will never change. Next, assume that we only knew the username. The following URL could be submitted.

https://api.flickr.com/services/rest/?method=flickr.people.findByUsername&api_key=27c196593dad58382fc4912b00cf1194&username=intellectarsenal

The response includes the following.

user id="8104823@N02"

Once you have identified the user number from either previous methods, we can submit the following URL.

https://api.flickr.com/services/rest/?method=flickr.people.getInfo&api_key=27c196593dad58382fc4912b00cf1194&user_id=8104823@N02

This returns the most details, including the following result from our target.

username>intellectarsenal
photosurl>https://www.flickr.com/photos/8104823@N02/
profileurl>https://www.flickr.com/people/8104823@N02/
mobileurl>https://m.flickr.com/photostream.gne?id=8084475

Navigating to the profile displays details such as the join date, followers, and photo albums. This may seem like a lot of work for a minimal amount of details, but this is quite beneficial. There is no native email address search on Flickr, but we can replicate the function within the API. You may not find young targets sharing images here, but the massive collection of photos spanning the past decade may present new evidence which was long forgotten by the target.

Exif Data

Every digital photograph captured with a digital camera possesses metadata known as Exif data. I have already explained several applications which extract this data from documents and images, but we need to have a better understanding of the technology. This is a layer of code that provides information about the photo and camera. All digital cameras write this data to each image, but the amount and type of data can vary. This data, which is embedded into each photo "behind the scenes", is not visible by viewing the captured image. You need an Exif reader, which can be found on websites and within applications. Keep in mind that some websites remove or "scrub" this data before being stored on their servers. Facebook, for example, removes the data while Flickr does not. Locating a digital photo online will not always present this data. If you locate an image that appears full size and uncompressed, you will likely still have the data intact. If the image has been compressed to a smaller file size, this data is often lost. Any images removed directly from a digital camera card will always have the data. This is one of the reasons you will always want to identify the largest version of an image when searching online. The quickest way to see the information is through an online viewer.

Jeffrey's Exif Viewer (exif.regex.info/exif.cgi)

I consider Jeffrey's Exif Viewer the online standard for displaying Exif data. The site will allow analysis of any image found online or stored on a drive connected to your computer. The home page provides two search options. The first allows you to copy and paste an address of an image online for analysis. Clicking "browse" on the second option will open a file explorer window that will allow you to select a file on your computer for analysis. The file types supported are also identified on this page. The first section of the results will usually provide the make and model of the camera used to capture the image. Many cameras will also identify the lens used, exposure settings, flash usage, date and time of capture, and file size. In one example, I could see that the camera used was a Canon EOS Digital Rebel with an 18 - 55mm lens at full 55mm setting. Auto exposure was selected, the flash was turned off, and the photo was taken at 2:30 pm on May 7, 2011. Not all of this will be vital for the researcher, but every bit of intelligence counts.

Scrolling down the analysis page will then identify many camera settings that probably provide little information to the researcher. These include aperture information, exposure time, sharpness, saturation, and other image details. Mixed in with this data is the serial number field. This is most common in newer SLR cameras and will not be present in less expensive cameras. These cameras usually identify the make, model, and serial number of the camera inside every photo that they capture. A serial number of a camera associated with an image can be valuable data. This can help an analyst associate other photos found with a target's camera. If an "anonymous" image was found online that included a serial number in the Exif data, and another image was found of a target of the investigation, these two photos can be analyzed. If the serial number as well as make and model of camera match, there is a good likelihood that the same camera took both images. However, it is important to know that this data can be manipulated. Using software such as ExifTool, a user can modify this data. While this is not a popular tactic to use, it is still possible. The difficult part of this is finding photos only knowing the serial number. Two services will help with that.

Camera Trace (cameratrace.com/trace)

This site was designed to help camera theft victims with locating their camera if it is being used by the thief online. For that use, you would find a photo taken with the stolen camera, and drop it into the previous site for analysis. This analysis identifies a serial number if available. If one is located, type the serial number into Camera Trace. It will attempt to locate any online photographs taken with the camera. This service claims to have indexed all of Flickr and 500px with plans to add others. A sample search using a serial number of "123" revealed several results. The website urges users to sign up for a premium service that will make contact if any more images appear in the database, but I have never needed this.

GPS

Many new SLR cameras, and almost all cellular telephone cameras, now include GPS. If the GPS

is on, and the user did not disable geo-tagging of the photos in the camera settings, you will get location data within the Exif data of the photo. Figure 20.04 (left) displays the analysis of an image taken with a camera with GPS. The data is similar to the previous analysis, but includes a new "Location" field. This field will translate the captured GPS coordinates from the photo and identify the location of the photo. Farther down this results page, the site will display an image from Google Maps identifying the exact point of the GPS associated with the photo. Figure 20.04 (right) displays this satellite view including a direction identifier. Since most cellular telephones possess an accelerometer, the device documents the direction the camera was facing. Most Android and iPhone devices have this capability. Your results will vary depending on the user's configuration of their GPS on the device.

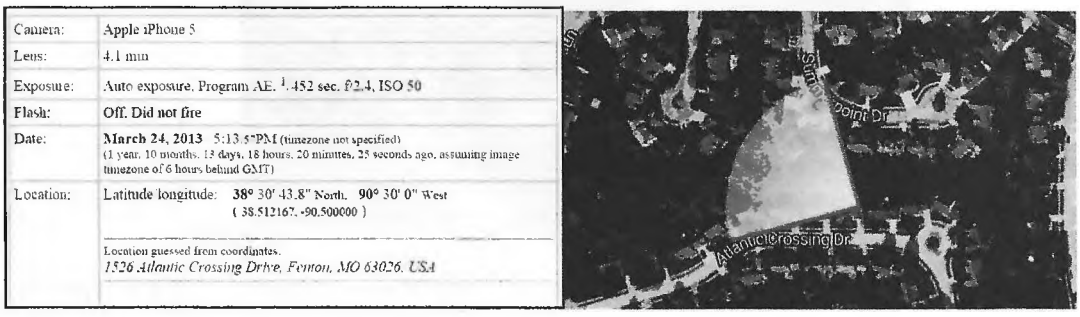


Figure 20.04: A Jeffrey's Exif Viewer result identifying location with map view.

Cropped Images

Another piece of information that we can look for inside the Exif data is the presence of a thumbnail image within the photograph. Digital cameras generate a small version of the photo captured and store it within the Exif data. This icon size image adds very little size to the overall file. When a user crops the image, this original smaller version may or may not get overwritten. Programs such as Photoshop or Microsoft Photo Editor will overwrite the data and keep both images identical. Other programs, as well as some online cropping tools, do not overwrite this data. The result is the presence of the original and uncropped image within the Exif data of the cropped photo. An example of this is seen in Figure 20.05. A cropped photo found online is examined through Jeffrey's Exif viewer. The cropped full-size large photo is seen on the left. The embedded smaller original photo was not overwritten when cropped. We can now see what the image looked like before it was cropped. This technique has been used by police to identify child pornography manufactures. These pedophiles will crop themselves out of illegal images to avoid identification. When photos of the children are found by police, an original uncropped image may be enough to identify and prosecute a molester. This is not limited to law enforcement. Some tech-savvy fans of television personality Catherine Schwartz examined a cropped photo on her blog in 2003. Inside the Exif data was the uncropped version which exposed her breasts and quickly made the rounds through the internet. We must remember this unfortunate lesson when we consider posting our own content to the internet.

Real World Application: In a civil litigation, a subject claimed an injury that prohibited him from work, walking, and a normal life. The suit claimed damages from pain and suffering and sought a monetary judgment for future lack of ability to work. A brief scan of the subject's online photo album revealed fishing trips, softball games, and family adventure vacations. With Exif information data intact, exact dates, times, locations, and cameras were identified and preserved. The subject withdrew his lawsuit.

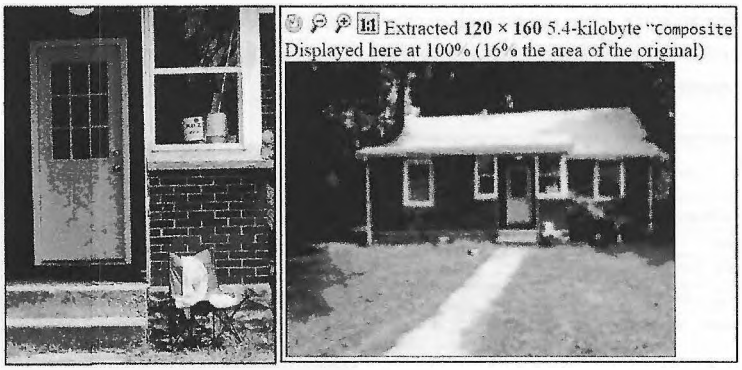


Figure 20.05: A Jeffrey's Exif Viewer summary result displaying an original uncropped photo.

Online Barcode Reader (online-barcode-reader.inliteresearch.com)

Barcodes have been around for decades. They are the vertical lined images printed on various products that allow registers to identify the product and price. Today's barcodes are much more advanced and can contain a large amount of text data within a small image. Some newer barcodes exist in order to allow individuals to scan them with a cell phone. The images can provide a link to a website, instructions for downloading a program, or a secret text message. I generally advise against scanning any unknown barcodes with a mobile device since malicious links could be opened unknowingly. However, an online barcode reader can be used to identify what information is hiding behind these interesting images.

Figure 20.06 displays the barcode search options from Online Barcode Reader. These include 1D, PDF417, Postal, DataMatrix, QR, and ID barcodes. After selecting the type of barcode image, you can select any PDF, TIFF, JPEG, BMP, GIF, or PNG file on your computer up to 4Mb in size. This could be a photo that possesses a barcode in the content or a digital code downloaded from a website. Screen captures of codes also work well. While sitting on a plane with Wi-Fi, I captured a photo of an abandoned boarding pass in the magazine holder in front of me. The barcode reader identified text information stored inside the code that was not present in text on the document.

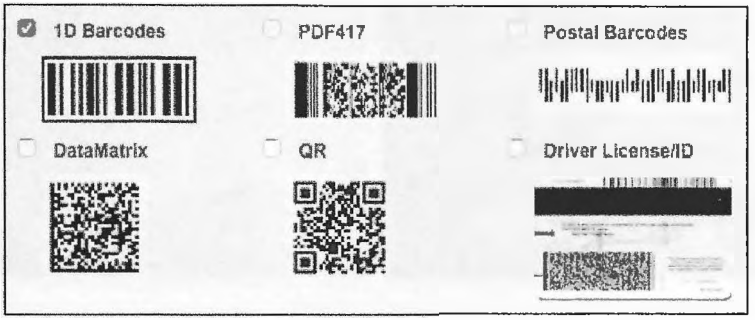


Figure 20.06: Barcode input samples from Online Barcode Reader.

Image Manipulation

It is common to find images on the internet that have been manipulated using software such as Photoshop. Often it is difficult, if not impossible, to tell if these photos have been manipulated by visually analyzing them. A handful of websites use a technique to determine not only if the photo has been manipulated, but which portions of the photo have changed. One site offers the following explanation of how the technology works.

"Error level analysis (ELA) works by intentionally resaving the image at a known error rate, such as 95%, and then computing the difference between the images. If there is virtually no change, then the cell has reached its local minima for error at that quality level. However, if there is a large amount of change, then the pixels are not at their local minima and are effectively original."

Foto Forensics (fotoforensics.com)

This site allows you to upload a digital image. After successful upload, it will display the image in normal view. Below this image will be a darkened duplicate image. Any highlighted areas of the image indicate a possible manipulation. While this site should never be used to definitively state that an image is untouched or manipulated, investigators may want to conduct an analysis for intelligence purposes only. Figure 20.07 displays original and manipulated images while Figure 20.08 displays the analysis of the images from Foto Forensics. This site will provide an analysis of an image from the internet or a file uploaded from a computer. It is important to note that any images uploaded become part of the website's collection and a direct URL is issued. While it would be difficult for someone to locate the URL of the images, it could still pose a security risk for sensitive photographs.



Figure 20.07: An original photograph (left) compared to a manipulated photograph (right).

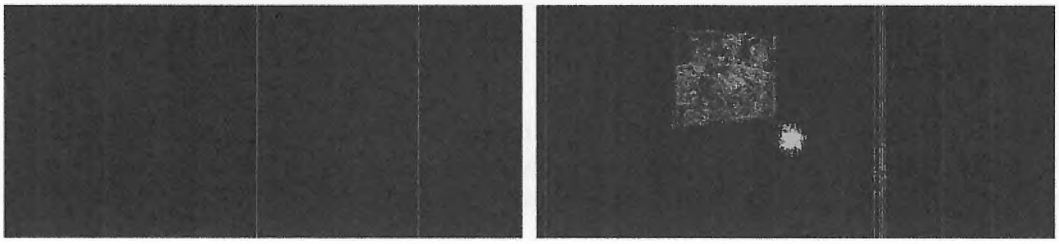


Figure 20.08: The original photograph (left) and manipulated image (right) on Foto Forensics.

Forensically (29a.ch/photo-forensics)

Forensically is a robust image analyzer that offers a huge collection of photo forensic tools that can be applied to any uploaded image. This type of analysis can be vital when image manipulation is suspected. Previous tools have offered one or two of the services that Forensically offers, but this new option is an all-in-one solution for image analysis. Loading the page will present a demo image, which is used for this explanation. Clicking the "Open File" link on the upper left will allow upload of an image into your browser for analysis. Images are NOT uploaded to the server of this tool, they are only brought into your browser locally. Figure 20.09 (left) is the standard view of a digital photo. The various options within Forensically are each explained and example images are included. Due to the black & white environment of this book, I have replicated all of this instruction in color on my blog at the following address.

https://inteltechniques.com/blog/2016/12/21/internet-search-resource-foresically/

The Magnifier allows you to see small hidden details in an image. It does this by magnifying the size of the pixels and the contrast within the window. There are three different enhancements available at the moment: Histogram Equalization, Auto Contrast, and Auto Contrast by Channel. Auto Contrast mostly keeps the colors intact; the others can cause color shifts. Histogram Equalization is the most robust option. You can also set this to none.

The Clone Detector highlights copied regions within an image. These can be a good indicator that a picture has been manipulated. Minimal Similarity determines how similar the cloned pixels need to be to the original. Minimal Detail controls how much detail an area needs; therefore, blocks with less detail than this are not considered when searching for clones. Minimal Cluster Size determines how many clones of a similar region need to be found in order for them to show up as results. Blocksize determines how big the blocks used for the clone detection are. You generally don't want to touch this. Maximal Image Size is the maximal width or height of the image used to perform the clone search. Bigger images take longer to analyze. Show Quantized Image shows the image after it has been compressed. This can be useful to tweak Minimal Similarity and Minimal Detail. Blocks that have been rejected because they do not have enough detail show up as black. Figure 20.09 (right) demonstrates this output.



Figure 20.09: A normal image view (left) and Clone Detector (right) in Forensically.

Error Level Analysis compares the original image to a recompressed version. This can make manipulated regions stand out in various ways. For example, they can be darker or brighter than similar regions which have not been manipulated. JPEG Quality should match the original quality of the image that has been photoshopped. Error Scale makes the differences between the original and the recompressed image bigger. Magnifier Enhancement offers different enhancements: Histogram Equalization, Auto Contrast, and Auto Contrast by Channel. Auto Contrast mostly keeps the colors intact; the others can cause color shifts. Histogram Equalization is the most robust option. You can also set this to none. Opacity displays the opacity of the Differences layer. If you lower it you will see more of the original image. Figure 20.10 (left) displays manipulation.

Noise Analysis is basically a reverse de-noising algorithm. Rather than removing the noise it removes the rest of the image. It is using a super simple separable median filter to isolate the noise. It can be useful for identifying manipulations to the image like airbrushing, deformations, warping, and perspective corrected cloning. It works best on high quality images. Smaller images tend to contain too little information for this to work. Noise Amplitude makes the noise brighter. Equalize Histogram applies histogram equalization to the noise. This can reveal things but it can also hide them. You should try both histogram equalization and scale to analyze the noise. Magnifier Enhancement offers three different enhancements: Histogram Equalization, Auto Contrast, and Auto Contrast by Channel. Auto Contrast mostly keeps the colors intact; the others

can cause color shifts. Histogram Equalization is the most robust option. You can also set this to none. Opacity is the opacity of the noise layer. If you lower it you will see more of the original image. The result can be seen in Figure 20.10 (right).

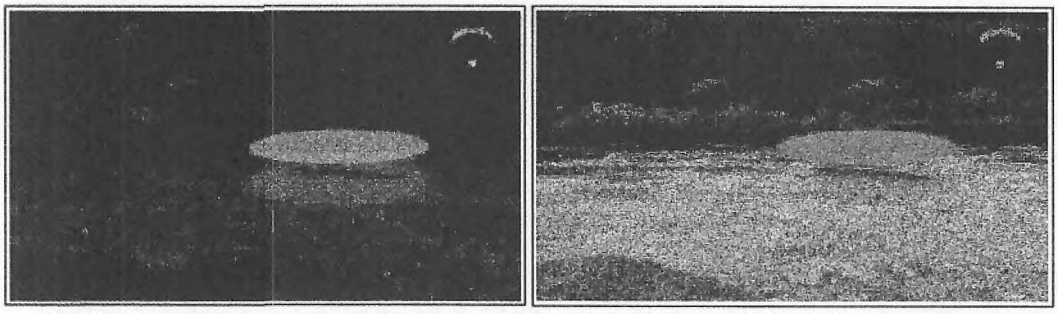


Figure 20.10: Error Level Analysis (left) and Noise Analysis (right) in Forensically.

Level Sweep allows you to quickly sweep through the histogram of an image. It magnifies the contrast of certain brightness levels. To use this tool simply move your mouse over the image and scroll with your mouse wheel. Look for interesting discontinuities in the image. Sweep is the position in the histogram to be inspected. You can quickly change this parameter by using the mouse wheel while hovering over the image, this allows you to sweep through the histogram. Width is the amount of values (or width of the slice of the histogram) to be inspected. The default should be fine. Opacity is the opacity of the sweep layer. If you lower it you will see more of the original image.

Luminance Gradient analyzes the changes in brightness along the x and y axis of the image. It's obvious use is to look at how different parts of the image are illuminated in order to find anomalies. Parts of the image which are at a similar angle (to the light source) and under similar illumination should have a similar color. Another use is to check edges. Similar edges should have similar gradients. If the gradients at one edge are significantly sharper than the rest it's a sign that the image could have been copied and pasted. It does also reveal noise and compression artifacts quite well. Figure 20.11 (left) displays this view.

PCA performs principal component analysis on the image. This provides a different angle to view the image data which makes discovering certain manipulations and details easier. This tool is currently single threaded and quite slow when running on big images. Choose one of the following Modes: Projection of the value in the image onto the principal component; Difference between the input and the closest point on the selected principal component; Distance between the input and the closest point on the selected principal component; or the closest point on the selected principal Component. There are three different enhancements available: Histogram Equalization, Auto Contrast, and Auto Contrast by Channel. Auto Contrast mostly keeps the colors intact; the others can cause color shifts. Histogram Equalization is the most robust option.

You can also set this to none. Opacity is the opacity of the sweep layer. If you lower it you will see more of the original image. Figure 20.11 (right) displays this view.

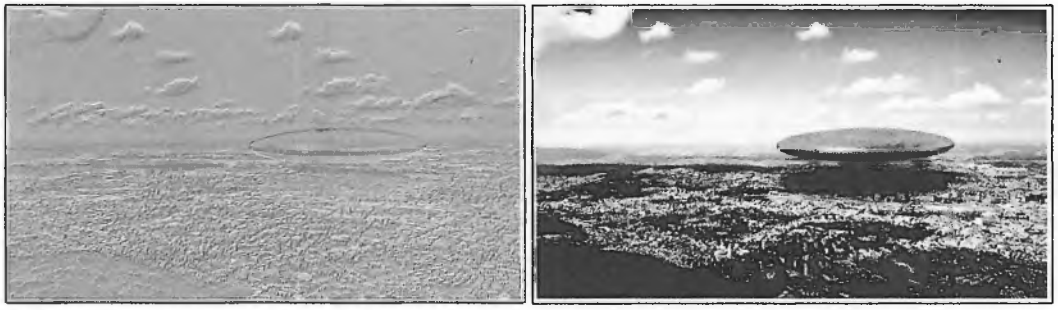


Figure 20.11: The Luminance analysis (left) and PCA analysis (right) within Forensically.

MetaData displays any Exif metadata in the image. Geo Tags shows the GPS location where the image was taken, if it is stored in the image. Figure 20.12 displays the result.

Thumbnail Analysis shows any hidden preview image inside of the original image. The preview can reveal details of the original image or the camera used. Figure 20.13 displays the online image (left) while the original thumbnail displays a different view (right).

Make	SONY
Model	ILCE 6000
Orientation	1
XResolution	300
YResolution	300
ResolutionUnit	2
Software	darktable 1.6.6
ModifyDate	2015:08:14 13:32:39
YCbCrPositioning	2
Rating	1
RatingPercent	20
DateTimeOriginal	Thu Jul 31 2014 09:05:43 GMT-0700 (PDT)
GPSVersionID	2,2,0,0
GPSLatitudeRef	N
GPSLatitude	47.3500
GPSLongitudeRef	E
GPSLongitude	8.4980

GPSVersionID	2,2,0,0
GPSLatitudeRef	N
GPSLatitude	47.35
GPSLongitudeRef	E
GPSLongitude	8.498

- View on OpenStreepMap
- View on Google Maps
- Other Images around here on Flickr



Figure 20.12: Metadata from Forensically.



Figure 20.13: An online image (left) and original thumbnail image (right) on Forensically.

The next time you identify a digital image as part of your online investigation, these tools will peek behind the scenes and may display evidence of tampering.

Bulk Image Download

Now that you have read about the various ways to locate images relevant to your investigation, you will likely encounter a situation which presents a large number of photos. I previously mentioned browser extensions which help automate this process, but they have limits. Loading an entire Subreddit or Twitter search can quickly exceed the resources available in your browser. Instead, we should consider an automated solution. I prefer Ripme (github.com/RipMeApp). This application requires Java to be installed on your machine. I refuse to install Java on any Windows or Mac host due to security concerns, as it adds an additional layer of vulnerability. However, I have no objection to installing it within a Linux virtual machine. Execute your OSINT Master VM and conduct the following steps.

- In Terminal, enter `sudo apt install default-jre` and click Y when prompted.
- Download "ripme.far" from github.com/ripmeapp/ripme/releases to "Downloads".
- Right-click this file and choose "Properties".
- In the Permissions tab, select "Allow executing file…".

You can now double-click this file to launch the program. Enter any target URL and click "Rip". Figure 20.15 (left) displays extraction of the OSINT Subreddit URL. The application downloaded all of the images to the "rips" folder in my Downloads. Figure 20.15 (right) extracted all of the images from a search of "OSINT" on Twitter. This application supports image downloads from Imgur, Twitter, Tumblr, Instagram, Flickr, Photobucket, and Reddit.

IntelTechniques Images Search Tool

I do not recommend manually typing this all into a web browser. It would be more efficient to navigate to each image search site and paste the photo URL. However, I have created an online

tool that automates this entire process. The first field allows input of the entire address of an online image. The Populate All button will supply this data to all fields. The next search options replicate the techniques explained here for Google, Bing, TinEye, Yandex, and Baidu. The final option on this page executes the above searches across all five networks into five separate tabs on your browser. Figure 20.14 displays the current state of this tool.

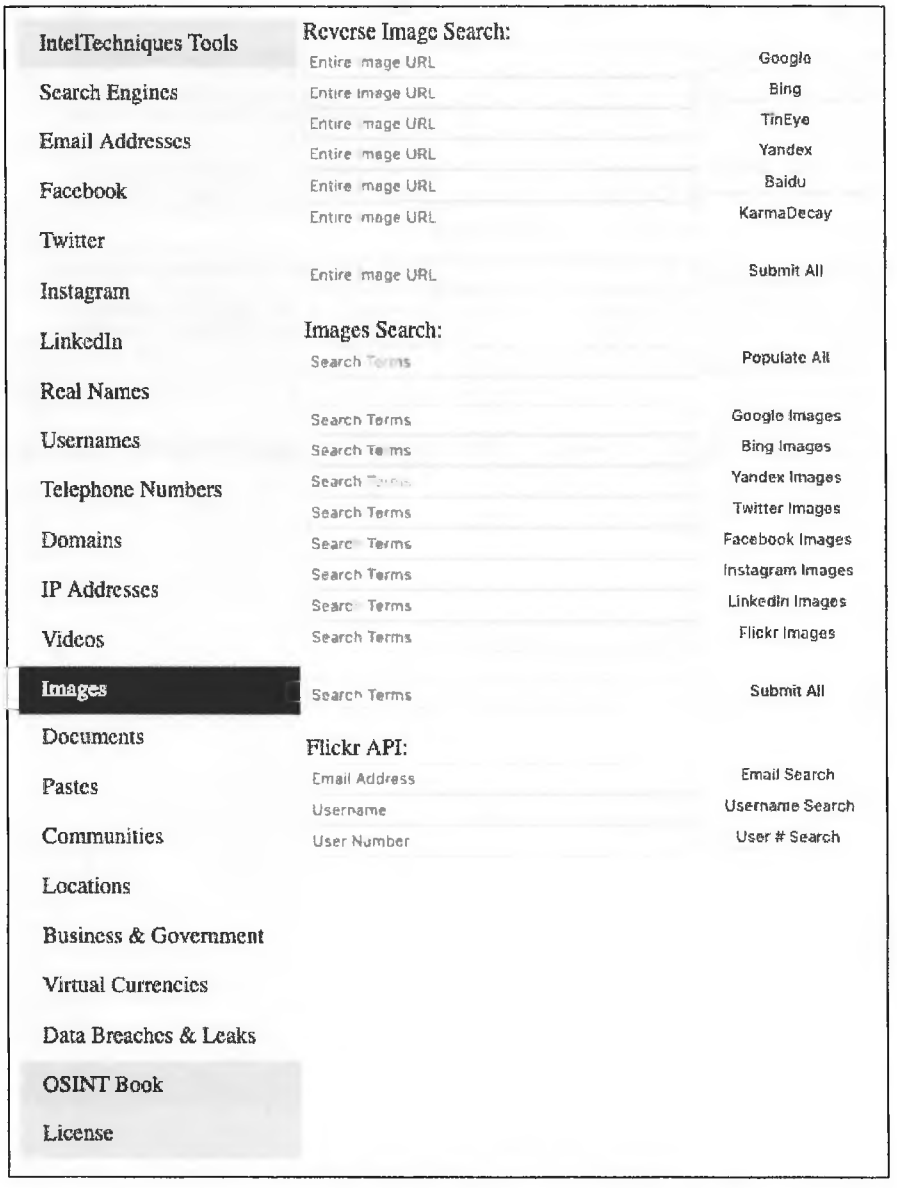


Figure 20.14: The IntelTechniques Images Search Tool.



Figure 20.15: The Ripme application extracting bulk images from Reddit and Twitter.

CHAPTER TWENTY-ONE
VIDEOS

Online videos are almost as common as online photographs. The cameras in smart data phones can act as video cameras. In some situations, uploading a video to the internet is easier than a photograph. Most social networks now act as independent video hosts for their platforms. Video sharing sites such as YouTube have made video publication effortless. For investigations, a video can contain a huge amount of intelligence. When any abnormal event happens, people flock to their phones and start recording. These videos may capture criminal acts, embarrassing behavior, or evidence to be used in a civil lawsuit. Obtaining these videos is even easier than creating them.

YouTube (YouTube.com)

The most popular of several video-sharing sites is YouTube. The official YouTube site declares that 500 hours of video are uploaded every minute, resulting in nearly 80 years of content uploaded every day. It further states that over a billion videos are viewed each day. These impressive statistics confirm the need to include videos as part of a complete OSINT analysis. YouTube is easy to search from the main search field on every page. This field can accept any search term and will identify video content or username. Users that upload videos to YouTube have their own "channel". Their videos are uploaded to this channel, and locating a user's channel will identify the videos uploaded by that user.

Many people use YouTube as a social network, leaving comments about videos and participating in discussions about various topics. If you locate a video of interest, it is important to also retrieve this text information. Each comment below a video will include the username that created the comment, which will link to that user's profile.

A search for "school bus fight" returned over 500,000 video links on YouTube. Adding a search term such as the city or school name may help, but it may also prohibit several wanted videos from appearing. The "filters" option can be expanded to help limit the search scope. This button is above the first video result. This provides additional filter options including the ability to sort by the upload date (date range), type (video vs. channel), duration (short or long), and features (video quality). In the "school bus fight" example, the "uploaded this week" option was chosen. This resulted in only 700 videos which could easily be examined for any intelligence. The lower left portion of any video page includes a link to the profile of the user who submitted this video. This profile page includes all of the videos uploaded by that user and additional profile information. Several YouTube "hacks" have surfaced over the years. Many of these stopped working as YouTube made changes to the environment. Of those still functioning, I find the following techniques helpful to my investigations.

Bypass Age and Login Restriction

Several YouTube videos have been tagged as violent, sexual, or otherwise inappropriate for young viewers. Others demand that you log in to a Google account in order to view the content for unclear reasons. Either way, this is an unnecessary roadblock to your investigation. As an OSINT investigator, I prefer to not be logged in to any personal or covert account while I am researching. Any time you are searching through a Google product while logged in to an account, Google is documenting your every move. This can be unsettling. One easy technique should remove this restriction. Navigate to the following website and notice the inability to view the video. If you are not logged in to a Google account with a verified age, you should see a warning about mature content. This video cannot be played.

https://www.youtube.com/watch?v=SZqNKAd_gTw

In this example, the YouTube Video ID is "SZqNKAd_gTw". In order to view this video through YouTube without a third-party service, and without supplying the credentials for your personal Google account, you can generate the following URL. Replace "SZqNKAd_gTw" with the ID of your target video. The result will be the restricted video in full screen view. Some users report that this technique will also bypass videos that have a viewing country restriction.

https://www.youtube.com/embed/SZqNKAd_gTw

If this technique should ever stop working, you can also use a non-Google service to achieve the same result. Navigate to the following website to access this same video (SZqNKAd_gTw).

http://www.nsfwyoutube.com/watch?v= SZqNKAd_gTw

Notice that all of the addresses are very similar. This final link will take you to NSFWYouTube, a third-party website, which will also remove the proof of age requirement. Please be warned that the content in this example contains very disturbing video, hence the blockage by YouTube.

Bypass Commercials with Full Screen

It seems lately that every long YouTube video I play possesses a 30 second commercial at the beginning. This is very frustrating when analyzing a large number of videos. The same URL trick will bypass this annoyance. Navigate to the following address and notice the commercial at the beginning.

http://www.youtube.com/watch?v=IEIWdEDFlQY

Alter this address slightly in order to force the video to play in full screen in your browser. This will also bypass any commercials. The URL should appear like the following.

https://www.youtube.com/embed/IEIWdEDFlQY

Display Thumbnails of Videos

When a user uploads a video, YouTube captures and displays a frame for that media. This is the still frame you see when searching videos and before a video is played. These possess a static URL, which will be helpful when we discuss reverse video searching. As an example, navigate to the following address to load a demo video.

https://www.youtube.com/watch?v=1nm1jEJmOTQ

Using that same video ID, navigate to the following address to view the main still frame. This is the image visible when a video is loaded within YouTube before playing.

https://i.ytimg.com/vi/1nm1jEJmOTQ/hqdefault.jpg

The address that displayed the main image is not your only option. An additional high-resolution image can usually be extracted from this specific video with the following address.

https://i.ytimg.com/vi/1nm1jEJmOTQ/maxresdefault.jpg

Furthermore, we can extract four unique frames with the following URLs.

https://img.youtube.com/vi/1nm1jEJmOTQ/0.jpg
https://img.youtube.com/vi/1nm1jEJmOTQ/1.jpg
https://img.youtube.com/vi/1nm1jEJmOTQ/2.jpg
https://img.youtube.com/vi/1nm1jEJmOTQ/3.jpg

In a moment, our tools will query all of these images for download and reverse image searching.

Identify and Bypass Country Restriction

Many videos on YouTube are allowed to be viewed in some countries and blocked in others. If you encounter a video that will not play for you because of a country restriction, you have options. Before proceeding, consider identifying from which geographical areas a video is restricted. For this, we will rely on the polsy.org.uk country restriction checker. After you have identified a video with possible country restrictions, paste the video ID into the following URL. Note that the following video ID (MGhMdT_C-vQ) is blocked in the U.S.

http://polsy.org.uk/stuff/ytrestrict.cgi?ytid=MGhMdT_C-vQ

The result is a page with a world map. Countries in grey are allowed to view the target video while countries in red are not. While I cannot natively play this video due to my location, I can easily

view the default and high resolution still frames with the technique described in the previous section. The following exact URLs display content otherwise not viewable.

https://i.ytimg.com/vi/MGhMdT_C-vQ/hqdefault.jpg
https://i.ytimg.com/vi/MGhMdT_C-vQ/maxresdefault.jpg

If a video is blocked from playing in your location, you can use a third-party proxy that should allow viewing. In both of these examples, MGhMdT_C-vQ is the target video that is blocked. The following URL should play the video.

https://clipmega.com/watch?v=MGhMdT_C-vQ

An alternative to this is HookTube, which would be the following URL for this video.

https://hooktube.com/watch?v=MGhMdT_C-vQ

The URL below will not stream the video, but should allow you to download the content.

https://en.savefrom.net/#url=http://youtube.com/watch?v=MGhMdT_C-vQ

Metadata and Reverse Image Search (citizenevidence.amnestyusa.org)

Most of the details of a YouTube video can be seen on the native page where the video is stored. Occasionally, some of this data may not be visible due to privacy settings or profile personalization. In order to confirm that you are retrieving all possible information, you should research the data visible from YouTube's servers. The easiest way to do this is through the YouTube Data Viewer from Amnesty International. Select any YouTube video and copy the entire address of the page. Paste the video address into this service. The result will include the video title, description, upload date, upload time, still images, and an option to conduct a reverse image search. Clicking the "reverse image search" option next to a still frame opens a reverse image search on Google for the selected image. While this works well on YouTube videos, complete reverse video searching across multiple networks will be explained later in this chapter.

Immediate Download Options

My preferred method for extracting YouTube videos was explained in Chapter Four while discussing YouTube-DL. However, if you have no software or browser plugins available to you, there is another easy option. While you are watching any YouTube video, you can add the letters "PWN" to the beginning of the address in order to download the video to your computer. To test this, navigate to http://www.youtube.com/watch?v= OmZyrynlk2w.

Now, add "PWN to the beginning, as indicated in the following address.

http://www.pwnyoutube.com/watch?v= OmZyrynlk2w

You will be presented a new page with many options including the ability to download the video; download only the audio; convert the video to a different format; and bypass the age restriction as discussed earlier. Additional options include yout.com, keepvid.com, and y2mate.com.

YouTube Comment Scraper (ytcomments.klostermann.ca)

While the video media of your target is essential to your investigation, the commentary associated with the posted content is equally important. Screen captures can usually identify the bulk of the comments attached to a video, but there are hurdles. Some viral videos will have thousands of comments, which will not appropriately fit into a screen capture. Other videos may have comments to comments, and those may need to be expanded manually in order to observe and collect the evidence. The solution to this is the YouTube Comment Scraper. This website allows easy use of open-source code available to extract YouTube comments from your own server. Simply provide the URL of your target video and allow the tool to process all data. The result will be a CSV spreadsheet with every comment text, including comment ID, timestamp, username, replies, and likes. This can all be viewed within the results screen or downloaded (my preference). Figure 21.01 displays a partial collection of captured data.

user	date	timestamp	commentText
Alois Hebenstreit	2 days ago	1514316998089	music is crazzy
James Lee	1 week ago	1513884998091	Finally a band that speaks the truth <3 Billy Talent
Yukio Nana	1 week ago	1513884998094	Tobito REEEEEEEEEEEEEEEEEEEEEEE
GERHARDTJEAN	1 week ago	1513884998096	Chapado e escutando isso!!!! Mas é bom demais!!!!

Figure 21.01: Data collected by YouTube Comment Scraper.

I believe that this type of documentation should be a part of every investigation associated with a YouTube video. If you ever find yourself downloading a target video, immediately jump to this free resource and collect the comments. If the video is actively discussed, you may need to return often and download new comments. If you use a program such as Excel, it will be easy to remove duplicates from all of your captures, leaving only the unique comments relative to your investigations.

Deleted Videos

It has become very common for people to remove their YouTube videos. This often happens when unwanted attention is generated and the user regrets the post. The following technique will not always work, but I have had many successes within my own investigations. Consider the following video which has been removed from YouTube.

https://www.youtube.com/watch?v=9ZmsnTDLykk

The Internet Archive has been mirroring YouTube videos for years and often possesses their own independent copies. We can look for this with the following URL.

https://web.archive.org/web/https://www.youtube.com/watch?v=9ZmsnTDLykk

This identifies 276 captures of this video page. However, these are HTML archives, and the video will not play within any of them. These pages are beneficial for locating comments associated with your target video, but not for the video content itself. We can use the following URL to play the full resolution version of the archived video.

https://web.archive.org/web/2oe_/http://wayback-fakeurl.archive.org/yt/9ZmsnTDLykk

We can now right-click on the video to save our own offline copy. You would only need to replace your target YouTube video ID with the one listed here (9ZmsnTDLykk). Our search tools will replicate this for you at the end of the chapter.

Reverse Video Searching

There was a brief mention earlier of conducting a reverse image search on the still captures of a YouTube video. This would use the same techniques as mentioned in Chapter Twenty for images. While there is no official reverse video search option, applying the techniques to still captures of videos can provide amazing results. This method is not limited to YouTube. We can conduct reverse image searches on videos from many sites. As the popularity of online videos is catching up to images, we must always consider reverse video searches. They will identify additional websites hosting the target videos of interest. Before explaining the techniques, consider the reasons that you may want to conduct this type of activity.

- School resource officers and personnel are constantly notified of inappropriate video material being posted online. Identifying these videos, consisting of fights, malicious plans, and bullying, may be enough to take care of the situation. However, identifying the numerous copies on other websites will help understand the magnitude of the situation.
- Global security divisions monitoring threats from protest groups will likely encounter videos related to these activities. However, identifying the numerous copies on various websites will often disclose commentary, online discussions, and additional threats not seen in the primary source video.
- Human trafficking investigators can now reverse search videos posted to escort providers similar to the now seized site Backpage. The presence of identical videos posted to numerous geographical areas will highlight the travel and intentions of the pimps that are broadcasting the details.

There are limitless reasons why reverse video searching should be a part of your everyday research. The following methods will get you started with the most popular video websites. These techniques can be replicated as you find new services of interest.

YouTube: As explained earlier, YouTube offers four still frames for every video uploaded plus a high resolution image. Obtain the URLs outlined during that instruction, and provide each to Google, Bing, Yandex, TinEye, and Baidu as a reverse image search as explained previously.

Facebook: Identifying Facebook screen captures does not require access to their API. Instead, we will view the source code of an individual video page. For this example, assume that you are viewing a video embedded into a Facebook profile. You can right-click the video while playing and choose the option "Show Video URL". This will present a small box with the address of the Facebook video that is playing. In this example, we will use the following.

https://www.facebook.com/billytalent/videos/10153157582551992/

Right-click on this new page and choose to view the source code. Conduct a search of .jpg and copy the entire URL of the first search result. In this example, the page includes the following image address.

view-source:https://scontent-dfw5-1.xx.fbcdn.net/v/t15.0-10/10604952_10153157584261992_10153157582551992_3872_390_b.jpg?_nc_cat=103&_nc_oc=AQkUqbgkm_1fpQXLUZBvfkocTbDK0-N1G8Wo9h8uoqwDPgjQQcw_iPpkC5rIrVXPLvM&_nc_ht=scontent-dfw5-1.xx&oh=24f851aee5624bb73ed5a32e7d98cd42&oe=5E004E65

Navigating to that URL presents a large primary screen capture of the target video. Conducting a reverse image search through various services will identify additional websites that possess the same content. These will usually be social network profiles of the target's friends. In this example, a reverse search of this image reveals 43 additional pages of interest. Each contain the target video and comments about the content.

Instagram: Instagram can host videos in the same way that it possesses images. It will require us to view the source code of a target video. In this, you will see a descriptor of meta property="og:image" near the top of the page. The URL that follows this is the address of the primary screen capture for this video. In this example, it is the following address.

https://scontent-dfw5-1.cdninstagram.com/vp/44688d997b44479c11b47284488a5bf9/5D84CFB5/t51.2885-15/e35/69052351_1351875194964840_2315106562461635628_n.jpg?_nc_ht=scontent-dfw5-1.cdninstagram.com&_nc_cat=100

Vimeo: Vimeo does not natively offer URLs with a video ID that display screen captures of multiple frames. However, they do provide a single high definition still capture for every video. This is stored in the Application Programming Interface (API) side of Vimeo, but it is easy to

obtain. As an example, consider your target is at https://vimeo.com/99199734. The unique ID of 99199734 is assigned to that video. You can use that number to access the API view of the video at https://vimeo.com/api/oembed.json?url=https://vimeo.com/99199734. This address will display a text only page with the following content.

type: "video",
version: "1.0",
provider_name: "Vimeo",
provider_url: "https://vimeo.com/",
title: "Billy Talent 'Try Honesty'",
author_name: "S M T",
author_url: "https://vimeo.com/user10256640",
is_plus: "0",
html: "<iframe src="https://player.vimeo.com/video/99199734" width="480" height="272" frameborder="0" title="Billy Talent 'Try Honesty'" webkitallowfullscreen mozallowfullscreen allowfullscreen></iframe>",
width: 480,
height: 272,
duration: 247,
description: "Music Video Directed by Sean Michael Turrell.",
thumbnail_url: "https://i.vimeocdn.com/video/513053154_295x166.jpg",
thumbnail_width: 295,
thumbnail_height: 167,
upload_date: "2014-06-25 21:29:06",
video_id: 99199734,

The portion relevant to this topic is the thumbnail URL. Following that description is an exact address of the "medium" image used at the beginning of each Vimeo video. Removing the "_295x166.jpg" from the end of the URL presents a full size option, such as the following.

https://i.vimeocdn.com/video/513053154

We will use this static URL within our tools. Also note that this view identifies the exact date and time of upload. The video page view only identified the date as "1 Year Ago". A reverse image search of the thumbnail URL, using our video search tools in a moment, will produce additional websites which host the same (or similar) video.

Liveleak: Similar to Instagram, LiveLeak displays a meta property="og:image" line near the top of every video page's source code view. As an example, consider that your target is a bus fight video located at http://www.liveleak.com/view?i=9d4_1447939701. The source code of that page reveals the still image to be located at the following address.

https://cdn.liveleak.com/80281E/ll_a_u/thumbs/2015/Nov/19/d5e7c911321a_sf_5.jpg

Others: Repeating this process for every video sharing website can get redundant quickly. Instead, consider the following. Practically every online video possesses a still image that is displayed to represent the video before being played in search results. This image is likely a direct link that can be seen in the source code. Providing this image to various reverse image search websites will likely display additional copies of the target video within unknown websites.

These sources are not the only video sharing services on the internet. Wikipedia identifies dozens of these sites, but searching each of them can become tedious. These sites are no longer restricted to pages of video files with standard extensions such as mp4, mpg, and flv. Today, services such as Instagram allow embedded videos that do not conform to yesterday's video standards. Many new services present the viewer with animated gif files that only appear as true videos. Fortunately, search engines like Google and Bing offer a search across all of the types.

Google Videos (google.com/videohp)

A search on YouTube for "school bus fight" returned over 342,000 results. However, Google Videos returned 3 million results. These include the results identified in the previous YouTube search plus any videos from other sites that meet the search criteria. This will often lead to duplicate videos that have been posted by news websites and social networks. Google can filter these results by duration time, date and time captured, and video source. The top menu of any Google video results page will display these options. A Google search for the term "female street fight", including filters for videos with a short duration that were posted this week from any source, returned over 900 results. These results could either be further filtered with search terms or quickly viewed by still frame to determine relativity to the investigation.

Bing Videos (videos.bing.com)

One feature that makes Bing a favorite site for searching videos is the instant video playback option. When viewing a video search results page, simply hovering the cursor over the video still shot will start the video playback from the beginning of the video. This eliminates the need to navigate to each video page for playback to determine the value of the video. Bing also offers filtering by length and source. The "select view" toolbar at the top of each search result page will allow you to sort the results by either the best match or the most recent. Whether using Google or Bing to locate videos, I recommend turning off the safe search feature. This feature is designed to prohibit some videos with adult content from displaying. With investigations, it is often these types of videos that are wanted.

Downloading Facebook Videos

Facebook hosts the videos that users embed into their Facebook profiles. If you have already located a target's Facebook page, scrolling through the wall posts will likely present any videos that have been uploaded to the account. If you do not have a specific target and want to search these videos by keywords, this is possible through search engines. Google Videos will include

Facebook videos in its video searches. Including the term Facebook in the search will place an emphasis on videos located on Facebook pages. This will not present every Facebook video fitting the criteria of the search. Using the site operator as discussed in Chapter Eight, you can specify a search to include only videos found on Facebook profiles. All Facebook profiles are hosted on the main Facebook domain under a subcategory of video. For example, the address of a video on Facebook would look like this:

http://www.facebook.com/video/video.php?v=1537495389593

The numbers at the end are associated with one unique video. A custom search on Google for a specific video may look like this:

site:facebook.com/video "school bus fight"

This search would attempt to locate any videos stored on the Facebook video servers that included the terms "school bus fight". Most of the videos located should link to the Facebook video page and play without being logged in to a Facebook account. After you locate your video of interest, downloading the video can appear difficult. While Facebook does not offer a native download option, you can use their API view to easily archive any content. In the previous Facebook reverse video search discussion, we used the following video page as an example.

https://www.facebook.com/billytalent/videos/10153157582551992/

10153157582551992 is the video ID for this media. Entering the following address into a web browser will reveal a text-only view of the Facebook data for this video.

https://www.facebook.com/video/video_data/?video_id=10153157582551992

At the bottom of this page, you should see a line of code similar to the following.

Hd_source: "https://video-lax3-1.xx.fbcdn.net/v/t43.1792-2/10575621_10153157584221992_398195468_n.mp4?efg=eyJybHIiOjIxNDUsInJsYSI6MjMxNCwidmVuY29kZV90YWciOiJsZWdhY3lfaGQifQ%3D%3D&rl="

The entire URL can be copied into a web browser address field. The result will be the full screen video playing natively without being embedded into a Facebook page. Saving this page within your browser's file menu will archive the actual video in MP4 format. This presents the best possible quality available without suffering video loss through screen captures or various online download tools. This method provides the purest copy of the target video possible.

Downloading Instagram Videos

Right-clicking a video within Instagram does not present an option to download the media file. There are plenty of third-party websites which will fetch a video for you, but we should understand how to do this ourselves. Earlier, you looked at the source code of an Instagram video in order to identify the thumbnail still image. Now, let's repeat the process to obtain a direct URL to a video. When we right-click the page at instagram.com/p/B2gy4Qwg6be and choose "View source", we see the same source code as explained previously. If we conduct a search for "video:secure_url", we will be presented with one hyperlink. If we copy and paste this entire address into a new tab, the video file opens and begins playing. We can now simply right-click this video and choose "Save video as". Alternatively, you can use the free websites **Blastup** (blastup.com/instagram-downloader) or **Instaview** (instaview.me).

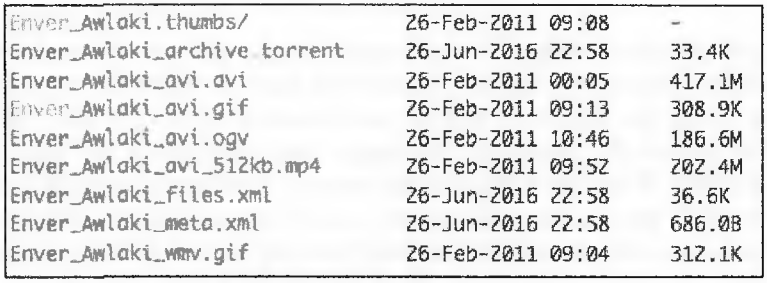
Downloading Twitter Videos

Many Tweets contain embedded videos from other services such as YouTube. In those scenarios, simply load the videos within the original host and proceed as instructed. However, videos hosted within Twitter's own servers present more of a problem. The source code will not display any links to mp4 videos or direct URLs to play the pure media within a new tab. Instead, we must rely on third-party options. The video download tools explained during the Linux Applications chapter are my preferred method, but we should have easy alternatives. I have had continued success with the websites **Save Tweet Vid** (savetweetvid.com) and **Twitter Video Downloader** (twittervideodownloader.com).

Internet Archive (archive.org/details/opensource_movies)

The premise of this site is to permanently store open source movies, which can include commercial and amateur releases. The search option at the beginning of every page allows for a specific section of the site to be searched. Selecting "community video" will provide the best results for amateur video. A large number of anti-government and anti-American videos are present and ready for immediate download. Unlike YouTube, this site does not make it easy to identify the user that uploaded the videos. Furthermore, it does not link to other videos uploaded by the same user. To do this, you will need to look for some very specific text data. As an example, consider that Internet Archive user Enver_Awlaki is your target. His video profile is located at http://www.archive.org/details/Enver_Awlaki. One of his video pages is stored at the address of https://archive.org/details/Awlaki_to_americans.

Below the video frame in the center of the page are several options on the lower right. These allow you to specify video files with different file types. Below these options is a link titled "Show All". Clicking the link provides a view of the files associated with the video as seen in Figure 21.02. The eighth link on this list forwards to the metadata associated with the video. This data includes the title, description, creator, email address used to upload, and the date of upload, as seen in the text below the example.



Enver_Awlaki.thumbs/	26-Feb-2011 09:08	-
Enver_Awlaki_archive torrent	26-Jun-2016 22:58	33.4K
Enver_Awlaki_avi.avi	26-Feb-2011 00:05	417.1M
Enver_Awlaki_avi.gif	26-Feb-2011 09:13	308.9K
Enver_Awlaki_avi.ogv	26-Feb-2011 10:46	186.6M
Enver_Awlaki_avi_512kb.mp4	26-Feb-2011 09:52	202.4M
Enver_Awlaki_files.xml	26-Jun-2016 22:58	36.6K
Enver_Awlaki_meta.xml	26-Jun-2016 22:58	686.0B
Enver_Awlaki_wmv.gif	26-Feb-2011 09:04	312.1K

Figure 21.02: An Internet Archive video options page.

<mediatype>movies</mediatype>
<collection>opensource_movies</collection>
<title>Awlaki_to_americans</title>
<description>UmmaNews</description>
<subject>UmmaNews</subject>
<identifier>Awlaki_to_americans</identifier>
<uploader>ibnumar@islamumma.com</uploader>
<addeddate>2012-03-31 22:47:36</addeddate>
<publicdate>2012-04-01 00:09:10</publicdate>

This view quickly identifies the email address of ibnumar@islamumma.com as the verified uploader of the video content. It also displays the exact date and time of upload and publication. In this example, notice that the author waited over an hour to publish the content. Since 2016, I have seen the Internet Archive become a very popular place to store video, especially from international subjects that may be blocked from traditional American services such as YouTube.

TV News Archive (archive.org/details/tv)

At the time of this writing, the TV News Archive, another part of archive.org, had collected 1,894,000 television news broadcast videos from 2009-2019. Furthermore, it extracts the closed captioning text from each video and provides a search option for this data. This allows you to search for any words verbally stated during these broadcasts in order to quickly locate videos of interest. A search of the term "Bazzell" resulted in 40 videos that mentioned someone with my last name within the broadcast. Selecting any result will play the video and all text from the closed captioning. The menu on the left will allow filtering by show title, station, date, language, and topic. I have found this resource valuable when vetting a potential new hire for a company.

Video Closed Captions (downsub.com)

YouTube and other providers attempt to provide captioning subtitles for as many videos as possible. This automated process transcribes any spoken dialogue within the audio of the video

file and documents the words to text. To see this text while watching a video, click on the closed captioning icon (cc) in the lower left area of the video box. When the icon changes to a red color, the subtitles will display. These subtitles are contained within a small text file associated with the video. It also includes time stamps that identify the frame in which each piece of text is spoken. YouTube does not provide a way to obtain this text file, but Downsub does. Copy an entire URL of any YouTube video with closed captioning. Paste this link into this website and execute the process. This will display download links for the captioning inside the video. Links for each language will download text files with an .srt file extension. These automated captions are not usually completely accurate. Slang and mumbled speech may not transcribe properly. Any time that you collect and submit a YouTube video as part of a report, I recommend obtaining this caption file as well. Even though the actual text may not be accurate, it can help during official proceedings with identifying a specific portion of a video.

Live Video Streams

If you are investigating any live event that is currently occurring, live streaming video sites can be a treasure of useful intelligence. These services offer the ability for a person to turn a cell phone camera into an immediate video streaming device capable of broadcasting to millions. The common set-up is for a user to download a host service's application to a smartphone. Launching the application will turn on the video camera of the phone and the video stream is transmitted to the host via the cellular data connection or Wi-Fi. The host then immediately broadcasts this live stream on their website for many simultaneous viewers to see. An average delay time of five seconds is common. There are now several companies that provide this free service. The following are listed in my order of preference for investigative needs. Each site has a search option to enter the keywords that describe the live event you want to watch. You may also see Twitter links to these services while monitoring targets.

LiveStream (livestream.com)
Twitch (twitch.com)
LiveU (liveu.tv)
YouNow (younow.com)
VaughnLive (vaughnlive.tv)

Real World Application: During several large events, I have used live streams to capture the majority of my intelligence. In one investigation, I was assigned the task of monitoring social networks during a large protest that had quickly become violent to both officers and civilians. While Twitter and Facebook occasionally offered interesting information, live streams provided immediate vital details that made a huge impact on the overall response of law enforcement, fire, and EMS. The live video streams helped me identify new trouble starting up, victims of violence that needed medical attention, and fires set by arsonists that required an immediate response.

Periscope (pscp.tv)

Periscope is a live video streaming app for iOS and Android acquired in January 2015 by Twitter before the product had been publicly launched. Its use requires a mobile device and there is no

web-based official search or player for Periscope streams. Your best option is to search within Twitter. After conducting a keyword search, click the "Broadcasts" option in the top menu. This will present several Periscope video streams. Many will be archived from previous streams, but any live content will be at the top of your results. Once you have identified the video of interest, you should open the media within an official Periscope page. Hover over the video and click on the small Periscope icon in the lower right. It appears similar to a marker on an online map. This will open a page similar to the following.

https://www.pscp.tv/w/1djGXMpDewzJZ

The last portion of this URL is the unique video ID of the target video. This should appear similar to previous instruction on YouTube and Facebook videos. The first search we should conduct in relation to this ID is a Periscope API query. The following URL presents the server data available about the video ID in this example. The Twitter Tools page mentioned in Chapter Ten possesses an option to translate an ID to the direct output page.

https://api.periscope.tv/api/v2/getBroadcastPublic?token=1djGXMpDewzJZ

Within the data from this query, we can see the following. Note that this is only a partial list of details, and focused only on the data most valuable to us. The commentary in parentheses explains the purpose of this data.

created_at: 2015-05-13 06:26:59 (Date and time of account creation)
twitter_screen_name: DVATW (Twitter username)
broadcast created_at: 2017-12-28 19:58:16 (Date and time of video stream creation)
updated_at: 2017-12-28 20:16:10 (Date and time of any profile changes)
friend_chat: false (Target is not using "friend chat")
private_chat: false (Target is not using "private chat")
language: en (Language setting of target's profile)
start: 2017-12-28 20:00:29 (Beginning date and time of video stream broadcast)
has_location: true (User has allowed location enabling)
city: houston (User provided city)
country: USA (User provided country)
country_state: tx (User provided state)
ip_lat: 38.4 (GPS coordinates of IP address in use*)
ip_lng: -90.5 (GPS coordinates of IP address in use*)
width: 320,height:568 (Size of the video stream)
camera_rotation: 270 (Identifies how the phone is rotated)
broadcast_source: periscope_ios_1.13.6 (Identifies the make of his device and app version)
available_for_replay: true (Displays whether user allows archiving of video)
tweet_id: 946470936403775488 (Twitter post announcing video stream)
n_watching: 112 (People currently viewing video)
n_watched: 1140 (Total people that viewed any portion of video)

Note that the GPS coordinates are never meant to be an exact location. These are generic numbers determined from the IP address of the target. These will usually identify the city of the broadcast, but never an accurate address. These settings can be disabled by the user. Worse, use of a VPN could generate unreliable location information. I have had limited success with free third-party Periscope search websites, including **Perisearch** (perisearch.xyz) and **Xxplore** (getxplore.com), but nothing can compare to a search within Twitter or on the Periscope page.

