EXPERT INSIGHT

Learn Computer Forensics

Your one-stop guide to searching, analyzing, acquiring, and securing digital evidence

Second Edition

William Oettinger

‹packt›

Learn Computer Forensics

Second Edition

Your one-stop guide to searching, analyzing, acquiring, and securing digital evidence

William Oettinger

‹packt›

BIRMINGHAM—MUMBAI

Learn Computer Forensics

Second Edition

Copyright © 2022 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

Senior Publishing Product Manager: Aaron Tanna

Acquisition Editor – Peer Reviews: Saby Dsilva

Project Editor: Amisha Vathare

Content Development Editor: Liam Draper

Copy Editor: Safis Editing

Technical Editor: Aniket Shetty

Proofreader: Safis Editing

Indexer: Manju Arasan

Presentation Designer: Pranit Padwal

First published: April 2020

Second edition: July 2022

Production reference: 2050822

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.

ISBN 978-1-80323-830-2

www.packt.com

Contributors

About the author

William Oettinger is a veteran technical trainer and investigator. He is a retired police officer with the Las Vegas Metropolitan Police Department and a retired CID agent with the United States Marine Corps. He is a professional with over 20 years of experience in academic, local, military, federal, and international law enforcement organizations, where he acquired his multifaceted experience in IT, digital forensics, security operations, law enforcement, criminal investigations, policy, and procedure development. He has earned an MSc from Tiffin University, Ohio. When not working, he likes to spend time with his wife and his three miniature schnauzers.

This book is dedicated to IACIS and the pioneers of this field whom I have had the privilege of meeting and learning from. Mike Anderson and Will Docken were some of the first professionals I met, and they had a significant impact on me as I started in this field. I want to thank Eric Zimmerman, Harlan Carvey, Brett Shavers, and Steve Whalen for their work for the forensics community. Your information sharing and work have impacted me and helped me grow as an examiner. There is a long list of people who contributed to my success that I want to thank: Larry Smith, David Papargiris, Tom Keller, Dave McCain, Steve Williams, Scott Pearson, Scot Bradeen, Matt Presser, Mike Webber, and everyone else who has helped me along the way.

About the reviewer

Steve Whalen is a Certified Computer Forensic Examiner (CFCE) with degrees in Psychology and Sociology and served as a Delaware State Trooper. As a state trooper, Steve worked as a detective with the Criminal Investigations Unit and served as their first full-time forensic examiner for digital evidence. Building off this experience, Steve helped the Delaware State Police develop its first High Technology Crimes Unit in 2001, where he processed thousands of electronic devices and media containing digital evidence from hundreds of cases relating to intrusion, financial crimes, child sexual exploitation, narcotics, stalking and homicides.

After retiring from law enforcement, Steve co-founded SUMURI, a leading provider of hardware, software, training and services relating to digital evidence and computer forensics worldwide. Steve was the designer of the successful Macintosh Forensic Survival Courses, RAPTOR, PALADIN, CARBON and RECON forensic software, and TALINO workstations.

Steve has developed and delivered forensic training to thousands of investigators and examiners around the world through organizations such as the International Association of Computer Investigative Specialists (IACIS), the High Technology Crimes International Association (HTCIA) and the US Department of State Anti-Terrorism Assistance Program. Steve has over 20 years of experience in computer forensics and has provided training throughout North America, Asia, Europe, the Middle East, the Caribbean, Africa and Oceania.

Wanting to do more, Steve founded the non-profit company Red Stapler Inc. and used his knowledge of digital forensics, psychology, sociology to create a “first of its kind” software solution (`https://www.catchapredator.org/`) to combat the sexual exploitation of children in a way that has never been done in all of history.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

Table of Contents

Preface xv

Chapter 1: Types of Computer-Based Investigations 1

Introduction to computer-based investigations **2**

Criminal investigations **4**

First responders • 4

Investigators • 5

Crime scene technician • 5

Illicit images • 7

The crime of stalking • 12

Criminal conspiracy • 14

Corporate investigations **16**

Employee misconduct • 17

Corporate espionage • 19

Security • 20

Threat Actors • 20

Social engineering • 21

Real-world experience • 24

Insider threat • 25

Case studies **28**

Dennis Rader • 28

Silk Road • 29

San Bernardino terror attack • 31
Theft of intellectual property • 32
Summary **34**
Questions **34**
Further reading **35**

Chapter 2: The Forensic Analysis Process 37

Pre-investigation considerations **38**
The forensic workstation • 38
The response kit • 40
Forensic software • 43
Forensic investigator training • 47
Understanding case information and legal issues **48**
Understanding data acquisition **50**
Chain of custody • 53
Understanding the analysis process **56**
Dates and time zones • 57
Hash analysis • 57
File signature analysis • 60
Antivirus • 62
Reporting your findings **66**
Details to include in your report • 67
Document facts and circumstances • 68
The report conclusion • 70
Summary **70**
Questions **71**
Further reading **72**

Chapter 3: Acquisition of Evidence 73

Exploring evidence **73**
Understanding the forensic examination environment **76**

Tool validation **77**

Creating sterile media **82**

Understanding write blocking • 87

Hardware write blocker • 88

Software write blocker • 89

Defining forensic imaging **90**

DD image • 91

EnCase evidence file • 93

SSD device • 94

Imaging tools • 95

FTK Imager • 95

PALADIN • 104

Summary **109**

Questions **110**

Further reading **111**

Chapter 4: Computer Systems 113

Understanding the boot process **113**

Forensic boot media • 116

Creating a bootable forensic device • 117

Hard drives • 119

Drive geometry • 121

MBR (Master Boot Record) partitions • 122

Extended partitions • 125

GPT partitions • 125

Host Protected Area (HPA) and Device Configuration Overlay (DCO) • 130

Understanding filesystems **131**

The FAT filesystem • 132

Boot record • 133

File allocation table • 134

Data area • 136

Long filenames • 138
Recovering deleted files • 139
Slack space • 141
Understanding the NTFS filesystem 141
Summary 154
Questions 154
Further reading 155

Chapter 5: Computer Investigation Process 157

Timeline analysis 158
X-Ways • 160
Plaso (Plaso Langar Að Safna Öllu) • 164
Media analysis 176
String search 177
Recovering deleted data 180
Summary 183
Questions 183
Further reading 184
Exercise 185
Data set • 185
Software needed • 185
Email exercise • 185
Data carving exercise • 185

Chapter 6: Windows Artifact Analysis 187

Understanding user profiles 188
Understanding Windows Registry 190
Determining account usage 193
Last login/last password change • 193
Determining file knowledge 200
Exploring the thumbcache • 200

Exploring Microsoft browsers • 202

Determining most recently used/recently used • 204

Looking into the Recycle Bin • 207

Understanding shortcut (LNK) files • 208

Deciphering JumpLists • 209

Opening shellbags • 211

Understanding prefetch • 213

Identifying physical locations 214

Determining time zones • 215

Exploring network history • 215

Understanding the WLAN event log • 217

Exploring program execution 218

Determining UserAssist • 218

Exploring the Shimcache • 219

Understanding USB/attached devices 219

Summary 222

Questions 223

Further reading 224

Exercise 224

Data set • 224

Software needed • 224

Scenario • 224

Chapter 7: RAM Memory Forensic Analysis 227

Fundamentals of memory 227

Random access memory? 228

Identifying sources of memory 231

Capturing RAM 233

Preparing the capturing device • 233

Exploring RAM capture tools • 234

Using DumpIt • 234

Using FTK Imager • 236

Exploring RAM analyzing tools 238

Using Bulk Extractor • 238

Using VOLIX II • 243

Summary 246

Questions 246

Further reading 247

Chapter 8: Email Forensics – Investigation Techniques 249

Understanding email protocols 250

Understanding SMTP – Simple Mail Transfer Protocol • 250

Understanding the Post Office Protocol • 251

IMAP – Internet Message Access Protocol • 252

Understanding web-based email • 253

Decoding email 253

Understanding the email message format • 253

Email attachments • 257

Understanding client-based email analysis 258

Exploring Microsoft Outlook/Outlook Express • 258

Exploring Microsoft Windows Live Mail • 259

Mozilla Thunderbird • 260

Understanding WebMail analysis 262

Summary 266

Questions 266

Further reading 267

Exercise 267

Data set • 267

Software needed • 267

Scenario • 267

Interviews • 268

Email accounts • 268

Question to answer • 268

Chapter 9: Internet Artifacts 269

Understanding browsers 269

Exploring Google Chrome • 270

Understanding bookmarks • 270

Understanding the Chrome history file • 274

Cookies • 276

Cache • 277

Passwords • 278

Exploring Internet Explorer/Microsoft Edge (Old Version) • 278

Bookmarks • 279

IE history • 279

Typed URL • 282

Cache • 283

Cookies • 285

Exploring Firefox • 287

Profiles • 287

Cache • 289

Cookies • 290

History • 290

Passwords • 292

Bookmarks • 293

Social media 294

Facebook • 296

Twitter • 298

Service provider • 299

P2P file sharing **300**

Ares • 301

eMule • 302

Shareaza • 304

Cloud computing **305**

Summary **308**

Questions **309**

Further reading **310**

Chapter 10: Online Investigations 313

Undercover investigations **314**

Undercover platform • 315

Online persona • 316

Background searches **322**

Preserving online communications **331**

Summary **337**

Questions **338**

Further reading **340**

Chapter 11: Networking Basics 341

The Open Source Interconnection (OSI) model **342**

Physical (Layer 1) • 343

Data link (Layer 2) • 343

Network (Layer 3) • 344

Transport (Layer 4) • 344

Session (Layer 5) • 345

Presentation (Layer 6) • 345

Application (Layer 7) • 345

Encapsulation • 345

TCP/IP **346**

IPv4 • 348

Port numbers • 350

IPv6 • 350

Application layer protocols • 352

Transport layer protocols • 353

Internet layer protocols • 354

Summary 355

Questions 356

Further reading 358

Chapter 12: Report Writing 359

Effective note taking 359

Writing the report 361

Evidence analyzed • 364

Acquisition details • 365

Analysis details • 365

Exhibits/technical details • 366

Summary 368

Questions 368

Further reading 369

Chapter 13: Expert Witness Ethics 371

Understanding the types of proceedings 372

Beginning the preparation phase 374

Understanding the curriculum vitae 375

Understanding testimony and evidence 377

Understanding the importance of ethical behavior 380

Summary 383

Questions 383

Further reading 385

Assessments **387**

Chapter 01 ..... **387**
Chapter 02 ..... **387**
Chapter 03 ..... **387**
Chapter 04 ..... **388**
Chapter 05 ..... **388**
Chapter 06 ..... **388**
Chapter 07 ..... **388**
Chapter 08 ..... **389**
Chapter 09 ..... **389**
Chapter 10 ..... **389**
Chapter 11 ..... **390**
Chapter 12 ..... **390**
Chapter 13 ..... **390**

Other Books You May Enjoy **395**

Index **399**

Preface

Welcome to the world of digital forensics! In this book, you will be going into the depths of the Windows operating system to determine the user's actions on the system. You will also learn about the different filesystems used by the Windows operating system. The role of the examiner is not only about the examination, but also about the report you generate and how you explain your findings. You will learn how to prepare for a digital investigation, including equipment selection, training, and planning a response to the crime scene. It is my hope that this book will be your resource if you are a novice examiner or an experienced examiner.

This book teaches forensic examiners and those who want to become forensic examiners about the various skills and tasks required to be a forensic examiner, completing forensic analyses in either criminal or civil matters. This book will deliver information through the lens of the author's experience in the United States of America so references to criminal matters will involve American law.

Who this book is for

This book is for the novice and experienced examiner in private or public employment sectors. While an understanding of operating systems, file systems is helpful, it is not required.

What this book covers

Chapter 1, *Types of Computer-Based Investigations*, introduces to the reader the different topics of computer-based investigations, from criminal acts investigated by the police to potentially illegal actions performed by an employee or third parties and examined by a non-governmental investigator. While the goal is the same—to present evidence about an incident—the methods of the two slightly differ. It is essential for the reader to understand the similarities, that is, being able to present evidence in judicial proceedings, and recognize the differences, that is, search warrant requirements for a government agent.

Chapter 2, *The Forensic Analysis Process*, details the critical thinking in the planning of providing digital investigative services. This topic will allow the reader to create a strategy to conduct an efficient investigation. The reader will learn to offer different approaches to conduct an investigation depending on the unique set of circumstances for each matter.

Chapter 3, *Acquisition of Evidence*, explains that digital evidence is one of the most volatile pieces of evidence an investigator can handle. The mishandling of digital evidence can severely impact an investigation. Additionally, you may destroy the entire dataset. This chapter will address how to minimize or eliminate these issues when using a validation process to create a forensic image.

Chapter 4, *Computer Systems*, explains that the investigator must control the computer processes while acquiring digital evidence. When dealing with the many combinations of operating systems and hardware, you must implement controls to protect the integrity of the evidence. This chapter will discuss the boot process in detail and identify the most commonly used filesystems.

Chapter 5, *Computer Investigation Process*, explains that being a forensic examiner is much more than pushing a button. Once the evidence has been collected, you have to analyze the dataset. It is not about finding artifacts but rather examining the data and putting it into a context that will either support or not support the hypothesis about the user's actions on the system.

Chapter 6, *Windows Artifact Analysis*, explains that Microsoft Windows is by far the most common operating system today. In this chapter, we will look at the different versions of Windows and will show the reader how to identify and recover common artifacts based on the release of Windows being examined.

Chapter 7, *RAM Memory Forensic Analysis*, covers the analysis of RAM, which is a source of evidence that has recently been recognized as containing vital information about the user's actions on the system. RAM is very volatile evidence and can provide data that cannot be found anywhere else on the computer system.

Chapter 8, *Email Forensics – Investigation Techniques*, discusses email, which is a part of everyday life. This communication vector can be one of the primary communication tools for the majority of the population. These communications can contain incredible amounts of data related to an investigation. The investigator must be able to reconstruct the path that email took from the source to the destination to determine its validity.

Chapter 9, *Internet Artifacts*, explains that using the internet is a daily activity for the majority of the population. Like any other activity, the internet can be used for legal, law-abiding business, or for criminal activity. The internet can be accessed in a variety of ways. The forensic investigator must be able to analyze all these different aspects of the internet to get to the truth of the matter.

Chapter 10, *Online Investigations*, discusses how to use open-source intelligence techniques to learn about the target of the investigations. Also discussed are the steps an investigator can take to hide their true identity and create an undercover online persona.

Chapter 11, *Networking Basics*, explains some of the common network protocols, hardware and models that are being used to connect devices and share information. The ability to understand how information is shared between devices is a critical skill for the online investigator.

Chapter 12, *Report Writing*, covers report writing, which is not the most exciting portion of the forensic exam process. The forensic examiner must be able to explain a technical topic to a non-technical user. As a forensic examiner, you must be able to place that artifact into a context that the audience understands. This ability is a critical skill that you need to master to be a competent forensic examiner.

Chapter 13, *Expert Witness Ethics*, explains that a forensic examiner must be objective, truthful, honest, and perform their due diligence when conducting an examination. The examiner will be providing testimony that may result in someone losing their freedom. The ultimate goal of the investigation conducted by the forensic examiner is to provide testimony or evidence in a judicial or administrative proceeding to stop the cybercriminal's activity.

Download the exercise files

You can download exercise files for this book from at `https://github.com/bill-lcf/Learn-Computer-Forensics`.

Employed academic faculty can also download PowerPoints for each chapter and a question bank after validation. Send an email to `verify@learncomputerforensics.com` from an `.edu` email address requesting access. If you do not have an `.edu` email address, please send proof that you are an instructor.

Once the files are downloaded, please make sure that you unzip or extract the folder using the latest version of:

- WinRAR / 7-Zip for Windows
- Zipeg / iZip / UnRarX for Mac
- 7-Zip / PeaZip for Linux

We also have other code bundles from our rich catalog of books and videos available at `https://github.com/PacktPublishing/`. Check them out!

Download the color images

We also provide a PDF file that has color images of the screenshots/diagrams used in this book. You can download it here: `https://static.packt-cdn.com/downloads/9781803238302_ColorImages.pdf`.

Conventions used

There are a number of text conventions used throughout this book.

`CodeInText`: Indicates code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles. For example: "Outlook stores email information in several file types, such as `.pst`, `.mdb`, and `.ost`."

A block of code is set as follows:

```
MIME-Version: 1.0
Content-Type: text/html; charset=us-ascii
Content-Transfer-Encoding: 7bit
```

When we wish to draw your attention to a particular part of a code block, the relevant lines or items are set in bold:

```
{"endpoint_info_list":[{"endpoint":"smtp:badguy27@yahoo.com",
"c_id":"d24c.2d00",
"c_name":"Joe Badguy Smith"},
{"endpoint":"smtp:badguyneedslove@gmail.com",
"c_id":"e80f.5b71","c_name":"John Badguy Smith"},
{"endpoint":"smtp:yahoo@mail.comms.yahoo.net",
"c_id":"624f.10f0","c_name":"Yahoo! Inc."}]}
```

Any command-line input or output is written as follows:

```
$USER$\AppData\Local\Google\Chrome\User Data\Default
```

Bold: Indicates a new term, an important word, or words that you see on the screen. For instance, words in menus or dialog boxes appear in the text like this. For example: "The **MSF** files are **Mail Summary files**, one part of the email."

Warnings or important notes appear like this.

Tips and tricks appear like this.

Get in touch

Feedback from our readers is always welcome.

General feedback: Email `feedback@packtpub.com` and mention the book's title in the subject of your message. If you have questions about any aspect of this book, please email us at `questions@packtpub.com`.

Errata: Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this book, we would be grateful if you reported this to us. Please visit `http://www.packtpub.com/submit-errata`, click **Submit Errata**, and fill in the form.

Piracy: If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at `copyright@packtpub.com` with a link to the material.

If you are interested in becoming an author: If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, please visit `http://authors.packtpub.com`.

Share your thoughts

Once you've read *Learn Computer Forensics, Second Edition*, we'd love to hear your thoughts! Please `click here to go straight to the Amazon review page` for this book and share your feedback.

Your review is important to us and the tech community and will help us make sure we're delivering excellent quality content.

1

Types of Computer-Based Investigations

Welcome to the 21st century, where almost everything in life is connected to an electronic device. There are digital cameras inside doorbells; your smartphone tracks your daily progress from work to home and back again; you get social media updates when you go to the gym, a show, or travel to a new city.

Your phone calls, bank access, and medical appointments are tracked via digital technology. If it tracks your mundane daily activity, what about criminal or unethical behavior? Of course, that activity is also followed, and if you are a digital forensic investigator, you must know the repositories of the digital evidence and how to analyze it. All activity, benign or criminal, will most likely generate some sort of digital evidence. As an investigator, it is your job to locate all data of interest, process it, and present the evidence to the finder of fact. This chapter will introduce you to the different topics of computer-based investigations, from criminal acts investigated by the police to civil and potentially illegal actions performed by an employee, and an external third party examined by a nongovernmental investigator.

While the goal is the same, to present evidence related to an incident, the methods for evidence gathering and for evidence presentation are slightly different. Therefore, you need to understand where there are similarities and where there are differences.

The topics that will be covered in this chapter are as follows:

- Differences in computer-based investigations
- Criminal investigations
- Corporate investigations

Introduction to computer-based investigations

This book is all about introducing a beginner to the realm of digital forensics. What is digital forensics? It is a division of forensics involving the recovery and analysis of data that has been recovered from digital devices. At one time, the term *digital forensics* was treated as a synonym for computer forensics, but now it involves all devices capable of storing digital data. No matter what term is used, the goal is to identify, collect, and examine/analyze digital data while preserving its integrity. Digital forensics is not only about finding the artifact; it is a formal examination/analysis of the digital evidence to prove or disprove whether the accused committed the violation.

It is not always about demonstrating that the suspect is guilty; as a forensic examiner, you also have that ethical obligation to find **exculpatory** evidence that will prove the subject's innocence. In addition, you must be an unbiased third party in presenting the investigation's findings. In a criminal examination, your findings could deprive someone of their liberty, and in a corporate investigation, your findings may lead to a criminal investigation or cost someone their livelihood. As a digital forensic examiner, your conclusions can have an extraordinary impact on the subjects of the investigation.

To be a digital forensic examiner, you need to have a desire to ask questions, have specialized equipment, and have the required training. From teaching people interested in the field, I have found the best students can critically examine the facts and circumstances being presented and, using that ability, can focus their efforts on efficiently reaching an accurate conclusion. Unfortunately, I find many students want to use a "find evidence" button, find all the artifacts, and print up a thousand-page report and call it a day. That is not digital forensics.

Digital forensics is not finding the artifact. By artifact, I am talking about an incriminating Google search in browser history, an incriminating email between the subject and a co-conspirator, and illicit images found in the filesystem. Artifacts are breadcrumbs leading to the identity of the person conducting the illegal activity. However, on their own, they do not identify the user who created these artifacts or the one who is responsible for their creation indirectly. One of the biggest challenges in this field is identifying the user who is physically operating the device. You want to tie the user to the specific subject, and to do that, you have to analyze – that is the keyword – the digital evidence to associate it with a particular user.

If you are in the IT field, you will understand networking and computer operating systems, but you will lack knowledge of how to preserve evidence, maintain a chain of custody, and present it in criminal/administrative proceedings.

If you are an investigator, you will understand the chain of custody, evidence preservation, and testifying in criminal/administrative proceedings. However, you may lack experience in the digital field. To be an effective digital forensic examiner, you must be part of both those worlds. You must understand how data is created, shared, and saved in the digital realm and preserve that evidence in a forensically sound manner and be able to testify in proceedings. Sometimes, the ability to talk in front of a large group while answering challenging questions posed to you by attorneys from both sides is the hardest part of the field.

As with any field, the way you get better and more effective is to practice, conduct real and mock examinations, receive training, and have the willingness to reach out to your peers for advice. Since you are reading this book, you are taking that first step. You could be reading the text on your own, using it as a textbook for a college course you are taking, or using it in a corporate training session. The reason does not matter. Reading this book will put you on the road to becoming a more effective digital forensic examiner.

What is cybercrime? What crimes does a digital forensic examiner investigate? A digital forensic examiner may investigate any alleged wrongdoing that touches the digital world. Nearly everyone possesses a mobile device. Sometimes, a person owns or uses multiple mobile devices, laptops, and the traditional desktop. All of these sources can maintain a significant amount of information related to the investigation. For example, I investigated a crime against a person where the victim was physically unable to communicate with the police. How does that become a crime that requires the use of a digital forensic examiner?

Well, in this case, she had maintained communication with the suspect of that crime via a website and instant messaging on her mobile device. So, while they did not directly have evidence relating to the crime being investigated, they had evidence about the relationship between the victim and the suspect. In the 21st century, almost any crime may have evidence stored in a digital format. Now, there are some crimes where someone will have used their computer as a tool to commit the crime, such as sending harassing emails, fraud and forgery, hacking, corporate espionage, or the trafficking of illicit images. Your occupation will dictate your response to a situation; if you are law enforcement, you will have one set of procedures to follow, while if you are in the corporate world, you will have a different set of procedures to follow. While some processes may overlap in different fields, each one has its unique differences, which is what we will discuss next.

Criminal investigations

As a law enforcement professional, your first consideration will be officer safety. Is the scene **safe and secure** to process and secure evidence? When the investigation starts, you may participate in one or more roles. The most basic positions are as follows:

- The first responder
- The investigator
- Crime scene technician

Depending on the size of your agency, you may fill one position or all three, and you may report to one or more supervisors. Now, with digital evidence, the person in charge of the crime scene should know the fragility of digital evidence. That allows personnel to enact the proper procedures to ensure that the evidence is not corrupted.

Let's talk about what each role does.

First responders

The first responders are the first ones on the scene. They secure what may be a chaotic scene. They will identify the following:

- Potential victims
- Witnesses
- Potential suspects
- How best to maintain control

They will do this until the investigator arrives. The first responder's primary mission is to make the scene safe and secure and ensure that no one can contaminate the evidence. As you can imagine, crime scenes can vary from a dynamic crime scene to a relatively static crime scene, depending on the nature of the crime. In both scenarios, the first responder must have basic knowledge of what items could contain digital evidence when they secure the scene. We would not want subjects grabbing cell phones or laptops and using them for any activity.

So, how does a first responder protect the crime scene? Like you see in TV shows and movies, yellow crime scene tape is the most common method. It is the most straightforward visible sign of a crime scene barrier, and in our culture, people recognize the barrier being presented by that thin piece of yellow plastic. One or more personnel will have to monitor the crime scene to regulate who can cross that line and enter the scene.

Investigators

The investigator will respond to the scene after being requested by the first responder. Upon arriving at the scene, the first responder and the investigator will coordinate, and information sharing will now start. The first responder will provide the basic information, which typically involves the five Ws and one H, specifically the who, what, when, where, why, and how, about the incident.

The first responder will also provide information about any actions they or anyone else had taken before the arrival of the investigator. For example, the investigator will want to know whether the first responder(s) touched anything, moved anything, or changed anything within the crime scene. This could be a physical action such as applying first aid to a victim or turning a computer on or off. I remember an examination I did where the first responders did not reveal that they had accessed the victim's computer. While conducting my examination, I did a timeline analysis and saw an abnormality in the activity after the victim had died. The abnormality was caused by the unreported actions of the first responders. What's important to understand here is that the first responders' actions were not wrong. What created complications was that they did not report the actions, which led to additional work and explanations.

The investigator takes charge of the scene and directs all activity. They will direct the other team members' investigative efforts to ensure the proper documentation is completed regarding the seizure of evidence. Sometimes, the first responder will seize evidence and turn it over to the investigator. A chain of custody document must be completed and maintained showing who found the item and who maintained control until the completion of the judicial or administrative proceedings.

Crime scene technician

Finally, we come to the crime scene technician. This can be a sworn or unsworn position within the law enforcement agency. They have specialized training in the collection of evidence. This could be physical evidence, such as fingerprints, tool comparison, the collection of biological fluids, and crime scene photography, all of which require specialized training and equipment. The collection of digital evidence requires the same level of expertise that the collection of physical evidence does.

Note

We can put law enforcement jobs into two basic groups.

Sworn: May take an oath to support the laws in their jurisdiction; they have the power to make arrests and carry firearms.

Unsworn: May take an oath but do not have powers to arrest. These positions are typically crime scene analysts or law enforcement support technicians (this will be dependent on your jurisdiction).

The crime scene technician is responsible for preserving evidence and starting the chain of custody. Some actions they could carry out include acquiring the volatile memory of a computer system, creating forensic images of the storage devices, or creating the logical forensic image of logical files from a server. Next, the evidence will be bagged, tagged, and transported to a secure location. What do I mean by *bagged and tagged*? They will place the physical evidence or the items holding the digital evidence in the appropriate storage container. A tag will then be filled out with identifiers to specify which investigation the evidence belongs to, who collected it, and what evidence is contained within the container.

As we go through the rest of this book, we will cover the duties of the crime scene technician in greater detail.

A law enforcement officer may be a first responder, investigator, or crime scene technician and, in all roles, is an agent of the government. Depending on your jurisdiction, the government may restrict how and when the property can be seized and searched. I will discuss the judicial process in the United States; your locality may have different laws and procedures.

In the United States, a citizen's rights to privacy are protected by the fourth amendment of the US Constitution, which states the following:

> *"The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized."*

At a basic level, this means that before the government can seize any evidence, there must be (a) a search warrant based upon probable cause or (b) the owner's consent. The consent given by the owner must be willingly given and must be able to be revoked, which can create an issue in some jurisdictions where the processing of digital evidence can take months and, in some jurisdictions, years. If the owner revokes their consent or refuses to give it, what options does law enforcement have? A search warrant.

How does a member of law enforcement get a warrant? As we learned from the preceding passage, it must be based on probable cause. The definition of probable cause is a reasonable standard that the applicant must reasonably believe that the items being searched for are at that location. Who determines what is reasonable? This would be the judicial official, such as a judge, Justice of the Peace, and so on.

The law enforcement officer makes the written request while the judge reviews it and will approve/disapprove it. If approved, the law enforcement officer can seize and search the property within the guidelines specified by the judicial official. The law requires only agents of the government to get a search warrant to seize and search property. This process will not pertain to you if you work in the corporate world.

Now, let's talk about some potential crimes someone might call you to investigate. This will be a high-level overview of the crime itself. Later in this book, we will address the specific artifacts we should analyze to determine whether criminal actions occurred.

Illicit images

Nearly everyone is connected to the many different forms of digital networks via our mobile devices, tablets, laptops, and computers–we are always connected in one manner or another. Depending on who you ask, it is either the best thing in the world or the worst. There are some excellent aspects; social media allows people/family members to stay in contact, no matter where they are in the world. The totality of the world's knowledge is just a few clicks away. You can read news reports from portions of the world that you previously did not know existed. It is an adventure waiting to happen. Now, it is not all unicorns and rainbows out there. Like any society, there are dark and dangerous portions of the internet where you should be hesitant to travel. That includes the sourcing and sharing of illicit images. For our purposes, an illicit image is an image whose subject matter is offensive or illegal, depending on your cultural or legal landscape.

Before the advent and widespread use of the internet, trafficking in illicit images was almost eradicated, so what changed? The consumer of illicit images no longer had to be physically present to pick up the physical images. The internet allows the user to be relatively anonymous and access illicit images with minimal exposure. I have read reports stating that the high-speed data network that most of us enjoy is because the consumer wants faster throughput speeds to download illicit images.

Consumers of illicit images have free access to terabytes of data with simple clicks of the mouse. If the consumer wants higher quality or a specific subject matter, then it is not a complicated process to find a vendor to meet the consumer's needs for a price.

Your jurisdiction will determine what is or is not an illicit image and the level of criminality associated with the contraband images' possession and/or distribution. I will not differentiate or specify a subject to define illicit images. Instead, I will discuss them using the generic title of illicit images or contraband images. You can use either phrase depending on what may be legal/illegal in your jurisdiction.

How do people share contraband images? At a basic level, a file is a file. A JPEG image of a sunset does not differ from a JPEG image of a contraband subject. Anyone can use any aspect of the internet to share files–the content of the files is irrelevant. If the system allows the user to share data, then the contents of those shared files can be legal or illegal content. Let's look at some media through which illicit images could be exchanged.

Email-based communications

Email is one of the easiest ways to share information through files between two or more people. An email address does not automatically point to a specific user. Some service providers actively advertise anonymity for users of their email accounts. The service provider states that they do not save transactional information, such as source IP, dates and times of connection, or billing information. The service provider may be located outside of the jurisdiction investigating the contraband, which will allow the service provider to ignore the judicial paperwork requesting the subscriber information.

Newsgroups/USENET

This is one of the first components of the internet and has fallen off the radar for the everyday user. Initially, the internet comprised the World Wide Web, with components such as web browsing, email, and USENET. Web browsing and email are known by nearly every internet user, while USENET has faded out of public perception. However, this does not mean it is not being used. USENET is like the old bulletin board system, where you had specific groups, and users could post messages, attach files, and other users could download the files and comments. The user can post just a text message or attach a file to the message. This attached file is known as a binary.

This USENET attachment will be a file type, such as digital images, video, audio software, or any other file type a user can access. The user must use a newsreader to access USENET. There are free and paid versions of newsreaders available in which the user can subscribe to a USENET service. Just like the email service providers that we discussed earlier, one selling point for USENET service providers is anonymity; they explicitly state that they maintain no user transactional data or billing records or they are in jurisdictions whose laws may not adequately address the contraband contained on the server:

Figure 1.1: Unison application

The preceding screenshot shows you the Unison program running on macOS and accessing the service provider Astraweb.

Looking from left to right, you can see the hierarchical system used by USENET. I have selected **alt** in the far-left column, which then populates the next column with many named folders. The folders' naming convention shows the subject of the group. I have selected **binaries**, which means I am looking for attached files to the postings. We can see folder icons in the third column and a brown folder icon with papers coming out the top. The folder icon shows that additional groups are contained within, while the brown folder icon indicates a newsgroup.

As you can see from the preceding screenshot, there are a variety of subjects for the user to explore; some groups may or may not contain contraband images/files. Your jurisdiction will determine what is legal or not as you conduct your investigation.

Peer-to-Peer file sharing

Peer-to-Peer (**P2P**) file-sharing is a decentralized method of file sharing. In traditional file sharing, a server hosts the file, and the client accesses the server to download the file. In the early days of Napster and music sharing, this became a liability for copyright violations. The service provider was served with judicial processes and was liable for hosting a directory of copyrighted files.

In response, the P2P method was changed; no longer was a centralized database created, but instead, users were able to directly search for other users' shared folders on the network. Users connected to a shared network and acted as servers and clients. In P2P file sharing, when users identify a file they want to download, the software reaches out to the other users who possess the desired file. Each user then provides a piece of the file to the recipient. When all the pieces are collected, the software returns them to the original configuration. The user could then participate as a node (the term "node," when discussing P2P, refers to the user's system connected to the P2P network and sharing files) and start sharing the file they just downloaded:

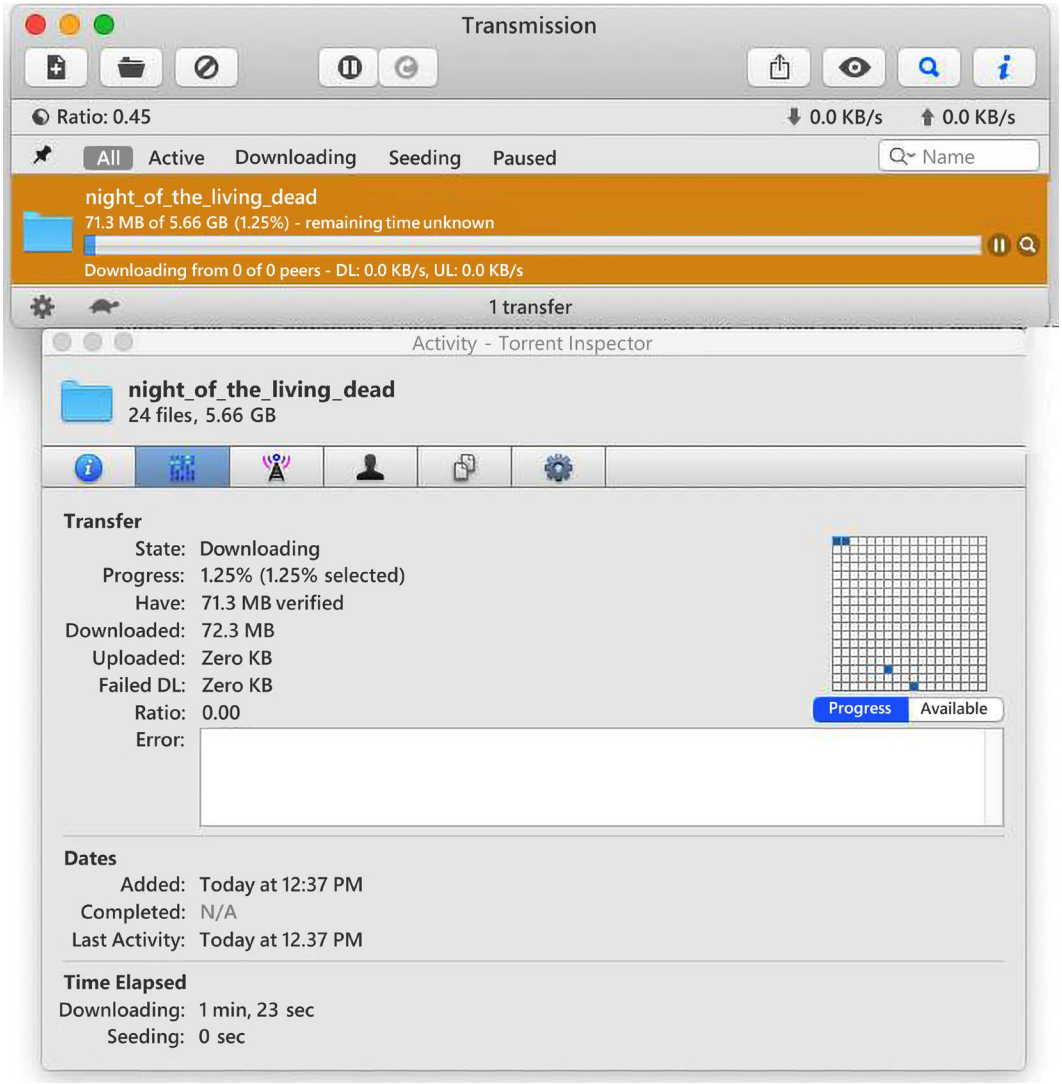


Figure 1.2: Transmission application

The preceding screenshot shows the **Transmission** program running on macOS. I am downloading a movie from the public domain (`archive.org`), and in the bottom portion of the preceding screenshot, you can see that the file has been broken into much smaller bits. The highlighted bits show which parts of the file I have downloaded. Later, we will go into much greater detail about P2P file sharing and the artifacts left in the filesystem.

The crime of stalking

For all of the good that the internet provides, it also provides a conduit for people to exploit, harass, and bully others. The victim could be known to the subject or could have interacted with the victim's online persona in some manner and felt the victim had wronged them. A lot of the bad behavior we see with online activities is because of the anonymity that the internet provides the attacker/subject. When eyes are watching or when we know the attacker's true identity, they change their behavior to conform to societal norms. Unfortunately, it takes time for society to recognize the criminality of specific actions via the digital medium.

Cyberstalking or cyberbullying is now being regulated and considered an actual crime. Depending on your jurisdiction, the definition will vary, and what resources the government will spend on prosecuting these crimes will differ. Remember, the user's identity at the other end of the digital world can be challenging to prove to the high standard required by a court of law.

According to the National Center for Victims of Crime, `https://web.archive.org/web/20201028110630/https://members.victimsofcrime.org/our-programs/past-programs/stalking-resource-center/stalking-information`, historically, in the United States, almost 1,500,000 people, the majority of them women, have been victimized, harassed, and bullied via the digital medium, with the attacks lasting more than two years. In addition, the attacks increased in length if the participants had been intimate partners.

The impact of this criminal behavior is immense; the victim may lose time from work, may have to move residences (several times, sometimes), and potentially suffer from the physical and mental effects such as the anxiety and depression that come from being targeted. In addition, the ability to stalk a former intimate partner in the digital world opens the door to the ability to inflict significant violence on a former partner and, in some cases, bring about their death.

What behaviors can make up cyberstalking? Generally cyberstalking is where the stalker engages in a series of actions, which can cause the subject of the efforts to be fearful and concerned about their well-being. An example of this is where a terminated employee has sent manipulated, compromising images of their supervisor to members of the organization and the general public. This activity continued for months before it was stopped. Despite the harassment ending and the perpetrator being identified, the supervisor still felt the need to leave their job, change their name, and move to another community.

So, where do we begin in our attempts to investigate this crime? The interview will be the best starting place. Asking the victim if they know or suspect who may be behind the harassment is the first question asked.

In my experience and most of the time, the victim will have a general idea of who the harasser is, especially if it is a former intimate partner. Now, some victims may suffer from mental health issues that could complicate the assessment. As an investigator, you must listen to the whole story to understand the totality of events. Just because someone may appear paranoid does not mean that their concerns or fears are unfounded. As an investigator, you must have an open mind and not allow your preconceptions to make you miss evidence or indicators that may be visible.

If the victim has an idea of who the harasser may be, make sure you record all the pertinent information they can provide you. Names, addresses, usernames, email addresses, screen names, and social media locations will all give you valuable information so that you can start your investigation.

Establish the method of the harassment and when it started. For example, was it a Facebook group? Snapchat? Text messages? Chat rooms? Is a mobile device involved in text messages, missed calls, and more? Has the harassment gone old-school with the use of the post office with physical letters?

Threats of violence may increase the severity of the crime and should not be discounted.

The investigator will need to ensure they get forensically sound copies of the digital evidence to start the investigation. This creates the chain of custody of the digital evidence and is the beginning of the investigation.

We will go into much greater detail about the specific artifacts found in digital evidence, but once you have account usernames and IP addresses that the attacker is using to facilitate their attacks, you have a starting point to identify them.

In the United States, a subpoena is required to obtain subscriber information. This information includes the user's first and last names, physical address, how often they access the account, and the IP address used to access the account. It varies between service providers as to how long this information is maintained. Sometimes, it could be as little as weeks or as much as years, depending on the provider. You can also submit legal paperwork asking them to "freeze" the account so that the user cannot disable it or delete any incriminating information.

To gain access to the information contained within the account, such as email content, contents of messages, or anything having to do with content, a search warrant signed by a judge will have to be served on the service provider. If the service provider is within the same jurisdiction as the judicial authority, there are typically no issues. However, when the service provider is in another jurisdiction within the United States or a jurisdiction outside the borders of the United States, this is when the process becomes much more difficult, and sometimes it's impossible to proceed.

Some subscriber information you get may or may not be accurate. It is not unusual for a user to complete the registration forms with false information. But what you can do, for example, if you have an email address, is you can do an open-source search and see whether the user used the email address anywhere else. For example, some online forums will use the email address as a username, and if so, the user may post identifying information in their communications with the other users. That forum now becomes a source of information for which you can issue a subpoena to get the subscriber information.

As you can see, following breadcrumbs of information may lead you to sources you never even considered. Moreover, it can be quite complicated and time-consuming.