Everything Else

As a final reminder, I have found YouTube-DL to work best when attempting to download video from thousands of supported resources. It is explained in the previous Linux Applications chapter. If you have encountered a video service which is not listed here and are looking for an online solution that does not require any software configuration, I have had success with the following online resources. These will attempt to download or convert any embedded video content.

Keepvid (keepvid.com)
KeepvidPro (keepvid.pro)
ClipConverter (clipconverter.cc)
OnlineConverter (onlinevideoconverter.com)

Video Associations Graph

If you have not yet found a target video of interest for your investigation, Yasiv may help. This has been a "last resort" option for me and has proved to be beneficial. Enter any search term and it creates a visual representation of any YouTube videos which match the search. It also displays connections to other related videos. These connections are determined by YouTube, and are created based on user input and activity. Figure 21.03 displays a result for the term "OSINT". It displays the exact thumbnails for each video, which was explained earlier. This technique has quickly identified videos of interest based on the thumbnail and association to other evidence.

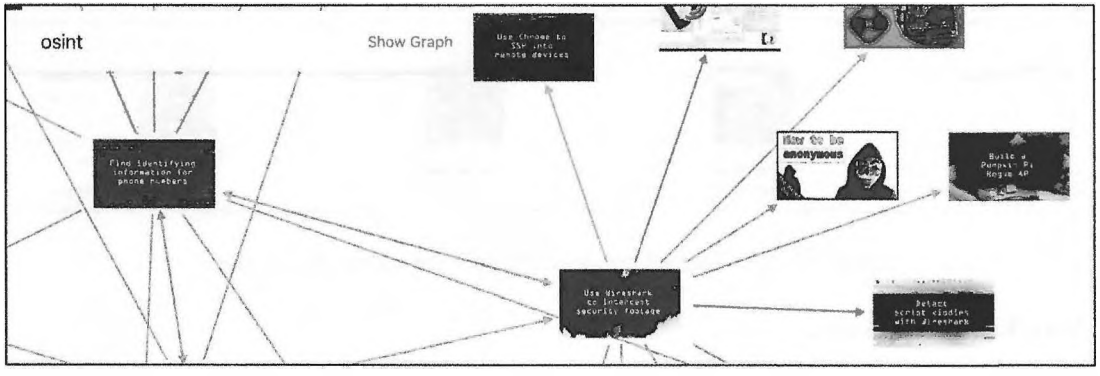


Figure 21.03: A Yasiv video associations graph.

YouTube Channel Crawler (channelcrawler.com)

As explained previously, anyone can search YouTube and filter by Channels. This allows you to only see results which possess one or more videos within a designated Channel. Unfortunately, the results place emphasis on the most popular channels. Within an investigation, it is much more likely that your target will not have thousands of views or followers. Instead, a channel with no subscribers is more common. Finding these poorly-visited channels is quite difficult with official search options. Instead, we can find these lesser-known collections with YouTube Channel Crawler. Let's conduct an example demonstration. I queried the term "Guns" within YouTube, clicked the filters option, and chose to only display Channels. I received numerous results, and every Channel featured over 100,000 subscribers. I would never find my target there. Now, let's use our crawler. Review the settings in Figure 21.04. I have chosen the term of "Guns", no limit to the results, a maximum of 40 subscribers and 40 total views, and did not specify a date range. Figure 21.05 displays partial results. As you can see, these Channels receive very little attention, but were at the top of my results due to the filters I applied.

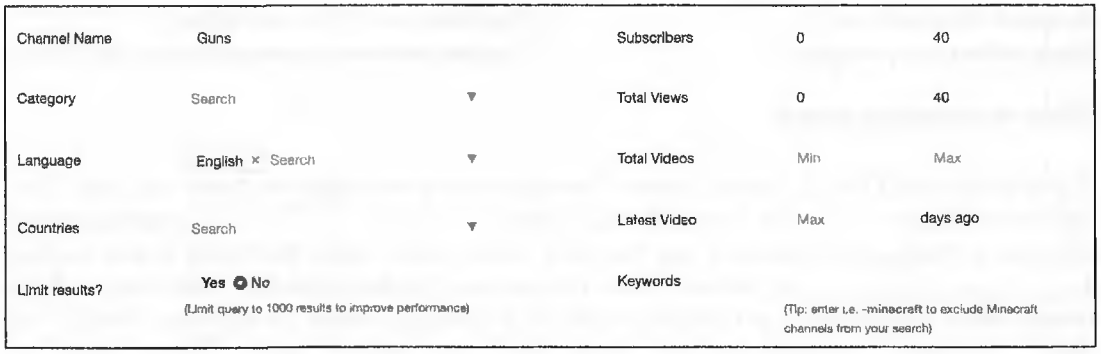


Figure 21.04: Search options on YouTube Channel Crawler.



Figure 21.05: Channel results on YouTube Channel Crawler.

Petey Vid (peteyvid.com)

This is nothing more than a simple video search engine, but I find it valuable. In a recent case, I searched a very unique keyword on Google and found no results. Petey Vid found a video from an otherwise unknown video host (bitchute.com). This is now in the tools and checked often.

TikTok (tiktok.com)

TikTok is a social media video application originally created for creating and sharing short lip-sync, comedy, and talent videos. Today, it is very similar to Vine, which was shut down in 2016. TikTok allows users to create standard videos of 3 to 15 seconds in length and looping videos of 3 to 60 seconds in length. It is popular globally and was the most downloaded app in the U.S. during various months in 2018. It should not be ignored when investigating younger targets. The data is public, and we need to focus on search, usernames, comments, and acquisition of videos.

TikTok Search: TikTok does not offer any native search options within its pages or profiles. However, third-party resources provide a hashtag search option. I have found **TikTokDom** (tiktokdom.com) helpful, but prefer **TikTokAPI** (tiktokapi.ga). Within this page, click on "Hashtag Search" and provide any complete or partial search terms. Searching "osint" provides results of "inosint", "osintamu", and any other hashtags containing the term. If you know your exact hashtag of interest, you can navigate to a direct URL as follows, replacing "test" with your target information.

https://www.tiktok.com/tag/test

TikTok Usernames: TikTokAPI also possesses a username search option which behaves exactly as the hash tag feature. If you know your target's username, you can navigate directly to the profile with the following URL. Replace "test" with your target's username.

https://www.tiktok.com/@test

TikTok Comments: Similar to Instagram, users leave comments on the TikTok posts of others. These can be publicly viewed next to the video, but the content is not located within the source code of the profile. You could select all of the text, copy it, and then paste into a report. However, I prefer to use **Export Comments** (exportcomments.com). Paste your target TikTok video URL into the search field and allow the process to complete. Below is an actual excerpt, which includes the exact usernames, date, time, likes, and comments from the target post.

Name	Date	Likes	Comments
daddy.devito	08/10/19 19:51:24	1	lmao sorry for this
alicetsybulski	08/10/19 21:20:31	1	😂 I hate you Trisha!
barbaquesauce	08/10/19 21:25:24	3	😂😂 He's so weird looking

TikTok Videos: When identifying a video of interest on TikTok, the original video file should be downloaded and archived. The video page does not offer a download feature, and right-clicking does not present any helpful options. Instead, we must either dig into the source code of rely on third-party websites. First, let's focus on the manual method. Right-click any video page and select the "View Source" option. When the source code is displayed, search for "muscdn" and locate a URL which includes "video" within the path. An actual example follows.

https://v16.muscdn.com/75872b2c9c565afcfce83a9d9779815e/5da50dc4/video/tos/maliva/tos-maliva-v-0068/bf6d76a268fd437b912512d5cdaf41a9/

Place a found URL similar to above into a new browser tab to view the full size video, usually in 720p resolution. You can now right-click on this video to save it as a standard mp4 video file. Alternatively, you can replicate this process using the services **TikTokDownload** (tiktokvideodownload.com) or **ExtertsPHP** (expertsphp.com/tiktok-video-downloader.html). The ExprtsPHP service also supports numerous additional video hosting websites.

Fluid Data (fluiddata.com)

This is not a video utility, but I believe it fits best within this chapter as a media service. Fluid Data queries a keyword through millions of audio podcasts and presents any results. As an example, I conducted a search of the term "privacy" and received 437,672 matches of that word within the spoken audio of all indexed podcasts. We can take this a step further with an email address search. I searched "privacypodcast@protonmail.com" and received no results. This is because an email address is not verbalized in this manner. A better search would be "privacy podcast at proton mail dot com". That search also revealed no results. However, a search of "privacy podcast at" presented one hit. Figure 21.06 displays this result which identifies the podcast title (101), author (Michael Bazzell), time stamp of the spoken words (53:52), and translation (privacy podcast at pro time male dot com). As you can see, the technology is not perfect. It is also possible that I slurred my words. Clicking the bar at the time slot plays the podcast from the point where the terms were spoken. I find this service fascinating. As the number of podcasts continue to grow, and audio to text technology improves, this technique will become more useful in our investigations.

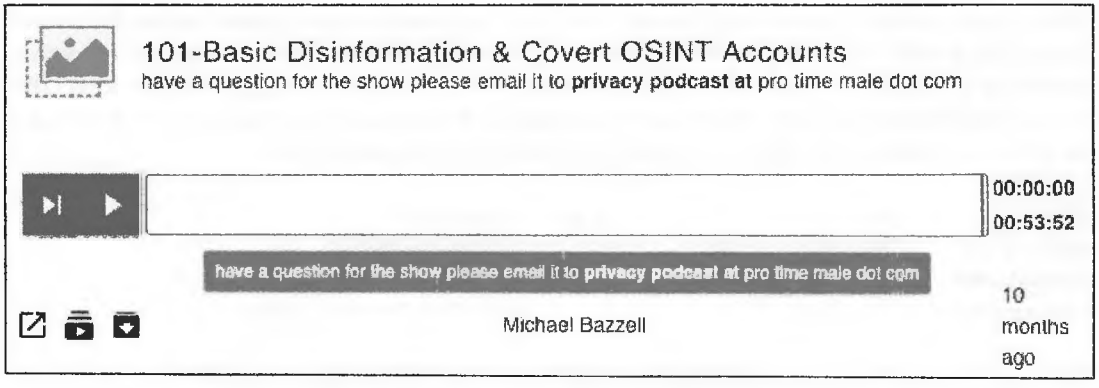


Figure 21.06: A Fluid Data search result.

IntelTechniques Video Search Tool

If you feel overwhelmed at this point, I understand. I found myself getting confused about which address was the most appropriate for each technique. I created a tool which will walk you through each process. Navigate to the Videos.html file in your Tools download to access an all-in-one option similar to the previous tools. This should replicate and simplify the processes that were explained throughout this chapter. I rely on this tool almost every day, and hope that you will also find it beneficial. Figure 21.07 displays the current configuration.

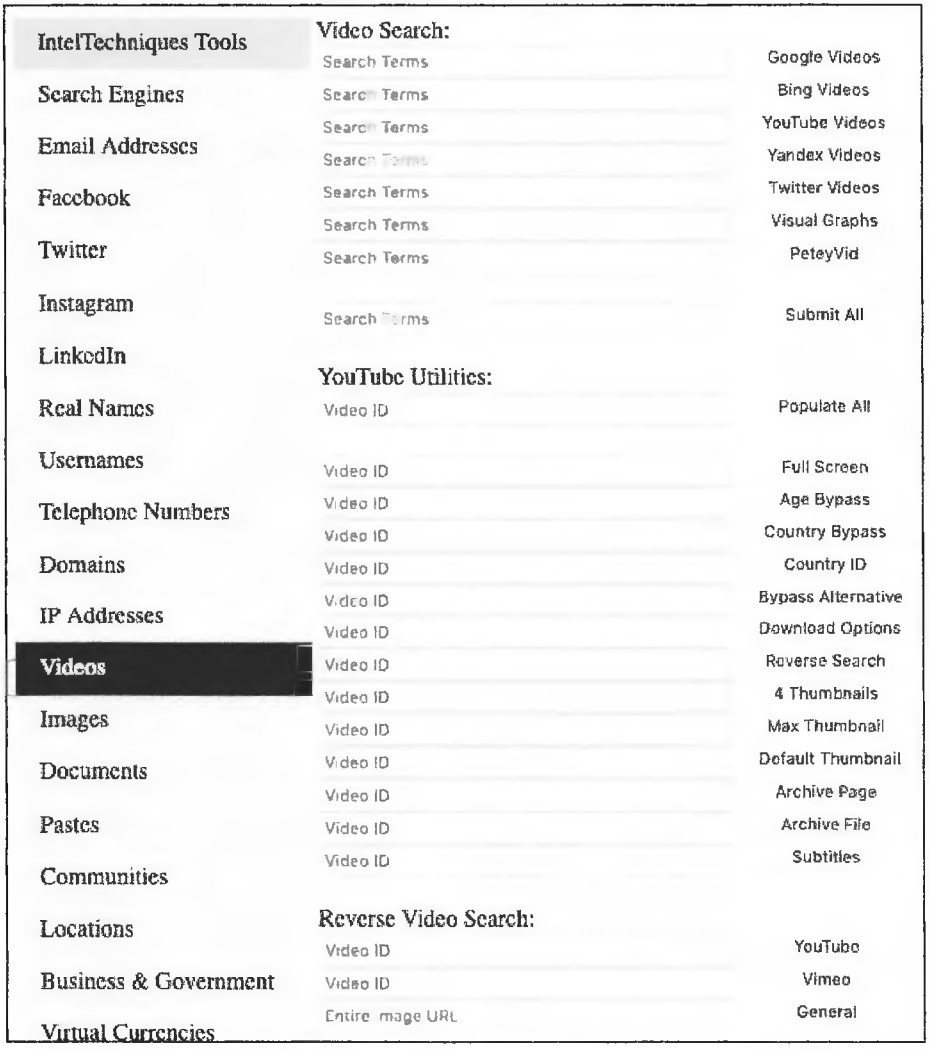


Figure 21.07: The IntelTechniques Video Search Tool.

CHAPTER TWENTY-TWO
DOMAIN NAMES

A specific web page may quickly become the focus of your investigation. Websites, also known as domains, are the main sites at specific addresses. For example, the website that hosts my blog, www.inteltechniques.com/blog is on the domain of inteltechniques.com. The "www" or anything after the ".com" is not part of the domain. These addresses should always be searched for additional information. If your target has a blog at a custom domain, such as privacy-training.com, the content of the site should obviously be examined. However, digging deeper into the domain registration and associated connections can reveal even more information. Every time that I encounter a domain involved in my investigation, I conduct the following types of research, and usually in the order specified. This chapter will explain techniques for each process.

Current Domain Registration
IP/DNS Configurations
Historical Domain Registration
Live & Historical Visual Depictions
Website Analytics Associations
Server and Content Details
Subdomain Locations
Robots.txt Information
Search Engine Marketing & Optimization
Replication of Content

Current Domain Registration

Every website requires information about the registrant, administrative contact, and technical contact associated with the domain. These can be three unique individuals or the same person for all. The contact information includes a full name, business name, physical address, telephone number, and email address. These details are provided by the registrar of the domain name to the service where the name was purchased. This service then provides these details to Internet Corporation for Assigned Names and Numbers (ICANN). From there, the information is publicly available and obtained by hundreds of online resources. While ICANN declares that the provided information is accurate, this is rarely enforced. While most businesses supply appropriate contacts, many criminals do not. While we must consider searching this publicly available information, often referred to as Whois details, we will also need to dig deeper into domain analysis in order to obtain relevant results. First, we will focus on the easy queries.

Whois queries (pronounced Who Is) are very simple searches, but are not all equal. While this data is public, it could change often. Some Whois search sites display the bare bones details while others provide enhanced information. There are dozens of options from which to choose, and I will explain those that I have found useful. After the demonstrations, I present my own custom online tool that automates many processes. For the first example, I will use a target domain of phonelosers.org, the website of the Phone Losers of America, a telephone hacking and prank calling organization. Assume that this website is the focus of your investigation and you want to

retrieve as much information as possible about the site, the owner, and the provider of the content. For the standard Whois search, as well as many other options, I prefer ViewDNS.info.

ViewDNS Whois (viewdns.info/whois)

This service provides numerous online searches related to domain and IP address lookups. Their main page (viewdns.info) provides an all-in-one toolbox, but the above website connects you directly to their Whois search. Entering phonelosers.org here presents the following information.

Updated Date: 2019-08-05T17:58:11Z
Creation Date: 1997-08-13T04:00:00Z
Registry Expiry Date: 2020-08-12T04:00:00Z
Registrant Name: Brad Carter
Registrant Organization: Phone losers of America
Registrant Street: PO Box 465
Registrant City: Albany
Registrant State/Province: Oregon
Registrant Postal Code: 97321
Registrant Country: US
Registrant Phone: +1.8144225309
Registrant Email: brad@notla.com

The administrative and technical contacts were identical to the registrant shown above. This data identifies Brad Carter as the owner of the site, it was created in 1997; expires in August of 2020; and he has a PO Box in Albany, Oregon. A telephone number and email address can be searched through the methods explained in previous chapters. This is a great start, if the provided details are accurate. I have found that ViewDNS will occasionally block my connection if I am connected to a VPN. Alternative Whois research tools are whois.net and who.is.

Many domain owners have started using private registration services in order to protect their privacy. These services provide their own data within the Whois search results, and only these companies know the true registrant. While a court order can usually penetrate this anonymity, I will discuss public resources to help in these situations. While we are discussing ViewDNS, you should be aware of the additional search options available from the main website.

ViewDNS Reverse IP (viewdns.info/reverseip)

Next, you should translate the domain name into the IP address of the website. ViewDNS will do this, and display additional domains hosted on the same server. This service identified the IP address of phonelosers.org to be 104.28.10.123 and stated the web server hosted 134 additional domains. These included domains from websites all over the world without a common theme. This indicates that he uses a shared server, which is very common. If I would have seen only a few domains on the server, that may indicate he is also associated with those specific domains.

ViewDNS Port Scanner (viewdns.info/portscan)

This online port scanner looks for common ports that may be open. An open port indicates that a service is running on the web server that may allow public connection. A search of phonelosers.org revealed that ports 21, 80, and 443 are open to outside connections. Port 80 is for web pages and port 443 is for secure web pages. These are open on practically every website. However, port 21 is interesting. ViewDNS identifies this as a port used for FTP servers, as was discussed previously. This indicates that the website hosts an FTP server and connecting to ftp.phonelosers.org could reveal interesting information.

ViewDNS IP History (viewdns.info/iphistory)

This tool translates a domain name to IP address and identifies previous IP addresses used by that domain. A search of phonelosers.org reveals the following details. The first column is the IP address previously associated with the domain, the second column identifies the current user and company associated with that IP address, and the last column displays the date these details were collected by ViewDNS.

104.28.10.123	Reserved	CloudFlare, Inc.	2016-01-24
104.28.11.123	Reserved	CloudFlare, Inc.	2016-01-24
104.28.10.123	Reserved	CloudFlare, Inc.	2016-01-24
208.97.152.79	Brea - United States	New Dream Network, LLC	2015-08-14
162.213.253.190	San Francisco - United States	Namecheap, Inc.	2015-01-15
104.28.10.123	Reserved	CloudFlare, Inc.	2015-01-11
104.28.11.123	Reserved	CloudFlare, Inc.	2014-10-29
104.28.10.123	Reserved	CloudFlare, Inc.	2014-10-17
192.185.46.66	Chelmsford - United States	WEBSITEWELCOME.COM	2014-08-08
74.208.175.23	Wayne - United States	1&1 Internet Inc.	2014-07-04
74.208.211.36	Alliance - United States	1&1 Internet Inc.	2011-05-02

The utilities hosted at ViewDNS are always my first stop for a couple of reasons. First, the site has been very reliable over the past ten years. More vital, it allows query via static URL. This is beneficial for submission directly from our tools. The following displays the URL structure for the previous four techniques, with my own site as the target.

https://viewdns.info/whois/?domain=inteltechniques.com
https://viewdns.info/reverseip/?host=inteltechniques.com
https://viewdns.info/portscan/?host=inteltechniques.com
https://viewdns.info/iphistory/?domain=inteltechniques.com

Historical Domain Registration

As stated previously, many domains now possess private registration. This means that you cannot see the owner of a domain publicly. Many hosts are now offering private registration as a free service, further complicating things for investigators. If you query a domain and see a name entry such as "WhoisGuard Protected", you know that the domain is protected. There are two ways to defeat this. The first requires a court order, which is outside the scope of this book. The second way to reveal the owner is through historical domain records. If the domain has been around a while, there is a very good chance that the domain was not always private. There are several free and paid domain history services, and I present my favorites in order of usefulness.

Whoxy (whoxy.com)

This is one of the very few premium services which offer a decent free tier. Searching my own domain, which currently possesses private registration, reveals valuable results. The general registration data confirms what I found on ViewDNS. Scrolling down the page reveals powerful historical records. These identify my real name, multiple email addresses used during various registrations, and a date next to each entry to tie it all together. Figure 22.01 displays one of the results. This confirms that during July of 2015, my site was briefly registered without privacy protection. The "9 Domains" link reveals even more information. Figure 22.02 displays this result which identifies numerous domains which I had previously created from 2007 through 2018. This resource has single-handedly exposed more private domain registrations than any other free service during my investigations throughout 2019. Furthermore, it allows submission via URL which will benefit our search tools. The following URL presents the structure.

https://www.whoxy.com/inteltechniques.com

The search option in the upper right allows us to query email addresses, names, and keywords. This can be extremely valuable when you do not know which domain names your target has owned. Searching "OSINT" reveals a surprising amount of people and companies purchasing domains associated with this topic. Whoxy offers many paid services if the free tier is not sufficient. Of the paid services, this one is the most affordable, allowing you to purchase small amounts of information without committing to any specific level of subscription.



Figure 22.01: A Whoxy historical domain registration result.

DOMAIN NAME	REGISTRAR	CREATED	UPDATED	EXPIRY
inteltechniques.com	NameCheap, Inc.	21 Jul 2013	14 Apr 2018	21 Jul 2019
inteltechniques.online	GoDaddy.com, LLC	31 Mar 2018	31 Mar 2018	31 Mar 2019
privacy-training.com	GoDaddy.com, LLC	11 Jun 2016	31 May 2017	11 Jun 2018
yourcomputernerds.com	GoDaddy.com, LLC	18 Dec 2007	20 Oct 2015	18 Dec 2016
missourlinternationalexports.com	GoDaddy.com, LLC	1 Apr 2008	25 Mar 2016	1 Apr 2017
illinoisinternationalexports.com	GoDaddy.com, LLC	1 Apr 2008	25 Mar 2016	1 Apr 2017
humphreyinternationalexports.com	GoDaddy.com, LLC	1 Apr 2008	25 Mar 2016	1 Apr 2017
osint-training.info	GoDaddy.com, LLC	8 Apr 2015	-	8 Apr 2016
computercrimeinfo.com	GoDaddy.com, LLC	1 Dec 2007	29 Oct 2014	1 Dec 2015

Figure 22.02: Additional domains owned by a target from Whoxy.

Whoisology (whoisology.com)

I feel like I am cheating a bit when I discuss Whoisology since I have a paid subscription. However, I will focus only on the free tier. This service appeared in 2014 and becomes more powerful every month. Like Whoxy, it provides historical domain records as a reverse-domain search utility. However, that description is not powerful enough to convey the search options available within this website. The home page of Whoisology presents a single search field requesting a domain or email address. Entering either of these will display associated websites and the publicly available Whois data. This is where the publicly available free features end. Any further details require registration. I encourage all readers to create a free account. Once logged in as a free user, you receive much more detail within your searches.

The first basic feature that you see is the display of standard Whois data that will identify the registered administrative contact, registrant contact, technical contact, and billing contact. These will often be the same individual for most personal websites. The advanced feature within this content is the ability to immediately search for additional domains associated within any field of this data. As an example, a search for the domain of phonelosers.org reveals the following data.

Name	Brad Carter (88)
Email	brad@notla.com (7)
Street	PO Box 465 (1,091)
City	Albany (42,428)
Region	Oregon (492,506)
Zip / Post	97321 (3,080)
Phone	8144225309 (4)

The name, address, and other data can be found on any Whois search website. However, the numbers in parentheses identify the number of additional domains that match those criteria. In this example, there are a total of 88 domains registered to Brad Carter, and seven domains registered to the email address of brad@notla.com. Clicking on any of these pieces of data will launch a new page with all of the matching domain information. As an example, clicking on brad@notla.com will display the 7 domain names associated with his email address. Clicking 8144225309 will display the 4 domain names associated with his telephone number. One of these is a new domain that is not directly associated with him. However, since it was registered with the same number, there is now a connection.

This type of cross-reference search has not been found through many other services. Another powerful feature of Whoisology is the historical archives. This service constantly scans for updates to domain registrations. When new content is located, it documents the change and allows you to search the previous data. As an example, a search of computercrimeinfo.com reveals the current administrative contact telephone number to be 6184628253. However, a look at the historical records reveals that on October 16, 2012, the domain contact number was 6184633505. This can be a great way to identify associated telephone numbers that have since been removed from the records. Whoisology will also provide details from the search of an email address. In my experience, Whoisology will provide a more detailed and accurate response than most other resources. A search of the email address brad@notla.com revealed the following domains associated with that account.

notla.com
phonelosers.com
albanyscavengerhunt.com
bigbeefbueno.com
callsofmassconfusion.com
snowplowshow.com
phonelosers.org

If you ever encounter an investigation surrounding a domain or any business that possesses a website, I highly encourage you to conduct research through Whoxy and Whoisology. They also offer access through their API at a cost. The individual queries through their website are free. Whoisology restricts free accounts to only one historical record and three searches every 24 hours. Because of this, Whoxy receives my overall recommendation. I believe Whoisology offers the most data for those looking to purchase a subscription in order to query numerous domains.

Domain Big Data (domainbigdata.com)

Domain Big Data is free and similar to Whoisology and Whoxy. However, it does not offer many options for cross-reference search of data fields. It does offer a limited historical view of the Whois registration data as well as related domains based on email addresses. A search of the domain notla.com revealed the standard Whois data, an associated email address of brad@notla.com, and two additional domains associated with that email address. There were over a dozen historical records of this domain's registration details. Most of them were very recent and identified redundant information. However, one historical record was from six

months prior and identified a previous domain registrar. Searching my own domain confirmed that I possess private registration. However, it exposed a public registration from 2013 including the following details.

Recorded : 2013-09-01
Historical Whois Record
Domain Name: INTELTECHNIQUES.COM
Registrar URL: http://www.godaddy.com
Registrant Name: Michael Bazzell

Historical Visual Depictions

While it may seem obvious, you should document the current visual representation of your target website. This may be a simple screen capture, but you may have advanced needs. When I am investigating a domain, I have three goals for visual documentation. I want documentation of the current site, any historical versions archived online, and then future monitoring for any changes. Previous chapters presented numerous methods for capturing live web pages and techniques for locating online archives. Let's briefly revisit these methods and then apply new resources.

Search Engine Cache

Searching your target domain within Google, Bing, and Yandex should present numerous results. The first is almost always the home page, and clicking the green down arrow usually presents a cached version of the page. This should be conducted first in order to identify any recent cached copies, as explained previously. In my experience, Google is going to possess the most recent cache, which can always be accessed via direct URL with the following structure.

https://webcache.googleusercontent.com/search?q=cache:inteltechniques.com

Carbon Dating (carbondate.cs.odu.edu)

This free service provides a summary of available online caches of a website, and displays an estimation site creation date based on the first available capture. The following static URL would be used to query my own domain.

http://carbondate.cs.odu.edu/#inteltechniques.com

This provided an estimated creation date of "2013-11-16T08:47:11". Furthermore, it displayed multiple direct URLs to historical archives of my website within the Wayback Machine and other services. After checking these links and visiting any other archives at the Internet Archive, the following services fill in the gaps with screen captures from various dates.

Website Informer (website.informer.com)

This site provides extremely vague and publicly available details about any domain. I am only interested in the screen capture available to the right of a search result. Figure 22.03 (upper left) displays their capture of my site in May of 2019.

URLScan (urlscan.io)

Similar to the previous option, this service provides very little valuable details. However, the screen capture is often unique. Figure 22.03 (upper right) displays the result from August of 2018.

Easy Counter (easycounter.com)

The screen capture presented here was very similar to Website Informer, but it was cropped slightly different. This appears to be from May of 2019 and is seen in Figure 22.03 (lower left).

Find Sub Domains (findsubdomains.com)

The goal of this service to identify subdomains of websites, which is explained later in this chapter. I find their collection of subdomains minimal and typically unhelpful. However, their screen captures are in high resolution and current. Figure 22.03 (lower right) displays the result from my domain captured in August of 2019.

DomainIQ (domainiq.com/snapshot_history)

If you were impressed with the historical screen captures presented previously, you should bookmark this new resource. DomainIQ provides numerous query options, similar to ViewDNS. Their historical snapshots are my favorite offering. Figure 22.04 displays the results for my domain. It presented seven unique views. More importantly, it disclosed the capture date, IP address, and host of each screenshot.

Real World Application: In 2019, I used this service to expose a known domain which was suspected of being associated with an extortion case. The suspect quickly deleted the content before the investigation began. The site had not been indexed by the Wayback Machine, Google, Bing, or Yandex. However, DomainIQ possessed a single screen capture which clearly displayed inappropriate images of my client, which was captured several months earlier.

The following direct URLs will be used for our search tool, displaying my own site as an example.

https://website.informer.com/inteltechniques.com#tab_stats
https://www.easycounter.com/report/inteltechniques.com
https://findsubdomains.com/subdomains-of/inteltechniques.com#
https://www.domainiq.com/snapshot_history#inteltechniques.com

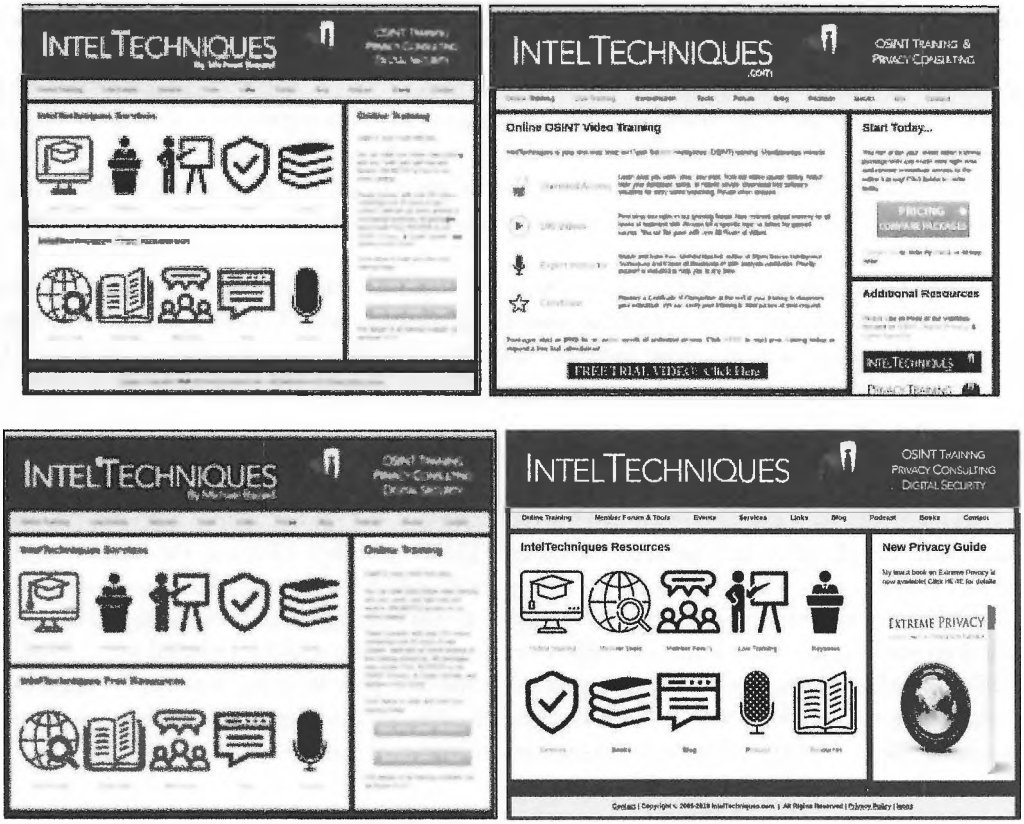


Figure 22.03: Historical screenshot captures from four resources.

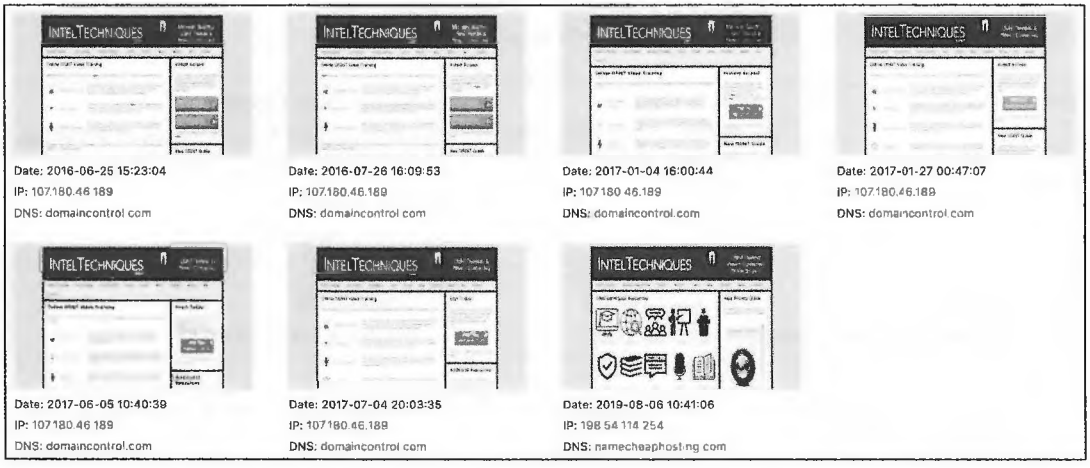


Figure 22.04: Several screen captures from DomainIQ.

Domain Registration History Archives

You now know there are many ways to identify the owner of a domain through historical Whois captures. You might get lucky through a free preview of previous registration data or be required to purchase access in order to see all records dating back decades. We have one last free option which has been successful for me throughout several investigations. We can query the Wayback Machine for the exact historical URL of a domain registration. Let's conduct a demonstration.

If you navigate directly to https://who.is/whois/phonelosers.org, you can see that the domain possesses WhoisGuard protection and the owner's identity is masked. This is displayed in Figure 22.05 (left). However, it is possible that the Wayback Machine captured this exact page. The following URL displays any results.

https://web.archive.org/web/http://who.is/whois/phonelosers.org

This URL defaults to a capture from 2017 which displays the owner name, address, telephone number, and email. This can be seen in Figure 22.05 (middle). Clicking the earliest archive presents a capture of this data from 2010, as seen in Figure 22.05 (right). We now have accurate historical domain registration data without a premium membership, and two additional telephone numbers to investigate. This is not the only domain registration service indexed by archive.org. The following direct links query domain registration history from Who.is, Domain Tools, Whoxy, Domain Big Data, and Whoisology. Replace cnn.com with your target domain. The domain tools presented at the end of this chapter replicate each of these options.

https://web.archive.org/web/http://www.who.is/whois/cnn.com/
https://web.archive.org/web/https://whois.domaintools.com/cnn.com
https://web.archive.org/web/https://www.whoxy.com/cnn.com
https://web.archive.org/web/https://domainbigdata.com/cnn.com
https://web.archive.org/web/https://whoisology.com/cnn.com

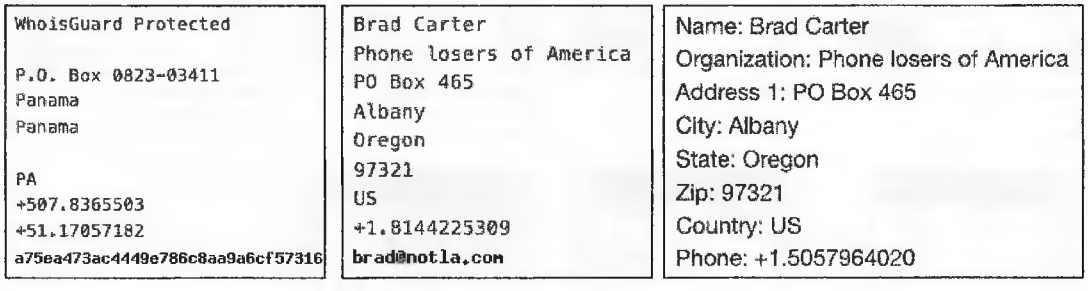


Figure 22.05: Results from who.is (left) and the Wayback Machine (middle and right).

Follow That Page (followthatpage.com)

Once you locate a website of interest, it can be time consuming to continually visit the site looking for any changes. With large sites, it is easy to miss the changes due to an enormous amount of content to analyze. This is when sites like Follow That Page come in handy. Enter the address of the target page of interest, as well as an email address where you can be reached. This service will monitor the page and send you an email if anything changes. Anything highlighted is either new or modified content. Anything that has been stricken through indicates deleted text. Parents are encouraged to set-up and use these services to monitor their child's websites. It does not work well on some social networks such as Facebook, but can handle a public Twitter page fine.

Visual Ping (visualping.io)

If you found the service provided by Follow That Page helpful, but you are seeking more robust options, you should consider Visual Ping. This modern Swiss website allows you to select a target domain for monitoring. Visual Ping will generate a current snapshot of the site and you can choose the level of monitoring. I recommend hourly monitoring and notification of any "tiny change". It will now check the domain hourly and email you if anything changes. If you are watching a website that contains advertisements or any dynamic data that changes often, you can select to avoid that portion of the page. Figure 22.06 displays the monitoring option for phonelosers.org. In this example, I positioned the selection box around the blog content of the main page. I also chose the hourly inspection and Tiny Change option. If anything changes within this selected area, I will receive an email announcing the difference.



Figure 22.06: A portion of a web page monitored by Visual Ping for changes.

Reverse Domain Analytics

Domain analytics are commonly installed on websites in order to track usage information. This data often identifies the city and state from where a visitor is; details about the web browser the person is using; and keywords that were searched to find the site. Only the owner of the website can view this analytic data. Analytics search services determine the specific number assigned to the analytics of a website. If the owner of this website uses analytics to monitor other websites, the analytic number will probably be the same. These services will now conduct a reverse search of this analytic number to find other websites with the same number. In other words, it will search a website and find other websites that the same owner may maintain. Additionally, it will try to identify user specific advertisements stored on one site that are visible on others. It will reverse search this to identify even more websites that are associated with each other. None of this relies on Whois data. A couple of examples should simplify the process.

Spy On Web (spyonweb.com)

Spy On Web is one of many sites that will search a domain name and identify the web server IP address and location. It also conducts a Whois query which will give you registration information of a website. More importantly, it identifies and cross-references website analytic data that it locates on a target domain. A search for the website phonelosers.org reveals a "Google AdSense" ID of ca-pub-3941709854725695. It further identifies five domains that are using the same Google AdSense account for online advertising. This identifies an association between the target website and these new websites. We now know that whoever maintains phonelosers.org places ads on the page. We also know that those same ads and affiliate number is present on five domains. This means that our target likely maintains all of the following domains.

www.notla.com
www.oldpeoplearefunny.com
www.phonelosers.com
www.phonelosers.org
www.signhacker.com

Analyze ID (analyzeid.com)

While Spy On Web is a strong overall analysis tool, there are additional options that should be checked for reverse analytics. Analyze ID performs the same type of query and attempts to locate any other domains that share the same analytics or advertisement user numbers as your target. This will provide new websites related to your target. During a search of phonelosers.org, it identified the Google AdSense ID previously mentioned. It also revealed an Amazon Affiliate ID of phonelosersof-20 and an Amazon product ID of 1452876169. These are likely present because the target sells books through his websites and receives compensation from Amazon through customer purchases. These new pieces of information can be very valuable. Clicking the Amazon Affiliate ID presents the five domains that possess that same code. The Amazon product ID displays the seven websites that also advertise a specific product. When you encounter this type of data, consider the following Google search. It identifies the actual product

that the target is selling within embedded ads. This search reveals that the product is the target's self-published book titled Phone Losers of America.

amazon product 1452876169

Analyze ID also identified a Clickbank affiliate ID that was associated with five additional domains, one of which was unique to any other results. Spy On Web and Analyze ID have provided us several new pieces of information about our target. If you ever see an ID that starts with "UA-", this is likely an identifier for Google to monitor viewers of the website. Searching that number within these tools will also identify related websites. We should visit each website and analyze the content for any relevant evidence. After, we should consider other resources.

DomainIQ (domainiq.com/reverse_analytics)

This service sticks out from the pack a bit because it offers isolated search options. With the previous tools, you search a domain name and hope that the service can extract any analytics IDs and provide details about any online presence. DomainIQ is unique in that it provides designated pages for submission of any IDs located. As an example, I submitted the Google Analytics ID of UA-1946201-13 to the address above. This number was identified as associated with our target website while using Analyze ID. In this example, DomainIQ immediately identified the owner as notla.com. Unfortunately, all other details were redacted within the results page. Fortunately, we could see them in the exported CSV file. Directly above the redacted results is an icon labeled as CSV. Downloading that file revealed an IP address (74.208.175.23), name (Brad Carter) and email address (brad@notla.com). Next, let's search for a Google AdSense ID.

We must first navigate to the URL of domainiq.com/reverse_adsense#. We can enter any AdSense accounts identified with the previous options or ask DomainIQ to search a domain. I provided the ID of ca-pub-3941709854725695, which was identified earlier. The service immediately identified multiple domains sharing the same AdSense account. However, the biggest drawback of DomainIQ is that you are limited to three free searches per day. This is monitored by IP address. If you have a VPN, you could bypass this restriction by switching IPs.

PubDB (pub-db.com)

Another free domain analysis service is PubDB. It is not as robust as others mentioned; however, it occasionally identifies information that was not available with other services. It identified our target's Google AdSense affiliate number within a website and cross-referenced that data to three other related domains.

Nerdy Data (search.nerdydata.com)

Nerdy Data is a search engine that indexes the source code of websites. If you have located a Google Analytics ID, AdSense ID, or Amazon ID of a website using the previous methods, you

should consider searching this number through Nerdy Data. A search of our target's Google AdSense number revealed five domains that possess the same data. The search of the Amazon number revealed three domains. If this service presents more results than you can manage, consider using their free file download option to generate a csv spreadsheet.

Real World Application: While investigating an "anonymous" website that displayed photo evidence of a reported felony, I discovered that the registration information was intentionally inaccurate. A search of the website on these services identified a Google analytics number and an additional website that possessed the same number. That additional website was the personal blog of the suspect. An arrest was made the same day.

Built With (builtwith.com)

A quick analysis of a target website may identify the technologies used to build and maintain it. Many pages that are built in an environment such as WordPress or Tumblr often contain obvious evidence of these technologies. If you notice the YouTube logo within an embedded video, you will know that the creator of the site likely has an account within the video service. However, the presence of various services is not always obvious. Built With takes the guesswork out of this important discovery. Entering the domain of phonelosers.org into the Built With search immediately identifies the web server operating system (Linux), email provider (DreamHost), web framework (PHP, WordPress), WordPress plugins, website analytics, video services, mailing list provider, blog environment, and website code functions. While much of this is geek speak that may not add value to your investigation, some of it will assist in additional search options through other networks. Another option for this type of search is **Stats Crop** (statscrop.com).

Pentest-Tools (pentest-tools.com/reconnaissance/find-subdomains-of-domain)

This unique tool performs several tasks that will attempt to locate hidden pages on a domain. First it performs a DNS zone transfer which will often fail. It will then use a list of numerous common subdomain names and attempt to identify any that are present. If any are located, it will note the IP address assigned to that subdomain and will scan all 254 IP addresses in that range. In other words, it will attempt to identify new areas of a website that may not be visible from within the home page. The following example may help to clarify.

The website at phonelosers.org is a blog that appears to have no further content to be analyzed. Searching for it on Pentest-Tools provides additional intelligence. It identifies the following subdomains present on the web server:

webmail.phonelosers.org
ssh.phonelosers.org
ftp.phonelosers.org
www.phonelosers.org
mail.phonelosers.org

We now know that this domain possesses a webmail server, SSH connection, FTP server, and mail server. This method has helped me locate "hidden" pages which contain several forum messages from users of the site. Previous editions of this book have discussed additional providers for this type of service. Pentest-Tools is the only provider that continues to function. The rest have disappeared.

Robots.txt

Practically every professional website has a robots.txt file at the "root" of the website. This file is not visible from any of the web pages at the site. It is present in order to provide instructions to search engines that crawl the website looking for keywords. These instructions identify files and folders within the website that should not be indexed by the search engine. Most engines comply with this request, and do not index the areas listed. Locating this file is relatively easy. The easiest way to view the file is to open it through a web browser. Type the website of interest, and include "robots.txt" after a forward slash (/). The file for Reddit can be found at the following address.

http://www.reddit.com/robots.txt

If this technique produces no results, you can conduct a Google or Bing query to identify any files. A search of site:twit.tv "robots.txt" on either search engine identifies robots.txt files from the entire website Twit.tv. The main robots.txt located at twit.tv/robots.txt appears as the following, and is not very helpful tour investigation.

```
#
#  88888888888 888        888 d8b 88888888888  888
#      888     888    o   888 Y8P     888      888
#      888     888  d8b   888         888      888
#      888     888 d888b 888 888     888      888888 888  888
#      888     888d88888b888 888     888      888    888  888
#      888     88888P Y88888 888     888      888    Y88  88P
#      888     8888P   Y8888 888     888  d8b Y88b.   Y8bd8P
#      888     888P     Y888 888     888  Y8P  "Y888   Y88P
#
```

```
User-agent: *
Crawl-delay: 10
Sitemap: https://twit.tv/sitemap.xml
```

This was used as an interesting way to show a message to curious web visitors while providing search engines a sitemap for accurate indexing of their site. However, the robots.txt file for twit.tv located on the Wayback Machine at https://web.archive.org/web/*/twit.tv/robots.txt provides more lucrative information, as seen on the following page.

```
# Squarespace Standard Robot Exclusion
# Access is disallowed to functional / filtering URLs

User-agent: *

Disallow: /display/admin/
Disallow: /display/Search
Disallow: /display/Login
Disallow: /display/RecoverPassword
Disallow: /login
Disallow: /contributor
Disallow: /blogold/category
Disallow: /blogold/week
Disallow: /blogold/month
Disallow: /blogold/recommend
Disallow: /blogold/author
Disallow: /login
Disallow: /blog/category
Disallow: /blog/week
Disallow: /blog/month
Disallow: /blog/recommend
Disallow: /blog/author
Disallow: /contests/category
Disallow: /contests/week
Disallow: /contests/month
Disallow: /contests/recommend
Disallow: /contests/author
```

The first line indicates that this website was created by Squarespace, and is likely stored on their servers. The rest of the file identifies online folders that include a live blog, previous blog, contest area, and login portal. Much of this content would not be found in a search engine because of the Disallow setting. These Disallow instructions are telling the search engines to avoid scanning the folders login, contributor, and RecoverPassword. It is likely that there is sensitive information in these directories that should not be available on Google or Bing. You can now type these directories after the domain name of your target to identify additional information. Based on the robots.txt file in this demonstration, typing the following addresses directly into a browser may generate interesting results.

http://twit.tv/blog
http://twit.tv/login
http://twit.tv/contests/week

Most robots.txt files will not identify a secret area of a website that will display passwords, raunchy photos, or incriminating evidence. Instead, they usually provide insight into which areas of the site are considered sensitive by the owner. If you have a target website and have exhausted every other search method, you should also visit this file. It may direct you toward a new set of queries to find data otherwise ignored by search engines.

Search Engine Marketing Tools

The ultimate goal of most commercial websites is to generate income. These sites exist to bring in new customers, sell products, and provide the public face to a company. This has created a huge community of services that aim to help companies reach customers. Search Engine Optimization (SEO) applies various techniques affecting the visibility of a website or a web page in a search engine's results. In general, the higher ranked on the search results page and more frequently a site appears in the search results list, the more visitors it will receive. Search Engine Marketing (SEM) websites provide details valuable to those responsible for optimizing their own websites. SEM services usually provide overall ranking of a website; its keywords that are often searched; backlinks; and referrals from other websites. SEO specialists use this data to determine potential advertisement relationships and to study their competition. Online investigators can use this to collect important details that are never visible on the target websites. Three individual services will provide easily digestible data on any domain. I will use my own domain for each example in order to compare the data. Only the free versions will be discussed.

Similar Web (similarweb.com)

Similar Web is usually the most comprehensive of the free options. However, some of these details usually contradict other services. Much of this data is "guessed" based on many factors. A search of inteltechniques.com produced the following partial information about the domain.

- The majority of the traffic is from the USA, followed by UK, DE, FR.
- There is no paid search or advertisements on search engines.
- There are 56 websites that possess links to the target, and 15 are visible.
- "Buscador" led more people to the site than any other search term followed by OSINT.
- Over 50,000 people visit the site monthly.
- There are five main online competitors to the target, and the largest is onstrat.com.
- 71% of the visitors navigated directly to the domain without a search engine.
- 12% of the traffic was referrals from other websites and search engines.
- The referrals included my other website (computercrimeinfo.com).
- 3% of the traffic to this site originated from social networks Facebook and Twitter.
- Similar websites include onstrat.com and automatingosint.com.

Alexa (alexa.com)

This service is considered the standard when citing the global rank of a website. Most of the collected data is targeted toward ranking the overall popularity of a website on the internet. The following details were provided about inteltechniques.com.

- It is ranked as the 104,033rd most popular site on the internet.

- The average visitor clicks on three pages during a visit.
- Popular searches used to find the domain include OSINT Links and IntelTechniques.
- Facebook, Twitter, and Google referred more traffic than any other source.

Both of these services provided some details about the target domain that were not visible within the content of the page. These analytical pieces of data can be valuable to a researcher. Knowing similar websites can lead you to other potential targets. Viewing sources of traffic to a website can identify where people hear about the target. Global popularity can explain whether a target is geographically tied to a single area. Identifying the searches conducted before reaching the target domain can provide understanding about how people engage with the website. While none of this proves or disproves anything, the intelligence gathered can help give an overall view of the intent of the target. Three additional websites that provide a similar service are listed below.

Search Metrics (suite.searchmetrics.com)
Majestic (majestic.com)
SpyFu (spyfu.com)

Shared Count (sharedcount.com)

This website provides one simple yet unique service. It searches your target domain and identifies its popularity on social networks such as Facebook and Twitter. A search of labnol.org produced the following results. This information would lead me to focus on Pinterest and Facebook first. It tells me that several people are talking about the website on these services.

Facebook Likes: 348
Facebook Total: 1034
Facebook Shares: 538
Twitter Tweets: 0
Facebook Comments: 148
Pinterest Pinned: 1

Small SEO Tools: Backlinks (smallseotools.com/backlink-checker)

After you have determined the popularity of a website on social networks, you may want to identify any websites that have a link to your target domain. This will often identify associates and people with similar interests of the subject of your investigation. There are several online services that offer a check of any "backlinks" to a specific website. Lately, I have had the best success with the backlink checker at Small SEO Tools. A search of my own website, inteltechniques.com, produces 264 websites that have a link to mine. These results include pages within my own websites that have a link to inteltechniques.com, so this number can be somewhat misleading. Several of the results disclosed websites owned by friends and colleagues that would be of interest if I were your target.

Small SEO Tools: Plagiarism Checker (smallseotools.com/plagiarism-checker)

If you have identified a web page of interest, you should make sure that the content is original. On more than one occasion, I have been contacted by an investigator that had been notified of

a violent threat on a person's blog. I was asked to track down the subject before something bad happened. A quick search of the content identified it as lyrics to a song. One of many options for this type of query is the plagiarism checker at Small SEO Tools.

You can use this tool by copying any questionable text from a website and paste it into this free tool. It will analyze the text and display other websites that possess the same words. This service uses Google to identify anything of interest. The benefit of using this tool instead of Google directly is that it will structure several queries based on the supplied content and return variations of the found text. Clicking the results will open the Google search page that found the text. Another option for this type of search is **Copy Scape** (copyscape.com).

Reddit Domains (reddit.com)

Reddit was discussed previously as a very popular online community. The primary purpose of the service is to share links to online websites, photos, videos, and comments of interest. If your target website has ever been posted on Reddit, you can retrieve a listing of the incidents. This is done through a specific address typed directly into your browser. If your target website was phonelosers.org, you would navigate to the following website.

reddit.com/domain/phonelosers.org/

This example produced 16 Reddit posts mentioning this domain. These could be analyzed to document the discussions and usernames related to these posts.

Hunter (hunter.io)

Previously, I explained how Hunter could be used to verify email addresses. This tool can also accept a domain name as a search term, and provides any email addresses that have been scraped from public web pages. The free version of this tool will redact a few letters from each address, but the structure should be identifiable.

Visual Site Mapper (visualsitemapper.com)

When researching a domain, I am always looking for a visual representation to give me an idea of how massive the website is. Conducting a "site" search on Google helps, but you are at the mercy of Google's indexing, which is not always accurate or recent. An alternative to this is to use Visual Site Mapper. This service analyzes the domain in real time, looking for linked pages within that domain. It provides an interactive graph that shows whether a domain has a lot of internal links that you may have missed. Highlighting any page will display the internal pages that connect to the selected page. This helps identify pages that are most "linked" within a domain, and may lead a researcher toward those important pages. Figure 22.07 displays a portion of the map for our previous target. I hovered over a single page, which identifies the URL and highlights

any internal pages with links pointing back to it. This visual representation helps me digest the magnitude of a target website.

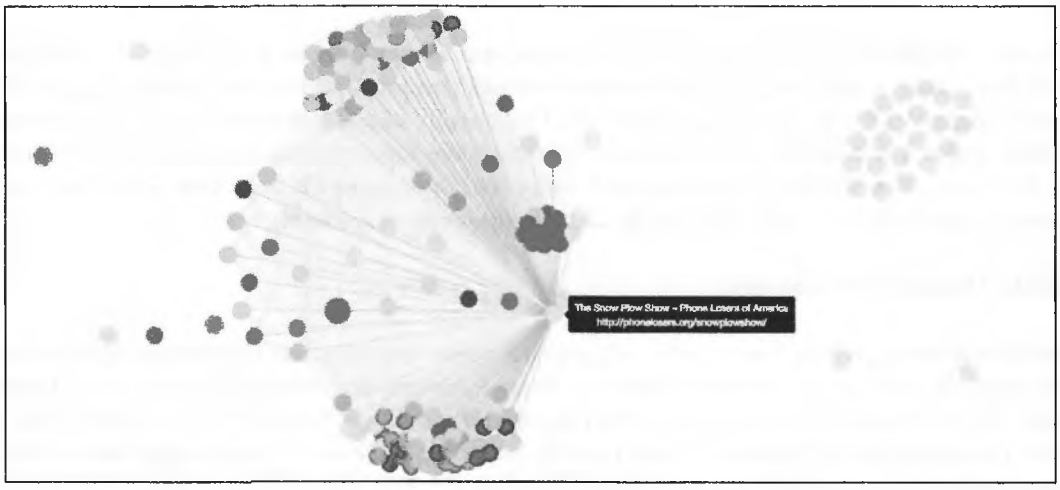


Figure 22.07: A partial view of a website mapping from Visual Site Mapper.

Threat Data

This represents a new category for this chapter, and I am quite embarrassed it took me so long to realize the value of the content. I use the term "Threat Data" to encompass the top four websites which monitor for malicious content. For a network security analyst, this data might identify potentially malicious sites which should be blacklisted within internal networks. For us OSINT researchers, the data represented here can provide a unique glimpse into our target domain. Instead of explaining every facet of these services, I will only focus on the new evidence received after a query of various domains. All of these services are available in your search tools.

Virus Total (virustotal.com)

This option displays the most useful information in regard to OSINT, and this is likely the most popular threat data service of the four. Much of the details presented here are redundant to the previous options, so let's focus only on the unique data. The "Details" menu provides the most public data. The Whois and DNS records should be similar to other sites. The "Categories" area provides the general topics of the target site. For mine, it displays "Information Technology". This can be useful to know a small detail about sites which have disappeared. The "HTTPS Certificate" section can become interesting very quickly.

The "Subject Alternative Name" portion of this section identifies additional domains and subdomains associated with the SSL certificate of your target site. When I search cnn.com, I received dozens of additional URLs which could prove to be valuable. Below is a partial view.

dev.client.appletv.cnn.com,dev.cnnmoney.ch,dev.content.cnnmoney.ch,dev.hypatia.api.cnn.io,dev.money.cnn.com,eightiesyourself.cnn.com,graphql.verticals.api.cnn.io,hypatia.api.cnn.io

I now know that cnn.io is directly associated with the target, which should then be investigated. The "Relations" tab identifies many new subdomains such as customad.cnn.com, go.cnn.com, and store.cnn.com. The "Files" section displays unique content from practically any other resource. It identifies files downloaded **from** the target site for analysis and files which have a reference **to** the target site. Let's analyze my own site as an example. Figure 22.08 displays two files which are present on my site. Both of these files have been analyzed by VirusTotal from either user submission or automated scanning. The first column displays the date of the scan and the second column identifies whether any virus databases detected the files as malicious (none out of 60 tested positive for a virus). The final two columns identify the file type and name. If I were to remove these files today, this evidence would stick around.