Criminal conspiracy

Criminal conspiracy and digital forensics: how do these aspects intersect in the world of the digital forensic investigator? First, let's define what a conspiracy is: when two or more people agree to commit an illegal act. However, just deciding to commit the unlawful act is not enough; actions also have to be taken to further the conspiracy. What does all that mean? For the physical crime of robbery, criminal A contacts criminal B to discuss robbing victim C. The conversation between criminals A and B does not meet the statutory definition of a conspiracy. However, suppose criminal A paid criminal B and agreed on the amount of funds in exchange for the service of the robbing of victim C. In that case, we have an act in furtherance of the conspiracy to commit robbery. So, what crimes can the digital forensic investigator find within the digital realm? Almost any crime imaginable. Let's take a look at an example of such a crime:

> *"Michelle Theer was convicted of a crime against a person. She conspired with John Diamond to commit the crime against her husband, Marty. Investigators had no direct evidence, no physical evidence, and no eyewitness evidence, but they had digital evidence showing the conspiracy to commit the crime. Investigators recovered over 80,000 emails and instant messages between Diamond and Theer that showed a personal relationship between the two and the messages showing the conspiracy between them to commit the crime."*

You can read about this case in more detail at `https://caselaw.findlaw.com/nc-court-of-appeals/1201672.html`.

Now more than ever, people are connected to their devices for their everyday activities. It is not a stretch of the imagination that criminals also use their devices to help organize their criminal activities. The digital forensic investigator has to know of all potential sources of digital evidence and recognize that the **Internet of Things (IoT)** is an untapped bonanza of digital evidence. What is the Internet of Things?

Home assistance programs such as Siri and Alexa, smartwatches, home security systems, and GPS devices – anything that has an app – might contain evidence and show the criminals' intent to commit the crime. Failure to recognize digital devices can result in significant damage to your investigation. For example, there have been instances where the subject of an investigation was placed in the interrogation room, and the investigator did not recognize the suspect was wearing a smartwatch. While they left the subject unattended in the interrogation room, the subject was able to communicate with their co-conspirators and direct their efforts to destroy evidence and interfere with the investigation. Once the investigators caught on to the subject's actions, they used the smartwatch to show the criminal conspiracy. They used the evidence to generate additional charges for the suspect in custody and their co-conspirators.

Social media is also a source of digital evidence for showing a conspiracy. For example, take the case of Larry Jo Thomas. The government convicted Thomas of committing a crime against Rito Llamas-Juarez. Initially, investigators only knew that a specific type of item harmed Llamas-Juarez. However, as investigators processed the crime scene, a bracelet that was "distinctive" was found and collected as evidence. The investigators examined Thomas's Facebook page and saw a photo of Thomas posing with an item similar to what was used at the crime scene. In a different photo, they found the "distinctive" bracelet being worn by Thomas. While the digital evidence did not directly impact the criminality being investigated, it showed how the subject had the means and had been at the crime scene.

Vehicles are also a source of evidence to prove the conspiracy. New vehicles are connected to the network and have their own Wi-Fi connection and sync data between mobile devices, GPS data, and the vehicle's black box. Potentially, the investigator can show the subjects performing reconnaissance on their targets, meetings between the conspirators at a shared location, or where they have traveled to and returned from using toll passes.

Technology is rapidly changing and advancing as the general population uses technology, and so do the criminals. The general population plans out their day by utilizing technology; criminals also plan out their day of criminal activity using the same technology. I am always amazed when criminals use their mobile devices to plan and execute criminal activity and then take pictures to memorialize their illegal business.

Now that we have learned about criminal investigations, the roles, and the means by which information is being shared, let's move on to the next type of investigation, which is corporate investigations.

Corporate investigations

We will now discuss computer forensics from a civilian or non-law enforcement perspective. Since you are not an agent of the government, the search warrant requirement does not pertain to you. (Your specific jurisdiction may be different.) While you may not have the search warrant requirement, you cannot seize and analyze private property. What do I mean by that? You are the investigator for a large multinational corporation; you have an employee you believe is harassing other employees and may have viewed illicit images on their company laptop. What is the legal requirement for you to examine the contents of the employee's laptop? If you are an agent of the government, the employee has an expectation of privacy. However, as an employee utilizing the company's equipment, in the United States the courts have held that the employee has a limited expectation of privacy on the data in the device.

Important note

This may differ, depending on your local jurisdiction. I was teaching a class in Germany and as I was teaching, the students explained that German law gave an employee a high expectation of privacy. In their jurisdiction, there were specific requirements that had to be met before they could examine an employee's computer.

Other than the search warrant requirement, the corporate investigator's duties are similar to law enforcement's. They still must acquire the evidence, analyze the evidence, and present their findings. They could present their findings in an administrative proceeding or, if necessary, forward them to law enforcement, where they may have to testify in a judicial proceeding. In either case, the digital forensic investigator must ensure that the digital evidence was collected in a forensically sound manner while maintaining the chain of custody of the digital evidence.

If the digital forensic examiner cannot authenticate the evidence, they cannot testify or present it in the administrative/judicial proceeding. The corporate digital forensic investigator also investigates a wide variety of allegations. Typically, they will not be investigating a crime where a person was hurt or killed. However, they can still investigate fraud, forgery, a violation of the company's policies and procedures, corporate espionage, or if they believe an employee has stolen intellectual property or is trying to harm the corporation itself. So, let's now talk about employee misconduct.

Employee misconduct

As a condition of the employee's employment, they must abide by the policies created by their organization. Typically, an employer has an "Employee Handbook" or has a set of policies and procedures that dictate what behaviors are acceptable and which ones are not acceptable. Such policies also include laying out specifications to ensure that the organization treats all employees with dignity and respect in the organization's daily operations. There may be rules that specify an acceptable use of the organization's desktop and laptop computers, and a violation of those rules could result in an investigation analyzing those devices, as we mentioned earlier.

Now, I use the term "policy and procedures," and I have found a large amount of confusion with those two terms, primarily when used together. A policy is a statement from the organization addressing a specific issue, while the procedure is the specific instructions regarding how to accomplish the goals of the policy. For example, the organization could enact a policy to restrict employees from accessing non-organizational emails using the organization's computers. The procedure would have two audiences: all the employees and the IT staff. The procedure would inform the employees of how to access the organization's email while directing the IT staff regarding how to block non-organizational emails from being accessed.

You need to follow some general guidelines as your organization drafts and implements policies and the accompanying procedures, as follows:

- The policy should be simple to understand. Short and sweet – do not overcomplicate it. If there is a way for an employee to "misunderstand" the policy, then they will dispute whether their actions violated the policy.
- The procedure should specify all the steps needed to implement the task outlined in the policy. Don't assume the reader will understand if you are not specific in what you want them to do.
- The organization must inform the employee of the potential consequences of violating the policy.
- The organization cannot implement policies that violate the law.
- The organization must enforce the policies. There have been many investigations I have conducted where multiple employees have violated the policy, but the organization never enforced the policy. If they do not enforce the policy for 51 weeks and then, during the 52nd week, the organization enforces the policy against some employees and not others, how can the employees be held accountable during week 52?
- There must be documentation that the employee knew and understood that the organization implemented the policy and the penalties for violating the policy.

If an employee violates the organizations' policies or procedures, does law enforcement have to get involved? Of course not. It would depend on the violation, whether it was a criminal act, and whether the organization had a responsibility to notify law enforcement. Sometimes, the law may mandate the organization to notify law enforcement if they discover the employee has committed a criminal violation. Make sure you know the statutory requirements in your jurisdiction and communicate with in-house counsel during the investigation.

As a digital forensic investigator, it is not typically your decision to notify law enforcement. Instead, after you consult the organization's legal counsel and C-level executives, they will make that decision. It does not matter whether the investigation relates to a criminal or non-criminal matter for the digital forensic investigator's purposes.

Remember, we treat *every* investigation as if we may have to go to court and testify. While the initial investigation may deal with policy violations, you may discover there have been criminal violations that mandate law enforcement involvement in the inquiry. The prosecution and defense will scrutinize all of your investigative endeavors before law enforcement involvement. If you do not maintain the standards of the investigative process, it could weaken the prosecution.

As a digital forensic investigator for a corporate organization, there are a variety of violations the organization may call on you to investigate. One of the more common incidents is the complaint of harassment or a hostile work environment. This is where one person causes one or more people to be intimidated, harassed, physically threatened, humiliated, or any other activity that makes the workplace offensive. How would you investigate someone for a hostile work environment? After conducting the interviews with the complaining employees, they may provide statements on how the subject created the harassment/hostile work environment, if at all.

Your investigation will determine whether the actions were physical, verbal, or carried out on digital media and the frequency of the offending conduct. Was there a single employee whose behavior was offensive, or is there a culture within the organization? If a supervisor was notified or asked the offender to stop, what resulted from the efforts to stop the offending behavior? The offending employee could send offensive text messages, emails, or instant messages utilizing the organization's communication network. Suppose the alleged behavior occurred on or was facilitated with the organization's devices. In that case, you should be conducting your examination to determine whether there is any digital evidence to support or refute the allegations since the property belongs to the organization, limiting the employee's expectation of privacy. (Remember, this may vary by jurisdiction.)

The investigation can proceed once you have supervisory approval to conduct the digital forensic examination. With the information at hand, you can filter out a large amount of additional data that may be contained on the storage device. To be efficient while dealing with the extraordinarily large datasets in today's high-capacity devices, you have to filter out data that is not pertinent to your investigation. For example, if we deal with harassing emails, you may restrict your examination to only email traffic.

Now, your investigation may grow based on your findings on the initial exam. For example, while viewing emails, you observe the subject sending illicit images to other employees. Your investigation has now increased based on the violation and the potential number of violators. Do not limit yourself to only the suspect's computer; you need to examine both the suspect and the complaining witness.

The complaining witness may have evidence of the offending email, while the suspect may have used anti-forensic techniques to remove the source email from their computer. Or you may find the complaining witness had changed the email to contain offensive material. You want to be as thorough as possible, which dictates an examination of the emails from both the sender and the recipient.

You are not typically called upon to determine whether the conduct was offensive – that is a very subjective determination. What one employee considers offensive, another employee may not. Your job will be to recover the artifacts to allow the fact finder to make a well-informed decision on whether the complaining witness' statement can be substantiated. Human resources or in-house legal counsel will determine whether the employee's conduct was offensive. Your job is to be an impartial third party and present the findings. This could be through an administrative proceeding such as a hearing, or you could make a presentation to a senior executive. Remember that the organization may be held liable when they have been informed of the employee's offensive behavior and did not take action.

Corporate espionage

In the corporate environment, no matter how large or small, there are specifics about your organization you don't want to share with the entire world. For example, you could provide a proprietary widget to another organization or have an exclusive recipe for a consumer food product. In almost every case, your organization provides a service, and they get paid to provide that service. If a competitor could look inside the organization's internal workings, that look may mitigate any advantage the organization has over the competition.

We can define corporate espionage as one organization spying on another to achieve commercial or financial gain. The same tactics that nation-states use against each other are utilized by corporate actors against each other; for example:

- Physical or digital trespassing to gain access to data or information
- Impersonating any employee to gain physical access to an organization's buildings or other facilities
- Intercepting voice or data communications or manipulating a competitor's website
- Manipulating social media against a competitor

Some actions I just listed are not in the digital realm, so how can a digital forensic investigator determine what occurred?

Security

It comes down to physical and digital security. The organization has to be proactive and identify the critical infrastructure that needs protection. Once the critical infrastructure has been identified, the organization can then implement controls for security and documentation. If an attacker is successful, the digital forensic investigator will have to determine how the attacker got past the established protocols. The organization's physical and digital defenses should be multifaceted and not rely on a single aspect. I mean that there should be a mixture of physical and digital mitigation efforts to protect the organization. For example, access control is essential; a locked door could be access control, such as controlling access to the server room. Now, the door could be locked and unlocked with biometrics or a physical token. The organization should maintain the access control logs at an off-site facility.

If the attacker compromised and used an employee's access control token, a digital forensic investigator can analyze the logs and determine which user identity accessed the server room. Implementing digital surveillance recordings will allow the investigator to observe the compromise and decide whether or not it was the employee or an unknown third party. With a digital attack, you will have to analyze the logs from the network security devices, for example, antivirus logs, authentication servers, routers, and firewalls, all of which are detective controls. While a detective control allows you to investigate what occurred, it doesn't prevent the incident, nor is it a deterrent. Access control is about protecting an asset; you control users and prevent unauthorized access.

Threat Actors

You may be the victim of an attack from a threat actor. What is a threat actor? Typically, it's a malicious user gaining access to information systems that belong to another.

You may see the terms "black hat" or "white hat" threat actor, where the color of the hat determines the threat actor's intent.

A "white hat" threat actor is a positive actor. This is a person or persons whose goal is to identify vulnerabilities in the system so that the organization's owner or vendor may correct them. A "black hat" threat actor is someone who is attacking the system with malicious intent; their goal is to violate and exploit the organization's data system. Finally, there is also the "activist threat actor," who is looking to exploit vulnerabilities in the system for political reasons. The attack could be compromising information maintained in the system or a distributed denial-of-service attack on the organization. The following is a table to help highlight the differences:

White Hat	**Black Hat**	**Activist**
They hack into systems to discover the liabilities before the bad actors.	They hack into systems for their own personal gain. (Bad actor)	They hack into the system to expose activities, harass the owner, or to promote a political agenda. (Bad actor)

A bad actor will not only rely on accessing the system through technical means; they will also attack an organization through the employees. This is known as using social engineering, which is what we will discuss next.

Social engineering

Social engineering is another attack that is relatively common in the corporate environment. One aspect is a "phishing attack," where the attacker attempts to trick the user into gaining access to confidential information such as a username and password. Typically, this attack is made via email, where the sender purports to be a bank, or someone in authority, where they're asking the user to provide biographical information, name, date of birth, governmental identification number, username, and passwords.

If the user believes the email and provides that information, the attacker can impersonate the user and attempt to gain a foothold into the organization's data systems.

There are automated tools designed to use social engineering, such as a phishing attack, against organizations. These tools do not require a significant amount of specialized knowledge to implement. The users of these tools are referred to as "script kiddies" and could attack your organization using these automated tools.

The vendors of the tools state they are to be used by the organization to test their defenses, but there is no method to control what the user does with the software once downloaded.

Gophish

Gophish is one such automated tool. It works on all three of the major operating systems and is freely available for anyone to download. It does not require significant installation skills; you can extract it and run the executable, and the program will be up and running. The following screenshot shows the initial login screen when the software is up and running:

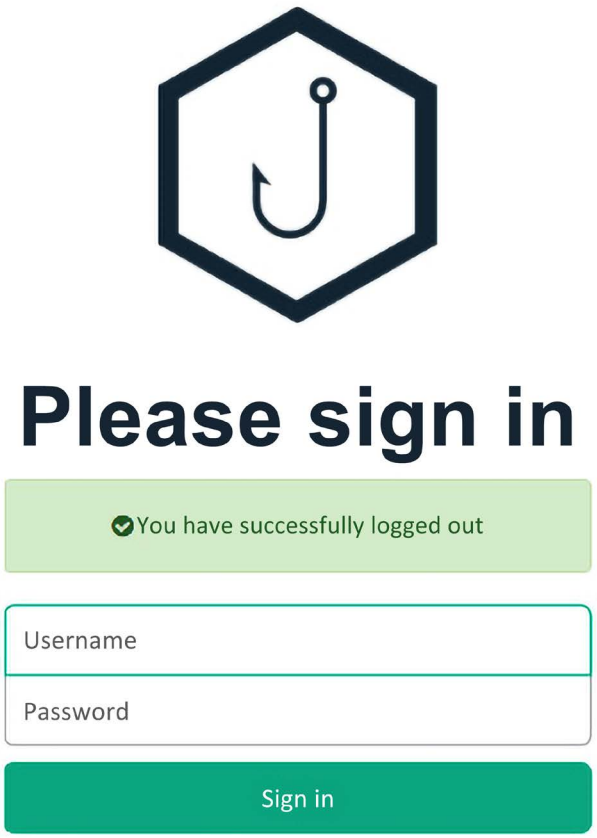


Figure 1.3: Gophish login

Once you log in, you will be presented with the **Dashboard** of the service.

Note

This book is not about running Gophish or any other program; it is merely to give you an idea of what is available out there.

Please follow all applicable laws and regulations.

You can create email templates that you can send out to organizations. You can capture members of the organization's emails using **open source intelligence techniques (OSINTs)** and import them into the program:

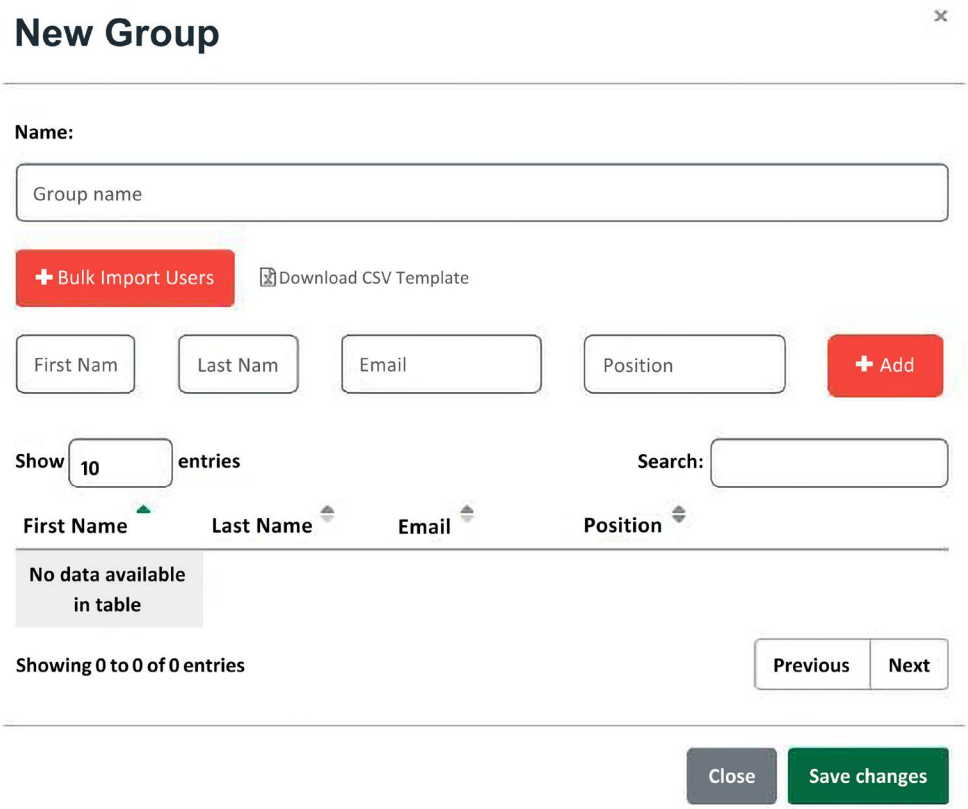


Figure 1.4: Gophish import emails

A common theme when it comes to phishing the user's credentials is to send them an email asking them to reset their password, and when they do so, it directs them to a clone of the official landing page. After the attackers capture the username and password, the user is redirected to the official page, and they never know what occurred.

Real-world experience

One time, I was hired to conduct a vulnerability analysis of an organization. As part of the scenario, they did not provide me with any information about the data network's internal workings or the building's physical security. The building had public access during regular business hours. During normal business hours, I walked around the organization and conducted my reconnaissance to see whether I could identify any vulnerabilities.

To go to the executive levels of the building, I was required to sign in at the security desk and receive a **radio frequency identification** (**RFID**) pass. As I signed in, they did not require me to show any identification or state my business or my destination. I signed in and was given a visitor RFID card and was sent on my way. I took the elevator to the top floor and walked around the executive level. I was dressed in the typical business casual clothing, carrying my laptop case. I found an unlocked training room where I entered and set up my laptop. I plugged into the network and accessed the system. Several employees walked in while I was inside the training room, but none of them questioned why I was there, sitting alone, typing furiously at my computer. I stayed in the room until four hours after the building closed. During that time, no one questioned why I was in there. I packed up my laptop and had full access to the executive level for the rest of the evening.

If I was an actual attacker, how would you be able to investigate what happened? What sources of evidence, maintained by the organization, could you process? The first step would be to identify a potential timeline for what occurred. One control for this vulnerability test was not to damage the network and to access the control file. A control file is a plain document of no value and can be safely manipulated to show unauthorized access. The manipulated file will contain the timestamps to show when the unauthorized access happened. The timestamps will give the investigator a starting point for starting the investigation.

This will be achieved by examining server logs and firewall logs and identifying my digital footprints within the network. Once they identify the physical device location where the compromise occurred, they can review the surveillance footage to work backward on how I gained access to the executive level, the RFID-protected elevator, and the physical security log I completed. Typing out the reaction to the compromise in the system does not address the enormity of the task facing the digital forensic investigator.

If the organization identifies the compromise within a timely fashion, that makes the investigation more straightforward, but consider if the compromise isn't recognized for days, weeks, or months. How hard would it be to determine what occurred months later, after the compromise?

Consider the compromise of Sony Pictures in 2014. While the exact duration of the attack is unknown, the attackers spent at least two months inside the network copying files, with some reports saying the attackers had access to the internal network for a year. Although it has never been confirmed, the attackers claim to have compromised and transferred over 100 TB of data from Sony Pictures. The compromise of information was not the only vector of attack; they made employees' computers inoperable and compromised some of the organization's social media accounts. In addition, the organization's employees were also victimized by the compromising of their personal information by the attackers.

Insider threat

An organization cannot assume the attack will come from an external threat. While the design of most protocols and mitigations is to safeguard the organization from the external threat, the internal threat can be more dangerous than the external threat. No longer can the organization rely upon outward-facing security such as firewalls, building access control systems, intrusion prevention systems, or intrusion detection systems; they must also assess and monitor internal vulnerabilities to mitigate the threat from the inside. This is not an easy task; the insider threat has knowledge of the security protocols, policies, and potential vulnerabilities that the external threat does not.

In 2016, almost 1/3 of all electronic crimes were known/suspected to be caused by an insider threat. The damage caused by the insider was more significant than an external attack. No sector is protected from the internal attacker; if you are a US federal agency or a defense contractor, the government requires you to create a formal insider threat program, which is not surprising since there have been nearly 100 insider threat incidents within the last ten years. (We are not talking about espionage incidents.) Almost 3/4 of the insider attackers were actively employed by the federal agency, while 1/3 were not directly employed, such as a contractor or an employee of another agency. Most of the federal cases dealt with fraud and were committed by the insider for financial gain.

Who typically commits insider attacks? Is it a new employee? A veteran? Remember, for an insider attack to be effective, the insider must be trusted. If we look at the federal government sector, nearly half of the insiders had been with the organization for over five years, with most of them abusing their access and creating fraudulent documents.

Now, in the information technology sector, the demographics of an insider attack are a bit different. Nearly 75 percent were former employees and were with the organization for less than a year. In addition, almost 20 percent did not have their accounts deactivated when they left the organization. That means they could use their credentials to access the confidential information, despite leaving their employment.

As an investigator, this should be a warning that there is an issue with that organization's policies and procedures that must be immediately corrected.

Having a procedure at hand to deactivate an employee's account either before termination or shortly after they give their resignation would have stopped 1/5 of the documented attacks.

Investigating an insider threat will be difficult. You are dealing with people/employees who, at some level, have gained the trust of the organization. The investigator has to try and determine what the insider's mindset is underneath the persona that is being shown every day. Are they an opportunist? Are they a disgruntled employee? Are they someone out for revenge against an executive? Those are the potential attackers you may have to deal with. You want to create the groundwork before the attack happens.

Various sections of the organization – Human Resources, Legal, and IT – will be part of planning any potential response as well as being part of the response. The response team will identify who may be involved in an insider threat, such as the following:

- Executive staff
- Directors
- Employees with access to data

If you have to identify any potential data source(s) for when we have an investigation, you will need to examine the following:

- Company-issued laptops
- Company-issued tablets
- Cell phones or mobile devices
- Any cloud account access

You will have to correlate the user and the user's devices with access to the critical data, and the team will have to identify the critical data beforehand. When should insider threat investigation be initiated? Typically, this will start with a notification from Legal or Human Resources. The organization could also implement a policy investigating when an employee leaves the organization.

If the employee's position gives them access to sensitive or privileged information, then a review of their activities within the organization should be conducted. This could start in a broad sense; you are looking to gather data from mobile devices, laptops, desktops, and potentially the cloud. Then, you take that dataset and filter it to reflect access to the critical information.

Once the employee has resigned or the organization has decided to terminate the employee, the data collection process should start. The data collection process should begin before the employee is told they will be terminated. I recommend that the organization collects between 30 and 90 days' worth of activity for the employee. The more data is acquired, the better informed the investigator will be of the employee's actions. Some of the artifacts that may help determine whether the employee has exfiltrated data are as follows:

- USB devices
- Cloud accounts
- Sharing of files via social media
- Burning a CD/DVD

You will also analyze the activity around the critical data. This should be a standard activity so that there is an understanding of what is normal. Then, you must monitor the data to get that normal baseline to understand when the unusual traffic occurs. For example, you could monitor the traffic to the critical data, and suddenly, access to that data spikes. Does an attack cause this spike, or is it normal because it is the end of the pay period and accountants access the data as part of standard processing?

Another example could be whether the data is accessed after regular business hours. Is there a legitimate reason for that access? These are the circumstances that need to be identified before the investigation starts. This foreknowledge will allow you to filter out all the baseline information and focus only on that data outside of the norms.

The investigation may show no malicious intent or indicate there was malicious intent. Either way, you report the findings to the team to determine the next steps. This could lead to a review of policies and procedures and new controls to mitigate future attacks.

How effective is digital evidence when used in criminal or civil proceedings? There are many variables in play during the trial, from the jury members (if there is one) to the ability of the lawyers to present the digital evidence in the most favorable light to help them accomplish their goal. Then you must consider the expert witnesses that will testify about the digital evidence.

The effectiveness of the expert witness to explain a highly technical subject to a non-technical audience is going to be critical.

Case studies

The following case studies are snapshots of what you may see during administrative or judicial proceedings. Be advised that there will be many proceedings where the court (or an official of the proceeding) will not release the digital evidence to anyone outside of the proceeding. Potential explanations can include the digital evidence that will contain contraband, such as **child exploitation material** (**CEM**), also known as **child sexual abuse material** (**CSAM**), or it may contain sensitive information, and the court has ruled to keep the material private.

Dennis Rader

One of the first national cases dealing with digital evidence I became aware of when I started my forensic training was the **Bind Torture Kill** (**BTK**) serial killer Dennis Rader.

Initially, as a youth Rader had sexual fantasies about women that he considered trapped and helpless. Rader also exhibited other troubling behavior such as killing and torturing small animals and voyeuristic behavior by spying on female neighbors. When Rader reached adulthood, he dropped out of college and joined the United States Air Force for four years. After being released from active duty, he moved to the Wichita, Kansas area. Rader was soon married and ultimately had two children with his wife.

Rader had a variety of employment types, including a security system installer for ADT, an animal compliance officer for Park City, Kansas, and an operations supervisor for the U.S. Census. In addition, Rader was involved in the community as a Cub Scout leader and was elected president of his Church Council.

Rader started his killing spree in January 1974, when he killed four members of the Otero family. The killings were discovered when the children returned home from school. In October 1974, Rader described the killings in great detail in a handwritten letter he placed in an engineering book in the public library. Rader continued his killing during the spring of 1974 until the end of 1977. During this timeframe, he killed three more women. In 1978, the television station KAKE received a letter written by Rader that claimed responsibility for the deaths of the Otero family and of the three women (Kathryn Bright, Shirley Relford, and Nancy Fox). The letter's contents included suggestions for a nickname that the new station could use when reporting on the murders. This is where the BTK nickname originated. A second letter was received by the television station, which demanded greater media attention. Rader killed his last victim, Dolores Davis, in January 1991.

In 2004, Rader started communicating with the local media. Numerous letters and packages were sent to the television station and placed in the community. Some items included identification cards, threats to law enforcement, and dolls posed with the limbs bound in a plastic bag over its head. One item left by Rader included a cereal box that he placed in the bed of a pickup truck that was parked in the parking lot of a Home Depot store. When Rader asked law enforcement about the cereal box, he realized they had not found the box. The pickup truck owner had thrown the cereal box into the trash. When law enforcement went to the parking lot, they were able to recover the cereal box, which contained a question that Rader had about using a floppy disk in his communications with law enforcement. Rader asked if he stored his writings on a floppy disk, would law enforcement be able to trace its origins. Rader told law enforcement to respond by posting a message in the local newspaper with the words "Rex, it will be okay." Law enforcement was able to find security CCTV footage that showed an unidentified man driving a black Jeep Grand Cherokee that stopped near the pickup truck and then the driver walking around the truck.

In February 2005, a television station, KSAS, received a package that contained a Memorex floppy disk, a letter, a necklace, and a copy of the cover for the book "Rules of Prey."

When the investigators conducted a forensic examination of the floppy disk, they were able to recover a previously deleted Microsoft Word document. The embedded metadata contained information about the organization that registered this version of Microsoft Word; in this case, the examiners found "Christ Lutheran Church" in the organization name in the embedded metadata. The metadata also included the last user to modify the document, which was identified as "Dennis." An Internet search identified Dennis Rader as the church council president for the Christ Lutheran Church.

Physical surveillance revealed that Rader owned a black Jeep Grand Cherokee. Law enforcement was able to get a search warrant to collect Rader's daughter's DNA from a Pap smear and compare the DNA found on the victims. The test showed there was a family relationship between the two samples. Rader was then arrested, tried, and convicted. Rader was sentenced to ten consecutive life sentences.

Silk Road

Silk Road was the first online black market hosted on the dark web. This required the use of the Tor browser, which allowed anonymous users to access the vendors without fear of their traffic being monitored by a third-party. The founder of Silk Road was known by the pseudonym "Dread Pirate Roberts," later identified as Robert Ulbricht.

In February 2011, Ulbricht launched Silk Road, taking its name from the historical trade routes between India, China, and Europe and using the Tor network combined with the cryptocurrency Bitcoin for anonymous transactions between anonymous users.

The success of Silk Road led to an article written by Adrian Chen titled "The Underground Website Where You Can Buy Any Drug Imaginable" and published on the website Gawker. As the public noticed, so did law enforcement and the federal government. First, multiple different agencies started their investigations, and the **Federal Bureau of Investigation (FBI)** started a deep examination of the Tor network to identify potential vulnerabilities. Next, the **Internal Revenue Service (IRS)** began to follow the money to understand how anonymous users could purchase services being offered on Silk Road. Finally, the **Drug Enforcement Administration (DEA)** and the **Department of Homeland Security (DHS)** focused their efforts on interdiction by identifying the packages of illegal drug shipments being sent to the country.

The IRS dug deep into the origins of Silk Road. Investigators started researching internet traffic, such as posts to message boards, newsgroups, and discussion forums, looking for information that a user or administrator may have posted at the same time as when Silk Road was open to the public. They were able to unearth a posting about Silk Road to a discussion forum by a user with the username "Altoid." As the investigators started following the history of Altoid, they found a posting that listed a Google Plus account that the investigators traced back to Robert Ulbricht. Unfortunately, there was no evidence linking Ulbricht to Silk Road or even that Ulbricht had computer systems or networking background.

In July 2013, **Homeland Security Investigations (HSI)** intercepted a package that contained counterfeit identification cards that had Ulbricht's picture. The intercepted package was intended to be delivered to Ulbricht's address in San Francisco, California. HSI agents followed up on the counterfeit identification investigation and spoke to Ulbricht. At that time, the agents were unaware of the connection between the Silk Road investigation and Ulbricht.

The FBI continued its investigative efforts to identify any vulnerabilities that could lead to other investigative endeavors to identify the operators and users of Silk Road. The agents were able to locate an IP address that a coding error exposed on the Silk Road website. The IP address returned to geolocation within the country of Iceland. The Icelandic government agreed to cooperate with the FBI and created a clone backup of the server, but unfortunately, the server's contents were encrypted. Ultimately, the FBI broke the encryption, and the server's contents were now available to be examined. Armed with this information, the FBI created a mirror of the Silk Road servers and identified employee information, accounting information, and copies of chats between users.

One chat included the user "Dread Pirate Roberts." The chat contained information that Dread Pirate Roberts had agreed to pay for the murder of an adversary. Additional chats found that Dread Pirate Roberts had often engaged with individuals to pay for them to kill people that were considered to be a danger to Dread Pirate Roberts.

Finally, in July 2013, the different agencies investigating the Silk Road website sat down and shared information. When the IRS brought up Robert Ulbricht's name, the four agencies were able to conduct a thorough background check. The background check identified that Ulbricht had traveled to Dominica, which is used by individuals wishing to hide their monetary proceeds from the US government. They were also able to locate an email address used by Ulbricht, which also matched a user account on the servers.

The FBI then started physical surveillance of Ulbricht and was able to match Dread Pirate Roberts's activity with Ulbricht's activities. In October 2013, they decided it was time to arrest Ulbricht.

There was a concern that Ulbricht could destroy the digital evidence before they took Ulbricht into custody. The FBI waited until Ulbricht went to the public library and opened his computer. The Silk Road had hired an FBI undercover agent as a new employee and sent Dread Pirate Roberts a message to check a post from an admin account that had been flagged. When the undercover agents saw Ulbricht interfacing with his computer, they created a distraction. A male and female undercover agent started a verbal argument that turned physical. When Ulbricht was distracted, an undercover agent came up and took the open laptop and immediately passed it to another agent. They then took Ulbricht into custody without further incident. When the investigators examined the laptop, the investigators found that full disk encryption was active. When agents completed the examination of the computer, the investigators had found nearly 150,000 bitcoins, accounting information for the Silk Road web page, a listing of all the Silk Road servers, and diary entries made by Ulbricht listing the creation and operational details of Silk Road.

The FBI then took down the Silk Road website, and Ulbricht was tried and convicted.

San Bernardino terror attack

Terrorists carried out a coordinated attack in San Bernardino, California, on December 2, 2015. The attack was a coordinated attack using semiautomatic rifles and explosives. A training event and Christmas party hosted by the Department of Public Health was the target of the attack conducted by Syed Farook and Tashfeen Malik. Farook and Malik were married and lived in Redlands, CA. The death count was 14, and 22 people were critically injured in the attack. Farook was born in the United States and was an employee of the Department of Public Health. Malik was born in Pakistan and was a legal resident of the United States.

The investigation labeled Farook and Malik as "homegrown violent extremists." They were not a member of any terrorist cell or terror network. It is believed that Farook and Malik became radicalized before the assault, declaring their devotion to jihadism and martyrdom in private conversations with each other before the incident took place. Farook and Malik had accumulated weapons, ammo, and bomb-making material in their house.

On February 9, 2016, a report from the FBI stated they could not unlock an iPhone 5C, which belonged to the county and was issued to Farook as a part of his employment. The **National Security Agency** (**NSA**) was asked to break into the phone, but they could not do so. The FBI then asked Apple to create a RAM-based operating system to bypass the iPhone's security. Apple declined the request because of its policy never to undermine the security elements of the software. Apple's response caused the FBI to request a United States magistrate judge issue a court order requiring Apple to develop and furnish the software to the FBI. The magistrate judge granted the request.

Apple considered constructing a backdoor of this type a significant security concern to its users and challenged the ruling.

The United States **Department of Justice** (**DOJ**), in response to Apple's denial, then requested the court to force Apple to comply with the court's order. The DOJ informed the court the FBI would deploy the software and would allow Apple to remove the software via a remote connection on the phone.

Apple reported they presented other options to the FBI to access the data stored in the iPhone. However, the FBI's actions had removed one of the more promising methods because of an operational error. When the FBI recovered the shooter's phone, they asked San Bernardino County to reset the user's iCloud account password. This would allow them to access the data stored in the iCloud backup. However, when the county reset the password, the phone could not be backed up to iCloud unless the user entered the passcode on the phone.

On March 28, the DOJ withdrew the lawsuit against Apple because it reported they had unlocked the iPhone. Several scenarios were reported on how access was granted to the phone's data: the Israeli business Cellebrite agreed to assist the FBI, or the FBI paid threat actors to exploit a zero-day vulnerability in iOS.

Apple's refusal to comply with the court order elicited a mixed response from the public. A CBS poll showed that 50 percent backed the FBI, and 45 percent supported Apple.

Theft of intellectual property

It is not only criminal matters that you may come across; civil matters also require a forensic investigator. The following case study is from the firm Cyber Diligence, Inc.

The firm was hired to assist the legal team of a world-renowned scientist who was accused of stealing intellectual property from his previous employer, with the matter being filed in federal court. This matter had a large amount of discovery dealing with computer forensics. There was a concern that the previous employer was abusing the process to keep the client from working with another employer. The previous employer did not have a non-compete agreement with the client. The previous employer wanted to get an injunction because they believed the client had stolen intellectual property. The client's law firm provided copies of all the court documents and discovery, including transcripts of depositions and expert statements. A review of the documents showed the previous employer did not suspect intellectual property theft before they filed the case.

The previous employer did not conduct a forensic analysis on the workstations used by the client before they filed in federal court. The statements of the forensic expert contained inaccurate information, such as the client removed emails from the employer because the modified date time stamps of the OST file showed the file had been modified the day before the client left the organization.

An OST file (`.ost`) is used with Microsoft Outlook. This file allows users to work offline. When the user regains connectivity, they can synchronize any changes with the Exchange server

The expert statement related that this was proof that the client had extracted emails that belonged to the organization. The expert could not think of a valid reason a user would create an OST file other than for illegal purposes. Outlook creates OST files when the client connects to the Exchange server. This is an automated process and not user-initiated. The modified timestamp indicates when the contents of the OST file have been changed, such as receiving and sending emails. An interview with the client was conducted to understand the network configuration of the previous employer. Consultations with the client's legal team helped develop a plan to address the expert statements, which were prolonging the matter and draining the client's financial resources.

The client's legal team requested the digital evidence for examination by the forensic team. The forensic team performed their analysis in preparation to depose the opposing forensic expert. It was also noted that the previous employer's legal team filed the expert's statements, which lacked a digital or physical signature of the expert. When the forensic images of the workstations used by the client were examined and compared to the "facts" presented in the six statements of the expert, they found the findings to be inaccurate. The next step was to interview the defense expert, which was much harder than it should have been. Ultimately, the federal judge, hearing the matter, ordered the previous employer to produce the expert and to make them available to the client's legal team.

Their expert stated he did not make the conclusions found in the reports, nor were they the reports he created. The expert reports filed by the previous employer's legal team were fabricated. The district judge threw out the case on a summary judgment, and they ordered the previous employer to pay for the client's legal fees.

Summary

In this chapter, you have gained an understanding of the different types of issues you may encounter during a digital forensic examination. You have learned how the digital world and the physical world interact and how to use the digital world to help prove or disprove allegations. You have gained an understanding of different procedures and how to collect and manage evidence when investigating allegations of wrongdoing.

In the next chapter, we will discuss the forensic analysis process to maximize the efficiency of your investigation.

Questions

1. Peer-to-Peer filesharing is used to share illegal files only.

   a. True

   b. False

2. What does a first responder identify?

   a. Potential victims

   b. Witnesses

   c. Subjects

   d. All of the above

3. You may find digital evidence in every type of investigation.

   a. True

   b. False

4. Which amendment of the U.S. Constitution protects the rights of citizens from unlawful search and seizure?

   a. First

b. Second

c. Third

d. Fourth

5. What is a “binary”?

a. A star

b. An attached file

c. A USENET post

d. A web browsing artifact

6. What is required in the United States to obtain subscriber information?

a. A search warrant

b. A subpoena

c. Consent

d. Hacking

7. Criminals use social media for illegal purposes.

a. True

b. False

The answers can be found in the back of this book, under *Assessments*.

Further reading

John Vacca and Michael Erbschloe. *Computer Forensics: Computer Crime Scene Investigation*. Charles River Media, 2002 (available at `https://www.amazon.com/Computer-Forensics-Investigation-CD-ROM-Networking/dp/1584500182`)

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

2

The Forensic Analysis Process

We will now discuss the forensic analysis process. As a forensic investigator, you will need to create a strategy that will enable you to conduct an efficient investigation. You also need to make sure you are familiar with your tools and the results that they will provide. Without a process, you will waste time examining data that will not impact your investigation, and you will not be able to rely on your tools. In addition, you want to make sure you get valid results from the tools you deploy. Finally, to be thorough and efficient, you must use critical thinking to determine the best investigation or exam method.

While there are similarities in every investigation, you will find differences that will require you to have an exam strategy to be efficient. I am not a fan of keeping an examination checklist because there will be areas that aren't relevant, such as different operating systems, physical topography of the network, criminal elements, and suspects. These variables ensure that no two examinations or investigations are the same and will require the investigator to execute a different strategy for each of them.

The forensic analysis process is made up of five subsets:

- Pre-investigation considerations
- Understanding case information and legal issues
- Understanding data acquisition
- Understanding the analysis process
- Reporting your findings

The upcoming sections will discuss each of these in greater detail.