Figure 22.09 displays the result in the "Files Referring" section. These are files and programs, which are not present on my site, that refer to my domain. All of these display positive results for being malicious. These are basically files stored on other websites which mention my domain. If you were investigating me, you should try to find these files for further analysis. The fourth file is a virus disguised as a digital copy of this book, attempting to fool would-be downloaders. If your target is mentioned within off-site files, you can learn a lot from analysis. Always use a dedicated virtual machine without network access if you plan to download or open anything found here.

2018-11-14	0 / 59	ZIP	ffmpeg.zip
2018-11-25	0 / 60	PDF	workbook.pdf

Figure 22.08: A Virus Total result identifying files from a domain.

Scanned	Detections	Type	Name
2019-08-29	1 / 48	Win32 EXE	OSCAR.exe
2019-08-04	1 / 60	ZIP	Doxing eBooks.zip
2019-05-14	3 / 73	Win32 EXE	HawkEye.exe
2019-05-09	1 / 60	Office Open XML Document	Open Source Intelligence Techniques
2019-02-11	1 / 60	ZIP	Hacklog2.zip
2019-01-15	2 / 58	PDF	Hacklog. Web Hacking - vol. 2.pdf
2019-01-06	1 / 60	ZIP	Doxing eBooks.zip

Figure 22.09: A Virus Total result identifying files referring to a domain.

Finally, the "Community" tab can be a treasure of details if anything exists about your target domain. This is where members of the Virus Total community can leave comments or

experiences in reference to the target site. While there are no comments currently on my profile, I have seen helpful details on target "hacking" related sites. These included owner details and new domains created for illegal phishing purposes. Most sites will not have any comments, but this option should be checked while browsing through the other sections.

Threat Intelligence (threatintelligenceplatform.com)

This service replicates many of the features already presented. However, I usually look to three specific sections in the domain report. The "Connected Domains" area identifies any external domains which are linked from your source. This can often display hidden links to third-party services otherwise unknown from previous queries. On my domain, you see a link to the icon service I use because I gave attribution within the footer of my page. In previous investigations, I have found additional domains owned by the suspect. From there, I focus on the "Potentially dangerous content" and "Malware detection" sections. Both of these offer a historical view into any malicious content hosted on the target domain. This can include suspicious files or phishing campaigns. While recently investigating a domain which currently possessed no content, this service confirmed the presence of a phishing page designed to steal credentials.

Threat Crowd (threatcrowd.org)

This service provides a unique view of the domains associated with your target. Figure 22.10 displays a partial result for my own domain. It displays my server IP, primary domain, and additional domains which were once associated with my account. The upper-right domain was a small website created for a friend which was hosted as demonstration for a short amount of time.

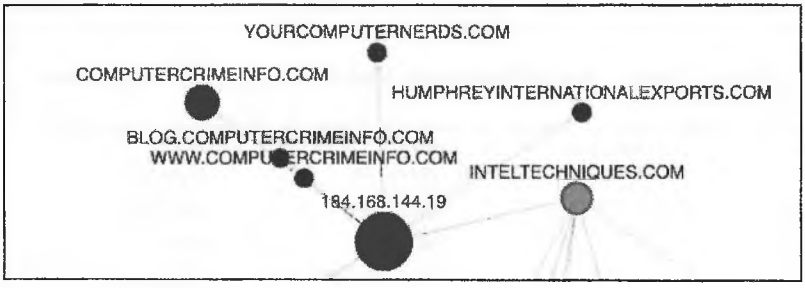


Figure 22.10: A Threat Crowd domain report.

Censys (censys.io)

Censys provides a detailed summary of the basics, which is quite redundant at this point. However, there are three key pieces of information I access here on most domains. I previously mentioned the importance of checking the Subject Alternative Names of a domain's SSL certificate. Most services conduct scans of the entire internet in order to retrieve this data. The moment a certificate is issued, it is provided in real-time to Censys. Censys thus does not need to

rely on internet scans to discover certificates, and more importantly Subject Alternative Names. Therefore, I always click the "Details" button on the summary page and look for any interesting data by searching "alt_name" within the results.

Next, I have relied on the HTTP Body text information stored within the "Details" page of the HTTP and HTTPS sections. This is basically the HTML code which makes the web page display properly within a browser. It is the same data you would see by viewing the source code of a target page. If the target website should disappear, this entire section of code could be copied; pasted into a text file; saved with an html extension; and opened within a browser to reveal the overall look and structure. I prefer to capture this data in the event of a modified or removed target web page.

Finally, I believe Censys has the overall best data about the security certificates associated with a domain. It provides hyperlinks to every certificate within the chain and extreme details about each. Much of this data is not valuable to an OSINT report, but I prefer to collect a screen capture for potential later analysis or comparison to live data.

Overall, threat data is often considered minor bits of information designated only for the digital security community. OSINT practitioners should also be aware of the content available within these sites. While the majority of details are not immediately useful, small nuggets of valuable information which cannot be found anywhere else awaits you.

Shortened URLs

Social networking sites such as Twitter have made the popularity of shortened URL services soar. When people post a link to something they want their friends to see, they do not want the link to take up unnecessary space. These services create a new URL, and simply point anyone to the original source when clicked. As an example, I converted a URL to a blog post from this:

https://inteltechniques.com/blog/2019/08/03/book-release-extreme-privacy/

to this: https://bit.ly/32Up8h7

You have likely seen these during your own investigations, and many people pay them little attention. There is actually a lot of information behind the scenes of these links that can reveal valuable information associated with your investigation. For a demonstration, I created the following shortened links, all of which forward to my home page. After, I will explain how to access the hidden data behind each service.

bitly.com/29A4U1U
http://tiny.cc/v973ez
goo.gl/Ew9rlh
bit.do/cbvNx

Bitly allows access to metadata by including a "+" after the URL. In our scenario, the direct URL would be bitly.com/29A4U1U+. In this example, the results only identified that 21 people have clicked on my link. However, creating a free account reveals much more detail. After logging in, I can see any websites that referred the user to the link and extremely generic location data, such as the country of the user. This is a good start.

Tiny.cc adds a "~" to the end of a link to display metadata. In our example, the direct URL would be tiny.cc/v973ez~. The results on this page identify the number of times the URL was clicked, the number of unique visits, the operating systems of those that clicked the link, and the browsers used. This service also displays generic location data, such as the country of the user.

Google gives us the same detail as above. It also uses the "+" at the end, and our direct demo URL would be goo.gl/Ew9rlh+. This demo notified me that 18 people have clicked my link from 7 different countries. They are mostly Windows users with the Chrome browser.

Bit.do provides the most extensive data. They use a "-" after the URL, and our direct demo address would be http://bit.do/cbvNx-. The results identify all of the details listed previously, plus the actual IP addresses of each visit. In this demo, I know the following about those that clicked on my bait. Note that I redacted the IP addresses to respect the privacy of those involved.

User's IP	Country/City	Access Date
173.244.48.111	United States (Los Angeles, California)	2017-12-29 12:58:55
213.133.92.111	Cyprus (Nicosia, Nicosia)	2017-10-31 04:00:19
125.31.39.111	Macau	2017-04-22 03:36:57
198.8.80.111	United States (Seattle, Washington)	2017-01-12 11:27:12
195.235.92.111	Spain (Madrid, Madrid)	2016-11-18 03:33:22

This type of service can be used in many ways. If you are investigating a viral Twitter post with a shortened URL, you may be able to learn more about the popularity and viewers. You could also use this offensively. During covert investigations, you could forward a shortened URL from Bit.do and possibly obtain the IP address being used by the suspect. I will explain more options for this type of use in the next chapter. If you are investigating a shortened URL link that was not mentioned, consider using the catch-all service at **CheckShortURL** (checkshorturl.com).

Carbon 14 (github.com/Lazza/Carbon14)

I end this chapter with an advanced technique which helped an investigation as I was writing this. It requires Linux, but you should be all set after the instructions within Section One. I was investigating an anonymous blog, with the intent of learning more about the age of the site. It was a brand new WordPress installation, and every post had the same recent 2019 date and time. It had obviously been created recently, but then received an import from an older blog. I wanted to know more about the true dates of the original posts. This is the perfect scenario for Carbon14.

Jump into your OSINT Master virtual machine, open Terminal, and enter the following.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/Lazza/Carbon14`
- `cd Carbon14`
- `python -m pip install -r requirements.txt`

You now have Carbon 14 installed with all dependencies met. If you ever want to update the program, you can enter the following at any time, or add it to your update script.

- `cd ~/Downloads/Programs/Carbon14`
- `git pull https://github.com/Lazza/Carbon14`

Once installed, you must navigate to the Carbon14 folder in order to execute the program. When opening Terminal, my commands appear as follows, if the IntelTechniques blog was the target.

- `cd ~/Downloads/Programs/Carbon14`
- `python carbon14.py https://inteltechniques.com/blog/`

This launches Carbon14 and executes a scan of inteltechniques.com/blog/. It searches for any images hosted within the page and analyzes the metadata for creation dates. Below are partial results from my own site. This indicates that the static page was last modified on August 2, 2019, and the images were posted in March 2018, May 2019, and July 2019.

```
Server: Apache
Last-Modified: Fri, 02 Aug 2019 21:53:52 GMT

# Internal images

Date (UTC)              URL
2018-03-18 16:41:15     <https://inteltechniques.com/img/blog.png>
2019-05-23 16:43:51     <https://inteltechniques.com/img/banner.png>
2019-07-15 17:25:36     <https://inteltechniques.com/img/EP-3D-Gray.png>
```

If this were my target, I would now have suspicion that the original blog posts were from an earlier date. These dates can be intentionally or unintentionally altered, so this is not a forensically sound method. It is simply an additional piece to the puzzle which may warrant further investigation. I find that Carbon14 works best on static websites and blogs. I have also had surprising success when targeting social network profiles. If desired, you could modify the scripts and techniques mentioned during the Linux chapters in order to automate this process.

IntelTechniques Domain Search Tool

Similar to the previous custom search tools mentioned here, I have created a page for easier domain searching. While it does not possess every service discussed in this chapter, it can automate queries across the most beneficial options. Each box allows entry of a target domain name. The "Submit All" option will open several tabs within your browser that present each query listed on the page. The final section provides easy access to shortened URL metadata. Figure 22.11 displays the current status of the tool.

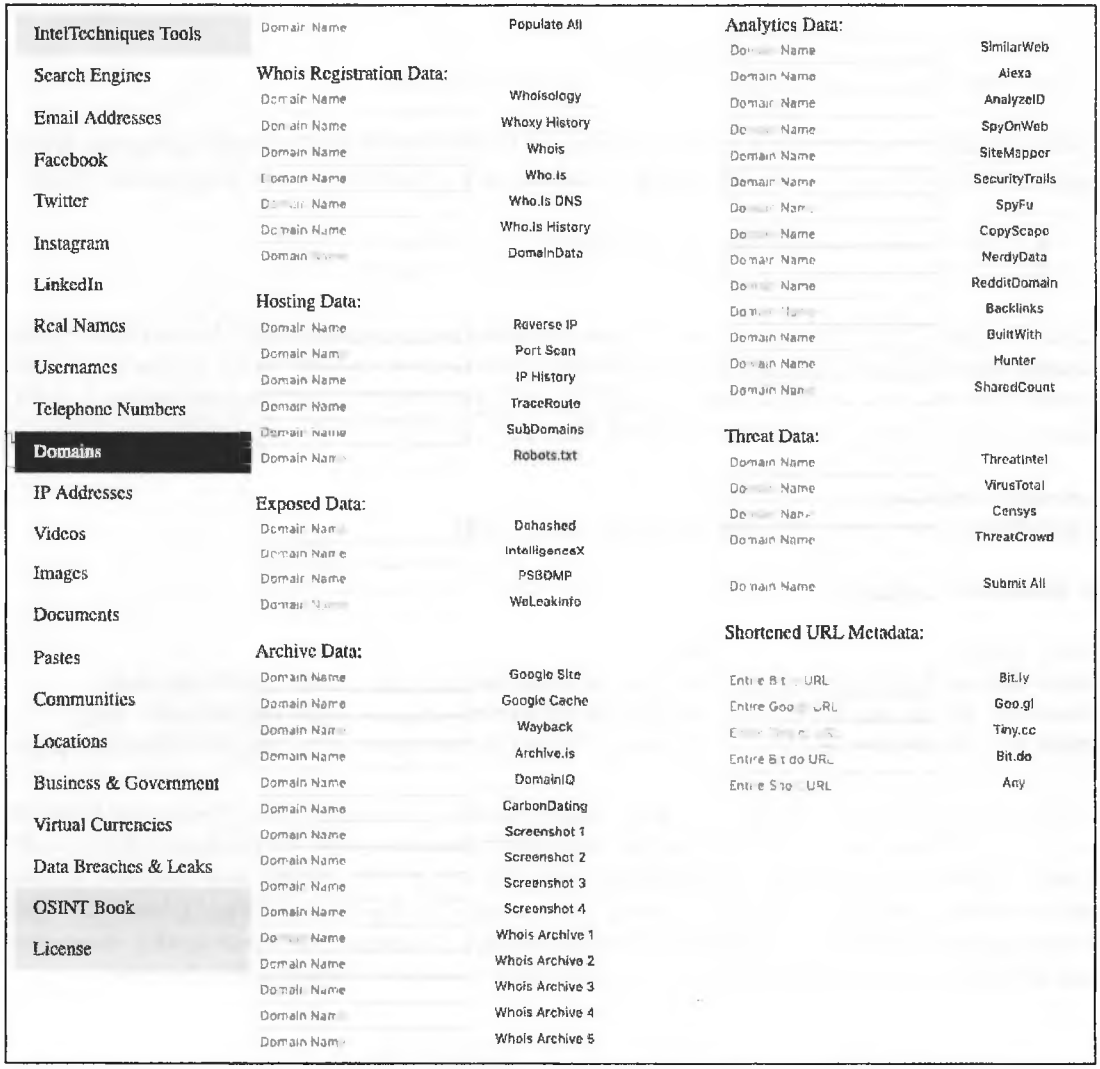


Figure 22.11: The IntelTechniques Custom Domain Search Tool.

CHAPTER TWENTY-THREE
IP ADDRESSES

IP addresses are often obtained from an internet investigation, email message, or connection over the internet. When legal process is served to online content providers, a list of IP addresses used to log in to the account is usually presented as part of the return of information. Serving legal orders to identify and obtain IP addresses is outside the scope of this book. However, several techniques for collecting a target's IP address using OSINT are explained in this chapter. The previous instruction assumed that you were researching a domain name. These names, associated with websites, simply forward you to a numerical address that actually hosts the content. This is referred to as an Internet Protocol (IP) address.

If you know the IP address of a website, it can be entered into an address field instead of the domain name. As an example, when you enter google.com into a browser, your connection forwards you to 74.125.224.72. Typing this IP address directly into your browser will present the Google landing page. The way that you encounter IP addresses as a target of your research will vary widely. Law enforcement may receive an IP address of an offender after submitting a subpoena to an internet provider. Any online researcher may locate an IP address while researching a domain with the previous methods. While only one website can be on a domain, multiple domains can be hosted on one IP address. The following resources represent only a fraction of available utilities. Note that many of the domain name resources mentioned in the previous chapters also allow for query by an IP address.

IPLocation (iplocation.net)

IPLocation offers unlimited free IP address searches, and queries five unique services within the same search results. The results are the most comprehensive I have seen for a free website. While GPS coordinates of an IP address are available, this most often returns to the provider of the internet service. This usually does not identify the exact location of where the IP address is being used. The country, region, and city information should be accurate. If an organization name is presented in the results, this indicates that the address returns to the identified company. The exception here is when an internet service provider is identified. This only indicates that the IP address belongs to the specified provider. Most results translate an IP address into information including business name, general location, and internet service provider. This can be used to determine if the IP address that a target is using belongs to a business providing free wireless internet. If you see "Starbucks", "Barnes & Noble", or other popular internet cafés listed in the results, this can be important intelligence about the target. This can also quickly confirm if a target IP address is associated with a VPN service. Alternative websites include **IP Fingerprints** (ipfingerprints.com) and **IP2Location** (ip2location.com).

Bing IP (bing.com)

Once you have identified an IP address of your target, you can search for websites hosted on that IP address. A specific search on Bing will present any other websites on that server. If your target is stored with a large host such as GoDaddy, there will not be much intelligence provided. It will only list websites that share a server, but are not necessarily associated with each other. If the user is hosting the website on an individual web server, this search will display all other websites that the user hosts. This search only works on Bing and must have "ip:" before the IP address. An example of a proper search on Bing would look like ip:54.208.51.71. The results of this search identify every local website hosted by a specific local website design company.

ViewDNS Reverse IP (viewdns.info/reverseip)

This page was previously used to translate a domain name into an IP address. It will also display additional domains hosted on an individual IP address. This service identified 134 domains hosted on 104.28.10.123. These included domains from websites all over the world without a common theme. This indicates that he uses a shared server, which is very common. If I would have seen only a few domains on the server, that may indicate that he is also associated with those specific domains.

ViewDNS IP Location (viewdns.info/iplocation)

This utility cross-references an IP address with publicly available location data connected to the server hosting any domains associated with the IP address. A search of 54.208.51.71 revealed the following information.

City: Ashburn
Zip Code: 20147
Region Name: Virginia
Country Code: US
Country Name: United States

ViewDNS Port Scan (viewdns.info/portscan)

This online port scanner looks for common ports that may be open. An open port indicates that a service is running on the web server that may allow public connection. A search of 54.208.51.71 revealed that ports 21, 53, 80, and 443 are open to outside connections. Port 21 is for FTP connections, 53 is for DNS settings, 80 is for web pages, and port 443 is for secure web pages.

ViewDNS IP Whois (viewdns.info/whois)

This service was used earlier to display registration information about an individual domain. Entering an IP address will attempt to identify details about any domain registrations associated

with the address. A search of 54.208.51.71 revealed it to belong to Amazon and provided the public registration details.

ViewDNS IP TraceRoute (viewdns.info/traceroute)

This tool identifies the path that ViewDNS took from their servers to the target IP address. This can identify IP addresses of servers that were contacted while you tried to establish communication with the target's address. These will occasionally identify associated networks, routers, and servers. Additional IP addresses can be later searched for further details. The numbers after the IP addresses indicate the number of milliseconds that each "hop" took.

That's Them (thatsthem.com/reverse-ip-lookup)

The previous resources rely on conventional IP address data, which is sourced from various registration documentation and scanning of servers. Very little information is sensitive or personal in nature. That's Them enters into an environment that is a lot more invasive. This service, mentioned previously during person, email, and telephone search, collects marketing data from many sources to populate its database. This often includes IP address information. These details could have been obtained during an online purchase or website registration. Regardless of the source, the results can be quite beneficial. At the time of this writing, I searched an IP address associated with a business email that I received. The result identified a person's name, home address, company, email address, and age range. All appeared accurate. This tool will work best when searching static business IP addresses, and not traditional home addresses that can change often. While I get no results much more often than positive results, this resource should be in everyone's arsenal.

I Know What You Download (iknowwhatyoudownload.com)

While discussing invasive websites, this resource might be the most personal of all. This service monitors online torrents (ways to download large files which often violate copyright laws) and discloses the files associated with any collected IP addresses. I searched the previous IP address collected from an associate, and received an immediate hit. Figure 23.01 displays the result. It identifies that the target IP address was downloading two specific movies on December 28, 2017 at 9:53 pm. Clicking on the movie title presents every IP address captured that also downloaded the same file. Again, this will work best with IP addresses that rarely change, such as a business, organization, or public Wi-Fi network. I have used this to determine the files being downloaded from the network with which I was currently connected. On one occasion, this revealed an employee that was downloading enormous amounts of pornography on his employee's network. He should have used a VPN, which would have masked his online activity from me. In order to see the power of this type of service, try searching a known VPN address such as an address provided by Private Internet Access (PIA) 173.244.48.163. While I know that no one reading this book has ever downloaded pirated content, this should serve as a reminder why VPNs are essential.

Dec 28, 2017 9:53:21 PM	Dec 28, 2017 9:53:21 PM	Movies	Looper
Dec 28, 2017 9:53:19 PM	Dec 28, 2017 9:53:19 PM	Movies	The Beguiled

Figure 23.01: A search result from I Know What You Download.

Exonerator (exonerator.torproject.org)

The Onion Router (Tor) was explained in Chapter Three. It is a network that provides anonymity by issuing IP addresses to users that connect to servers in other countries. If you possess an IP address of your target, but cannot locate any valuable information using the previous techniques, it is possible that the address was part of the Tor network and there is no relevant data to be located. Exonerator is a tool that will verify the usage of an IP address on the Tor network. Provide the IP address and a date of usage, and the service will display whether it was used as a Tor connection. While a date is required, you could provide the current date if your target time frame is unknown. Most IP addresses are either always a part of the Tor network or not connected at all.

Wigle (wigle.net)

Wigle is a crowd-sourced database of wireless access points. Users in all areas of the country conduct scans of wireless devices in their area; identify details of each device; and submit this data to Wigle in order to map the found devices on the site. This allows anyone to browse an area for wireless access points or search an address to locate specific devices. Additionally, you can search for either a specific router name or MAC address and locate any matching devices. The results will include links that will display the results on an interactive map. Most of the world has been covered. In order to take advantage of the search features, you will need to register for a free account. Generic or misleading information can be used that does not identify you.

There are many investigative uses for this service. You can identify the wireless access points in the immediate area of a target's home. As an example, a search of the address of a gas station revealed a map of it with the associated routers. In this view, I can identify the router names including potential sensitive information. It displays wireless router SSID's of AltonBPStore, tankers_network, Big Toe, and others. Clicking View and then Search in the upper left of the page presents a detailed query engine. A search of tankers_network, as identified previously in the map view, displays details of the wireless access point. It has a MAC address of 00:1F:C6:FC:1B:3F, WPA encryption, was first seen in 2011, and operates on channel 11. An investigator could also search by the target's name. This may identify routers that have the target's name within the SSID. A search of "Bazzell" identifies seven access points that probably belong to relatives with my last name. These results identify the router name, MAC address, dates, encryption method, channel, and location of the device. This can easily lead an investigator to the home of a target.

Many internet users will use the same name for their wireless router as they use for their online screen name. Assume that your target's username was "Hacker21224". A search on Wigle for "Hacker21224" as a router name might produce applicable results. These could identify the router's MAC address, encryption type, and GPS coordinates. A search on Google Maps of the supplied GPS coordinates will immediately identify the home address, a satellite view of the neighborhood, and a street view of the house of the target. All of this intelligence can be obtained from a simple username. These results would not appear on any standard search engines.

Shodan (shodan.io)

Shodan is a search engine that lets you find specific computers (routers, servers, etc.) using a variety of filters. General search engines, such as Google and Bing, are great for finding websites; however, they do not search for computers or devices. Shodan indexes "banners", which are metadata that a device sends back to a client. This can be information about the server software, what options the service supports, or a welcome message. Devices that are commonly identified through Shodan include servers, routers, online storage devices, surveillance cameras, webcams, and VOIP systems. Network security professionals use this site to identify vulnerabilities on their systems. Criminals use it to illegally access networks and alter devices. We will use it to locate specific systems near a target location. In order to take advantage of Shodan's full search capabilities, you must create a free account. Only a name and email address is required. The following example will identify how to locate live public surveillance cameras based on location. The target for this search is Mount Pleasant, Utah. The following search on Shodan produced 9,684 results.

country:US city:"Mount Pleasant"

There are two flaws with this search. First, you may receive results from other cities named Mount Pleasant. Second, you will likely receive too many results to analyze effectively. A search of "geo:39.55,-111.45" will focus only on the specific GPS location of interest (Lat=39.55, Long=-111.45). There were 238 results for this search. This is much more manageable and all of the results will be devices in the target area. Adding more specific search criteria will filter the results further. A search of "geo:39.55,-111.45 netcam" identified only one device.

The result identifies this device as a "Netcam". It also identifies the internet service provider as "Central Utah Telephone" indicating the user has a DSL connection. To connect to the device, you would click on the IP address identified as 63.78.117.229. Clicking through each of these results may be time consuming. You can add a search term to filter your results. Replicating this search for a GPS location in a large city will produce many results. Clicking the IP address will take you to the page that will connect to each device. You must be careful here. Some devices will require a username and password for access. You could try "admin" / "admin" or "guest" / "guest", but you may be breaking the law. This could be considered computer intrusion. However, many of the webcam and netcam results will not prompt you for a password and connect you to the device automatically. There is likely no law violation when connecting to a

device that does not prompt you for credentials. Your local laws may prohibit this activity. **Shodan Maps** (maps.shodan.io) allows you to conduct any of these searches based on location alone while **Shodan Images** (images.shodan.io) displays collected webcam captures from open devices. Figure 23.02 displays a home using an automated lighting and climate control system in Missouri located with Shodan Maps. These two options are premium services and require a modest fee. All Shodan features allow input of the following types of information for filtering.

City: Name of the city (ex. city:"San Diego")
Country: 2-letter country code (ex. country:US)
GPS: Latitude and longitude (ex. geo:50.23,20.06)
OS: Operating system (ex. os:Linux)
IP Address: Range (ex. net:18.7.7.0/24)
Keyword: (ex. webcam)

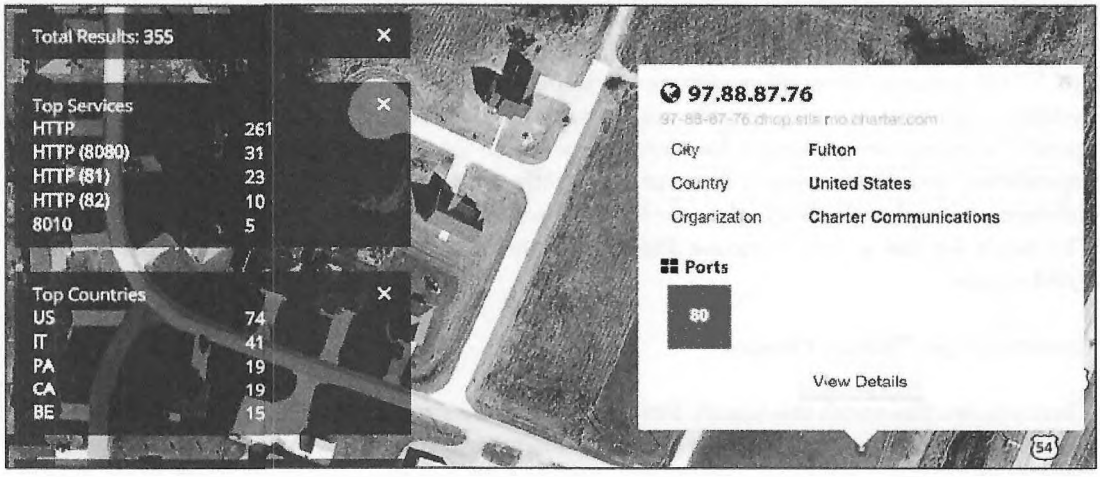


Figure 23.02: A Shodan Maps search result.

ThreatCrowd (threatcrowd.org)

As mentioned in the previous chapter, ThreatCrowd is a system for finding and researching artifacts relating to cyber threats. Searching an IP address can reveal an association to malicious software being spread over the internet. A positive result will display the type of malware, associated domain names, dates of discovery, and any comments by other researchers. Most readers that actually need this type of service likely already know more about it than me. However, it should be a consideration when investigating suspicious IP addresses.

Censys (censys.io)

Similarly, Censys is a search engine that enables researchers to ask questions about the hosts and networks that comprise the internet. Censys collects data on hosts and websites through daily scans of the internet, in turn maintaining a database of how hosts and websites are configured. Researchers can interact with this data through a search interface. As an example, a search of 173.189.238.211 reveals it to be associated with a Schneider Electric BMX P34 2020 device through a Windstream provided internet connection, located near Kansas City, Kansas.

IPv6 Addresses

The IP addresses previously mentioned were all version four (IPv4), such as 192.168.1.1. Due to limited availability, many providers are switching to IPv6, which allows many more addresses. A typical example may appear as 2001:0db8:85a3:0000:0000:8a2e:0370:7334. While many of the utilities mentioned here are adapting to this input, we should query these types of addresses through designated IPv6 engines. I have tested dozens, and I find the following two to work best, both of which are included in the search tool.

Ultra Tools (ultratools.com) **WhoIs** (who.is)

Email Headers

I no longer teach email header analysis in my live courses. The vast majority of users rely on web-based email such as Gmail or Yahoo. These services do not disclose the IP address of an individual user within the email headers. The only email headers that I have encountered over the past three years that contained valuable IP addresses were business users that sent emails within a desktop client such as Outlook. If you would like to analyze an email header in order to identify the IP address and sender information, you have two options. You can look through a few sites and teach yourself how to read this confusing data, or you can use an automated service.

IP2Location (ip2location.com/free/email-tracer) provides a large text box into which an entire email header can be copied for analysis. The response includes the IP address and location of the sender, interactive map identifying the originating location, internet service provider, and links to additional information from an IP search. Anyone wanting more information from an email threat should start here. An alternative site that conducts similar actions is **MX Toolbox** (mxtoolbox.com/EmailHeaders.aspx).

Obtaining a Target's IP Address

You may want to know the IP address of the person you are researching as provided by their internet service provider. This address could be used to verify an approximate location of the person; to provide law enforcement details that would be needed for a court order; or to

determine if multiple email addresses belong to the same subject. All of those scenarios will be explained here while I explain the various services that can be used.

IP Logger (iplogger.org)

For many years, this was my favorite option for identifying the IP address of a target. There are many options now, and most of them will be explained here. This specific technique involves some trickery and the need to contact the target from a covert account. For this demonstration, assume that your target has a Facebook page that he checks regularly. You can send him a private message that includes "bait" in the form of an online link. A detailed set of instructions should explain the processes. The main website presents several options, but only the "URL & Image Shortener" service will be explained.

Link: You can generate a URL which will redirect to any website that you provide. IP Logger will save the IP address of each user who clicked the link. In the box provided, enter any address that you want the target to see when clicking on a link. This could be something generic such as cnn.com. After submitting, you will receive a series of links. This page also serves as the log of visitors, and I recommend documenting it. In an example, I received the following link at the beginning of this list.

http://www.iplogger.org/3ySz.jpg

Although the link appears to be a jpg image, clicking this link or typing it into a browser forwards the target to cnn.com. This action collects his or her IP address, operating system, and browser details. These details, along with the date and time of capture, can be viewed at the link generated previously. A URL shortening service such as Bitly (bit.ly) would make the link look less suspicious.

Image: You can provide a digital image to this service, and it will create a tracker out of it for placement onto a website, forum, or email message. I provided an image that is present on my website at inteltechniques.com/img/bh2016.png. This presented a page similar to the previous example. I was provided the following links.

http://www.iplogger.org/23fq.jpg <img src="http://www.iplogger.org/23fq.jpg">

The first link forwards to the image that I provided. During this process, the IP address, operating system, and browser details are collected and stored on the page that stored the links. The second link could be inserted directly into a web page or email message. Upon loading either, the image is present and collects the same data.

Real World Application: I was once communicating with an unknown subject about illegal matters on a web forum about hacking and stolen credit card numbers. I wanted to find out his IP address in order to discover his true identity with a court order. I told the hacker that I had an

image of a freshly stolen debit card that I was willing to share. He requested proof, so I created an IP Logger link based on a generic online image, and embedded that link into the web forum where we were communicating. Within a few moments, I visited the log for this image and discovered his IP address in Newark, New Jersey.

Canary Tokens (canarytokens.org)

A newer option for IP identification is Canary Tokens. It offers redundant functionality as the previously mentioned product, but may be more useful to you. Ultimately, you should familiarize yourself with all options and choose which works best for you. Lately, I have found Canary Tokens to be the superior option of all. It allows creation of a PDF or DOCX file that contains a tracker, and is the most user-friendly of the services. After choosing a tracking option, it walks you through the process. I maintain a few Canary Token files at the following address. They are used as traps for people that conduct Google searches attempting to find my home address. Opening any of these alerts me to your IP address and general location. At the time of this writing, the most recent opening of one of these documents occurred only two days prior. The culprit lives in Matawan, New Jersey, possesses MCI as an internet provider, and had recently downloaded an Xbox 360 game through a torrent.

https://inteltechniques.com/canary

Always remember that technologies such as VPNs, Tor, and other forms of IP masking may create inaccurate results. Always use caution when sending these types of trackers, and make sure you are not violating any laws or internal policies. Due to the heavy usage of VPNs within the communities in which I investigate, I find these services slowly becoming less useful.

URL Biggy (urlbiggy.com)

If you want to add some flare to an IP logging link, consider this service. You can supply a link created with the previous instruction, but specify a new custom forwarding link. In other words, I can convert the file at https://inteltechniques.com/canary into urlbiggy.com/michael-bazzell-home-address. This could entice the target to click a link that may otherwise appear suspicious.

Get Notify (getnotify.com)

Get Notify tracks the opening of email messages and presents the connection information of the target. This service is completely free and does not require Gmail as your email provider. You will need to create an account through the Get Notify website and you will be limited to five email messages per day. After you have registered the email address that you will be using, you can send emails from that account as usual. However, you will need to add ".getnotify.com" after each email recipient. Instead of sending an email address to the valid account of Michael@inteltechniques.com, you would send the message to a modified email address of Michael@inteltechniques.com.getnotify.com. This will force the email message to go through

Get Notify's servers and route the message to the valid address. When your target reads the email message, Get Notify will track the user's IP address, geographical location, and notify you whether your message was viewed for a length of time or deleted right away.

Get Notify works by adding a small invisible tracking image in your outgoing emails. When your email recipient opens your message, this image gets downloaded from a Get Notify server. Get Notify will know exactly when your sent email was opened and it notifies you through an email that your sent message was read by the recipient. You can also view log files within your online account. The tracking image inserted by Get Notify is invisible to the recipient. Optionally, you can specify your own images to be used as tracking images by going to the preferences section after signing in to GetNotify.com. Your recipient will not see ".getnotify.com" at the end of his or her email address. If you want to send a single email to multiple recipients, you should add ".getnotify.com" at the end of every email address.

There are countless scenarios that may make these techniques beneficial to your online research. While I used it for law enforcement, especially in tracking down stolen goods on Craigslist, civilians can use it for many different things. Private investigators have used it on dating websites while hunting cheating spouses. Singles have used it to verify that the potential mate they have been chatting with for weeks is really local and not in another state or country. The possibilities are endless.

IntelTechniques IP Address Search Tool

Similar to the domain tool mentioned previously, this page automates some of the most common IP address searches. The first box accepts any IP address. Clicking the Populate All button will insert this address into all of the search options where manual queries can be conducted. The final option will open several tabs within your browser that present each query listed on the page. Figure 23.03 displays the current status of the service.



Figure 23.03: The IntelTechniques IP Address Tool.

Chapter Twenty-Four
Government & Business Records

Open source government and business information has never been easier to obtain. A combination of a more transparent government, cheaper digital storage costs, and marketing data leaks has placed more information online than ever before. There is no standard method of searching this data. One county may handle the queries much differently than another county and business records vary by state. The following resources and techniques should get you started in the United States.

County General Records (www.blackbookonline.info/USA-Counties.aspx)

Counties all over America have digitized the majority of their public records and allow unlimited access over the internet. Searching for your county's website will likely present many information options. This can become overwhelming and it can be easy to get lost within the pages of the site. My initial preference when just beginning an investigation is to use Black Book Online's free county public records page. It allows you to drill-down from state to county. The resulting page isolates all available records for viewing. As an example, I chose Illinois and then Madison County as my target. I was presented with the following databases, each linking directly to the source.

Coroner Reports
Delinquent Tax Sale
Government Expenditures
Property Tax Search
Public Employee Salaries
Recorded Documents
Registered Lobbyists
Press Releases
Voter Registration Verification
Voter Registration Address Search
Unclaimed Property
Crime Map
Building Contractors
Building Permits
Foreclosed Properties

County Court Records (www.blackbookonline.info/USA-County-Court-Records.aspx)

A Google search of your county of interest should identify whether an online court records database is available. As an example, St. Clair County in Illinois possesses a website that has their entire civil and criminal court records online (circuitclerk.co.st-clair.il.us/courts/Pages/icj.aspx). Searching only a last name will present profiles with full name, date of birth, physical identifiers, case history, fines, pending appearances, and more. Navigating the website will expose charged crimes even if they were dismissed. This can be extremely useful in civil litigation. There are several websites that help connect you to publicly available county government records, such as Black Book Online. It allows you to drill-down to your local records. The main page will prompt for the state desired. The result will be a list of links that access each county's court information.

Some rural areas are not online, but an occasional search should be done to see if they have been added. Repeating my previous search of Madison County, Illinois revealed the following court related databases.

Circuit Court Complete Docket
Circuit Court Attorney Docket
Family and Civil Pro Se Dockets
Felony State's Attorney Jury Trials
Traffic, Misdemeanor, DUI Docket
Traffic Citations
Crash Reports
Police Blotter
Daily Crime Log
Jail Inmate Search

If the Black Book Online options do not provide optimal results, please consider **Public Records Online** (publicrecords.onlinesearches.com).

PACER (pacer.gov)

PACER is an acronym for Public Access to Court Electronic Records. It is an electronic public access service of United States federal court documents. It allows users to obtain case and docket information from the United States district courts, United States courts of appeals, and United States bankruptcy courts. As of 2013, it holds more than 500 million documents. PACER charges $0.10 per page. The cost to access a single document is capped at $3.00, the equivalent of 30 pages. The cap does not apply to name searches, reports that are not case-specific, and transcripts of federal court proceedings. Account creation is free and if your usage does not exceed $15 in a quarter, the fees are waived. I have possessed an account for several years and have never been billed for my minimal usage. PACER has been criticized for being hard to use and for demanding fees for records which are in the public domain. In reaction, non-profit projects have begun to make such documents available online for free.

RECAP (courtlistener.com/recap)

RECAP (PACER backwards) allows users to automatically search for free copies during a search in PACER, and to help build up a free alternative database at the Internet Archive. It is an extension for the Firefox and Chrome browsers. Each PACER document is first checked if it has already been uploaded by another user to the Internet Archive. If no free version exists and the user purchases the document from PACER, it will automatically upload a copy to the Internet Archive's PACER database. While the browser extension assists greatly with searching, a search page exists on RECAP at the address above.

UniCourt (unicourt.com)

This court document search site appeared in 2018 and currently possesses an impressive database. The search options are straightforward, and results appear quickly, but you are limited to three searches without a paid membership. In my experience, clearing your browser's cache and selecting a new VPN IP address seems to reset this restriction. This is a premium site with full

court documents hidden from guests, but we can obtain a fair amount of free information. Let's conduct an actual example.

Searching "Facebook" within the official site provides numerous results. However, after clicking through three case summaries, you will likely receive a notification to purchase a monthly premium membership. This may be acceptable if you plan to use the service heavily, but there are many additional free resources available. Consider conducting your queries through Google instead. Since the case summaries are publicly available, Google appears to have indexed most, if not all, of the pages. The following Google search produced over 10,000 results.

site:unicourt.com "facebook"

The first result connected to the following static URL.

https://unicourt.com/case/ca-la23-adam-blumenkranz-vs-facebook-123515

This page returned the following case details without a subscription or registration.

Case Summary	Case Type	State
Case Number	Jurisdiction	Defendant Names
Filing Date	Judge Name	Plaintiff Names
Update Date	Courthouse	Respondent Names
Case Status	County	Docket Entries

The documents area is restricted, but the docket entries are public. In this case, the details provide dates and actions taken within the litigation. Below is a partial example. UniCourt is now a staple within my investigations, but I rely on Google to find a direct link to the data.

03/14/2018 Case Management Statement; Filed by FACEBOOK, INC. (Defendant)
03/14/2018 Reply Filed by FACEBOOK, INC.; MARK ZUCKERBERG (Defendant)
02/21/2018 NOTICE OF CONTINUANCE; Filed by Attorney for Defendant
02/21/2018 Continuance of Hearing and Order; Filed by FACEBOOK, INC.

FOIA Search (foia.gov/search.html)

The Freedom Of Information Act (FOIA) allows us to demand government records from various entities. Filing a specific request exceeds the scope of this book, but you should know that a lot of public information is already available online. This site appears to rely on Bing to index government websites, so I assumed that this resource was unnecessary. However, my attempts to replicate some of the results within Google were unsuccessful. As an example, I searched for "Darrell Bazzell" on the FOIA site and received a specific result from a 2001 committee session. I replicated this search within Google and could not retrieve the same entry.

Therefore, this service should be available within your arsenal of tools. A static submission URL is as follows, and present within the search tool.

https://search.foia.gov/search?affiliate=foia.gov&query=osint

Open Corporates (opencorporates.com)

Practically every state offers a searchable database of all businesses created or registered within the state. This will usually identify the owner(s), board members, and other associated subjects. **Dun & Bradstreet** (dnb.com) offers a business search within the home page, but many registered companies do not participate with them. In my experience, the best overall business lookup entity is Open Corporates. This free service indexes all 50 states in the U.S. plus dozens of additional countries. The records usually identify corporate officers' names, addresses, and other contact details. The basic search option allows queries by business name, entity registration number, or officer name. Clicking the advanced option allows query by physical address, but requires you to create a free account. This website is superior to targeted Google queries, because it indexes and scrapes data directly from government websites. This can assist with identifying historical results that no longer appear within the original source. I visit this resource every time I encounter a business name during my research, or identify a target that would likely be associated with an organization. Open Corporates allows search submission via URL as follows.

https://opencorporates.com/companies?q=inteltechniques
https://opencorporates.com/officers?q=bazzell

AIHIT (aihitdata.com)

This service is unique in that it uses artificial intelligence (AI) in order to populate and update business records. I do not know how much of this is legitimate process and not marketing hype, but I consistently find unique data here. The default search options query business names and registration numbers, and usually provides the same results as the previous options. However, the "More Fields" area allows entry of an email address, telephone number, or name of an individual. I have used this option to search personal email addresses and retrieve associated businesses. Fortunately, AIHIT allows search submission via URL as follows.

https://www.aihitdata.com/search/companies?k=email@gmail.com
https://www.aihitdata.com/search/companies?t=Michael Bazzell
https://www.aihitdata.com/search/companies?c=inteltechniques

Birthday Database (birthdatabase.com)

This site should identify the full name, date of birth, and city and state of birth of your U.S. target. The only available search fields are first name, last name, and approximate age. The age field is not required, but may help eliminate multiple results. A direct URL can be submitted as follows.

http://www.birthdatabase.com/cgi-bin/query.pl?textfield=JOHN&textfield2=DOE

SSN Validator (ssnvalidator.com)

A simple way to verify if a social security number is valid is at SSN Validator. This does not provide the personal information attached to the number, only verification that the number is valid. A typical response will include the state that issued the number, the year issued, verification that the number was assigned, and confirmation of death if applicable.

Social Security Death Index (genealogybank.com/explore/ssdi/all)

This public index of death records is stored on a genealogy site. The only required information is the first and last name. The results will identify birth year, death year, state of last residence, and state of SSN issue.

Legacy (legacy.com/search)

There are many websites that search for death-related information such as social security indexes and ancestry records. A leader in this area is Legacy. This site indexes online obituaries and memorials from approximately 80 percent of all online newspapers. The search on this site is straightforward and results can identify family members and locations.

Asset Locator (www.blackbookonline.info/assetsearch.aspx)

Black Book Online's Asset Locator is the most comprehensive list of sources for the search of real estate, judgments, bankruptcies, tax liens, and unclaimed funds. This page will allow you to select the type of asset you are researching and the state of the target. This will then create a new page with all of the options for that state. It will provide direct links to the sites for a search of the target. This often includes online databases of public employee salaries, vehicle registrations, property tax records, and dozens of other categories.

Vehicles

Many people assume that information related to vehicle registration and licensing is only available to law enforcement through internal networks. While a full driver's license search and complete license plate query is not publicly available, a surprising portion of related data is online for anyone to view. The following methods will display all publicly available details.

Department of Transportation (vpic.nhtsa.dot.gov)

The DOT has a website which provides free access to vehicle identification number (VIN) data. All information on this website is public information, and the data comes from vehicle

manufacturers. You can search by VIN within their database to find detailed vehicle information. I submitted a unique VIN and received the following response.

Year: 2010
Make: VOLKSWAGEN
Model: JETTA
VIN: 3VWRL7AJ6AM13xxxx
Trim Level: TDI
Style: SEDAN 4-DR
Manufactured: Mexico
Weight: 4,500 lbs

The following options also allow you to enter any VIN and retrieve the year, make, and model of the vehicle associated. The first option will often display estimated mileage based on service records.

Vin decoderz (vindecoderz.com)
CarFax (carfax.com/vehicle-history-reports)
Check That VIN (checkthatvin.com)
Search Quarry (searchquarry.com)
FaxVin (faxvin.com)
Vehicle History (vehiclehistory.com)
VinCheck (vincheck.info)

NICB VIN Check (nicb.org/vincheck)

While the previous searches will identify details about vehicles, they will not display any information about theft or salvage records. The National Insurance Crime Bureau (NICB) allows a search of any VIN and will display two unique pieces of information. The VINCheck Theft Record will identify vehicles that have been reported stolen, while the VINCheck Total Loss Records identifies VINs that belong to salvaged vehicles.

Cycle VIN (cyclevin.com)

VINs from motorcycles may not be searchable on standard VIN engines due to the amount of characters in them. Cycle VIN will display a year and make, as well as any indication that the VIN exists in its proprietary database. If it does, $25 will obtain title and mileage information. I only use this as a free resource for verifying motorcycle VINs to the correct year and make.

Vehicle Registration

Several free services identify the year, make, and model of a vehicle after supplying the VIN. However, it is more likely that we know the license plate registration details rather than a VIN. Fortunately, we have many options for researching these plates. The following services provide a search option based on the vehicle registration plate and state. Results are hit-or-miss, and rarely include a name, but many will identify the VIN for further research.

Reverse Genie (reversegenie.com/plate.php)
Auto Check (autocheck.com)
Records Finder (recordsfinder.com/plate)

CarFax (carfax.com/vehicle-history-reports)
Search Quarry (searchquarry.com/vehicle_records)
Free Background Search (freebackgroundcheck.org)
Carvana (carvana.com/sellyourcar/getoffer/vehicle)
VinCheck (vincheck.info/free-license-plate-lookup/)
Vehicle History (vehiclehistory.com/license-plate-search)

As a test, I submitted a vehicle's registration number which was displayed on a television show playing at the airport while I wrote this section. The result from almost all of these sites correctly identified the vehicle as a 2010 Dodge Avenger. While only a small piece of information, it works in conjunction with other search techniques. After exhausting all of these searches, you should be able to obtain the VIN, make, model, year, engine, and style of the vehicle. These options will not typically provide the name of the owner.

O'Reilly Auto Parts (oreillyauto.com)

If the previous search options fail to identify the VIN, year, make, and model of a vehicle based on the license plate, try O'Reilly. As a service to potential customers, it allows you to enter your license plate in order to identify applicable parts for your vehicle. Select the "Shop by Vehicle" in the upper-right and enter the license plate and state. I provided the state of California and the plate of "HACKER". I received the following response.

2007 GMC Sierra 1500 SLE
V8,6.0,5967,364,Electronic,SFI,GAS,FI,MFI

When I searched the same plate on VinCheck, I received a VIN of 3GTEK13Y87G527460. I now have a decent beginning to my investigation of a license plate.

Progressive (progressive.com)

While not an official vehicle search, the insurance provider Progressive offers an interesting piece of information. I first learned about this technique from S.L., a member of my online OSINT forum. When you view the home page at progressive.com, you are prompted to request a free insurance quote. If you provide the zip code and address of any target, you receive a summary of the year, make, and model of all vehicles registered at that address. You can supply any random data besides the physical address and receive the results. This was likely designed to make the quote process more efficient and accurate, but investigators should appreciate the free utility.

Marine Traffic and Boat Information

There is an abundance of details available about global marine traffic within ownership records and real-time monitoring. **Marine Traffic** (marinetraffic.com) provides an interactive map that displays the current location of all registered ships and boats. Clicking on any vessel provides the

name, speed, collection time, and destination. **Boat Info World** (boatinfoworld.com) allows the search of a boat name and provides the following details.

Boat Name
Boat Owner
Record Date
Registered Address
Hull ID
Hailing Port
Lloyd's Registry Number
Call Sign
Coast Guard Vessel ID
Service Type
Boat's Length
Boat's Gross Tons
Vessel Build Year
Ship Builder
Hull Shape
Propulsion Type

Aircraft Information

Monitoring aircraft during flight and searching historical ownership records is relatively easy. Commercial planes constantly announce their location with automated reporting systems and tail numbers act similarly to a vehicle's registration plate. Today, this information is publicly available on multiple websites. **Plane Finder** (planefinder.net) displays an interactive global map identifying all known aircraft currently in flight. Hovering over a selection displays the carrier, flight number, originating departure, destination, speed, and altitude. Historical ownership records are available on multiple websites and none are completely accurate. I recommend **Black Book Online's** aviation page (blackbookonline.info/Aviation-public-Records.aspx). At the time of this writing, it provided direct links to the following databases.

Aircraft N Number Search
Aircraft Ownership Search
Airline Certificates
Airport Profiles
Certified Pilots
Cockpit Voice Recorder Database
Flight Tracker
Military Aviation Crash Reports

Campaign Contributions

Any contributions to political campaigns are public record. Searching this is now easy thanks to three separate websites. These sites will search using information as minimal as a last name. Including the full name and year will provide many details about the target. This includes occupation, the recipient of the contribution, the amount, the type of contribution, and a link to the official filing that contains the information. After an initial search is conducted, you will receive additional search tabs that will allow you to filter by zip code, occupation, and year. Melissa Data allows you to search a zip code and identify all political donations for a specified year. The results from these sites may be redundant, but often contain unique data.

Open Secrets (opensecrets.org)
Money Line (politicalmoneyline.com)
Melissa Data (melissa.com/v2/lookups/fec/index)

Criminal Information

If a target has a criminal past, there is probably evidence of this on the internet. County court searches will identify most of this information, but this requires a separate search on each county's website. There are a handful of services that attempt to locate nationwide information by name.

Family Watch Dog (familywatchdog.us)

This is one of the leading sites in identifying public criminal information about sex offenders. The main page includes a "Find Offender" area on the left side. You can search here by address or name. The name search only requires a last name to display results. This will identify registered sex offenders that match the criteria specified. This will include a photograph of the target and details of the offense.

Felon Spy (felonspy.com)

This site can appear difficult to navigate at first. Most of the search fields for information forward to a sponsored result that will demand a fee for the information. The only free way to search this data is to click on the "Begin Search" button overlapping the map in the middle of the page. The only fields that should be searched on this page are in the top row and include address, city, and state. Entering any address will display markers on a map of convicted felons in that area.

Inmate Searches

Both federal and state prisons offer prisoner details online. The amount of detail will vary by state, but most will include photographs of the target and details of the crime. In most states, this information is maintained in public view after the target is released, if the subject is still on probation or parole. Federal prisoners can be located at www.bop.gov/inmateloc. A first and last name is required for a search. Each state maintains its own database of prisoner information. Conducting a search on Google of "Inmate locator" plus the state of interest should present official search options for that state.

VINELink (vinelink.com)

VINELink is an online portal to VINE, a victim notification network. VINE has been providing victims and concerned citizens with information for decades, allowing individuals to access reliable information about custody status changes and criminal case information. After choosing the state of interest, you can select from the following options.

Find an Offender: Get info and register to be notified of custody status changes.
Find an Offender Court Case: Get info and register to be notified of offender court dates.
Find Sex Offender Registry Status: Get info about sex offender registry status changes.
Find a Protective Order: Get info and register to be notified of protective order status changes.

Selective Service Verification (sss.gov/Home/Verification)

This website requires a last name, social security number, and date of birth of the target. The result will identify the person's full name, selective service number, and date of registration.

Voter Registration Records

The election of 2016 caused a lot of controversy in regard to the use and collection of voter registration data. While these personal details are public record, many people did not believe it was appropriate for politicians to use this personal data as a part of their campaign strategies. Regardless of your opinion on these matters, much of the voter registration details are available online. The most beneficial site I have found is **Voter Records** (voterrecords.com). You can search by name or browse by state. Any results will identify full name, home address, mailing address, gender, party affiliation, age, and relatives. Currently, databases are available for Alaska, Arkansas, Colorado, Connecticut, Delaware, Florida, Michigan, Nevada, North Carolina, Ohio, Oklahoma, Rhode Island, Utah, and Washington.

High Programmer (highprogrammer.com/cgi-bin/uniqueid)

Most states use some type of algorithm to create a driver's license number for a person. Often, this number is generated from the person's name, sex, and date of birth. After you have determined your target's middle initial and date of birth from the previous websites mentioned, you can use this data to identify the target's driver's license number. High Programmer will automate this process for the following states:

Florida	Michigan	New York
Illinois	Minnesota	Washington
Maryland	New Hampshire	Wisconsin

BinDB (www.bindb.com/bin-database.html)

While not technically government data, I felt that this option fits best in this chapter. This website will allow you to enter the first six digits of any credit card number and identify the brand, issuing bank, card type, card level, country, bank website, and customer care line.

Real World Application: While working in the homicide division, I often identified credit or debit card numbers of my victims. If the actual card was located, I did not need this service. However, if only the number was located, this service helped to identify the financial institution and a contact number. In one specific investigation, I had learned that the victim had eaten at a local restaurant the evening prior to her suspicious death. Visiting the restaurant allowed me to acquire the billing details of her dinner, which identified the debit card that she used for payment. Searching this number through BinDB identified the issuing bank and customer care number. Calling the number presented an automated self-service feature for members of that bank.

Entering the newly found debit card number and the zip code of the victim allowed me to access the previous 30 days of charges to her account. This quickly identified an otherwise unknown ATM withdrawal on the day of her killing. Retrieving video from that ATM machine displayed a passenger in her vehicle. This initiated a new investigation which eventually led to the killer.

Lead Ferret (leadferret.com/search)

This service provides interactive access to a typical "B2B" (Business to Business) database. These data sets often include detailed business and employee records. Sales people use this content as leads for potential new clients. We will use it to query data about our own targets. This search page is massive, and I will only focus on a few key areas. Section 6 presents company search options such as the business name, address, website domain, and telephone number. The search button for any query is in the upper-left, which is a bit awkward. Figure 24.01 displays a typical result when searching by company. Clicking on any name presents the format seen in Figure 24.02. It displays full details including social network links, telephone numbers, titles, and revenue. The name search in the upper right allows query of any employee.

☐	Mark	Wright	Creative Director	CNN News Group	Atlanta	GA
☐	Sam	Feist	Executive Producer	Time Warner Inc.	Washington	DC
☐	Dan	Crane	Vice President Operations	Time Warner Inc.	Washington	DC

Figure 24.01: A Lead Ferret company result.

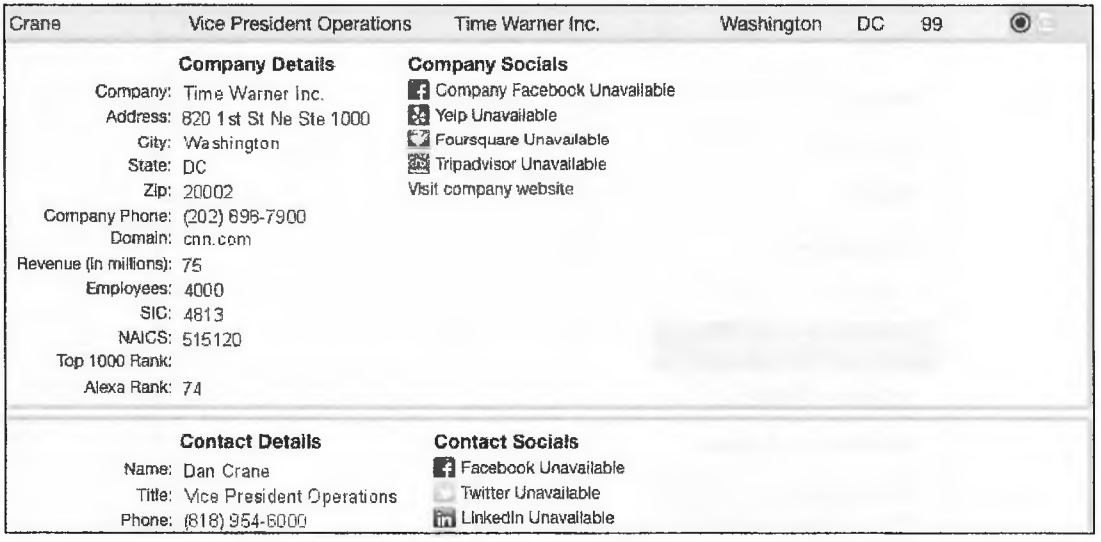


Figure 24.02: A Lead Ferret name result.