Pre-investigation considerations

The pre-investigation is where you determine your capabilities and equipment specifications to conduct a forensic exam, regardless of whether it is in the field or a lab environment. Now is the time to determine your hardware, personnel, and training budget. Some of those costs will not be a one-time expenditure but will be an ongoing budget expenditure. The equipment must be updated, personnel training must be maintained, and the purchase of new technology as it becomes available.

Being a digital forensic investigator is not about buying the equipment, going to a training class, and never updating either of these afterward. As technology changes, so do the methods of hiding data or conducting criminal activities, so the investigator must be ready to adjust to these changes.

Before you are ready to begin the investigation, you must prepare yourself. This will allow for greater efficiency and a better work product. This includes preparing your equipment and becoming familiar with the current laws and legal decisions and the organization's policies and procedures.

Some equipment will be reusable, and some will not. For the single-use items, make sure someone replaces them as soon as the incident concludes.

Note

I cannot tell you how many times I have responded to the scene with my "to go" kit only to find that another detective had already used it and not replaced the consumable equipment. It was my mistake for not checking it before I departed to go to the crime scene, and it was my partner's mistake for not replacing the items.

We will now discuss the equipment you will use as an investigator.

The forensic workstation

Whenever you get forensic investigators together, a common topic of conversation is the forensic workstation. How much **Random Access Memory (RAM)**? How many **Solid State Drives (SSD)** drives? Which **Central Processing Unit (CPU)**? Which **Operating System (OS)**? These are all questions that you might commonly hear. There is always a difference of opinion about the configuration of a forensic workstation. None of the views are incorrect because the investigator's workstation configuration depends on their budget and the cases that are being investigated.

Forensic workstations are not cheap. Depending on the skill level of the investigator, they can either build their own or purchase a pre-made forensic workstation. Several vendors will configure a workstation to your specification. For example, consider the vendor SUMURI (`https://sumuri.com`) and their TALINO workstations. The base model costs approximately $8,000 and comes with:

- Intel Core i9-10900X 3.7 GHz 10-Core LGA 2066 processor
- 32GB of DDR4 2666 MHz RAM
- 500GB M.2 NVMe SSD

That is a basic forensic workstation, and you still must add storage for the forensic images. The high-end version costs over $18,000 and comes with:

- Dual Intel Xeon Gold 5220 18-Core processors
- 128GB DDR4 RAM
- 1TB SSD for the operating system
- 1TB M.2 NVMe SSD for temporary files and processing
- 2TB M.2 NVMe SSD for databases
- Eight 6TB hard drives configured in RAID 10 for evidence
- A 30-series GDDR6 **Graphics Processing Unit (GPU)** such as the NVIDIA RTX 3070 or 3080

One bottleneck that a forensic investigator may face with their forensic workstation is data transfer. I suggest using SSDs because they have much higher throughput than the typical spinning disk does. A fast CPU and a large amount of RAM enable maximum performance for forensic analysis. However, these machines are not portable, and you are not always able to perform the analysis or to acquire the data from the relative comfort of your workstation. A forensic laptop is also an expensive piece of equipment. At the time of printing, the TALINO OMEGA comes with:

- Intel Core i9-11900K processor
- 64GB DDR4 2933 MHz RAM
- 500GB M.2 NVMe SSD for the operating system
- 250GB M.2 NVMe SSD for temporary files and processing
- 1TB M.2 NVMe SSD for database
- 2TB M.2 NVMe SSD for evidence files
- NVIDIA GeForce RTX 3080 GPU with 16GB GDDR6 video memory

Note

You will need to include Gigabit Ethernet on both workstations to communicate on the local area network.

As you can see, you can never have too much CPU, RAM, or storage space on your forensic workstations. The equipment I described is on the higher end; you can conduct digital forensic examinations with less expensive equipment and still achieve the same results. In addition, the more high-end equipment will decrease the time involved. If you are a member of a multinational corporation or a large law enforcement agency, you may have the budget for high-end equipment. A smaller law enforcement agency, a smaller organization, or a single practitioner will have to determine what cost is more appropriate for their situation.

Sometimes you must leave the lab, which means you need additional portable equipment. We will now discuss the equipment required in your response kit.

The response kit

The digital evidence is not always delivered to your workspace. Sometimes, you may have to respond to a third-party location to acquire that evidence. The collection of that evidence is the basic building block for any digital forensic examination you may conduct. Like conducting an examination in your workspace, you need the proper tools and supporting equipment to accomplish this task. You need to create a response kit that includes documentary paperwork, pens, and storage containers to store digital evidence.

A response kit is unique to each digital forensic investigator. No kit is perfect; all kits are always subject to improvement. The goal of your response kit is to have everything you need to collect digital evidence, and we will go over some equipment that, in my experience, I have found helpful:

- **Digital camera**: Capable of still and video recording. You need to document the scene as it was when you arrived. If you testify in official proceedings, you will show the fact-finder precisely what you saw as you arrived. Some organizations also video record all the actions of the digital forensic investigator's activities as they collect digital evidence.

Note

A word of advice: I would disable the microphone so as not to record audio. You may have extended discussions about how to proceed using language that may be regarded as less professional. These discussions and use of language could be used as a distraction by the opposing side in the presentation of evidence.

- **Latex or nitrile gloves**: These protect several aspects of the evidence collection — you are not leaving your fingerprints, and you are also protecting yourself from potential biohazards that may be on the scene. I am talking about blood, urine, feces, and any other biological fluid you can think of.
- **Notepads**: You need to document your actions on the scene. A notepad is a perfect repository to maintain that information. You can take notes about who you talk to, who secured the scene, and the basic facts of the case. When you begin the investigation, a lot of information will come at you, and it could be easy for you to forget a specific action if you do not record it. Some organizations also make a hand-written sketch of where the digital evidence is being collected. Your organization's policies and procedures will determine whether a sketch is required.
- **Organizational paperwork**: This could be a property report for seizing evidence, and it lists exactly what was taken, where it was taken from, and any specific identifying marks or serial numbers on the item being taken. You can also include labels or tags to identify items that contain digital evidence.
- **Paper storage bags/antistatic bags**: You have to put the containers of digital evidence somewhere to prevent any unauthorized access. Digital evidence is very fragile, and you want to make sure you do not store it in a manner where static electricity can be generated. Static electricity can render the storage media inoperative, and you will lose access to any data.
- **Storage media**: Hard drives can be a traditional spinning disk or SSD and USB devices. A corporate digital forensic investigator will not shut down a server to create a forensic image. Instead, they will collect the specific datasets in the form of log files, RAM, or user directories and store them on the appropriately sized storage media.

- **Write blocking devices**: This could be a hardware device, such as the Tableau TK8u USB 3.0 forensic bridge (`https://security.opentext.com/tableau/hardware/details/t8u`), which allows you to access a storage device without changing its contents. We will discuss the acquisition of evidence in much greater detail in *Chapter 3*, *Acquisition of Evidence*. Alternatively, you can use a forensic boot disk, such as SUMURI's PALADIN, a Linux distribution based on Ubuntu that allows the collection of digital evidence in a forensically sound manner. SUMURI offers PALADIN as a free download at `https://sumuri.com/software/paladin`.
- **Frequency shielding material**: This could include commercial aluminum foil, Faraday bags, or any container that will block radio transmissions. You will use this when you seize a mobile device to prevent the user from remotely wiping or resetting the device. Be aware, however, that when you place the device in these containers, the battery will quickly deplete, as it will attempt to reconnect to the network. If you have access to the mobile device's menu, you can put the device into airplane mode. Then, the device will no longer attempt to connect to the network. Ensure you document any changes you make to the device.
- **A toolkit**: A small precision toolkit with multiple screwdriver bits is used to disassemble laptops, desktops, or mobile devices to access the digital storage container. You want to make sure you have a variety of screw heads to match what the various manufacturers use. Sometimes, the manufacturers will use two or three different screw heads when assembling their devices.
- **Miscellaneous items**: This can include extra power cables, data cables, USB hubs, screws, or anything else that might be difficult to acquire when you are at the subject's location in the middle of the night, and no stores are available for you to purchase the missing item. If you are responding to a commercial site, keep a spare mouse and keyboard in case you need to access a server and they are not available. (If you are conducting network-based investigations, you may also want to include a network tap.) This subset comprises items you don't think are needed until you are onsite and need them.
- **A forensic laptop**: Make sure all your software is up to date. I recommend creating a folder containing digital versions of any forms you will use, any processes you need to document, and any applications you find helpful in carrying out your tasks.
- **Encryption**: If you are traveling out of the country to get to the target site, you might want to encrypt the target drives that contain the acquired data you need to analyze. It is not uncommon for security services or customs to seize devices. This will ensure the data you acquired will not be compromised.

- **Software security keys**: This is also referred to as a dongle. You will find commercial versions of software that require you to insert a USB-based security key to use it. You want to make sure you have them with you because the software cannot be used without the security key inserted.

Note

A program called VirtualHere (`http://virtualhere.com/home`) allows you to use your USB devices remotely. This will require a network connection at your destination and at your home location where the USB keys are plugged in. If you are unsure about the quality of your network connection, I recommend taking the keys with you.

Now, the important question is this: how do you carry all of this from one location to another?

My recommendation is a Pelican-type case that is watertight and crush-proof to protect the equipment. Also, include a TSA-compliant locking device if you must travel via commercial air in the United States.

The list of items we have just discussed is only a recommendation. You will add/subtract from this list to meet the needs of the task at hand. There is no right or wrong answer when stocking your response kit. The budget, the organization, and the task at hand will dictate what equipment is needed.

A government/law enforcement digital forensic investigator may acquire full forensic images at the scene, and they will need larger storage capacity devices. As you become more experienced, you will accurately determine what equipment you need to perform your duties.

The result is that you need to have a response kit when leaving the office to acquire digital data or respond to any incident. How you stock that kit is entirely up to you as the forensic investigator. This is all about making your job easier and more efficient.

That has covered some of the hardware and physical items needed. We will now move on to discussing software.

Forensic software

This is the software that you will use to analyze data. You have a choice of utilizing commercial software designed for the forensic process or open-source tools. You want to make sure that you use fully licensed software in your work environment.

There is nothing more embarrassing than an organization using pirated software to investigate and have that fact come out in the administrative or judicial proceeding. It will be a severe hit to your reputation if you use pirated software to conduct your investigation, and it will call into question your integrity, your ethics, the results of your inquiry, and the results provided by the forensic tool. I cannot stress this enough: you must use fully licensed software in the forensic process. So, what is the difference between open-source and commercially available tools?

Vendors make open-source software freely available for anyone to use. Typically, there are no restrictions on its use; you can use it for educational, profit, or testing purposes. The positive aspect is that it is available at no cost in most situations. The downside is that you will have little or no technical support if something goes wrong. It will depend entirely on your skillset and level of comfort working with these tools. In addition, many open-source tools use a **command-line interface** (**CLI**) and not a **graphical user interface** (**GUI**), which can intimidate new users.

A commercial tool will typically have better customer support, documentation, and timely updates. The downside is that you are paying for those services. In reality, most of the time anything that a commercial forensic tool can do, an open-source tool can do the same thing. A commercial tool may carry out multiple functions, while with an open-source framework you may have to use different open-source tools to accomplish the same task.

Neither choice is wrong. As a digital forensic investigator, you must know where the data came from and ensure that the tool provides an accurate representation of the data. It does not matter if the tool is an open-source or commercial version; you must validate the results provided by any tool. We will talk about validation a little further on in this chapter.

I often get questions about whether a particular piece of software is court-approved. Forensic software is not court-approved, but you need to explain in the administrative/judicial process whether the tool you used produces reliable results and is accepted within the forensic community.

In the United States, this is known as the Daubert standard, which comes from the Supreme Court case Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579 (1993). This standard is used to determine whether an expert witness's testimony is based on scientifically valid reasoning and can be appropriately applied to the facts of the matter. The factors the court considered are as follows:

- Whether the theory or technique can be or has been tested
- Whether it has been subjected to peer review and publication
- The known or potential error rate
- The existence and maintenance of standards
- Its acceptance within the scientific community

Initially, the courts only used the standard for scientific testimony. That changed with the Kumho Tire Co. v. Carmichael 526 U.S. 137 (1999) case; the Supreme Court clarified that the factors used in the Daubert decision could also apply to non-scientific testimony, that is, the testimony of engineers and other experts who are not scientists. So, as you can see, it is not so much the software being used but the expertise of the digital forensic investigator. Commercial forensic tools simplify the process and sometimes have a **find evidence** button. However, as the digital forensic investigator, you still must know where the forensic tool extracted the artifact from within the filesystem. (Your local jurisdiction may have different opinions.)

The **National Institute of Standards and Technology** (**NIST**) has sponsored the **Computer Forensic Tool Testing Project** (**CFTT**) (`https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt`), which has established a methodology for testing computer forensic software tools through the development of general tool specifications, test procedures, test criteria, test sets, and test hardware. This project provides a source for testing the results of forensic tools on its website. They also offer a collection of testing media to conduct your validation of forensic software. It is part of your best practices to validate the results of your forensic tools at least annually or whenever the tool is updated. It does not matter whether you are a government or private sector digital forensic investigator: you need to have confidence in your tools and be able to testify that you have tested and validated the process.

In 2011, this validation process was called into question during the trial of Casey Anthony. Casey Anthony was being tried on the following charges: first-degree murder, aggravated child abuse, aggravated manslaughter of a child, and four counts of providing false information to police, who were investigating the death of her child. During the trial there was a significant assertion by the prosecution was that someone searched for the term “chloroform” 84 times on Anthony’s computer. While the trial was ongoing, it was discovered that the forensic tool used by the digital forensic investigators had misinterpreted the values in the internet history database. The user had only visited the site one time, not 84 as reported. The software designer of the forensic tool realized the mistake while the trial was ongoing and notified the trial team of the error. My recommendation is that you have multiple forensic tools to validate your findings. For example, you could have two commercial forensic tools, one commercial and one open-source forensic tool, or two open-source forensic tools, but you need to validate your findings.

Some open-source forensic tools include the following:

- **Autopsy**: Autopsy is a fully functioning suite of forensic tools that allows you to conduct a complete forensic examination. It costs nothing and can be found at `https://www.autopsy.com`.

- **SIFT Workstation**: SIFT is a virtual machine that uses the Ubuntu operating system with multiple forensic tools pre-installed. It is free and can be found at `https://digital-forensics.sans.org/community/downloads`.
- **PALADIN Forensic Suite**: PALADIN is a live Linux distribution based on Ubuntu and has implemented several open-source forensic tools in a user interface called the PALADIN toolbox. It is free and can be found at `https://sumuri.com/software/paladin/`.
- **CAINE**: **Computer-Aided Investigative Environment** (**CAINE**) is a digital forensics project that provides a GUI and many open-source forensic tools for free. You can find it at `https://www.caine-live.net/`.

These are just a few of the open-source forensic suites available. There may be others out there that I haven't mentioned, or you may wish to use single-purpose tools. As long as you achieve the goal of finding the artifact to reveal the truth about the matter being investigated, it does not matter which tool you use. The key is to use your training and experience to explain the pertinence of the artifact and how you determined the tool is providing reliable results.

Here are some commercial forensic tools available for Windows-based users:

- **X-Ways Forensics**: `https://www.x-ways.net/`
- **EnCase**: `https://www.guidancesoftware.com/encase-forensic`
- **Forensic Toolkit** (**FTK**): `https://accessdata.com/products-services/forensic-toolkit-ftk`
- **Forensic Explorer** (**FEX**): `http://www.forensicexplorer.com/`
- **Belkasoft Evidence Center**: `https://belkasoft.com/ec`
- **Axiom**: `https://www.magnetforensics.com/products/magnet-axiom/`

Here are some Macintosh-based tools:

- **Cellebrite Inspector**: `https://cellebrite.com/en/inspector/`
- **RECON LAB**: `https://sumuri.com/software/recon-lab/`
- **RECON ITR**: `https://sumuri.com/software/recon-itr/`

A Linux-based tool is **SMART** (`http://www.asrdata.com/forensic-software/smart-for-linux/`).

This is just a sample of the commercial forensic tools available for use. Each tool will have its strengths and weaknesses, which can be debated endlessly with your fellow practitioners.

Right now, I prefer X-Ways as my primary tool, and I supplement it with FEX and Belkasoft Evidence Center.

You can have all the tools, software, and hardware, but how effective will you be without training? So next up are some training options for you to consider.

Forensic investigator training

If you travel on the path of a career in digital forensics, you will need to continually upgrade your skills and training, which must be considered an ongoing expense. Just because someone goes through a 40-hour course does not automatically make them a digital forensic investigator. Instead, they are taking the first steps down that career path, but they will need to continue to attend training sessions and associate with other like-minded peers.

Certification is not a guarantee that the user knows what they are doing. Instead, certification shows that the user met the minimum level to achieve that certification. There are many certifications available, and some are more worthwhile than others. Before joining an organization and participating in its certification process, you must do your due diligence and research the costs, availability, and whether that certification is accepted within the forensic community. Most certifying organizations will require annual dues and a yearly training requirement to recertify the certification. There are tool- and vendor-specific certifications where you are being tested on your ability to use the vendor's forensic tool and an understanding of the fundamentals of digital forensics. At the other end of the spectrum is tool-agnostic certifications. You can use any tool to complete the certification process.

This is a list of some of the certifications available:

- **Certified Forensic Computer Examiner (CFCE) (Tool-Agnostic)**: `https://www.iacis.com/`
- **EnCase Certified Examiner** (**EnCE**) (tool-specific): `https://www.Opentext.com/products-and-solutions/services/training-and-learning-services/encase-training/certifications`
- **ACE** (tool-specific): `https://training.accessdata.com/exams`
- **Computer Hacking Forensic Investigator** (**CHFI**) (tool-agnostic): `https://www.eccouncil.org/programs/computer-hacking-forensic-investigator-chfi/`
- **Global Information Assurance Certification** (**GIAC**) (tool-agnostic): `https://www.giac.org/certifications`
- **Certified Forensic Mac Examiner (CFME)** (tool-agnostic): `https://sumuri.com/mac-training/`

Now that we have explored the equipment and training options, you still must prepare by understanding the legal and case information pertaining to the specifics of an investigation. So, we will discuss legal issues next.

Understanding case information and legal issues

Let's talk about case information and legal issues. You must get this information before you even power up your workstation to look at the digital evidence. You will have to gather information from the person requesting your services. It would be best if you asked the following questions:

- What is the nature of the investigation? For example, is it a narcotics case, homicide, or employee misconduct? As you listen to this information, you formulate your plan on how you want to proceed.
- What digital evidence do you expect to find at the scene? I've had responses where the investigator was only looking for a single laptop, and once we were at the scene, we found multiple laptops, multiple desktops, and many mobile devices. Just remember the information you get may not always be accurate, so you also must be prepared for that eventuality.
- What is the legal justification? For law enforcement—what is the rationale behind the search? Consent? A search warrant? It doesn't matter whether it is written consent or a written search warrant: you need to read the search warrant and consent to understand the limits placed on the search. It may be physical limits within the scene or digital limits on what you can search for on digital devices.
- As a government and corporate digital forensic investigator, I have had limits on what I can search for or view on digital devices many times. Be aware of those limits; if you find relevant artifacts outside of the scope of the search authority, they cannot be used in the proceedings, and you may face sanctions if you do use them.
- Who are the subjects and suspects, and what roles do they play in the investigation? Now, depending on your role, you may or may not have any contact with the subjects and suspects involved. However, if you do have that ability, try talking to them. If you can have a civil conversation with them, you may get additional information about the digital containers and the data.

If you're thinking, "We have gathered information from the first respondents, and we have gathered information on the other subjects involved; now we can jump right in and collect evidence!"—well, not yet. You want to make sure the crime scene has been adequately documented and safe. For law enforcement, this will include removing extraneous personnel from the scene, restricting access, and allowing someone to record the scene.

The easiest way is to photograph everything. They may call you to testify in a proceeding 12, 18, 24, or even more months in the future. Lawyers may ask you where a specific item was and, unless you have a photograph (or sketch) of the scene, you may not be able to answer the question.

For a corporate investigation—for example, a hidden camera found in a confidential location—what do you do? The finder's actions may hamper your ability. For example, I investigated a hidden camera in a unisex restroom. A restroom user found the camera when the tape holding it to the bottom of the shelf released, and the camera fell to the ground. The user gave the camera to their supervisor. The supervisor opened the camera and removed the digital storage card. They then placed it into a card reader and plugged it into their computer. At least five other people handled the camera and the SD card, putting it into multiple computers before contacting me. Every time they plugged the SD card into a computer system, they changed the evidence. When you access the data on an SD card, you change the date and time stamps on the files you access. An organization has to train its members not to look at digital evidence when there is an incident and to call a professional. This will ensure that the evidence is contained in a state that allows it to be presented in a judicial or administrative proceeding.

This case required interviewing all the people involved, processing the digital camera and the SD card, and examining the five workstations. Since this was a corporate environment and, initially, law enforcement would not be involved, I took photographs of the workstations and the connections to identify the specific workstations and their users. Remember, we are in a corporate environment, and there are multiple versions of the same make and model of computers everywhere.

There will be times when you have been presented the digital evidence after someone else collected it. You still must ask questions, and the source of your answers may only be the investigative reports. You will want to know the following:

- Why was this item seized?
- Does it contain evidence of criminal activity or evidence considered exculpatory?
- Is there a chain of custody for this item?
- How many people have had access to it?
- Where was the item found?
- Was it found in a secured location or a common area of the site?
- Are there any date and time references?
- What should the investigation focus on?
- When does the investigator need the findings of the digital forensic exam?

You need to review the documentation before you start the evidence-collection process. When investigators bring you digital evidence containers such as computers, you need to ensure the search warrant authorized its seizure. There have been several cases where devices containing digital evidence were seized, but there was a grey area around the use of digital evidence.

The search warrant will come with limitations on your search. For example, if it is an illicit images investigation, you may be restricted to only viewing images. It is your responsibility to read all the judicial paperwork and understand what it authorizes and does not. Only then can you create a plan for how you stay within limits.

You also must anticipate what problems you may encounter as you conduct the digital forensic examination. For example, is there an aspect of the investigation where your training and experience could be lacking? This is not something to be ashamed of but should be acknowledged so you can reach out for help to increase your training and experience. What resources do you have available to assist you?

Once the legal portion of your preparation is done, we can move on to the next portion of the process. You must now deal with acquiring the data in a forensically sound manner.

Understanding data acquisition

So, let's recap: you have received training as a digital forensic investigator and may have received certification. You have built or purchased a digital forensic workstation and a forensic laptop and have created your response kit. You have responded to the scene and ensured that it had been made secure. You have verified that no one has altered the scene, and you have documented the scene with photographs. Now, it is time to process the scene and collect that digital evidence. We will now discuss the acquisition of data, otherwise known as evidence.

There are multiple scenarios where someone may call on you to acquire data for a digital forensic investigation. For example, as a law enforcement officer, you may respond to the scene, identify potential sources of digital forensic evidence, and then seize those items. As a private sector or corporate investigator, you may be called on to take an employee's workstation or respond to the server room (either physically or remotely) to collect the data you need to analyze. The procedures we will discuss in the next section can be utilized in every environment.

A source of potential evidence is volatile memory. In the past, the data contained within volatile memory was ignored with a "pull the plug" mentality. This was based on whether officers responded to a scene and the computer was up and running. Best practice required officers to pull the plug to shut the system down.

However, volatile memory is only available while a system is up and running. Therefore, when the investigator pulled the plug, they lost all that data, including any potential evidence. As the field of digital forensics has matured, we have learned that what we once considered best practice was, in reality, not.

To collect volatile evidence, we should start from the most to the least volatile. This is called the **order of volatility**, and it goes like this:

1. Live system
2. Running
3. Network
4. Virtual
5. Physical

We approach volatile data collection with the same mindset as creating forensic images. You must document the steps you take because you will interact with the machine to collect volatile data, which will change the evidence. In reality, the changes you make typically do not affect what you are investigating. But you should know that changes are being made to the system; you may get asked a question about potential changes to the evidence while testifying at the administrative or judicial proceeding. If you don't know the answer, it could be professionally embarrassing.

The changes you make while collecting the volatile data will impact the processes found in RAM. That is why you need to take notes and document everything you do. Some examples of volatile data we collect are the current state of the system networking information (the ARP table, connections, routing table, and name cache), the logged-on users, running services, running processes, shared drives, remote activity, and open encrypted containers.

We have to balance our changes versus the evidence that may be potentially lost forever. The term "forensically sound manner" means leaving the smallest possible footprint during collection to minimize the amount of data being changed with the collection. The order of collecting volatile data is significant because if you collect volatile data in the wrong order, you may destroy the evidence you are looking for. RAM is considered to be the most volatile of all volatile data, so we would want to collect that first.

Keep the following in mind:

- Collecting the volatile data may not always be possible, depending on the specific set of circumstances you encounter on the scene.

- If you find there is a destructive process running on the machine and the information you want to collect is being altered or overwritten, you may not want to take the time to collect the RAM as evidence is being manipulated.
- If it is a remote connection causing the destructive process, you need to document the connection, sever the connection, and then collect the RAM. Again, it depends on your investigation and the information you are trying to acquire.
- If the attacker is connected remotely and is accessing highly sensitive data, do you want the attacker to maintain access while you collect the RAM, or do you want to interrupt the connection? What if it is not critical information?
- Do you want to let the attacker continue to have access while you continue your processing?

Ultimately, the goal of digital forensics is to create a forensic image for analysis. Therefore, under normal circumstances, it is not appropriate to change digital evidence during collection.

In today's environment, that is not always possible. Due to the easy availability of full disk encryption or full volume encryption, it is no longer acceptable to pull the plug on computer systems.

Let's take a slight detour and talk about what encryption is. At a basic level, encryption is encoding information to protect the confidentiality of the information and allow only the person with the decryption key to access it. All encryption can be broken if the attacker has enough time.

With today's level of equipment, that time factor is measured in hundreds of years. As technology advances with increases in processing power, the time taken to decrypt top-level encryption decreases. So, what was considered secure encryption in the 1990s is now regarded as weak. That is why it is imperative not to pull the plug on a system where it is possible that encryption is being used. Without gaining access to the decryption key, you cannot get to the data.

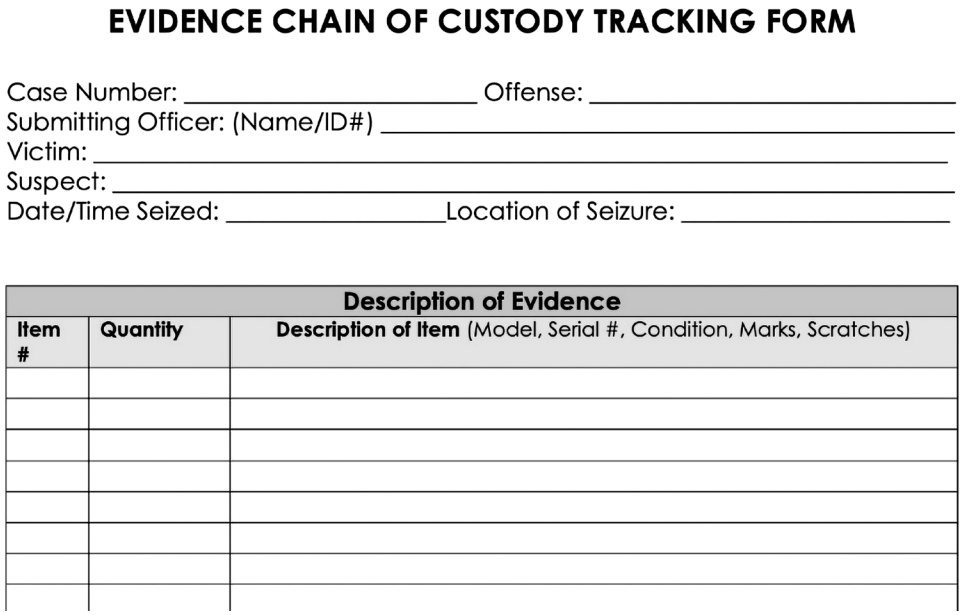
Every situation, every crime scene, and investigation will be different, which means the actions you take will be based on the specific set of circumstances you encounter. Utilize your problem-solving skills and make quick decisions based on the limited information you have available.

Now we have the evidence, how do we keep control of it? Let's talk about the chain of custody.

Chain of custody

Maintaining the **chain of custody** is an integral part of preserving and authenticating physical and digital evidence for an administrative or judicial proceeding. The chain of custody documents all access to the evidence, who accessed it, when it was accessed, and for what purpose it was accessed.

NIST provides a chain-of-custody document, shown in the following figure. It is a generic chain-of-custody form for you to use and adjust as needed and can be downloaded at `https://www.nist.gov/document/sample-chain-custody-formdocx`. The form is used to track the chain of custody and will be maintained every time evidence changes hands:



EVIDENCE CHAIN OF CUSTODY TRACKING FORM

Case Number: _____ Offense: _____
Submitting Officer: (Name/ID#) _____
Victim: _____
Suspect: _____
Date/Time Seized: _____Location of Seizure: _____

Description of Evidence		
Item #	**Quantity**	**Description of Item** (Model, Serial #, Condition, Marks, Scratches)

Figure 2.1: An evidence form

As you can see, some fields may not be pertinent to you. For example, as a corporate digital forensic investigator, you may not need the **Victim** field, so you can change it or remove it altogether.

The goal of this form is to track the digital evidence and maintain control so that you may authenticate the evidence later. In the **Description of Evidence** field, you describe the container holding the digital evidence. It could be non-reusable media, such as a DVD with log files burned for later examination.

In the following figure, you can see the **Description of Evidence** section. The **Item** number refers to a sequential numbering system to help track the items. **Quantity** is the physical number of items, and the **Description of Item** field is self-explanatory:

Description of Evidence		
Item #	**Quantity**	**Description of Item** (Model, Serial #, Condition, Marks, Scratches)
CD-001	1	Ultimate DVD contains server logs from AD001
HDD-001	1	Samsung SSD 1TB Ser#ABC9876
HDD-002	1	Samsung SSD 512 MB Ser#DEF4567
CP-001	1	Pixel XL 128 GB Ser# A5 12 D3 AC FD
TD-001	1	Generic Thumb drive 32GB (green) Unknown Serial Number
MD-001	1	Apple iPad 512 GB Ser# 09 E3 4D AB Rose Gold

Figure 2.2: A description of the evidence

For example, in the previous figure, a DVD is listed as item **CD-001**. You might impound several CDs or DVDs and have the problem of trying to differentiate one disk from another. It's not just CDs or DVDs but also hard drives. It won't often be that you will impound a single item of a specific media type.

I use the following numbering system as a part of my process:

- CD/DVD: **CD-XXX**
- Hard drive: **HD-XXX**
- Thumb drive: **TD-XXX**
- Cell phone: **CP-XXX**
- Mobile device (not a cell phone): **MD-XXX**

Note

As a side note, you also need to make a permanent mark on the items being seized, but you should try to do so in a manner that will not reduce the value of the item.

You can see in the following figure that the hard drive is marked as **HDD001** with the date and the initials of the officer seizing the device:

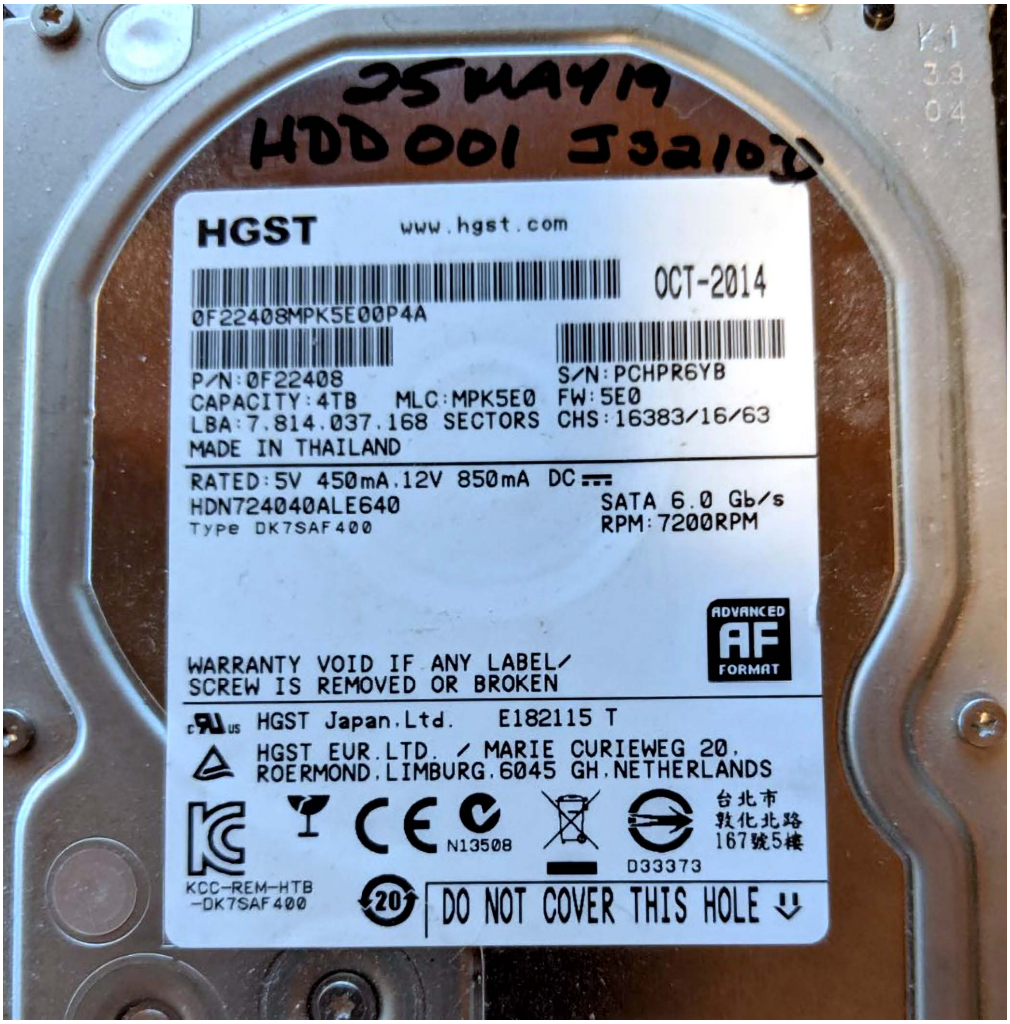


Figure 2.3: A hard drive

When the forensic image is created, the device will be referred to as **HDD001** for the rest of the process.

If you cannot write on a device without permanently reducing its value, such as an iPad, do not use a permanent marker to write **MD-XXX**. Instead, use an adhesive label to mark the information.

Note

Use a system that works for you. When you have developed your system, make sure you use it every time. It will save you from losing evidence or mismarking evidence.

When we are on the scene and seizing evidence and containers containing digital evidence, we want to make sure we do so in a forensically sound manner. Therefore, we do not analyze the original evidence; we create a copy to do the exam to ensure we do not make any changes to the original evidence.

We have three choices for making a working copy:

- **A forensic copy**: This is a straight bit-for-bit copy of the source to the destination. This is not common in today's environment. Ensure that your destination device has no old data from previous investigations. You do not want to cause cross-contamination between the current digital forensic investigation and a past investigation. We will recover deleted files, file slack, and partition slack. We will discuss wiping hard drives later on in this book.
- **A forensic image or forensic evidence file**: We create a bit-for-bit copy of the source device, but we store that data in a forensic image format. This could be a **DD** image, an **E01** image, or an **AFF** image. We take that source data and wrap it in a protective wrapper of the forensic image. We will recover deleted files, file slack, and partition slack.
- **A logical forensic image**: Sometimes, we are restricted to only accessing specific datasets. They do not allow us to access the entire container. We cannot create a bit-for-bit copy forensic image/forensic evidence file or a forensic copy. This can be used when we extract data from a server, and we cannot shut the server down to create a forensic image from the source hard drives. So, we can make logical copies of the files and folders pertinent to the investigation. We will not recover deleted files, the file slack, and partition slack.

Later on in *Chapter 3*, *Acquisition of Evidence*, we will address creating a forensic image from the devices we have seized or the data seized at the scene.

Now that we have discussed what you need to consider when acquiring a dataset, we will discuss what you need to understand when analyzing data.

Understanding the analysis process

Once you have collected data from the scene, you return to your lab, and it is now time to start your forensic analysis. You will find yourself quickly overwhelmed by the sheer amount of data you will find in storage devices. You must promptly determine whether the information contained within the storage containers is pertinent to your investigation.

This is where the information gathering in the case information and legal issues step of the process will play an essential part.

Therefore, you must capture the five Ws of the investigation (previously mentioned in *Chapter 1*, *Types of Computer-Based Investigations*). First, associate the activity on the computer system with a specific user and identify that user as a real-life person.

If the investigation already has a live suspect identified, you correlate that suspect with the user on the computer system. We will discuss some guidelines that can be used with commercial or open-source forensic tools. The goal is to understand the process without resorting to any specific forensic tool.

Now that we have discussed what you need to consider when acquiring a dataset, we will discuss what you need to understand when analyzing the data.

Dates and time zones

Dates and time zones can cause issues for the digital forensic investigator if they forget to consider them. For example, if you only conduct exams in a specific time zone and all your seized data comes from the same time zone, the issues you face are minor. But if the data comes from multiple time zones or you travel to various time zones, they can cause some confusion if you do not consider the time zone issue.

Setting the forensic machine and tools to use **universal time** (**UTC**) as a standard frame of reference helps solve this problem. Also, ensure that you adjust any timeframe where the criminal activity occurred in UTC. It does not help that operating/file systems save metadata in different time zone formats. You also must consider that the suspect may have changed the time zone settings on the computer to hide their illicit activity. Timeline analysis is critical when conducting a forensic exam.

Next, we will need to be able to identify files we know are irrelevant, as well as instantly identify contraband images. We can do that with hash analysis.

Hash analysis

What is a hash value? A hash is a digital fingerprint for a file or piece of digital media. It is generated using a one-way cryptographic algorithm.

The standard cryptographic algorithms used in digital forensics are **Message Digest 5** (**MD5**) and the **Secure Hashing Algorithm** (**SHA-1**). MD5 creates a 128-bit digital fingerprint, while SHA-1 produces a 160-bit digital fingerprint. Using a hashing algorithm allows using a variable input to create a fixed-length output.

If one bit changes in the variable input, it will cause a different output. Additional hashing information will be provided later in the book. Let's see how this works in the following exercise:

1. Create a text file containing the words `This is a test` with a filename of `Hash Test.txt`:

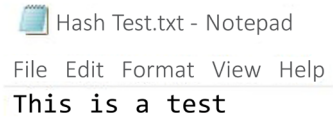


Figure 2.4: The hash text

2. Use the free Jacksum utility (`https://jacksum.loefflmann.net/en/index.html`) to obtain the hash values:

Figure 2.5: The Jacksum values

As you can see in the preceding figure, the `ce114e4501d2f4e2dcea3e17b546f339` value is the MD-5 standard length output for the `F:\Hash Test.txt` file.

The second value, `a54d88e06612d820bc3be72877c74f257b561b19`, is the SHA-1 output. It doesn't matter which forensic tool I use—these values are the digital fingerprint for this specific file.

3. Change a single part of the contents of the file:

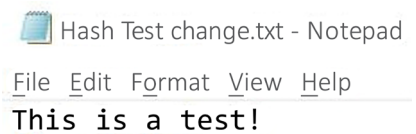


Figure 2.6: The change in the text

I have added an exclamation point to the end of the sentence—a very small change—but any change will change the hash values.

4. Use Jacksum again and you will get a totally different hash value:

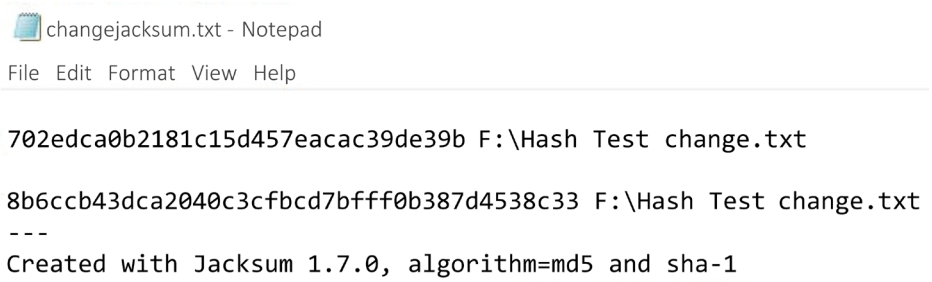


Figure 2.7: The change in the Jacksum values

The MD5 value is now `702edca0b2181c15d457eacac39de39b`, which is different from the original value of `ce114e4501d2f4e2dcea3e17b546f339`.

The standard output generated by the hashing algorithm is a one-way process. You cannot input the alphanumeric value to reverse the process to get the original dataset used in the hashing process. If you have a hash set of known illicit images, the values within that hash set cannot be used to re-create the illicit images.

There are hash sets (sets of multiple hash values) that identify known good files. These are files that are of no interest to an investigator. These can be the standard files used in an operating system or application. Using a known good hash set allows you to filter out files with no evidentiary value. On the other hand, if you have identified files of interest, such as illicit images or known documents that have been stolen, any data that may interest the investigator can also be highlighted. For the known bad files, someone needs to have access to the original file to create the hash value used to identify the file.