IntelTechniques Business & Government Search Tool

While minimal, this tool should assist with replicating some of the searches mentioned within this chapter. Figure 24.03 displays the current state of the tool.

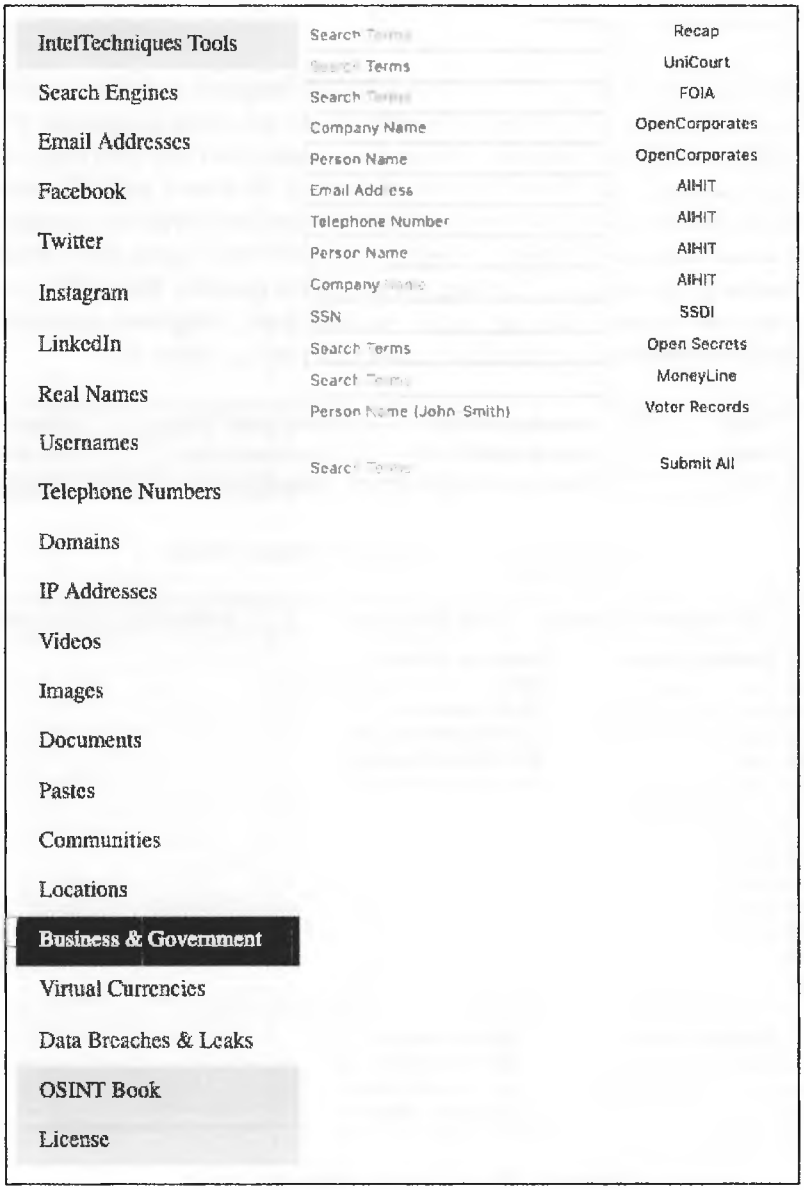


Figure 24.03: The IntelTechniques Business & Government Search Tool.

Chapter Twenty-Five
Virtual Currencies

In simplest terms, virtual currencies can be spent for goods and services without connection to a person or bank account. It has no physical presence, and is mostly used online as digital payment. Bitcoin is virtual currency. A bitcoin address, which is an identifier you use to send bitcoins to another person, appears similar to a long string of random characters. In our demo, we will use 12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw, which is the real address that was used to collect ransom from victims after malicious software had taken over their computers. Think of a Bitcoin address as an email address. That address stores their "virtual" money.

Blockchain (blockchain.info)

This website allows search of a Bitcoin address and displays the number of transactions, total amount of Bitcoin received ($), final balance, and a complete transaction history. We can track every incoming and outgoing payment. This will almost never be associated with any real names, but it provides a great level of detail about the account. We learn that this account has received 19.12688736 Bitcoin worth $ 287,391.14 USD at the time of this writing.

Bitcoin Who's Who (bitcoinwhoswho.com)

Our next stop is a service that provides a bit more analysis about the suspect account. We immediately learn that it is a suspect ransomware account, and that the address has appeared on various news outlet websites. Furthermore, we see transaction IP addresses, which are likely behind VPNs. Overall, I use Blockchain for transaction details and Bitcoin Who's Who to get a better idea of why I might care about the account.

BlockChair (blockchair.com)

This service is very similar to Blockchain, but I find it has better representation across multiple virtual currencies. Additionally, we can query each currency via URL, which will assist in our tools. Let's start with a search of a Bitcoin address at the following URL.

https://blockchair.com/bitcoin/address/1EzwoHtiXB4iFwedPr49iywjZn2nnekhoj

The results are typical, and include balance and transaction data. The power of BlockChair is the ability to search Bitcoin, Ethereum, Ripple, Bitcoin Cash, Litecoin, Bitcoin SV, Dash, Dogecoin and Groestlcoin. We will use the following URLs for each, replacing "xxx" with the target address.

https://blockchair.com/ethereum/address/xxx
https://blockchair.com/ripple/address/xxx
https://blockchair.com/bitcoin-cash/transaction/xxx
https://blockchair.com/litecoin/address/xxx
https://blockchair.com/bitcoin-sv/address/xxx
https://blockchair.com/dash/address/xxx
https://blockchair.com/dogecoin/address/xxx
https://blockchair.com/groestlcoin/address/xxx

Bitcoin Abuse (bitcoinabuse.com)

This service focuses on one feature. It notifies you if others have reported a target virtual currency address as associated with malicious activity. This often provides valuable information about an investigation. Consider an actual report located at the following URL.

https://www.bitcoinabuse.com/reports/1KUKcwCv64cXQZa4csaA1cF3PPTio6Yt2t

The results include a summary of the activity and the email addresses sending malicious email.

Sep 21, 2019	sextortion	peter6389dd@excite.co.uk	"Hacked computer email"
Sep 21, 2019	ransomware	addntfogjnfi@activeware.com	Claims to hack computer

Wallet Explorer (walletexplorer.com)

The previous utilities examined an individual virtual currency account, such as a Bitcoin address. Many people possess numerous addresses and store them all within a virtual wallet. This is where Wallet Explorer can be extremely beneficial. While researching one of our target Bitcoin addresses within this free service, the results identified a wallet of "00037fd441" which contained the target address. Clicking on the link to this wallet revealed multiple new transactions from additional Bitcoin addresses previously unknown. This step is vital in order to track all transactions associated with your suspect. The following URLs search an address and a wallet.

https://www.walletexplorer.com/address/1EzwoHtiXB4iFwedPr49iywjZn2nnekhoj
https://www.walletexplorer.com/wallet/00037fd441938ba4

Virtual Currency APIs

In order to create the custom search tool presented at the end of this chapter, I needed a very simple way to query virtual currency addresses for various tasks. Many of the websites which allow searching of Bitcoin addresses do not permit submission via URL. Instead, I will take advantage of various Application Programming Interfaces (APIs) which allow us to query directly

and receive a text-only result. The following URLs are used within the tool, with an explanation of each. Each display of "xxx" is where the virtual currency address or amount would be inserted.
Validation: The following URL provides an indication whether a provided address is valid or invalid. This is a great first search to make sure you have a proper address.

http://codacoin.com/api/public.php?request=validate&address=xxx

Value: The following URL presents the current value of one Bitcoin.

https://blockchain.info/q/24hrprice

Received: This URL displays the total amount of Bitcoin received by a specific address. It is important to note that this amount will be in "Satoshi". A Satoshi is equal to 0.00000001 Bitcoin. Put another way, one bitcoin contains 100 million Satoshis. This unit of measurement is popular because a single Bitcoin is currently worth approximately $10,000. The Satoshi is a more precise number. In a moment, we will convert Satoshi to USD.

https://blockchain.info/q/getreceivedbyaddress/xxx

Sent: This URL displays the total amount of Bitcoin sent by a specific address. It is important to note that this amount is also presented in "Satoshi" (0.00000001 Bitcoin).

https://blockchain.info/q/getsentbyaddress/xxx

Balance: This utility displays the current balance of an address in "Satoshi".

https://blockchain.info/q/addressbalance/xxx

BTC > USD Value: The following URL will always display the current value of any amount of Bitcoin in USD. This price fluctuates hourly. Replace "xxx" with your value of Bitcoin.

http://codacoin.com/api/public.php?request=convert&type=btctofiat&input=xxx&symbol=enabled&decimal=2&exchange=average&currency=USD&denom=bitcoin

USD > BTC Value: The following URL will always display the current Bitcoin value of any amount of USD. This price fluctuates hourly. Replace "xxx" with your value of USD.

http://codacoin.com/api/public.php?request=convert&type=fiattobtc&input=xxx&symbol=enabled&decimal=10&exchange=average&currency=USD&denom=bitcoin

Satoshi > USD Value: The following URL will always display the current value of any amount of Satoshi in USD. This price fluctuates hourly. Replace "xxx" with your value of Satoshi.

http://codacoin.com/api/public.php?request=convert&type=btctofiat&input=xxx&symbol=enabled&decimal=2&exchange=average&currency=USD&denom=satoshi

USD > Satoshi Value: The following URL will always display the current Satoshi value of any amount of USD. This price fluctuates hourly. Replace "xxx" with your value of USD.

http://codacoin.com/api/public.php?request=convert&type=fiattobtc&input=xxx&symbol=enabled&decimal=10&exchange=average&currency=USD&denom=satoshi

Summary: This URL displays a brief summary of a Bitcoin address including total received, total sent, balance, total transactions, first transaction, and most recent transaction. Replace "xxx" with your virtual currency address.

https://chain.api.btc.com/v3/address/xxx

First seen: This Blockchain query displays the date which a virtual currency address transaction was first seen within the public blockchain. Note that this result will appear in Unix time format, but our tools will allow you to convert this to traditional time format. Replace "xxx" with your virtual currency address.

https://blockchain.info/q/addressfirstseen/xxx

IntelTechniques Virtual Currency Tool

This tool simplifies the various techniques explained in this chapter. Note that the Virtual Currencies tool possesses an empty window to the right, as seen in Figure 25.01. This is technically an Iframe, and allows results from the first section of queries to appear directly within your tool. For those interested in the technology behind this, below is a sample line within the search tool.

window.open('http://codacoin.com/api/public.php?address=' + Search01, 'frame1')

Within previous tools, the target URL from a search opened as a new tab (or window), such as "Search01window". Here, we have specified to open the content within "frame1". The following code represents the empty frame.

<iframe name='frame1' width='400px' height='550px' frameborder=1> </iframe>

This designates the Iframe name (frame1), width (400 pixels), height (550 pixels), border presence (1), and end of the script (</iframe>). Viewing the source code behind Currencies.html may help understand this concept for future replication and modification.

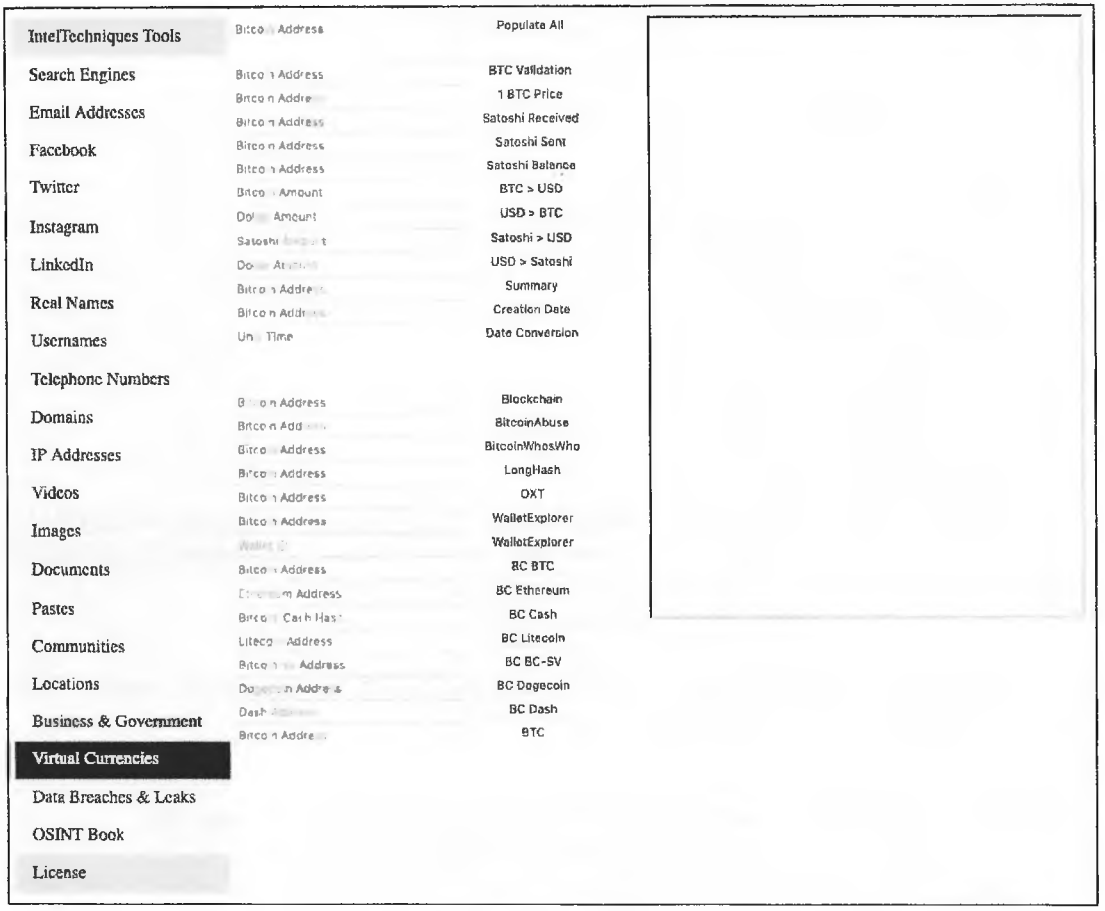


Figure 25.01: The IntelTechniques Virtual Currency Tool

Investigation Summary

Now that you understand the details available about a virtual currency address, let's use the custom tools to run through a typical investigation. Assume you are investigating a Bitcoin address of 1EzwoHtiXB4iFwedPr49iywjZn2nnekhoj. It was used as part of an extortion email, and you have been tasked to find any information about the address. First you input the address into the search tool. The following information would be presented after each of the options within the first section.

BTC Validation: Valid (The address is a proper format)
1 BTC Price: $9,978.23 (The current value of one Bitcoin)
Satoshi Received: 716409285544 (The total amount of received currency)
Satoshi Sent: 716371585974 (The total amount of sent currency)

Satoshi Balance: 37699570 (The total amount of the current balance)
Satoshi > USD: (Used to convert Satoshi to USD as follows)
Received: $71,408,167.18
Sent: 71,405,913.85
Balance: 3,757.79
Summary:
"address": "1EzwoHtiXB4iFwedPr49iywjZn2nnekhoj",
"received": 716409285544,
"sent": 716371585974,
"balance": 37699570,
"tx_count": 3534,
"unconfirmed_tx_count": 0,
"unconfirmed_received": 0,
"unconfirmed_sent": 0,
"unspent_tx_count": 3,
"first_tx": "6cc1542feb7abcff6364e0d31fc75097e0ecf7dae897ad6de6a2c1c5a1261316",
"last_tx": "e11525fe2e057fb19ec741ddcb972ec994f70348646368d960446a92c4d76dad"
Creation Date: 1331482301 (Unix time when the address was first seen)
Date Conversion: Mar-11-2012 10:11:41 (Time in UTC)

The second section will query the various services mentioned throughout the chapter. The following highlights some of the findings.

Blockchain: A detailed list of all transactions.
BitcoinAbuse: One report of malicious activity and a new email address to research.
BitcoinWhosWho: Links to online references to the address on Reddit.
WalletExplorer Address: Transaction details and wallet ID of 00037fd441.
WalletExplorer Wallet: Several pages of additional Bitcoin addresses within the suspect's wallet.

You could now repeat the process with the new Bitcoin addresses with hopes of identifying more email addresses. The email address search tool may identify further information about your target. While this is all very time consuming, the tools should simplify the queries.

CHAPTER TWENTY-SIX
ADVANCED LINUX TOOLS

The first section of this book focused heavily on a Linux virtual machine, and I explained numerous applications which assist our investigations. Those programs demanded a single piece of information, which allowed me to present scripts to automate the processes. As an example, the Instagram tools prompt you for a username and then execute the proper commands in order to simplify the process. This chapter is quite different. It contains some advanced applications which would be impossible to automate with simple scripts. These programs require heavy user input and all features must be tackled within Terminal. Some may skip this chapter until they become more comfortable within Linux. However, I believe those within any skill level can replicate the tutorials.

Recon-ng

Recon-ng is a full-featured web reconnaissance framework written in Python. Complete with independent modules, database interaction, built-in convenience functions, interactive help, and command completion, Recon-ng provides a powerful environment in which OSINT research can be conducted quickly and thoroughly. This utility provides automation to many of the redundant tasks that OSINT examiners find themselves performing on a daily basis. I offer a warning before proceeding. This is a technically complicated portion of this book. Please don't let that scare you off, we will approach each step slowly. First, we need to install Recon-ng into our OSINT Master virtual machine. Type the following into Terminal.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/lanmaster53/recon-ng.git`
- `cd recon-ng`
- `sudo pip3 install -r REQUIREMENTS`

This should install Recon-ng, and typing ./recon-ng from within this directory launches the application. This screen will display the current version, and I am writing this based on version 5.0.1. In a moment, I will present a desktop shortcut which can be used to launch the program, which eliminates the need to navigate to ~/Downloads/Programs/recon-ng every time you want to use the application. The following command would be used to update your version of Recon-ng, and I recommend adding this to your update script mentioned previously. It is not in there by default because many readers may not choose to install the programs mentioned in this chapter. I will present the entire update script toward the end.

- `cd ~/Downloads/Programs/recon-ng`
- `git pull https://github.com/lanmaster53/recon-ng.git`

Recon-ng does not possess many online tutorials. The guides that I have found are mostly an index of commands with little explanation. Instead of trying to summarize how the program functions, I will walk you through actual usage and explain as we go. I will start with the basics and then conduct numerous actual searches. In lieu of screen captures, I will include all text input and output in **9 pt Terminal font**. Upon executing Recon-ng, you will be notified that no modules are installed. This is normal, and we will add them as we need them. At this prompt, let's begin with the help command. Typing **help** reveals the following commands and explanations.

```
back            Exits the current context
dashboard       Displays a summary of activity
db              Interfaces with the workspace's database
exit            Exits the framework
help            Displays this menu
index           Creates a module index (dev only)
keys            Manages third party resource credentials
marketplace     Interfaces with the module marketplace
modules         Interfaces with installed modules
options         Manages the current context options
pdb             Starts a Python Debugger session (dev only)
script          Records and executes command scripts
shell           Executes shell commands
show            Shows various framework items
snapshots       Manages workspace snapshots
spool           Spools output to a file
workspaces      Manages workspaces
```

Typing **marketplace search** will reveal the current functions available. Think of the marketplace similar to a list of utilities within Recon-ng, and each option as a "resource". Just like Bing is a website resource that we can use through a web browser, "bing_domain_web" is a specific resource that we can use in Recon-ng. The following modules were available at the time of this writing. We will use some of these during the instruction.

```
discovery/info_disclosure/cache_snoop
discovery/info_disclosure/interesting_files
exploitation/injection/command_injector
exploitation/injection/xpath_bruter
import/csv_file
import/list
import/nmap
recon/companies-contacts/bing_linkedin_cache
recon/companies-contacts/pen
recon/companies-domains/pen
recon/companies-domains/viewdns_reverse_whois
recon/companies-multi/github_miner
recon/companies-multi/shodan_org
recon/companies-multi/whois_miner
recon/contacts-contacts/mailtester
```

```
recon/contacts-contacts/mangle
recon/contacts-contacts/unmangle
recon/contacts-credentials/hibp_breach
recon/contacts-credentials/hibp_paste
recon/contacts-domains/migrate_contacts
recon/contacts-profiles/fullcontact
recon/credentials-credentials/adobe
recon/credentials-credentials/bozocrack
recon/credentials-credentials/hashes_org
recon/domains-companies/pen
recon/domains-contacts/metacrawler
recon/domains-contacts/pen
recon/domains-contacts/pgp_search
recon/domains-contacts/whois_pocs
recon/domains-credentials/pwnedlist/account_creds
recon/domains-credentials/pwnedlist/api_usage
recon/domains-credentials/pwnedlist/domain_creds
recon/domains-credentials/pwnedlist/domain_ispwned
recon/domains-credentials/pwnedlist/leak_lookup
recon/domains-credentials/pwnedlist/leaks_dump
recon/domains-domains/brute_suffix
recon/domains-hosts/binaryedge
recon/domains-hosts/bing_domain_api
recon/domains-hosts/bing_domain_web
recon/domains-hosts/brute_hosts
recon/domains-hosts/builtwith
recon/domains-hosts/certificate_transparency
recon/domains-hosts/findsubdomains
recon/domains-hosts/google_site_web
recon/domains-hosts/hackertarget
recon/domains-hosts/mx_spf_ip
recon/domains-hosts/netcraft
recon/domains-hosts/shodan_hostname
recon/domains-hosts/ssl_san
recon/domains-hosts/threatcrowd
recon/domains-hosts/threatminer
recon/domains-vulnerabilities/ghdb
recon/domains-vulnerabilities/xssed
recon/domains-vulnerabilities/xssposed
recon/hosts-domains/migrate_hosts
recon/hosts-hosts/bing_ip
recon/hosts-hosts/ipinfodb
recon/hosts-hosts/ipstack
recon/hosts-hosts/resolve
recon/hosts-hosts/reverse_resolve
recon/hosts-hosts/ssltools
recon/hosts-hosts/virustotal
recon/hosts-locations/migrate_hosts
recon/hosts-ports/binaryedge
recon/hosts-ports/shodan_ip
recon/locations-locations/geocode
```

```
recon/locations-locations/reverse_geocode
recon/locations-pushpins/flickr
recon/locations-pushpins/shodan
recon/locations-pushpins/twitter
recon/locations-pushpins/youtube
recon/netblocks-companies/whois_orgs
recon/netblocks-hosts/reverse_resolve
recon/netblocks-hosts/shodan_net
recon/netblocks-hosts/virustotal
recon/netblocks-ports/census_2012
recon/netblocks-ports/censysio
recon/ports-hosts/migrate_ports
recon/profiles-contacts/bing_linkedin_contacts
recon/profiles-contacts/dev_diver
recon/profiles-contacts/github_users
recon/profiles-profiles/namechk
recon/profiles-profiles/profiler
recon/profiles-profiles/twitter_mentioned
recon/profiles-profiles/twitter_mentions
recon/profiles-repositories/github_repos
recon/repositories-profiles/github_commits
recon/repositories-vulnerabilities/gists_search
recon/repositories-vulnerabilities/github_dorks
reporting/csv
reporting/html
reporting/json
reporting/list
reporting/proxifier
reporting/pushpin
reporting/xlsx
reporting/xml
```

At any time, you can type **`marketplace info`** into Recon-ng to receive details about a specific item. As an example, typing **`marketplace info profiler`** displays the following description.

```
description    Takes each username from the profiles table and searches a variety of
web sites for those users. The list of valid sites comes from the parent project at
https://github.com/WebBreacher/WhatsMyName |
required_keys  []
dependencies   []
status         not installed
```

This provides the detailed description, and whether the utility requires an API key or other dependencies. It also confirms we have not installed the module. We will execute this option later in the chapter. For now, we must set up our first investigation.

Before we can conduct any research within this program, we must create a workspace. A workspace is a container that will isolate your work from one investigation to another. Think of a workspace as a case file. You may have a stack of cases on your desk, each with its own folder.

All of your work on a case stays within the folder associated. Workspaces are similar. You should create a new workspace for each investigation. They can be deleted later or preserved for additional work. You can type **workspaces list** at any time to see the currently used workspaces. For now, we will create a new workspace titled OSINT by executing a command of **workspaces add OSINT**.

After creation, you will automatically begin using the new workspace. If you have created more than one workspace, such as one titled OSINT2, you can switch to it by typing **workspaces select OSINT2**. You might have a workspace for every target suspect or a single workspace for an entire case. Each situation will be unique. Now that you have a space created, we can begin.

Let's start with a very simple yet powerful query, using the Profiler module previously mentioned. First, we must install the module with the following command within Recon-ng.

```
marketplace install profiler
```

The module is now installed, but is not loaded. The following loads the module.

```
modules load profiler
```

Now that the module is loaded, we can add any input desired. Since this module queries usernames, we will add our target of "inteltechniques" with the following command. Note that SOURCE is uppercase, which is required.

```
options set SOURCE inteltechniques
```

We should test our input with the following command.

```
input
```

The response should now be the following.

```
+-----------------+
|  Module Inputs  |
+-----------------+
| inteltechniques |
+-----------------+
```

Finally, we can launch the module with the following command.

```
run
```

This script should query the username of inteltechniques against numerous online services. It does not present any results when complete, but did locate and store valuable data. To view the results, type the following command.

```
show profiles
```

The results appear similar to the following.

```
+---+---------------+-----------+-----------------------------------------------+--------+
| 1 | inteltechniques | Gravatar | http://en.gravatar.com/profiles/inteltechniques | images |
| 2 | inteltechniques | reddit   | https://www.reddit.com/user/inteltechniques     | news   |
| 3 | inteltechniques | Twitter  | https://twitter.com/inteltechniques             | social |
+---+---------------+-----------+-----------------------------------------------+--------+
```

In just a few seconds, we queried dozens of online services and immediately received only the three which contained the presence of our target username. This demonstrates the ability to save a substantial amount of time by using Recon-ng. If you were tasked to locate online profiles of ten suspects, this could be completed in a few minutes. Let's repeat the process, but with another username, with the following commands.

```
set SOURCE humanhacker
run
show profiles
```

The following result displays the additional online profiles collected during this second query. Recon-ng will continue to store target data as you receive it. This is one of the more powerful features of the application.

```
+---+---------------+-----------+-----------------------------------------------+----------+
| 1 | inteltechniques | Gravatar | http://en.gravatar.com/profiles/inteltechniques | images   |
| 2 | inteltechniques | reddit   | https://www.reddit.com/user/inteltechniques     | news     |
| 3 | inteltechniques | Twitter  | https://twitter.com/inteltechniques             | social   |
| 4 | humanhacker     | Blogspot | http://humanhacker.blogspot.com                 | blog     |
| 5 | humanhacker     | Disqus   | https://disqus.com/by/humanhacker/              | discussion |
| 6 | humanhacker     | Flipboard | https://flipboard.com/@humanhacker             | tech     |
| 7 | humanhacker     | GitHub   | https://api.github.com/users/humanhacker        | coding   |
| 8 | humanhacker     | Instagram | https://www.instagram.com/humanhacker/          | social   |
| 9 | humanhacker     | Kongregate | http://www.kongregate.com/accounts/humanhacker | gaming   |
| 10 | humanhacker    | Kik      | https://kik.me/humanhacker                      | social   |
| 11 | humanhacker    | Medium   | https://medium.com/@humanhacker/latest          | news     |
| 12 | humanhacker    | Technet  | https://social.technet.microsoft.com/humanhacker/ | tech  |
| 13 | humanhacker    | Minecraft | https://namemc.com/name/humanhacker            | gaming   |
| 14 | humanhacker    | Pornhub  | https://www.pornhub.com/users/humanhacker       | XXX PORN |
| 15 | humanhacker    | scratch  | https://scratch.mit.edu/users/humanhacker/      | coding   |
| 16 | humanhacker    | reddit   | https://www.reddit.com/user/humanhacker         | news     |
| 17 | humanhacker    | Twitch.tv | https://passport.twitch.tv/usernames/humanhacker | gaming |
| 18 | humanhacker    | Twitter  | https://twitter.com/humanhacker                 | social   |
| 19 | humanhacker    | Xbox     | https://www.xboxgamertag.com/search/humanhacker/ | gaming   |
+---+---------------+-----------+-----------------------------------------------+----------+
```

Let's conduct another example within a different module. First, we must leave our current module by typing **back**. This returns us to our workspace. Next, install four additional modules with the following commands.

```
marketplace install bing_domain_web
marketplace install google_site_web
```

```
marketplace install brute_suffix
marketplace install pgp_search
```

We will use these in order and target the website cnn.com. First, we will load the bing_domain_web option with the command of **modules load bing_domain_web**. Next, we will set our source with **options set SOURCE cnn.com** and execute the script with **run**. This command queries the Bing search engine for hosts connected to the domain cnn.com. The result identified over 70 unique hosts, including the following.

```
[*] [host] internationaldesk.blogs.cnn.com (<blank>)
[*] [host] crossfire.blogs.cnn.com (<blank>)
[*] [host] reliablesources.blogs.cnn.com (<blank>)
[*] [host] lightyears.blogs.cnn.com (<blank>)
[*] [host] commercial.cnn.com (<blank>)
[*] [host] collection.cnn.com (<blank>)
```

We can replicate this type of search on Google to make sure that we are not missing any hosts that could be valuable by typing **back**, then **modules load google_site_web**, then **options set SOURCE cnn.com**, and finally **run**. This notifies us **38 total (15 new) hosts found**, which indicates that Bing found more hosts than Google, and Google found 15 hosts that we did not have in our collection from Bing. Since Recon-ng can parse out duplicates, we should have a list of unique hosts with a combined effort from both Google and Bing. Typing **show hosts** will display all of them. Below is a small portion.

```
news.blogs.cnn.com
m.cnn.com
buzz.money.cnn.com
thechart.blogs.cnn.com
globalpublicsquare.blogs.cnn.com
tech.fortune.cnn.com
```

Next, let's type **back** to leave the current module and then **modules load brute_suffix** to load our next demo. Since there is no domain set as our source for this module, we will add one with **options set SOURCE social-engineer.org**. There are many top-level domains (TLDs) aside from .com and .org. Executing **run** will scour the various TLDs such as .net, .tv, and others. After completion, typing **show domains** again will display our updated set of target addresses ready for further searching. In this example, I was notified that 11 additional domains were located, including the following.

```
social-engineer.be
social-engineer.ch
social-engineer.com
social-engineer.de
social-engineer.dev
social-engineer.info
social-engineer.me
social-engineer.net
social-engineer.se
social-engineer.training
social-engineer.us
```

These are all new leads that should be analyzed later. We could now repeat our previous module execution of **bing_domain_web** and **google_site_web** to likely grow our list of hosts substantially. This is a good time to pause and consider what is happening here. As we find data, Recon-ng stores it within our workspace. Every time we conduct a new search, or repeat a previous search, all of the new data is appended. This prevents us from documenting everything that we locate because Recon-ng is keeping good notes for us. This can allow us to collect an amount of data otherwise impossible to manage manually. Let's move on to individual contacts.

Typing **show contacts** will display any contacts stored within the current workspace. You likely do not have any, so let's add some. First, type **back** to make sure you are out of the previous module. Next, load our next module with **modules load pgp_search**. This will scan all of the stored domains that we have located and search for any email addresses associated with public PGP keys within those domains. We have not set a source for this module, but you likely already have some ready for you. In a previous example, you searched social-engineer.org within other top-level domains and received numerous results. If you type **input** within this module, you should see those same domains listed. This is because Recon-ng is constantly storing found data and making it available for future use. If we type **run**, this list will be searched, but no results will be found. Note that this list does not possess our target domain of social-engineer.org, and only the additional names found previously. Therefore, you may wish to remove these sources, and then add a new source, with the following commands.

```
options unset SOURCE
options set SOURCE cnn.com
```

Typing **run** and striking enter executes the process, while submitting **show contacts** afterward displays the results. The following is the partial output with new email addresses identified. Each of these addresses are now stored in your workspace, ready for the next round of research.

```
| 2   | barsuk   |   |          | barsuk@cnn.com
| 3   | Tristan  |   | Helmich  | tristan.helmich@cnn.com
| 4   | Paul     | P | Murphy   | paul.p.murphy@cnn.com
```

Let's reflect on how this can be beneficial. Assume that you are investigating numerous websites. Recon-ng provides dozens of utilities which automate queries and provides immediate information. Magnify this by tens or hundreds of domains, profiles, or other target data, and you have an easy way to replicate several hours of work. In another scenario, you are investigating a list of potential email addresses connected to a case. Entering these into Recon-ng allows you to execute your searches across all accounts. The effort to check one address is the same to check thousands. This impressive capability is only a small fraction of what can be done with this application.

This seems like a good time to back away, create a report, and start a new set of actions. The following commands will back out of our current module; install the reporting feature; instruct

Recon-ng that we want to use the reporting tool; mandate a graphical html (web) template be used; set the "Customer" as IntelTechniques; set the "Creator" as M.Bazzell; and execute the process.

```
back
marketplace install html
modules load html
options set CUSTOMER IntelTechniques
options set CREATOR M.Bazzell
run
```

Note the output after the final command. It identifies that the report is complete, and provides the storage location. Since I am running Recon-ng from my OSINT virtual machine, the default location is /home/osint/.recon-ng/workspaces/OSINT/results.html. Therefore, I can open the home folder on my desktop; double-click the ".recon-ng" folder; double-click the "workspaces" folder; double-click the "OSINT" folder; and then open the "results" file. Figure 26.01 displays the partial file from this example. Note that the Domains, Hosts, and Contacts sections are not expanded, but contain a lot of information. At the bottom of this file, the "Created by", date, and time clearly identify these report details.

Hopefully this demonstration explained the usage of Recon-ng. Executing **exit** in the window closes everything, but removes nothing. Before our next example, let's delete our previous work and start fresh. Note that deleting a workspace removes all associated data and reports. Make sure that you have exported your evidence if needed. First, relaunch Recon-ng. The following commands display the current workspaces; delete the OSINT workspace; and create a new workspace titled location.

```
workspaces list
workspaces delete OSINT
workspaces add location
```

This chapter explains only a small portion of the capabilities of Recon-ng. Please consider revisiting the modules listed at the beginning and experiment with the execution of each. Overall, it would be very difficult to break the application, and any errors received are harmless. You will receive best results by requesting free API keys from the services available within the modules marked with "*" in the "K" column. Many API keys are free and open new possibilities. Overall, an entire book could be written about this application alone. The goal of this section was simply to familiarize you with the program and demonstrate the power of automated queries.

If you would like more information about Recon-ng, please visit the official Github page at https://github.com/lanmaster53/recon-ng. From there, you can join a dedicated Slack group in order to participate in group discussions about errors, features, and overall usage.



IntelTechniques

www.recon-ng.com

Recon-ng Reconnaissance Report

[-] Summary

table	count
domains	12
companies	0
netblocks	0
locations	0
vulnerabilities	0
ports	0
hosts	39
contacts	0
credentials	0
leaks	0
pushpins	0
profiles	19
repositories	0

[+] Domains

[+] Hosts

[-] Profiles

username	resource	url	category	notes	module
humanhacker	Blogspot	http://humanhacker.blogspot.com	blog		profiler
humanhacker	Disqus	https://disqus.com/by/humanhacker/	discussion		profiler
humanhacker	Flipboard	https://flipboard.com/@humanhacker	tech		profiler
humanhacker	GitHub	https://api.github.com/users/humanhacker	coding		profiler
humanhacker	Instagram	https://www.instagram.com/humanhacker/	social		profiler
humanhacker	Kongregate	http://www.kongregate.com/accounts/humanhacker	gaming		profiler
humanhacker	Kik	https://kik.me/humanhacker	social		profiler
humanhacker	Medium	https://medium.com/@humanhacker/latest	news		profiler
humanhacker	Microsoft Technet Community	https://social.technet.microsoft.com/profile/humanhacker/	tech		profiler
humanhacker	Minecraft	https://namemc.com/name/humanhacker	gaming		profiler
humanhacker	Pornhub Users	https://www.pornhub.com/users/humanhacker	XXX PORN XXX		profiler
humanhacker	scratch	https://scratch.mit.edu/users/humanhacker/	coding		profiler
humanhacker	reddit	https://www.reddit.com/user/humanhacker	news		profiler
humanhacker	Twitch.tv	https://passport.twitch.tv/usernames/humanhacker	gaming		profiler
humanhacker	Twitter	https://twitter.com/humanhacker	social		profiler
humanhacker	Xbox Gamertag	https://www.xboxgamertag.com/search/humanhacker/	gaming		profiler
inteltechniques	Gravatar	http://en.gravatar.com/profiles/inteltechniques.json	images		profiler
inteltechniques	reddit	https://www.reddit.com/user/inteltechniques	news		profiler
inteltechniques	Twitter	https://twitter.com/inteltechniques	social		profiler

Created by: m. Bazzell
Sun, Sep 22 2019 10:44:40

Figure 26.01: A partial Recon-ng report.

Skiptracer

This program attempts to automate some of the OSINT resources mentioned throughout the book. It uses web scraping techniques to bypass the need for paid APIs. Installation is simple, and can be conducted with the following commands in Terminal.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/xillwillx/skiptracer.git skiptracer`
- `cd skiptracer`
- `pip install -r requirements.txt`

We can now launch the program with the command `python skiptracer.py` while still within the path of the skiptracer folder. That will present the following menu.

```
[1] Email - Search targets by email address
[2] Name - Search targets by First Last name combination
[3] Phone - Search targets by telephone number
[4] ScreenName - Search targets by known alias
[5] Plate - Search targets by license plate
[6] Domain - Search targets by Domain
[7] Help - Details the application and use cases
[88] Report - Generates a docx report from queries
[99] Exit - Terminate the application
```

Let's start with an email search by entering 1 and then striking enter. This presents a new menu as follows.

```
[1] All - Run all modules associated to the email module group
[2] LinkedIn - Check if user exposes information through LinkedIn
[3] HaveIBeenPwned - Check email against known compromised networks
[4] Myspace - Check if users account has a registered account
[5] AdvancedBackgroundChecks - Run email through public page of paid access
[6] Reset Target - Reset the Email to new target address
[7] Back - Return to main menu
```

At the time of this writing, only the AdvancedBackgroundChecks feature would function. I entered 5, followed by striking enter, and supplied my target email address. Minimal data was obtained. Overall, I find the email search to lack any useful features. However, the name and telephone searches are quite beneficial. Type 7 to leave this menu, and then 2 to launch the name search. Choose 1 to run your query through all services and enter your target name. You will be asked several questions in order to focus on the appropriate target.

The final result should display the data acquired from your target details. While searching my own name, it displayed several associates, telephone numbers, and email address associated with a relative having the same name. Be sure to copy and paste any desired data before hitting the enter key, as that takes you back to the previous menu. I have found the Report (88) option to be

unstable and have never received a valid report. Pressing enter, and then 6 to go back, takes you to the main menu. Choosing the Telephone Number option repeats the process of providing target data and executing a search across all options. I submitted an old cellular number and received numerous accurate results.

The only benefit to this application over a web search is speed. If you need to query multiple targets and collect only the associated text, this program works very well. It does not obtain content which you cannot find manually. In order to update the application, execute the following, which we will add to our script later.

- `cd ~/Downloads/Programs/skiptracer`
- `git pull https://github.com/xillwillx/skiptracer.git`

Please note that Skiptracer currently runs on Python 2.x, and there are plans to migrate it to Python 3.x. If this happens before you install, you may need to tweak the installation a bit. Details can always be found at https://github.com/xillwillx/skiptracer.

Sherlock

In the earlier tutorials for Recon-ng, I demonstrated a module which identified online profiles based on the username of the target. We can replicate this with Sherlock, and eliminate the need to launch Recon-ng for this task. First, install Sherlock with the following commands.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/sherlock-project/sherlock.git`
- `cd sherlock`
- `python3 -m pip install -r requirements.txt`

We must now construct a proper execution of the command. We must be within the program folder in Terminal (~/Downloads/Programs/sherlock) and launch Python 3 followed by the script. We must also specify the target username and any additional options. My command is as follows, with an explanation immediately after.

```
python3 sherlock.py inteltechniques --csv -o ~/Documents/Report.csv
```

python3 (Specifies Python version 3.x)
sherlock.py (The python script)
inteltechniques (the target username)
--csv (Creates a CSV spreadsheet file)
-o ~/Documents/Report.csv (Specifies save location and file name)

I find that Profiler with Recon-ng works better than Sherlock, but we should possess many options within our arsenal. Sherlock seems to display many false positives. The commands to update Sherlock are as follows.

- `cd ~/Downloads/Programs/sherlock`
- `git pull https://github.com/sherlock-project/sherlock.git`

Spiderfoot

This program introduces more valuable utilities than any other single resource within this book. This will take some effort to install and configure, but the benefits justify the work. First, open Terminal within your master VM and install Docker with the following commands.

- `sudo apt-get update`
- `sudo apt install docker.io`
- `sudo systemctl start docker`
- `sudo systemctl enable docker`

Next, we need to install Spiderfoot with the following instructions.

- Navigate to https://www.spiderfoot.net/download.
- Download the Linux package file directly to your desktop.
- Right-click on the file and choose "Extract here".
- Open Terminal and navigate to the Spiderfoot folder. (In my example, I downloaded Spiderfoot version 2.12, and my command was cd ~/Desktop/spiderfoot-2.12.0-src/spiderfoot-2.12/)
- In Terminal, enter `sudo docker build -t spiderfoot .`
- In Terminal, enter `sudo docker run -p 5009:5001 -d spiderfoot`

The final command executes the Spiderfoot program, but there is no icon to click. Instead, you must navigate directly to http://127.0.0.1:5009/ within Firefox. This opens the Spiderfoot menu, and you are ready to explore. You might want to bookmark this address for immediate access in the future. Note that you will need to launch `sudo docker run -p 5009:5001 -d spiderfoot` within Terminal before each use of Spiderfoot. Let's take a look at the interface and conduct an example query.

After navigating to http://127.0.0.1:5009/ within Firefox, click the "New Scan" option. Provide a name for your session (IntelTechniques) and a domain (inteltechniques.com). Choose your desired "Use case" and click "Run scan". I chose "All" in order to test the features, but this can be intrusive toward your target site. Choose the level of access appropriate for your investigation. The scan will launch and may take a while to complete, possibly hours. The amount of data acquired will be substantial, and I will only focus on a few areas of interest.

The default screen displays the current progress and a log file. The "Browse" button in the upper left allows you to start peering into the data found about your target. Below are the sections of interest to my own site and the results displayed.

Account on external site: Four online profiles connected to my brand
Human Name: Identification of my full name and three associates
Leak site content: Ten Pastebin files referencing my domain
Similar domain: Two domains with similar spelling, attempting to hijack traffic
URL (Accepts passwords): Pages on my site which possess a login form
Web technology: Pages on my site which rely on PHP, and the version running on my server

The "Graph" button displayed a detailed chart of connections from my domain to external sources. Figure 26.02 displays my result, identifying associations to a Reddit username, Gravatar profile, and email server.

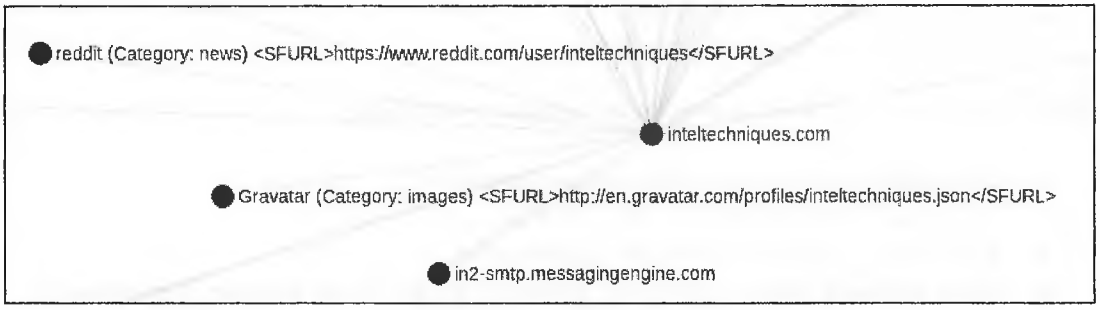


Figure 26.02: A Spiderfoot graph displaying external associations.

I cannot overstate that I am only presenting a handful of interesting nuggets. This application scours a domain, IP address, or email address for hundreds of data points which may provide value. Clicking the "Scans" button provides options to stop, re-run, or delete a scan result. It also provides a status summary of each current scan, and you can execute numerous scans simultaneously.

Update Script

If you added these four new Linux utilities, and Metagoofil as explained within a previous chapter, to your OSINT Master VM, you should consider adding the update steps to the "updates.sh" script within the file you downloaded during the previous section of this book. The following represents the current file in your "scripts" folder with amendments to include these new applications. This entire content is available within the linux.txt file which you previously downloaded.

```
#!/usr/bin/env bash
##Updates Script

sudo apt-get update
sudo apt-get upgrade
sudo -H pip install --upgrade youtube-dl
sudo pip install instalooter -U
sudo pip3 install Instaloader -U
sudo pip3 install Twint -U
cd ~/Downloads/Programs/EyeWitness
git pull https://github.com/ChrisTruncer/EyeWitness.git
cd ~/Downloads/Programs/Sublist3r
git pull https://github.com/aboul3la/Sublist3r.git
cd ~/Downloads/Programs/Photon
git pull https://github.com/s0md3v/Photon.git
sudo snap refresh
cd ~/Downloads/Programs/metagoofil
git pull https://github.com/opsdisk/metagoofil.git
cd ~/Downloads/Programs/recon-ng
git pull https://github.com/lanmaster53/recon-ng.git
cd ~/Downloads/Programs/skiptracer
git pull https://github.com/xillwillx/skiptracer.git
cd ~/Downloads/Programs/sherlock
git pull https://github.com/sherlock-project/sherlock.git
```

It is now your responsibility to keep your Linux applications updated. Additionally, you should continue to seek new options. Practically anything you find can be installed, configured, and launched using the methods discussed throughout this book.

Chapter Twenty-Seven
Data Breaches & Leaks

The techniques that you will read about in this chapter are for educational use only. Many, if not all, of the methods described here could violate organization policy if executed. While the chapter will only discuss publicly available data, possession could violate your security clearance or be determined illegal by state or federal law. Distribution of the types of content that will be discussed here will be illegal in most cases. However, I will explain ways that you can apply these practices in a legal way, and keep yourself employed. Overall, please do not take any action from this instruction unless you have verified with your organization's legal counsel or supervisory personnel that you have the authority to do so. Let's begin.

In previous chapters, I discussed the online services Have I Been Pwned and Dehashed as two amazing resources for email search. These services possess a huge database of publicly available leaks that were stolen from the host companies and distributed over the internet. In order to emphasize the concern, I will repeat that this data was originally stolen by criminals. When you search an email address on these services and are informed that it was compromised within the LinkedIn data breach, this means that a partial copy of this stolen data resides within these services. HIBP and others are often applauded as a great resource to monitor your own accounts for any reported compromises. Well, what if we created our own collection of this data?

This is where things get tricky. While you can find copies of thousands of stolen databases all over the internet, what are the legalities of downloading and possessing the data? First, let me say that I am not an attorney and I offer no legal advice. I have spoken with many attorneys and prosecutors about this, and the feedback was similar from each. If the data is publicly available, possession alone would likely be legal. This is similar to viewing an email stolen from Hillary Clinton posted on WikiLeaks or an internal document stolen from Google posted on Medium. If you do not violate the laws and policies applicable in your city, county, state, or country when you view this publicly available, yet stolen, data, then you could likely get away with viewing stolen credentials existing in the various database leaks online.

What matters most is that you never take any illegal action with the data that you possess. In a moment, I will explain how to access valid email addresses and passwords of billions of accounts. Using this data as a search technique is one extreme, but attempting to use this data to access someone's account is another. There is no situation where gaining access to a target's account is acceptable, unless you have a valid search warrant or court order to do so. Since many of you are in law enforcement, this chapter may identify new strategies for you when you have the legal authority to access an account. We will start with some very basic data which is legal to possess.

I previously presented websites that allowed you to search a real name and identify the voter registration record of the individual. This usually identifies a home address, telephone number, party affiliation, and occasionally an email address. These websites are committing no crime. Some states make the data freely available, and some charge a fee for digital access. All states' voter registration details are public record, which is why it is overly analyzed and scrutinized during election seasons. At the time of this writing, entire state databases were available for download at the following addresses.

Arkansas:	http://arkvoters.com/download.html
Colorado:	http://coloradovoters.info/downloads/20170401/
Connecticut:	http://connvoters.com/download.html
Delaware:	http://delawarevoters.info/downloads.html
Florida:	http://flvoters.com/download/20171130/
Michigan	https://michiganvoters.info/download.html
Oklahoma:	http://oklavoters.com/download.html
Rhode Island:	http://rivoters.com/downloads.html
Utah:	http://utvoters.com/

These sites are all operated by the same individual. He files Freedom of Information Act (FOIA) requests for copies of the public data, and discloses the content he is given by the government. If you are thinking that this cannot be legal, consider the situation with Ohio. If you navigate to https://www6.sos.state.oh.us/ords/f?p=111:1 and click on the "Statewide Voter File" tab, you can download the entire database of registered voters directly from the official Ohio government domain. This is truly public data. The data set for Washington state is also available from the government if you promise not to further distribute or sell the content. You can download the entire database directly at https://www.sos.wa.gov/elections/vrdb/extract-requests.aspx. Other states' voter databases are also out there, and Google should get you to them. You may not see the value of downloading this data versus conducting queries on the various websites. Let's walk through an example to see where we can benefit from the full data set.

Assume that you have downloaded the dozens of county files from the Florida link above. You now have a folder titled Voter-FL on your desktop of your Linux VM that was discussed in Section I. The reason we want the data inside Linux is because we can take advantage of built-in commands that will help us sort and collect our data. Furthermore, we will add Ripgrep to Linux, which allows extremely fast searching of large data sets. Personally, I do not like having a separate text file for each county in Florida. I would rather combine all of them into one single file and title it appropriately. If you start ~~hoarding~~ collecting data, you may want to keep your collection tidy. Therefore, I would open Terminal and type the following commands. The first navigates you into the folder, and the second combines all of the files into one file titled Voter-FL.txt.

```
cd Desktop/Voter-FL
cat * > Voter-FL.txt
```

You could now move that file to another folder on your desktop titled Databases, and delete the Voter-FL folder. This new large text file may take a long time to open, which is quite common with these datasets. Therefore, let's clean it up. First, we need to install a small utility titled Ripgrep. The following instructions explain the process for Mac and Ubuntu Linux. Ripgrep does not play well with Windows, hence the need for a Linux VM.

Ubuntu: Conduct the following steps within your OSINT Master VM.

- Navigate to https://github.com/BurntSushi/ripgrep/releases/.
- In the "Assets" section, download the file ending with ".deb".
- Open the Files application, click on Downloads, and double-click this .deb file, which was ripgrep_11.0.2_amd64.deb at the time of writing.
- Click "Install" and follow the prompts.

Mac: Enter the following command within Terminal. You must have Brew installed as explained in the beginning of the book.

- `brew install ripgrep.`

Now that we have Ripgrep installed, we can use it to peek into very large files. First, let's take a look at the structure. The following commands within Terminal in Linux or Mac will navigate you to the Databases folder on your desktop, and conduct a search for the name "Williamson" within the new file. The search parameters will be explained in a moment.

```
cd ~/Desktop/Databases
rg -a -F -i -N Williamson
```

The output appears sporadic and similar to the text below. This is because there are unnecessary "tabs" in each line which causes the text to appear broken.

```
ALA     100397608       Williamson              Glenda Dianne N       217218  NW 48Th
TER             Gainesville             32606                         08/17/1951
        08/14/1980      DEM     22      0       22C3            ACT     3       21
```

The following optional command will replace each tab with a colon, which compresses the data and makes it easier to read. The same search from the previous example now produces the text visible below the command. This is much cleaner and will make future search results easier to digest. Note that the spaces in the command are created by pressing the control, v, and tab keys at the same time. This represents a "tab" to the command. When sed is executed, you will not see any output or indication it is working. However, you will be prompted when complete.

```
sed -i 's/    /:/g' Voter-FL.txt
```

```
ALA:100397608:Williamson::Glenda:Dianne:N:217218  NW 48Th TER : :Gainesville:
:32606:::::::F:5:08/17/1951:08/14/1980:DEM:22:0:22C3::ACT:3:21:8:0:0::::
```

The presence of the unnecessary colons bothers me, so I will execute the following command to eliminate any two consecutive colons, and replace them with a single colon.

```
sed -i 's/::/:/g' Voter-FL.txt
```

Repeating the original search for Williamson now reveals the following output.

```
ALA:100397608:Williamson:Glenda:Dianne:N:217218  NW 48Th TER :Gainesville:
:32606:F:5:08/17/1951:08/14/1980:DEM:22:0:22C3:ACT:3:21:8:0:0:
```

If you have your Terminal window maximized, this all fits quite nicely on one line. I may have created more confusion than provided answers, so let's dissect each process. This will be beneficial in our future examples. In order to search the data collected, we will be using Ripgrep, as previously mentioned. The commands we will be using are explained below.

rg	The command for Ripgrep
-a	The switch to search all data as text
-F	The switch to treat the pattern as a literal string
-i	The switch to ignore case
-N	The switch to exclude the line number

rg -a -F -i -N Test@Gmail.com	Search EXACTLY test@gmail.com or TEST@GMAIL.com
rg -a -i -N Test@Gmail.com	Search ALL test and gmail and com
rg -a -F -N Test@Gmail.com	Search ONLY Test@Gmail.com and not test@gmail.com
rg -a -F -i Test@Gmail.com	Search EXACTLY test@gmail.com and show line #
rg --help	Show Ripgrep help menu

Overall, we will almost always use the combination of rg -a -F -i -N 'target'. The commands that begin with "sed" tell the operating system to modify data. In the previous command of `sed -i 's/::/:/g' Voter-FL.txt`, the following should help explain.

sed	The command for "Sed"
-i	Modify the file and save it in real-time
's/::/:/g'	Replace every occurrence of :: with : throughout the entire file
Voter-FL.txt	The file to modify (* would modify all files in the folder)

If you feel overwhelmed, please don't let that convince you to abandon this chapter. I promise that the commands will start to make more sense as we progress. Let's jump back into the search results we received. They clearly identify the full name, home address, gender, date of birth, date of last registration change, and party affiliation. I find this extremely useful, but all of this could have been obtained from the previously mentioned websites. Therefore, let's conduct a search

for any registered voters that live on the street of our previous example. The following would be the query, providing our search in single quotes because it contains a space. Instead of specifying the file, we can leave that off and it will simply search every file in the folder.

```
rg -a -F -i -N 'NW 48Th TER'
```

The response contains the same type of details as seen previously, but displays hundreds of people. You cannot replicate these results with a traditional manual search on the sites themselves. Furthermore, we could conduct these searches across all of our databases, which we will do later. You could replicate all of this instruction throughout the various voter registration databases available for download. You might eventually have a single file for each state, with all of the unnecessary content removed from within the files. Some readers may choose not to modify the files while others may sanitize them to the point of ultimate conservation of file size. Each user's collection will be unique. You might try the following search options.

rg -a -F -i -N @gmail.com	Display all Gmail accounts within the voter files
rg -a -F -i -N rita@gmail.com	Search a specific email address within all of the files

Hopefully, you agree that possessing your own collection of this type of data could be useful. Let's add another collection of publicly available data. An individual previously purchased the entire Social Security Death Index, which is public data, and uploaded it for public consumption. This entire database can be downloaded from http://cancelthesefunerals.com. After downloading the multiple files, you can combine them as stated previously into a single file and name it SSDI.txt. Assuming you placed the text file in the Databases folder, the following commands from the Terminal would navigate to the folder and conduct a search for anyone in the file with my last name.

```
cd ~/Desktop/Databases
rg -a -F -i -N bazzell SSDI.txt
```

The results will include multiple entries with the following format.

```
433220353 BAZZELL                DOROTHY        S              0118201303191921
```

The first set of numbers is the Social Security Number of the deceased individual, followed by the last name, first name, and middle initial. The last set of numbers represents the date of death (01/18/2013) and the date of birth (03/19/1921). You could choose to replicate the "tab" removal process described earlier, or just leave this file as it was created. Let's recap where we are at right now. You might have single, very large, text files that each include all of the registered voter data for various states. These often include dates of birth, names, telephone numbers, addresses, and email accounts. You also have the entire SSDI. You can now search through all of that data with a simple search. If you wanted to identify any entries within all of the files, with a telephone number of 618-555-1212, the following would be the most appropriate searches. You could also now search by social security number, date of birth, or date of death.

```
rg -a -F -i -N 6185551212
rg -a -F -i -N 618-555-1212
```

Now that you understand the basics of searching and modifying data, we should discuss the true value of this technique. Until this point, we have only experimented with truly public data that is legal to possess. The real power lies within the stolen databases that are now publicly available which I am about to discuss. There are usually two types of people that download these databases. The first are amateur criminals that use the data to illegally obtain account access to victims' accounts by supplying the usernames and passwords from one service into another service. As an example, if a criminal learns your username and password that you used on LinkedIn, he or she may try that same combination on Tumblr, Twitter, or Gmail. We will never do this, as it is illegal and unethical.