Using hash analysis can save you some time and effort during your investigation:

- You can use it to verify the evidence has not changed.
- It can be used to exclude files.
- It can be used to identify files of interest.

NIST has created the **National Software Reference Library** (**NSRL**) (`https://www.nist.gov/software-quality-group/national-software-reference-library-nsrl`), where they have collected software from many sources and created a **Reference Data Set** (**RDS**).

The RDS is a large hash set to help identify known good files when conducting your examination. The RDS is freely available to law enforcement, the government, and private industries. Some files identified in the RDS may be considered malicious, such as hacking tools. The investigator still has to put the files in context to see if they were being used for an unlawful purpose. The RDS does not contain hash values of illicit data, such as illegal images.

A collision occurs when two different variable inputs result in the same fixed-length output. This means that two different files have the same hash value, which you will realize is not good for identifying evidence based on our previous discussions. However, nation-states have manipulated variable inputs to create the same fixed-length output, and they have been successful.

Does that mean hashing is dead? No, it isn't. There have been no two different files found in the wild with the same hash value. All the collisions that have occurred have been files that have been manipulated. When independent examiners analyzed the manipulated files, they did not have any user-readable content. While there has been concern that this would negatively affect the admissibility of digital evidence, in 2009, the court case of US versus Schmidt ruled that the odds of a collision of two files were insignificant and were not an issue.

Now that we have determined the digital fingerprint, let's make sure the files are correctly identified.

File signature analysis

Your next step is to carry out a file signature analysis to ensure the file extension matches the file type. Many file types you will find in the filesystem have been standardized and possess unique file signatures to identify themselves to the filesystem. This is not the file extension, such as a Microsoft Word document with a file extension of `.doc` or `.docx`.

A user can change the file extension to hide incriminating evidence. The intention behind carrying out a file signature analysis is to determine whether the file signature and file extension match.

The following screenshot shows how X-Ways flags a file when the file extension does not match the file signature:

Name	10534.gif
Type	jpg
Description	existing
Existent	✓
Size	3.0 KB (3,081)
Modified	07/12/2008 21:51:38 +0
Ext.	gif
Type status	mismatch detected, OK
Type descr.	JPEG

Figure 2.8: A file signature mismatch

The file extension identifies the file as a GIF, but X-Ways has identified the file as a JPEG. The next figure shows the file header for the GIF file in question:

```
 Offset    0  1  2  3  4  5  6  7   8  9 10 11 12 13 14 15      ANSI ASCIT
00000000  FF D8 FF E0 00 10 4A 46  49 46 00 01 01 00 00 01  ÿØÿà  JFIF
```

Figure 2.9: A file header

A GIF file should have a hexadecimal **47 49 46 38** file signature, not hexadecimal **FF D8 FF E0**. In some cases, the mismatch is through normal usage of the filesystem and not from user interaction. You must examine the data to ensure the mismatch can be attributed to a specific user.

Gary Kessler has created a website that allows you to search a database based on the file extension or signature. You can refer to this website at `https://filesignatures.net/`:

Figure 2.10: filesignatures.net

You can search by file extension or file signature. Once you input the file extension, in this case JPG, you will be returned the file signatures associated with the JPEG standard:

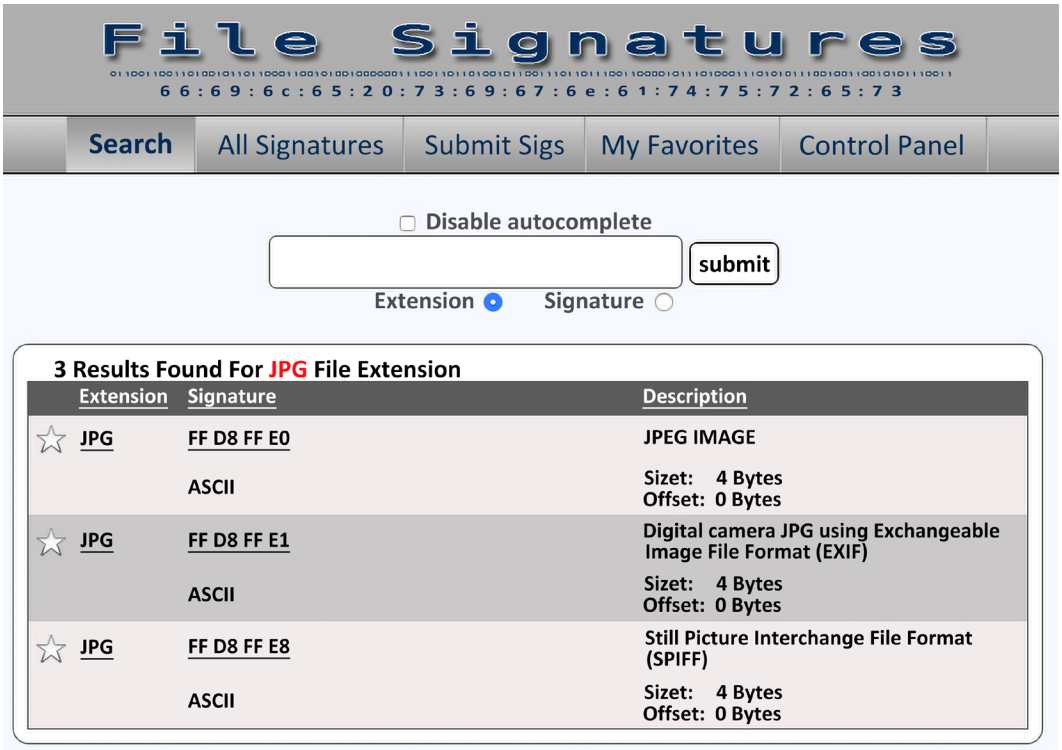


Figure 2.11: The results for a JPG file signature

After we have ensured that the files have been properly identified, we need to identify any malware that may be on the system. We can do that with antivirus.

Antivirus

A common claim of innocence from a subject accused of wrongdoing that "a virus did it" has occurred in nearly every investigation I have done. Have you determined whether that is a valid claim? For example, is there malware on the system, and did it cause the behavior you are investigating without the user's interaction or knowledge?

This is one reason we collect the volatile data to see what was occurring on the system at the time of collection. If someone else has collected the evidence and all you have is a forensic image, you can still scan that forensic image to help determine whether someone has installed malware. Several forensic tools allow you to "mount" the forensic image as a read-only drive, and you can then scan the filesystem to help determine whether there is malware installed.

FTK Imager is a free tool offered by AccessData, available at `https://accessdata.com/product-download/ftk-imager-version-4.2.0`, which allows you to mount the forensic image.

Image mounting allows you to mount a forensic image as a drive or physical device. Your viewing is in read-only mode. You will find many benefits to mounting a forensic image, such as using the file explorer to view it as if it were a device attached to the computer. In addition, you can natively view different file types, use antivirus against the forensic image, share the mounted forensic image over a network, and copy files from the mounted forensic image.

We will now cover how to mount a forensic image with FTK Imager in the following exercise:

1. To mount a forensic image in FTK Imager, you need to select the **File** menu and then select **Image Mounting...** from the menu, as in the following screenshot:

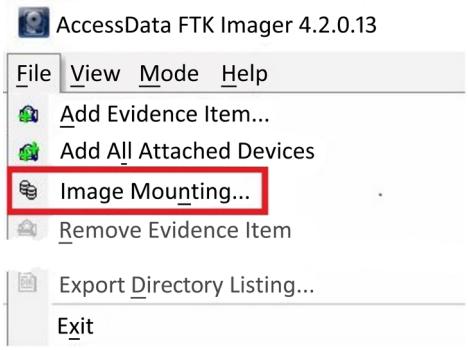


Figure 2.12: Image mounting

2. It will then present you with the **Mount Image to Drive** menu:

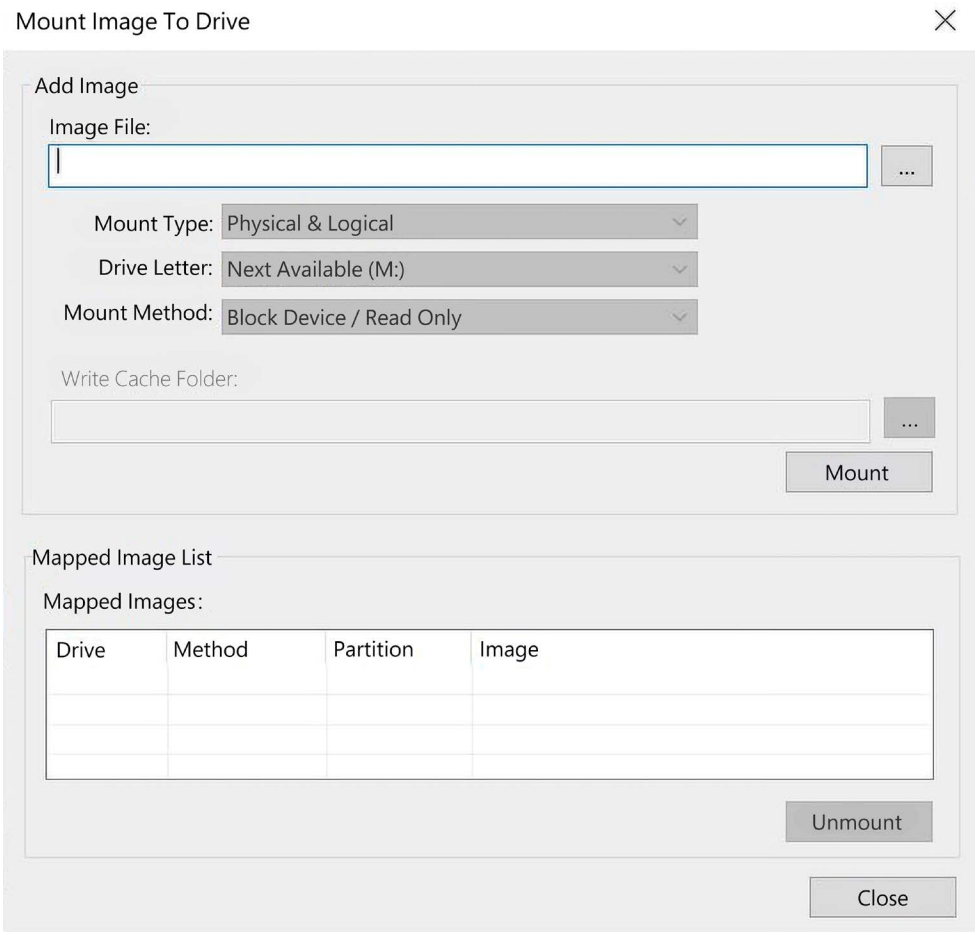


Figure 2.13: Mount the image

In the dialog box, you will have to select the forensic image you want to mount. If this is a segmented forensic, you only need to point it at the first segment:

- **Mount Type**: You have a choice of **Physical & Logical**, just **Physical**, or just **Logical**. If you select **Physical & Logical**, the software will mount the forensic image as a physical device and mount any logical partitions.
- **Drive Letter**: This is where you want to see the forensic image. In the previous figure, it shows that the next available drive letter is **M**. You can select any open drive letter you desire.

- **Mount Method**: You have the following choices:
    - **Block Device / Read Only**: This will read the device as a block device, which means a Windows application that performs physical name querying can view the mounted device.
    - **Block Device / Writable**: No changes are made to the original evidence. It will save any changes you attempt to make in a cache file.
    - **File System / Read-Only**: The device as a read-only device that someone can view using Windows Explorer.

In the following screenshot, you can see we have mounted a forensic image and the forensic image has partitions in it:

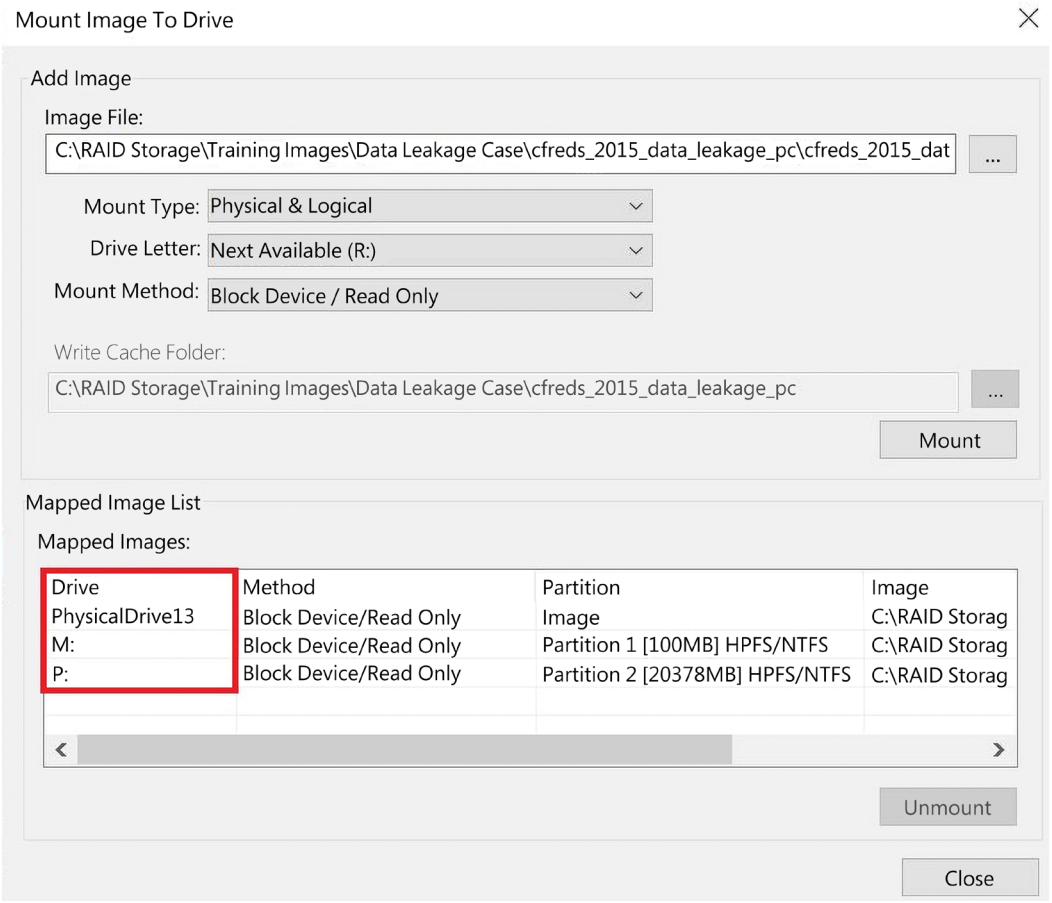


Figure 2.14: A mounted image

The system has mounted the partitions as drive **M** and drive **P**. Now, you can run the antivirus of your choice against those volumes to see whether they have installed any malware.

If malware has been installed, that is still not an alibi for the suspect. Determine whether the found malware can do the actions the suspect claims. I have investigated many illicit images investigations where the accused claims the malware downloaded the images. I have yet to find any malware that searches, finds, downloads, and sorts by content the illicit images found on a subject's computer. You still must analyze the content to determine the context of the digital evidence.

Now, you can begin your analysis of the filesystem and operating system. We will discuss the specific artifacts in the rest of this book. The OS is in place to communicate between the applications and the hardware. Some common operating systems are Microsoft Windows, Macintosh, and Linux. Almost every action conducted within an OS, whether user- or machine-generated, will leave a footprint somewhere within the OS. You want to analyze these artifacts controlled by the OS to determine whether the user committed any wrongdoing.

A filesystem is the storage mechanism for the data. A filesystem is independent of the OS. The filesystem tracks where the data is stored and what space is available. There are many filesystems, such as NTFS, HFS+, FAT 32 and Ext 4. Some formats are compatible with multiple operating systems, and some are not. For example, NTFS is utilized by Microsoft Windows as the filesystem of choice.

Once we are sure there is no malware on the system, we can then move on to report the investigation findings.

Reporting your findings

We are at the final step of the process: your report. You did all the work of preparing, purchasing the equipment, going to training, and creating your response kit, and when the call came, you responded to the scene. You successfully got the case information and navigated any potential legal issues when you arrived. You collected the volatile data, identified containers of digital evidence, and duly seized the digital evidence while maintaining the chain of custody when transporting it back to your lab. You then conducted your analysis and found artifacts that show that the suspect did or did not do what they were accused of.

Now what? You must be able to explain your findings to a non-technical person. You must take a very technical topic and talk about it in a manner that a non-technical person will understand. This is one of the most challenging aspects of being a digital forensic investigator to master.

You may have to create different report versions depending on the audience. Your intended audience will read and interpret your report, and a third party might question you on it in a judicial or administrative hearing.

Details to include in your report

You need to include enough details so that you can remember what occurred. Taking notes as you traverse the process will be your friend. There have been many times where I have failed to take that advice and had to go back and redo the process because I did not write something down. Your notes can take many forms, such as handwritten notes, typed notes, screenshots, or notes made with the built-in blogging function of your favorite forensic tool. There is no right or wrong rule on how to take notes, only that you take notes during the process.

So, what do you want to document? The following gives you an idea:

- Communication between the primary investigator and prosecutor/C-Suite executives
- The condition of the evidence containers
- The specifics of the storage device (the make, model, serial number, and condition)
- Personal identifiers of the suspect, victim, and witness (if a criminal matter)
- Personal identifiers of the witness(es), response team, responsible executive (if a civil matter)
- The forensic hardware used
- The forensic software used
- What you examined (even if the examination turned up nothing of evidentiary value)
- Your findings
- Glossary (to define technical terms)

Take all the pieces and put them together so that a non-technical reader will understand the investigations, the steps you have taken, and why you made the conclusions you did. As with everything else in digital forensics, there is not a set standard for the format of your report. Instead, you will have input from your employer, the recipients of the report, and your personal preferences.

I would recommend you include the following in your report. You should break your report into three primary sections:

- Your narrative
- Pertinent exhibits
- Supporting documentation

The narrative is what it sounds like. This is where you explain what occurred, what you did, and what it means. You should include an executive summary to hit the key points and conclusions and then move on to a detailed narrative. In your narrative, you should provide screenshots of the artifact you are talking about. Do not add a screenshot without an accompanying narrative. Do not assume the reader will understand what is pertinent about the screenshot. You will have to explain it to the reader. Make sure you focus your screenshot on the artifact you are discussing.

Suppose your report contains screenshots of contraband, such as illicit images. In that case, you will need to maintain control of that report to not cause an accidental release of the contraband images. You will also need to create a second report with the contraband images redacted for readers who cannot legally possess the contraband images.

After the executive summary, you should include basic administrative information. Next, identify the subjects involved, including the victim, suspect, witness, and other investigators.

Document facts and circumstances

You have two options regarding listing the evidence that you analyzed. In some larger cases, the listing of digital evidence can take two or more pages. Having a long, drawn-out list does not help the reader understand your report. More likely, the reader will skip the evidence listing and move on. If the investigation does not have a large number of digital devices being examined, then you can list them here, including the devices where you found nothing of evidentiary value. If you have many digital devices, I recommend you only list the devices with artifacts of evidentiary value while listing the entire evidence list at the end of the report.

You should also list details about the creation of the forensic images. I typically include a summary of the acquisition details in the narrative portion. I then create a detailed step-by-step process of the forensic image's creation as an exhibit. Once again, having a step-by-step process in the report's narrative does not help the reader understand the process. Giving the reader the high-level details of the forensic image process and then providing the details at a different location improves the readability of the report.

The analysis of the digital evidence will make up the bulk of your report. This is where you will walk the reader through the step-by-step process of the incriminating artifacts you found and why the artifact is important. I have often seen reports where a specific image is highlighted as important, but then it never explains why the image is important. Is it the location of where the image was found, or is it the image itself? Explain why that specific artifact is important and how you determined it was important.

Note

Remember, you are taking a technical subject and explaining it to a non-technical reader. Do not create a list of important files and assume the reader will know what is important.

I find that it's best to present the artifacts in chronological order. For example, if you are examining the illegal downloading of copyright-protected material, you would start by potentially identifying the owner of the computer and any artifacts that can identify a specific user. You can then show any browser searches the user performed when looking for the copyright-protected material and then the steps taken to download that material. Suppose the user had any ongoing communications with other users about the copyright-protected material. In that case, you could then use these communications to support your hypothesis about the user's activity of downloading the copyright-protected material.

You can also present the artifacts by subject. For example, if you are investigating the possession and distribution of illicit images, you can present the artifacts showing that the user viewed the images. This will show that the user knew about the images on their system and whether the user actively shared them with other users. Just the image alone is not enough; you must also find the OS artifacts to support your hypothesis about the image. When creating the analysis section, you will need to avoid making any absolute statements. I have seen forensic reports dealing with illicit images where the investigator made the unequivocal statement that the user knew about the illegal image. They found the image in question in the thumb cache database. The location of an image in a thumb cache database is not absolute proof that the user knew about the image. The system can include images in the thumb cache database without the user's knowledge. So, you want to be very careful with your language. Do not include opinions—only provide factual information.

I have seen reports describing artifacts as "a disturbing image of a child." The term "disturbing image" is not factually based—it is an opinion. It would be best to describe the artifact as it is without projecting your feelings about it. A better description could be "an image depicting a young-looking male, nude, standing in a wooded area." Be careful how you describe the artifacts attributed to a specific user or person. The most challenging item to prove is who is behind the keyboard. You can never say with 100 percent certainty that suspect **A** did the criminal activity unless you have a video showing suspect **A** was at the keyboard at that specific time. This is not the place for you to offer your opinion; do not assume ownership of an item or the identification of a user.

The report conclusion

The final portion of your narrative is your conclusion. This is the section where you can offer your opinion based on the artifacts you described in the analysis section of the report. You must still be careful about presenting your opinions. Try to look at the artifacts with no preconceived notions and determine whether the facts again meet your hypothesis. If you cannot decide, include that opinion. Remember, it is not always about proving the subject's guilt or liability. You must also provide evidence if the subject did not do what they are being accused of.

You will probably create an electronic report for distribution; a standard format is PDF. No matter what format you use, make sure you digitally sign the report. The digital signature will show that no one has altered the report since you signed it.

Note

Remember, the report is a representation of you and the investigation. If you create a poor report, that will reflect poorly on you, the investigation, and your organization.

Proofreading is essential. Do not proofread the report yourself, use the peer review process. You will miss typographical errors, poor sentence structures, and unclear findings. What may be clear to you in your mind may not always be accurately transcribed in written form. Suppose the investigation proceeds to administrative or judicial proceedings. In that case, I can guarantee the opposition will dissect your report line by line, looking for inconsistencies and places where you were not objective.

Remember, if the reader does not understand what you are saying about the artifacts you found, your entire investigation effort has been wasted.

Summary

In this chapter, we have discussed the forensic analysis process. You now know how to prepare to conduct a digital forensic examination, from getting the proper equipment to the training and getting certification. In addition, you now understand the importance of obtaining information before seizing digital evidence and ensuring you talk to other investigators or personnel involved in the situation.

I cannot stress the importance of collecting volatile data enough; if you do not, you will lose a large amount of potential evidence. Next, we discussed some strategies for conducting your examination and the differences between an OS artifact and a filesystem artifact. Lastly, we discussed reporting your findings so that the reader easily understands them.

The next chapter will go into the specifics of the acquisition of evidence and how to validate your tools to create an error-free forensic image.

Questions

1. Which of the following should be included in your response kit?
    a. A digital camera
    b. Latex gloves
    c. A write-blocking device
    d. All of the above
2. You must use commercial software to perform a valid forensic examination.
    a. True
    b. False
3. What questions need to be asked when you receive digital evidence?
    a. Why was the digital evidence seized?
    b. Where is the chain of custody?
    c. Who has accessed the evidence?
    d. All of the above.
4. RAM is the most volatile of evidence.
    a. True
    b. False
5. The chain of custody documents _____.
    a. Who controlled the evidence
    b. Who witnessed the crime
    c. The suspect's fingerprints
    d. None of the above
6. Which of the following is best for a digital forensic exam?
    a. A forensic copy
    b. A forensic image
    c. A logical forensic image
    d. Both B and C

7. Which of the following is a hashing algorithm?

    a. CDC
    b. FBI
    c. MD5
    d. LSD

The answers can be found at the end of the book in the *Assessments* section.

Further reading

Warren Kruse and Jay Heiser, Computer Forensics: Incident Response Essentials (*Addison Wesley, 2001*)

You can purchase the book from `https://www.amazon.com/Computer-Forensics-Incident-Response-Essentials/dp/0201707195`.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

3

Acquisition of Evidence

Digital evidence is one of the most volatile pieces of evidence an investigator can handle, and the slightest error or mishandling on the investigator's part can severely affect the investigation. For example, you may lose the data forever or lose pieces of it. In addition, the unintentional manipulation of data can cast doubt on your ability to investigate or question the integrity of the data in the investigation. This chapter will address minimizing or eliminating any of these issues by using a tool validation process to create an error-free and validated forensic image.

We will cover the following topics in this chapter:

- Exploring evidence
- Understanding the forensic examination environment
- Tool validation
- Creating sterile media
- Defining forensic imaging

Exploring evidence

What is evidence? The dictionary definition is the available body of facts or information indicating whether a belief or proposition is true or valid. Now that seems to be a short, simple, common-sense answer to a simple question. In reality, the question becomes far more convoluted when you consider regulations, the law, and rules of evidence in one jurisdiction, which grows exponentially when considering multiple jurisdictions. Evidence is a determination made by the trier of fact. The trier of fact will determine if the evidence meets the standards for that proceeding and jurisdiction.

I offer the following example: Let's say you are investigating a murder and you find the victim's and suspect's blood in the suspect's vehicle; the victim's blood on the suspect's socks; and a bloodied glove at the scene, and its matching mate found in the suspect's house.

You could believe the government had an airtight case against the suspect based on this evidence. But in this case, the defense was able to successfully argue and challenge the evidence, which resulted in the suspect's acquittal. As you can see, just because something is evidence, if it cannot withstand the challenge of the opposition, then it becomes a liability.

I have worked on both sides of the judicial process regarding digital evidence, and every time, the sheer amount of digital evidence that never sees the light of day amazes me. If we do not present the evidence to the trier of fact, it does not exist as far as the proceedings are concerned. Neither side will reference it or offer it during the proceedings. It simply will not exist.

How does the opposition attack evidence that the trier of fact has admitted? Either by attacking the evidence itself and/or by attacking the process and personnel associated with collecting and analyzing the evidence.

Consider the following example:

An examiner analyzes the thumb cache of the system and sees a URI (the **URI** is a **uniform resource identifier** based on the standard created by the Internet Engineering Task Force; in this instance, it is a file path) pointing to the location of the original image. The original destination folder no longer exists on the system, nor does the source image for the thumbnail in the cache.

As shown in the following screenshot, the source image for the thumbnail in the cache was located in the **New** folder on the **Picture Drive** of the user account of **bob**. *Figure 3.1* displays the URI that was found in the thumbnail's metadata:

```
URI: file:///media/bob/Picture Drive/New
```

Figure 3.1: URI from thumbnail

In the following screenshot, you can see the URI found in the metadata of a different thumbnail in the same thumb cache. The path is very similar to the one found in the URI image. However, there are significant differences here–the user account is **bobby**, not **bob**, and the **New** folder does not exist:

```
URI: file:///media/bobby/Picture Drive/
```

Figure 3.2: URI image: bobby

On the system that was being analyzed, there was not a **bob** user account, nor were there any artifacts showing the **bob** account was ever created or deleted from that system. The digital forensic examiner amended their report and incorrectly stated that the **Picture Drive** was the same in both instances based on the similarities of the URIs. Initially, the digital forensic examiner noted that the metadata's URIs represent file paths that cannot be verified.

The digital forensic examiner conducted a second exam and found a deleted folder called **New** on the **Picture Drive** and amended the report to reflect that. The URIs found within the metadata represent evidence item HDD 001. The **New** folder was deleted on this date and time. (I am not using exact names or dates for obvious reasons.)

Based on the file path and the current users, there was no way to determine if the **New** folder referenced in the URI was the same as the deleted **New** folder. When the lawyer confronted the digital forensic examiner about these discrepancies, they admitted they had made an error. I believe they made the error because of the similarities of the file paths and not paying attention to the specific details. I absolutely believe the error was not malicious or intentional but an honest mistake by the opposition's digital forensic examiner. As you can see, sometimes, a simple mistake can lead to additional questions being asked about the collection of the evidence and the process used to generate the report and the evidence.

In a different case that I was brought in on, the subject was charged with attempting to lure a child. In this specific set of circumstances, the subject communicated with an **undercover agent (UC)** and sent many illicit images to the UC. When law enforcement took the subject into custody, the subject was interrogated, confessed, and wrote an apology letter.

The confession, over 400 pages of chats, and a dozen illicit images were submitted as evidence in the judicial proceedings. Once again, you would expect that there would be a conviction based upon this evidence.

During the trial, it was revealed the government had deleted some text messages and edited the video file of the recording of the confession. The judicial authority informed the jury of the manipulated evidence. Additionally, the jury was told the only conclusion they could consider was that the government's agents altered the digital evidence to hide facts that would hinder the government's prosecution. The jury then found the subject not guilty of all charges.

If you do not follow your organization's best practices, policies, and procedures, the evidence will not see the inside of the courtroom. If the flawed evidence is admitted, the opposition's attacks will mitigate its effectiveness. These attacks can create enough reasonable doubt to generate an acquittal.

So, what can we do to mitigate the attack of the opposition? First, it does not matter which side of the matter you are on; the opposing counsel will attack your findings if it is harmful to their case.

Do not forgo proper evidence-handling procedures. Proper evidence handling does not end with collecting evidence in the field. As the evidence is transported from the field to the secure location, and whenever someone checks over the evidence, you must maintain the evidence's chain of custody and security.

Do not forgo utilizing proper procedures, methodologies, or processes when conducting your digital forensic investigation. Do not take shortcuts.

Validate any procedure, methodology, or processes. You must go through the validation process; you cannot rely on third-party validation. Your validation must repeatedly reproduce the same results when performed by you or anyone else.

Prepare and conduct your digital forensic examination with the mindset that someone will go through every step you take and question every finding you make. With this mindset, you should be able to mitigate any attack against your digital forensic examination. The key is that you must prepare. If you are unprepared for the attack, then you may be made to look incompetent while testifying in judicial/administrative proceedings.

We have discussed the evidence, but what about the environment in which you will conduct the investigation? We will now discuss how you should control the examination environment.

Understanding the forensic examination environment

A term that has been pounded into my head since I first went to training with IACIS is the *forensically sound examination environment*. While it sounds complicated, it is a relatively simple concept:

- The digital forensic examiner controls the working environment of the digital forensic examination
- No actions will occur unless the digital forensic examiner intends the action to occur
- When the action has been completed, the examiner will reasonably know what the expected outcome is

This concept does not merely apply to a physical location, but anywhere we complete a digital forensic examination or perform actions to support the digital forensic investigation. This could be a lab, office, or in the field where the digital evidence has been collected.

The forensically sound examination environment is a mindset of the digital forensic examiner. You want to be methodical and thorough to support the digital forensic examination. This mindset will help eliminate some mistakes that may occur during the process.

For example, the organization sent two colleagues to a remote location to acquire several workstations. They were able to complete data acquisition within 2 to 3 days. The investigators did not perform triage on the dataset or examine the dataset while on the scene, but it was expected to be completed when they returned to the central lab. The remote location was several hundred miles away, and once my colleagues left, they could not return to gain access to the source devices. Upon arrival at the central lab, my colleagues started to conduct their digital forensic examination. Colleague A started to examine one of the forensic images, and as a part of their process, they viewed the folder structure of the filesystem. As they were looking at the installed programs, they were shocked to find a commercial forensic tool installed on the suspects' system. As they drilled down further into the filesystem, they started to find documents with their names on them. Again, they were shocked; how did the suspect gain access to Colleague A's information?

The suspect didn't have access to the information.

Colleague A made an error when creating a forensic image. Instead of imaging the suspects' device, they imaged the system drive of their forensic laptop. They ignored the details as they were creating a forensic image. Luckily, the procedure was for each colleague to make a forensic image of the source device, for a total of two forensic images.

While this story is embarrassing, there were no lasting repercussions because we could use the second copy. Imagine how you would feel if you were Colleague A, and there was no second backup to use. How do you explain to your supervisor or the client that you could not complete the task as given, and now you do not have access to the source device?

To help stop that from occurring, we will look at tool validation.

Tool validation

Earlier, we discussed potential attacks on you, your exam, and your findings. The opposing counsel will focus on how you did the exam and what tools you used to perform the exam. Your ability to mitigate the opposing counsel attacks is directly related to your preparation and the documentation you created during the exam. Being aware and following best practices is critical in your ability to defend your actions successfully. How do you do this? By continuing your education. The field is constantly changing, and you must keep aware of those changes.

The level of detail can easily overwhelm new digital forensic investigators as they need to know how to successfully mitigate the opposing counsel's attack. While you need not know the specific programming or code a particular tool uses, you need to know where the artifact found by the tool is located within the filesystem/operating system so you can adequately explain it as you testify or create your report. I have often seen an examiner rely on a checklist provided by a colleague or one they found on the internet and yet have little to no understanding of why the items are on the checklist or the process used to recover the artifact. It can be as simple as recovering a deleted file. If the digital forensic investigator cannot explain the process of how the filesystem processes the user's request to delete a file and how the tool recovered the deleted file, their time testifying will be very uncomfortable. If you cannot explain the basics, the opposition will question your findings.

You need to determine if your tools produce a valid result. As we saw in our previous discussion in *Chapter 2*, *The Forensic Analysis Process*, in the matter of Casey Anthony, the opposing counsel successfully mitigated the digital evidence because of an error reported by the forensic tool. If the forensic tool has been found to be faulty, then the tool may be used to discredit the integrity of the exam and the competence of the examiner.

How do you mitigate the attacks on your process or your tools?

- Understand their functionality
- Document your training
- Take notes during the exam
- Validate the tools

Your testimony about your exam, your findings, and the use of the tools is based on your personal experiences. You cannot testify about someone else's validation. You do not know all the parameters the third party used. This process is something you must do personally. Use the tool against a known dataset to see whether it performs as expected. If you do not validate your forensic tool, how can you testify that it provides an accurate result? How do you answer the question if you get questioned on the stand? It is not uncommon for the opposing counsel to recreate the forensic exam you did. The opposition will attempt to use the same forensic process and forensic tools to determine whether they can get the same result. What happens if they get a different result using the same method and tools? What happens if they get a different result using the same process but different tools? How can you prepare for that attack against yourself or your examination if you do not validate your methods and forensic tools?

As I mentioned earlier, **NIST** has created the **Computer Forensic Reference Dataset**. In addition, you can follow this link to assist you in validating your tools: `https://www.cfreds.nist.gov`. These datasets *provide an investigator with documented sets of simulated digital evidence for examination*. NIST has also provided resources for the creation of your test images.

We can use these datasets in a variety of ways:

- Validation testing
- Proficiency testing
- Training

When using your dataset or a third-party dataset, you must ensure there is documentation on what is contained in the dataset and where the testing data is located within the dataset. In the following example, we will use the DCFL control image provided by NIST.

The following example will use two forensic tools: the Autopsy open source tool, and the X-Ways commercial tool. As shown in the following screenshot, the documentation states there should be two logical files:

```
The following non-system files should be present on the logical level of the disk:

039C8A00  Scientific control.mp3    MD5:   e73a608dfb422a206ce7a62deb90ff9b
029D4A00  Export_me.JPG    MD5:   c0c3892606849fd76a8534ef80956705
```

Figure 3.3: DCFL control image hash values

The documentation provides the logical filename and extension, the hexadecimal offset, and the MD5 hash value for the file (remember that the hash value is the file's digital fingerprint).

In the following screenshot, we are looking at the interface of Autopsy, which shows that there are two logical files (identified by their file extensions): one image file and one audio file.

So far, that matches the documentation we have been given for the control image:

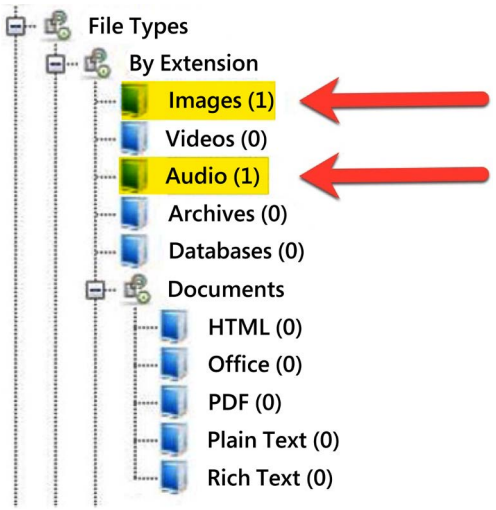


Figure 3.4: DCFL control image – file types

In the following screenshot, we are looking at the interface of X-Ways, and it has also identified two logical files whose filenames match the control:

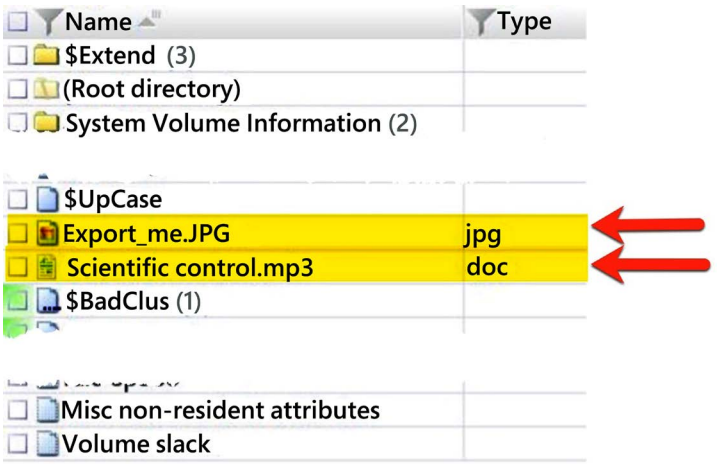


Figure 3.5: DCFL control image – X-Ways logical files

In the following screenshot, we are looking at the metadata of the image file as provided by Autopsy, and we can see that the filename, extension, and hash values match the information provided in the control documents:

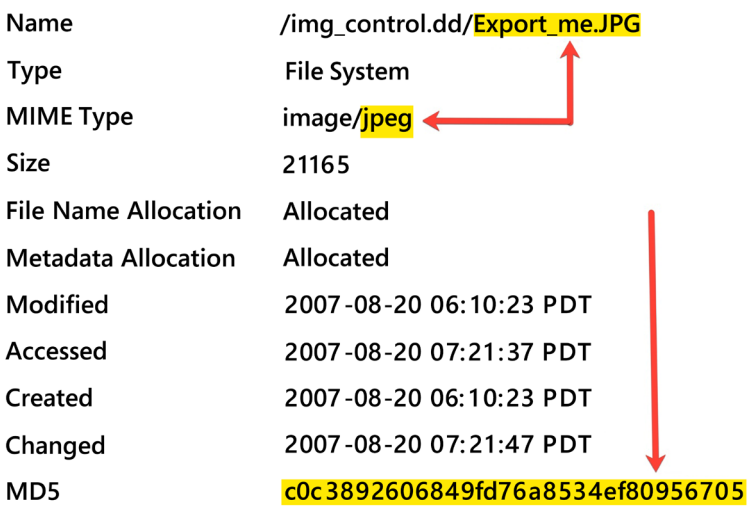


Figure 3.6: DCFL control image – metadata of jpg

In the following screenshot, we are looking at the metadata for the same file in the X-Ways interface, and find it also matches the information provided in the control documents:

Evidence object	control
Name	Export_me.JPG
Type	jpg
Description	existing
Existent	✓
Size	20.7 KB (21,165)
Created	08/20/2007 13:10:23 +0
Modified	08/20/2007 13:10:23 +0
Accessed	08/20/2007 14:21:37 +0
Record changed	08/20/2007 14:21:47 +0
Record changed²	08/20/2007 13:10:23 +0
Ext.	JPG
Pixels	0.4 MP
Analysis	0% skin tones
Hash¹ (MD5)	C0C3892606849FD76A8534EF80956705
Hash² (SHA-1)	4F90640F999271C41A1E77804FD7AAA4F0340D9D
Generator signature	60F38468 (U:Standard 75 Edited)
Device type	unknown
Relevance	3.59

Figure 3.7: DCFL control image – X-Ways metadata of JPG

You can work your way through the rest of the control image to ensure the forensic tool of choice functions correctly and produces accurate results. There are multiple control datasets you can use to validate your tools. You cannot guarantee your tool works appropriately until you conduct the validation test. Your organization should have a policy dictating when the validation should occur and how to document and record the validation test results. If you do not log the validation tests, the opposing counsel can call it into question when they request those records.

That covers the validation of your tools, but what about the storage containers? Let's discuss sterile media and define what it is.