The second group of individuals that download this data are security researchers. Numerous private organizations authorize employees to collect this stolen information and use the data to make their own systems more secure. I personally know of many researchers who download this data, protect it from further distribution, and only use it in an ethical manner. This is the only acceptable use of this data. For those that still believe that we should not access stolen databases released to the public, consider one last argument. Tens of thousands of criminal hackers use this content to "dox" people every day. Doxing is the act of exposing personal information about a victim online. This is a very common tactic used against law enforcement officers when an event such as an officer-involved shooting happens. The government employees' passwords often get leaked on Pastebin or other similar sites. This stolen data is already out there, and it can never be secured. We should embrace the same techniques used against us, when researching criminals. OK, enough warnings. You can use your best judgment for the rest of this chapter.

By now you may just want to know where to obtain these databases. The source websites that allow download of this stolen data get shutdown often. However, the magic of internet archives can help us regain the data. Let's start with something in a grey area. There are many companies that collect millions of email addresses and create spam lists. These then get sold to companies selling fake pharmaceuticals or designer handbags, and those unwanted messages end up in your inbox. We are all on a spam list somewhere. These are very expensive to buy, especially the good ones. One such database is the Kyle Data, or "Special K" spam list. We used this example in the browser chapter when DownThemAll easily acquired all of the data. This database contains 178,288,657 email addresses. There are no passwords, but it does contain other sensitive data. In November of 2015, a web page at updates4news.com temporarily possessed active links to this entire data set. This page is long gone by now. Fortunately for us, the Wayback Machine archived all of it. The following direct link displays hundreds of csv spreadsheets, each containing millions of email addresses.

https://web.archive.org/web/20151110195654/http://www.updates4news.com:80/kyledata/

If desired, you could use the Firefox extension mentioned previously to automatically download all of this data overnight into a folder on your Linux desktop titled "SpecialK". You could then execute a command through Terminal within that folder of `cat * > SpecialK.txt`. The result would be a single file, 19.73 gigabytes in size. That is very big, but also very powerful. Let's take a look at the content. Assume I was looking for a target who possessed an email address of robyn2977@yahoo.com. Assuming that I had made a single file titled SpecialK.txt within my Databases folder on my Linux desktop, my commands in Terminal would be as follows. The result of the search is directly after the search command.

```
cd ~/Desktop/Databases
rg -a -F -i -N robyn2977@yahoo.com SpecialK.txt
```

```
MANHATTAN BEACH, robynsnest2006@yahoo.com,yahoo.com, Robyn Bazzell, 2015-04-07,
72.129.87.179, paydayloaneveryday.com, CA, 90266
```

This tells me that my target lives in Manhattan Beach, CA 90266. Her name is Robyn Bazzell, and on 04/07/2015 her marketing data was collected from the website paydayloaneveryday.com. At the time, her computer IP address was 72.129.87.179. We basically just converted an email address into a full dossier on our target. This type of data is simply not available on public websites. If you doubt the validity of this data, please consider searching your own content. This is only one of dozens of interesting databases stored within the Wayback Machine. It is up to you to exercise your skills from previous chapters to locate them all.

The next exercise also involves archive.org. Specifically, data from LinkedIn. This data set surfaced in 2017 and does not possess any passwords. It contains only LinkedIn user ID numbers and the associated email addresses. While this type of data was included within the original 20 GB LinkedIn breach, it was excluded from the popular password dumps that can be found online today. Since it does not contain passwords, it can be found on the Wayback Machine. At the time of this writing, the direct link to download the entire user ID database could be found at https://archive.org/details/LIUsers.7z.

Decompress this file by right-clicking within Linux and choosing "Extract here". The text file contains the email addresses and user ID numbers extracted from the full LinkedIn breach. It is nothing more than each email address used to create a specific profile ID number. The following search within Terminal would display the results of a target email address, and the response is directly below the command.

```
rg -a -F -i -N janewilson@microsoft.com linkedin_users.txt
```

```
1332567, janewilson@microsoft.com
```

You now know that your target's user ID number is 1332567. This data alone is not very helpful. Let's consider another perspective. Assume you find your target's LinkedIn profile page, and you want to know the email address used to create the account. Right-clicking on the profile allows

the option to display the "Source Code" of the profile. Searching the term "member:" within this code presents numerous occurrences of that term. The third or fourth from last occurrence should appear similar to "member:1288635". This tells you that the member ID for your target is 1288635. The following search displays the email address associated with that number, with the results below.

```
rg -a -F -i -N 1288635 linkedin_users.txt

1288635, John.h.007@gmail.com
```

Real World Application: I have used this technique numerous times over the past couple of years. During one investigation, I located the LinkedIn profile of my target. I had previously found no email addresses connected to him. By looking at the source code of the LinkedIn profile, I could see his user ID. Searching that user ID within this LinkedIn data provided the personal email address used to create the account. That email address led to old passwords, which led to additional email addresses, as explained in a moment. Obtaining a confirmed personal email address can lead to numerous new investigation techniques.

Let's conduct one more archive.org demonstration before moving on to credentials. In January 2014, Snapchat had 4.6 million usernames and phone numbers exposed. The breach enabled individual usernames, which are often used across other services, to be resolved to phone numbers. This file is titled SnapChat.7z and can be found at archive.org/download/SnapChat.7z. This link contains the entire breach inside of one text file. Conducting the following search would query a specific Snapchat username for any leaked data. Directly after this command is the result.

```
rg -a -F -i -N mikenap115 SnapChat.txt

('21220392XX', mikenap115, '', '')
```

This identifies the cellular telephone number of our target to include 212-203-92xx. While the last two numbers are redacted, you could try all 100 combinations within the previously discussed telephone search options. It is a great start. You could have also conducted this query on websites which possess this data. However, those sites require an absolute username. It cannot search partial names. If we only knew that our target had a Snapchat name that included the term "topher4", this database would provide the following users of interest.

```
('30351923XX', 'topher451', '', ''),
('41572499XX', 'topher413', '', ''),
('71974140XX', 'topher456', '', ''),
('75426428XX', 'topher481', '', ''),
```

We can also now search by telephone numbers. If you know your target has a cellular number of 303-519-2388, you could conduct the following search, with the partial result appearing after.

```
rg -a -F -i -N 30351923XX SnapChat.txt
```

```
('30351923XX', 'topher451', '', ''),
('30351923XX', 'ben_davis', '', ''),
('30351923XX', 'rosemcdonald', '', ''),
('30351923XX', 'cuzzycook', '', ''),
('30351923XX', 'angelgallozzi', '', ''),
('30351923XX', 'kmo85', '', ''),
('30351923XX', 'rinisob', '', ''),
```

Your target may not appear in the results, but minimal investigation could result in a powerful new lead. You should note that the full numbers were never publicly leaked, and the group that conducted this attack redacted the results to exclude the last digits.

Real World Application: In 2017, I was provided a Snapchat username, similar to "yourdaddy3", of a suspect in a human trafficking investigation. A quick search revealed the first eight digits of his cellular telephone number which did not prove to be beneficial to various online searches. However, conducting a search for "yourdaddy" without the digit revealed a new partial number associated with "yourdaddy1" and "yourdaddy2", all within the same area code. Using the reverse caller ID APIs mentioned previously, I attempted to identify the owners of all 100 possible numbers. 58 of them were not registered, which left 42 valid numbers returning to individuals living in the area. Of those, only 19 of them were associated with males. Of those, only three fit a physical description of the suspect. A cooperating witness easily identified the suspect from these options. I truly believe in the power of this data.

We have only discussed a few databases stored within archive.org. I promise you there are many others containing sensitive details which would help greatly within your investigations. Conducting searches for any specific breached databases displayed on notification websites such as https://haveibeenpwned.com/PwnedWebsites should reveal interesting results. You have learned about numerous search methods throughout this book. What can you find?

Now that we are dipping our toes into the waters of "stolen" data, we should consider the big breaches. You have likely heard about LinkedIn, Dropbox, Myspace and Adobe being hacked. All of those breaches have been publicly released by various criminal organizations. At the time of this writing, most remaining sites that possessed this huge collection of stolen data had been shut down. However, the data lives on. Numerous "hacking" groups have collected tens of thousands of breached databases, both large and small, and combined them into credential lists. These lists contain only the email addresses and passwords of billions of accounts. They do not identify which breach each credential originated, but the data is extremely valuable for investigators. Before we analyze the content, we must obtain a full copy of the data. This is where things get tricky (again).

My attorney says I cannot HOST links to any of this content, which I understand. She also says that I cannot display any direct links to an identifiable breach, such as LinkedIn or Adobe. This would make me a target for a lawsuit, and seems like great advice. My only option is to "identify public resources which link to data without attribution of a specific originating source". In other

words, I can tell you where you may find this "Combo List" content, but it is up to you to take action to obtain it. Let's head to Reddit while connected to our VPN. The following URL presents a discussion about a huge collection of credentials posted online.

reddit.com/r/hacking/comments/akb6mf/any_torrent_for_the_recent_773_million_data/

Within this post is a link to the file host sendspace.com. The link downloads a .rar file. Right-click this file within Linux and choose "Extract here". This should present two torrent files, titled as follows.

Collection 1.torrent
Collection #2-#5 & Antipublic.torrent

If this link is dead, similar links can be found within the following discussion boards.

news.ycombinator.com/item?id=18925818
reddit.com/r/pwned/comments/agsjie/troy_hunt_the_773_million_record_collection_1/

The titles of Collection 1-5 refer to combo list collections created in order to combine all known stolen credentials into one data set. Collection 1 contains 773 million user credentials obtained from all known public data breaches from 2009 through 2018. Collections 2-5 contain newer credentials. For the purposes of demonstration, let's focus only on Collection 1.

Double-clicking the "Collection 1.torrent" file should launch Transmission, the default torrent manager within Ubuntu Linux. This collection contains over 40 gigabytes of data (even larger uncompressed), so you likely have a storage issue. The moment that the torrent file loads within Transmission, you should be prompted to "Open" the torrent. Before doing so, deselect "Start when added" in the lower left. This prevents your VM from filling to capacity by adding these files. Figure 27.01 (left) displays this screen. By unchecking "Start when added", the torrent should load in the paused state. Click "Open" and allow the torrent to load within Transmission. Double-click the file within Transmission to see the details menu. Click on the "Files" menu (fourth tab from left) in order to see the content of this torrent. Figure 27.01 (right) displays my result. I have deselected everything except "Collection # 1_BTC", "Collection # 1_Dumps", and "Collection # 1_EU Combo". This will download less than 4 gigabytes of data to my VM when I click the Start Torrent (green arrow) within Transmission.

If all of the links posted to Reddit and YCombinator are no longer available, consider a search of "Collection #1-5 & Zabagur & AntiPublic" on Google. At the time of this writing, there were over twenty additional copies of these torrent files. I want to make it clear that I do not host any of these files. They are all readily available to the public, linked from Reddit, Google, and many other places. If possessing user credentials violates your security clearance or organizational polices, do not proceed with this download.

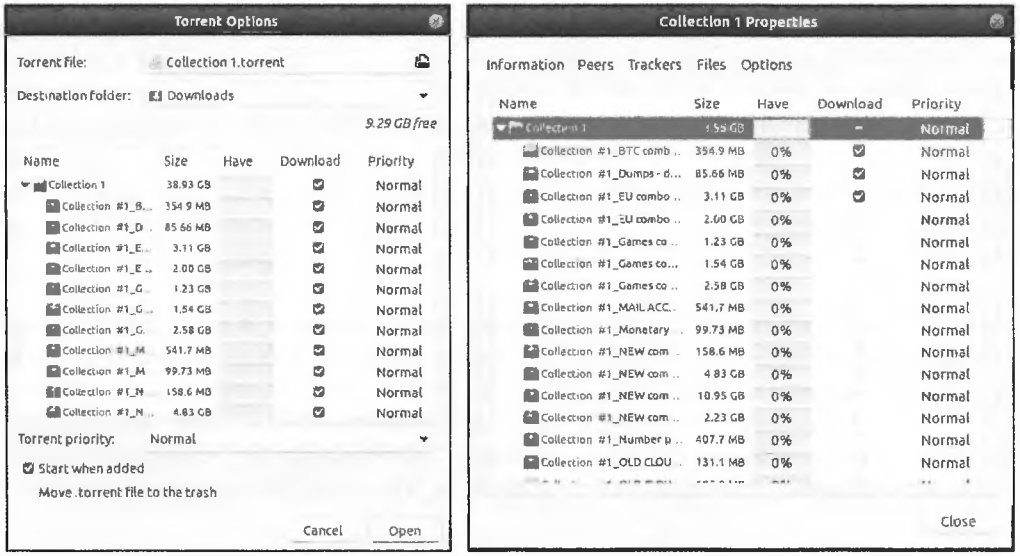


Figure 27.01: Menus from the torrent downloading manager Transmission.

As you can see within the previous screen captures, this is a lot of data. It may be more content than your internal hard drive can store, and is definitely more than your VM is configured to handle. If you expand each of these collections within Transmission before downloading, you can see that individual files can be selected. If you plan to download this data within your VM, you will need to select a few files, download them, pause the torrent, move the files to an external source, and then repeat. This is a huge annoyance. If appropriate, you may consider downloading the entire torrents to an external hard drive. During testing of this tutorial, I connected a 1TB external USB drive to my VM and chose it as the destination for the downloaded files. This kept my VM clean and all data was available on the external drive. This allows use within multiple VMs or host computers. Depending on your internet connection, this entire download can take hours or weeks to finish. You may want to download only a few files in order to continue with the demonstrations.

Let's assume you now have a hard drive with the entire contents from Collection 1. Each of these files are compressed in order to save space. We must first decompress every file which you downloaded. The easiest way is to select all of the files (ctrl-a), right-click, and choose "Extract here". This will result in numerous folders, each containing multiple text files. Once complete, you could delete all of the compressed files if desired. You now possess hundreds or thousands of text files within numerous folders. Some text files contain informative names describing the exact website as the source, while others contain practically no details. Hopefully, a few demonstrations of how this data could be used within investigations will warrant consideration by your organization to adopt policies allowing its use.

Assume you possess a folder called Collection1 within your external hard drive connected to your Linux VM. Opening Terminal and navigating to that folder may not be easy. You would need to know the exact path. Instead, open the Files application and you should see your external hard drive in the left menu. Click on it and select the Collection1 folder. Open Terminal, type cd, then space, then drag the Collection1 folder in the external storage from the Files application to Terminal. Strike enter and it should populate the path of the drive. Mine appeared as `cd '/media/osint/1TBExternal/Collection1'`. You could also find the external drive within the Files application in Ubuntu, right-click, and choose "Open in Terminal".

You should now be within your new data collection inside Terminal. We can conduct searches to start evaluating the benefit of this data. The following queries will all assume that you have already opened Terminal and have navigated to the folder where your data is stored. The next search would attempt to identify a specific email address within all of the files. Note that results were modified to protect the privacy of the individual. As a reminder, this command applies our parameters and identifies the target data to be searched. The result follows the command.

```
rg -a -F -i -N mikewilson@microsoft.com
```

```
Dumps-dehashed/Poloniex.txt
mikewilson@microsoft.com:bigbucks55
```

We now know that at some point a password of bigbucks55 was used in conjunction with an online account associated with our target email address. The data was obtained from the text file titled "Poloniex.txt" within a folder titled "Dumps-dehashed". A search of "Poloniex" reveals it is a crypto-currency service. A search of "Poloniex breach" reveals details about the data breach which likely created this result. "Dehashed" indicates that someone cracked the passwords, and that the included results are in plain text.

Would this password still work with any online accounts? Possibly, but we will never know. Attempting to log in to an account with this information is a crime. Instead, think about other ways that we could use this data offline. We know our target's work email account, but we may want his personal account. Since many people recycle the same password across multiple accounts, let's conduct a search for the password.

```
rg -a -F -i -N bigbucks55
```

The results include accounts associated with our target, and some that are not. The more unique your target's password, the better quality of associated content you will receive. The following data was presented during my search. Notice the ways that data was displayed based on the search, and how any portion of the results were included. Since we specified our search parameters, we received results regardless of case.

```
bigbucks551@yahoo.com:towboat@1
bigbucks55@hotmail.co.uk:towboat@1
```

```
bigbucks55@hotmail.com:towboat@1
prizeunit@yahoo.com:bigbucks55
mike.wilson5@gmail.com:BigBucks55
mikewilson@microsoft.com:bigbucks55
```

This tells me that the last two results are very likely my target, since the names are similar and the passwords are almost identical. I can now assume that my target's personal email account is mike.wilson5@gmail.com. The first three accounts could be my target, but are probably not. This is likely just a coincidence. However, we can assume that the owner of one of those accounts is the owner of all three since the passwords are identical. The fourth response is a toss-up, but worth further investigation.

Real World Application: The tactic of searching leaked passwords and recovering associated accounts is by far the most successful database leaks strategy that I have applied to my investigations. In 2017, I was assisting a federal agency with a child pornography investigation where a suspect email address had been identified. This address was confirmed as being connected to the original person that had manufactured illegal videos of children being molested, but an alias name was used during creation. A search of this address through the Breach Compilation revealed a unique password associated with the account. Searching this password revealed only one additional email account, which was the personal Gmail account of the suspect (in his real name). The suspect used the same password for both accounts. The primary investigator made an arrest the next day after confirming the connection. While some will say that we should never download leaked databases that contain personal login credentials, I disagree. There is great potential value in these data dumps that could solve cases on a grand scale, and make a huge impact on the prosecution of serious criminals.

I cannot overstate that this instruction is the very tip of the iceberg. There are thousands of compromised databases online, and more being published every day. If you find "Collection 1" to be valuable, you may consider downloading "2-5" within the same torrent files mentioned previously. The entire collection is over 1 terabyte in size, so be prepared. If you invest some time into seeking the sources of this data, you will quickly become overwhelmed at the mass amounts of content to properly obtain, clean, and store. I can only discuss the basics here. It is your job to proceed if you choose.

As a start, you may consider focusing on public credential leaks on Pastebin (pastebin.com). When you search an email address on psbdmp.com, the results often include paste dumps. Clicking on these links will take you to the original paste document, which likely has many additional compromised credentials. A query of test@test.com on this site will likely present hundreds of data sets ready for download. Do you remember the "Index Of" technique mentioned previously? This is very useful in searching for leaks. Any time I see on HIBP that a new public leak has surfaced, I search for that specific data with a custom Google search. If I saw that Myspace has been exposed, I would use the following.

"myspace" ext:rar OR ext:zip OR ext:7z OR ext:txt OR ext:sql

There are many online database resources that will sell you the data. Please avoid these. First, you will likely get ripped off a few times before you find an "honest" seller. Second, you are giving money to criminals, and I don't like encouraging that behavior. Many researchers that I know possess over 10,000 databases which contain over two terabytes of total information, all found on public sites. A search of your collected data can take a few minutes. Searching "TiVETVovV", "NdUrkHvUr" or "vvtMffTCa" on Google may lead you to gold.

Hashes

Most websites store passwords in "hashed" form. This guards against the possibility that someone who gains unauthorized access to the database can retrieve the plain-text passwords of every user in the system. Hashing performs a one-way transformation on a password, turning the password into another string, called the hashed password. "One-way" means that it was practically impossible to go the other way and turn the hashed password back into the original password. This was true many years ago, but not so much today. There are several mathematically complex hashing algorithms that fulfill these needs. Some are very insecure and others are nearly impossible to crack. Let's look at a few examples.

MD5: Many older databases which have been breached possessed simple MD5 password hashes. The password of "password1234" is as follows as an MD5 hash.

BDC87B9C894DA5168059E00EBFFB9077

SHA1: The Sha-1 hash of the same password, "password1234", is as follows. Notice this is substantially longer, and a bit more secure. However, it will be quite easy for us to crack these passwords in just a moment. Below this example is the same password in Sha-256 and Sha-512 format. As you can see, these become increasingly complicated.

E6B6AFBD6D76BB5D2041542D7D2E3FAC5BB05593

SHA256: B9C950640E1B3740E98ACB93E669C65766F6670DD1609BA91FF41052BA48C6F3

SHA512: 8C7C9D16278AC60A19776F204F3109B1C2FC782FF8B671F42426A85CF72B102
1887DD9E4FEBE420DCD215BA499FF12E230DAF67AFFFDE8BF84BEFE867A8822C4

In regard to the various breached databases which you are likely to find on the internet, you will most commonly see MD5 and SHA1 hashed passwords. Some of them possess a "salt". This is a small amount of data added to the hashing which makes cracking the password more difficult. If a breach does not possess the salt, the passwords are nearly impossible to crack. If the salt is present, it takes considerable additional resources in order to display the text password. The methods of "cracking" passwords exceed the scope of this book. Fortunately, we do not need the knowledge and computer horsepower to convert these hashes into valuable passwords. We simply need a third-party resource.

Hashes (hashes.org)

Per their website, Hashes.org attempts to reveal the plain text of your submitted password hash. This is usually done in an effort to assist security professionals to evaluate the security provided by the relevant hash submitted. For us, it provides a new lead to follow. The database currently contains 3,316,303,754 cracked hashes available via web search and API. Typically, password cracking volunteers spend their time reversing hashes in small individual batches spread across multiple communities, frequently duplicating each other's work. Alternatively, Hashes.org provides a single point of reference for both hashes and their plain-text passwords. This large community possesses the most robust database of "dehashed" passwords that I have found. Let's conduct a few real examples in order to understand this better. Within the "Collection 1-5" files, several of the breaches have not been "dehashed", or converted into plain text. This means that no passwords are visible within the files. The following examples are all slightly modified in order to protect any personal data, but the methods and results are accurate.

MD5

Some older data breaches possess passwords hashed with MD5, an extremely insecure method. These hashes can be cracked very quickly. Below are two entries retrieved from various files.

174@gmail.com:482C811DA5D5B4BC6D497FFA98491E38
185@gmail.com:22F4182AAE2784FB3D1A432D44F07F46

Everything before the colon is the username or email address, and everything after is the MD5 hash. Copying these hashes into hashes.org/search.php produces the following results. We now know that the first password is "password123" and the second is "reader12".

482c811da5d5b4bc6d497ffa98491e38:password123
22f4182aae2784fb3d1a432d44f07f46:reader12

MD5 + Salt (Vbulletin)

Some hashes will contain a "salt". This is usually a small piece of data after the hash, and both pieces are required to translate into plain-text. One of the most popular examples of this are the hundreds of Vbulletin online forums which have been infiltrated by various hacking groups. In the following examples, the salt is the final three characters after the last colon.

Aanas,menball@aol.com:9d9e3c372d054c0769bd93181240be36:tye
Traor,tranr@optusnet.com.au:9274583d060b3efb464115e65a8c1ead:vv#

Searching these hashes and salts on hashes.org/search.php provides the following results.

9d9e3c372d054c0769bd93181240be36:tye:eliza!%
9274583d060b3efb464115e65a8c1ead:vv#:runner

In the first example, you see the target password is "eliza!%". This proves that semi-complex passwords containing special characters are often dehashed and ready for conversion. This is where Hashes sticks out from the crowd. There are many online reverse-hashing tools, but most of them focus only on one format, such as MD5, and use minimal dictionaries to crack the hashes. Let's try one more demonstration.

SHA1

This format is slightly more secure than MD5, but still extremely easy to crack. The passwords obtained from the LinkedIn breach were all hashed in SHA1. Below are two examples.

174@gmail.com:403E35A2B0243D40400AF6BB358B5C546CDDD981
185@gmail.com:B1C4BBC4D7546529895CFABF8C1139CA7E486E18

The results from Hashes follows. These two passwords are identical, but with different case. The first is all lowercase while the second contains uppercase. Notice that these create a completely different hash because of this. Keep this in mind when investigating passwords of your target.

403E35A2B0243D40400AF6BB358B5C546CDDD981:letmein!
B1C4BBC4D7546529895CFABF8C1139CA7E486E18:LetMeIn!

Searching these through hashes.org is not difficult, but can be time consuming. Each search requires a Captcha be completed and there is no direct URL submission to their site. Therefore, I encourage you to register for a free API key at https://hashes.org/api.php. I have already added a valid key to the search tools under an alias account, but this is likely to be terminated due to abuse once the tools become public. If it works, you are all set. If you receive errors, obtain your own free API key and replace mine in the tools. First, let's take a look at the URL structure. The following address contains my key provided within the tools and the first SHA1 hash mentioned previously. The result immediately follows.

https://hashes.org/api.php?key=wQbxQlPJ0LZtLUVcBJKmDc8hnY5lcc&query=B1C4BBC4D7546529895CFABF8C1139CA7E486E18

{"status":"success","result":{"B1C4BBC4D7546529895CFABF8C1139CA7E486E18":{"plain":"**LetMeIn!**","hexplain":"4c65744d65496e21","algorithm":"**SHA1**"}}}

This method provides us three benefits. First, it eliminates the Captcha and need to submit through a website. Second, the result is immediate and displayed in plain text. Finally, it discloses the type of hash submitted (SHA1). This is the technique I rely on daily for submission of hashes. Now, let's take a look at the code within your search tool.

```
<script type="text/javascript">
function doSearch21(Search21)
{window.open('https://hashes.org/api.php?key=wQbxQlPJ0LZtLUVcBJKmDc8hnY5lcc&query=
' + Search21, 'Search21window');}
</script>
<form onsubmit="doSearch21(this.Search21.value); return false;">
<input type="text" name="Search21" size="30" placeholder="Hash" value="" />
<input type="submit" style="width:120px" value="Hashes.org" /><br /></form>
```

If this tool should stop working, you can create a free account at hashes.org and obtain your API key from your "Settings" menu. You would then replace `wQbxQlPJ0LZtLUVcBJKmDc8hnY5lcc` in the previous code with your own API key. This ensures long-term access to this powerful tool. In a moment, we will run through a few scenarios which takes advantage of this technique.

Online Search Resources

Throughout this book, I have mentioned various websites which allow query of target data against breached and leaked content. These do not display passwords, but will confirm presence within a specific breach. I will not revisit each option, as they have all been explained previously, but I will provide a summary with direct URL submission options. This will be vital for our Data Search Tool mentioned at the end of the chapter.

Email Address (test@test.com)
https://haveibeenpwned.com/unifiedsearch/test@test.com
https://dehashed.com/search?query=test@test.com
https://weleakinfo.com/search?type=email&query=test@test.com
https://intelx.io/?s=test@test.com
https://psbdmp.ws/api/search/test@test.com
https://leakprobe.net/ajax.php?query=searchLeaks&email=test@test.com
https://portal.spycloud.com/endpoint/enriched-stats/test@test.com

Username (test)
https://haveibeenpwned.com/unifiedsearch/test
https://dehashed.com/search?query=test
https://weleakinfo.com/search?type=username&query=test
https://intelx.io/?s=test
https://psbdmp.ws/api/search/test
https://leakprobe.net/ajax.php?query=searchLeaks&username=test

Domain (inteltechniques.com)
https://weleakinfo.com/search?type=domain&query=inteltechniques.com
https://dehashed.com/search?query=inteltechniques.com
https://psbdmp.ws/api/search/inteltechniques.com
https://intelx.io/?s=inteltechniques.com

Telephone (6185551212)
https://dehashed.com/search?query=6185551212
https://weleakinfo.com/search?type=phone&query=6185551212
https://psbdmp.ws/api/search/6185551212

IP Address (1.1.1.1)
https://dehashed.com/search?query=1.1.1.1
https://weleakinfo.com/search?type=ip&query=1.1.1.1
https://intelx.io/?s=1.1.1.1
https://psbdmp.ws/api/search/1.1.1.1

Name (Michael Bazzell)
https://dehashed.com/search?query=michael bazzell
https://weleakinfo.com/search?type=name&query=michael bazzell
https://psbdmp.ws/api/search/michael bazzell

Password (password1234)
https://dehashed.com/search?query=password1234
https://weleakinfo.com/search?type=password&query=password1234
https://psbdmp.ws/api/search/password1234

Hash (BDC87B9C894DA5168059E00EBFFB9077)
https://dehashed.com/search?query=BDC87B9C894DA5168059E00EBFFB90
https://weleakinfo.com/search?type=hash&query= BDC87B9C894DA5168059E00EBFFB90
https://psbdmp.ws/api/search/BDC87B9C894DA5168059E00EBFFB90

Sample Investigation

We have covered a lot so far within this chapter. Let's pause and conduct an investigation using this data. Assume that you possess the databases mentioned previously, especially the "Collection 1-5". Your investigation identifies a suspect with an email address of johndoe1287@gmail.com. This appears to be a burner email account, as you find no evidence of it anywhere online. It was used to harass your client and likely created only for devious activity. Our first step is to query the address through all of the databases you have acquired. Assume these are all stored on your external hard drive, which is plugged into your host computer. You have already connected this drive to a cloned OSINT virtual machine through the "Devices" menu in VirtualBox. Let's take each step slowly.

Open Terminal, type cd, then space, and drag-and-drop the external drive from your Files application into Terminal. This makes sure you are in the correct path. My command appears similar to the following.

```
cd '/media/osint/1TBExternal/'
```

I can now conduct a query within Terminal against all of my collected data. The following searches our target email address. Each query could take several minutes if you possess a lot of data and a slow drive.

```
rg -a -F -I -N johndoe1287@gmail.com
```

This results in one entry as follows.

```
Dumps-dehashed/nulled.to
johndoe1287@gmail.com:H8teful0ne45
```

This result is associated with a file titled "nulled.to", which is a hacking community which has suffered numerous breaches. We now know one of his interests and that he used the password of H8teful0ne45 on this site. We should next conduct a search of that password to see if it is used anywhere else. The following query is appropriate.

```
rg -a -F -I -N H8teful0ne45
```

Unfortunately, this returned no results. However, this search only queries for this exact term. If you possess a database which has not been dehashed, your target password could be present within an MD5, SHA1, or other hash. Therefore, let's convert this password into the most commonly used hashes with the following websites, displaying the output below each.

https://passwordsgenerator.net/md5-hash-generator/
9EF0EC63E2E52320CB20E345DCBA8112

https://passwordsgenerator.net/sha1-hash-generator/
D15FB15C1BC88F4B7932FD29918D1E9E9BBE7CA5

https://passwordsgenerator.net/sha256-hash-generator/
37A790A268B9FE62B424BABFC3BCAB0646BFB24B93EC1619AAE7289E0D7086DB

We can now submit each of these throughout all of our data with the following three commands.

```
rg -a -F -I -N 9EF0EC63E2E52320CB20E345DCBA8112
rg -a -F -I -N D15FB15C1BC88F4B7932FD29918D1E9E9BBE7CA5
rg -a -F -I -N 37A790A268B9FE62B424BABFC3BCAB0646BFB24B93EC1619AAE7289E0D7086DB
```

The first query returned the following result.

```
michael.bazzell411@gmail.com:9EF0EC63E2E52320CB20E345DCBA8112
```

This tells us that a user with an email address of michael.bazzell411@gmail.com possesses the password of "H8teful0ne45" within a breached database with MD5 hashed passwords. Is this the same person? It could be. It could also be a coincidence. The more unique a password is, the

more confidence I have that it is the same individual. This definitely warrants further investigation. My next query targets this new information.

```
rg -a -F -I -N michael.bazzell411@gmail.com
```

The results include the following.

```
michael.bazzell411@gmail.com:Hatefulone45
michael.bazzell411@gmail.com:johndoe
```

These results give me more confidence that these two accounts are owned by the same person. The variant of the "hateful" password and presence of "johndoe" within the original email address and the new password convinces me we are on the right track. I would now target this new email address and replicate the searches mentioned within previous chapters. We should also check our Pastes search tool and the online breach resources previously explained.

Your biggest frustration may be the speed of each query. I possess all of my data within an internal solid-state drive (SSD) with amazing read speeds. It still takes a few of minutes to parse through all of my data (2TB). If you are using an external spinning drive, expect that time to triple. If this technique proves to be valuable, you might consider a dedicated machine for this sole purpose. Personally, I never conduct these queries within a virtual machine due to these speed issues. I have a dedicated MacBook Pro with a 4TB internal drive to store and query my content. This may be overkill for your needs.

Data Leaks

A data leak is much different than a data breach. Breaches are usually deliberate criminal intrusions which result in stolen data. Leaks are usually unintentional public exposure of data. The next time you hear about a "misconfigured server" which exposed millions of customer details, this is likely a data leak. Our best tool to find this data is Shodan (shodan.io), and you will need to be logged in to an account to conduct these queries. I will focus only on open Elasticsearch databases, which are extremely easy to access. Our Shodan search is as follows.

product:elastic port:9200 [target data]

As a demonstration, I want to search for open Elasticsearch databases which contain an index titled "Customer". Think of an index as the title of a table or section of the database. The following search on Shodan produces over 200 results including an index titled "customer".

product:elastic port:9200 customer

The first result is an open database with 401GB of data. I have redacted the IP address and modified each address throughout this example. I will use a fictitious address of 34.80.1.1 throughout the entire demonstration. Figure 27.02 displays this result. Clicking the red square

with arrow next to the redacted IP address connects to the database within your browser in a new tab. Figure 27.03 displays this result. This confirms that the database is online and open.

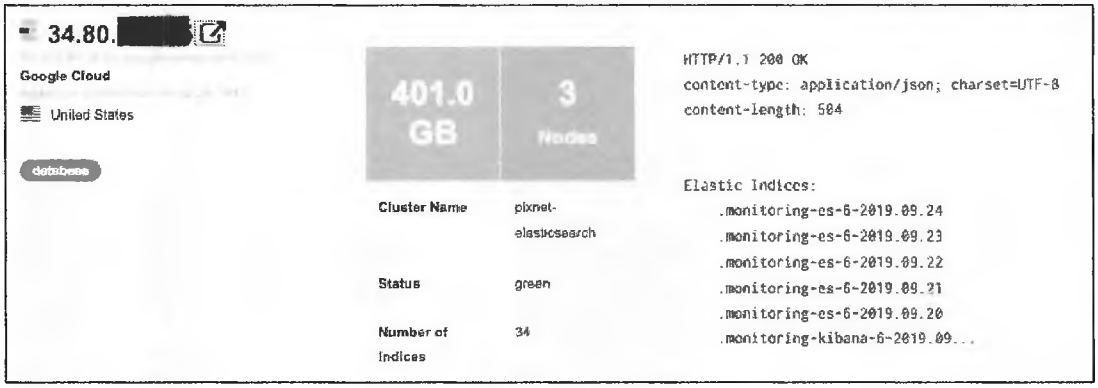


Figure 27.02: A Shodan result of an open Elasticsearch database.

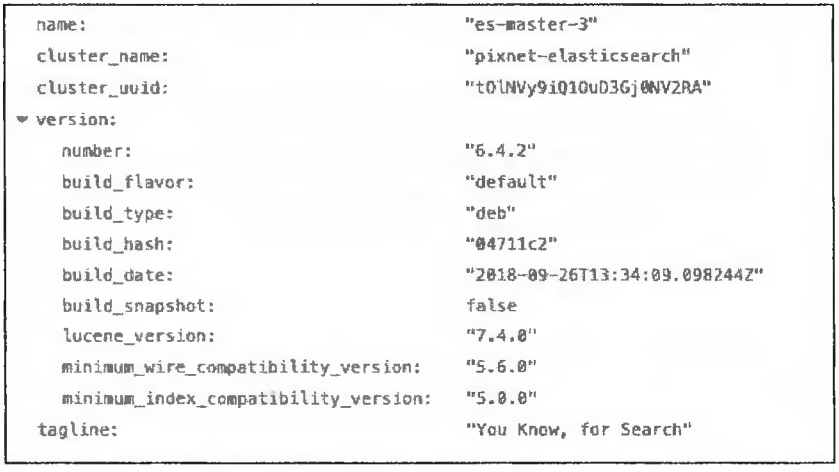
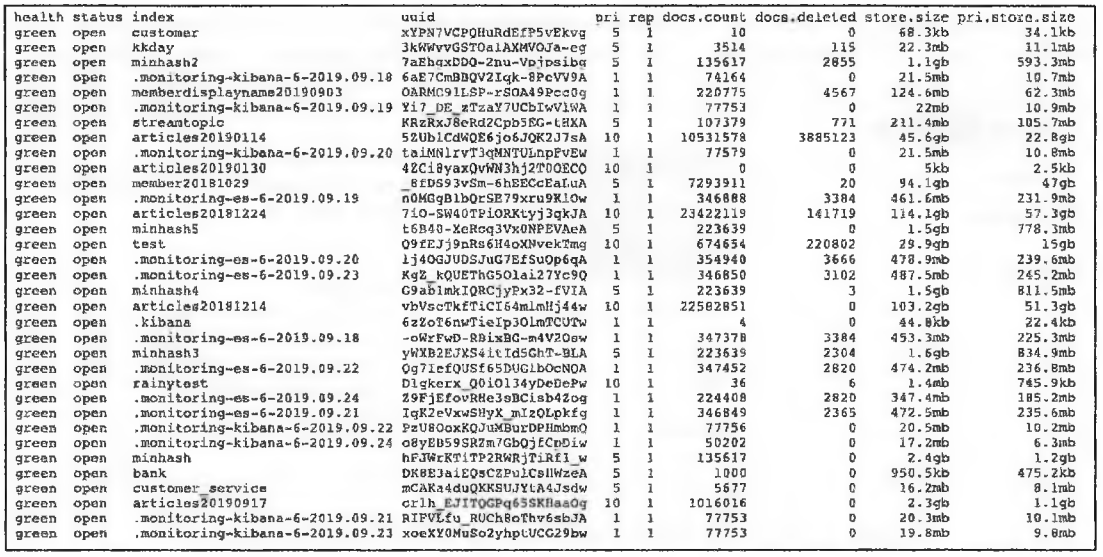


Figure 27.03: An open Elasticsearch server.

Next, we want to obtain a list of all indexes within this database. These titles are often indicative of the content. The following URL queries this data based on a target IP address of 34.80.1.1. The result can be seen in Figure 27.04.

http://34.80.1.1:9200/_cat/indices?v



health	status	index	uuid	pri	rep	docs.count	docs.deleted	store.size	pri.store.size
green	open	customer	xYPN7VCPQHuRdEfP5vEkvg	5	1	10	0	68.3kb	34.1kb
green	open	kkday	3kWWvvGSTOalAXMVOJa-eg	5	1	3514	115	22.3mb	11.1mb
green	open	minhash2	7aEhqxDDQ-2nu-VpipsibQ	5	1	135617	2855	1.1gb	593.3mb
green	open	.monitoring-kibana-6-2019.09.18	6aE7CmBBQV2Iqk-8PcVV9A	1	1	74164	0	21.5mb	10.7mb
green	open	memberdisplayname20190903	OARMC91LSP-rSOA49PccOg	1	1	220775	4567	124.6mb	62.3mb
green	open	.monitoring-kibana-6-2019.09.19	Yi7_DE_zTzaY7UCbIwVlWA	1	1	77753	0	22mb	10.9mb
green	open	streamtopic	KRzRxJ8eRd2Cpb5EG-tHXA	5	1	107379	771	211.4mb	105.7mb
green	open	articles20190114	52Ub1CdWQE6jo6JQK2J7sA	10	1	10531578	3885123	45.6gb	22.8gb
green	open	.monitoring-kibana-6-2019.09.20	taiMNlrvT3qMNTULnpFvEw	1	1	77579	0	21.5mb	10.8mb
green	open	articles20190130	4ZCi8yaxQvWN3hj2T0OECQ	10	1	0	0	5kb	2.5kb
green	open	member20181029	_8fDS93vSm-6hEECcEaLuA	5	1	7293911	20	94.1gb	47gb
green	open	.monitoring-es-6-2019.09.19	n0MGqB1bQrSE79xru9K1Ow	1	1	346888	3384	461.6mb	231.9mb
green	open	articles20181224	7iO-SW40TPiORKtyj3qkJA	10	1	23422119	141719	114.1gb	57.3gb
green	open	minhash5	t6B40-XeRcq3VxONPEVAeA	5	1	223639	0	1.5gb	778.3mb
green	open	test	Q9fEJj9nRs6H4oXNvekTmg	10	1	674654	220802	29.9gb	15gb
green	open	.monitoring-es-6-2019.09.20	1j4OGJUDSJuG7EfSuQp6qA	1	1	354940	3666	478.9mb	239.6mb
green	open	.monitoring-es-6-2019.09.23	KgZ_kQUEThG5Olai27Yc9Q	1	1	346850	3102	487.5mb	245.2mb
green	open	minhash4	G9ablmkIQRCjyPx32-fVIA	5	1	223639	3	1.5gb	811.5mb
green	open	articles20181214	vbVscTkfTiCI64mlmHj44w	10	1	22582851	0	103.2gb	51.3gb
green	open	.kibana	6zZoT6nwTieIp3OlmTCUTw	1	1	4	0	44.8kb	22.4kb
green	open	.monitoring-es-6-2019.09.18	-oWrFwD-RBixBG-m4V2Osw	1	1	347378	3384	453.3mb	225.3mb
green	open	minhash3	yWXB2EJXS4itId5GhT-BLA	5	1	223639	2304	1.6gb	834.9mb
green	open	.monitoring-es-6-2019.09.22	Qg7IefQUSf65DUGlbOcNQA	1	1	347452	2820	474.2mb	236.8mb
green	open	rainytest	Dlgkerx_Q0iOl34yDeDePw	10	1	36	6	1.4mb	745.9kb
green	open	.monitoring-es-6-2019.09.24	Z9FjEfovRHe3sBCisb4Zog	1	1	224408	2820	347.4mb	185.2mb
green	open	.monitoring-es-6-2019.09.21	IqK2eVxwSHyX_mIzQLpkfg	1	1	346849	2365	472.5mb	235.6mb
green	open	.monitoring-kibana-6-2019.09.22	PzU8OoxKQJuMBurDPHmbmQ	1	1	77756	0	20.5mb	10.2mb
green	open	.monitoring-kibana-6-2019.09.24	o8yEB59SRZm7GbQjfCpDiw	1	1	50202	0	17.2mb	6.3mb
green	open	minhash	hFJWrKTiTP2RWRjTiRf1_w	5	1	135617	0	2.4gb	1.2gb
green	open	bank	DK8EJaiEQsCZPulCsIWzeA	5	1	1000	0	950.5kb	475.2kb
green	open	customer_service	mCAKa4duQKKSUJYtA4Jsdw	5	1	5677	0	16.2mb	8.1mb
green	open	articles20190917	crlh_EJITQCPq65SKHaaOg	10	1	1016016	0	2.3gb	1.1gb
green	open	.monitoring-kibana-6-2019.09.21	RIFVLfu_RUCh8oThv6sbJA	1	1	77753	0	20.3mb	10.1mb
green	open	.monitoring-kibana-6-2019.09.23	xoeXY0MuSo2yhpUCG29bw	1	1	77753	0	19.8mb	9.8mb

Figure 27.04: A list of indexes within an open Elasticsearch database.

We now know there are several indexes which can be viewed. The first is titled "customer", but only contains a small amount of data (68kb). Near the bottom, we see an index titled "memberdisplayname20190903" which contains 124mb of data including customer usernames. Some of these indexes contain email addresses and other sensitive data. Let's focus on the index titled "bank". We can view the content with the following URL.

http://34.80.1.1:9200/bank/_search?size=100

This combines the IP address (34.80.1.1), necessary port (9200), name of the index (bank), and instructions to display the first 100 records (/_search?size=100). If we wanted to view the maximum number of entries visible within the browser, we would query 10,000 records with the following URL.

http://34.80.1.1:9200/bank/_search?size=10000

Figure 27.05 displays actual redacted results from this query. It identifies the first name, last name, email address, gender, city, state, bank account number, and balance associated with a customer. The rest of the index contains the same information for additional customers.

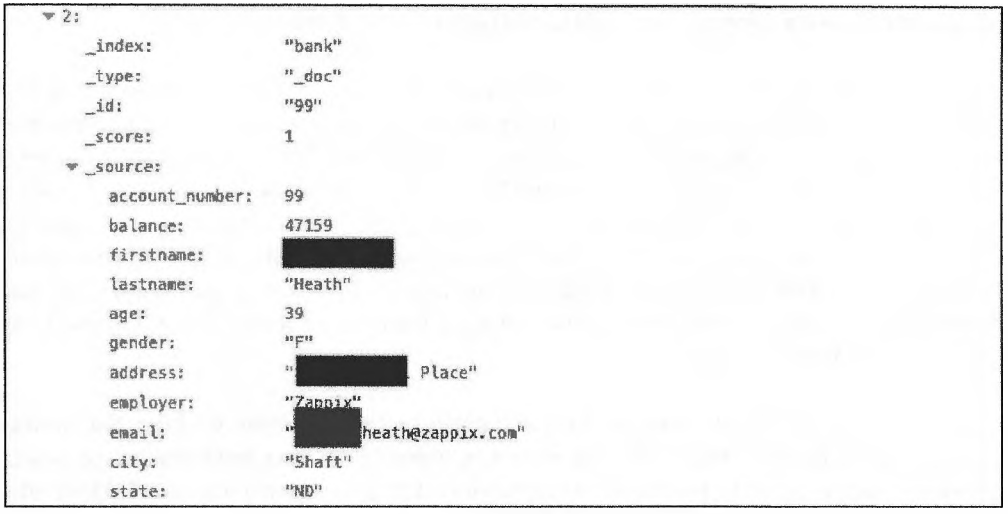


Figure 27.05: An entry from an open Elasticsearch database.

Issues

There are many complications with acquisition of open Elasticsearch data. The first is the record limitation. Displaying results within a URL limits the results to 10,000 records. Modifying the URL as described previously presents all data possible within the browser. Saving this page stores all of the content for future queries. However, many of the data sets I find contain millions of records. A Python script which parses through all results and stores every record is most appropriate, and is explained in a moment.

Next are the legal considerations. Technically, this data is publicly available, open to anyone in the world. However, some believe the act of manipulating URLs in order to access content stored within a database exceeds the definition of OSINT. I do not agree, but you might. I believe most of this data should have been secured, and we should not be able to easily collect it. The same could be said for FTP servers, paste sites, online documents, and cloud-hosted files. I believe accessing open databases becomes a legal grey area once you decide to use the data. If you are collecting this content in order to sell it, extort the original owner, or publish it in any way, you are crossing the line and committing a crime. I remind you that the Computer Fraud and Abuse Act (CFAA) is extremely vague and can make most online activity illegal in the eyes of an aggressive prosecutor. Become familiar with the data access laws in your area and confirm that these techniques do not violate any laws or internal policies. My final reminder and warning is that I am not an attorney. I am not advising that you conduct any of these methods on behalf of your own investigations. I am simply presenting techniques which have proven to be extremely valuable to many investigators. If you believe that accessing an open (public) Elasticsearch database is legal in your area, and does not violate any internal policies, it is time to parse and download all contents.

Elasticsearch Crawler (github.com/AmIJesse/Elasticsearch-Crawler)

In early 2019, I was made aware of an open Elasticsearch database which exposed sensitive data associated with 57 million people. In most cases, these records contained personal information such as first name, last name, email address, home address, state, ZIP code, phone number, and IP address. These types of records are extremely helpful to me as an investigator. Connecting personal email addresses with real people is often the best lead of all online research. I had found the database on Shodan using the methods discussed here. Specifically, I searched for an index titled "Leads" and sifted through any results of substantial size. Once I had located the data, I was desperate to download the entire archive. With a browser limit of 10,000, I knew I would need the help of a Python script.

I reached out to my friend and colleague Jesse and explained my scenario. Within a few moments, he sent me a small Python script. This file is now a staple in my data leaks arsenal of tools. He has agreed to share it publicly, please use it responsibly. First, let's install it to our OSINT Master VM. Enter the following commands within Terminal.

- `cd ~/Downloads/Programs`
- `git clone https://github.com/AmIJesse/Elasticsearch-Crawler.git`
- `cd Elasticsearch-Crawler`
- `pip install nested-lookup`

You are now ready to download an entire Elasticsearch open database and specify which fields should be acquired. Note that you must open Terminal and navigate to your script in order to launch this utility. I have intentionally omitted a desktop shortcut for this program due to the sensitive nature of the data. I do not want people experimenting with a clickable icon without understanding the technique. Therefore, this will only be presented as a Terminal execution. However, you could easily use the previous tutorials and scripts in order to create your own shortcut. The following commands from within any Terminal session will launch the script.

- `cd ~/Downloads/Programs/Elasticsearch-Crawler`
- `python crawl.py`

This will present several user prompts to enter the target IP address, index name, port number, and fields to obtain. Let's walk through a real demonstration in order to understand the application. You must be logged in to a free or premium Shodan account in order to complete this tutorial. Using the IntelTechniques Data Breaches & Leaks Tool, seen in Figure 27.06 at the end of the chapter, I entered "leads" into the field titled "Elasticsearch". This conducted an exact search of "product:elastic port:9200 leads" within Shodan, which displayed a handful of results. One of these was an Elasticsearch server in India. This database appeared to contain test data, so I will not redact any of the results. The IP address was 111.93.162.238 and the database was approximately 1GB in size. Clicking the red square within the result on Shodan opened a new tab to the following address.

http://111.93.162.238:9200/

The brief response confirmed that the server was online. The URL discloses the IP address (111.93.162.238) and port (9200). This is the default port and is almost always the same. Now that I had the IP address, I entered it within the option titled "Index List" within the tool. Clicking that button launched a new tab at the following address.

http://111.93.162.238:9200/_cat/indices?v

This connects to the IP address (111.93.162.238) and port (9200), and then conducts a query to display all public indexes (/_cat/indices?v). The result included the following.

index	store.size
imobilestore	1.3mb
testcsv	190.7kb
easyphix	1.8mb
index_test	7.2kb
lambda	1.2kb
crazyparts	503.5kb
lead	1.2kb
from_lambda	8.3kb
valueparts	2.3mb
surtestcsv	40.1kb
.reporting-2018.11.25	19.3kb
mobilemart	1.7mb
surtest1csv	237.5kb
sw	69.9kb
sample_data_flights s	6.4mb
my-index	226.4mb
scrape_results	45.7mb
comparision_sample1	2.3mb
leads	280.8kb
demo-index	69.9kb
csv-import	190.7kb
.kibana	91.9kb

I usually look at both the names and the sizes. If I see an index titled "Customers", I know it usually contains people's information. If it is only 1kb in size, I know it is too small to be of any use. When I see any index with multiple gigabytes of data, my curiosity kicks in and I want to see the contents. For this demonstration, let's focus on the index of our original search of "leads". Within our search tool, the next option is labeled "Index View". This requires the IP address of our target (111.93.162.238) and the name of the index you wish to view (leads). This opens a new tab at the following URL.

http://111.93.162.238:9200/leads/_search?size=100

This connects to the target IP address (111.93.162.238) and port (9200), loads the desired index (leads) and displays the first 100 results (/_search?size=100). This is usually sufficient to see enough target content, but this can be raised to 1000 or 10000 if desired. Below is a record.


```
"_index": "leads",
"_type": "leads",
"_id": "PXIhqmUBcHz5ZA2uOAe7",
"_score": 1,
"_source": {
 "id": "86",
 "email": "test80@agencyinnovations.com",
 "first_name": "test80",
 "last_name": "test80",
 "phone": "32569874",
 "ip": "0.0.0.0",
 "orgname": "Sales Arena",
 "tags": "",
 "isDeleted": false,
 "created_at": "2018-09-05 19:57:08",
 "updated_at": "2018-09-05 19:57:08",
 "fields": {}
```


This is obviously test data, but assume it was a record containing a real person's name, email address, and phone number. Also assume there were over a million records within this index, which is quite common. We could save this page, but would be forced to save the undesired fields such as "tags" and "_source". Also, the data would be in a difficult format to search. This is where our new Python script is helpful. You have already launched the crawl.py script, and should have been presented with a prompt for the IP address of the target. The following displays each entry I submitted for this demonstration.

```
IP address: 111.93.162.238
Index name: leads
Port (Default is 9200): 9200
Field values to obtain (submit an empty line when finished):
Value: email
Value: first_name
Value: last_name
Value: phone
Value:
```

After being prompted for the IP address (111.93.162.238), it asked me for the target index name (leads) and port number (9200). It then prompted me to enter the first field I wanted to acquire (email). Since I entered a field, it then prompted for the next field (first_name). The tool will continue to ask for field names for as long as you provide them. Notice there is an empty line in the last "Value". This empty result tells the script you are finished, and it begins collecting the data. When finished, a text file will be saved in the same directory as your script. In this example, it was at ~/Downloads/Programs/Elasticsearch-Crawler/111.93.162.238-leads.txt. The title of the file was automatically created to reflect the IP address and name of the index. The following are the first three lines of this text file.

test65@agencyinnovations.com,test65,test65,987485746
test22@agencyinnovations.com,test22,test22,124958616
test69@agencyinnovations.com,test69,test69,2145968

If this were real data, you would see millions of people's email addresses, names, and telephone numbers. There are likely hundreds of legitimate databases on Shodan right now, just waiting to be found. The next time you see a news article about a security researcher who found an exposed database containing millions of sensitive records, it is very likely that Shodan and a similar script was used. If your downloaded file contains random text, you have likely encountered a patched version of Elasticsearch. At the time of this writing, Elasticsearch databases version 6.4.0 and newer were blocking the script. Anything older worked fine. There may be a new release of this crawler, and you may need to update your script as follows. Please note these commands are also included in the linux.txt file previously downloaded.

- `cd ~/Downloads/Programs/Elasticsearch-Crawler`
- `git pull https://github.com/AmIJesse/Elasticsearch-Crawler.git`

The final option within the search tool allows you to search a target IP address, index, and term. Let's conduct a new demonstration. Assume you have already searched "customer" within the first "Elasticsearch" option within the search tool. You have located an open database of interest and viewed the list of indexes with the "Index List" feature. You copied the name of the index and IP address to the "Index View" search. However, the file contains millions of records, and you can only see the first 10,000. You might want to search within the remaining records to see if your target is present. We can do this with the "Index Search" feature, as seen as the last option in Figure 27.06.

Enter your target Elasticsearch IP address in the first field and the target index name in the second. Enter any desired search term in the last. This could include the email address or name of your target. It is also less invasive than downloading all of the content. In May of 2019, I located an extremely sensitive open Elasticsearch database. It contained Social Security Numbers and medical records. I did not want to download the data, but I did want to search my own name. I entered the IP address and index name in the first two fields and "Bazzell" in the last. The query returned dozens of patients' records associated with my last name, but nothing connected to me.

This was all done within the web browser through their open server, and I did not archive any data onto my own machine. I identified the host and reported the leak anonymously. I never received a response, but the data was removed the next day.

I predict we will see fewer open databases within 2020. While we still hear about sensitive leaks every week, word is spreading and companies are locking down their data. This is a good thing for all of us. Until then, I will continue to search.

SQL Files

Finally, some old-fashioned Google queries might find more data breaches and leaks than one can manage. Let's conduct a few examples with SQL files. SQL, often pronounced "sequel", is an acronym for Structured Query Language. It is a computer language used in programming and designed for managing data held in a relational database management system. In other words, most SQL files are databases of some sort. Many of the most popular database breaches were originally released as SQL files. WordPress backups, web forum exports, and other online maintenance files are also stored in this format. Searching for public SQL files can reveal surprising results. First, let's search Google for these files with the following commands using the tutorials discussed within previous chapters.

ext:sql

This returns millions of results. While some are files with the extension of ".sql", most of the results are web pages ending with ".sql", and are not helpful. Let's modify the search as follows.

ext:sql "create table"

This returns 55,000 results which include the exact terms "create table". This is a standard statement within SQL files which specifies the names of tables within the database. This filters most of the website names from our search and displays only valid SQL files. Next, let's add to our search with the following modification.

ext:sql "create table" "@gmail.com"

This returns 10,000 results. Each is an SQL database with at least one entry of "gmail.com" inside. This indicates that active email addresses are within the files, which is indicative of a true breach or leak. The following search should reveal data worth analyzing.

ext:sql "create table" "@gmail.com" "'password'"

This returns 5,400 results. Each is an SQL database with at least one entry of "gmail.com" inside and a table titled "password". Note the single quotes around password within double quotes. This tells Google to search specifically for 'password' and not the word alone. 5,400 results are

overwhelming, and include a lot of test files stored on Github. Since many people use Gmail addresses as the administrator of a test databases, we should add another email domain as follows.

ext:sql "create table" "@gmail.com" "'password'" "@yahoo.com" -site:github.com

This reveals 228 results. Each are SQL files which contain at least one Gmail address, one Yahoo address, and a table titled password. Further, all results from Github are removed. Most of these results will load as text files within your browser. However, some will be quite large and may crash your browser while loading. The following is a modified excerpt from an actual result, which I found by searching "@gmail.com" within the browser text.

(62,'RichardWilson','admin','richard@redactedemail.com','4d5e02c3f251286d8375040ea2b54e22','Administrator',0,1,25,'2008-05-28 07:07:08','2009-04-02 13:08:07')

This is a very typical structure within SQL files. The following explains each piece.

62, (User ID)
'RichardWilson', (Name provided)
'admin', (Username)
'richard@redactedemail.com', (Email address)
'4d5e02c3f251286d8375040ea2b54e22', (Hashed password)
'Administrator', (Usertype)
0,1,25, (Variable internal codes)
'2008-05-28 07:07:08', (Registration date)
'2009-04-02 13:08:07', (Last visit date)

You could save the entire page as a .txt file (right-click > Save page as...) for your personal data archive. An alternative query for text files is as follows.

ext:txt "create table" "gmail.com" "'password'" "yahoo.com" -site:github.com

There may be a trove of sensitive information within these files, so use caution and always be responsible. Exercise good defense when browsing any of these sites or downloading any data. Trackers, viruses, and overall malicious software are always present in this environment. Using a Linux virtual machine and a reputable VPN will provide serious protection from these threats. I search and store leaked and breached databases as a part of every investigation which I conduct. I can say without hesitation that these strategies are more beneficial than any other online investigation technique of which I know. Some investigators within my circles possess several terabytes of this data from tens of thousands of breaches and leaks. Querying your own offline archive during your next investigation, and identifying unique data associated with your target, can be extremely rewarding. Again, I ask you to be responsible. Never use any credentials to access an account and never allow any data obtained to be further distributed. Use this public data, stolen by criminals, to investigate and prosecute other criminals.

Public Data Sets

Many large data sets which are beneficial to investigations are not "breaches" or "leaks". This chapter has focused mostly on data which was never meant to become publicly available. However, there are numerous archives full of public information which should be considered for your data collection. Most archives cannot be searched through traditional resources such as Google. Instead, we must acquire the data, condense it, and then conduct our own queries. As a demonstration, I will explain how I utilize Usenet archives as a vital part of my investigations.

Usenet was my first introduction into newsgroups in the early nineties. My internet service provider allowed full access to thousands of topics through Outlook Express. I could subscribe to those of interest and communicate via email with people from all over the world. This sounds ridiculously common today, but it was fascinating at the time. I located a newsgroup about my favorite band, and I was quickly trading bootleg cassettes and exchanging gossip about the music industry. I was freely sending messages without any consideration of any abilities to permanently archive everything.

Today, the Internet Archive presents huge repositories of data containing over thirty years' worth of Usenet messages. It is presented in over 26,000 archives, each containing between 1 to 20,000 files. It is massive, and would take years to download manually. Fortunately, we can use a download script created by the Internet Archive to automatically obtain all desired data. The following tutorial will install the necessary software into our Linux virtual machine; download the entire Usenet archive; extract email addresses and names of each member; acquire newsgroup data to associate with each person; and condense the data into a useable format. This will take some time and may be overwhelming, but the final product is worth the effort. Let's begin.

First, we need to install the Internet Archive script within out OSINT Master VM. Within Terminal, enter the following.

- `pip install internetarchive`

We now have the script installed and ready to use from any folder. Next, we need to identify the Internet Archive collections which we want to acquire. For this demonstration, I will focus on the following two data sets.

https://archive.org/details/giganews
https://archive.org/details/usenethistorical

These are two independent Usenet archives. Each contain unique records and some redundant data. Let's take a look at the second collection. The first folder is titled Usenet-Alt and contains over 15,000 files extracted from decades of conversations within the "Alt" communities. Opening the file titled alt.2600.fake-id.mbox reveals names, email addresses, dates, and entire messages dating back to 1997. The following is an excerpt.

From: "David N." <david.nine@juno.com>
Subject: Need Fake Texas DL
Date: 1998/07/11
Newsgroups: alt.2600.fake-id

I need a quality bogus TX DL for my 17 year old sister. Can you help me out?

Possessing a database of every email address and name from thirty years of Usenet posts can be very beneficial. Downloading every message can be overkill. This entire data set is over a terabyte in size. Instead of trying to download everything, I only want specific portions of the data. The Giganews collection includes two files for each archive. The first is the "mbox" file which includes the full messages along with user information. These are very large and could take months to download. The second is a "csv" file which only includes the date, message ID, name, email address, newsgroup, and subject of each post. This is a much more manageable amount of data which includes the main information desired (name and email address). We will only download the minimal information needed for our purposes.

First, we must create a text file which includes every archive within each collection. The following commands navigate to your Desktop and create text files for the Giganews and Usenet Historical archive.org data sets.

- `cd ~/Desktop`
- `ia search 'collection:giganews' -i > giganews1.txt`
- `ia search 'collection:usenethistorical -i > usenethistorical1.txt`

Let's dissect the Internet Archive commands. "ia" is the application, "search" identifies the type of query, "collection:giganews" identifies the target data on archive.org, "-i" instructs the application to create an item list, and " > giganews1.txt" provides the desired output. These text files on your Desktop contain the names of all of the archives within the collections. The following is an excerpt.

usenet-alt.2600
usenet-alt.2600a
usenet-alt.2600crackz

We can now instruct Internet Archive to begin downloading the necessary files. The following command downloads only the "csv" files from the Giganews collection. It can take several hours if you have a slow internet connection. If you do not have sufficient space within your VM, consider saving these to an external drive as previously instructed.

- `ia download --itemlist giganews1.txt --glob="*.csv.gz"`

You should now have thousands of folders, each containing multiple compressed CSV files. This is not very useful or clean, so let's extract and combine all of the valuable data. The following command will decompress all of the files and leave only the actual CSV documents. It should be executed from whichever directory contains all of the downloaded folders. In my demonstration, it is in the Desktop.

- `gunzip -r .`

"gunzip" is the command to extract the data from the compressed files, "-r" conducts the command recursively through any sub-folders, and " ." continues the action through every file. Below is a modified excerpt from one of the files. It identifies the date of the post (12/04/2003), name of the author (John Smith), email address associated with the account (John-smith@smh.com), specific newsgroup used (microsoft.windowsxp), and the subject of the post (Help me!).

#date	from	newsgroups	Subject
20031204	John Smith <John-smith@smh.com>	microsoft.windowsxp	Help Me!

We still have thousands of files, which is not ideal. The following command will combine every CSV file into a single text file titled Giganews2.txt. Furthermore, it will only extract the columns of data most valuable to us, as explained afterward.

- `find . -type f -name \*.csv -print0 | xargs -0 cut -f1,3,4,5 > Giganews2.txt`

Let's break down each portion of this command, as it can be quite useful toward other data sets.

find	This is the command to "find" data to manipulate.
.	This instructs the command to find all files.
-type f	This searches for a regular file type.
-name *.csv	This filters to only find a specific file extension (csv).
-print0	This directs output to a file instead of the screen.
|	This is a "pipe" character which separates the commands for a new instruction.
xargs -0	This builds our next command from the previous data as input.
cut -f1,3,4,5	This extracts only the data from columns 1,3,4 and 5 from each CSV.
>	This instructs the command to send the data to another file.
Giganews2.txt	This identifies the output file name.

The final result should be a very large file which contains all of the valuable content from within every downloaded file. In a moment, we will use this data to research our targets. The Usenet Historical collection is stored in a different format, so these instructions will need to be modified. The following steps will extract the most beneficial data from that collection, and should appear similar to the previous actions. First, we must download the entire archive with the following command.

- `ia download --itemlist usenethistorical1.txt --glob="*.zip"`

Next, we must extract the "mbox" files from their compressed containers with the following.

- `find . -name "*.zip" -exec unzip {} \;`

Finally, we must extract the "From:" line from each file with the following command.

- `rg -a -F -i -N "From: " > UsenetHistorical2.txt`

This leaves us with a single large file titled UsenetHistorical.txt. Below are a few lines.

```
alt.2600.mbox:From: "Bill Smith" <momop@mvc.biglobe.ne.jp>
alt.2600.mbox:From: "yosinaga jackson" <purizou@geocities.co.jp>
alt.2600.mbox:From: "yosinaga jackson" <purizou@geocities.co.jp>
```

We do not have as many details within this data set as we did with the Giganews option. However, possessing the names, email addresses, and newsgroups provides a lot of value. Both the Giganews2.txt and UsenetHistorical2.txt files possess many duplicate entries wasting valuable space. You might consider the following command which combines both files into one file while removing all duplicate lines.

- `sort -u -f Giganews2.txt UsenetHistorical2.txt > UsenetFinal.txt`

We now possess a single file, quite large in size, which possesses the names, email addresses, and interests of most users of the Usenet system over a thirty year period. Now, let's discus ways this can be used during investigations. Assume you are researching a target with my last name. When using Ripgrep, your command is as follows.

- `rg -a -F -i -N bazzell UsenetFinal.txt`

The result includes the following partial data.

```
microsoft.public.pocketpc,Steven Bazzell steveill@yahoo.com
sci.bio.microbiology,cbcWiW@UJOMHEH.EDU.TW Wayne A. Bazzell,M.P.S.E
sci.crypt,cbcWiW@UJOMHEH.EDU.TW Wayne A. Bazzell,M.P.S.E
sci.crypt,General Darcy J. Bazzell corporation@incidentally.mi.us
```

The first line identifies Steven Bazzell in the Usenet group of microsoft.public.pocketpc while using an email address of steveill@yahoo.com. You could search by names, emails, partial emails, domains, etc. I have successfully used my own Usenet archive in the following scenarios.

- When searching a unique target name, I have uncovered old email addresses which led me to otherwise unknown social network profiles.

- When searching an email address, I have identified various interests of the target, determined by the newgroups to which he or she has posted.
- When searching an email address, it serves as a great way to verify a valid account and establishes a minimum date of creation.
- When searching a domain, I often identify numerous email addresses used by the owner.
- When conducting background checks on targets over the age of 35, I often identify email addresses connected to questionable interests. While investigating a potential police officer, I located evidence he had previously posted images to a child pornography newsgroup. This was confirmed during the interview.
- When searching a username, I often locate confirmed email addresses from several years prior. These open new possibilities once provided to our search tools. In 2019, I located an otherwise unknown email address associated with my target username. It had been used to post to a hacking forum in 2002. The name, email, and username only identified "JohnDoe". This new email account was associated with a unique password within the Myspace breach data. Searching the password identified additional email accounts within other breach data sets. Searching those accounts within the Usenet data displayed the true name of my target. Everyone makes a mistake eventually.
- Any time I see a serial killer, active shooter, or otherwise deranged individual in the news, I seek any presence within old newsgroups. More often than not, if the subject was tech-savvy in the 90's, I find interesting evidence missed by most news media.

These collections possess over 43 million unique email addresses. You may be questioning the justification of the time it would take to create your own archive. While I encourage you to replicate the steps here, I also respect the difficulty involved. Therefore, I have made my own Usenet file available for download at the following address and shortened URL.

https://mega.nz/#!qDYlma4B!vdRl1AMGTYLFfPhqfUHZ3pyiL9cbPWMlpkTLMJof0D8
https://bit.ly/32exiQO

The file contains only the name, email address, and newsgroup fields from both collections in order to keep the size minimal. I have removed all duplicates and cleaned the file formatting to exclude any unessential data. I hope you find it to be useful.

I present a final warning about disk space. If you replicate all of these steps within a VM which is only allocated a small disk (less than 100GB), expect to run out of space. If you plan to include data breaches and leaks into your daily routine, you might consider a dedicated Linux host. Expect frustration, complications, and slow queries if using an external drive. The methods presented in this chapter can be conducted within any operating system as long as the proper utilities are installed. I use a dedicated Linux laptop with a 4TB SSD internal drive for my data collection. My queries are fast, and I never worry about disk space. My Linux OS provides protection from malicious software, and this machine is never used to conduct online investigations. Once you see the benefits of this data, you might be willing to make the jump.

Scylla (scylla.sh)

I include this resource at the end of the chapter because I suspect its longevity is questionable. Scylla is the only current service which allows free query of compromised data while displaying full password details. It requires no account, and the results are impressive. Many users flock to this page, search an email address or domain, and become frustrated when the only result is an error message. This is because the query syntax for Scylla is unique from other sites. You must include the type of content, followed by a colon, followed by the target data. If I wanted to search test@inteltechniques.com, my query would be the following.

email:test@inteltechniques.com

The following results are fictional, but represent a typical response after an email search.

Domain	linkedin.com	(The source of the breach)
IP	98.221.15.4	(The IP address of the user)
Password	p@$$w0rd12	(The plain-text password from the breach)
Name	Michael Bazzell	(The user's name)
Email	MB@inteltech.com	(The user's email address)
PassHash	hlk2h34jhg2346236	(The hash of the password, if available)
PassSalt	;87sr	(The salt of the hash, if available)
User	MBIntel	(The username of the breached account)
UserId	87611344	(The user number of the breached account)

Now, let's look at the URL structure for each query. This can be used if you decide to add this resource to your own custom search tools. I only include the queries which display results.

IP	https://scylla.sh/search?q=ip:98.221.15.4
Password	https://scylla.sh/search?q=password:password123456789
Name	https://scylla.sh/search?q=name:Michael+Bazzell
Email	https://scylla.sh/search?q=email:MB@inteltech.com
PassHash	https://scylla.sh/search?q=passhash:hlk2h34jhg2346236
PassSalt	https://scylla.sh/search?q=passsalt:;87sr
User	https://scylla.sh/search?q=passsalt:MBIntel
UserId	https://scylla.sh/search?q=userid: 87611344

I intentionally omitted Scylla from the following custom search tool. First, it displays full credentials, which may violate some users' clearances or organizational policies. Second, I suspect this service will disappear or become a paid product soon, and I do not want a non-functioning resource in your tools template. Third, and most important, the page launches a cryptocurrency "miner" which uses your computer's resources to obtain Bitcoin for the owner (which can be disabled with uBlock Origin as explained previously). Finally, this is an excellent opportunity to test your own tool modification skills by adding this service if you desire.

IntelTechniques Breaches & Leaks Tool

This final search tool combines most of the online search options mentioned throughout the chapter. The breach data resources are split into categories based on the target data (email, username, etc.). The Hash Conversion opens a new tab when a hash is presented. It queries the hashes.org API and a key is already provided. Finally, the last feature allows entry of any found password and immediately generates an MD5, SHA1, and SHA256 hash for further research. Figure 27.06 displays the current view.

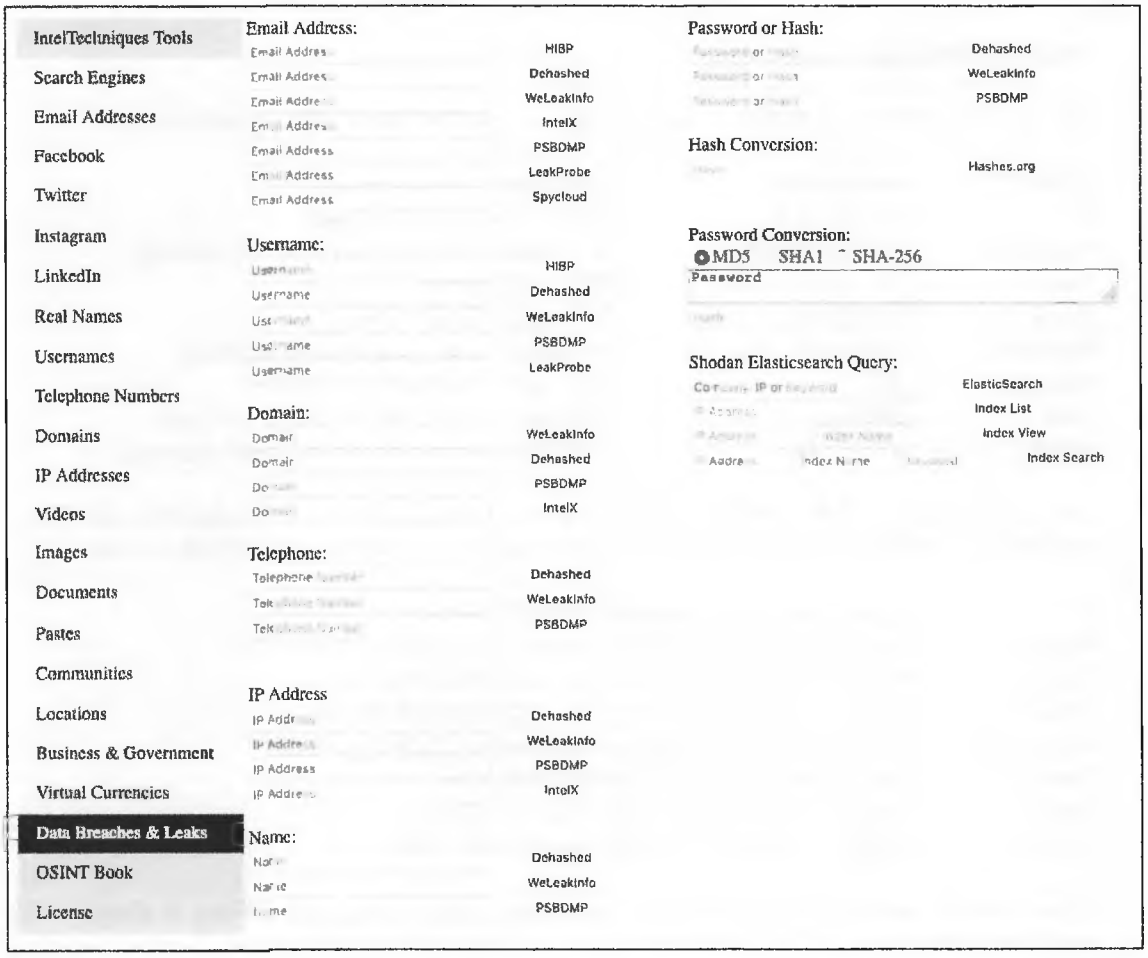


Figure 27.06: The IntelTechniques Breaches & Leaks Tool.

SECTION III
OSINT METHODOLOGY

In the late 90's, I was tasked to investigate a computer-related crime involving inappropriate online contact from a registered sex offender to children in his neighborhood. The internet was new to most people, AOL dial-up connections were common, and there was very little monitoring or enforcement in place. I contacted the subject at his home and conducted an interview. He admitted to inappropriate behavior and showed me the evidence on his computer. I had no forensic imaging machine or acquisition methods. I didn't even have a digital camera. I had my notepad and pen. Months later, I testified about the illegal activity this suspect conducted with local children. I verbally explained what I observed on his computer without any digital evidence. It was a very different time, and would never be acceptable today. Current prosecution would require forensic acquisition, detailed logs, and pictorial proof of every step. This is a good thing, but presents a higher demand toward your own documentation and overall OSINT methodology. Without digital evidence, the computer crime or online incident you are investigating never happened. Without proper training and policies, your evidence may never be considered. Without confidence in your work, you may not be taken seriously.

This section enters territory I have always avoided in previous editions. We can no longer ignore discussions about workflow, documentation, and other formalities of our investigations. It is also time that we tackle the ethics surrounding online investigations. These are not easy conversations, and many people will have their own opinions. I do not claim to have all of the answers. I only have my own experiences and lessons learned from many mistakes.

I rely heavily on assistance from my friend and colleague Jason Edison throughout this entire section. Jason is a 20-year veteran of a major U.S. police department where he serves as the investigative lead for the agency's Digital Crimes Unit. He has trained thousands of students in the public and private sectors on various topics related to open source intelligence and cyber-crime investigations. In fact, he is an official IntelTechniques OSINT instructor who travels the world presenting my methods. Most of the content in this section is directly from him. I maintain the first-person usage of "I" throughout the section. It is a collective "I" from both of us.

Jason and I do not agree on everything presented here. This is why you see alternative tools and methods which may contradict each other. This is a good thing. We need as many thoughts and opinions as possible in order to present ideas applicable to many situations. As an example, I try not to use Microsoft or Google products unless absolutely necessary. I have forced myself to use Linux whenever possible, and avoid closed-sourced tools which "call home". Jason prefers Microsoft OneNote, which is extremely robust. His need for an ideal note-taking solution outweighs my paranoia of metadata collection by Microsoft. He prefers Chrome while I insist on Firefox. Neither of us are right or wrong. We simply have strong preferences. We only hope to

present numerous options which may help you choose the best methods for your own investigations. Only you can decide what is most appropriate for your daily workload.

It is now time to take a breath and get back to basics. You may have been overwhelmed with the techniques discussed throughout the previous sections. You may wonder how you will present your findings, create a report, and defend your right to access public information from the internet. This section tackles these issues. Throughout this section, we present numerous document templates and workflow diagrams. All of these are available to you via digital download at the following URL.

https://inteltechniques.com/osintbook

CHAPTER TWENTY-EIGHT
METHODOLOGY & WORKFLOW

An often overlooked component of open source intelligence gathering is the importance of establishing an efficient and repeatable workflow. You need to be thoughtful and deliberate in how you proceed on an investigation rather than wading in haphazardly. As an instructor, one of the most common stumbling blocks with which I see new practitioners struggle with is putting the tools into action in a fashion that results in a professional looking work product. This section provides a step by step walkthrough of the entire investigative process, from receiving an OSINT assignment all the way to submitting a professional case report.

This chapter assumes that you have already completed the steps laid out in the previous sections of this book. You will need familiarity with each of those tools and techniques if you wish to take full advantage of the recommended workflow. If you have an existing investigative process, there will likely be pieces shared here that can be folded into your current procedures. The examples used here were chosen purely for demonstration purposes and not due to any association with ongoing criminal investigations.

Receiving the OSINT Mission

The first step in most investigations is what we in law enforcement refer to as "intake". This is the process of receiving a mission assignment from a supervisor or fielding a request for investigative support from another internal unit or outside agency. For those in the private sector, this might be accepting a contract investigation from a client or conducting a security assessment as part of your normal duties. The following are examples of OSINT requests that we receive on a regular basis:

- **Threat Assessments (Individuals):** Online threats to carry out an act that we wish to prevent. Who is this person? Where are they? What is their capability and true intent?
- **Threat Assessments (Events):** Monitor intelligence prior to and during a significant event that impacts the organization or region of responsibility. Who is involved? What are their intentions? What is the scale of impact on available resources?
- **Target Profiles (Individuals):** Uncover the target's entire online presence, including email addresses, home addresses, friends, hobbies, etc.
- **Target Profiles (Organizations):** Uncover an organization's online footprint and/or entire technological infrastructure. This can be a business, criminal enterprise, or group of individuals organized to pursue a shared goal.
- **Subscriber Identification/Account Attribution:** Identify the real person associated with a domain, IP address, or online account. Who runs a malicious web site? Which child predator has web traffic through this IP address?

The following recommendations can be applied to any of these common investigative scenarios. More than anything else, the key to success is staying organized and having a repeatable process.

Triage

Triage is the practice of assessing a situation or mission to calculate an approach that is likely to result in the best possible outcome. A common mistake that is made when working OSINT investigations is to rush to action with no clear plan or direction. You should take time at the beginning of a mission to ensure that you are following a productive path to relevant answers. Depending on the urgency of the situation, this step could be 30 seconds or 30 minutes. The important thing is to make a plan of attack and move forward with purpose rather than just bush-whacking your way through the internet. Here are some of the key considerations during the triage phase of your investigation.

Be certain of the mission objectives. If you ask a professional analyst to describe the first step they take in any assessment, they will tell you that it is to identify the question. This of course could be multiple questions, but the important thing is that you articulate the investigative goals. This can be a verbal or written confirmation, depending on your situation, but written is preferred should the other party later misremember the conversation.

The first benefit of articulating the questions is establishing a clear set of expectations with the person asking you to do the work. This could be a supervisor, contract client, colleague, or victim of a crime. Do not overthink it. An example could be: "To be clear, you want to know the real name and physical addresses associated with the person in control of the email account of kirby@rockerzandrollerz.com, and we have two hours to accomplish this. Is this correct?"

Include in your verification any specific identifiers (email addresses, names, phone numbers, IP addresses, etc.) that were originally provided by the requestor. It gives them a chance to catch any typos or miscommunications. They may have given you the email address of the victim rather than the suspect. Those types of mix-ups occur frequently and can waste a lot of valuable investigative time and resources if not caught and corrected early on. That quick clarification also defines the primary questions for our investigations, similar to the following.

Find the legal name of the real person associated with kirby@rockerzandrollerz.com.
Find any home and/or work addresses for kirby@rockerzandrollerz.com.

When it comes time to write your investigative report, these questions should be clearly addressed in the summary of key findings. Taking the time to articulate and clarify mission goals up front lays the groundwork for your final work product. You should also ask questions regarding the source of any initial leads or other intelligence on your target. Why do we believe that email address belongs to our suspect? How was that lead obtained and how certain are we that it is correct? **Information is not intelligence until it has context.** You need to ask questions up front to establish any available context for the target. Do we know anything about their location

or culture? Are they into video games or an avid pro-baseball fan? Once you get to the research phase of the investigation, you will have a far easier time locating pages, accounts, and identifiers related to your target if you start learning about his or her day to day life. Never assume that the person tasking you with this work is giving you all available information. Ask questions and be persistent.

Legal Service & Preservation Letters

If you work in support of law enforcement, you should consider if there is likely to be a legal request made to any known social media platforms. For example, if a Gmail address was involved in a crime, you might want to issue a preservation letter to Google requesting that they retain any data related to the specified address. The preservation letter is issued in anticipation of a future subpoena or search warrant for account data, such as subscriber information. If you are unsure of whom to contact at a particular provider in order to submit this request, a good starting point is the ISP list at https://www.search.org/resources/isp-list/. Try to get a live person on the phone rather than just sending an email. Build rapport with the support person or legal contact and shepherd them into doing the right thing based on the urgency of the situation. No one wants to be responsible for a teen suicide or the next school shooting. Often, they will be much more cooperative and productive if they feel invested in the situation.

Deconfliction

Not all investigations involve infiltration into criminal organizations. However, when they do, you may want to check with colleagues in other agencies to make sure you are not stepping on any ongoing investigations. This could also save you time should you locate an investigator who has already laid groundwork into the online communities in question. We always want to be respectful of another professional's work and collaborate whenever operationally appropriate. In the past, I have concluded long-term investigations only to find out later that other teams were running operations in that community at the same time. While reviewing the case, we found that we had wasted time working to gain reputation with users who, unbeknownst to us, were other undercover investigators. This is not only a waste of time and focus, but can complicate the individual cases where they overlap. If nothing else, ask the person for whom you are doing the work if anyone else is working on the case. You would be surprised how often two branches of the same agency are unknowingly pursuing a common target.

Note Taking

The triage stage is the appropriate time to begin your note taking. I will discuss specific tools in the next chapter, but at its core you will need a paper scratch pad and digital notebook, such as Microsoft OneNote. A paper notepad allows for quickly noting key details without having to move out of your browser during searches. This is even more crucial if you are on a laptop or otherwise have limited screen real estate with which to work. Your digital notetaking application is for pasting content as you copy it from your browser. Keep in mind using Microsoft products

does allow them to collect user information, so make sure that this is within the operational security requirements of your organization.

At the top of your legal pad, list out the questions you are trying to answer and any initial investigative steps. This does not need to include extreme details, but it establishes your plan.

Key Questions/Goals

- Find the real name associated with kirby@rockerzandrollerz.com
- Find any home and/or work addresses for kirby@rockerzandrollerz.com

Investigative steps

- Using Chrome - Google search kirby@rockerzandrollerz.com
- Query kirby@rockerzandrollerz.com using custom email tools

In your digital notebook, create a new section and title it logically based on the date of request, incident type, or case number if there is one. For example, "OSINT Request Oct13_2019" or "Robbery 19-486544". Any emails or other written correspondence received leading into the case should be copied into your digital notebook.

Finally, before moving on, ask yourself if OSINT is the right tool for the job. I have made the mistake of investing hours into online searches, only to realize later that a two-minute phone call would have likely given me the same information. Do not make the mistake of overlooking traditional investigative resources. Would a phone call to a postal inspector identify the occupants of a residence quicker than an online search? The strongest investigators are ones who think outside the box while also using every tool in it.

Knoll Your Tools

Now that you have established a plan and a clear understanding of the mission goals, you need to prepare your workspace for the investigation. "Knolling" is the process of organizing your resources so they are ready to go and easily accessible once you start the actual work. Think of how a surgeon's instruments are sanitized and laid out in an organized fashion. The time spent preparing up front will result in a more efficient overall operation while also reducing the chances of unnecessary mishaps.

If you followed the instructions and recommendations in previous chapters, you should already have a custom Linux VM. It should be patched and preloaded with your preferred OSINT applications. Additional recommended preparations prior to the search phase include the following.

- Check the status of your VPN on your host machine. If not already connected, join a node in a geographical area close to where you believe the target is located.
- Start VirtualBox and load your custom OSINT virtual machine.

- If you are using a Windows-based digital notebook, such as OneNote, you will need to switch back to your host environment (Windows) when adding content to that notebook. In the next chapter we will look at a Linux compatible notebook that has some of OneNote's desired functionality.
- Once in your OSINT VM, run your browser by selecting it from your docker bar on the left or from the application window which is accessed by clicking on the square-shaped set of nine dots at the bottom of the docker bar.
- If you have not already done so, log in to your covert social network accounts used for OSINT research. We will likely need to conduct searches on these platforms and pre-authenticating in this browser session will save us time later.
- If you need to make new covert social media accounts for research, you should disconnect from your VPN prior to doing so. It should also be noted that running several queries on freshly made accounts is highly likely to result in security challenges and/or suspended accounts. Try to always have a supply of semi-mature accounts available for time-sensitive investigations.
- Open the custom OSINT tools which you built in previous chapters of this book.
- Create a new folder in the shared directory of your VM and rename it to match your digital notebook including date, target name, or case number. This is where we will store any manually captured digital evidence such as saved images or pdf captures. I keep a directory in my Master VM that is prepopulated with folders titled to reflect my custom tools. This gives me an organized destination for anything I recover and saves me from having to repeat this step every time I open a new case. Figure 28.01 displays an example of this, and the digital templates download contains the same structure which can be copied if desired.
- If you use a digital notebook that accepts the pasting of multi-media filetypes, you have the option of storing files within your digital notes.

Your knolling is complete. You have a virtual machine preloaded with the most useful OSINT tools, and you are on a secure and private connection. We are prepared to search quickly, collect pertinent content, store it logically, and track our progress within our notes.

Closed-Source & Premium Data

You should now be ready with all your tools and note-taking resources. Begin the research phase of the investigation by querying your target against any in-house, premium, or proprietary data sources. This includes any of the following.

- Commercial aggregators such as Accurint (Lexis-Nexis), TLO, Clear, or others.
- Premium products such as BeenVerified, Intelius, Spokeo, Pipl-pro, and WhitepagesPro.
- Government and LE databases such as Department of Licensing, Criminal Records, Department of Corrections, Insurance Bureau, and Agency Records Management Systems.

Whereas using purely open-source tools typically requires visiting dozens of sites in order to find just a few leads, paid services often quickly provide a list of possible addresses, associates, and accounts. If you have services like Lexis-Nexus or Clear available, use them early for easy additional leads on your target. These services obtain much of their data from credit histories and utilities. Therefore, they tend to be good sources for residential address history, land-line phone numbers, employers, roommates, and family members. They tend to work very poorly with email addresses, usernames, and social media accounts.

This is also when you should run any premium people-search services such as Pipl, Spokeo, or BeenVerified. These types of services range from $5-$100 a month depending on the subscription tier, but tend to offer a much richer, immediate return on queries than their free counterparts. Although Pipl Pro formerly offered some of the best premium results, they are also one of the most expensive. Additionally, they have moved to a complex per-record pricing model. Spokeo is one the cheapest at $8-$14 a month depending on your plan, but they have a modest amount of data for a paid service and charge extra to make reports easily printable. BeenVerified allows you to run an unlimited number of fairly extensive reports for $53 quarterly and they are print friendly. Many investigators kickstart the early stages of their open source investigations using one of these cheap premium aggregators, but keep in mind everything you get from paid people search sites is available for free elsewhere (although with considerably more time and effort).

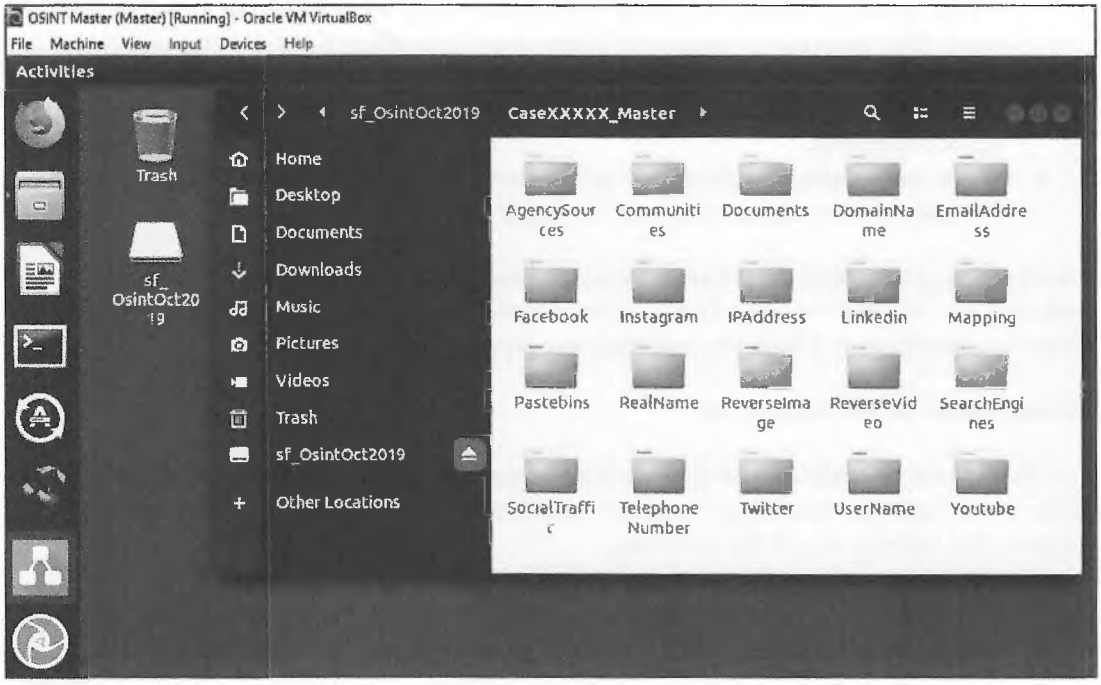


Figure 28.01: A logically structured case directory.

For those on the government or law enforcement side of the house, internal agency records systems, department of licensing requests, and criminal history queries can be very powerful additions to your early digging. An advantage that government investigators have is that many of these systems will provide access to a photo of the target, which can be used to verify or rule out possible social media accounts. These records also typically include recent associates, phone numbers, and residential addresses. Even if the subject did not use their own address during a previous contact with government agents, they likely used one where they can receive mail, such as a relative's house. Most people are not trained to think on their feet and will use familiar data when cornered with hard questions.

Any promising identifiers from your premium or government searches should be added to your notepad, and generated reports should be dropped into your digital notebook as pdfs. Photos can be copied and pasted into your digital notes or dropped into your designated directory within the shared folder on your desktop. This reflects our workflow going forward. Any valuable page, image, or identifier is noted, and a corresponding pdf or image capture is placed either in our digital notebook or investigative folder.

Open-Source Research & Collection

Once you have exhausted your in-house and paid data resources, it is time to dive into your OSINT tools and resources. This tends to be the most involved stage of research due to the large number of sites that you will check for your target's information. Tab management is critical in staying organized. If you have not already done so, add the **OneTab** (one-tab.com) extension to Chrome and Firefox within your VM.

- In your VM, conduct a quick Google search on your known identifiers.
- Open your custom OSINT tools and use the tool category that corresponds to your known identifiers, such as the email and search engine tools.
- Perform any additional queries on sites not included in your custom toolset. For example, a colleague may have very recently recommended a new email search site. If that resource provides good results, consider adding it to your custom toolset.

At this point, you should be in your VM looking at several open tabs in your browser. These tabs represent the results from the Google and custom tools queries which you have executed. The rule going forward is to deal with each tab completely, and intentionally keep or discard it before moving on to the next. A common misstep is to start clicking on leads that look interesting prior to completely reviewing the page on which you are currently visiting. Therefore, tab discipline should be in the forefront of your mind as you parse through your first batch of search results. Consider the following.

- Review the first tab of Google results, looking for anything that stands out as a likely valid lead on your target. For any results that look promising, right-click on the link and choose "Open link in new tab".

- Continue to scroll through the first page of Google results and when you get to the image results, right-click on it and choose "Open link in new tab". If Google does not include an "Images for..." section in the first page of results, you may need to select "Images" from the tabs at the top of the page. The image results are always worth reviewing as you can quickly scan the page for potential profile images or other photos of your target.
- Once you are satisfied that you have fully reviewed the first page of Google results and have opened any promising leads in their own tabs, you can move on to the next tab.
- As you start to do more OSINT work, small efficiencies compound to save a lot of time in the overall investigation. Learning keyboard commands for frequently used browser actions will be very beneficial. In this case, you can use Ctrl-Tab (Windows) or Command-Tab (Mac) to move to the next tab to the right. Holding down "Shift" with the previous key combinations will cycle through tabs in the opposite direction, from right to left.

This system of exhausting leads on the current page before moving on to other tabs is crucial in ensuring that you do not overlook potential intelligence or lose your way by moving too quickly from lead to lead. That is called "rabbit holing" and is the biggest pitfall with which new investigators inevitably struggle. You also need to be disciplined about closing any tabs that are false positives or otherwise present no fruitful results. This will help to control browser clutter and reduce the load on your workstation resources.

As you move through your tabbed results methodically, you may come upon a page of results which is a jackpot of links to potential target data. This is a good problem to have, but a problem nonetheless. The same rules apply, but with one additional recommendation, which is that any lead that warrants its own full set of queries should be opened in a new window rather than a new tab. Consider the following example.

- You have located a social media account containing several strong leads which require their own full set of queries using Google and your custom OSINT tools. This can be seen in Figure 28.02.
- The Twitter handles need to be queried through the Twitter tools and the email addresses through the email tools. Think of each of these as a new path that needs to be followed independently.
- Any strong leads should be represented in your notes. Write down any account identifiers on your legal pad, and for each create a new page in your digital notebook. Figure 28.03 displays the documentation in OneNote.
- Much like a family tree forks into new branches which create even more branches, following online leads often presents new leads. Browser tabs and windows can help you categorize and isolate all leads, providing a sense of structure to your investigation.



Figure 28.02: An example of several account identifiers as strong leads.

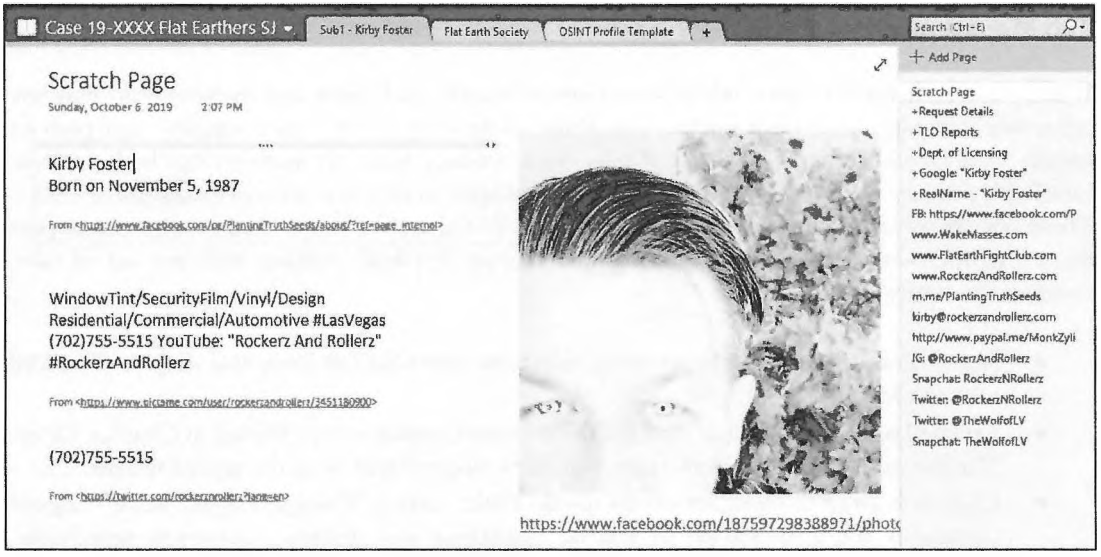


Figure 28.03: Structuring digital notes in OneNote.

This OneNote digital notebook is logically structured to organize intelligence leads as they are uncovered. The notebook title on the top left reflects the case number and name of the target organization. I should mention that this example was chosen arbitrarily, and the group depicted is not likely criminal in nature. I have tabbed sections for the target individual and the organization. I also have a tab which contains fresh copies of my OSINT templates should I need them. The visible section represents a "real-name" investigation into my primary target. On the right, I have added pages that reflect each strong lead, which was created for each account identifier. This ensures that every new investigative path I open has a place to paste relevant data, while also making it easier to visualize my leads as a whole. The following explains some of the options present in Figure 28.03.

- The Scratch Page is for quickly pasting links and reminders for items which I want to have quick access or revisit later.
- The Request Details page is where I paste the details of the investigative request along with any other important details gleaned during triage.
- The various Premium/Government Data resource pages contain pasted reports or snippets from closed-source, in-house, and paid services.

Strong leads are given a new browser window and their own page in my digital notebook. This ensures that every new investigative path I open has a place to paste relevant data while also making it easier to visualize my leads as a whole. The list of pages can also be used as a check or "to-do" list. Once you have fully exhausted that lead, you can add a + symbol to the title to indicate that it is complete, as seen in my example.

Tab Management

When you reach the last open tab in your current search, look back and make certain that any open tabs are pages that have useful data. Prior to moving on to a new window and path of inquiry, you should preserve your list of tabs. This is where your tab manager can be beneficial. OneTab's primary purpose is its ability to quickly collapse all tabs into an exportable list of URLs. These bookmarks can then be pasted into your notes or shared with a colleague who can import them into their own OneNote instance. Once you are finished working with any set of tabs, conduct the following.

- Right-click anywhere in the window, select the blue OneTab icon, and click on "Send all tabs to OneTab".
- You will now be looking at the OneTab manager, which was explained in Chapter Three. The list of bookmarks for the tabs you just collapsed will be at the top of the list.
- Click in front of the number of tabs to add a title, such as "Google Target Name". Logical naming is the cornerstone of staying organized and making your work searchable. Eventually this set of bookmarks will get pushed farther down the page as you add newer

tab sets. To find it again, hit ctrl-f (Windows) or command-f (Mac) to conduct a keyword search for the custom title you added.

The list of tabs is saved locally in your VM within your browser extension data, but you will want it in your notes for easy access. Click on "Export/Import URLs" on the top right of the page. Figure 28.04 displays an example. The export page is missing titles, but each set is separated by a space and the URLs are in the same order as the lists on the OneTab management page. Consider the following steps.

- Left-click and drag your mouse to highlight your set of URLs. Press ctrl-c (Windows) or command-c (Mac) to copy the list.
- Move to your digital notebook, select the appropriate page, and press ctrl-v (Windows) or command-v (Mac) to paste the list into your notes. Figure 28.05 displays this content within OneNote.

Although OneTab is the tab manager I recommend for most people, if you require online sync or advanced features, some other tab extensions of note are Toby, Tabs-Outliner, Workona, and Graphitabs. As discussed earlier in this book, extensions always come with a security cost, so use them sparingly and only when the added functionality is mission critical.

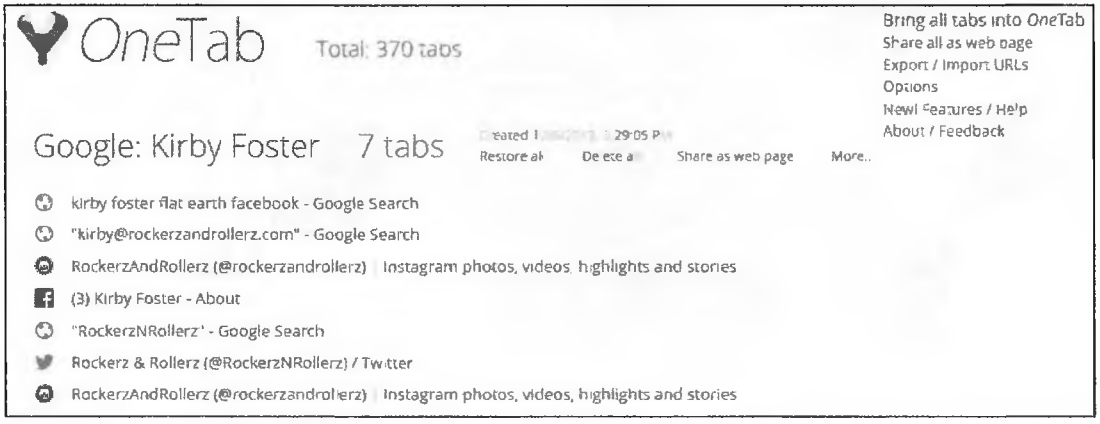


Figure 28.04: The OneTab management page.

Context Menu Queries

Speed and efficiency are key if you are conducting OSINT at the professional level. Context search extensions, such as **ContextSearch** (github.com/ssborbis/ContextSearch-web-ext), allow you to query a keyword or link just by right-clicking on it. These extensions come with a predefined set of search options, such as Google reverse image, and allow you to also add your own custom queries. Figure 28.06 displays ContextSearch in action.

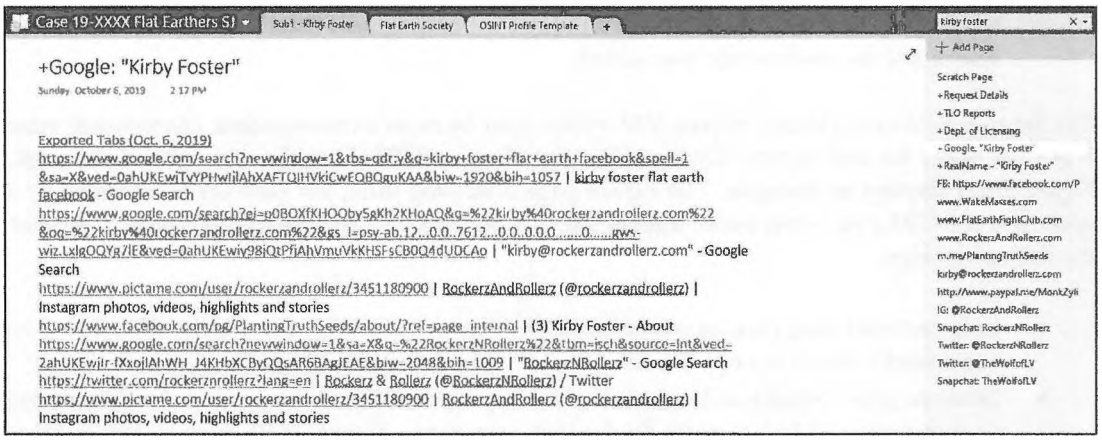


Figure 28.05: An OSINT case template in OneNote.



Figure 28.06: The ContextSearch Extension available for Chrome & Firefox.

Queries are customized via the options menu by adding structured URLs to the existing list of popular search engines. To add your own custom search strings, conduct the following.

- Left-click on the ContextSearch extension in your toolbar and then click on the gear icon to go to settings.
- Select the Search Engines tab and click on the "Add" button.

- Type in a name for your new search and click on "OK".
- In the template field, paste the URL query for the search you want to execute. These can be some of the same queries that you have added to your custom OSINT toolset. At the end of the URL add "{SEARCHTERMS}". Figure 28.07 displays a custom search option for Dehashed.com.

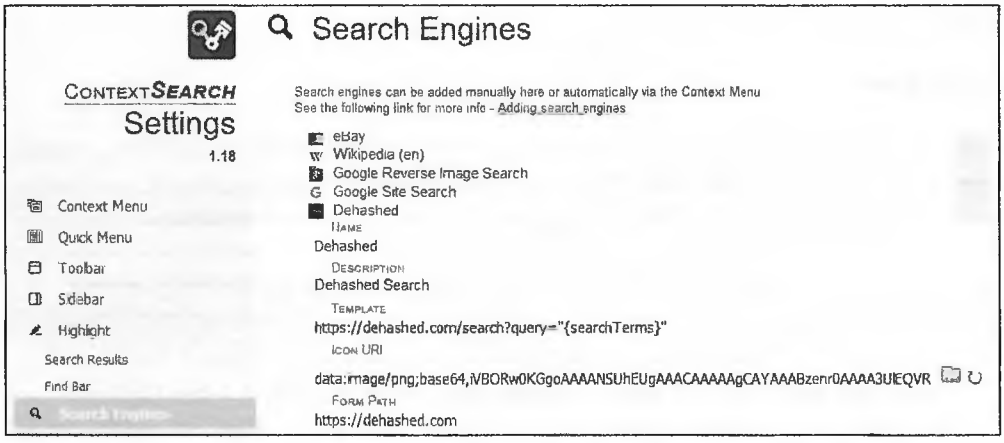


Figure 28.07: A custom ContextSearch dialogue.

If I highlight a phrase on a page, such as "tom jones" and right click on that phrase, I now have a Dehashed option in my ContextSearch menu. Clicking on that Dehashed search will provide the results seen in Figure 28.08.



Figure 28.08: ContextSearch Dehashed.com search results.

Context-based search capabilities complement, but do not replace your custom toolset. They offer speed and convenience, but lack the full level of control and customizability that you have with your own tools.

Capture and Collection

You should now have a reliable, repeatable process for working through your OSINT queries and capturing results in your paper and digital notes. The final piece of the research phase is capture and collection. This is made up of the steps involved in preserving content for evidentiary or reporting purposes. There are three approaches to collection depending on the tools you have available.

Manual Capture

Manual capture includes any technique that is user-initiated and results in capturing a single file or multiple files within a single page. These are often browser-based tools such as javascript or extensions. Here are the steps to integrating manual capture into your workflow.

- Create a case folder in your VM shared directory named logically for your case.
- Open that case folder and create a new folder matching the corresponding category from your toolset. If I am working on an email address this will be "Email Addresses".
- Open that directory and create a folder titled appropriately for the identifier you are querying. If an email address is the target, the folder may be titled similar to "kirby@rockerzandrollerz.com".
- Repeat this for any other strong leads such as Twitter handles, real names, domains, etc.
- Now you have a logically structured set of folders to store any saved digital content.
- As you work through your set of tabs specifically for that lead, capture any pages that support your findings and save those to this folder.
- Any time you save an image or video related to that lead, also save it to this directory using any of the tools referenced earlier in the book.
- Any time you save a specific image or video, you should also save a screen capture of the page from where you obtained it. This shows visually how that image or video was laid out on the page. The capture of the entire page is saved to the same folder where you placed the image or video it references.

Wash, rinse, repeat. If you followed the prior steps for tab management, these subfolders should match up with any strong leads that you have opened in their own windows. At the conclusion of your investigation, the digital evidence is nicely organized by its corresponding identifier and toolsets. Figure 28.09 displays an example of this type of documentation.

Passive Capture

The best example of a passive capture tool is Hunchly. It records pages loaded in Chrome at the source code level, as well as any images on those pages. It is providing a wide safety net, but you should be more intentional in taking advantage of its capture capabilities. The following steps assume that you have Hunchly and Chrome installed in your custom OSINT VM. If you are not

a Hunchly user, you may skip this section and move on to scripted capture. Figure 28.10 displays a Hunchly tagging option.

- Create a new Hunchly case named the same as your investigative notebook and your digital evidence directory.
- Click on the Hunchly extension icon on the top right of your Chrome browser and make sure that it is set to capture and that it is set to the correct case name.
- Proceed with your research in Chrome as described in the previous sections. Any time you find an image that is key to your case, right-click on it, select the Hunchly entry on the context menu, and choose "Tag Image". Provide a logical caption and hit "Save".

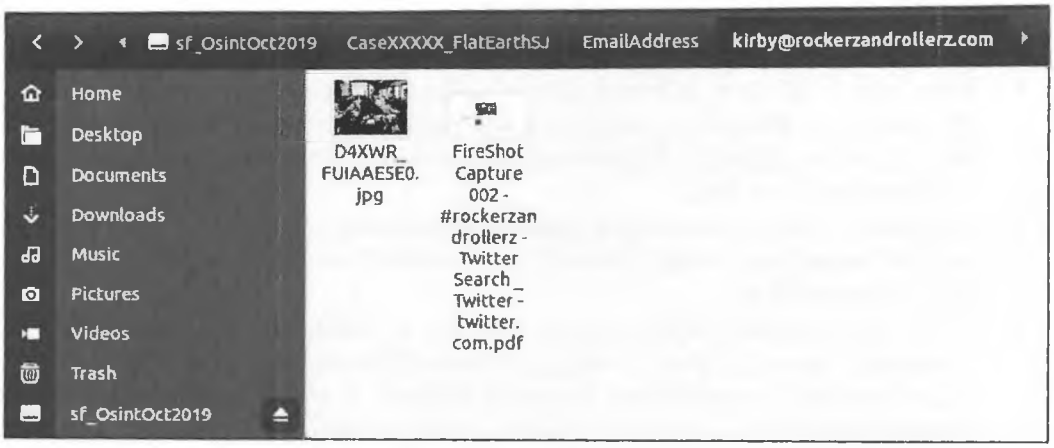


Figure 28.09: An example of a properly structured case folder.

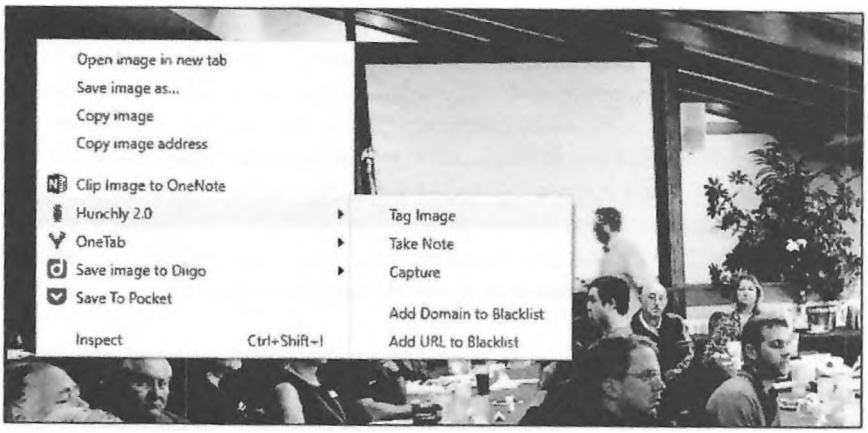


Figure 28.10: An example of Hunchly image tagging.

When it comes time to prepare the appendix of our case report, Hunchly will generate a forensically sound report containing all tagged images. We will learn more about reporting in the next chapter.

Scripted Capture

Scripted capture is made up of the manually activated programs that collect or "mine" digital content on our behalf. A good example of this is using Instaloader to rip all of the photos from a specified Instagram account. These types of tools were covered earlier in the book, and there are only a few things to keep in mind on how they fit in our workflow, as explained below.

- For scripts that prompt you for a save location, you should use the same system that was described for manual capture: a series of logical nested folders.
- Some scripts will store collected data in a default directory, such as Downloads or Documents. In these cases, complete the collection and then manually move the files over to your case directory. When reasonable, move rather than copy files to limit clutter and abandoned case data.
- Add a line to your notes indicating which script was used, what it was directed to collect, and the date and time. Unlike Hunchly, most of these tools do not annotate or generate a log of their actions.
- If you are collecting digital evidence for use in court, you should consider also conducting a manual capture of any crucial items. The problem with scripts is that you may not be able to explain how they work in court. A simple manual save is easy to explain confidently to a jury when the time comes to testify.

Analysis

Whether you are working your case independently or have the support of a dedicated team, the research phase will include some level of multimedia analysis. This is the process of examining the visual and metadata characteristics of recovered images and video. Visual examination is exactly what it sounds like. View each image or clip at the highest resolution possible and methodically examine media for any intelligence that was unintentionally included in the frame. You are looking for things like business signage in the background of your target's profile photo. Identify anything that narrows down who or where they might be, and include this in your case notes. This process can be very time consuming, but it remains one of the best methods of locating an elusive target who has otherwise covered their online tracks.