Creating sterile media

Sterile media is also a concept that was emphasized when I first trained. There is an ongoing discussion regarding whether sterile media is still needed in today's forensic environment. The decision to use sterile media to store the forensic data will be based on the acquisition and the type of examination you will use. Sterile media can be used before the start of the forensic process and at the end of the forensic process. There are multiple reasons for using sterile media, which we will now discuss. When digital forensics was first starting, we could not create a forensic image; we were forced to make a forensic copy to perform our examination on. Remember, we talked about a forensic copy in *Chapter 2*, *The Forensic Analysis Process*, and defined a forensic copy as follows:

> *"A straight bit-for-bit copy of the source to the destination. This is not common in today's environment; ensure that your destination device has no old data from previous investigations. You do not want to cause cross-contamination between the current digital forensic investigation and a past investigation. We will recover deleted files, file slack, and partition slack."*

If your source and destination were the same make, model, and capacity, then you would potentially not have an issue. In real life, this rarely happens, so to be safe, make sure that you use a larger-capacity device as your destination device. After you copy the data from the source device to the destination device, you will have unallocated space on the destination device.

Suppose you did not wipe or use sterilized media as your destination device. In that case, it is possible that there would be pre-existing data on your destination device, and this creates the possibility for the co-mingling of data. So, when using the forensic copy process and looking for data in unallocated space or slack space, you must use sterilized media.

There have been cases where the examiner has used a newly purchased storage device or had the storage device provided to them; they still must wipe the drive and sterilize it of all pre-existing data. Suppose you do not, and the destination device is provided to the opposing counsel, and they find data not relevant to the matter at hand. In that case, it can call into question the integrity of the exam and the examiner's competence.

What do you do with old storage devices that contain digital evidence? Do you destroy them? Do you recycle them? Do you turn them over to your organization and not worry about it? Before the storage device leaves your control, you must wipe that device to ensure no confidential information or contraband is released to an unauthorized entity. That way, you can be positive that no one can find any data relating to any digital forensic exam on the devices.

So, what exactly is sterile media? It is simply where every byte on the device is overwritten with a hexadecimal `00`. Technically, you can use any character you wish. It is much easier to verify whether the sterilization of media was successful if you use the hexadecimal `00`. We use the 64-bit checksum to validate the sterilization process. If you run the 64-bit checksum against the sterile media, you will get zeros as the generated checksum value. I do not recommend using the MD5 or SHA–1 hashing algorithms to verify the sterilization process. They will not give you a value you can use to immediately identify the successful sterilization process.

Let's look at the sterilization process. We will use PALADIN from SUMURI Forensics. PALADIN is a live bootable version of Ubuntu. This means you have to have PALADIN installed on a USB or a DVD/CD. Using a USB or CD/DVD will allow the computer to boot to the operating system contained on the USB/CD/DVD. PALADIN will enable you to access the host computer while not modifying digital evidence. The PALADIN toolbox allows us to create forensic images, convert forensic images, and create sterile media.

In the following screenshot, I have opened the PALADIN toolbox and selected **Disk Manager**:

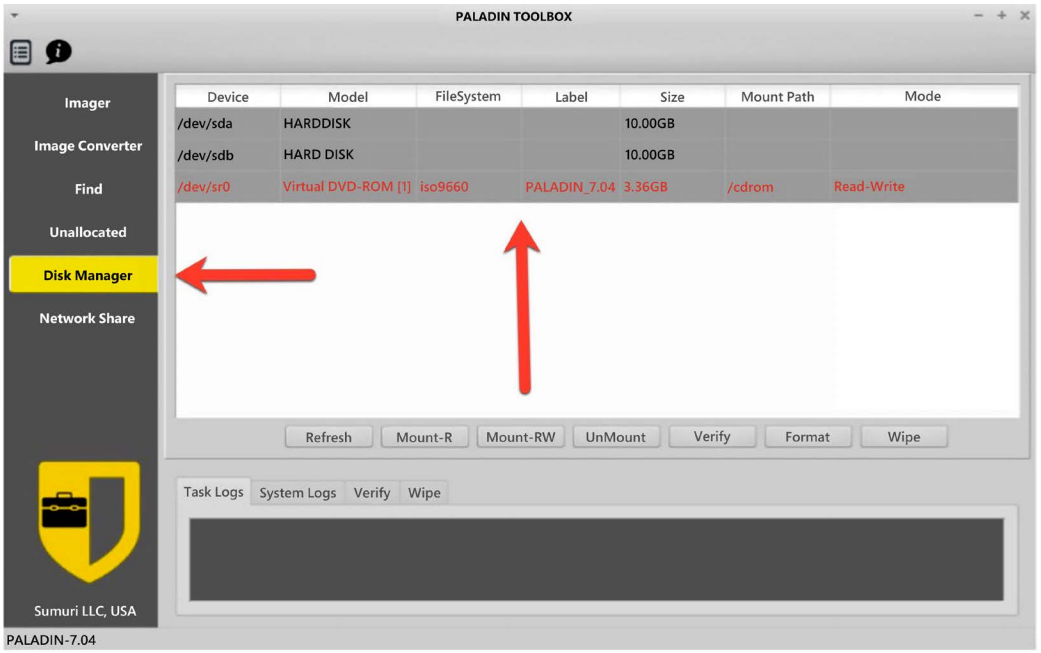


Figure 3.8: PALADIN toolbox

As we look at the preceding screenshot, we see three devices on the system: two 10 GB hard drives and a CD-ROM drive. The CD-ROM is the PALADIN operating system, while the two hard drives are the storage drives on the computer. We will wipe one of these storage drives, in this case, `/dev/sdb`. As you look at the interface below the device listing, you will see various options. At the far right, we have a button titled **Wipe**. We will select this button after we left-click on the device we want to wipe. You do not want to mount the device before wiping it.

Once PALADIN has completed the wiping/sterilization process, it will show you a log of the processes used. The following screenshot shows that it input the pattern `00` and the number of sectors that were overwritten. The last line tells us when the operation was completed. You need to save this log and store it with the storage device you have just wiped:

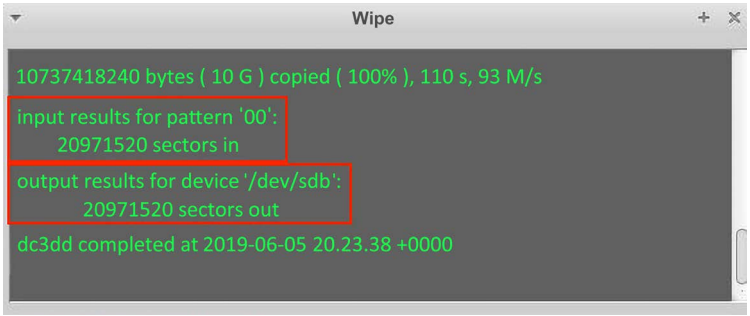


Figure 3.9: PALADIN toolbox – Results of wiping

But how do we verify the results to ensure the tool works as expected? Here, we will use the commercial tool X-Ways Forensics. X-Ways Forensics is a commercial tool offered by X-Ways Software Technology AG and is my go-to tool when conducting a digital forensic exam. I find its ease of installation, price, and the ability to use it on multiple platforms attractive features of this tool. It's not that other tools are not worthwhile; this is just my personal preference.

We have added the device to X-Ways, and now we want to verify the sterilization process we used with PALADIN. Follow these steps to do so:

1. Right-click on the device and select **Properties**:

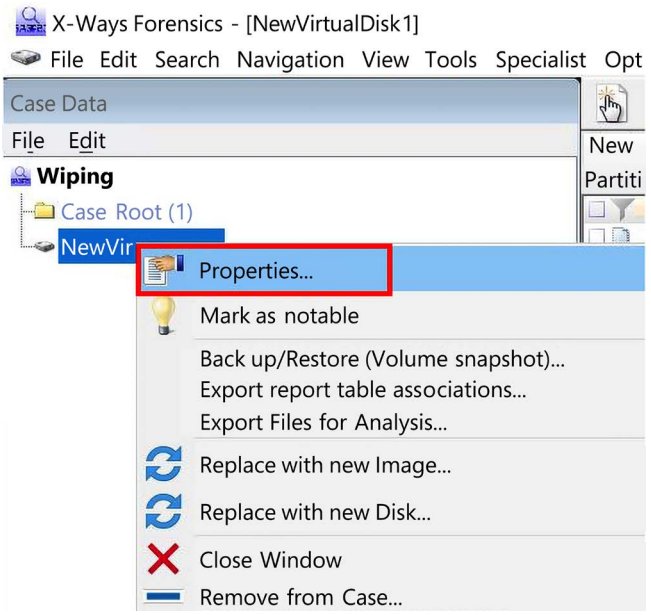


Figure 3.10: X-Ways – Properties menu

2. The **Properties** window of the device will appear. Toward the bottom right, you will find the **Compute hash** button. When we left-click it, we will see the hashing options available to us:

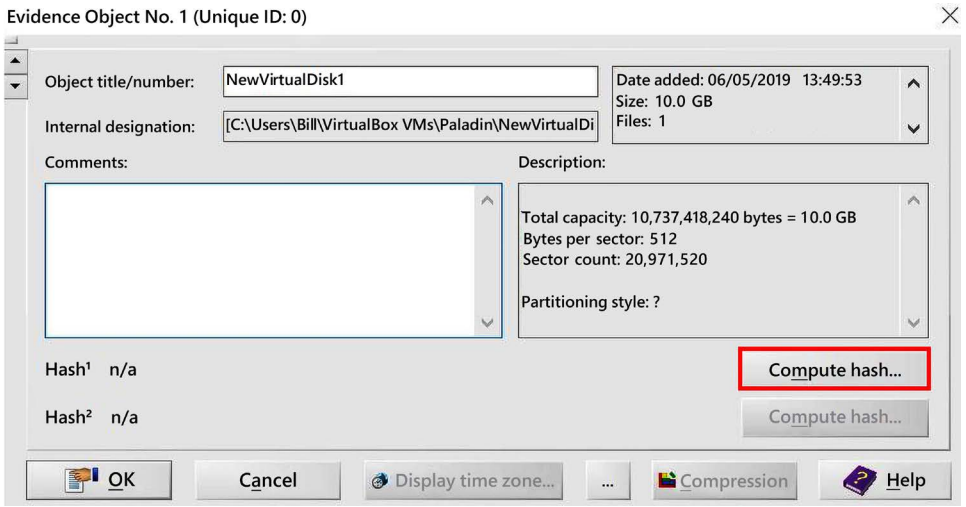


Figure 3.11: X-Ways – Hashing configuration

3. You will want to select **Checksum (64 bit)**, which will return zeros if the sterilization process worked correctly:

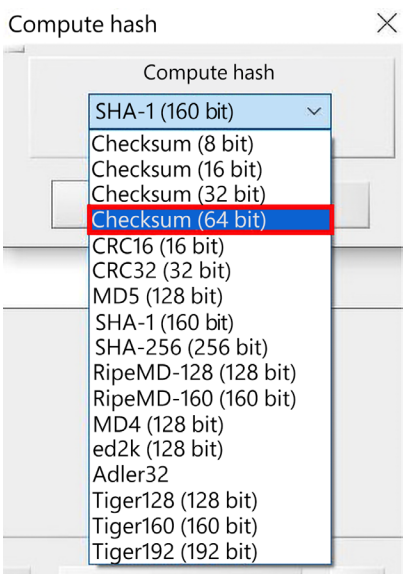


Figure 3.12: X-Ways – Selecting Checksum (64 bit)

If you choose MD5, SHA–1, or any other hashing algorithm, you will get a value for the device, but that value will not let you determine whether there is any residual data left on the device.

4. As shown in the following screenshot, the checksum result is a string of zeros. This informs us that the media sterilization process has worked correctly. We have also just validated another aspect of our forensic tools:

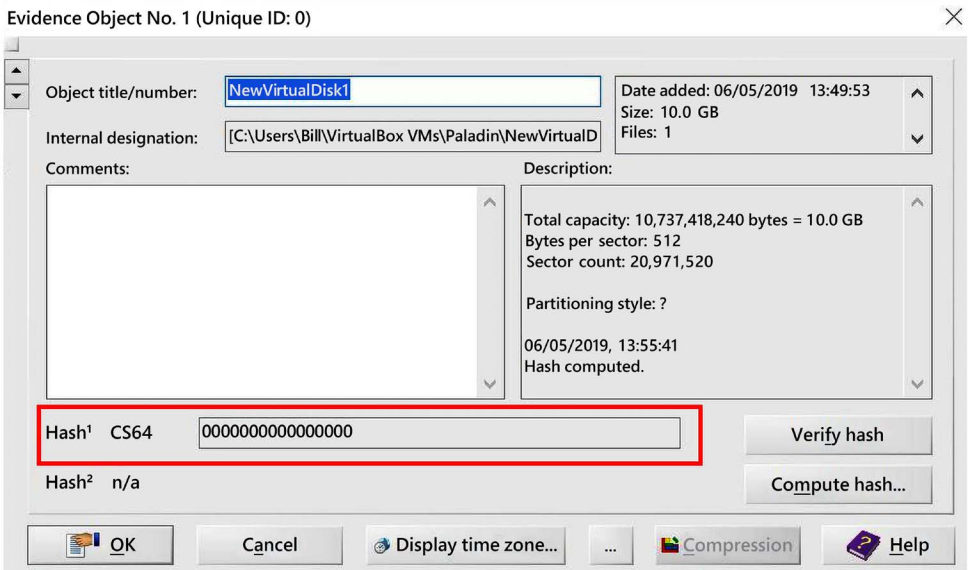


Figure 3.13: X-Ways – Checksum result

We now have sterile media, but how do we protect the original evidence? The answer is to perform write blocking, which we will discuss next.

Understanding write blocking

Write blocking is at the core of the forensic examination environment. With the fragility of digital evidence, we want to ensure we do not change a single bit of data on the source device. Evidence handling is an essential function of the examination process, and we must ensure that we meet all the requirements to avoid altering or damaging the evidence. For example, if I plug the device into a Windows-based computer system, to enhance the user's experience, the operating system scans and makes writes on that device that change the evidence. To prevent the alteration of the source device, we must use a write blocker.

You have a choice of utilizing a "hardware write blocker" or a "software write blocker."

Hardware write blocker

As the operating system issues commands, it will read/write from the source device. A hardware write blocker is a device that intercepts and prevents any modification to the source device. It is physically connected between the computer and the source device to accomplish this. There are standalone hardware write blockers that are self-contained that allow you to attach the source and destination device and then create the forensic image.

The following image shows the Tableau Forensic SATA/IDE Forensic bridge T35u that the Department of Homeland Security tested in October 2018. This device allows you to forensically acquire SATA and IDE devices by using the computer's USB 3.0 connection:

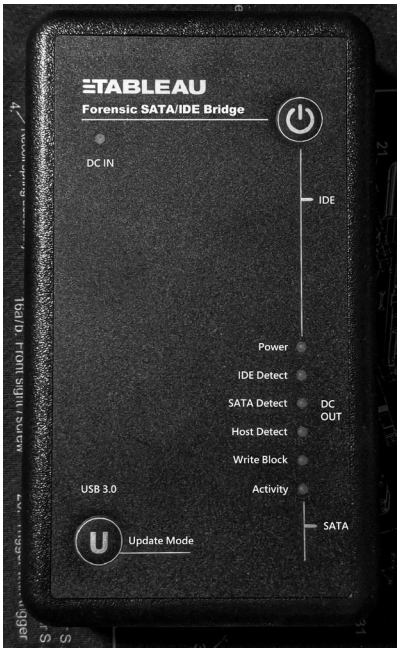


Figure 3.14: Tableau Writeblocker

NIST has created the Computer Forensics Tool Testing Program, which lists the testing results for hardware write blockers (`https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/hardware`). Here, you can find the report on the T35u and other devices.

Software write blocker

Software write blocking is where a change is made to the operating system to stop it from making writes to the device. For example, there is a registry change you can make for a Windows-based system to prevent writes to attached USB devices.

Another option is to utilize a bootable operating system, such as PALADIN or Win FE.

In the following screenshot, we can see the PALADIN toolbox, which lists the drives in the system. By default, PALADIN does not automatically mount attached storage devices. This means it makes no modifications and doesn't look at the devices until you tell the software to mount the device:

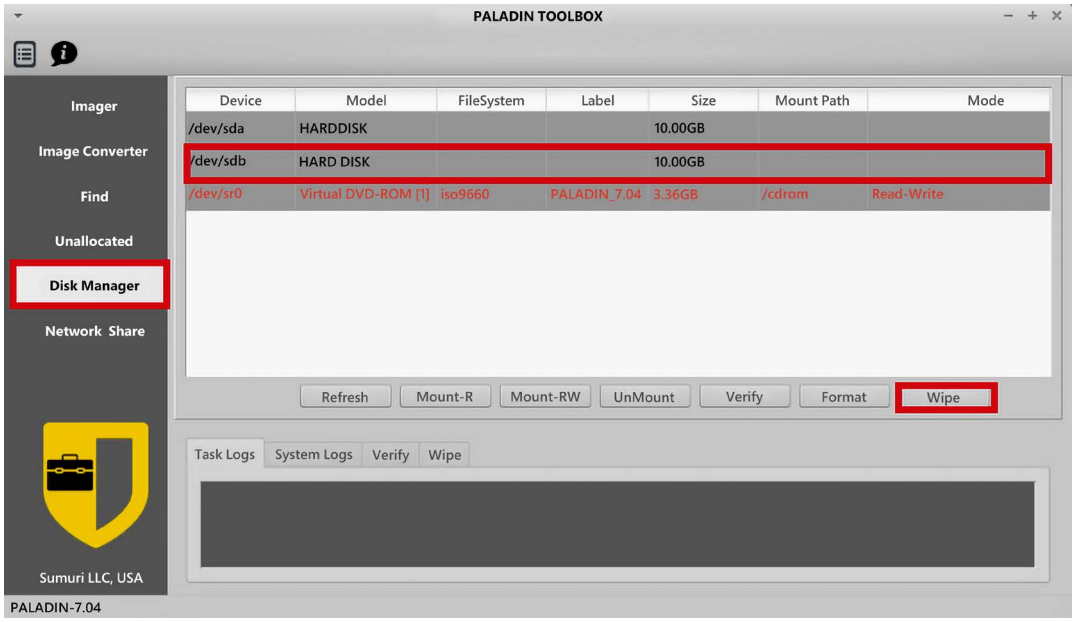


Figure 3.15: PALADIN toolbox – Disk Manager

There are two options when you mount a device:

- Read-only
- Read/write

It would be best if you did not mount a device read/write unless you want to change the device. For example, if you're going to create a forensic image of the device, you should mount the device read-only.

As shown in the following screenshot, there is a column listed as **Mode**, and we can see that the CD-ROM is mounted as read/write and is highlighted in red, while the hard disk is green and shows read-only:

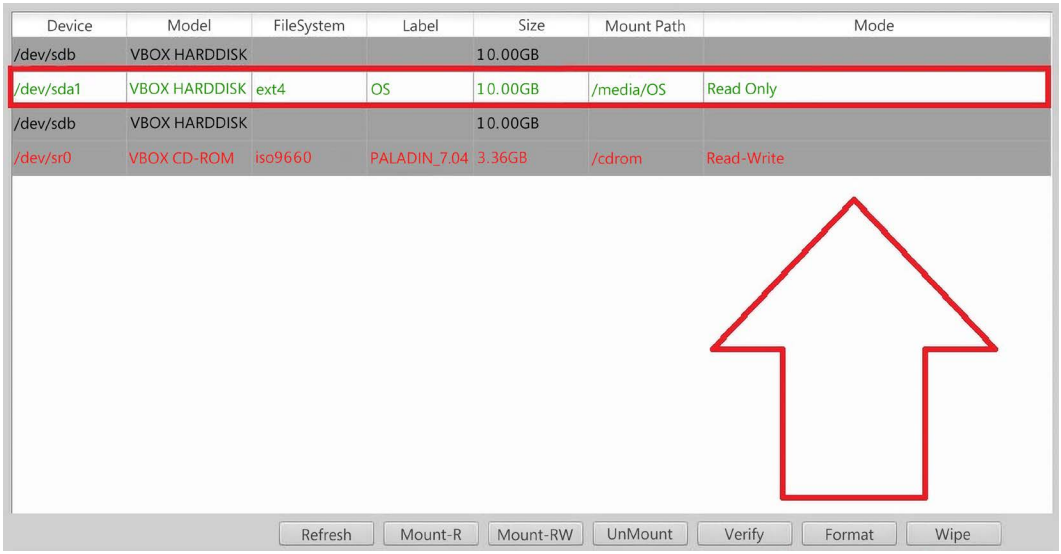


Figure 3.16: PALADIN toolbox – Disk Manager Mode status

Now that we have protected the source device with the original evidence, let's move on to creating the forensic image.

Defining forensic imaging

I continue to stress that we never want to change the source device/digital evidence. That is why we never conduct a digital forensic examination on the original device. You should only conduct your digital forensic analysis on a copy, not the original device. You must remember the forensic copy you make will also be considered the evidence and will have the same evidentiary weight as the source device in terms of evidence. What are we transferring from the source device into our forensic copy? Everything! I want to look at allocated files, deleted files, slack space, unallocated space, and unpartitioned space. I want to collect every bit on the source device. Earlier in this book, in *Chapter 2*, *The Forensic Analysis Process*, I gave you the following definitions:

- **Forensic copy**: This is a straight bit-for-bit copy of the source to the destination. This is not common in today's environment, so ensure that your destination device has no old data from previous investigations. You do not want to cause cross-contamination between the current digital forensic investigation and a past investigation.

We will recover deleted files, file slack, and partition slack. We will discuss wiping hard drives later on in this book.

- **Forensic image or forensic evidence file**: We are creating a bit-for-bit copy of the source device, but we store that data in a forensic image format. This could be a DD image, an E01 image, or an AFF image. We take that source data and wrap it in a protective wrapper of the forensic image. We will recover deleted files, file slack, and partition slack.
- **Logical forensic image**: Sometimes, we are restricted to only accessing specific datasets. They do not allow us to access the entire container. We cannot create a bit-for-bit copy of a forensic image or a forensic copy. This could be when we are extracting data from a server, and we cannot shut the server down to create a forensic image from the source hard drives. Due to this, we can make logical copies of the files and folders pertinent to the investigation. We will NOT be able to recover deleted files, file slack, and partition slack.

For the forensic copy and the forensic image, we will acquire every bit on the source device; if there are restrictions, then we will only be able to copy the logical files. We will then put the logical files into a forensic container, which will then encapsulate them in a protective format to prevent any alteration to the data after we've collected it. These are not backups, as you might see in the corporate environment. In the corporate environment, they have not created those backups in a forensically sound manner. They will not contain any information about file slack, unallocated space, deleted files, or any piece of data that is not maintained by the filesystem. I do not recommend doing digital forensic examinations where the agent used commercial/open source backup software to collect the evidence. Only use a trusted and verified forensic tool to collect the dataset in a forensically sound manner.

There are two common formats for a forensic image (there are others, but DD and E01 are the two formats I consistently see used by government and corporate digital forensic investigators). Let's look at them now.

DD image

DD is a UNIX command, and some call it the oldest imaging tool available that has migrated to other platforms. You can find versions of DD that work on Linux, Windows, or Mac, and they all work in relatively the same manner. They designed it to copy data from a source device to a destination device. Fairly simple, is it not?

With the DD command, you can create a forensic copy, where every byte from the source device is sent to the destination device. You also have the option of creating a flat file/RAW image of the source device. The image file can be a single file or segment into multiple file pieces.

The DD command does not compress the forensic image, so you must ensure that your destination device has the same capacity or a capacity greater than the source device.

The following screenshot shows an example of a DD image that has not been segmented and is 21 GB in size. Depending on the format of your storage device, you may have to segment the forensic image to meet the filesystem's constraints. You may also see different file extensions for the DD image: `.001`, `.dd`, and `.img` are common file extensions:

Figure 3.17: DD image example

`dcfldd` (`http://dcfldd.sourceforge.net/`) is a version of the `dd` command that has incorporated additional features, such as the following:

- Hashing on the fly
- Status output
- Disk wiping
- Verifying an image or wipe
- Multiple outputs
- Splitting outputs
- Piped output and logs

`dcfldd` was written by Nick Harbor (former employee of DCFL).

Note

`dcfldd` has an issue with imaging faulty drives. NIST reported that `dcfldd` will misalign the data in the image after a faulty sector is encountered on the source device. You can visit this link to find out more: `https://www.dhs.gov/sites/default/files/publications/DCFLDD%201%203%204-1%20Test%20Report_updated.pdf`.

`dc3dd` (`https://sourceforge.net/projects/dc3dd/`) is another version of the `dd` command. Whereas `dcfldd` is a fork of the `dd` command, `dc3dd` is a patch of the `dd` command. While these options are similar, they have a slightly different code base and feature sets. When the `dd` command is updated, `dc3dd` is also automatically updated.

Some features available on `dc3dd` include the following:

- The ability to have on-the-fly hashing
- The ability to write errors directly to the file
- The ability to create error log pattern wiping
- The ability to verify the mode
- The ability to create progress reports
- The ability to split outputs

Jesse Kornblum developed `dc3dd` at the DoD Cybercrime Center. The next format we will discuss is the EnCase evidence file.

EnCase evidence file

The other forensic image format we will discuss is the EnCase evidence file, commonly referred to as `e01/ex01`, or the expert witness file format. The `dd` command is a direct bit-for-bit copy; the `e01/ex01` format is also a bit-for-bit copy, but includes additional data within the forensic image format. EnCase Forensics is a commercial forensic tool created by Guidance Software (now Open Text) and was one of the first commercial digital forensic tools available for use. Andy Rosen created the forensic image file known as the `e01/ex01` format, **Expert Witness Format (EWF)**, or the EnCase image file format. The current version of the EnCase Evidence file is **Ex01** and should be used, wherever possible, instead of the `e01/ex01` format. AES256 encryption, LZ compression, MD5, and SHA-1 hashing are included in the `ex01` format, which is an upgrade from the `e01/ex01` format.

The `e01/ex01` file format is a forensic image that encapsulates the raw data from the source device to prevent changes from occurring after acquisition. While the `dd` image only contains the data from the source device, the `e01/ex01` forensic image contains header information, including evidence name/number, acquisition dates and times, investigator notes, and information about the forensic tool used to create the forensic image. The `e01/ex01` forensic image also has additional security features to ensure the validity of the forensic image. There is a CRC calculation every `64` sectors as the forensic image file is created. It stores the CRC value within the forensic image so that your forensic tool can verify it every time the forensic image is utilized.

As seen in the following screenshot, you can see the layout of the `e01/ex01` file format. The **Case Information** is at the head of the file, a **CRC** is created from the header information, and then a `64`-sector block is added to the image file, and a **CRC** value is created for that `64`-sector block and added to the forensic image.

The **data** block and the corresponding **CRC** block process continue until it acquires the entire source device. Once the process has reached the end of the source device, an **MD5** hash value is generated of all the data blocks (and *only* the data blocks) and attached to the end of the forensic image. With the `eo1/ex01` forensic image, you also can enable compression; you do not have that ability with a `dd` image:

Figure 3.18: Expert Witness Format file layout

Next, we need to discuss SSD drives because SSD devices have some special considerations when it comes to imaging.

SSD device

Solid-State Storage (SSD) is a newer storage device that is becoming more prevalent in the business and consumer market. As the price of solid-state storage devices comes down, their use will increase. SSDs create a unique issue regarding digital forensics. There are automated processes that are run through the firmware of the device. The digital forensic examiner has no way to stop or intercept the firmware commands on the storage device. Wear leveling is a feature that ensures the storage blocks on the device are used at a similar rate. If some blocks on the storage device are overused or if the blocks are not equal, this can lead to the premature failure of some storage blocks. The firmware will decide where to move the data on the storage device. Plugging in the solid-state device can cause the firmware to move data around.

Garbage collection is the other firmware function that causes concern in the digital forensic world. When a user deletes a file, formats a partition, or deletes partitions, the firmware starts the garbage collection process with the trim command. Unfortunately, this causes the now unallocated space to be wiped, and the deleted data will no longer be accessible.

It is possible that after you create the forensic image of the source device and have your pre- and post-hash values, after days, weeks, or even months later, when you hash the source device again, it may come back with a different hash value. It also may be possible with large capacity drives with long imaging times that the pre-and post- imaging hash values do not match.

If you can explain the issues with SSD drives, you should not have any problems.

Next, we will move on to imaging tools.

Imaging tools

Remember that you do not want to conduct your investigation on the original media, especially SSD devices. As I mentioned in the prior section, the wear leveling and **trim** commands will change the original evidence. There are many forensic tools for you to use for your imaging needs; we will now discuss two freely available tools and how to create a forensic image.

FTK Imager

FTK Imager is a free tool offered by AccessData. You can visit `https://accessdata.com/product-download/ftk-imager-version-4.2.0`, which will help you create a hash value for the source device, image it, and then create the post hash value to verify that no changes were made to the source device during the imaging process:

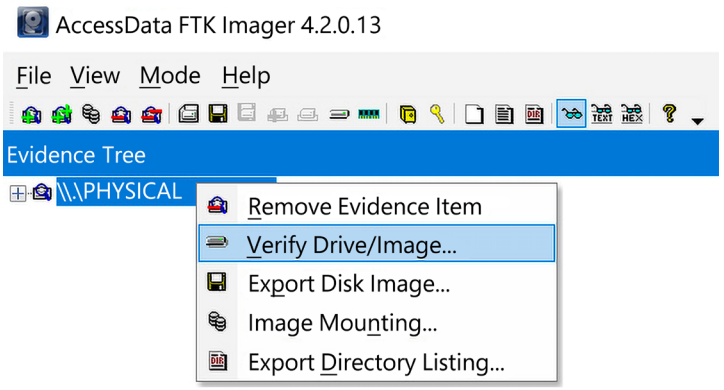


Figure 3.19: FTK Imager – Creating a hash value

I will now walk through the steps to create a forensic image. After using the appropriate write blocker, we attached a 2 GB USB thumb drive (Kingston Data Traveler) to the system. We will now obtain the pre-hash value of the device. This hash value gives us the starting value of the device. This value will be used to determine whether any alterations have occurred on the source device. In the preceding screenshot, you can see we have loaded the physical device into FTK Imager and then right-clicked it to bring up the **Verify Drive/Image** menu. Simply click on **Verify Drive/Image** and let FTK Imager do its work.

The results will be displayed after, as shown in the following screenshot:

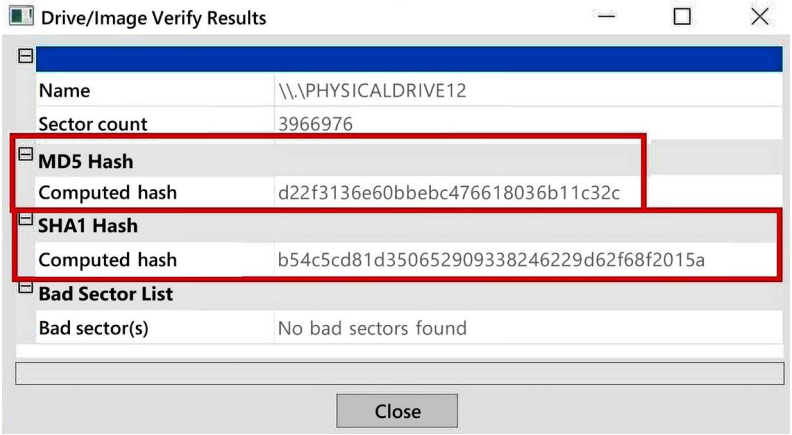


Figure 3.20: FTK Imager – Drive/Image Verify Results

Now that we know what the starting hash value is, we can proceed with the creation of the forensic image:

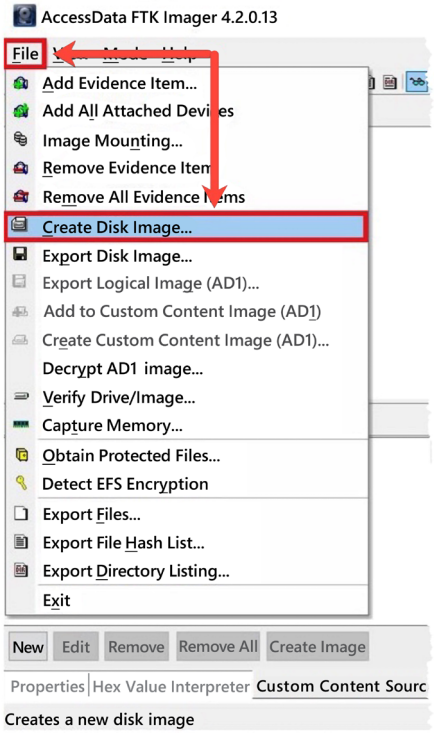


Figure 3.21: FTK Imager – Create Disk Image menu

As shown in the preceding screenshot, click on the **File** menu and select **Create Disk Image**.

From here, you will select your source. With FTK Imager, we have some choices to make:

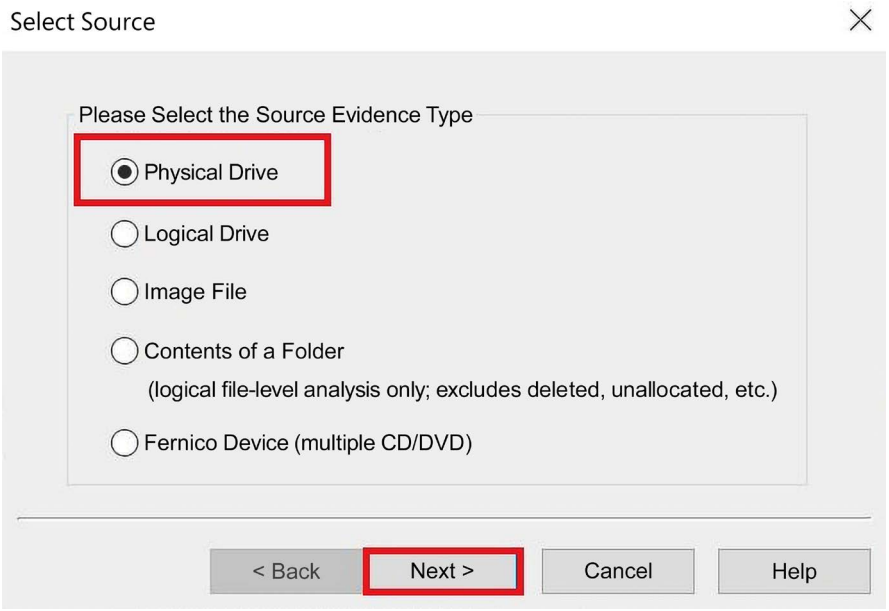


Figure 3.22: FTK Imager – Select Source menu

Let's discuss each option in detail:

- **Physical Drive**: The physical device will give us every bit of data on the source.
- **Logical Drive**: You will only get the data within the partition boundaries. If there are deleted partitions or data outside of the boundaries on the source device, you will not be able to recover that data.
- **Image File**: If you want to change the format of the forensic image; for example, change it from an `eo1/ex01` to a `dd` image.
- **Contents of a Folder**: You will only get the logical data. You will not get deleted data or unallocated space. Sometimes, you may not be able to shut the system down to create a physical image, such as a server, so you have to grab the logical files for analysis.
- **Fernico Device**: Use this option if you have a Fernico FAR system.

Since we want to get all the data on the device, we will select **Physical Drive**:

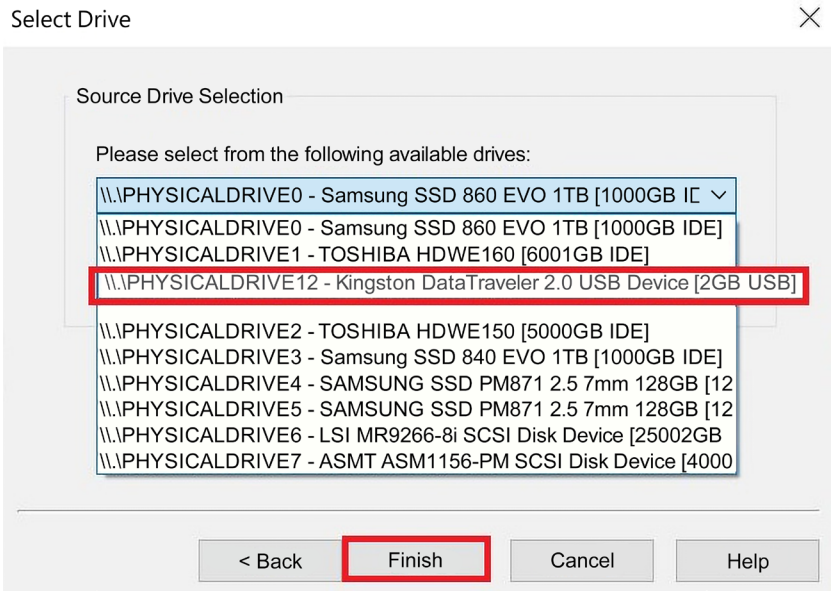


Figure 3.23: FTK Imager – Select Drive menu

You will then be presented with the **Select Drive** dialog. In the preceding screenshot, there are a lot of physical drives being presented, so you must take care that you select the correct device!

Note

You can use Windows (or your OS of choice) Disk Manager to get the physical device number.

We want to select physical device 12, which is the Kingston Data Traveler:

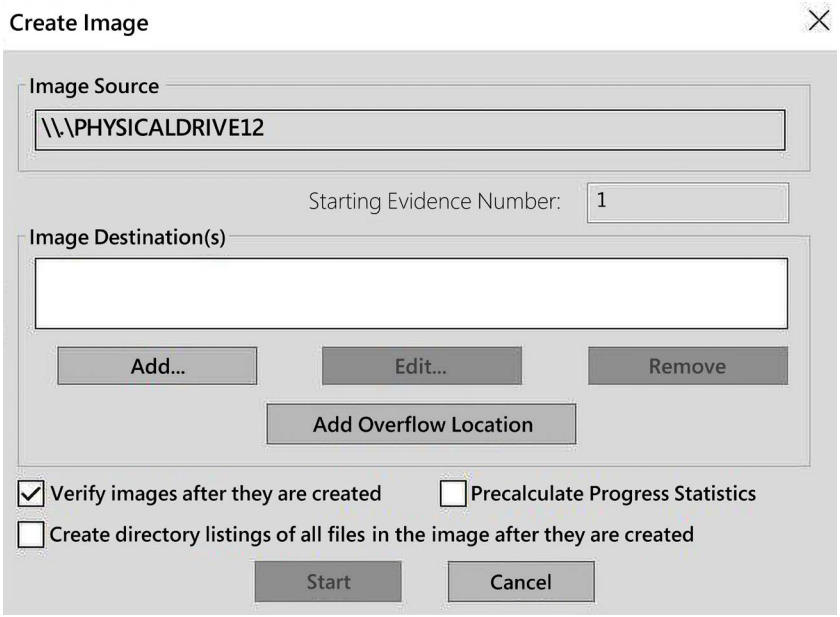


Figure 3.24: FTK Imager – Create Image menu

Now, left-click on the **Add** button to select where you want to save the forensic image and what kind of forensic image you wish to create:

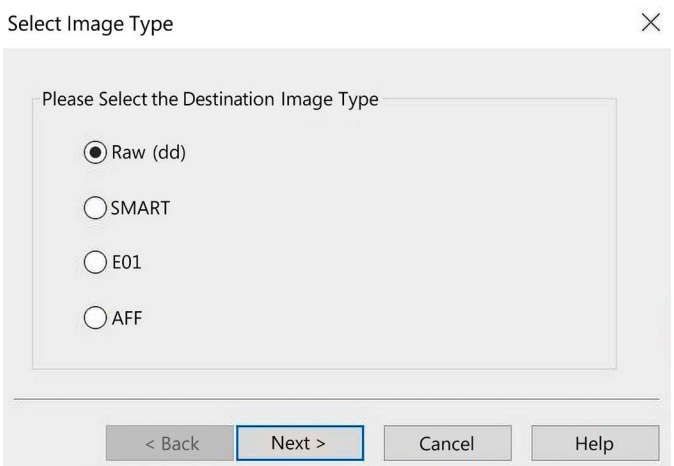


Figure 3.25: FTK Imager – Select Image Type menu

What kind of forensic image do you want to create? You have the choice of four options:

- **Raw (dd)**
- **SMART**
- **E01**
- **AFF**

We have already discussed the two most common formats: `dd` and `e01/ex01`. You can also create two other types of forensic images:

- **SMART**: SMART forensics is a commercial forensic tool on the Linux platform offered by ASR that can be found at `http://www.asrdata.com`. It can create compressed or uncompressed forensic images and can segment forensic images.
- **AFF**: **Advanced Forensics Format** (**AFF**) is an open source format for the creation of forensic images. The goal of the designers was to create a non-proprietary forensic imaging format. Simson Garfinkel and Basis Technology originally developed AFF (you may find multiple proprietary non-standardized versions of AFF4. This is because commercial organizations created the non-standardized versions by adding/changing the AFF standard and have not shared the changes with the community).

I do not recall creating a forensic image that was not in EnCase format or a `dd` image. My preference is to create a `dd` image because it is faster than creating an `e01/ex01` forensic image. Once the examination is complete, I convert the `dd` image into the `e01/ex01` format with high compression to help reduce the file size.

Once you've selected the forensic image format, you will be asked to enter the evidence item information (as shown in the following screenshot), which comprises the following:

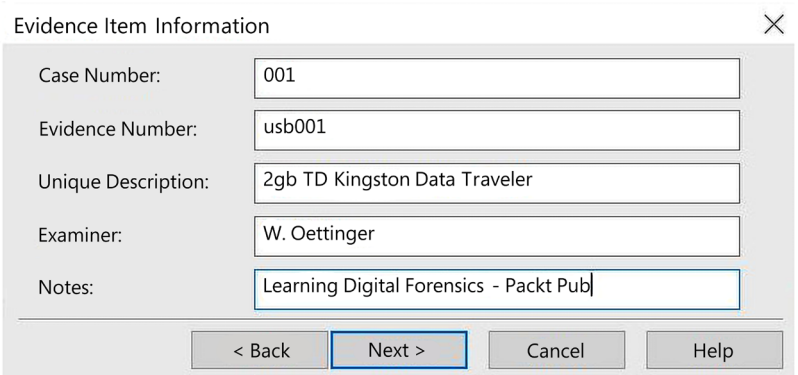


Figure 3.26: FTK Imager – Evidence Item Information window

Let's discuss each option in detail:

- **Case Number**: This should be the overall identifier for the investigation.
- **Evidence Number**: This should be an identifier to help you track the digital evidence. If you have an extensive investigation with multiple source devices, this will help you accurately identify what forensic image you are working on.
- **Unique Description**: This is where I would add the make, model, capacity, and serial number of the source device.
- **Notes**: This is where I would add some specific details about where the source device came from, such as a laptop or desktop.

Your next option is to select the destination (as shown in the following screenshot) for the forensic image in the image destination folder. This could be a storage device attached to the local computer, a connected RAID device, or a form of **network-attached storage (NAS)**:

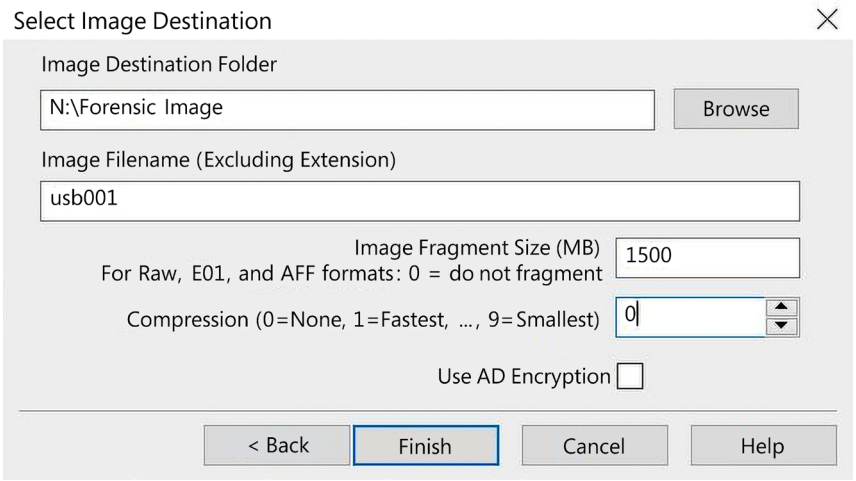


Figure 3.27: FTK Imager – Select Image Destination window

Next, you need to make a selection regarding the filename. I recommend using a similar identifier as the evidence number to help avoid confusion.

Image Fragmentation Size will come into play, depending on the filesystem on the storage device and how you will archive the data. In the past, I used a 2 GB fragment size to ensure the forensic image could be used with multiple filesystems. If I do not expect the forensic image to leave my environment, I will not use a fragmented image.

Some filesystems have a limitation on the maximum file size. For example, FAT32 has a 4 GB maximum file size, while ExFAT, HFS+, APFS, and NTFS do not. You must know which filesystems have a file size limitation.

I rarely use compression because of the increase in time used to create the forensic image.

Your last option is to encrypt the forensic image. If you encrypt the forensic image, make sure you use a password you will not forget. If you forget the password, you cannot use the forensic image.

Once you have completed answering the requested information, as depicted in the preceding screenshot, you will see the **Create Image** window, showing the options you have selected. You also have the option to add a second destination to create two forensic images at a time:

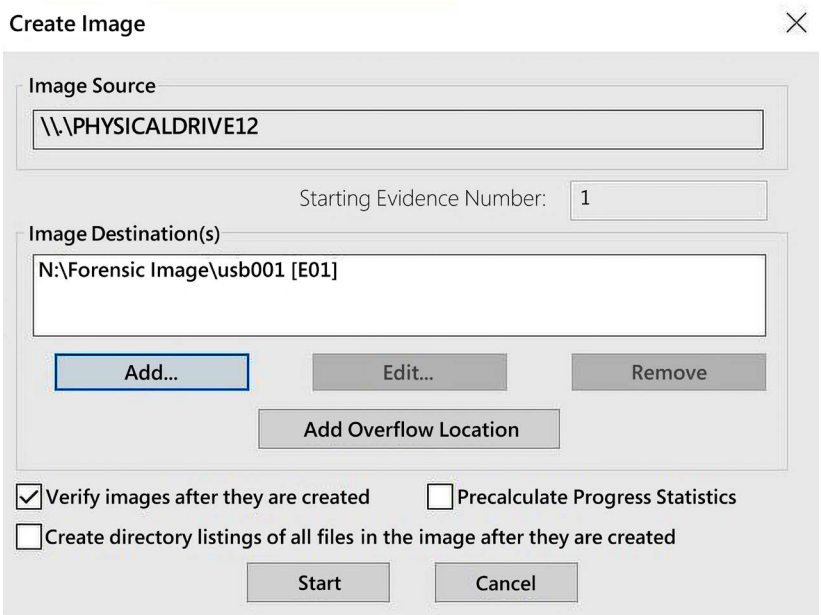


Figure 3.28: FTK Imager – Create Image window

Once FTK Imager has completed creating the forensic image, it will provide you with a status update showing the elapsed time:

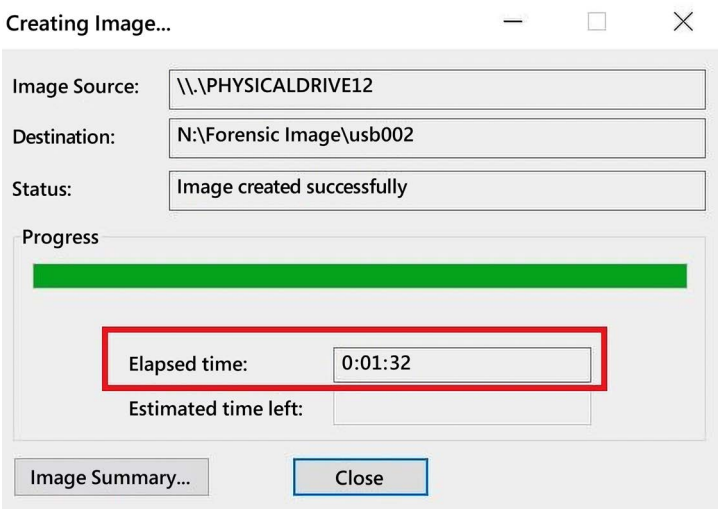


Figure 3.29: FTK Imager – Completed Creating Image window

This will also show you the results window, as shown in the following screenshot (a text file is also automatically created and stored in the same location as the forensic image):

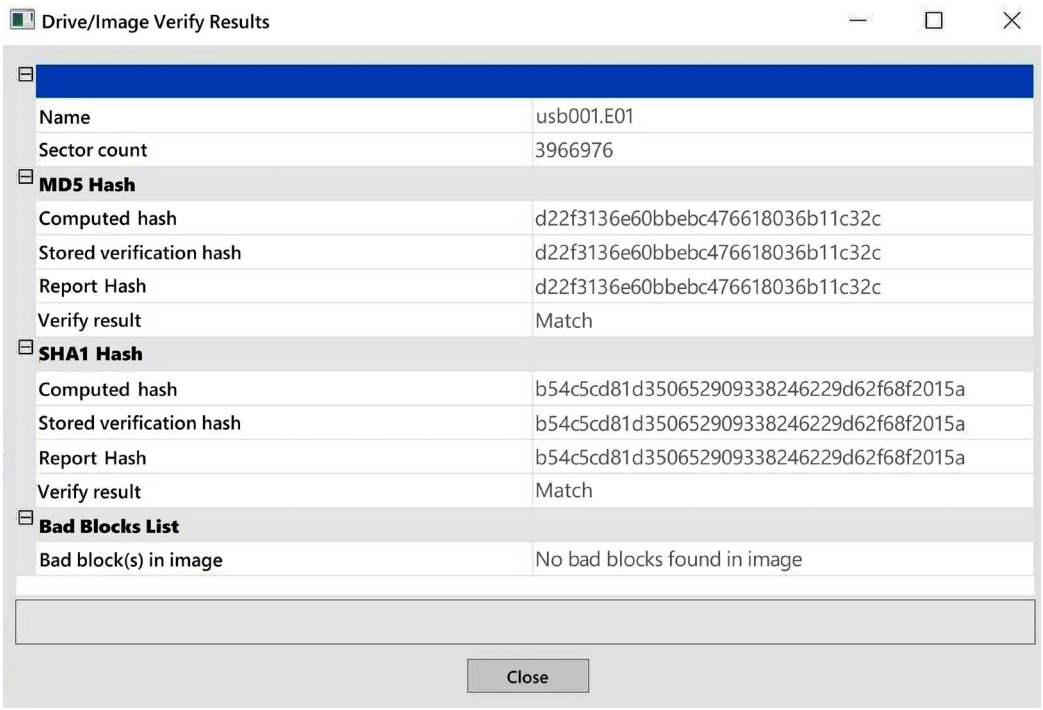


Figure 3.30: FTK Imager – Final verification window

FTK Imager is not the only tool you can use to create a forensic image. One open source forensic tool that you can use is PALADIN. PALADIN has many features, but we will only discuss how it creates a forensic image here.

PALADIN

SUMURI's PALADIN is a Linux distribution based on Ubuntu that allows the collection of digital evidence in a forensically sound manner. The following screenshot shows the desktop you will see when you boot up PALADIN:

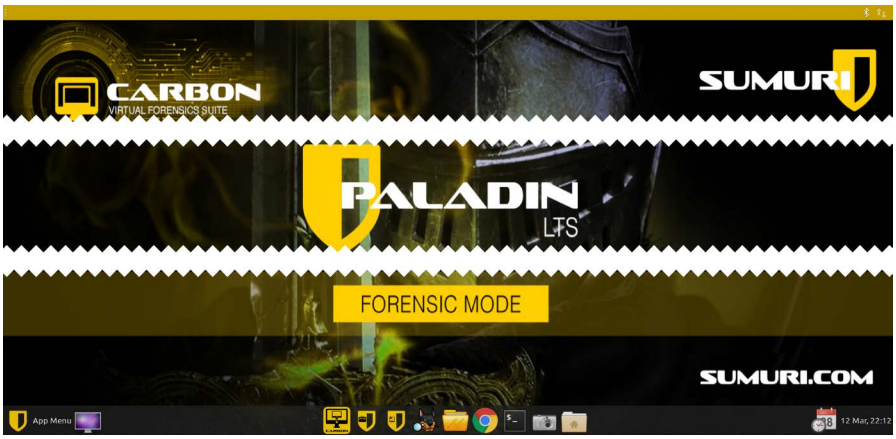


Figure 3.31: PALADIN – Desktop

To create a forensic image with PALADIN, we will follow the same general steps that we did for FTK Imager, with the exception that we do not have to use a hardware write blocker. PALADIN is a live distribution of Ubuntu, so you will have to boot your computer to either a USB device or a CD/DVD. Once you see the desktop shown in the preceding screenshot, you are ready to start imaging:

1. Left-click on the PALADIN toolbox icon to get started.
2. Once the PALADIN toolbox opens, left-click on **Disk Manager** (as shown in the following screenshot) to see what devices are attached to the system. You will see there are three SATA devices on the system:

    - SDA–20 GB hard drive.
    - SDB–256 GB thumb drive with one partition (sdb1).
    - SDC–2 GB thumb drive.

    All three devices are represented in black text.

3. Once PALADIN mounts the device, the text will change to green for read-only access and to red for read/write access:

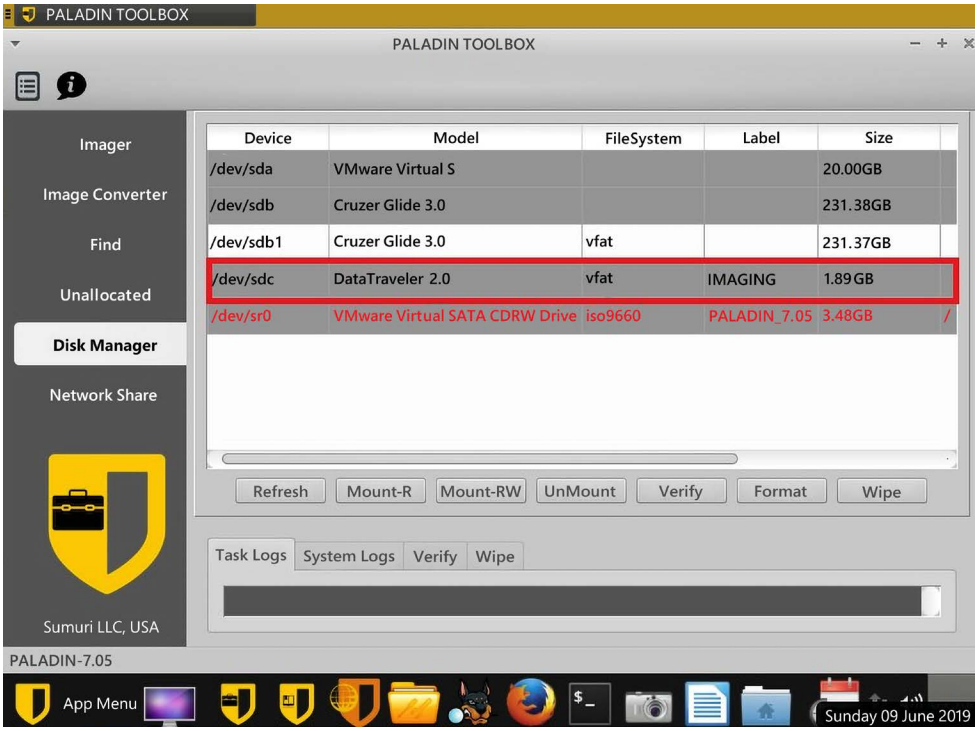


Figure 3.32: PALADIN toolbox

4. Before starting the forensic imaging process, we must pre-hash the source device. Just select the source device and then click the **Verify** button while in **Disk Manager**. You will see the output shown in the following screenshot:

```
dc3dd 7.2.641 started at 2019-06-09 16:43:43 +0000
compiled options:
command line: dc3dd of=/dev/null hash=md5 hash=sha1 if=/dev/sdc hlog=/tmp/
000AEBFFB4C45B8903020517_06-09-2019-16-43-43_verify.log

input results for device `/dev/sdc':
  d22f3136e60bbebc476618036b11c32c (md5)
  b54c5cd81d350652909338246229d62f68f2015a (sha1)

output results for file `/dev/null':
```

Figure 3.33: PALADIN – Hash results

5. In the following screenshot, you now have the option to choose the source device, the forensic image type you want to create, and the destination location:

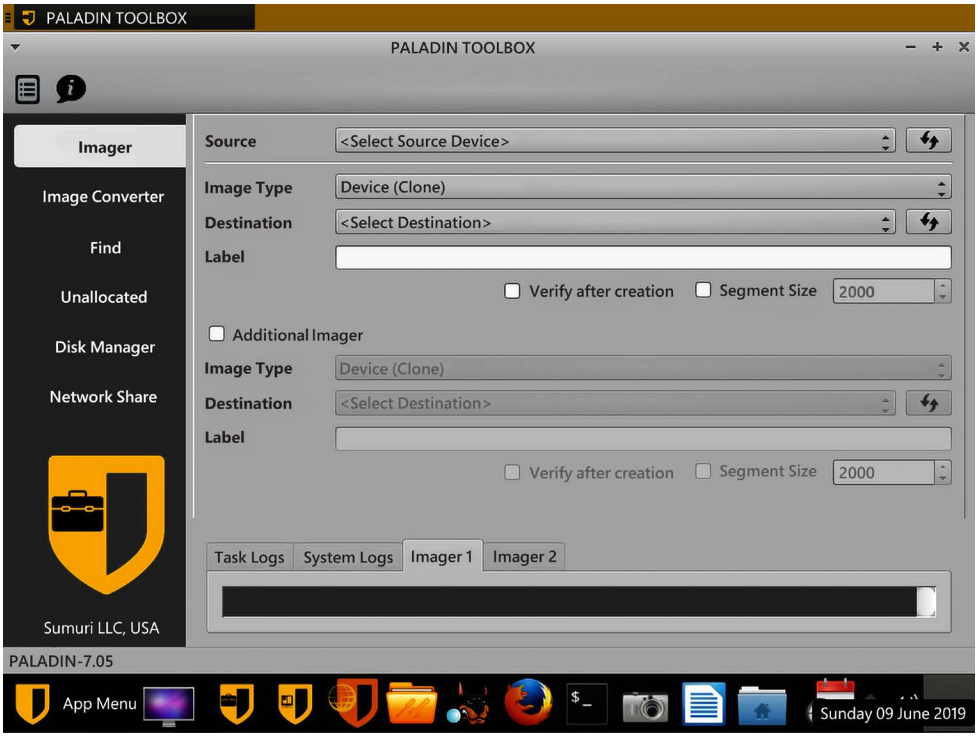


Figure 3.34: PALADIN – Toolbox imaging screen

6. When you select the dropdown for the source device, you will see a list of devices recognized by the system. This is the same list of devices we saw in Device Manager. It is essential to choose the correct device when creating your forensic image. Here, we will select the `sdc` device:

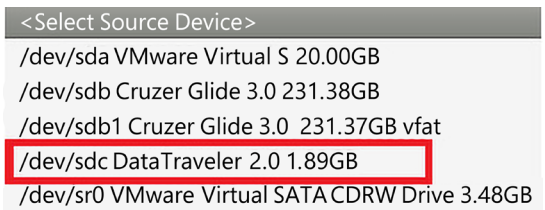


Figure 3.35: PALADIN – Toolbox Select Source Device drop-down menu

7. When you select the image format drop-down menu, you will be presented with more choices. We have discussed `dd`, `e01`, and `SMART` forensic images, so let's consider the remaining options shown in the following screenshot:

Figure 3.36: PALADIN – Toolbox Image Format drop-down menu

Let's discuss them in detail:

DMG: This is a proprietary Apple disk image file. It is a RAW forensic image.

VMDK: **VMware Virtual Disk Format**. This is a virtualization disk image.

VHD: **Virtual Hard Disk**. This is a virtual hard disk format typically used by Microsoft Virtual PC, Virtual Server, and Hyper V Server.

8. Your next option is to select the destination. With PALADIN, you must ensure the destination device is mounted as read/write. I have ensured that `sdb1` has been mounted as read/write and has sufficient capacity to store the forensic image:

Figure 3.37: PALADIN – Toolbox Destination drop-down menu

9. All that remains is to add a label, that is, a filename. I recommend using the same naming convention to identify the different pieces of evidence. Since it is a USB device and it is the first device I have imaged, I will label it `usb001`:

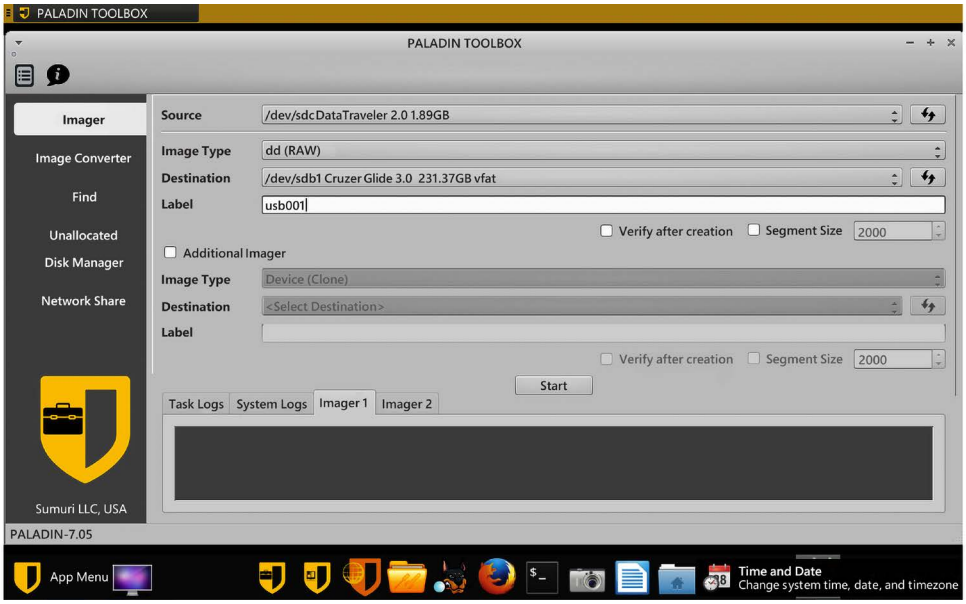


Figure 3.38: PALADIN – Toolbox imager

You also have the option to **Verify after creation** and whether you want to create a forensic image with segments.

You also have the option to create a second forensic image at the same time.

Once the forensic image creation process has been completed, PALADIN will present you with a log of the process. As shown in the following screenshot, PALADIN is using `dc3dd` to create the forensic image:

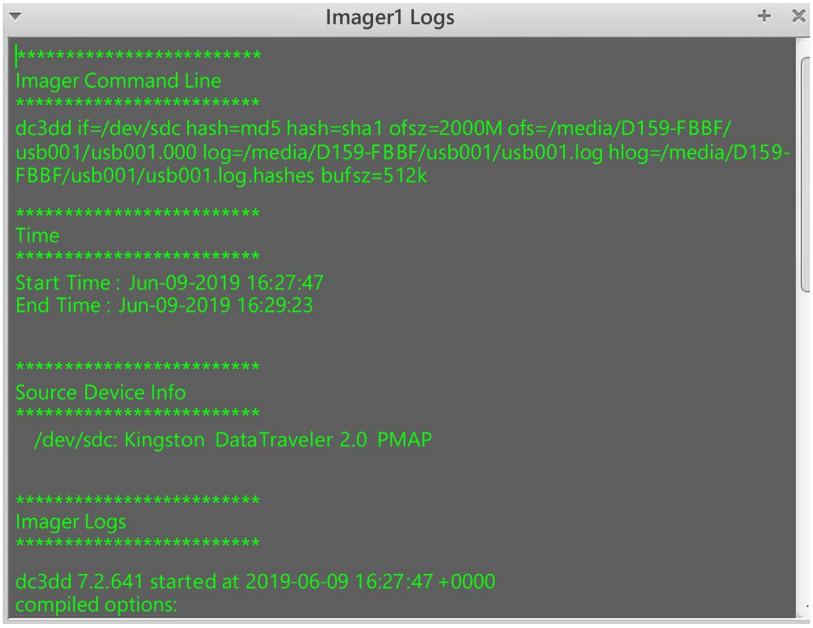


Figure 3.39: PALADIN – Completed imaging screen

With that, you have just created a forensic image with PALADIN.

Summary

In this chapter, we have discussed evidence and how you need to validate your processes and your forensic tools to ensure accurate results. You learned about the forensically sound examination environment and how you must maintain control of the environment. The environment is not just in the lab, but encompasses when you start the forensic analysis process. We have gone over how to validate your forensic tools, create sterile media, and explored the different write blocking options that are available. Next, we have gone through the process of creating a forensic image utilizing forensic tools such as FTK Imager and PALADIN and gone into detail about the different formats available to create a forensic image. Now, we can move on and explore how the computer operates and explore different filesystems.

In the next chapter, we will go into the workings of the computer system and the storage devices you may encounter.

Questions

1. Digital evidence is _____.

    a. Volatile
    b. Non-volatile
    c. Good to have
    d. Not needed when you have a confession

2. Why would it be a good idea to wipe a drive before reusing it to store evidence?

    a. Chain of custody
    b. To make sure it is formatted correctly
    c. To ensure no prior data exists on the device
    d. It's the examiner's choice (the examiner can decide the course of action)

3. You must use a write blocker on the source device when creating a forensic image.

    a. True
    b. False

4. Who controls the forensically sound examination environment?

    a. Suspect
    b. First responder
    c. Examiner
    d. Depends on the situation

5. The examiner must validate all tools before use.

    a. True
    b. False

6. When creating a forensic image, which is the best option?

    a. Forensic copy
    b. Forensic image
    c. Logical forensic image
    d. Backup copy

7. A `dd` image can be compressed.

    a. True

    b. False

The answers can be found at the back of this book, under *Assessments*.

Further reading

Zatyko, K., 2011. *Commentary: Defining Digital Forensics*. Retrieved from `http://www.forensicmag.com/`.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

4

Computer Systems

As we discussed in previous chapters, digital forensic investigators must control the environment they operate in. The diversity of computer hardware, operating systems, and filesystems requires the digital forensic investigator to have a firm understanding of all the different and potential configurations they may encounter. This requires the digital forensic investigator to have procedures or controls to protect the integrity of digital evidence and the processes used to examine it. If you do not understand the boot process and how the system reacts when it starts or which filesystem is used on storage devices, you could make a fatal mistake. In addition, you must understand how they work together. Failure to understand these essential components could lead you to alter the digital evidence. You will also find that you will be less effective when you testify in judicial or administrative proceedings.

In this chapter, we will cover the following topics:

- Understanding the boot process
- Understanding filesystems
- Understanding the NTFS filesystem

Understanding the boot process

To control the environment as we start our investigation, we must understand the environment. Here, digital evidence is being stored, created, and accessed. In most cases, this will be a computer system. I use the term "computer system," which comprises the operating system, the filesystem, and the hardware bundled together to create a computer. To be effective, you must understand the physical media the data is stored on, the filesystem used on the storage device, and how that data is tracked and accessed while on the storage device.

Once you understand the process, you can then implement controls to protect the integrity of the digital evidence.

So, what is the boot process? When you push the power button and electricity energizes the system, commands are issued. As it executes the commands, the system is taking steps (like on a ladder) to achieve the goal of a running operating system. If something breaks any of those steps, the system will not load.

The first step is the **Power-On Self-Test (POST)**; the CPU will access the **Read-Only Memory (ROM)** and the **Basic Input/Output System (BIOS)** and test essential motherboard functions. This is where you hear the beep sound when you turn the power on to the computer system. If there is an error, the system will notify you of the error through beep codes. If you do not have the motherboard manual, a Google search will help determine the meaning of the specific beep code.

Once the **POST** test has been successfully completed, the BIOS is activated and executed. Note that the system has not accessed the storage media. This is because all the program executions occur at the motherboard level and not in the storage devices. The user can access the BIOS by using the correct key combination displayed on the screen.

Note

The time allowed for you to hit the correct key can sometimes be relatively short. If you are unsuccessful, the system will continue booting and accessing the storage device. If you are trying to access the suspect's computer system, disengage the storage devices if they are accessible before starting the process. This will ensure that you are not booting to the suspect's storage device and destroying evidence.

The BIOS will have the basic information of the system: the amount of RAM, the type of CPU, information about the attached drives, and the system date and time. The easiest way to document this information is to photograph it as it is displayed on the screen. This is also where you can change the boot sequence. Typically, the system checks the CD/DVD first and then the designated hard drive. This is where you would be able to change the setting of the boot device when we create the boot media later in the chapter. Changing the boot device tells the BIOS to access the device we provide and not the suspect's device.

In 2010, the BIOS function was replaced by the **Unified Extensible Firmware Interface (UEFI)**. It provides the same service as the BIOS, but has been enhanced, as follows:

- By providing better security at the preboot process
- Faster startup
- Will support drives larger than 2 TB
- Support for 64-bit device drivers
- Support for the **GUID Partition Table (GPT)**

The Secure Boot feature allows us to use authenticated operating systems when booting the computer system. This can be an issue if you attempt to use an alternative booting device.

As you can see in the following diagram, once the power is turned on and it has completed the POST test, depending on the system, it may boot with the BIOS, or it may boot with the UEFI scheme:

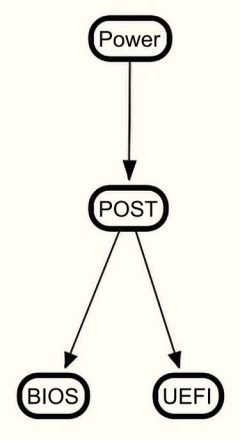


Figure 4.1: Boot process

The BIOS will look for the **Master Boot Record (MBR)** of the boot device. The MBR is located at sector zero and holds information about the partitions, the filesystems, and the boot loader code for the installed operating system. Once the MBR is found in the boot loader and activated, control is then passed over to the operating system to complete the booting process.

The UEFI will look for the GPT; the GPT will have a protective MBR to ensure legacy systems will not mistakenly read this as being unpartitioned and overwrite the data. It will also contain the partition entries and backup partition table header. A GPT disk can contain up to 128 partitions for a Windows operating system. Like in the BIOS scheme, once the active partition and boot loader have been found, the operating system will take over the booting process.

Since you now understand the boot process, we still want to control the boot environment by creating forensic boot media, which we will discuss next.

Forensic boot media

It is a widespread practice to remove the hard drive from the system to create a forensic image. However, sometimes, the investigator cannot remove the storage device from the system, and they need to create a forensic image of the storage device. To accomplish this task, you need to use a bootable CD/DVD or USB device to create a forensic environment to create a forensic image.

Using boot media, you will want to ensure that it will create that sound forensic environment and not cause any changes to the source device. As we discussed during the boot process, we want to intercept any potential changes to that source device, and we want to have the system boot inside an environment we control. While it is still possible to boot using a CD/DVD, finding systems without an optical drive is becoming more common. Without an optical drive, we must use a boot USB device to create a sound forensic environment to access the storage device.

Linux is a standard operating system that is used to create a USB-based (live) operating system to create the forensic environment needed to examine these devices. As discussed in *Chapter 3*, *Acquisition of Evidence*, PALADIN is one such tool. It is freely available to download and purchase if you wish to have it preinstalled on a USB device. Sumuri also provides some limited technical support in the operation of PALADIN.

There is also a Windows-based bootable environment known as **WinFE** (**Windows Forensic Environment**). WinFE was developed by Troy Larson in 2008 and has spawned other tools such as Mini-WinFE, which was developed by Brett Shavers and Misty (`http://reboot.pro/files/file/375-mini-winfe/`). The benefit of using the Windows bootable environment is that you now have Windows-based forensic tools. It is possible to run X-Ways or FTK Imager from this secure environment. I would not recommend using a tool that is resource-heavy. What I mean is that some forensic suites such as EnCase Forensic or FTK require significant resources to run effectively. X-Ways can be run from a USB device, as can some artifact-specific tools like RegRipper.

As with any tool or procedure, you must validate it to ensure you are getting the expected results. This means that before you go out into the field and boot a suspect's computer utilizing a forensic USB device, you must test it in the laboratory environment to ensure no changes are made.

Some of the challenges that you, as the examiner, need to be concerned with when using a bootable USB device include the following:

- Ensuring the system will boot to the device and not the internal hard drive by changing the boot order in the BIOS
- In some systems, it's difficult to access the BIOS in the time provided during the boot process
- Ensuring the system can boot to a USB device – some older systems cannot
- Knowing which filesystems the bootable device can write-protect and which ones it cannot
- Dealing with the secure boot feature of the UEFI boot process

As mentioned earlier, secure boot is a security feature of the UEFI process that allows trusted operating systems to boot the system. Therefore, if we want to use a bootable forensic operating system, the secure boot feature must be disabled.

You must enter the UEFI environment by pressing the catch key, such as *F2* or *F12* (this will vary depending on the computer manufacturer). Once you have entered the setup utility, navigate to the **Security** menu (this might vary depending on the computer manufacturer) and disable the secure boot option. Some Linux distributions and WinFE have received signed status and will boot a secure boot-enabled system.

You must document your steps as you go through this process. For example, if you miss hitting the catch key and start the boot process in the host operating system, you must document that it occurred. Even beginning a partial boot will change the timestamps and make entries in various logs in the operating system.

Now that you understand what a bootable forensic device is let's go ahead and create one in the next section.

Creating a bootable forensic device

To create a bootable forensic device, you will need a USB (I recommend using an 8 GB, or larger, device) and an ISO file for the operating system you wish to install. I will demonstrate using an ISO for PALADIN and free software called Rufus (`https:/rufus.ie/`). Rufus is a utility used to create bootable USB devices.

Once you download Rufus, execute the executable and the program will run:

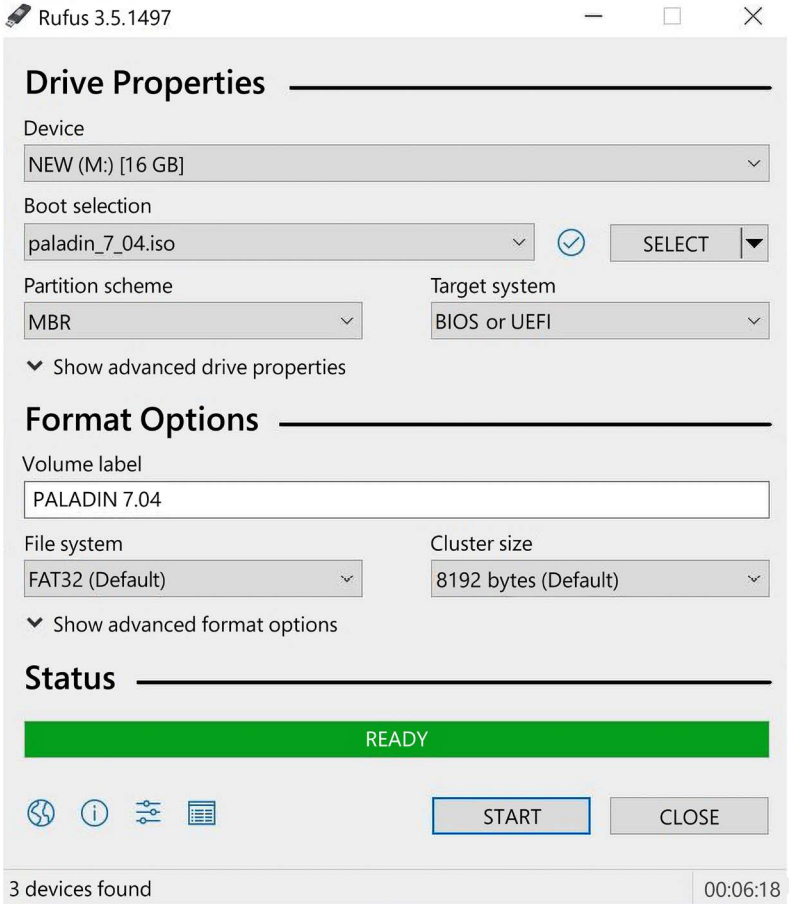


Figure 4.2: Rufus

Something similar to the preceding screenshot (Rufus) will appear, and you will have to select the appropriate choice from the drop-down menus:

- **Device**: This is the destination. It is the USB device you want to host the bootable operating system on.

- **Boot selection**: This will be the "live" operating system. Here, I am using an ISO file for PALADIN 7.04.
- **Partition scheme**: You have a choice of using MBR or GPT. Using MBR will give you greater flexibility in the devices you can boot.
- **Target system**: With the MBR selection for the partition scheme, you can use the device on either a BIOS or UEFI system. If you select GPT for the partition scheme, you can only target UEFI systems.

Under **Format Options**, accept the default values and then click on the **START** button. Once the program completes, you will have a fully functioning, bootable forensic environment.

We have created a forensic boot environment; let's discuss the storage media you will encounter. We will now discuss hard drives.

Hard drives

The term "physical drive storage device" refers to the hard disk drive itself. That is a physical device that contains platters or solid-state storage that holds data. The term "logical device/volume/partition" refers to the formatting of the physical device. A physical device can contain one or more logical devices/volumes/partitions. It is a common misconception that the term "C drive" refers to the physical device when, in actuality, it refers to a logical partition on the physical device.

Several components make up the interior of the hard drive (as shown in the following figure). If you were to open the case, you would find the hard drive comprised of one or more platters. One or more platters could be stacked together with a spindle in the center. The platters, made of a metal alloy or glass, are coated with a magnetic substance in which the heads magnetically encode information on the platters. The heads can write data on both sides of the platter. The spindles of the hard disk cause the disks to rotate at thousands of revolutions per minute; the faster the spindle causes the platters to spin, the higher the efficiency of accessing the data encoded on the platters. To read or write data to the platters, the heads are positioned less than **.1** microns from the platter's surface. Additionally, the actuator controls the heads; it swings across the platter, placing the head in the correct position to read/write the data.

The storage devices are manufactured with tight tolerances and can be damaged by sudden sharp movement or a mechanical shock:

Figure 4.3: Hard drive

A hard drive can have different interfaces, for example, you may run into some of the following:

- **Small Computer System Interface (SCSI)**: An older standard that is typically seen in the corporate environment. Limited to 16 chained devices and will have a terminator at the end of the chain.
- **Integrated Drive Electronics (IDE/EIDE)**: An old standard but may still be found in older consumer computer systems.
- **Serial Advanced Technology Attachment (SATA)**: A current standard found in many consumer and commercial environments.
- **Serial Attached SCSI (SAS)**: A current standard that is typically found in commercial environments.

Solid state drives (SSDs) are storage devices that contain no moving parts. Instead, they are made up of memory chips. As we discussed earlier, a traditional hard drive has several moving parts in which to read/write data to the spinning platters. With an SSD storage device, all the data is stored in memory chips, allowing for the following:

- Less weight
- Increased reliability
- Improved data access speed
- Reduced power consumption

For an SSD to function reliably, there are several operations controlled by the firmware of the device. These functions are as follows:

- **Wear leveling**: This spreads the writes across the different chips so that it uses the chips at the same rate.
- **Trim**: This will wipe the unallocated space of the device.
- **Garbage collection**: As the firmware scans the memory modules, it may identify pages within the data blocks that have been deleted. The firmware will move the allocated pages to a new block and will wipe the data block so that it can reuse the blocks. The firmware can only delete data in blocks.

The real-world effect on forensics is that we can no longer recover data that is or was in unallocated space. Since these operations are conducted at the firmware layer, these operations start automatically as soon as power is given to the device.

Drive geometry

The drive geometry of a platter drive details how data is stored on the device; the drive geometry defines the number of heads, the number of tracks, the cylinders, and the sectors per track. The manufacturer performs what it refers to as a low-level format, which creates the basic structure of the disk by defining the sectors and tracks. A track is a circular path on the platter's surface, as indicated in the following diagram. The red circle (**A**) is a single track, and each side of the platter will have its own set of tracks. They then subdivide the track into sectors. A sector (**B**) is the smallest storage unit on the device.

Initially, a sector used to be 512 bytes in size; however, newer disks are being formatted with a sector size of 4,096 bytes:

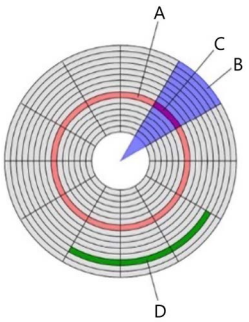


Figure 4.4: Drive diagram

The platters have an addressing scheme to locate the data; originally, **Cylinder, Head, Sector (CHS)** was used. **Cylinder** refers to the vertical axis of the same sectors on all the platters. **Head** refers to the read/write heads; each platter has two heads. Finally, **sector** refers to the number of sectors per track. This addressing scheme worked for large-capacity hard drives; however, as the storage capacity increased, the CHS scheme could not scale because of file size limitations, so **Logical Block Addressing (LBA)** was created. With the LBA scheme, you can address the sectors with a sector number starting from zero.

So, we have discussed the physical components of the device. We will now dive deeper and examine some of the internal aspects.

MBR (Master Boot Record) partitions

Three steps are required before the computer system can use the storage device. First, we have discussed the low-level format conducted by the manufacturer, but now we will discuss partitioning.

Partitioning occurs when we divide the physical device into logical segments called “volumes.” With the MBR partitioning scheme, we are restricted to four primary partitions. For example, with one physical device, you can have a primary partition used to host the Windows operating system. You can also have a second primary partition that hosts a Linux operating system. Note that you must have a primary partition to boot into an operating system. When a user selects the booted operating system, this is known as the **active partition**.

To get around the partition limit, developers created the extended partition. One of the four partition records is designated as an extended partition, which can then be divided into logical volumes.