The primary goals of link analysis is to understand how information is connected, and a way to visually represent these connections. These can be people, locations, websites, phone numbers, or any other identifiers you see associated with online accounts. Figure 28.11 displays a link analysis showing how an email address was linked to a social media profile. You will see specific examples of link analysis tools in the following chapter. Not all cases require a link chart, but you

should consider its value when faced with complex organizations or anytime your case might benefit from a visualization of how entities or accounts are connected.

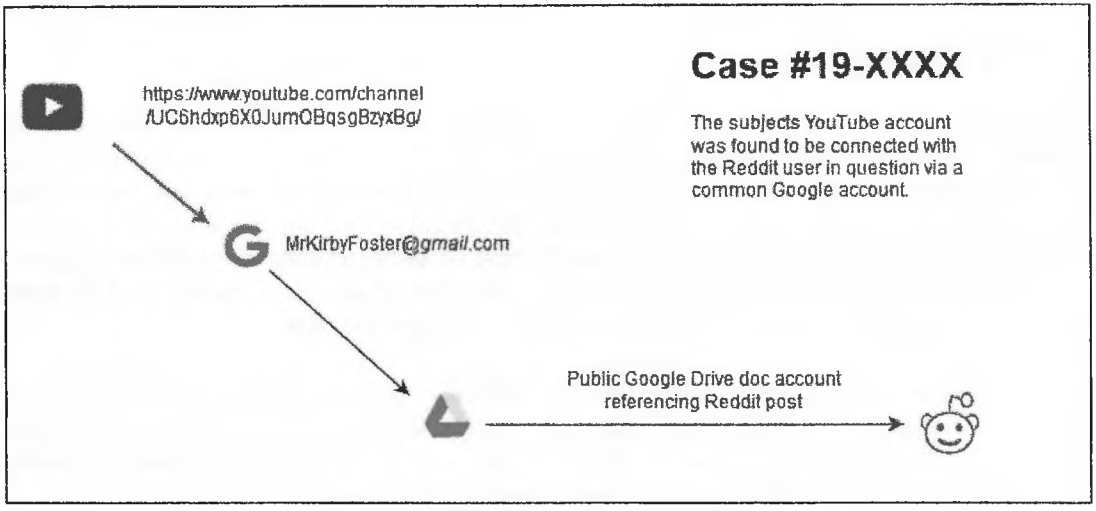


Figure 28.11: A link analysis example with Draw.io.

Once your research is complete, you will need to prepare your report. Several of the steps in this workflow were in preparation for the reporting phase. If you have followed along diligently, your efforts will be rewarded by painless report creation. Chapter 29 is dedicated to taking all of the intelligence that you have collected and using it to build a professional case report.

Submission and Cleanup

It is not unusual to move from one investigation to another very quickly. Just as we took time to properly set up our research, we also need to appropriately close out our work before moving on to the next task. The following may be beneficial.

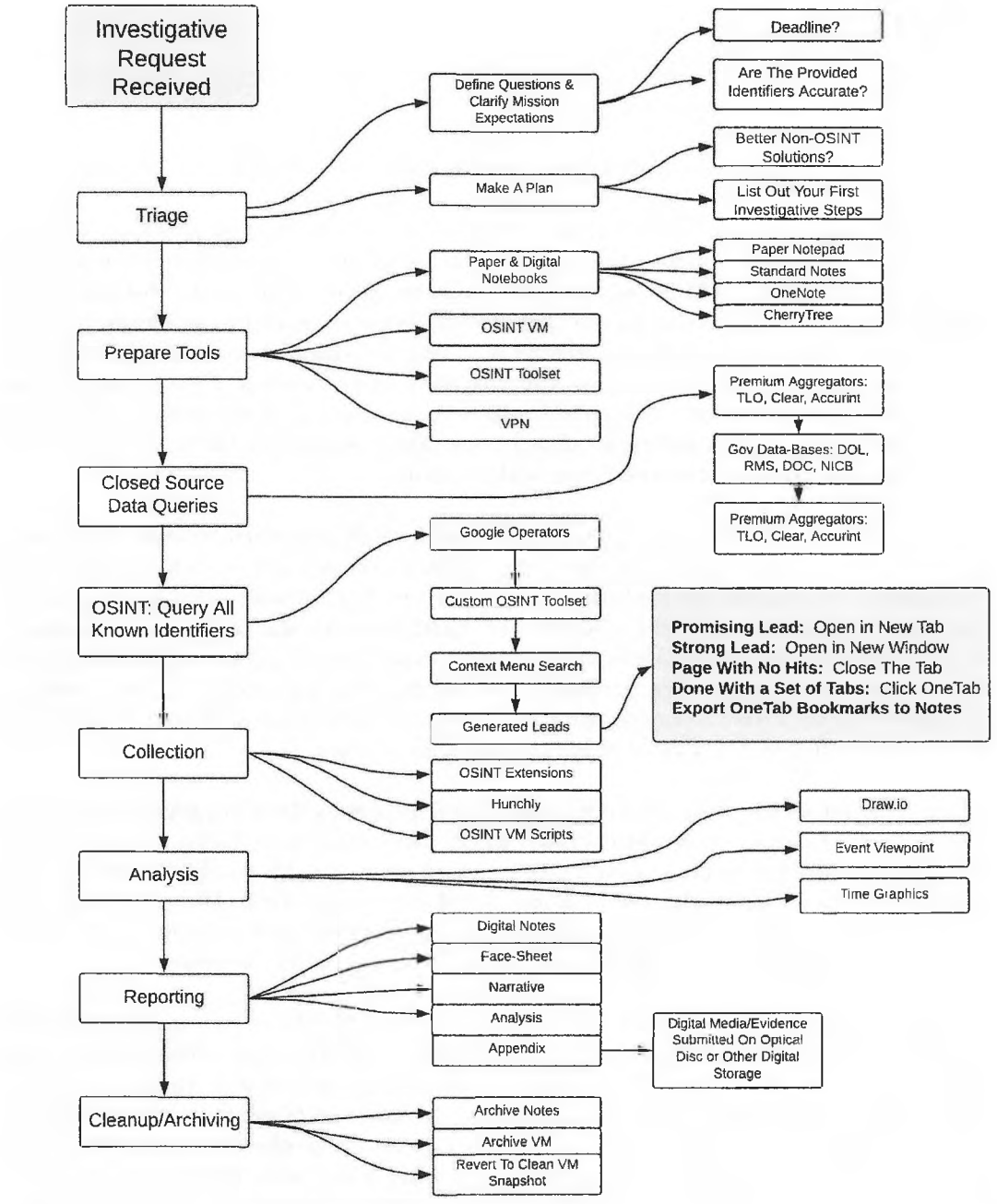
- Transfer any handwritten notes to either your digital notes or final report. If you prefer, you can scan the notes as a pdf using your scan enabled printer or scanning straight to OneNote. Any paper notes or printouts are then either filed in a secure location and in compliance with your agency's policies or they are shredded.
- Do not leave case assets scattered about or they will get mixed in with future case work. Your investigative directories should be archived in accordance with your agency's evidence submission and retention policies. Some teams retain copies of these "working files" on network attached storage or optical discs for 3-6 months. If the subject should resurface, as criminals tend to do, having historical notes from previous incidents can be a huge time saver.

- If appropriate, export a copy of the VM that you used for the case. Then return to a clean snapshot or clone as described in earlier sections of this book. Consider preparing fresh accounts for the next investigation and find replacements for broken tools.

20 min vs 20 days

Some major case investigations take place over a series of days or even months, while critical incidents may require you to give a threat assessment in 20-30 minutes. Your workflow will remain the same for each situation, but the time spent on each step will obviously be reduced. Any time you reduce the "time-to-solve" drastically, there will be compromises made to the quality of work and security. A common scenario where we use a streamlined workflow is a threat assessment, such as a person threatening suicide in an online chatroom. Consider the following threat assessment steps and Case Workflow chart on the following page.

- **Triage:** Verbally clarify the known identifiers and expected intelligence, such as: "User D1ckTraC" on 4chan is threatening to kill himself in a post". We want to know who he really is, where he lives, and whether he is likely to carry out the threat. Identify if there is a non-OSINT solution such as contacting a human resource.
- **Knoll your tools:** Grab your legal pad and a pen. Ideally you will have a fresh instance of your OSINT VM ready to use. If you do not have this prepared ahead of time, use "utility" social media accounts. Utility social media accounts are those on hand for non-criminal assessments where speed is essential and cross-contamination is a reasonable concession. Using fresh accounts would be preferable, but that just isn't always possible.
- **Collaboration:** If you are collaborating with a team on a platform such as OneNote, create a page for each user to paste key findings so that you don't confuse one another. Keep in mind that OneNote in a browser syncs almost instantly, whereas several users on OneNote desktop will have syncing issues. Assign one person to keep track of everyone's progress and build out the big picture.
- **Premium/Government Resources:** Run your target through any commercial aggregators and government databases. These checks should go very quickly and return low hanging fruit.
- **OSINT:** Begin with Google searches such as: site:4chan.org "username". Then query your target's known identifiers through your custom OSINT tools.
- Only open very promising links in new tabs and visually scan each page quickly for words or images that jump out at you. The images results can be especially useful on time sensitive assessments because your brain can process them exponentially faster than text.
- For anything useful, make a note on your legal pad and leave the corresponding tab open.
- Be prepared to give a briefing at the deadline, even if it is just a situational report similar to "we've located and preserved the original posting, there's a history of similar suicidal threats from that user, but we do not know who or where he/she is yet".
- Take care of any additional collection, analysis, and reporting once the crisis has passed. At that point you will fall back into the normal workflow and documentation steps.



IntelTechniques.com OSNT Workflow Chart: Case Workflow

Target Flowcharts

I have conducted numerous OSINT training programs over the past few years. Regardless of the audience, I receive one question at every event.

"Is there a standard process or workflow for each type of OSINT target?"

My short answer was always "no". I had always looked at each investigation as unique. The type of investigation dictated the avenues and routes that would lead me to valuable intelligence. There was no cheat-sheet that could be used for every scenario. While I still believe that there is no complete template-based solution for this type of work, I now admit that some standards can be developed. This section will display my attempt at creating workflows that can quickly assist with direction and guidance when you possess a specific piece of information. These documents are presented in six views based on the information being searched. Each example should be considered when you are researching the chosen topic. The categories are Email, Username, Real Name, Telephone Number, Domain Name, and Location.

Each example will try to show the standard path that I would take when provided the chosen type of data, such as an email address. The goal with my investigations is to get to the next topic. For example, if I am given an email address, my goal is to find any usernames and real names. When I have a username, my goal is to find any social networks and verify an email address. When I have a real name, the goal is to find email addresses, usernames, and a telephone number. When I have a telephone number, my goal is to verify the name and identify a physical address and relatives. When I have a domain name, my goal is to locate a real name and address. The cycle continues after each new piece of information is discovered.

Each example will identify only the services used. It will not display the actual address to navigate to the website. However, every method listed within these charts is explained throughout this book. These documents do not contain every avenue that may provide good information. They only display the most beneficial resources at the time of this writing. Think of them as the obvious steps to take when you receive a target to search. The results of your queries can lead you to more places than can display on a single page in this book. These are just the first priorities.

Consider the Email flowchart presented in two pages from now. The written translation of this would be to take the email address and search it within the TruMail email validation tool. Next, conduct searches of the address within quotation marks on the main search engines. After that, check the compromised databases and all options on the IntelTechniques Custom Email Search Tool. These options are likely to lead to a presence on social networks. Following these to the bottom of the chart encourages you to conduct the email assumptions mentioned previously, which you can verify and start over with the newly acquired information. You would then continue through the remainder of the chart. If you find the following information beneficial, you are welcome to download digital copies at inteltechniques.com/osintbook/workflow.zip. I also recommend visiting osintframework.com. While it is not a traditional workflow, it does

provide numerous online resources within an interactive tree. Many of the websites mentioned here are replicated on osintframework.com, which was created by Justin Nordine.

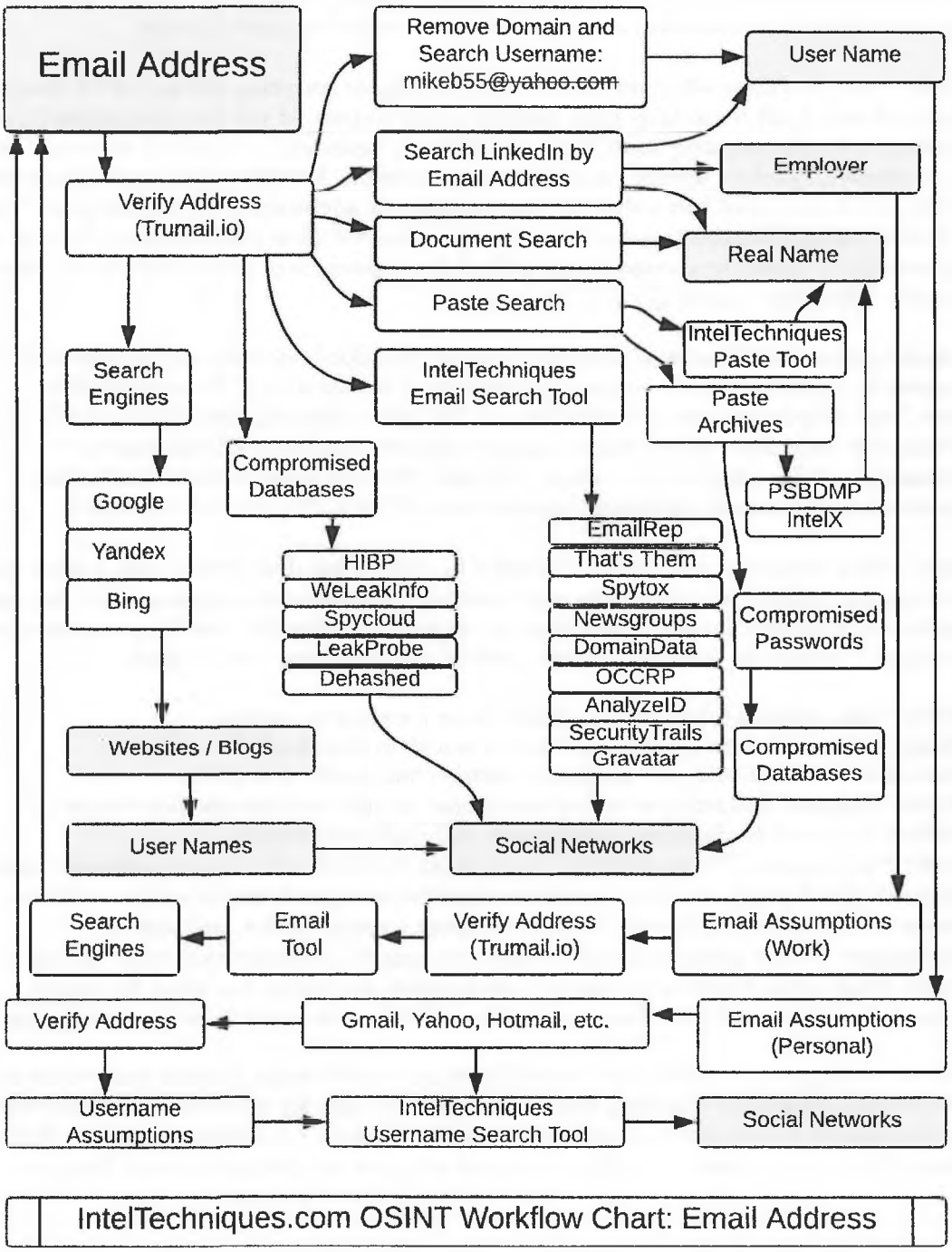
I believe that all of these will always be a work in progress. As everything else in OSINT changes, these will too. I will try to keep them updated on the website. If you have suggestions, I am honored to receive and apply them. If you would like to create better formulas, I encourage you to get creative. I used the website LucidChart.com to create each of these. I also made all of these public within the LucidChart website so that you can take advantage of my starting point. The following links will connect you to a live environment that will allow you to replicate the work in seconds. If you would like a similar service without the requirement of a registered account, please consider **MindMup** (mindmup.com).

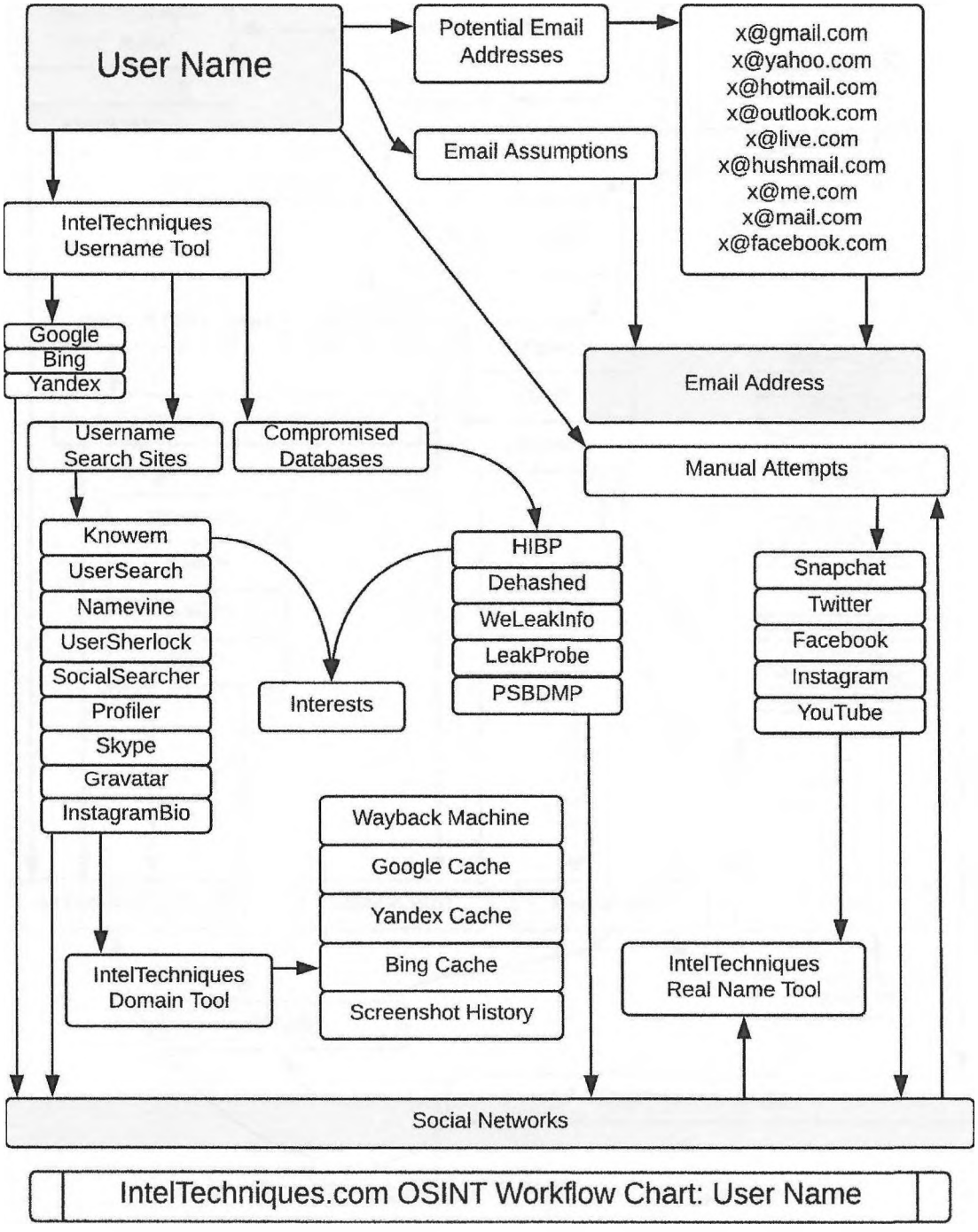
Email Address: lucidchart.com/invitations/accept/5282ad5a-b0dc-4442-a4a5-4a440a00dd05
Username: lucidchart.com/invitations/accept/5282ad70-58dc-4546-8758-0a460a00c875
Real Name: lucidchart.com/invitations/accept/5282ad8b-c4d0-4db3-98f2-25d00a00c875
Telephone: lucidchart.com/invitations/accept/5282ad9a-64a4-4435-9073-3ce80a00c875
Domain Name: lucidchart.com/invitations/accept/5282acc9-f324-43b2-af40-04c00a00c875
Location: lucidchart.com/invitations/accept/9d446294-580e-49ba-a88f-2437cc392b6f

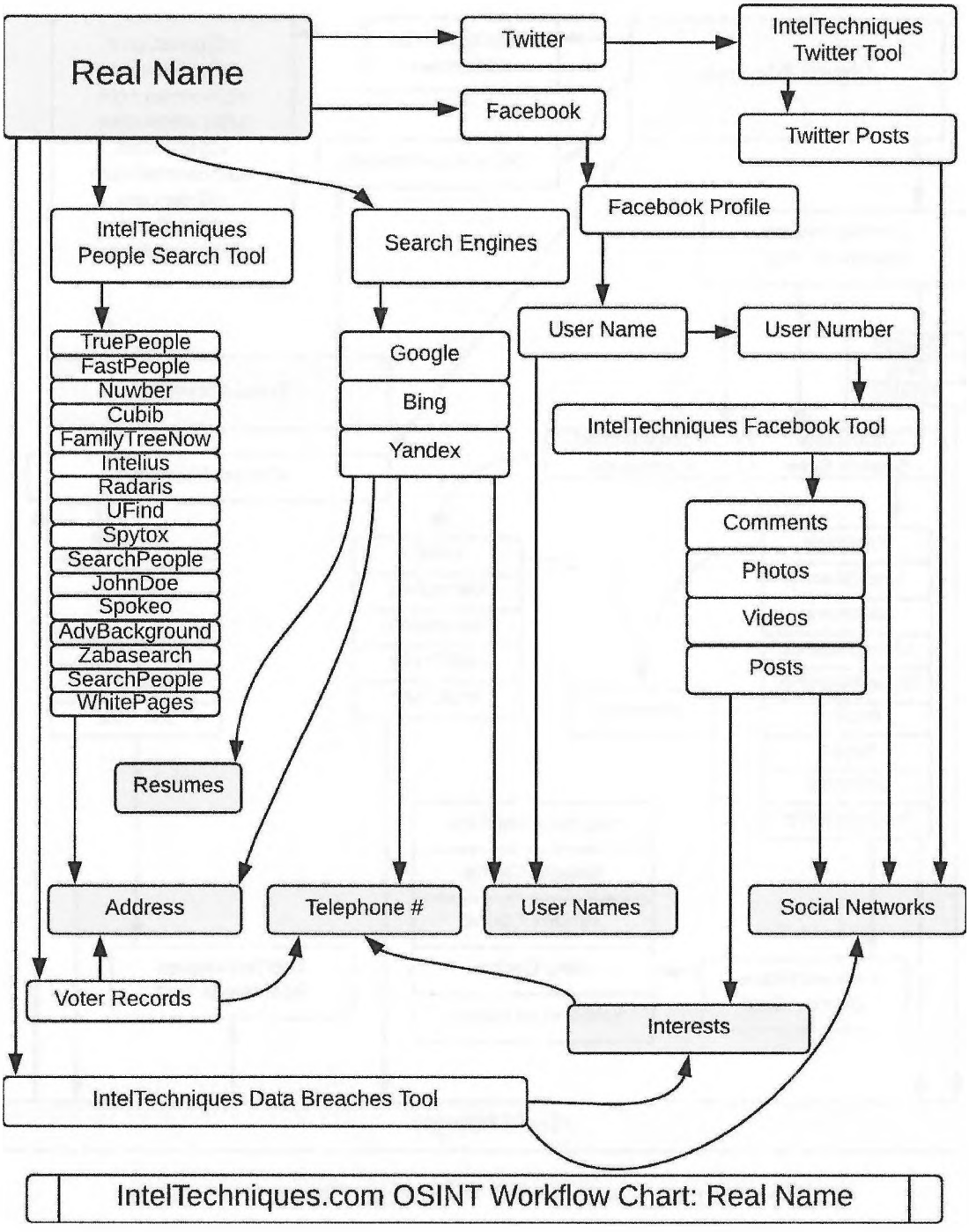
Many readers have requested practical exercises in order to test their OSINT skill. I agree that this would be helpful, but maintaining active and accurate online demonstrations with live data can be overwhelming. Instead, I encourage you to test your skills with real data, unknowing to the target. Consider the following scenarios, and use the flowcharts here as a guide.

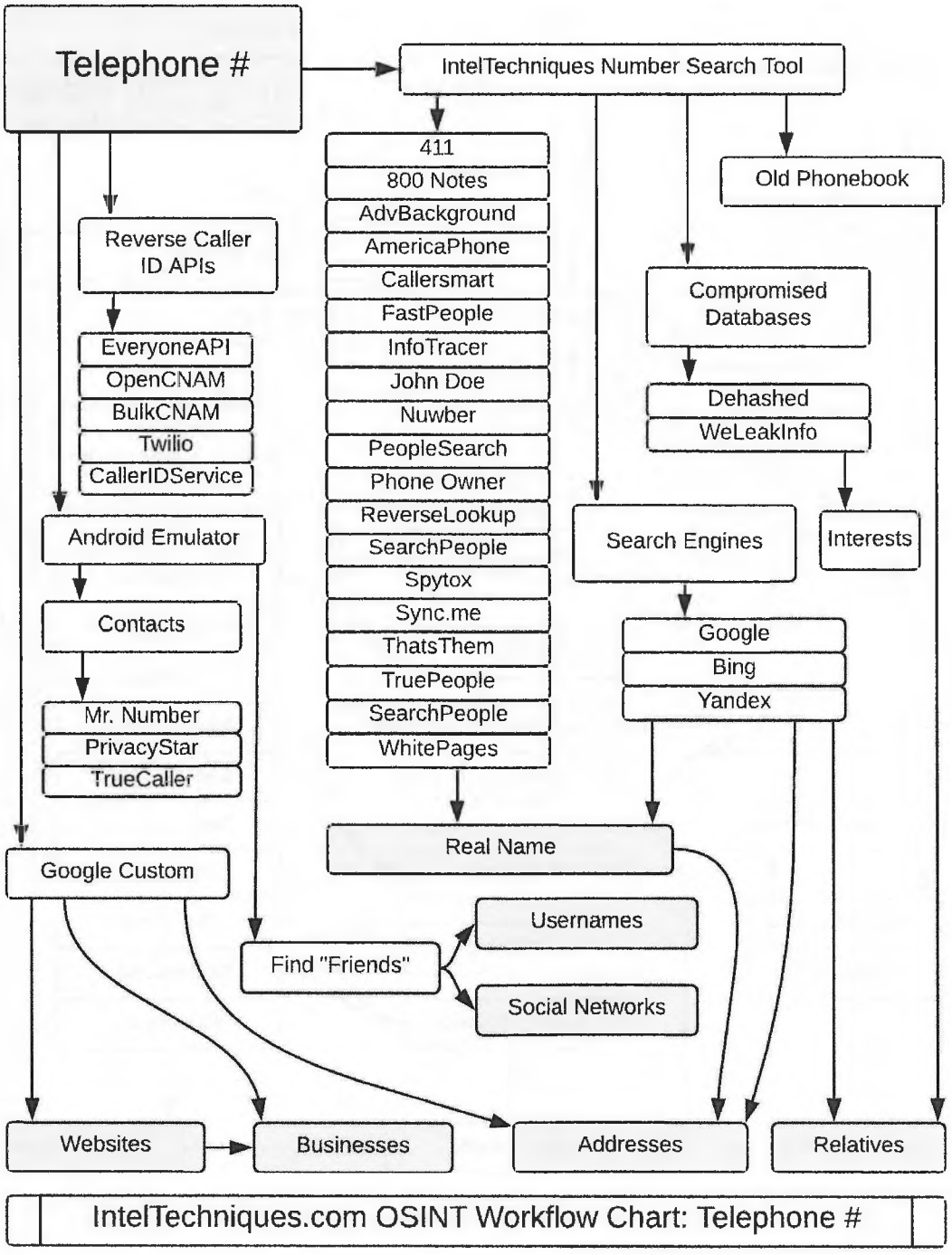
Zillow: Pick a random home and find all info about the previous owners.
Wrong Number (incoming): Reverse-search it, text them their details.
Wanted Criminals: Locate any significant others' online profiles with photos.
Waiter/Waitress: Research your server from dinner last night and identify their vehicle.
AirBnB: Locate all details about a host (owner) and email them directly.
Radio: Pick a morning "Happy Birthday" target, obtain full DOB and relatives' comments online.
Reviews: Find 5 people that have patronized a local business and locate their home addresses.
Game Show Contestant: Identify full address, phone number, photos, and relatives.
Newspaper: Choose a person quoted in today's newspaper and identify their social networks.
News: When a local Facebook comment is cited, explore the hidden data about the person.
Library: Locate an employee's Amazon wish list and buy them the book he or she wants (creepy).

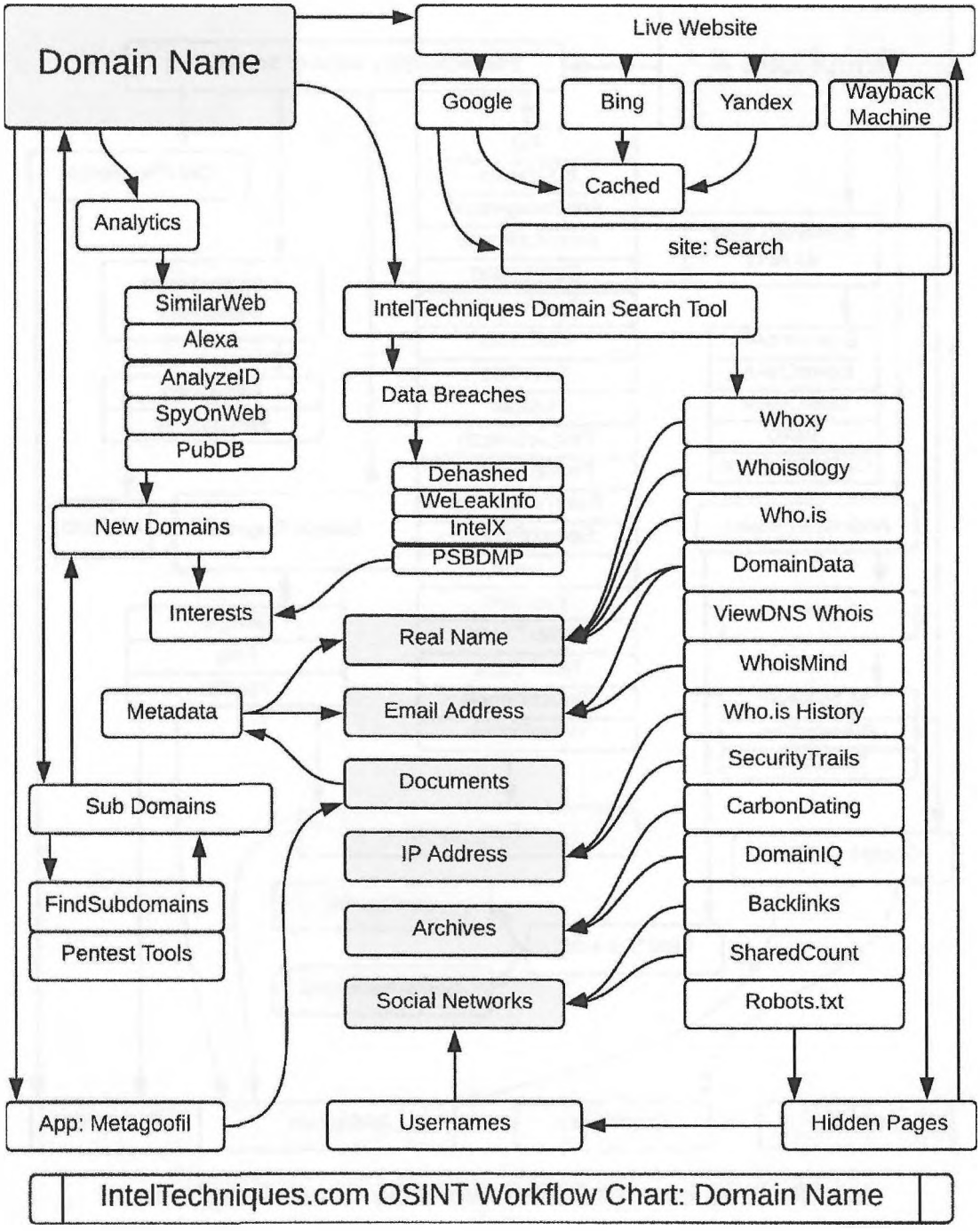
This list could grow for many pages. Overall, there are endless targets available that provide the best practice possible for exercising these techniques. This practice will increase the confidence in your research during an actual investigation. The hard part is not disclosing what you find to them. While you may think they will be impressed with your new skills, they won't. Trust me...

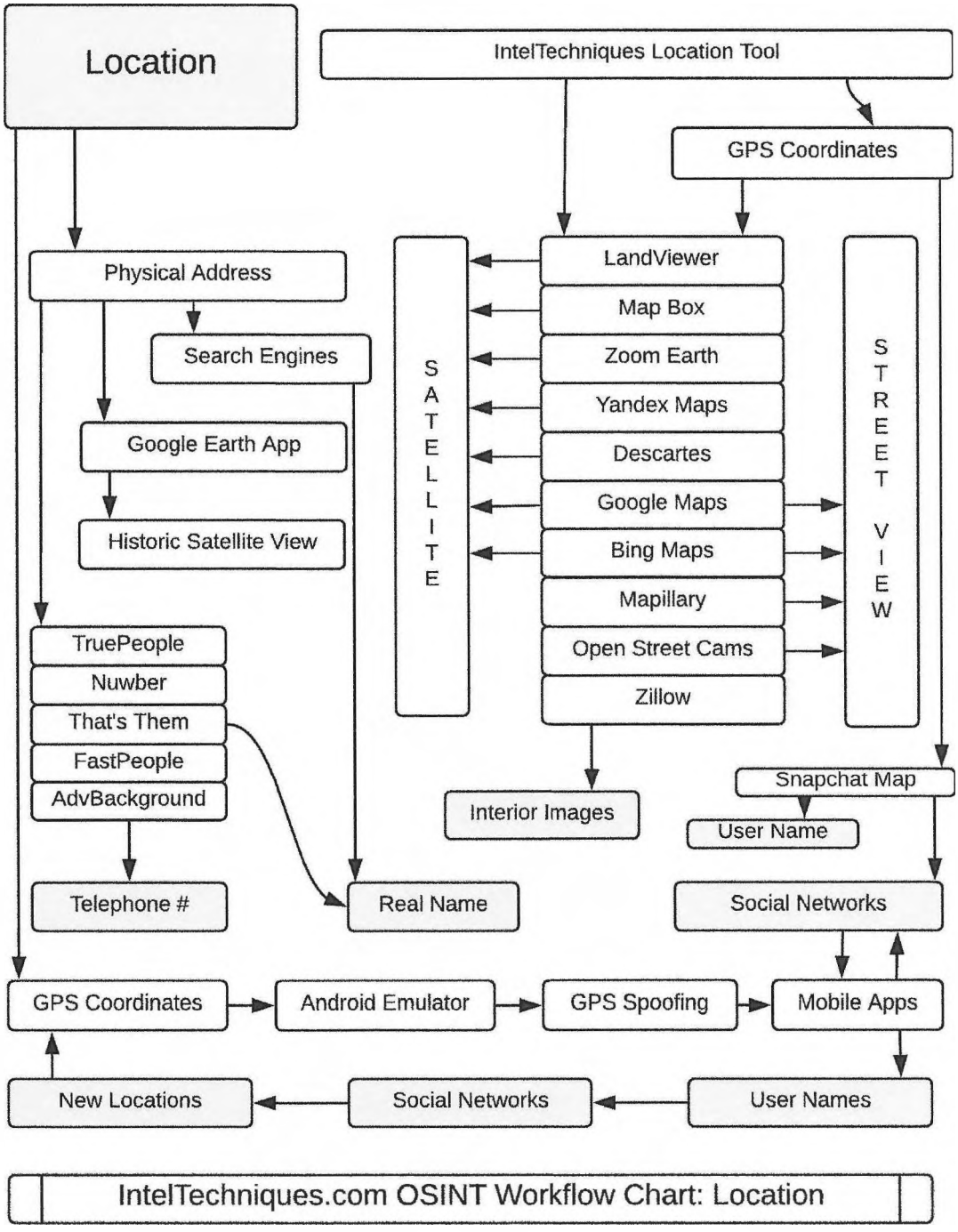

Email Address
Remove Domain and Search Username: mikeb55@yahoo.com
User Name
Search LinkedIn by Email Address
Employer
Verify Address (Trumail.io)
Document Search
Real Name
Paste Search
IntelTechniques Paste Tool
Search Engines
IntelTechniques Email Search Tool
Paste Archives
Compromised Databases
PSBDMP
IntelX
Google
Yandex
Bing
EmailRep
That's Them
Spytox
Newsgroups
DomainData
OCCRP
AnalyzeID
SecurityTrails
Gravatar
HIBP
WeLeakInfo
Spycloud
LeakProbe
Dehashed
Compromised Passwords
Compromised Databases
Websites / Blogs
User Names
Social Networks
Search Engines
Email Tool
Verify Address (Trumail.io)
Email Assumptions (Work)
Verify Address
Gmail, Yahoo, Hotmail, etc.
Email Assumptions (Personal)
Username Assumptions
IntelTechniques Username Search Tool
Social Networks
IntelTechniques.com OSINT Workflow Chart: Email Address


User Name
Potential Email Addresses
x@gmail.com
x@yahoo.com
x@hotmail.com
x@outlook.com
x@live.com
x@hushmail.com
x@me.com
x@mail.com
x@facebook.com
Email Assumptions
IntelTechniques Username Tool
Google
Bing
Yandex
Email Address
Username Search Sites
Compromised Databases
Manual Attempts
Knowem
UserSearch
Namevine
UserSherlock
SocialSearcher
Profiler
Skype
Gravatar
InstagramBio
HIBP
Dehashed
WeLeakInfo
LeakProbe
PSBDMP
Snapchat
Twitter
Facebook
Instagram
YouTube
Interests
Wayback Machine
Google Cache
Yandex Cache
Bing Cache
Screenshot History
IntelTechniques Domain Tool
IntelTechniques Real Name Tool
Social Networks
IntelTechniques.com OSINT Workflow Chart: User Name


Real Name
Twitter
IntelTechniques Twitter Tool
Facebook
Twitter Posts
IntelTechniques People Search Tool
Search Engines
Facebook Profile
TruePeople
FastPeople
Nuwber
Cubib
FamilyTreeNow
Intelius
Radaris
UFind
Spytox
SearchPeople
JohnDoe
Spokeo
AdvBackground
Zabasearch
SearchPeople
WhitePages
Google
Bing
Yandex
User Name
User Number
IntelTechniques Facebook Tool
Comments
Photos
Videos
Posts
Resumes
Address
Telephone #
User Names
Social Networks
Voter Records
Interests
IntelTechniques Data Breaches Tool
IntelTechniques.com OSINT Workflow Chart: Real Name


Telephone #
IntelTechniques Number Search Tool
411
800 Notes
AdvBackground
AmericaPhone
Callersmart
FastPeople
InfoTracer
John Doe
Nuwber
PeopleSearch
Phone Owner
ReverseLookup
SearchPeople
Spytox
Sync.me
ThatsThem
TruePeople
SearchPeople
WhitePages
Old Phonebook
Reverse Caller ID APIs
EveryoneAPI
OpenCNAM
BulkCNAM
Twilio
CallerIDService
Compromised Databases
Dehashed
WeLeakInfo
Android Emulator
Search Engines
Interests
Contacts
Google
Bing
Yandex
Mr. Number
PrivacyStar
TrueCaller
Real Name
Google Custom
Usernames
Find "Friends"
Social Networks
Websites
Businesses
Addresses
Relatives
IntelTechniques.com OSINT Workflow Chart: Telephone #


Domain Name
Live Website
Google
Bing
Yandex
Wayback Machine
Cached
site: Search
Analytics
SimilarWeb
Alexa
AnalyzeID
SpyOnWeb
PubDB
IntelTechniques Domain Search Tool
Data Breaches
Dehashed
WeLeakInfo
IntelX
PSBDMP
New Domains
Interests
Whoxy
Whoisology
Who.is
DomainData
ViewDNS Whois
WhoisMind
Who.is History
SecurityTrails
CarbonDating
DomainIQ
Backlinks
SharedCount
Robots.txt
Real Name
Email Address
Metadata
Documents
Sub Domains
IP Address
Archives
Social Networks
FindSubdomains
Pentest Tools
App: Metagoofil
Usernames
Hidden Pages
IntelTechniques.com OSINT Workflow Chart: Domain Name


Location
IntelTechniques Location Tool
GPS Coordinates
Physical Address
Search Engines
Google Earth App
Historic Satellite View
TruePeople
Nuwber
That's Them
FastPeople
AdvBackground
Telephone #
Real Name
SATELLITE
LandViewer
Map Box
Zoom Earth
Yandex Maps
Descartes
Google Maps
Bing Maps
Mapillary
Open Street Cams
Zillow
STREET VIEW
Interior Images
Snapchat Map
User Name
Social Networks
GPS Coordinates
Android Emulator
GPS Spoofing
Mobile Apps
New Locations
Social Networks
User Names
IntelTechniques.com OSINT Workflow Chart: Location

Chapter Twenty-Nine
Documentation & Reporting

Once you have completed your research, you will need to compile your intelligence into a professional report. Using the correct tools and techniques for documentation throughout your investigation will make generating your final report a relatively painless process.

Investigative Notetaking

Like any form of research, note-taking is essential when conducting a professional OSINT investigation. As mentioned in the previous chapter, I prefer to take my notes in both paper and digital formats. The paper scratchpad is for quick identifiers, reminders, and at times a rough diagram. The digital notes are for pasting copied text, images, video, or other files. My hand-written notes tend to be unstructured, whereas collecting digital content often benefits from a pre-structured destination. The following are my recommendations for solid digital notebook applications for OSINT work, along with some of their respective pros and cons.

Standard Notes (standardnotes.org)

Standard Notes is the best privacy focused notetaking application with AES-256 encryption and has a very clear privacy policy (https://standardnotes.org/privacy). It has open-source versions for Mac, Windows, IOS, Android, Linux, and web browsers making it a great fit for OSINT work. You can use Standard Notes completely offline, but if you choose to make an account even the free tier supports sync and end-to-end encryption. The premium version adds advanced features such as multi-factor authentication, automated backups to your own cloud service of choice, and some aesthetic options, such as themes.

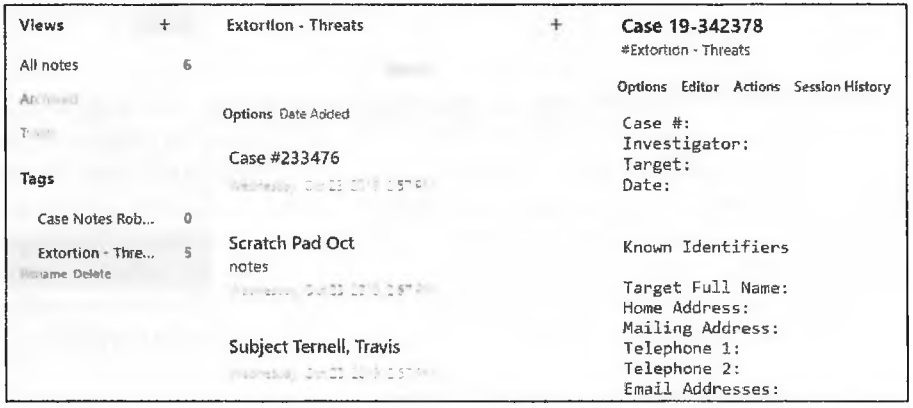


Figure 29.01: A Standard Notes application.

OneNote (onenote.com/download)

OneNote is the power option for digital notetaking at the price of a privacy compromise, but it is worth mentioning because many of us work for organizations that use Microsoft Office as their enterprise office platform. I recommend installing the standalone version of OneNote unless your agency requires use of the cloud-based Office 365 ecosystem. The default version is 32bit. If you want the 64bit build, look for "Other download options" at the bottom of the page listed above. Creating a fresh Microsoft account will prevent the installation from cross contaminating your other MS Office projects. Microsoft will offer to provide you with a new Outlook.com address, but a Protonmail address would be the better choice for most people.

The desktop version of OneNote is supported on Mac, IOS, Android, and most major browsers, but each will require authentication with the account you used for installation. Microsoft, like other major platforms, forces you to sync to their ecosystem whenever possible. I have had little luck getting even the "offline" versions of their Office products to load without an internet connection. What makes OneNote unique is its ability to organize notes in a hierarchical structure, similar to pages within sections within notebooks. From an organizational perspective, it functions exactly like an old-school case binder of paper reports and printouts. I can create a new notebook titled "CopperThieves" with tabbed sections for each suspect. Each section is then broken down into my desired pages: Bio, Facebook, Twitter, Vehicles, Employment, etc. These pages can be populated from scratch, or I can preload them with my case templates or a simple spreadsheet.

Another feature that makes OneNote beneficial for data collection is its ability to accept just about any filetype as insertions or pastes into the pages. I can drag a pdf report from the Department of Licensing and drop it on the "Vehicles" page of a target and OneNote will ask if I want it added as a printout or file. The printout option adds a human readable image of the file, while the file option embeds the pdf file itself. I like to add both image and file by dragging in the pdf twice. That gives me a visual representation of the pdf contents while also saving a copy of the report in its native format. Double clicking on that pdf file after import would open it in a browser window or pdf reader of choice.

I keep a master section prepopulated with my templates. For each new case, I right-click on the section tab at the top of the window and select "Move or Copy". I select the destination notebook and click on "Copy". If I select the same notebook that my master is in it will paste the copy into a new tab, but add a number to the section name, such as "Template 1". Now I can double-click on that tab and rename it appropriately. I can also export entire notebooks by clicking on "File" and then "Export". Notebooks can be exported in a proprietary format, whereas sections and pages may also be exported as pdfs, docs, and several other file formats. This is convenient should you want to export sections from your digital notes to include in your report appendix.

CherryTree (giuspen.com/cherrytree)

CherryTree is my note-taking application of choice for Linux. It ticks all the boxes we like such as being open-source and offline, but what separates it from other Linux options is its ability to support both hierarchical notes and some limited support for storing images, tables, and other filetypes. The following steps will install CherryTree to your Linux Master VM.

- Open a terminal window by clicking on the shortcut in your docker.
- Type the following into terminal: `sudo apt-get install cherrytree`
- Press enter and you will see CherryTree and all dependencies installed.

Open CherryTree from the Applications menu and add to your favorites, if desired. CherryTree has vast functionality, but the core feature of which we want to take advantage of is the hierarchical node structure. Think of Nodes as notebook sections and SubNodes as pages within those sections. On the left side of the CherryTree window is the "node tree"", as seen in Figure 29.02, which shows everything in your current notebook. To add a new section, click on "Tree" in the top toolbar and then "Add Node". Give your node a logical name such as "Facebook" and select "OK". To add a page or "SubNode" to that section, right-click on the node and select "Add SubNode". Name it appropriately and click on "OK". Most of the functions we care about can be completed via the "Tree" menu in the toolbar or by using the right-click context menu.

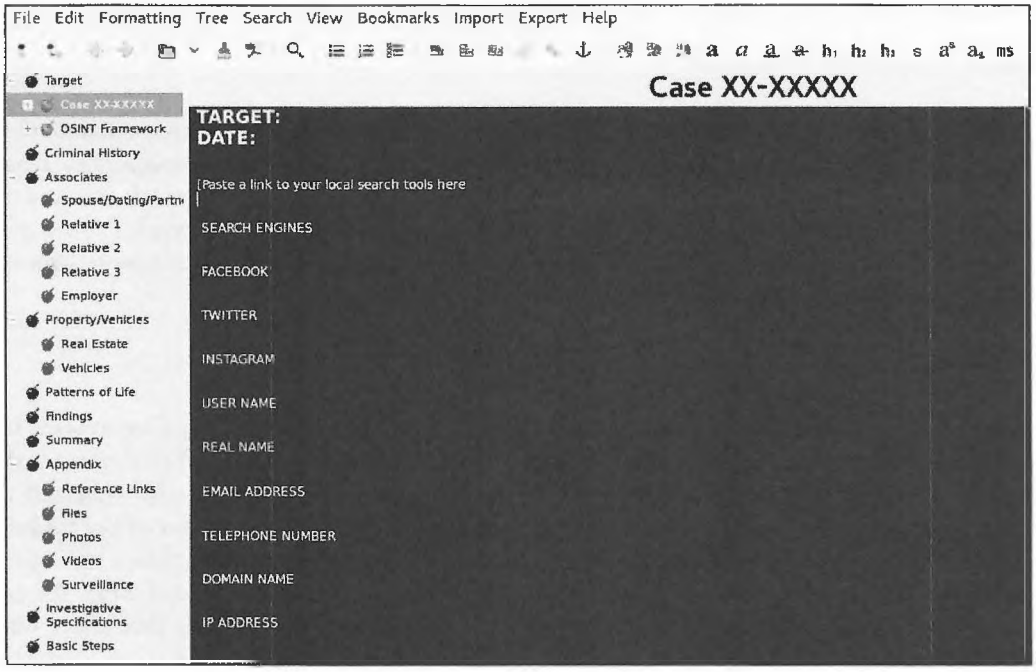


Figure 29.02: A CherryTree OSINT Template.

You can create your own templates with the export feature. Once you have a notebook structure complete, click on "Export" and then "Export To CherryTree Document". In the next window, choose "All the Tree" and click "OK". For storage type, leave the top option selected and click "OK". Type in a name for your template, such as "OSINT_TemplateSept19" and click on "Save". If launching a new investigation on a clean VM, the following steps will add your template within the application.

- Download or copy your template file to the desktop of your VM. The template file will have a file extension of .ctb.
- Within CherryTree, click on "File" and "Open File".
- Browse to your desktop, select your template, and click on "Open".
- Click on "File" then "Save As".
- Leave the storage type as "SQLite, Not Protected (.ctb)" and click on "OK".
- Type in your case name/number as the new filename and click on "Save".

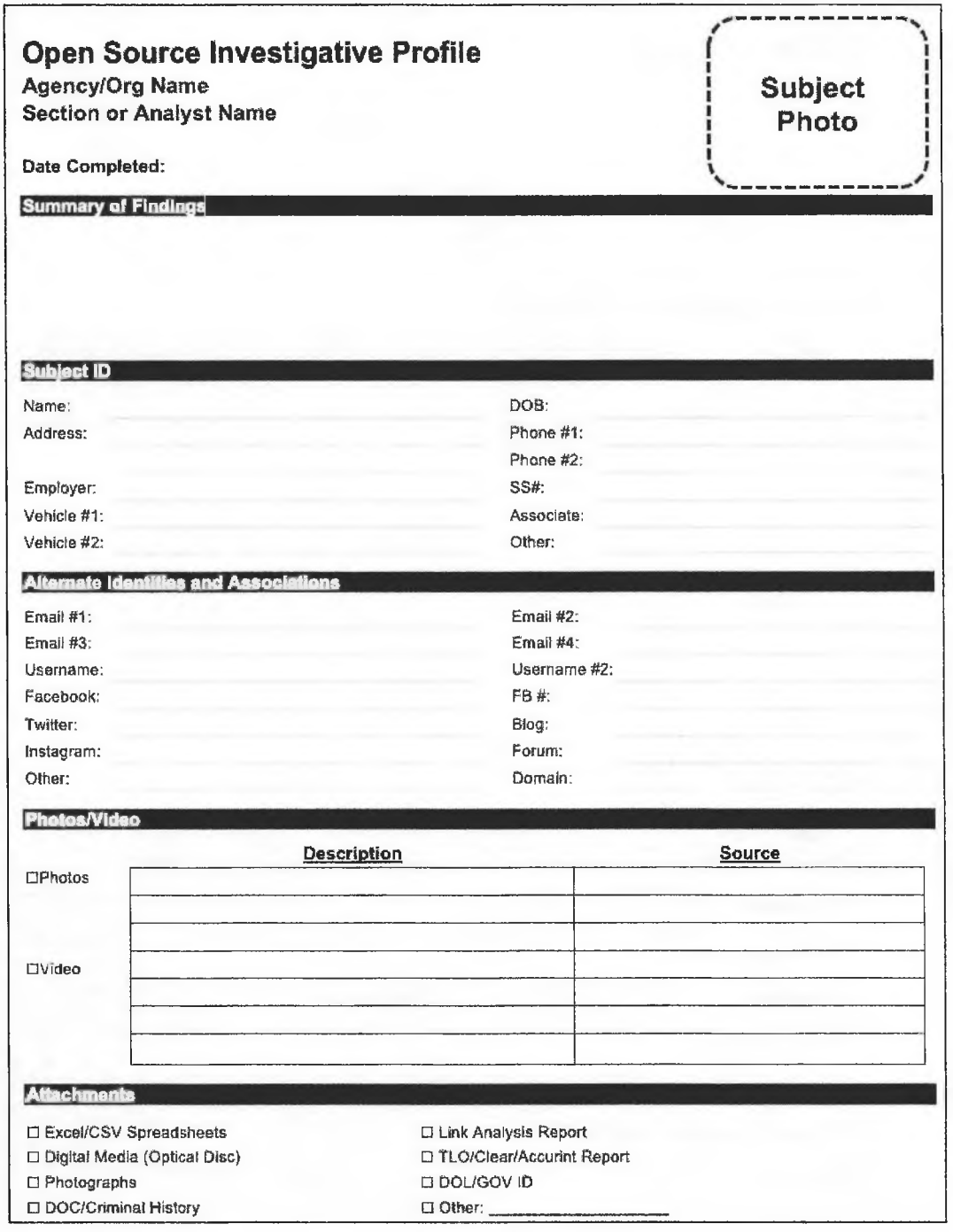
CherryTree is not without its downsides. It is not nearly as user-friendly as OneNote when it comes to drag-n-drop and its support for inserting non-text filetypes is inconsistent. However, it is a hierarchical digital notebook within our Linux VM and one with much greater privacy than we could ever obtain using a Microsoft or other closed-source office application. There is a windows installer available as well, although I recommend using the Linux version in your VM whenever possible. A CherryTree template is included in the digital files download page.

Anatomy of an OSINT Report

Although some investigators choose to write their reports as they go, I prefer to use handwritten and digital notes to complete the formal report at the conclusion of the investigation. I have created a series of templates in Microsoft Word and Adobe Acrobat, each of which are available in the digital files download, which provide a polished documentation framework for the most common mission scenarios. Although they vary in format, each contains a common structure including face-sheet, narrative, analysis, and appendix.

Face-Sheet

The face-sheet is typically one or two pages, and is the most important piece of any report. It is meant to quickly convey the most important intelligence. The mark of a good face-sheet is that a boss or client can glance at it and instantly know the case name/number, who collected the intelligence, investigative timeframe, primary subject identifiers, and a concise set of key findings. The key findings can be in bullet or paragraph format, but should be no more than a half of one page. If you choose to include a table of contents for the report, it is often placed under the case title and prior to key findings. Figure 29.03 displays an example of a simple face-sheet while Figure 29.04 displays a full report version.



Open Source Investigative Profile

Agency/Org Name
Section or Analyst Name

Date Completed:

Subject Photo

Summary of Findings

Subject ID

Name:	DOB:
Address:	Phone #1:
	Phone #2:
Employer:	SS#:
Vehicle #1:	Associate:
Vehicle #2:	Other:

Alternate Identities and Associations

Email #1:	Email #2:
Email #3:	Email #4:
Username:	Username #2:
Facebook:	FB #:
Twitter:	Blog:
Instagram:	Forum:
Other:	Domain:

Photos/Video

	Description	Source
☐Photos		
☐Video		

Attachments

☐ Excel/CSV Spreadsheets	☐ Link Analysis Report
☐ Digital Media (Optical Disc)	☐ TLO/Clear/Accurint Report
☐ Photographs	☐ DOL/GOV ID
☐ DOC/Criminal History	☐ Other: _____

Figure 29.03: A simple face-sheet.

Case #2020-XXXXXX

Subject Name

R/G/DOB
Phone Number(s):
Last Known Address:

AGENCY LOGO

Requested By:
Unit:
Date:

Report Prepared By:
Approved By:
Date:
Date:

Investigative Summary

Clear and concise synopsis of the case findings. Information critical to understanding can be included as bullet points or short paragraphs. Detailed evidence will be included in subsequent sections of the report.