As we discussed previously, we can find the MBR at sector zero. The MBR contains the information needed by the system to boot. The MBR will be in sector zero, so it will be no longer than 512 bytes. The partition table will show us which partition is the active partition. Once the starting sector of the active partition is located, the boot process will continue:

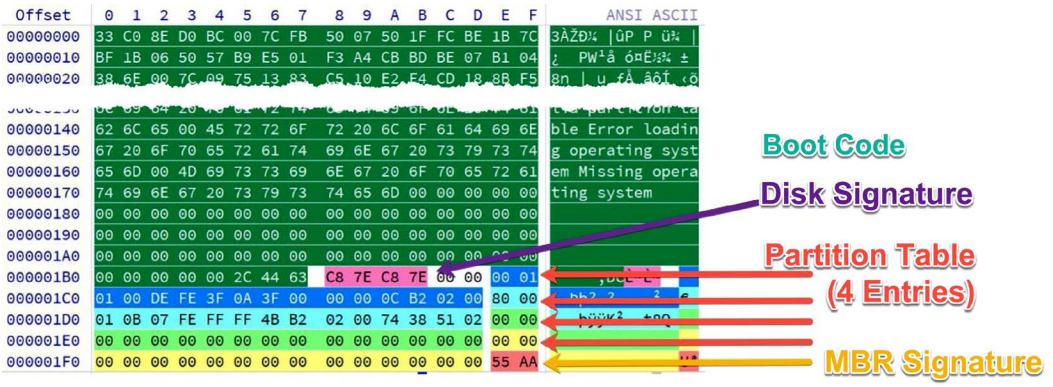


Figure 4.5: MBR map

The preceding MBR map depicts sector zero of a hard disk. This is the MBR for the physical disk. The first 440 bytes are highlighted; this is the boot code. The next 4 bytes are the disk signature and identify the disk to the operating system. The following 64 bytes comprise the partition table. Each 16-byte entry refers to a specific partition. Remember, it restricts us to 4 primary partitions utilizing the MBR partitioning scheme. The final 2 bytes is the signature for the MBR. It identifies the ending of the MBR and will be the last 2 bytes of the sector.

In the following table, I have extracted the four partition tables and reformatted the hex values for easier reading. The first byte will designate which partition is the active partition. A value of `x/80` identifies the active bootable partition.

A value of `x/00` shows the non-active (bootable) partition:

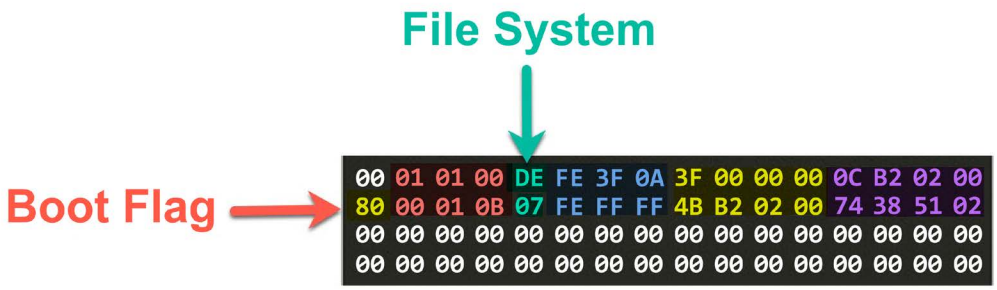


Figure 4.6: Partition tables

Typically, you would see the first partition marked as the active partition; in this case, it is the second partition, which is bootable. The next 3 bytes represent a starting sector for the CHS calculation. So, when we examine the partition table, we can see that the physical device has partition 0 and partition 1. The entries for partitions 2 and 3 are zeroed out. This tells us that there are only two partitions on this physical device.

The fifth byte represents the filesystem on the partition. For partition 0, we can see the hex value of **DE,** which tells us that it is part of the Dell PowerEdge Server utilities. Partition 1 has a hex value of **07**, which shows the NTFS filesystem.

If I found the hexadecimal values of **05** or **0f**, that would show an extended partition. We would then have to look into the extended boot records of the extended partitions.

Note

You can find a full list of partition identifiers at `https://www.win.tue.nl/~aeb/partitions/partition_types-1.html`.

The next 3 bytes are the values for the ending sector of the CHS calculation. The next 4 bytes show the starting sector of the partition, and the last 4 bytes show the size of the partition.

The sector values used in the CHS calculation are legacy values for older storage devices. The values showing the start sector and the total number of sectors (partition size) are being used for the current drives using LBA.

Each partition will have a **Volume Boot Record** (**VBR**) at sector zero of the partition. The system uses the VBR to boot the operating system in that volume. It is an operating system-specific artifact and is created when the partition is formatted.

It will also appear on unpartitioned devices, such as removable media, for example, a USB or floppy disk.

Primary partitions are not the only partitions that you may encounter; you can also encounter an extended partition, which is the subject of the next section.

Extended partitions

The limitation of the MBR of only allowing four primary partitions resulted in the creation of the extended primary partition. Here, it takes the place of one (and only one) primary partition and enables the user to create additional logical partitions over the four primary partitions.

The following partition map illustrates the replacement of a primary partition with an extended partition:

Figure 4.7: Partition map

The following diagram shows the extended partition. Here, the user has created multiple logical partitions within the extended partition boundary:

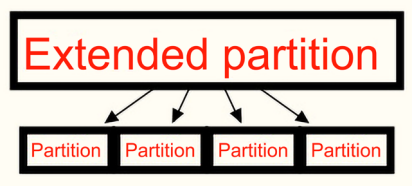


Figure 4.8: Extended partition map

The extended partition will not have a VBR. It will have an **extended boot record** (**EBR**), which will point to the first extended logical partition. The first extended logical partition will contain information about itself and a pointer to the next extended logical partition. In effect, this will create a daisy chain of pointers from one extended logical partition to the next.

We have now covered the aspects relating to the MBR; let's now go over the GPT-formatted aspects.

GPT partitions

A GUID is a **globally unique identifier** and uses a 128-bit hexadecimal value to identify different aspects of the computer system. A GUID comprises five groups and is formatted as `00112233-4455-6677-8899-aabbccddeeff`, and, while there is no central authority to ensure uniqueness, it is doubtful that you would get a repeating GUID.

RFC 4122 defines the five different GUIDs as follows:

- **Version 1**: Date-time and MAC address: The system generates this version using both the current time and client MAC address. This means that if you have a version 1 GUID, you can figure out when it was created by inspecting the timestamp value.
- **Version 2**: DCE security: This version isn't explicitly defined in RFC 4122, so it doesn't have to be generated by compliant generators. It is like a version 1 GUID except that the first four bytes of the timestamp are replaced by the user's POSIX UID or GID, and the upper byte of the clock sequence is replaced by either the POSIX UID or GID domain. (**UID** stands for **User Identifier**. **POSIX** stands for **Portable Operating System Interface**, which is a set of standards to ensure compatibility between operating systems.)
- **Version 3**: MD5 hash and namespace: This GUID is generated by taking a namespace (for example, a fully qualified domain name) and a name, converting it into bytes, concatenating it, and hashing it. Once it has specified the special bits such as version and variant, it then converts the resulting bytes into hexadecimal form. The special property regarding this version is that the GUIDs generated from the same name in the same namespace will be identical even if they were generated at different times.
- **Version 4**: Random: The system creates this GUID using random numbers. Of the 128 bits in a GUID, it reserves 6 for special use (version + variant bits) giving us 122 bits that can be filled at random.
- **Version 5**: SHA-1 hash and namespace: This version is identical to version 3 except that SHA-1 is used in the hashing step in place of MD5.

The GPT is a partitioning scheme that is used for newer storage devices and is part of the new UEFI standard. The UEFI standard replaces the BIOS, while the GPT replaces the MBR partitioning scheme.

The GPT petitioning scheme uses LBA, and a protective MBR is found in the physical sector zero. The protective MBR allows for some backward compatibility and helps to remove any issues when dealing with legacy utilities that do not recognize the GPT partitioning scheme. There is no boot code available in the protective MBR. As you can see in the following diagram, this is the first partition entry of the partition table of the protective MBR. The partition is identified by hex value **EE**, which shows it is a GPT partition disk, as shown in the following GPT hex:

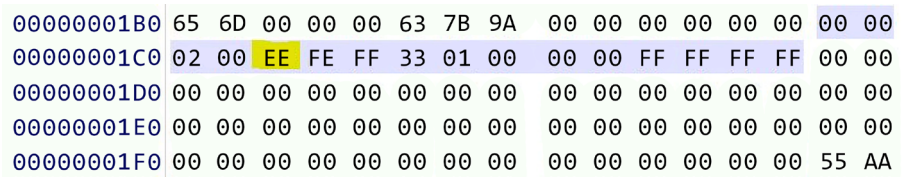


Figure 4.9: GPT hex

While the MBR contains the partition table within physical sector `0`, GPT houses the partition table header at physical sector `1`. The GPT header can be identified by the **EFI** signature of hexadecimal values `45 46 49 20 50 41 52 54`, as shown in the following diagram:

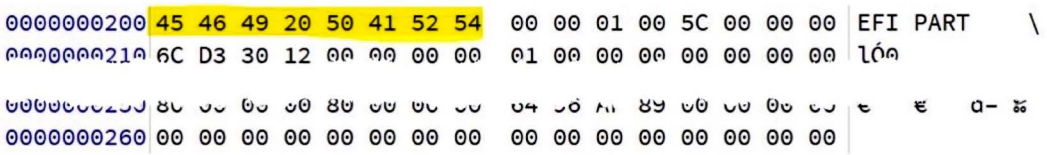


Figure 4.10: EFI PART

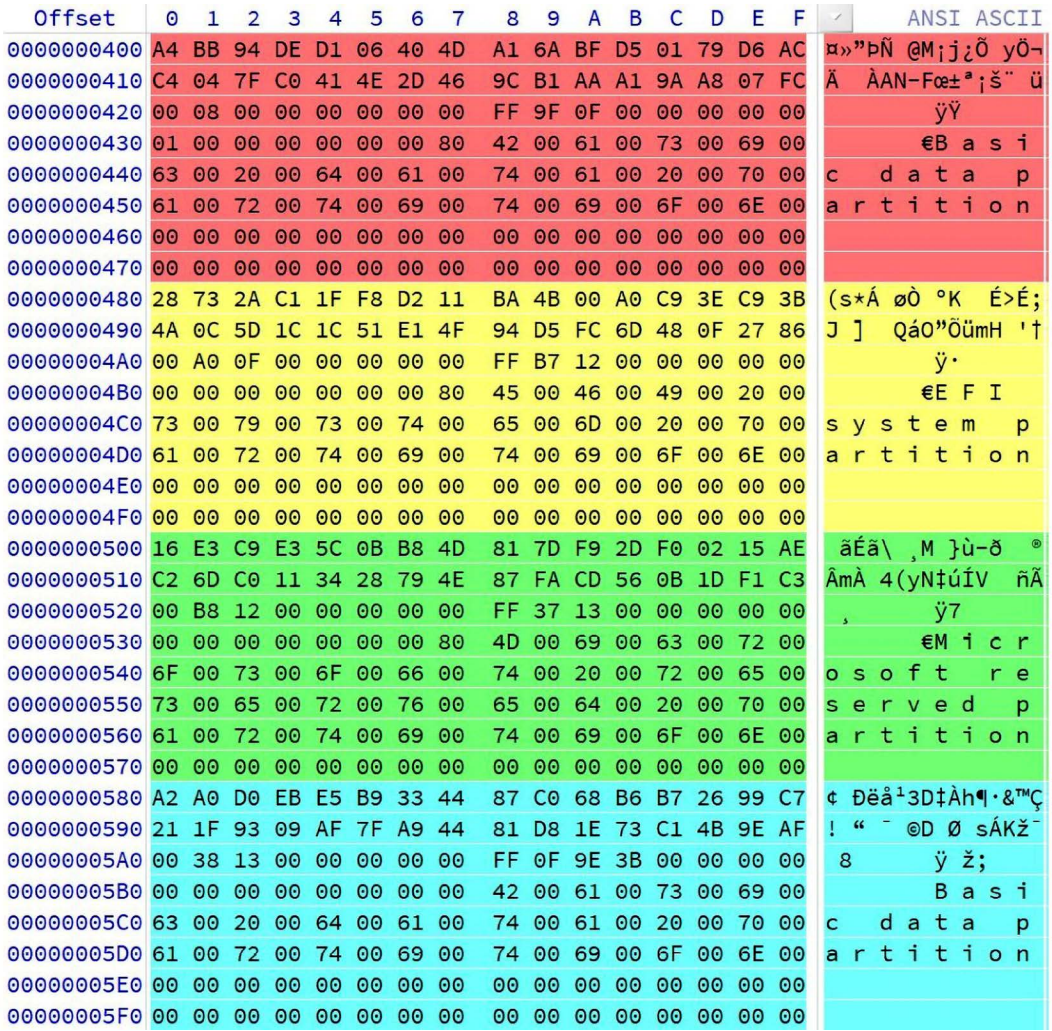
The following table shows the layout of the GPT header, which you can use to identify the layout of the desk:

GPT header format

Offset	Length	Contents
0 (0x00)	8 bytes	Signature ("EFI PART", 45h 46h 49h 20h 50h 41h 52h 54h)
8 (0x08)	4 bytes	Revision (for GPT version 1.0 (through at least UEFI version 2.7 (May 2017)), the value is 00h 00h 01h 00h)
12 (0x0C)	4 bytes	Header size
16 (0x10)	4 bytes	CRC32 checksum of the GPT header
20 (0x14)	4 bytes	Reserved; must be zero
24 (0x18)	8 bytes	Current LBA (location of this header copy)
32 (0x20)	8 bytes	Backup LBA (location of the other header copy)
40 (0x28)	8 bytes	First usable LBA for partitions (primary partition table last LBA + 1)
48 (0x30)	8 bytes	Last usable LBA (secondary partition table first LBA − 1)
56 (0x38)	16 bytes	Disk GUID in mixed endian
72 (0x48)	8 bytes	Starting LBA of array of partition entries (always 2 in primary copy)
80 (0x50)	4 bytes	Number of partition entries in array
84 (0x54)	4 bytes	Size of a single partition entry (usually 80h or 128)
88 (0x58)	4 bytes	CRC32 checksum of the of the partition table
92 (0x5C)	*	Reserved; must be zeroes for the rest of the block (420 bytes for a sector size of 512 bytes; but can be more with larger sector sizes

Figure 4.11: GPT header format

The GPT partition entries are typically found in physical sector 2. The following diagram shows the GPT partition table entries:



Offset	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	ANSI ASCII
0000000400	A4	BB	94	DE	D1	06	40	4D	A1	6A	BF	D5	01	79	D6	AC	¤»”ÞÑ @M¡j¿Õ yÖ¬
0000000410	C4	04	7F	C0	41	4E	2D	46	9C	B1	AA	A1	9A	A8	07	FC	Ä ÀAN-Fœ±ª¡š¨ ü
0000000420	00	08	00	00	00	00	00	00	FF	9F	0F	00	00	00	00	00	ÿŸ
0000000430	01	00	00	00	00	00	00	80	42	00	61	00	73	00	69	00	€B a s i
0000000440	63	00	20	00	64	00	61	00	74	00	61	00	20	00	70	00	c d a t a p
0000000450	61	00	72	00	74	00	69	00	74	00	69	00	6F	00	6E	00	a r t i t i o n
0000000460	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0000000470	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0000000480	28	73	2A	C1	1F	F8	D2	11	BA	4B	00	A0	C9	3E	C9	3B	(s*Á øÒ °K É>É;
0000000490	4A	0C	5D	1C	1C	51	E1	4F	94	D5	FC	6D	48	0F	27	86	J] QáO”ÕümH '†
00000004A0	00	A0	0F	00	00	00	00	00	FF	B7	12	00	00	00	00	00	ÿ·
00000004B0	00	00	00	00	00	00	00	80	45	00	46	00	49	00	20	00	€E F I
00000004C0	73	00	79	00	73	00	74	00	65	00	6D	00	20	00	70	00	s y s t e m p
00000004D0	61	00	72	00	74	00	69	00	74	00	69	00	6F	00	6E	00	a r t i t i o n
00000004E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000004F0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0000000500	16	E3	C9	E3	5C	0B	B8	4D	81	7D	F9	2D	F0	02	15	AE	ãÉã\ ¸M }ù-ð ®
0000000510	C2	6D	C0	11	34	28	79	4E	87	FA	CD	56	0B	1D	F1	C3	ÂmÀ 4(yN‡úÍV ñÃ
0000000520	00	B8	12	00	00	00	00	00	FF	37	13	00	00	00	00	00	¸ ÿ7
0000000530	00	00	00	00	00	00	00	80	4D	00	69	00	63	00	72	00	€M i c r
0000000540	6F	00	73	00	6F	00	66	00	74	00	20	00	72	00	65	00	o s o f t r e
0000000550	73	00	65	00	72	00	76	00	65	00	64	00	20	00	70	00	s e r v e d p
0000000560	61	00	72	00	74	00	69	00	74	00	69	00	6F	00	6E	00	a r t i t i o n
0000000570	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
0000000580	A2	A0	D0	EB	E5	B9	33	44	87	C0	68	B6	B7	26	99	C7	¢ Ðëå¹3D‡Àh¶·&™Ç
0000000590	21	1F	93	09	AF	7F	A9	44	81	D8	1E	73	C1	4B	9E	AF	! “ ¯ ©D Ø sÁKž¯
00000005A0	00	38	13	00	00	00	00	00	FF	0F	9E	3B	00	00	00	00	8 ÿ ž;
00000005B0	00	00	00	00	00	00	00	00	42	00	61	00	73	00	69	00	B a s i
00000005C0	63	00	20	00	64	00	61	00	74	00	61	00	20	00	70	00	c d a t a p
00000005D0	61	00	72	00	74	00	69	00	74	00	69	00	6F	00	6E	00	a r t i t i o n
00000005E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	
00000005F0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	

Figure 4.12: GPT sector 2

Each partition entry is 128 bytes and provides information about the partitions. The following table shows the contents of the partition entries, which include the partition type GUID, the GUID that is unique to that specific partition, the starting and ending sectors, and the partition name in Unicode:

GUID partition entry format		
Offset	**Length**	**Contents**
0 (0x00)	16 bytes	Partition type GUID
16 (0x10)	16 bytes	Unique partition GUID
32 (0x20)	8 bytes	Starting LBA
40 (0x28)	8 bytes	Ending LBA
48 (0x30)	8 bytes	Attribute flags
56 (0x38)	72 bytes	Partition name

Figure 4.13: GUID

A partition should hold all the data on the disk within the partition's boundaries; however, there are spaces on the disk outside of the normal partition boundaries where a technical user may hide data. We will discuss those areas next.

Host Protected Area (HPA) and Device Configuration Overlay (DCO)

HPA and DCO are hidden areas on the hard drive created by the manufacturers. The manufacturer uses the HPA to store recovery and diagnostics tools and it cannot be changed or accessed by the user. The DCO allows the manufacturer to use standard parts to build different products. It will enable the creation of a standard set of sectors on a component to achieve uniformity. For example, the manufacturer might use one set of parts to create a 500 GB hard drive, and while using the same components, it can also create a 600 GB hard drive. Once again, the user would usually not have access to this location. However, some utilities are freely available and could be used by a user to access these locations and store data.

The following screenshot shows you how an HPA may appear in X-Ways:

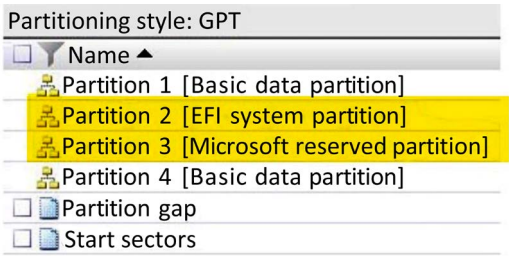


Figure 4.14: HPA 1

The following screenshot shows you how an HPA may appear in FTK Imager:

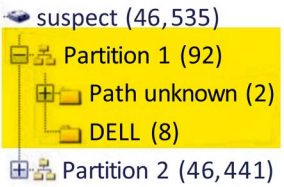


Figure 4.15: HPA 2

Let's move on and discuss some potential filesystems that you may encounter.

Understanding filesystems

A hard drive can have multiple partitions on it, and, in each partition, there will be (in most cases) a filesystem. There might be hundreds of thousands to millions of files within a partition. The filesystem tracks where every file is and how much space is available within the partition boundaries.

We discussed sectors earlier in the *Hard drives* section; they are the smallest units available to store data. The filesystem stores data based on clusters. Clusters are comprised of one or more sectors. A cluster is the smallest allocation unit the filesystem can write to. There are many filesystems available, and some are restricted to specific operating systems unless the user enables drivers that will allow the operating system to read the filesystem.

We will now look at some of the common filesystems you may encounter.

The FAT filesystem

The **File Allocation Table** (**FAT**) filesystem has been around since the early days of home computing, and it is one of the few filesystems that nearly all operating systems can read. It is the de facto standard filesystem for removable devices.

As time has gone by, the FAT filesystem has gone through numerous changes:

- **FAT12**: The first version was created in 1977 and used 12 bits (hence, the FAT12 designation) to address available clusters. This limited its use to only storage devices that could contain 4,096 clusters. It is rarely seen nowadays, but you might find it on a floppy diskette.
- **FAT16**: This was created in 1984 and used 16 bits (I see a pattern) to address the available clusters. It had the same issues as the FAT12, as it could not be scaled to be used with larger-capacity devices.
- **VFAT**: This was introduced with Windows 95 and added the Virtual File Allocation Table. It added the use of the **long filename** (**LFN**) and additional timestamps.
- **FAT32**: This uses 28 bits to address available clusters, theoretically allowing for a maximum volume size of 2.2 TB. Microsoft implemented restrictions that limited the volume size to 32 GB with a maximum file size of 4 GB. It is still in use today and can be found on most removable devices.

We will discuss the FAT32 filesystem for the remainder of this section on the FAT filesystem.

The FAT filesystem is laid out in two areas (as shown in the following diagram, Figure 4.16 – FAT areas):

- **System Area**: This stores the volume boot record and FAT tables
- **Data Area**: This stores the root directory and files:

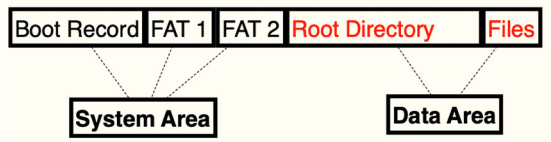


Figure 4.16: FAT areas

Next, we will discuss what falls under **System Area**.

Boot record

We have the **VBR** in the system area. We can find it in logical sector 0 (LS 0), the first sector within the partition boundaries. The boot process creates the VBR when the partition is formatted and contains information about the volume and boot code to continue the boot process for the operating system. If it is a primary partition, the VBR will consist of several sectors, typically sectors 0, 1, and 2, with a backup in sectors 6, 7, and 8. The VBR and backups are stored in a "reserve area," which is typically 32 sectors before the first file allocation table begins:

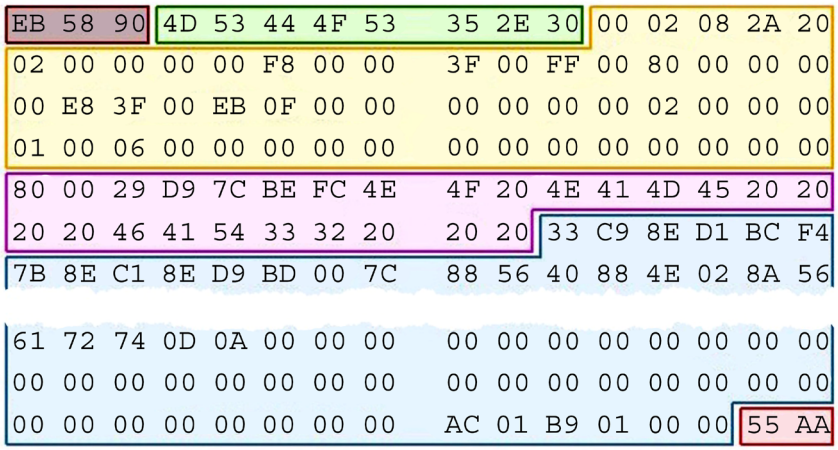


Figure 4.17: VBR

In the preceding diagram, we can see a volume boot sector, which helps to decipher the following information:

- `x00`: We will find the jump instructions for the system to continue booting
- `x03`: The OEM ID shows which operating system was used to format the device
- `x0B`: Bytes per sector
- `x0E`: Number of reserve sectors
- `x10`: Number of FATs (this should be 2)
- `x11`: Unused root entries (for FAT32, this should be 0 because the root directory is in the data area)
- `x13`: Number of sectors (this will be 0 if the number of sectors exceeds 65,536)
- `x15`: Media descriptor (`xF8` will show a hard disk, while `xF0` will show a removable device)
- `x16`: Number of sectors per FAT (for FAT32, this should be 0)

- `x18`: Number of sectors per track (this should be 63 for hard disks)
- `x1A`: Number of heads (this should be 255 for hard disks)
- `x1C`: Number of hidden sectors (the number of hidden sectors before the start of the FAT volume)
- `x20`: Number of total sectors (that is, the total sectors for the volume)
- `x24`: Logical sectors per FAT
- `x28`: Extended flags
- `x2A`: FAT version
- `x2C`: The starting root directory cluster (usually, cluster 2)
- `x30`: Location of the filesystem information sector (typically, this is set to 1)
- `x32`: Location of the backup sector(s) (usually, this is set to 6)
- `x34`: Reserved (set to 0)
- `x40`: Physical drive number (`x80` for hard drives)
- `x41`: Reserved
- `x42`: Extended boot signature (this should be `x29`)
- `x43`: Volume serial number (a 32-bit value is usually generated from the date and time; this can track removable devices)
- `x47`: Volume label (this might not be accurate; different OSes may not use this field)
- `x52`: Filesystem type

Next, we will take a look at the file allocation table.

File allocation table

The next component of the FAT filesystem is the file allocation table, which immediately follows the VBR. There are two file allocation tables (FAT1 and FAT2) by default. FAT2 is a duplicate of FAT1.

The purpose of the file allocation table is to track the clusters and track which files occupy which clusters. Each cluster is represented within the file allocation table starting with cluster 0. The file allocation table uses 4 bytes (32 bits) per cluster entry. The file allocation table will use the following entries to represent the cluster's current status:

- **Unallocated**: `x0000 0000`
- **Allocated**: The next cluster that is used by the file (for example, it represents cluster 7 as `x0700 0000`)

- **Allocated**: The last cluster that is used by the file (`xFFFF FFF8`)
- **Bad cluster**: Not available for use (`xFFFF FFF7`)

A cluster is the smallest allocation unit the filesystem can address. A sector is the smallest allocation unit on the disk. A cluster is made up of one or more sectors. It is very easy to get confused if you are co-mingling those terms. Consider the following cluster example:

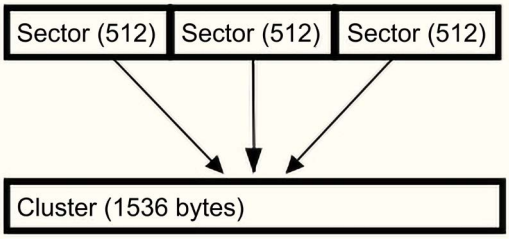


Figure 4.18: Cluster example

As users add files to the data area, the system will update the file allocation table. A file may occupy one or more clusters. Additionally, the clusters may not be sequential, so you could have the data of a file spread in different physical locations on the disk; we typically refer to this as fragmentation.

In the following diagram, we can see a representation of the file allocation table; in this scenario, we have a single file occupying three clusters: **Cluster 4**, **Cluster 5**, and **Cluster 6**. You can see that **Cluster 4** is pointing to **Cluster 5** and **Cluster 5** is pointing to **Cluster 6**. **Cluster 6** has the hexadecimal value for **end of file (EOF)**:

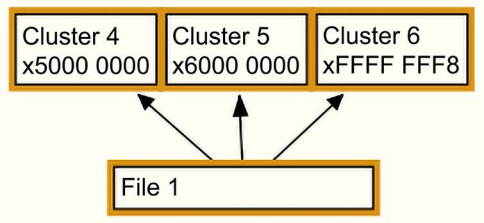


Figure 4.19: Non-fragmented file entry

In the following diagram, we can see a similar representation of the file allocation table with some changes. We now have two files, with File 1 occupying clusters **4** and **6**. We can see that **Cluster 4** is pointing to the next cluster containing the file data, which is **Cluster 6**. This is an example of file fragmentation. File 2 is wholly contained within the cluster boundaries of **Cluster 5**.

Cluster 5 will not point to a subsequent cluster; instead, it has the EOF hexadecimal value:

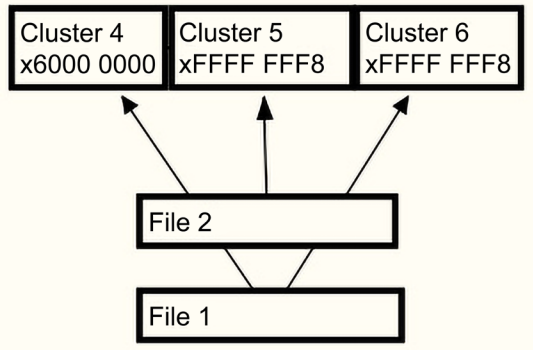


Figure 4.20:Fragmented file entry

We have covered the system area of the FAT; we will now discuss the data area of the FAT filesystem.

Data area

The root directory is housed in the data area because, when the system stored it in the system area, it could not grow enough to work with larger-capacity devices. The critical component of the root directory is the directory entry. If there is a file, directory, or subdirectory, there will be a corresponding directory entry.

Each directory entry is 32 bytes in length and helps track the file's name, starting cluster, and file size in bytes.

In the following diagram, we can see a FAT32 directory with multiple file entries. The filesystem will stop looking for file entries when it runs into a hexadecimal `00`, and all values following the hexadecimal `00` will be ignored:

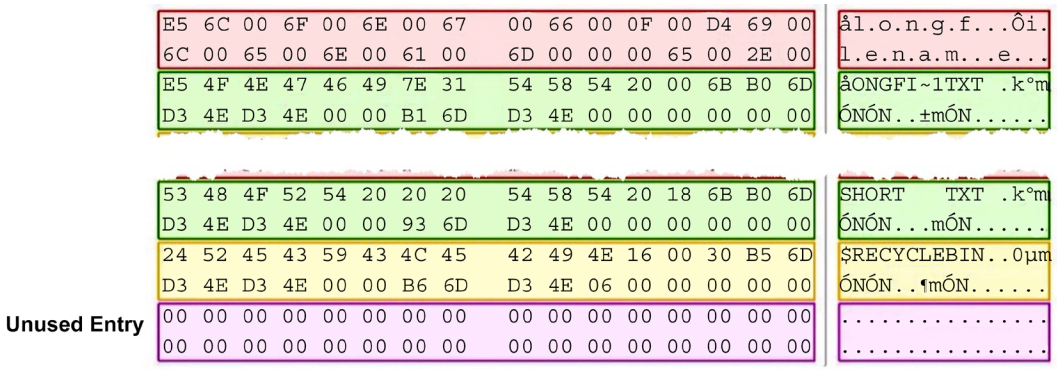


Figure 4.21: FAT directory entry

In the following FAT directory map, we can see the layout of the directory entry and a **short file-name (SFN)** directory entry with the specific offsets highlighted:

Offset (hex)	Size (Bytes)	Description
x00	1	The first character of the file name or status byte
x01	7	Filename (padded with spaces if required)
x08	3	Three characters of the file extension
x0B	1	Attributes
x0C	1	Reserved
x0D	1	Created time and date of the file
x0E	2	File creation time
x10	2	File creation date
x12	2	Last accessed date
x14	2	Two high bytes of FAT32 starting cluster
x16	2	Time of the Last Write to File (last modified or when created)
x18	2	Date of the Last Write to File (last modified or when created)
0x1A	2	Two low bytes of the starting cluster for FAT32
0X1C	4	File size (zero for a directory)

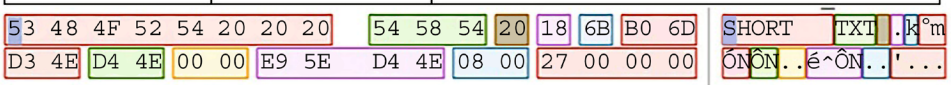


Figure 4.22: FAT directory map

If the first byte is `xE5`, then the filesystem will consider that entry as deleted. The remaining bytes of the file or directory name will remain, as will the other metadata.

The short filename must conform to the specifications as follows:

- Eight characters are allowed; if there are less than eight characters, then the name will be padded with `x20`.
- Three characters are allocated for the file extension (if there are less than three characters, then the name will be padded with `x20`).
- Spaces and the following characters are not permitted: **“+ * , . / : ; < = > ?[\]|.**

The directory entry will always be stored in uppercase. The attribute byte (offset `x0B`) is considered a packed byte, which means the different values have different meanings.

The following diagram shows that bit values in the `Attribute` flag can be combined, and the resulting hex value will reflect the combinations. If a file had the **READ ONLY** flag and the **HIDDEN** flag, then that would give us a value of `0000 0011`, and, when converted to hexadecimal, we get the value of `x03`:

0000 0001	READ ONLY
0000 0010	HIDDEN FILE
0000 0100	SYSTEM FILE
0000 1000	VOLUME LABEL
0000 1111	LONG FILENAME
0001 0000	DIRECTORY
0010 0000	ARCHIVE

Figure 4.23: Packed byte

When we look at the example at the bottom of the preceding FAT directory map, we find the hexadecimal value of 20 at the offset `x0B`; when we convert the hexadecimal into binary, we get `0010 0000`. This tells us that the file is an archive.

We can also encounter an **LFN**; the technique for handling an LFN is a little bit more complicated. We will discuss LFNs in the next section.

Long filenames

When a user creates an LFN, the system will generate an alias that conforms to the SFN standard. It will format the alias so that the first three characters will become the extension after the file extension dot. The first six characters will be converted to uppercase and used for the alias. The alias will then add a ~ character with a number following it. It will start with the number 1 and increase incrementally if additional files have the same alias name.

The following diagram shows a directory entry for a file with an LFN; the filename is `long filename.txt`:

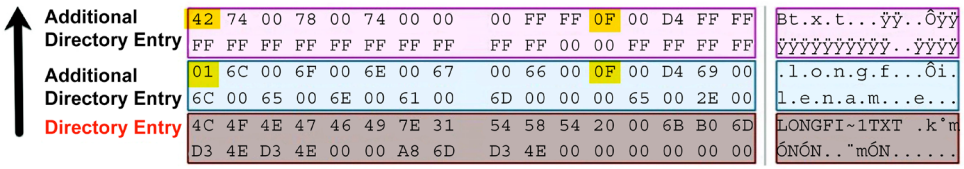


Figure 4.24: LFN

Since this is an LFN, the filesystem will create additional directory entries. In this specific case, there will be two additional directory entries to facilitate the use of the LFN. The first byte of each additional directory entry is the sequence byte. The right nibble is the sequence number. As we look at the directory entry depicted in the preceding diagram, the directory entry above the SFN entry has a hexadecimal value of `x01`. Here, the value of `1` tells us that this is the first value in the sequence. When we move up to the second directory entry, we can see that it has a hexadecimal value of `x42`, the right nibble informs us this is the second directory entry for this LFN file. The left nibble of the value, `4`, tells us this is the last directory entry for the file. In each of the LFN directory entries, you will find that the attribute byte is `x0F`.

But what happens when a file is deleted? Well, you may be able to recover the file and its associated metadata. In the next section, we will discuss recovering deleted files.

Recovering deleted files

When a file is deleted in the FAT filesystem, the data itself does not get changed. Instead, the first character of the directory entry will change to `xE5`, and the file allocation table entries are reset to `x00`. When the filesystem reads the directory entries and encounters `xE5`, it will skip that entry and read from the subsequent entries.

To recover deleted files, we need to reverse the filesystem's process to delete the files. Remember, it has not changed the file contents, and they still physically reside in their assigned clusters. So we now need to reverse engineer the deletion and recreate the file entry and the entries in the file allocation table. To do this, we need to find the first cluster of the file, the file size, and the size of the clusters in the volume.

In the following diagram, we have a directory entry showing that a file has been deleted. We can see `xE5` at the start of the directory entry. (Note that this will require the use of a hex editor to make the changes.)

Then, we must determine the starting cluster, `x00 x08` (but shown as `x08 x00` in the diagram). This value refers to cluster number `8`. Then, to determine the file size, take a look at the last `4` bytes, `x27 x00 x00 x00` (remember that the FAT filesystem stores data in little-endian, which means the least-significant byte is on the left, so we would read that value as `x00 x00 x00 x27`, and when we convert it into a decimal, we have a value of `39` bytes for the file size):

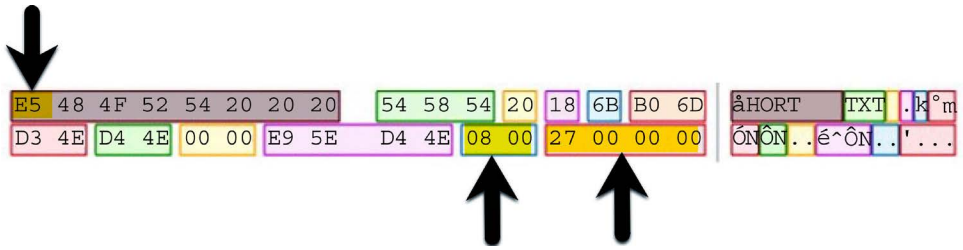


Figure 4.25: Deleted entry

Now we must determine how many sectors make up a cluster and what the sector size is. You will need to go to the boot record to get that information. The boot record shows us that there are 512 bytes per sector, and there are 8 sectors per cluster, which gives us a cluster size of 4,096 bytes (as shown in the following diagram):

Bytes per sector	011	512	15
Sectors per cluster	013	8	114

Figure 4.26: Boot record

This means that our file will only occupy a single cluster. We then go to the file allocation table and look at the entry for cluster 8 and see that it is zeroed out:

```
00 00 00 00 FF FF FF 0F   0B 00 00 00 0C 00 00 00
0D 00 00 00 0E 00 00 00   0F 00 00 00 10 00 00 00
```

Figure 4.27: Deleted FAT

To recover the deleted file, perform the following steps:

1. You need to change the entry in the file allocation table from `x0000 0000` to `xFFFF FFF8` or `xFFFF FF0F`. If this were a larger file, you would need to change the file allocation table entry to point to the next cluster until you reach the last cluster and the end of the file size. As you are rechaining the entries, if you come to an entry marked as allocated when expecting to find the entry unallocated, you may be dealing with a fragmented file. Another alternative is when the clusters were made available to the filesystem. The system placed a new file in the now-available sectors, which would cause the data to be overwritten. There are not many options available if you run into either one of these situations. If the data is overwritten, then you are stuck. If it is fragmented, you must guess where the next cluster will be, which is not very likely with a large-capacity device.

2. The next step is to return to the directory entry and replace `xE5` with another character. When replacing the `xE5` character of the filename in the directory entry, be careful not to guess what the character is. If you select the incorrect character, you could change the meaning or create a bias with the new filename, which would be improper. When recovering a deleted file, it is recommended that you replace that first character with an underscore or a dash, so there is no misunderstanding about the filename.

When recovering a file with an LFN, it is essential to relink the LFN to the SFN. This is because when the additional directories are created to accommodate the LFN, the system creates a checksum based on the data of the SFN. Therefore, when you change the `xE5` value on the SFN entry, you also want to use the same replacement character for the subsequent `xE5` entries for the LFN directory entries. You link the LFN to the SFN because the SFN directory entry contains information such as the date and time, the starting cluster, and the file size.

It is still possible to recover scraps of data that existed on the disk but no longer have any artifacts in the filesystem. This information will be stored in slack space, discussed in the next section.

Slack space

Now is the time to bring up slack space. Remember that the smallest unit the filesystem can write to is a cluster and that clusters are made up of one or more sectors. I keep repeating this because I have seen people who are new to the field get confused about the difference between the two. This is important because files come in a variety of sizes; almost no files will conveniently fit within the cluster boundaries. So, you will have files that spill over into the next cluster. The space between the end of the logical file and the cluster boundary is called "file slack." This slack space can contain data from the previous file. Until it is overwritten, that data will remain for you to examine.

You might find evidence of documents, digital images, chat history, or emails; any data that has been stored on the device, you may find remnants in slack space after the user has deleted the file.

This concludes the *FAT filesystems* section; next up is NTFS.

Understanding the NTFS filesystem

The **New Technology File System (NTFS)** is the default filesystem for Microsoft Windows operating systems. FAT32 had some significant shortcomings, which required a more reliable and efficient filesystem, along with additional administrative improvements to help Microsoft remain viable in a corporate environment. They initially designed NTFS for a server environment; however, as the hard drive capacity has increased, it is now the default filesystem in the commercial and consumer market for the Windows operating system.

NTFS is far more complicated than the FAT filesystem; however, the overall purpose remains the same:

- To record the metadata of a file, that is, the filename, the date timestamps, and the file size
- To mark the clusters the file occupies
- To record which clusters are allocated and which clusters are unallocated

The NTFS filesystem comprises the following system files:

$MFT	**Describes all files on the volume, including file names, timestamps, stream names, and lists of cluster numbers where data streams reside, indexes, security identifiers, and file attributes.**
$MFTMirr	**Duplicate of the first vital entries of $MFT, usually 4 entries (4 kb).**
$LogFile	**Contains transaction log of file system metadata changes.**
$Volume	**Contains information about the volume, namely the volume object identifier, volume label, file system version, and volume flags.**
$AttrDef	**A table of MFT attributes that associates numeric identifiers with names.**
$ (Root file name index)	**The root folder.**
$Bitmap	**Tracks the allocation status of all clusters in the partition.**
$Boot	**Volume boot record.**
$BadClus	**A file that contains all the clusters marked as having bad sectors.**
$Secure	**Access control list database.**
$UpCase	**Converts lowercase characters to matching Unicode uppercase characters.**
$Extend	**A file system directory containing various optional extensions, such as $Quota, $ObjId, $Reparse, or $UsnJrnl.**

Figure 4.28: NTFS table

To identify a partition with NTFS, we need to look at the MBR or the GPT, depending on which formatting scheme was used. In the following diagram, we can see the MBR for the hard drive and the partition table highlighted after the boot code:

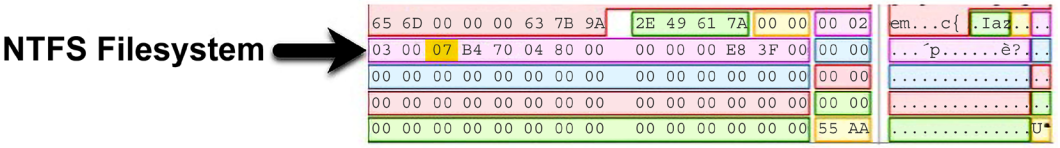


Figure 4.29: NTFS MBR

Looking at the partition table, we can see that there is a single partition, and, at offset decimal 11 from the start of the partition table, we can see the hexadecimal value of `07`. As we discussed earlier in this chapter, this is the filesystem identification for NTFS.