[Intelligence detailed in this report was collected from publicly available sources on the internet and in compliance with agency policy as well as local and federal law.]

[Finding 1]

[Finding 2]

[Finding 3]

[Finding 4]

[Finding 5]

Subject Profile

Add target details such as personal identifiers and account names/numbers. Retitle and/or delete cells as needed.

Target Full Name:	**Age:**	**DOB:**
Home Address:		
Mailing Address:		
Telephone:	**Telephone:**	
Target Email:		
Target Usernames:		
Target Social Network Profiles		
Facebook:	**Twitter:**	

Figure 29.04: A full report face-sheet.

If you are limited on time or supporting a field operation, such as fugitive apprehension, the face-sheet may be the only form of documentation required. It is not unusual to receive "profile" requests while other investigators are purely looking for assistance in identifying accounts, addresses, phone numbers, and associates. This also makes for a convenient quick reference sheet during future cases or incidents involving the same subjects. Events also get intelligence sheets which concisely address need-to-know details such as locations, hashtags, pertinent groups, and operational contacts. It should have everything you would want to effectively monitor public social media posts and live video streams. Figure 29.05 displays an event intelligence sheet.

Narrative

The Narrative follows the face-sheet and it tells the story. This is your opportunity to describe the path you took to reach your key findings. I like to organize my narrative in either chronological order or match the order of findings as listed on the face-sheet. Write in the first person and as concisely as possible, just as you would for any written statement for a police report or other legal document. Write just enough to provide context and a clear understanding for each piece of discovered intelligence. This section can be a couple of paragraphs to several pages, depending on the complexity of the case.

The narrative should address any discoveries that impact the investigative findings. It is important to acknowledge any exculpatory evidence and ensure that it is represented in your reporting. This is especially true for those working in the criminal justice and government sectors. Likewise, if your report takes the form of an intelligence briefing, your audience may expect confidence levels associated with your conclusions. Use terms such as "unsubstantiated", "likely", or "highly likely", rather than expressing the chances in percentages.

Consider adding a section describing the best-practices used during the online investigation. Key points to be made here are compartmentalization from other casework, thoroughness of research, software and services used, and a brief statement of training and qualifications. Think of this section as demonstrating why your work should be trusted. Figure 29.06 displays an example of a report narrative.

LOGO HERE

Event Name
Section or Analyst Name

Threat Assessment - Event

Event/Assessment Details

Event ID

Event Site:	Date(s):
Location:	Official #:
	Official @
Security Contact:	Contact #:
Incident Command:	Contact #:

Hashtag, Users, Sites

Hashtag:	Site:
Hashtag:	Site:
@User:	Site:
@User:	Site:
@User:	Site:
@User:	Site:

Groups/Individuals of Interest

Description	Details

Intel Feeds

Video Feeds	Maps

Figure 29.05: An event intelligence sheet.

Case Narrative

Intelligence detailed in this report was collected from publicly available sources on the internet and in compliance with agency policy as well as local and federal law.

Training & Qualifications – Jason Edison is a 20-year veteran of the Springfield Police Department where he serves as the investigative lead in the departments Digital Crimes Section. He has completed the departments advanced online investigations and digital forensics training. He has conducted hundreds of internet-based investigations and served as an expert witness regarding digital evidence.

On 11/14/2019 Detective Johansen with the Homicide Unit requested my assistance in identifying and locating a possible witness to a shooting death that occurred at 4200 N Jackson St. on November 12th, 2019. Detective Johansen provided me with a tip sheet wherein an anonymous caller purported that the Twitter handle "@Jakijumpjorp66" had posted photos of the shooting as it took place. No further information was provided by the anonymous complainant.

I researched user "@Jakijumpjorp66" using a fresh Chrome browser within a newly configured virtual machine. These best practices ensure that the online research is free from cross contamination with other casework. I conducted keyword searches of the username against the site Twitter.com using Google, Bing, Yandex, Baidu, Sear, Yahoo, Duckduckgo, and Exalead. Google returned a result that showed a photo that was clearly from the intersection in question. I browsed to the corresponding page on twitter (https://twitter.com/Jakijumpjorp66/media) and preserved a copy of the page using the Fireshot extension in my Chrome browser (see appendix item 3.46). The photo depicted a man clearly firing a handgun in front of a signage for "Tom's Waterbed Warehouse." I saved a digital copy of the photo at the highest resolution available and placed it in the digital media archive which is included in the optical media attached to this report.

Figure 29.06: A report narrative.

Link Analysis & Timelines

Not all reports will contain an analysis section, but this is where I present graphical components that support understanding of the preceding narrative. These include timelines of key events and link charts mapping out relationships between individuals, groups, locations, and internet sites. In my organization, this type of specialized work is often handled by civilian analysts who are proficient in software such as Maltego or I2 Analyst's Notebook. I provide them with a copy of my face-sheet, narrative, and a hand drawn map of the case entities and corresponding relationships. They use these to construct a more professional, visually appealing, and understandable graphical representation of the criminal organization or series of events. Not every investigation warrants this level of investment and not every investigator has these resources available, but they are a nice touch on major cases. When I do not have analyst resources available, I leverage one of the following free and user-friendly link visualization tools.

Draw.io (about.draw.io/integrations/#integrations_offline)
Visual Investigative Scenarios (vis.occrp.org/account/metro)
Gephi (gephi.org)
Paterva CaseFile (paterva.com/buy/maltego-clients/casefile.php)
Visual Site Mapper (visualsitemapper.com)
MindMup (mindmup.com)
NodeXL (nodexlgraphgallery.org/Pages/Registration.aspx)
Investigative Link Analysis (lampyre.io)
NWU (knightlab.northwestern.edu/projects)

I formerly used the free, stripped-down version of Maltego called CaseFile for my link charts. I have since moved on to better open-source options that are far less resource intensive and less of a privacy concern. Remember, when working in our VM, we are borrowing resources from our host machines, so we need to use lightweight applications whenever possible. I prefer Draw.io for most investigations. It is most commonly used online as a browser-based diagram solution. When you browse to the site it will prompt you to create a new diagram, but first click on "Change Storage". Select "Device" and check "Remember this setting" to establish that we will be saving our work locally rather than with Google or Microsoft. Once you select either a blank diagram or one of the many templates, you will be ready to start building a link chart representing your investigative diagram. Before moving on, consider a more private option by installing the offline desktop application in your Linux VM. The following steps should be conducted in your OSINT Master VM before launching a cloned investigation machine.

- Navigate to https://github.com/jgraph/drawio-desktop/releases.
- Click the Linux "deb" link and download the installation file.
- Right-click on the downloaded file and choose "Open with Software Install".
- Click the "Install" button and provide your password.
- Open the application from the Applications menu and add to favorites, if desired.

Once the application is running, choose to create a new diagram, and a window will open offering you several templates. Some of the options in the network category work well for OSINT, or you can start with a blank project. Draw.io is very intuitive, allowing you to select icons on the left and drag them into your chart. Any linear icons (lines or arrows) are connectors and can be used to represent relationships between entities on the diagram. Double-clicking on text will allow you to edit the labels. The toolbar at the top and the right-click context menu offer many additional options for editing or adding components to the link chart. Save and export options are in the "File" menu. Draw.io supports export in the most common filetypes such as pdf, docx, png, and html. When you first install Draw.io the default settings will present charts which appear a bit outdated due to the art style of the standard icons. Changing the following settings can fix this.

- Click on "Extras" at the top and select one of the additional themes. I use the lighter options (Figure 29.07) for anything being printed, but I find that dark works well visually (Figure 29.08).
- At the bottom of the "Shapes" panel on the left, click on "+More Shapes...".
- Browse through and check any sets that look useful and click on "Apply". One I always include is "Web Icons" under "Other". The "Web Icons" and "Web Logos" work very well for OSINT charts and the icon styles tend to be more modern than the default selections included in the offered templates.
- There is also a + button in your toolbar that will allow you to import your own images into the chart or even draw shapes freehand.

One small quirk that I have noticed when working with Draw.io in my VM is that when I choose to save my project the pop-up window is hidden behind the chart window. The easiest way to bring it to the front is to click on the "Activities" button at the top left of the VM window which will show you all open windows. Click on the file section window and it will pop back on top of all other open applications. The Activities button is a great way to find lost windows if you have several applications or terminal windows open at once.

Timelines and Event Maps

Draw.io is very flexible and you could use it to create timelines or event maps. However, I like to have multiple options for any task. The following descriptions are of two other applications for dedicated timeline and mapping tools.

Event Viewpoint (eventviewpoint.com)

Event Viewpoint is a free, browser-based timeline and mapping application. It will allow you to create events made up of a location, designated span of time/date, and event notes. You may add images and view your case as a list, timeline, or geographical map. An example is seen in Figure 29.09. You will need to sign up for a free account using a non-attributable email address. Event Viewpoint is not open source and does collect some user data so make sure to use a clean browser and VPN. I never use these types of browser-based applications to work with sensitive data.

Time Graphics (time.graphics)

Time Graphics is a premium browser-based timeline tool and you will need to make an account using a valid email address. Only premium users can save privatized projects, but you can select a two-day free trial. The interface is driven by an intuitive right-click context menu and will allow you to add events, pictures, video, and any notes. Figure 29.10 shows an example. You can export your project in several formats including pdf, docx, and json.

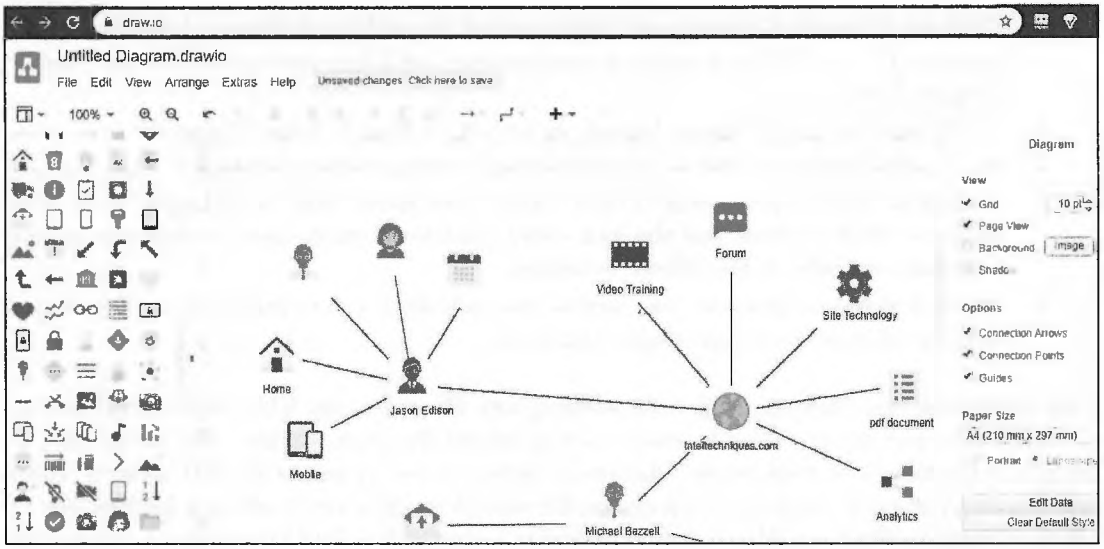


Figure 29.07: The Draw.io light theme.

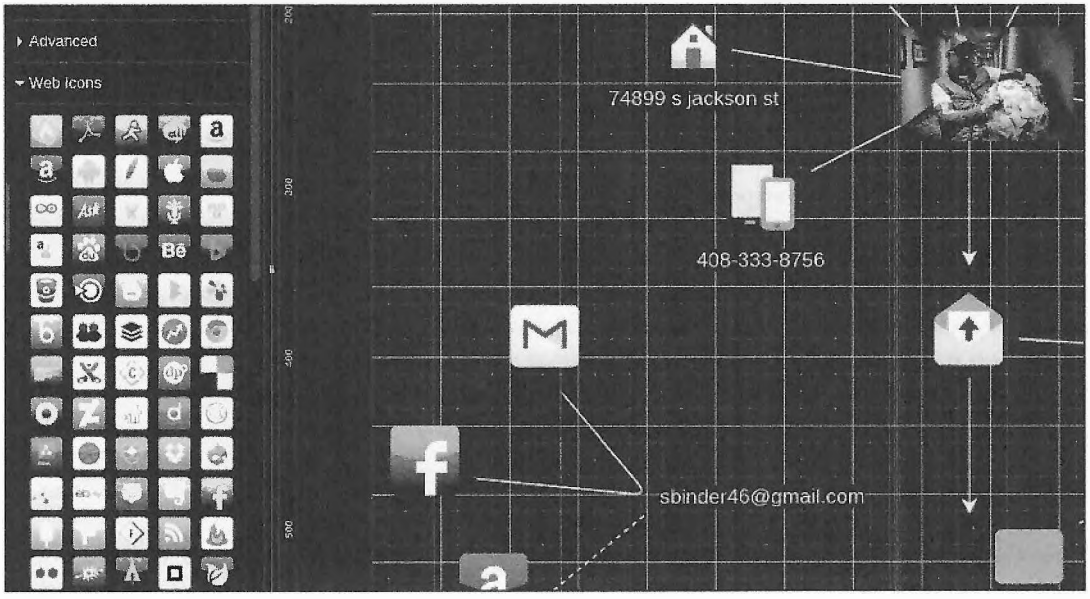


Figure 29.08: The Draw.io dark theme and web icon set.



Figure 29.09: The Viewpoint timeline and mapping application.

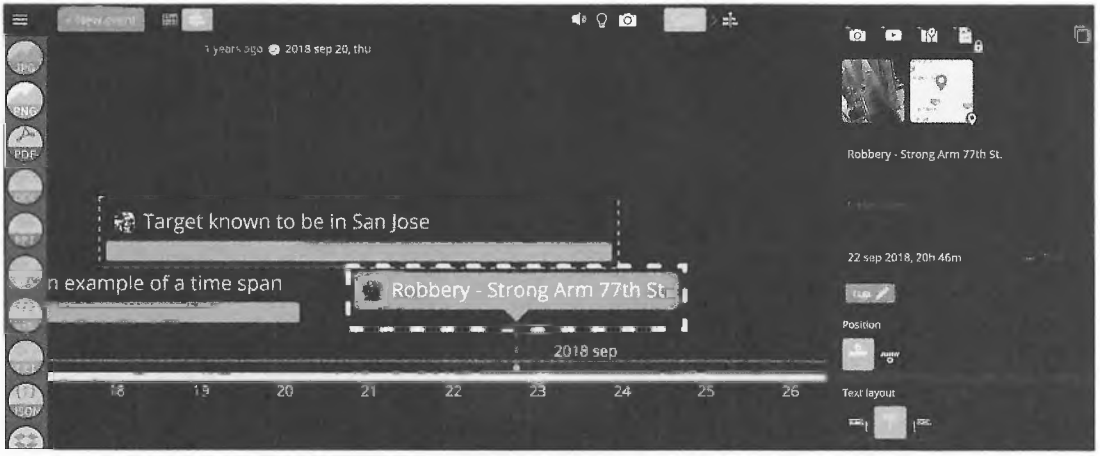


Figure 29.10: The Time Graphics application.

Appendix

The appendix of a report is made up of all the supporting documentation and media captured during the investigation. This will be screen capture files, still frames from YouTube evidence, spreadsheets of social media contacts, and any other digital evidence that can be displayed in a document format. These are working copies only, with the digital originals preserved on external media or other format approved by your agency's digital evidence policy. Many of your tools are going to generate pdf files which you will then need to include in your report appendix. If you have Adobe Acrobat DC you can easily combine the working copies of your pdfs into a single document for easier distribution, but the free versions of Acrobat do not support combining or conversion of multiple documents. I prefer an open-source option that does not require an account with Adobe, such as the following.

PDFsam Basic (pdfsam.org/pdf-merge)

PDFsam Basic is a free and open-source application for merging multiple pdf files. It is available for Linux, Mac, and Windows, but the Linux version is the perfect fit for our investigative VM. Installation in Ubuntu Linux is straight forward, as follows.

- In your Linux VM navigate to https://pdfsam.org/download-pdfsam-basic/.
- Click on "Deb package for Debian based Linux distributions" to download the app.
- Right-click the file and choose "Open with Software Install".
- Click the "Install" button.

This installation provides a new icon in the Applications menu. Opening the application presents a user-friendly menu with many options. The "Merge" function will compile pdfs into a single file with the following steps.

- Choose "Add" or drag multiple pdf files onto the window.
- In "Merge Settings" > "Table of Contents", select "Generate from file names".
- Click "Run" on the bottom left side.

A new combined pdf document will be created in the same directory as the source documents. PDFsam Basic has many more features that will allow you to manipulate your pdf documents, such as removing or rotating a single page within a long document. The main menu is accessible by clicking on the house shaped icon on the top right of the application. To the right of that, the icon made up of three stacked horizontal lines will take you to the settings page where you can change the default output directory, should you want to make it your case folder. The premium versions of PDFsam are closed source, costing roughly $60 a year, and primarily offer more options to edit and convert the documents into other filetypes such as docx. I find the free version sufficient for my needs. Figure 29.11 displays a typical menu layout.

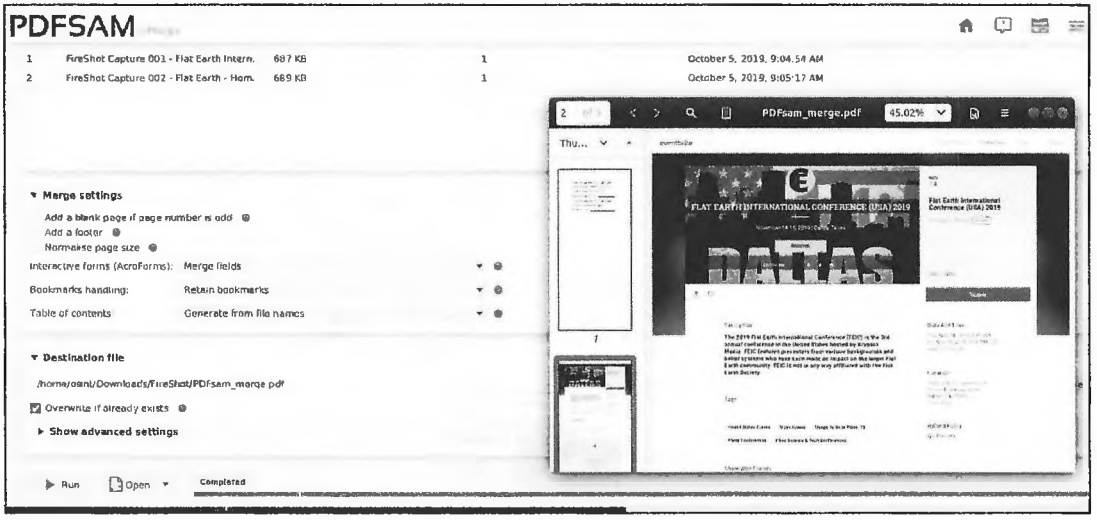


Figure 29.11: The PDFsam interface.

Hunchly (hunch.ly)

Hunchly was covered in prior chapters and is one of the only paid OSINT tools I use daily. If you are a Hunchly user, the report builder can be used to create your appendix even if you are not planning to use Hunchly to prepare the entire report. The first step is to set aside the evidentiary copy of your recovered data. This is accomplished by clicking on "Export" and then "Export Case". This will organize the raw data into an archive which can then be transferred to optical or other storage media. We may want a visual representation of this data included in the appendix of our report, which can be accomplished with the following steps.

- On the main toolbar, choose "Export" and then "Open Report Builder". You will see optional report elements on the left and an empty title bar on the right.
- Title the report to match your case, such as the target name followed by the case number.
- Under the title bar, there is a box with an editable title of "Section 1". Click it and rename as "Photo Evidence" or something similar.
- Find "Tagged Photos" in the left panel. Click on it once to expand a list of your tagged photos. Click and drag "All Tagged Photos" over to the box that you retitled "Photo Evidence".
- If you wish to add a list of all captured pages, drag "Page History Table" to the right.
- On the top right, you can toggle between docx and pdf export options and then select "Export". This will open a "save as" window and allow you to browse to your case directory where you can rename it "Photo Evidence" and click save.

The primary advantage of using Hunchly to generate your appendix is that it will add the capture date, hash, and URLs to each preserved page and image. This is beneficial when using the appendix as a hard-copy visual reference for the full evidentiary case export. Figure 29.12 displays the Hunchly Report Builder.

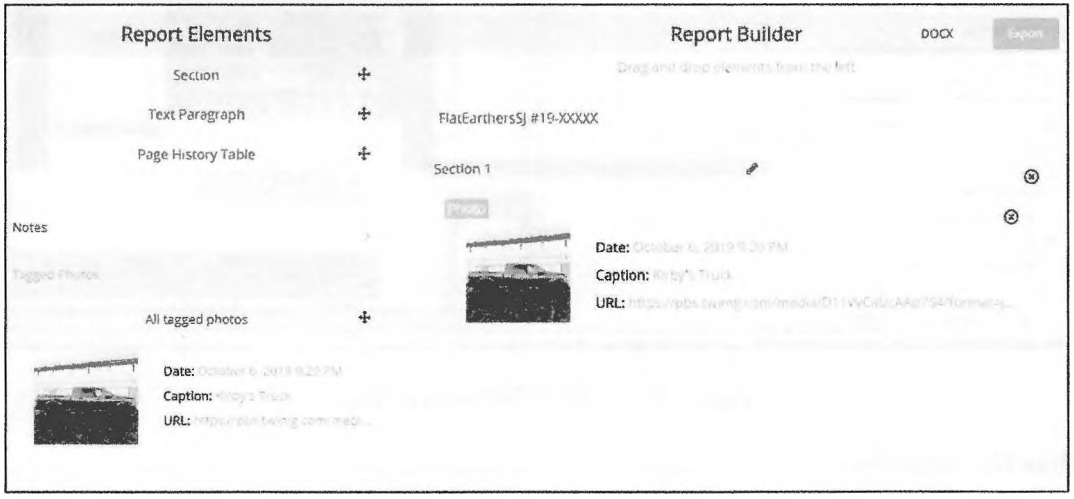


Figure 29.12: The Hunchly Report Builder interface.

Document Sanitization

Although I prefer to submit paper reports with an attached optical disc of supporting digital evidence, you may be constrained to the documentation procedures defined by your organization. If your agency uses any of the full-feature proprietary products such as Word or Acrobat, there is going to be metadata included in the digital version of your report that could unnecessarily expose your account details. While we would never want to remove metadata from evidentiary documents, removing hidden author data from your investigative report prior to submission is a recognized best practice.

To accomplish this in Microsoft Office products select File > Check for Issues > Inspect Document. This will open the document inspector window with a list of possible metadata. Check all boxes and click on "Inspect". This will execute a scan and any concerns will display a red exclamation point. Typically, these will be in the "Properties" or "Header, Footers, and Watermarks" sections. I often want to keep footnotes and do not remove this section. However, next to "Properties" click "Remove all". This will delete most of the metadata such as the username of the author. This is by no means a complete cleanse, but is an easy way to remove low hanging fruit. Figure 29.13 displays this dialogue.

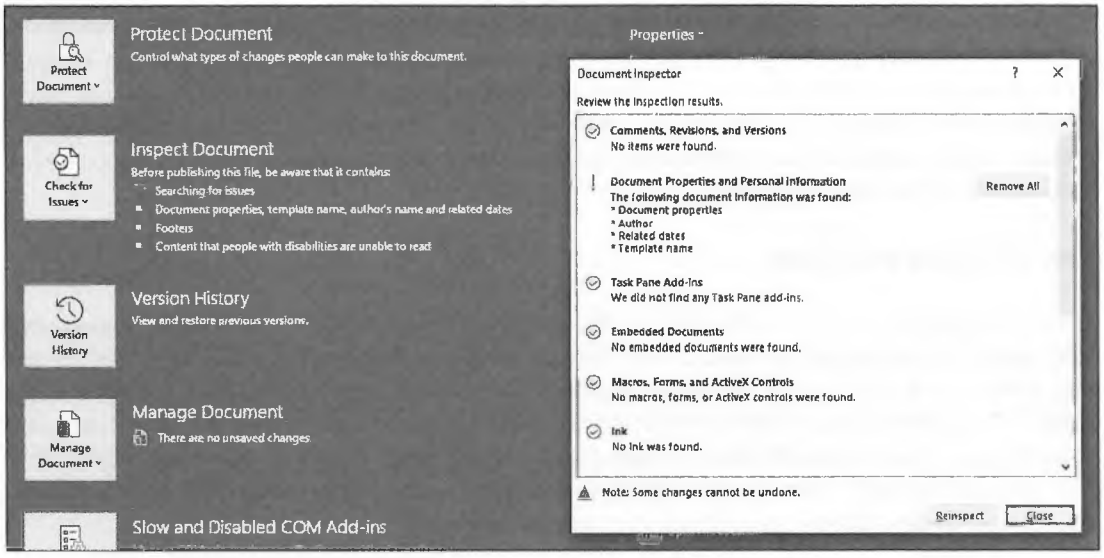


Figure 29.13: Metadata removal options within OneNote.

To perform a similar metadata removal in Adobe Acrobat, click on "Tools" and under "Protect & Standardize", click on "Redact". A new set of options will show up just below the top toolbar. Click on "Remove Hidden Information" and a panel will open on the left side of the application showing a running scan of your document. Under "Results" you can see what types of data were found. You should uncheck boxes next to any items that you do not want removed. When you click on "Remove", anything with a check mark will be deleted permanently from your document. These changes are not applied until the next time you save the document. To do this, click on "File", then "Save As" and give your cleaned document a unique filename, such as Policy_v6_san.pdf. Whenever possible, I try to use naming conventions that convey the state of the file such as adding "san" or "clean" to indicate it has been sanitized.

In the toolbar, you will see an option to "Sanitize Document". Think of this as the express option for removing metadata as it will use a set of default removal choices to perform the same task that we previously did. I prefer the former method for most projects, as I like to have more granular control over what is removed.

It should be noted that removing metadata can potentially break formatting or corrupt the entire file; therefore, it is wise to create a backup copy prior to metadata removal. You can also use third-party scripts and tools to clean documents, such as the Metadata Anonymization Toolkit (https://github.com/jubalh/MAT) for Linux. I find that these tools are more prone to breaking documents created in Microsoft and Adobe products. Thus, for those types of documents, I use the built-in removal tools described above.

Keep in mind an old, but still very effective, method for removing metadata from non-evidentiary documents is using a scanner. Just print your report and scan the paper report back to a digital format using a scanner. The new document will have metadata based on the scanner/copy machine rather than your own accounts or workstation. If this scanner is an enterprise grade copier, it likely saves this data temporarily to a hard-drive and you would never want to use this technique with sensitive documents.

Case Archiving & Cleanup

Upon completing a report, I collect my analog and digital case work and archive it for future use. The length of retention of these notes, virtual machines, screen captures, and reports are dependent on the policies of my organization or the expectations of the client for contracted work. When appropriate I keep handwritten notes for six months and digital documentation for at least three years. Should the case be later adjudicated, I have the documentation available. It is also not uncommon to have a current target resurface in a future investigation, in which case those digital notes will give you a head start on that new case.

A final warning for users of OneNote or other digital notebooks. The search capabilities built into OneNote are very useful when querying subjects for any previous work your team has completed. However, hoarding data long-term can leave you with an awful mess if you do not have a system for archiving old cases. I have an archive notebook for each year and within it sections broken down by month. Any completed cases get moved to the archive, and at the end of each month I review any remaining sections to assess their status. Then, at the end of each year, I go through all my notebooks to tidy up any loose ends and consider purging any archived notebooks that have gone untouched for longer than three years. Another approach is to export older notebooks and save them to optical discs or other long-term storage. In your own agency, please ensure that you implement an organization system that works for your workflow and one that is also compliant with retention/audit policies.

Smaller Investigations

Much of this chapter has focused on large investigations which require in-depth note-taking applications and visual analysis tools. This may be overkill for you. I end the chapter with reporting options for smaller investigations. This section attempts to convert your minimal evidence into digestible reports. I believe the first consideration for smaller OSINT reports should be the "Executive Summary". This document should be limited to one page, and it should only focus on the absolutely vital evidence found during your investigation. Think of it as the "elevator pitch". You have only a few minutes to present what you found. There is no time to explain how a reverse-video search works or about the ways you can exploit a Facebook account with a user ID number. The following page displays a typical Executive Summary.

Investigation Number: 2017-54887 Date: November 17, 2017
Investigator: Michael Bazzell Suspect: John Davis

On November 12, 2017, I was requested to conduct an investigation into John Davis, a former employee that has made claims against your company in reference to being injured on the job. Upon completion of my investigation, I have collected 143 relevant screen captures of online evidence that validates your resistance to his claims. The following pages include detailed analysis of the online content, but this cover page summarizes the most damaging evidence found that contradicts his previous deposition testimony. I find the following facts to be most useful.

- On June 13, 2017, Mr. Davis claimed to sustain a back injury while working at the South Plant on June 12, 2017. He was sent home on the 13th, and has not returned to work since.

- On June 11, 2017, Mr. Davis received an invitation through a Facebook Event from his sister, Jane Davis to attend a retirement celebration at Lake of the Ozarks, Missouri. This event was scheduled to take place on Saturday, June 17, 2017.

- According to his deposition, Mr. Davis' chiropractor is Neil Stevens in Ladue, Missouri. Dr. Stevens ordered Mr. Davis on bedrest until further notice on June 14, 2017, and continued this order on July 27, 2017.

- On June 18, 2017, Jane Davis posted numerous photos to her Facebook profile, including an image of Mr. Davis lifting a keg of beer above his head.

- On June 27, 2017, Mr. Davis participated in a 5K race in St. Louis. Online records indicate that he placed 113 out of 1,237 with a time of 00:28:16.

- On August 12, 2017, Mr. Davis posted a review on Amazon for a steel rack for mounting onto an ATV. The review included, "This thing survived five miles of rough terrain last week, and my rifles didn't get one scuff ... I will never go hunting without this on my 4-wheeler".

- Dr. Stevens' daughter possesses an Instagram profile indicating that she is best friends with the daughter of Mr. Davis. Mr. Davis' daughter attended a sleepover at Dr. Stevens' home on March 12, 2017. Dr. Stevens is an avid hunter, as is Mr. Davis. On September 22, 2017, Mr. Davis and Dr. Stevens placed second in a duck hunting competition in Boone County, Missouri. Online records confirm they were on the same team.

The following pages represent the evidence of these findings. Please note that all screen captures were preserved in digital format, and are included within the DVD attached to this report.

The previous example provided just enough detail to give an overall synopsis of the case. Most clients will not read past this page until necessary. They may thumb through the entire report and look at any screen captures, but ultimately, they possess the information they need. Obviously, you still need to provide your evidence, which is what we will do in the next portion of the report. I use a specific "Suspect Details" template, but you should create your own that best represents your needs. Consider my example below.

Investigation Number:		Date:
Investigator:		Suspect:
Full Name:	Age:	DOB:
Home Address:		Telephone:
Mailing Address:		Telephone:
Spouse:		
Child # 1:		
Child # 2:		
Suspect Email Addresses:		
Spouse Email Addresses:		
Child # 1 Email Addresses:		
Child # 2 Email Addresses:		
Suspect Usernames:		
Spouse Usernames:		
Child # 1 Usernames:		
Child # 2 Usernames:		
Suspect Social Network Profiles:		
Facebook:	Twitter:	
Instagram:	Google:	
Other:	Other:	
Spouse Social Network Profiles:		
Facebook:	Twitter:	
Instagram:	Google:	
Other:	Other:	
Child # 1 Social Network Profiles:		
Facebook:	Twitter:	
Instagram:	Google:	
Other:	Other:	
Child # 2 Social Network Profiles:		
Facebook:	Twitter:	
Instagram:	Google:	
Other:	Other:	
Other:	Other:	

This partial template includes explicit details located during my investigation. I do not cite any sources, and I use this as an easy reference for account information that someone may use later. On some investigations, the Suspect Details section is several pages. After I have completed my Executive Summary and Suspect Details, I write the report narrative. The following are two small portions of how this report might appear. Note that I already possess screen captures of all online evidence, titled as previously explained.

Investigation Number: 2017-54887 Date: November 17, 2017
Investigator: Michael Bazzell Suspect: John Davis

On November 17, 2017, I was assigned an investigation into potential fraudulent medical claims made by John Davis, a former employee of INSERT COMPANY HERE. The following represents a detailed summary of my findings.

I located the Facebook profile of the suspect at facebook.com/JohnDavis9. I generated a screen capture of each section of this profile, as publicly visible to any user. These files were saved to disk, and titled as follows.

001-https__facebook.com_JohnDavis9 | 2017-11-17-10-15-11.pdf
002-https__facebook.com_JohnDavis9_photos | 2017-11-17-10-16-12.pdf
003-https__facebook.com_JohnDavis9_about | 2017-11-17-10-17-18.pdf
004-https__facebook.com_JohnDavis9_friends | 2017-11-17-10-19-31.pdf
005-https__facebook.com_JohnDavis9_events | 2017-11-17-10-22-11.pdf

Most notable within these captures is the "Photos" section identifying photos associated with hunting, including images with both Mr. Davis and his doctor within the same photo. This specific evidence is titled as follows.

006-https__facebook.com_photo.php?fbid=1828444224119932 | 2017-11-17-10-33-11.pdf

Note that I did not place a screen capture of this evidence within the report itself. There are two main reasons I do not place screen captures within printed report text. First, I believe it is better to keep the report concise and clutter-free. The client can easily view the evidence within the provided disc or drive. Second, it can be quite difficult to include an entire screen capture within an 8 ½ x 11 page. I would likely need to crop any images, which also means that I am redacting evidence. I would rather my client view the digital screen capture which displays the entire page. If I do want to include printed online screen captures, I will do so at the end of the report, and as a supplement. Also notice that after each set of screen captures, I summarized the value. In this example, the most beneficial evidence was a specific image. I have found that presenting the client with every possible detail results in an overwhelming report. I believe it is our job to tell the client what he or she may care about relevant to the case. After all, we are the analysts. Anyone can collect a bunch of screen shots. The true value is understanding why these captures are important. In the next example, I outline findings on Twitter.

I located the Twitter profile of the suspect's daughter, Kylie Davis, at twitter.com/kdavis722. I exported the most recent 3,200 posts (Tweets), and saved this as kdavis722.csv on the attached disc. I found the messages between Kylie Davis and Patricia Stevens (pstevens6655) of most interest. I isolated these messages with the following two queries.

from:kdavis722 to:pstevens6655
from:pstevens6655 to:kdavis722

Screen captures of these messages were saved as the following.

045-https__twitter.com_from:kdavis722 to:pstevens6655 | 2017-11-17-11-15-45.pdf
046-https__twitter.com_from:pstevens6655 to:kdavis722 | 2017-11-17-11-16-42.pdf

Of these messages, I found three references to the suspect and his doctor participating in a hunting trip. These specific references were cropped and saved as follows.

045a-Cropped Messages.pdf
046a-Cropped Messages.pdf

Note that I included details of the search technique, the specific evidence files, and information as to the importance of the content. I like to be as brief as possible. The digital screen captures provide all of the evidence necessary, and explicit detail of each capture is overkill. In most investigations, I have several pages of this type of narrative. Finally, I include a one-page Summary Report at the end. This also identifies future investigation needs and whether the incident is resolved. The following is a partial example.

This investigation was conducted with the hopes of identifying the participation of medical fraud by the suspect. I believe that this claim of fraud has been proven true. I advise continuous monitoring until the workman's comp claim is settled. Specifically, this investigation reveals the following as fact.

- Online evidence proves a personal association between the suspect and his doctor.
- Online pictorial evidence proves the suspect to have been physically fit enough to lift heavy objects within the time period of the disability claim.
- Online evidence proves the suspect to have been physically fit enough to run 5 kilometers within 28 minutes within the time period of the disability claim.
- Online evidence proves the suspect to be able to hunt in rugged conditions within the time period of the disability claim.

Note that I did not make any claims I could not prove, and I did not inject much opinion into the matter. I always try to keep the reports factual and unbiased. If I locate any online evidence that supports the suspect, I include it as well. When this happens, I make sure to emphasize this with digital screen captures and a brief summary of the content. Overall, I try to include the following with each report.

- Executive Summary: One-page synopsis of vital evidence.
- Suspect Details: Specific data such as all personal identifiers, usernames, etc.
- Narrative Report: Detailed findings with references to digital evidence and summaries.
- Summary Report: One-page summary of facts and need for future work.
- Digital Evidence: A DVD or drive that contains all screen captures and files.

As stated previously, I believe that every OSINT investigation should be conducted within a virtual machine. At the end of the investigation, the entire machine should be exported as a single digital file and included with the digital evidence. I would also consider including the following paragraph within your narrative report.

> This entire investigation was conducted within a Linux virtual machine. This operating system was created on (insert date) and saved as a master copy. All security updates were applied at that time and no online investigation was conducted within this master copy. A clone of this system was created and titled (case number). This clone was used as the only operating system resource for the entire investigation. No other investigations were conducted within this clone. At the end of the investigation, this virtual machine was exported as (file name). This file can be used to recreate the entire investigation environment exactly as it appeared during the actual investigation.

This verbiage announces your competence to the prosecution and defense. It may stop some scrutiny toward your work during a trial or hearing. Ultimately, it shows that you conducted your investigation fairly with great concern for the integrity of your evidence. Additionally, this may make you stand out to your supervisors or the office of prosecution. I have found that consistent dedication to accurate reporting can go a long way toward your reputation and promotions.

This chapter has presented many reporting options, some of which may contradict the others. Your reports may be extremely complex and contain dozens of pages, or consist only of the executive summary. My goal here was to simply provide documentation considerations and their impact on the success of your investigation. Once you have developed a report template that works well for your investigations, recreating a report for each case will save time and energy. Everyone's reports are unique, and you should find the best way to explain your findings to an audience of any technical level. It is now your mission to identify the best documentation and reporting practices for your needs. I close this chapter with one final consideration for your next large-scale investigation. We should isolate a burner SIM card with a cellular data account and be prepared to disclose our information. The following should explain.

Large-Scale Continuous Investigations

Many of my investigations span weeks or months. Most are extremely sensitive and must possess complete isolation from any other investigations. This goes beyond a clean virtual machine. They demand management of dedicated Facebook, Twitter, email, and other accounts, which can become difficult when providers detect my behavior as suspicious. In these scenarios, I assign a dedicated mobile device and cellular data account to each investigation. This may sound ridiculous and expensive, but we can provide this extreme layer of protection at minimal cost. My process is as follows, and only applies to investigations which cannot afford to be compromised by case-contamination or suspended accounts.

First, I need an unlocked mobile device. This will never be used outside of the investigation or for any personal use. I usually buy refurbished Android devices at local cell phone repair shops for $20 each. These are generic, low-powered, and overall undesired units which have very little value. You may also find similar new phones in grocery stores, pawn shops, or online through eBay or Amazon. I then purchase Mint Mobile SIM cards from either Amazon or mintmobile.com. These are $2.50 each, but include a $5.00 credit for service and a one-week free trial. Three months of service is $15 per month ($45 total), and T-Mobile is the data provider. I purchase the phones and service with prepaid gift cards or Privacy.com virtual cards.

I insert the SIM into the device, download the Mint Mobile application over Wi-Fi, and register an account under an alias name, and start my trial. Since this is prepaid service, there are no verifications or credit checks. I choose a local telephone number issued by Mint Mobile which can be used for the countless verification text hurdles I am likely to face over the course of the investigation. When Facebook demands a real cellular number, I give this out freely. When Gmail blocks my account as suspicious, I can unlock it with a verification text. I no longer dread the suspension notices typically received when relying on VOIP numbers, VPN connections, and burner email accounts. This cellular number is my ticket out of most negative situations.

I can also place applications on the device when an emulator is not appropriate. As an example, Snapchat and Tinder usually block Genymotion and other virtual Android environments. With this device, I can install the native apps, launch a GPS spoofer, and conduct my investigation without roadblocks. My device appears "real" and I bypass scrutiny from the providers.

At the end of the investigation, I remove the SIM and place it and the phone in a sealable plastic bag with holes punched for use in a binder. These can be found in any office supply store. The phone and SIM are part of the investigation. The service will expire on its own and I have a clean digital trail. If necessary, I can provide the device and account details as part of discovery in court. I have no concerns if an expert witness wants to clone the machine for their own analysis. If the number should become exposed in a data breach, it is not a problem. It will never be used again. If you plan to replicate this technique, I advise preparing now. You do not want to be shopping for a device and waiting for delivery of a SIM card while you should be investigating your target.

Chapter Thirty
Policy & Ethics

There is much controversy over the use of social media and personal data for investigative purposes. Therefore, it is critical that your organization has a clear and concise technical investigations policy in place. This should be a one or two-page document and it must include at a minimum a training standard, an approval process, and an appropriate use policy.

Avoid using language tied to specific technologies or platforms, as those will change rapidly over time. In-house council may request language pertaining to Facebook or another specific third-party platform. They may recommend building policy around specific tools or technologies. It is our responsibility to demonstrate how the rapid changes in technology will make such a policy ineffective and almost immediately irrelevant. Moreover, limitations imposed by overly specific regulations will likely confine us to difficult options.

Responsible investigative policy should focus on the appropriate use of techniques and technologies. Review and borrow heavily from the mission statement and boilerplate language that your organization already applies to traditional investigative procedures. Any existing policy relating to training, equipment, enforcement, supervision, or chain of custody will have language that will fold easily into the framework of your online investigations policy.

Most organizations place online investigations policies within a section covering the overall social media policies. That section is usually broken down into a statement of intent, social media definitions, official use, personal use, and finally investigative use. I believe that operationally it makes more sense to group online investigations policies under the "Investigations" section of your agency manual. This is consistent with our argument that purpose and use should drive policy, rather than platform. Open source intelligence gathering is just another form of lawful investigation and we want to align it as such. An online undercover operation should be conducted and scrutinized in a manner similar to a covert assignment on the street. Comparing online investigative procedures with how the agency might use social media for public relations makes little sense and yet most organizations group them together.

I believe that for most agencies an online investigations policy should be no longer than two pages. The following is the basic framework for an appropriately non-specific policy. The statement of intent and definition sub-sections may be omitted if those are addressed within your general social media regulations.

SECTION 1.0 - SOCIAL MEDIA - INVESTIGATIONS - STATEMENT OF INTENT

[This statement articulates the importance of establishing guidelines for appropriate use of online accounts, services, and software for the purpose of online criminal investigations. It should be brief and reflect the mission statement of your organization. Consider borrowing language from existing policies related to appropriate investigative tactics.]

1.1 DEFINITIONS

[If not addressed elsewhere in the section]
Browser – Software which translates various forms of internet code into human understandable formats including, but not limited to, plain text, graphic images, and animations.

Private – Content is private when transmitted or conveyed in a manner that shows reasonable measures and intent to limit access to specific individuals, and where it is reasonable to expect that only those individuals will have access.

Public – Content is public when shared on a site or service in a manner wherein a reasonable person would expect that it is accessible to a broad or non-specific audience.

Post – Submitting information to the internet or a social media site.

Site – A page or series of pages hosted on the internet.

Social Media – Various online platforms, sites, and services facilitating the "posting" or sharing of information on the internet. Examples of social media content include text, images, audio, and video.

1.2 APPROPRIATE USE

Investigative accounts and tools will be used for agency purposes only. Investigative social media profiles, software, and services will not be used for personal purposes. Agency resources will be used in compliance with local, state, and federal law. Personal equipment, services, and accounts will not be used for investigative purposes.

Investigations will abide by all legal restrictions relating to private vs. public data and consistent lawful search and seizure. Where required by law, legal orders will be obtained from the appropriate magistrate and/or jurisdictional authority.

1.3 APPROVAL PROCESS

Social media accounts and online identities used in an investigative capacity will be reviewed and approved by the unit supervisor to ensure they are within policy. Ongoing casework will be

periodically evaluated by the unit supervisor for compliance with agency policies and investigative best practices.

1.4 TRAINING

Personnel will only participate in the creation and use of investigative social media accounts after successfully completing an approved Covert Online Investigations training course. Unit supervisors are responsible for ensuring that all personnel utilizing investigative technologies are provided with appropriate training and supervision.

1.5 RETENTION/AUDIT

Data obtained during online investigations is handled and stored per the agency's existing digital evidence and retention policies. Where operational constraints necessitate unique data management procedures, the unit supervisor will review, approve, and log any non-standard measures along with a written justification.

All operational documentation and investigative logs are subject to review by the Office of the Inspector General [or appropriate body of oversight].

1.6 AGENCY FORMS (if applicable)

- Authorization & Training Verification – Covert Online Investigations
- Unit Compliance Log – Online Investigations

1.7 ACKNOWLEDGEMENT OF TECHNOLOGICAL ADVANCEMENT

The state of the internet and social media technologies evolve rapidly, and it must be acknowledged that the pertinence of overly specific policy will limit the effectiveness of said policy. It is for these reasons that appropriate and responsible use, rather than specific products or technologies, are the foundation of this section.

1.8 REVISION HISTORY

*Legacy versions of this policy which are typically presented in reverse chronological order.
Rev. 19.32 [Link to official archive if applicable] [Date of inception]-[Date Rescinded]

1.9 APPENDIX

[Appendices are used for informational material that is helpful, but not directly related to the implementation of the policy. These can be references to related policies/procedures or case law providing the foundation for investigative best practices.]

For smaller teams or organizations lacking an official policy structure, an alternative format could be a set of standard operating procedures (SOPs). The content will be essentially the same, but less formal in structure. Take the previous policy example, delete the definitions and headings, and paste the remaining content into your agency's memorandum template. Have it approved by a commander or manager higher in rank than the front-line supervisor. The important thing is to have some type of documented standard for these types of investigations. Eventually a controversial incident or high-profile case will expose your team to scrutiny and a well-structured policy will go a long way to demonstrating transparency and professionalism.

If your organization requires language around "interaction", such as "friending", make this as non-specific as possible. Some agencies also restrict the use of covert social media accounts, which will extremely limit the effectiveness of your OSINT work. You have to educate management and legal advisors on the importance of covert accounts in building actionable intelligence. Decades of case law supports the use of "undercover" operations for traditional investigations and online investigations should be treated no differently. If you are forced to accept some restrictions related to covert accounts or infiltration, push for them to be contained within SOPs rather than agency policy. Continue to lobby for a reasonable standard utilizing appropriate concepts and language that is not overly specific. Common phrases of this type of language is: "for lawful purposes", "consistent with agency training and standards", and "when feasible and reasonable".

For those in the public sector, specifically law enforcement, always be thoughtful regarding any reasonable expectation of privacy. U.S. courts have ruled consistently in favor of online undercover infiltration tactics based on the fact that a reasonable person knows that others on the internet may not be who they purport to be. Where we can get ourselves into trouble is when we cross over into what could be considered private communication. One example is recovering an unlocked smart phone from an arrestee and noticing that the phone is receiving social media messages. You would not want to pretend to be the phone's owner and engage in conversations without appropriate court granted authority, such as a search warrant. Another scenario is a victim allowing you to use their social media account for direct communication. Some jurisdictions consider any communication outside a "group chat" private and you may be required to obtain a legal order to continue any type of private or direct messaging using a third party's identity, even if it is with consent. Much depends on the jurisdiction in which you are working, which can get complicated when your cooperating victim and the suspect are in two different states, or in some cases, different countries.

I am not an attorney and the best path is always to solicit advice from legal counsel that is familiar with the local, state, and federal laws affecting your jurisdiction. Common sense goes a long way and I always ask myself if there is any way a target could argue a reasonable expectation of privacy. Additionally, being good stewards of privacy is important when building trust with our communities. We never want to portray any hint of reckless during intelligence gathering. Practice good documentation during the process in order to build favorable case law and maintaining as many of our tools as possible.

Ethics

Privacy is certainly the cornerstone of ethical considerations related to open source intelligence gathering, but I will also be discussing issues of intent and deception. Privacy is a broad issue and will pertain to almost every portion of our case. Deception becomes pertinent when we engage in active covert measures such as undercover accounts or "friending" targets.

We each value privacy in our own lives and we should always be as mindful as possible in regard to the data we collect during our investigations. Early in an investigation, you should clarify the scope of engagement. How deep are you going to dig and what is the appropriate balance between purpose and level of intrusion? Here are two scenarios from the law enforcement world that represent opposing ends of the spectrum.

Scenario #1: I receive a report of online threats against a public official and am asked to make an assessment and if necessary, identify the owner of the account. I locate the account and posts in question and immediately see that although they are mean spirited, the comments do not articulate a true intended threat. A common example of this is similar to, "I hope you get cancer". It is a terrible thing to say, but there is no threat implied. At this point, I can report back to my boss that there is no threat. If necessary, I can show her a screenshot of the post. I have accomplished the mission and there is no justification to dig into the details of that person's life.

Scenario #2: We receive a tip from an internet service provider that a specific IP address is pushing traffic containing a large amount of child pornography. I investigate the IP address and find that it is a listed TOR exit node. That means the person operating the router at the associated residence or business is allowing people on TOR to funnel their internet traffic through their device. So now I know there is a fair chance that the person at that residence is not directly involved in the child pornography and may be completely unaware of its presence. I do not stop my investigation. I dig up every piece of public data that I can on the people controlling that router.

My decision to press forward on the second scenario is based on the seriousness of the crime. For a situation like crimes against children, you need to exhaust all reasonable means up until there is a certainty that the persons of interest bear no public threat. I am still obeying all laws, but it is far more reasonable for me to sort through the entirety of someone's public data if failing to do so might cause extraordinary harm to someone. The seriousness of the crime being investigated justifies a far broader scope of investigation and level of intrusion.

While some data on the internet is clearly public, there is questionable data that is publicly accessible and legally obtained, despite the owner intending it to be private. An example of this might be an Elasticsearch database that was improperly configured. If we can access it via a browser and providing no credentials, we should be in good legal standing. However, we are taking advantage of a mistake on behalf of the owner rather than something being intentionally shared. My feeling on these types of data sets is that our intentions make all the difference in

deciding if it is appropriate to use them. Examples of strong use cases would be defensive vulnerability assessments or investigations into crimes against children. We always want to balance the degree of impact on another individual's privacy with the greater good being served by the mission. If I am merely investigating a misdemeanor crime or gathering intelligence about what looks to be a peaceful community event, I am going to be far less aggressive in accessing any sites or data that were clearly made public in error.

In the workflow section, we talked about identifying your mission goals at the beginning of any OSINT engagement. Establish the reason why you are collecting intelligence on this target. This not only helps you to get organized in your approach, but also raises the question of intent. It might be that I am following orders or fielding a request from a colleague. The target might be related to a person of interest on a larger investigation. They might even be a future employee going through a background investigation. All of these are typical and reasonable business purposes for exercising our intelligence gathering skills.

I don't always know all of the details of a case, but I need to be able to articulate the reasonable operational justification for digging into a person's life. This is not a high bar, but just as I do not follow a stranger on the street for no reason, I too show care and consideration when conducting investigative work on the internet. A good rule of thumb is that any OSINT work that provides personal gratification is a red flag. We don't stalk exes or someone that caught our eye on the train, even out of innocent curiosity. Likewise, we don't cyber-stalk the guy in the car behind us in heavy traffic, who expresses his feelings towards us with a crude gesture. We are professional investigators and always leverage our skills in a professional manner.

One of the clearest examples of using OSINT for ill-intent is the practice of doxing. Doxing is using online research to profile an individual and then posting that information publicly on the internet with the intent of creating fear or embarrassment. The goal is to make the person feel exposed and show that you have power over them. It is cyber-harassment and it can even rise to the level of being a chargeable criminal offense.

Typically, my OSINT work is the result of a public safety event or statutory obligation, which allows for well-defined ethical ground and clear justification. When I do run into "gray areas", it is usually when doing contract work for private sector clients. It is perfectly acceptable to conduct investigative work on behalf of a for-profit organization, but we do need to be clear on the intended use of any gathered intelligence. Often this is very straightforward. Common use cases are the vetting of vendors or potential employees. Another is vulnerability assessments on both the organization or C-level management. How the client intends to use our work product matters. We never want to be an unwitting party to a harassment or cyberstalking situation.

Deception is any behavior that misleads another person to believe something that is not true. This seems wrong when taken at face value, but there are occasions when deception is ethical and warranted. Take for example using social engineering to get your home address removed

from a site-run data-mining company. You are doing no harm to others, but you are most certainly using deception to accomplish your goal.

The two most common forms of deception in open source intelligence are the use of covert accounts and the practice of infiltration or "friending" target individuals or groups. Let's look at each of these and their ethical considerations.

Covert Accounts: The use of covert accounts violates the "terms of service" of most social media platforms and online services. However, covert accounts are critical to successful queries into social media platforms, and not authenticating with these services would hamper the effectiveness of our work. If our intent is good, we have articulated the reasonable justification for the use of these accounts and stayed within the confines of our investigation, this deception will not be invasive to the other users. We are one of many anonymous accounts.

Infiltration: Infiltration, such as "friending" individuals or joining social media groups, is far more invasive than merely using covert accounts to run queries. When you join a group, you are, to a small degree, changing the dynamics of that group. When you interact with a target on Facebook or in a forum by commenting on their posts, you are directly affecting them, even if it is in a very small or positive manner. Just as we demonstrated when making a comparison of scope, infiltration has its place in an ethically conducted OSINT investigation. Joining a criminal forum to gain the confidence of, and to deanonymize, child predators is a case where the greater good significantly outweighs the level of intrusion. I do not like limiting these types of tactics by policy, as that type of framework tends to be too rigid. However, I do want to always be able to articulate my justification before interacting with targets online.

Finally, I believe that we have an ethical responsibility to engage in ongoing training in both OSINT and privacy. Lack of familiarity with both technique and legal requirements leads to mistakes. Sloppy casework does a disservice to us all. We owe it to the people we serve, whether they are clients or victims of crimes, our best efforts and discretion in making sure that we uncover the truth and bring it to light. A big piece of protecting our tradecraft is showing that we wield our tools and talents with restraint and thoughtfulness.

SQL Files, 492
Telegram, 231
Telephone Numbers, 303
  Caller ID Databases, 304
  Carrier Identification, 303
  Craigslist, 314
  Historical Records, 313
  Search Engines, 310
  Search Tool, 317
TikTok, 389
Tinder Profiles, 249
Tinder, 247
TOR Browser, 58
TOR Search Engines, 157
Tor2Web, 158
TruMail, 263
Tumblr, 230
TweetBeaver, 197
TweetDeck, 203
Twilio, 305
Twitter, 187
  Analytics, 208
  Date Range Search, 191
  Deleted Messages, 193
  Fake Followers, 207
  Location, 190
  Search Operators, 189
  Search Tool, 212
  Search, 187
  Videos, 383
Usenet Archive, 494
Usernames, 277
  Assumptions, 281
  Compromised Accounts, 280
  Search Tool, 282
Vehicle Records, 435
Vehicle Registration, 436
Videos, 373
  Associations, 387
  Closed Captions, 384
  Live Streams, 385
  Search Tool, 391
Virtual Box, 21
Virtual Currencies, 443
  APIs, 444
  Search Tool, 446
Virtual Machines, 19
Virtual Private Network, 15
Virus Total, 412
VLC Media Player, 60
VM Clones, 27
VM Exports, 27
VM Snapshots, 25
Voat, 244
Voter Registration, 300, 440, 466
Wayback Machine, 146
We Leak Info, 266
Web Archives, 145
Web Browsers, 29
Wigle, 422
Windows Rebuild, 12
Windows VM, 102
Yahoo Groups, 151
Yandex, 159
  Cache, 145
YouTube, 373
  Channel Crawler, 388
  Comments, 377
  Deleted Videos, 378
  Download, 376
  Restrictions, 374
  Thumbnails, 375

Made in the USA
Coppell, TX
11 July 2020



Completely Rewritten Seventh Edition Sheds New Light on Open Source Intelligence (OSINT) Collection and Analysis

It is time to look at OSINT in a different way. For many years, and within the previous six editions of this book, we have relied on external resources to supply our search tools, virtual environments, and investigation techniques. We have seen this protocol fail us when services shut down, websites disappear, and custom resources are dismantled due to outside pressures. This book aims to correct our dilemma. We will take control of our investigative resources and become self-reliant. There will be no more need for online search tools; we will make and host our own locally. We will no longer seek pre-built virtual machines; we will create and configure our own. This book puts the power back in your hands. The new OSINT professional must be self-sustaining and possess their own tools and resources. You will become a more proficient subject matter expert who will be armed with the knowledge and readiness to articulate the sources of your findings. Aside from eleven brand new chapters, hundreds of pages have been updated to keep your OSINT investigative methods fresh. Furthermore, an entire new section featuring Methodology, Workflow, Documentation, and Ethics provides a clear game plan for your next active investigation. All-new custom search tools, report templates, and detailed documents are included via download. Today, we start over.