With an NTFS-formatted partition, there is no system or data area like we saw with a FAT-formatted partition. Everything in NTFS is considered a file to include the system data. When we look at the VBR, we can see that it contains information for the system to continue the boot process:

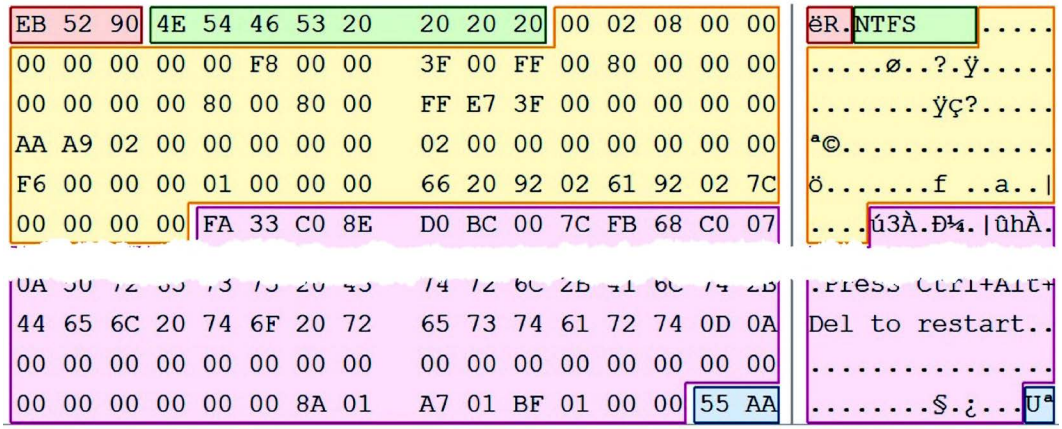


Figure 4.30: NTFS VBR

The information in the VBR is a file; the $Boot record contains all the information that we would expect to find in the VBR. The following $Boot diagram shows the data structure for the $Boot file:

JMP instruction	000	EB 52 90
OEM ID	003	NTFS
BIOS Parameter Block	**00B**	
Bytes per sector	00B	512
Sectors per cluster	00D	8
Reserved sectors	00E	0
(always zero)	010	00 00 00
(unused)	013	00 00
Media descriptor	015	248
(unused)	016	00 00
Sectors per track	018	63
Number of heads	01A	255
Hidden sectors	01C	128
(unused)	020	00 00 00 00
Signature	024	80 00 80 00
Total sectors	028	4,188,159
SMFT cluster number	030	174,506
SMFTMirr cluster number	038	2
Clusters per File Record Segment	040	246
Clusters per Index Block	044	1
Volume serial number	048	66 20 92 02 61 92 02 7C
Checksum	050	0
Bootstrap code	054	FA 33 C0 8E D0 BC 00 7C
Signature (55 AA)	1FE	55 AA

Figure 4.31: $Boot record

Arguably, the most important system file in the NTFS filesystem is the `$MFT` (master file table). The MFT tracks all the files in the volume to include itself. It tracks each file within the MFT through file entries called a file record. Each file record is uniquely numbered and is 1,024 bytes. Each file record starts with a header, with the ASCII text **"FILE"**, and has an EOF marker of hexadecimal `FF FF FF FF`. A new file record is created when files are added to the volume. If a file has been deleted, the record will zero out and make it available for reuse. The MFT will look for an empty file record and use it before creating a new record. The file record can be reused rather quickly, which would overwrite the previous data in the file record.

As shown in the following NTFS file record example, we can see a file record and file header starting with the ASCII values of **FILE**. If the record were corrupted or had an error, you would see the ASCII value of **BAAD**. The file header is 56 bytes:

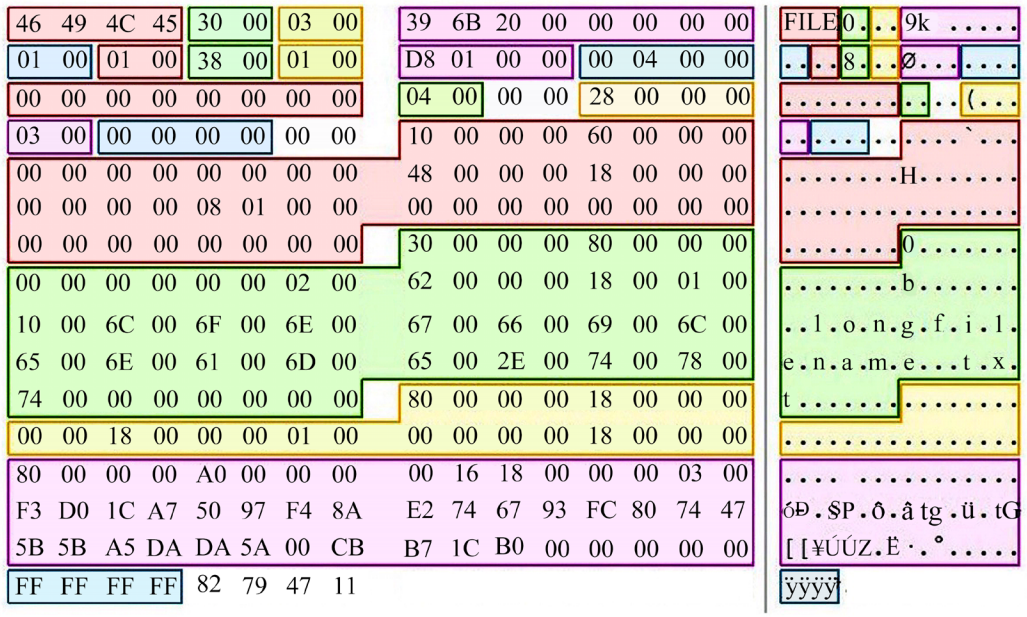


Figure 4.32: NTFS file record

In the following NTFS file record map, we can see the data structure of a file record header:

Signature (must be 'FILE')	000	FILE
Offset to the update sequence	004	0x30
Update sequence size in words	006	3
$LogFile Sequence Number (LSN)	008	2,124,601
Sequence number	010	1
Hard link count	012	1
Offset to the first attribute	014	0x38
Flags	016	01 00
Real size of the FILE record	018	472
Allocated size of the FILE record	01C	1,024
Base FILE record	020	0
Next attribute ID	028	4
ID of this record	02C	40
Update sequence number	030	03 00
Update sequence array	032	00 00 00 00
Attribute $10	**038**	
Attribute $30	**098**	
Attribute $80	**118**	
Attribute $80	**130**	
End marker	1D0	0xFFFFFFFF

Figure 4.33: NTFS file record map

The file record also contains defined data blocks called file attributes. These store specific types of information about the file. The following file attributes table shows several common file attributes that you are likely to see in almost every record:

$Standard Information - 0x10	Includes information such as timestamp and link count.
$Attribute List - 0x20	Lists the location of all attribute records that do not fit in the MFT record.
$File Name - 0x30	A repeatable attribute for both long and short file names. The long name of the file can be up to 255 Unicode characters. The short name is the 8.3 case-insensitive name for the file. Additional names, or hard links, required by POSIX can be included as additional file name attributes.
$Security Descriptor - 0x50	Describes who owns the file and who can access it.
$Data - 0x80	Contains file data. NTFS allows multiple data attributes per file. Each file type has one unnamed data attribute. A file can also have one or more named data attributes.

Figure 4.34: File attributes table

Let's look at each of these attributes in detail.

$Standard_Information Attribute (0x10): The file attributes follow the file header and contain information about the file and, sometimes, the actual file itself. The following diagram depicts a file attribute. The first four bytes show the attribute type; in this case, it is the $10 Standard Information Attribute, which contains general information, flags, accessed, written, and created times, the owner, and security ID. It is identified by the hexadecimal header: `x/10 00 00 00`. The file attribute map contains the decoded values:

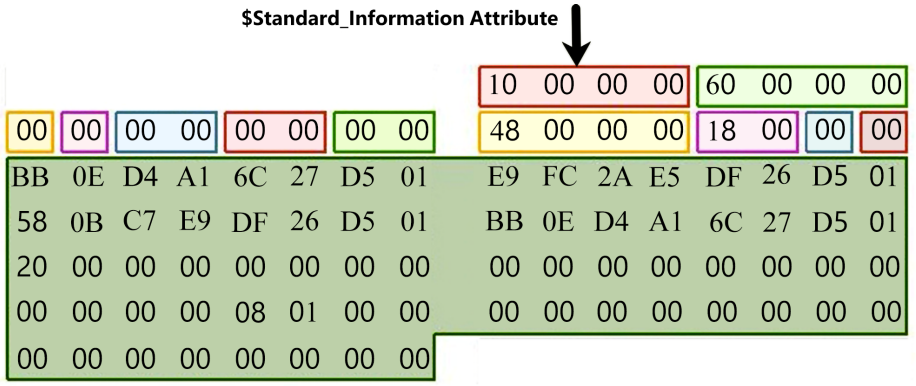


Figure 4.35: $Standard_Information Attribute

Here is a map of the values you will find in the attribute:

Attribute $10	**038**	
Attribute type	038	0x10
Length (including header)	03C	96
Non-resident flag	040	0
Name length	041	0
Name offset	042	0x00
> Flags	044	00 00
Attribute ID	046	0
Length of the attribute	048	72
Offset to the attribute data	04C	0x18
Indexed flag	04E	0
Padding	04F	0
v **$STANDARD_INFORMATION**	**050**	
File created (UTC)	050	6/20/2019 1:32 PM
File modified (UTC)	058	6/19/2019 8:45 PM
Record changed (UTC)	060	6/19/2019 8:45 PM
Last access time (UTC)	068	6/20/2019 1:32 PM
> File Permissions	070	20 00 00 00
Maximum number of versions	074	0
Version number	078	0
Class Id	07C	0
Owner Id	080	0
Security Id	084	264
Quota Charged	088	0
Update Sequence Number	090	0

Figure 4.36: File attribute map

$File_Name Attribute (0x30): The next attribute is the `$30` File_Name Attribute. This attribute stores the name of the file attribute and is always resident. The maximum filename length is 255 Unicode characters. It is identified by the hexadecimal header of `x/ 30 00 00 00`:

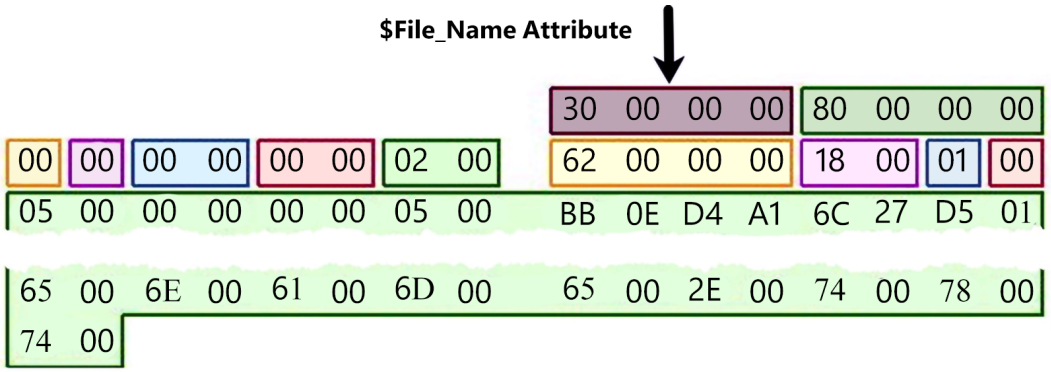


Figure 4.37: $File_Name Attribute

The following is a map of the values you will find in the attribute:

Attribute $30	**098**	
Attribute type	098	0x30
Length (including header)	09C	128
Non-resident flag	0A0	0
Name length	0A1	0
Name offset	0A2	0x00
Flags	0A4	00 00
Attribute ID	0A6	2
Length of the attribute	0A8	98
Offset to the attribute data	0AC	0x18
Indexed flag	0AE	1
Padding	0AF	0
$FILE_NAME	**0B0**	
Parent directory file record number	0B0	5
Parent directory sequence number	0B6	5
File created (UTC)	0B8	6/20/2019 1:32 PM
File modified (UTC)	0C0	6/20/2019 1:32 PM
Record changed (UTC)	0C8	6/20/2019 1:32 PM
Last access time (UTC)	0D0	6/20/2019 1:32 PM
Allocated size	0D8	0
Real size	0E0	0
File attributes	0E8	20 00 00 00
(used by EAs and reparse)	0EC	0
File name length	0F0	16
File name namespace	0F1	0
File name	0F2	longfilename.txt

Figure 4.38: Filename attribute map

$Data Attribute (0x80): The following attribute for this entry is the `$80` Data Attribute. The data attribute contains the file's contents or points to where the contents are in the volume. This attribute is the file data itself.

If the data attribute is resident, we only see the attribute header and the resident content header. The resident content of the attribute is the file's data. Only tiny files have a resident data attribute. We will discuss resident versus non-resident data later on in this chapter.

You may find multiple data attributes per file. In this record, the second $80 Data Attribute, Dropbox, has added some information to the file:

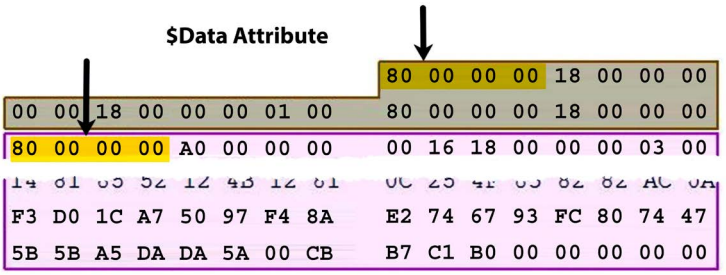


Figure 4.39: $Data Attribute

The following is a map of the values you will find in the attribute:

Attribute $80	**130**	
Attribute type	130	0x80
Length (including header)	134	160
Non-resident flag	138	0
Name length	139	22
Name offset	13A	0x18
Flags	13C	00 00
Attribute ID	13E	3
Length of the attribute	140	83
Offset to the attribute data	144	0x48
Indexed flag	146	0
Padding	147	0
Attribute Name	148	com.dropbox.attributes
$DATA	**178**	
Data	178	78 9C AB 56 4A 29 CA 2F 48
End marker	1D0	0xFFFFFFFF

Figure 4.40: Data attribute map

When examining the `$Data Attribute 0x80`, the filesystem may store the file's contents within the MFT file record itself. Since the file record is 1,024 bytes long, it would have to be a tiny file. When the data content of the file fits within the file record, it is called "resident data":

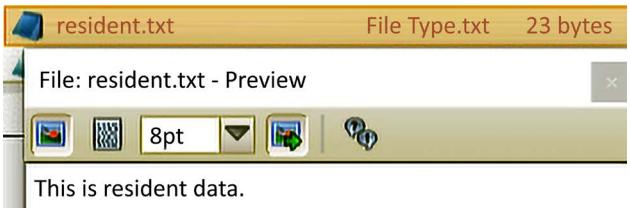


Figure 4.41: Resident data file

In the current example, we have a file named `resident.txt` that is 23 bytes. This is smaller than the 1,024 bytes of the file record. To look at the data of the file, we need to look at the `$Data Attribute 0x80` of the file record, as follows:

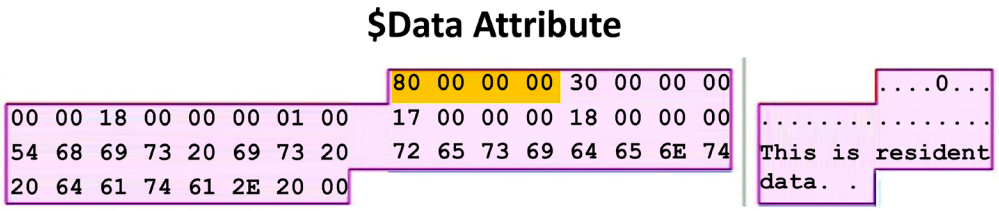


Figure 4.42: Resident data example

On examining the attribute, we can see the ASCII and hex representation of the file content we observed in the preceding resident data example. When dealing with a non-resident file, such as the one depicted in the following diagram, we can see that the `nonresident.txt` file, which is 145 KB in size, is larger than the 1,024-byte file record:

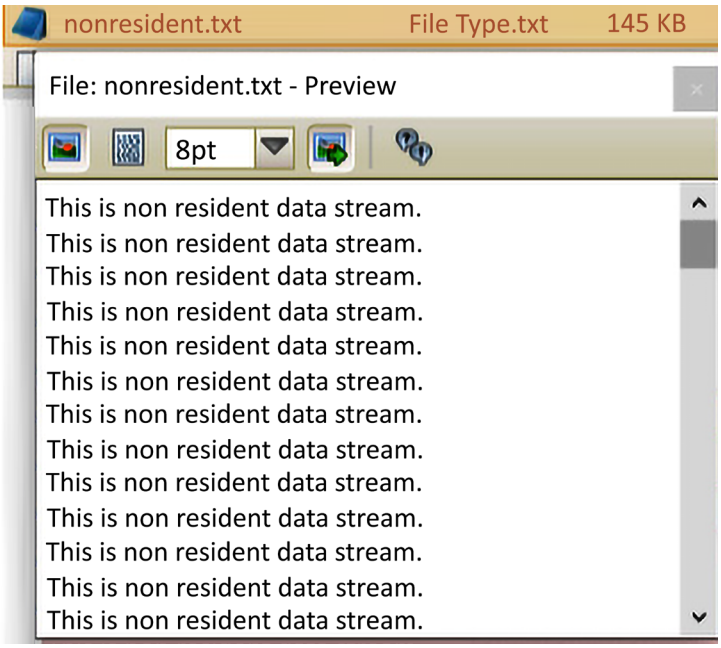


Figure 4.43: Non-resident data

When you look at the `$Data Attribute 0x80` of the file, as shown in the preceding diagram, we do not see the contents of the file, but we have pointers to the location of the file within the volume boundaries. We consider this to be non-resident content. Once the content of the attribute becomes non-resident, it can never become resident again. We commonly refer to the pointers in the file record of the attribute as a "run list" for the data runs of the non-resident data:

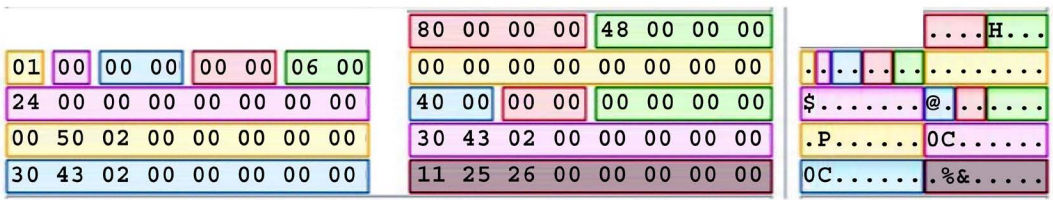


Figure 4.44: Non-resident data example

You can have a single data run, or multiple data runs, within the `$Data Attribute 0x80`. Deciphering the run list for the data runs can be tricky. In the following run list, we have the `$Data Attribute 0x80` with two run lists:

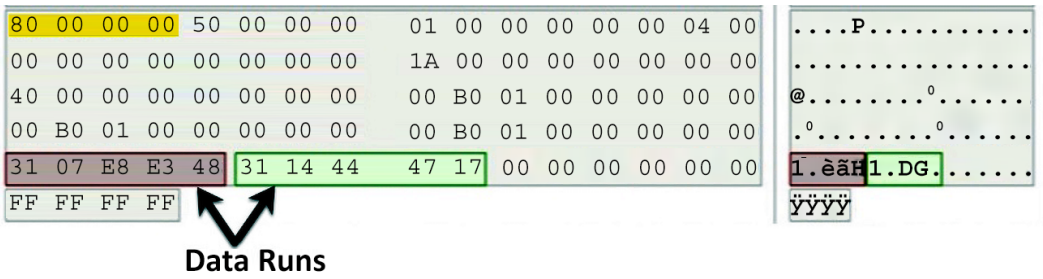


Figure 4.45: Run list

If the file is not fragmented, then you will have one run list pointing to the data run in the volume. If the file is fragmented (which is very common), then you will have multiple run lists providing information about the starting cluster for each fragment. I have taken the two run lists highlighted in the preceding list and created the following chart:

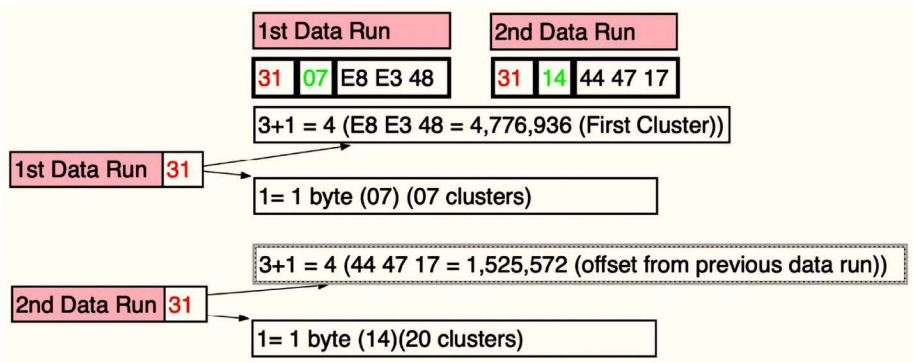


Figure 4.46: Run list map

The first run list comprises the hexadecimal values of `31 07 E8 E3 48`. Take the first byte of the header (`x/31`) and add the left and right nibbles (3+1 = 4). 4 is the number of bytes in the run list entry (this is `x/07 E8 E3 48`).

The right nibble (`x/1`) tells us that 1 byte represents the number of clusters being used for this fragment. We find a value of `x/07` in the length field, which represents 7 clusters for this fragment. The left nibble (`x/3`) informs us that 3 bytes (`x/E8 E3 48`) will represent the logical starter cluster of the fragment. At the end of the first run, we have a second run list of `x/31 14 44 47 17`.

Like the prior run list, we take the first byte of the header (`x/31`) and add the left and right nibbles (3+1 = 4). 4 is the number of bytes in the run list entry (which is `x/14 44 47 17`). The right nibble (`x/1`) tells us that 1 byte represents the number of clusters being used for this fragment. We find a value of `x/14` in the length field, which represents 20 clusters for this fragment.

The left nibble (`x/3`) informs us that 3 bytes (`x/44 47 17`) will represent the offset from the previous run list cluster. This process will keep going until the system hits `x/ 00 00 00 00`, which shows the end of the run lists.

That concludes our adventure into the world of NTFS. If you find yourself with a headache, you are not alone! This is just the basics of the filesystem. You can find entire books that have been written about NTFS, if you want to go into much greater detail.

Summary

In this chapter, we looked at how physical disks are constructed and prepared in order to store data. We discussed different partition schemes and how they address the creation of logical partitions. We also learned how filesystems differ and how data is organized.

In the next chapter, we will learn about the computer investigative process and how to analyze timelines, analyze media, and perform string searching for data.

Questions

1. Newer computer systems utilize the BIOS booting method.
    a. True
    b. False
2. A UEFI-based computer system will utilize _____ to boot from.
    a. MBR
    b. VBR
    c. GPT
    d. LSD
3. A cluster is the smallest storage unit on a hard drive.
    a. True
    b. False
4. An MBR-formatted disk can have more than four primary partitions.
    a. True
    b. False

5. A FAT32-formatted partition is laid out in two areas: a system area and a _____ area.
    a. Disk
    b. Doughnut
    c. Data
    d. Designer
6. In a FAT32-formatted partition, the root directory is in the system area.
    a. True
    b. False
7. In a NTFS-formatted partition, the filename is stored in the _____ attribute.
    a. Standard information
    b. Filename
    c. Data
    d. Security descriptor

The answers can be found at the end of the book under *Assessment*.

Further reading

Carrier, B. *File System Forensic Analysis*. Addison-Wesley, Reading, PA., Mar. 2005 (available at `https://www.kobo.com/us/en/ebook/file-system-forensic-analysis-1`).

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

5

Computer Investigation Process

Being a digital forensic examiner requires you to have a plan to conduct the investigation. For instance, there is the kitchen sink approach – where the person requesting the examination states, *I want it all*. However, this is not practical when the smallest drive might contain hundreds of thousands of pages or events. So while the kitchen sink approach is a plan, it may not be the most efficient.

In reality, your search method will depend on the crime you are investigating and whether there are limitations to the scope of the search. For example, in some investigations, the judicial authority may restrict an investigator's access to digital evidence to only email messages, or you may be limited to a specific date and time within the forensic image.

This chapter will first go through timeline analysis, where a user's activity is analyzed *temporally*. Then, we will examine the storage containers used by the user. You will also learn about string search, in which you search a dataset using matching strings of characters. Finally, in the last section, we will analyze data that has been deleted from the filesystem.

In this chapter, we will learn about the following topics:

- Timeline analysis
- Media analysis
- String search
- Recovering deleted data

Timeline analysis

During the investigation, you may find artifacts that appear to show the accused's guilt or innocence. However, we cannot construe the mere presence of the artifact as a sign of the suspect's guilt or innocence. Instead, the artifact needs to be placed within the user and system activity context.

For example, I was brought in as a consultant on a case; they accused the suspect of physically abusing their child. One piece of evidence that was considered against the suspect was the high number of Google searches about how to treat an injury. They attributed the searches to the accused, who was the father. The most challenging piece of evidence is to prove the user's identity behind the keyboard when the contested actions occurred. Since the items were present in the internet history (we will go into much greater detail in *Chapter 9*, *Internet Artifacts*), I wanted to check the context of when the searches were made. The wife was the primary owner of the laptop, but the husband was also a frequent user of the laptop. So, how do you attribute the searches to a specific user, especially when you have multiple people using the same laptop with the same user account?

A person's internet viewing habits can almost be as distinctive as a fingerprint. As I reviewed the one million-plus lines of internet history, I could differentiate the two different users on the laptop. I could correlate social media use with each user and attribute the Google searches to the child's mother. When she was confronted with the findings, the mother admitted that she searched for how to treat her child's injuries. After being presented with the evidence and testimony of the mother, the jury found the client not guilty of child abuse.

Suppose they had done a timeline analysis before making the charging decision. In that case, I believe the father would not have been charged, as the only evidence against him was the digital evidence found on the wife's laptop.

Your ability to create a timeline to analyze the system and actions of the user allows you to develop a much deeper and more thorough understanding of digital evidence. When I first started in the field, timelines were rudimentary and were typically based on the MAC times of the filesystem. **MAC** times refer to the **Modified, Accessed, and Created** times that are records created by the filesystem as created, edited, or accessed. The downside to only using MAC times for timeline analysis is that the recorded times may not be accurate. For example, this can happen when files are moved from one volume to another or if a user uses a third-party tool to change the timestamps and the timestamps are dependent on the system time.

We will now use multiple sources to help us to determine the context of what is happening on a system regarding a specific artifact. These additional sources may not be as easily manipulated as the MAC times and can determine any irregularities in the timestamps. For example, using multiple resources found within the forensic image, we can see when the user logs in, launches an executable, and accesses a file associated with the executable. This method of accessing multiple sources helps us confirm and validate the information provided by the MAC times.

Applying multiple frames of reference to the event being investigated allows us to support our hypothesis about the event. For example, can we determine whether the investigated incident results from user activity or is it a system process? In addition, using all of the available sources such as event logs, filesystem logs, or internet history captured by the system allows us to get into the small details to see the context of the event.

By gathering data points from multiple sources, you can create what Rob Lee from the SANS Institute calls a super timeline because of the sheer amount of data points you will have to sort through.

Hard drive capacity is not getting smaller. Instead, it is increasing at a phenomenal rate. Users and developers use this increased capacity to store more data and increase the number of logs that can track what occurs in a system. In some investigations, you may not need to examine the content of the files; for example, in an investigation dealing with illicit images, I need not see the visual depiction of the file. Instead, to answer whether a user knew about the existence of a specific file, I can use timeline analysis to make that determination.

Commercial forensic (and open-source) tools have made many advances when it comes to creating timelines. For example, at one time, you had to use many tools to extract data to create a timeline. Now you can use just a single tool to create a timeline.

Note

In this chapter, we will be discussing date-times, which will be converted into UTC/GMT. Always be aware of which time zone your dataset is operating in and the time zone it is stored. I use GMT/UTC as a standard when conducting an examination.

In this chapter, I will demonstrate the use of several tools for you to see the difference in the outputs and discuss where the tools pull the information from.

X-Ways

X-Ways Forensics has a very robust timeline-creation utility built in, called an **event list**. X-Ways compiles multiple sources such as timestamps at the filesystem level, internal timestamps, browser histories, event logs, registry hives, emails, and many other sources. When you start an event list, the data will be presented chronologically, creating a timeline. The event list is a very detailed timeline with copious amounts of information, which allows you to see the sequence of events of the incident you are investigating.

Note

As you explore the features of a new tool, remember to validate the tool against a known dataset. We will use a forensic image offered by Digital Corpora for this lab. You can visit `https://digitalcorpora.org/` and go to the 2008 M–57 Jean scenario for more information.

In this scenario, you are investigating a data leak. Someone has posted a spreadsheet containing an organization's confidential information onto a competitor's website, and the spreadsheet came from the computer of the CFO, Jean. During her interview, Jean stated that she emailed the spreadsheet to the president, Allison, at her request. The spreadsheet is `m57plan.xls` and can be found on the desktop of Jean's account. It has an MD5 hash value of `e23a4eb7f2562f53e88c9dca8b26a153` and a modified time of **2008-JUL-20 01:28:03 GMT** that also corresponds to Jean's statement regarding when she emailed the spreadsheet.

The filename and time frame give us a starting point for conducting the timeline analysis. When you are in the user environment of X-Ways Forensics, select the icon for the event list:

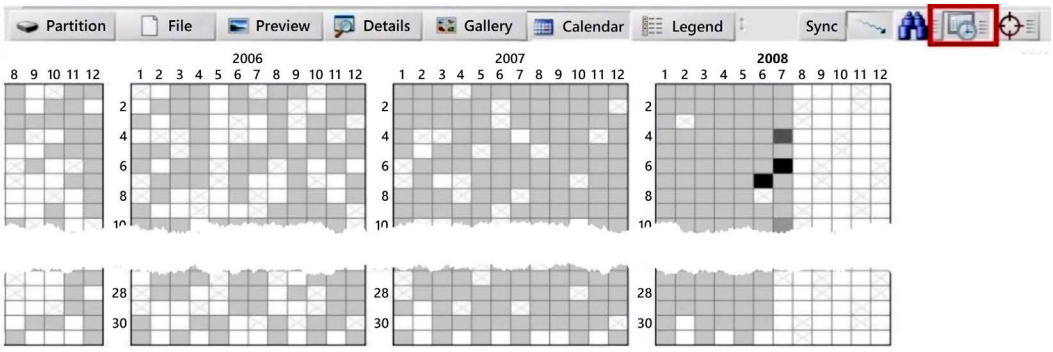


Figure 5.1: X-Ways

As you can see in the preceding screenshot, when you select the **Calendar** option, it will show you the calendar interface so that you can drill down to a specific day. If I do not filter any of the results on the event list, I have over one million entries that I will need to parse through. My preferred workflow method is to start big and then filter the results to meet the needs of my investigation.

When I filter down to July 20, I have reduced my results to a much more manageable 4,052 events.

Once we filter the results, let's search for the filename and see what activity has occurred. One of the first results shows that at 01:27:42, the system created a link file for the spreadsheet. In the following screenshot, you can see the user activity from 01:27 to 01:28. A pre-fetch file (`EXCEL.EXE-1C75F8D6.pf`) was created for Excel at 01:27, which shows the user starting the Excel program and then opening the spreadsheet, which corresponds to the creation of a link file:

Figure 5.2: Filter results

When you view the event list, you can see where the forensic tool is getting the information that is being displayed. The creation of the pre-fetch file starts with a change in the NT `user.dat` file. The tool follows along from gathering information from the internal file metadata to the operating system artifact. We can follow along and observe what occurs at the user and system levels as the user activity is being recorded.

If you look at timestamp 01:28:00, you can see that Jean sent a message out. In the **Name** column, we can see the subject of the email, and when we double-click on it, we can view the email itself:

Figure 5.3: Jean's email

We can see that Jean has emailed what appears to be `allison@M57.biz`, but, in reality, it is going to `tuckgorge@gmail.com`. We can then filter by file type, in this case, the `.eml` files, and you can see the results as follows:

Subject	**RE: Please send me the information now**
Date	07/20/2008 01:28:47 +0
Sender	Jean User <jean@m57.biz>
Recipients	tuckgorge@gmail.com
Attachments	m57biz.xls

I've attached the information that you have requested to this email message.

----- Original Message -----
From: alison@m57.biz [mailto:tuckgorge@gmail.com]
Sent: Sunday, July 20, 2008 2:23 AM
To: jean@m57.biz
Subject: Please send me the information now

Hi, Jean.

I'm sorry to bother you, but I really need that information now --- this VC guy is being very insistent. Can you please reply to this email with the information I requested --- the names, salaries, and social security numbers (SSNs) of all our current employees and intended hires?

Thanks.

Alison

E-mail Header
Date: 20 Jul 2008 01:28:47 -0000 From: Jean User <jean@m57.biz> Sender: Jean User <jean@m57.biz> To: <tuckgorge@gmail.com> Subject: RE: Please send me the information now Importance: Normal Mime-Version: 1.0 Content-Type: multipart/mixed; boundary="----=_NextPart_0"

Figure 5.4: Jean's email header

When you look at the **Sender** and **Recipients** columns, and when the data is sorted chronologically, you can get a good idea about the email communication between the attacker and Jean. It appears they have compromised Allison's account, as we can see the name "Alex" and the email account `tuckgorge@gmail.com` associated with the account.

Using the event list feature of X-Ways Forensics allows us to pinpoint when the file was compromised and from what vector. Now we can direct our investigation to Allison's computer to determine whether the attacker compromised her system. Based on these initial results, I believe the attacker targeted Jean in a phishing attack.

What I like about X-Ways Forensics is its ability to gather the dates and times from traditional sources and combine them with the actual artifacts, in this case, the emails. This gives you another level of granularity and context for your investigation.

The X-Ways Forensics documentation lists the following as sources of information for the event list feature:

Index.dat file(s)	Browser history
LNK file(s)	USNJrnl
Registry	Event log(s)
Metadata of Microsoft Office file(s)	Email message(s)
Recycle Bin file(s)	Shadow copy file(s)
Prefetch file(s)	Restore point(s)
Cookie(s)	MAC timestamp(s)

As you can see, this shows a very diverse list of sources. However, when used for analysis, it can give the investigator the confidence to rely on the date timestamps they are reporting in their investigation.

I have found that forensic suites also include timeline analysis with their products. I have discussed X-Ways Forensics and its ability to create a timeline for analysis with its event list feature. I have included a list of some additional forensic suites that you may use to analyze timeline data. The following list is not inclusive of all the forensic suites that are available:

- Belkasoft Evidence Center: `belkasoft.com/ec`
- Autopsy: `www.sleuthkit.org/autopsy`
- Recon Lab: `sumuri.com/software/recon-lab`
- PALADIN: `sumuri.com/software/paladin`

X-Ways is not the only tool you can use to create timelines; there are also several open-source tools that you can utilize. One of the most common is **Plaso/log2timeline**, which we will discuss next.

Plaso (Plaso Langar Að Safna Öllu)

Plaso (Plaso Langar Að Safna Öllu) is a Python backend and framework for the `log2timeline` tool. `log2timeline` is a forensic tool that pulls out timestamps from a system and creates a database of all the events, also known as a super timeline.

Note

You can download Plaso at `https://github.com/log2timeline/plaso`.

Plaso will work on most operating systems and was initially designed to replace the Perl version of `log2timeline`. However, the development has now shifted to modules, and they have created several CLI tools supported by the Plaso backend.

The tools supported by Plaso are activated by the **command-line interface (CLI)**. While the CLI can intimidate the user, if you take your time and proceed slowly, you will take the mystique out of the CLI. Many open-source tools use the CLI instead of the **graphical user interface (GUI)**. The very core of the CLI consists of two parts: the executable and the modifiers. Once you learn the specific modifiers for the CLI command, you will see that it all falls into place.

Let's talk about the tools included with Plaso:

- `image_export`
- `log2timeline`
- `pinfo`
- `psort`
- `psteal`

image_export

`image_export` will export file content from a device, media image, or forensic image. There are several parameters that you can use to define the information you wish to extract.

In the Windows version of the executable, the executable will end with `.exe`. With macOS, you may see it end in `.sh`.

Using –h or `--help` will give you the full list of parameters:

```
c:\tools\plaso>image_export.exe -h
usage: image_export.exe [-h] [--troubles] [-V] [-d] [-q]
                        [--artifact_definitions PATH]
                        [--custom_artifact_definitions PATH] [--data PATH]
                        [--logfile FILENAME] [--partitions PARTITIONS]
                        [--volumes VOLUMES] [--no_vss] [--vss_only]
                        [--vss_stores VSS_STORES]
                        [--artifact_filters ARTIFACT_FILTERS]
                        [--artifact_filters_file PATH]
                        [--date-filter TYPE_START_END] [-f FILE_FILTER]
                        [-x EXTENSIONS] [--names NAMES]
                        [--signatures IDENTIFIERS] [-w PATH]
                        [--include_duplicates]
                        [IMAGE]
```

Figure 5.5: image_export

Further down the screen, you will see detailed explanations for the modifiers. Note that I will only cover the most used options; there is additional documentation that we will not discuss here:

- `--names NAMES`: The filter on filenames. This option accepts a comma-separated string denoting all filenames, for example, `x NTUSER.DAT,UsrClass.dat`.
- `-w PATH`, `--write PATH`: The directory in which extracted files should be stored.
- `--data PATH`: The path to a directory containing the data files.
- `-x EXTENSIONS`, `--extensions EXTENSIONS`: The filter on filename extensions. This option accepts multiple comma-separated values, for example, `csv`, `docx`, and `pst`.

If you use the following command, it will export the `.xls` file to the `files` folder:

```
image_export --names 'm57plan.xls' C:\tools\plaso\image\jean.001 -w C:\
tools\plaso\export\files
```

You can see the breakdown of the preceding command as follows:

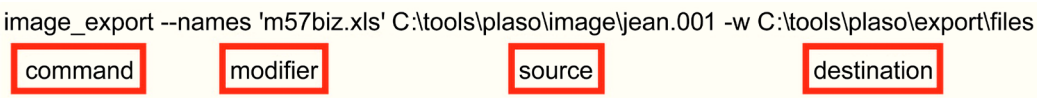


Figure 5.6: CLI map

Here, with the `image_export` command, we are using the `names` modifier to look for a specific file. In this case, it is `M57plan.xls`.

Now, you can tell the executable where to search; in this command, we are searching in the forensic image, `jean.001` (make sure that you include the full path to where the forensic image is located). Next, you can indicate where you want the exported files to be sent. The `-w` modifier will specify the write location.

You will find that the modifiers have some commonality with the commands within the Plaso framework.

log2timeline

`log2timeline` is a CLI tool that is designed to extract chronological-based events from files, directories, forensic images, and devices. It will create a database file (`.plaso`) that can then be analyzed by a variety of tools.

As you can see in the following screenshot, the `-h` modifier (help) will display the options for the command. As before, there are detailed explanations not displayed that will give you additional context for these commands. You should be able to recognize some of them from the previous command we looked at:

```
c:\tools\plaso>log2timeline.exe -h
usage: log2timeline.exe [-h] [--troubles] [-V] [--artifact_definitions PATH]
                        [--custom_artifact_definitions PATH] [--data PATH]
                        [--artifact_filters ARTIFACT_FILTERS]
                        [--artifact_filters_file PATH] [--preferred_year YEAR]
                        [--process_archives] [--skip_compressed_streams]
                        [-f FILE_FILTER] [--hasher_file_size_limit SIZE]
                        [--hashers HASHER_LIST]
                        [--parsers PARSER_FILTER_EXPRESSION]
                        [--yara_rules PATH] [--partitions PARTITIONS]
                        [--volumes VOLUMES] [-z TIMEZONE] [--no_vss]
                        [--vss_only] [--vss_stores VSS_STORES]
                        [--credential TYPE:DATA] [-d] [-q] [--info]
                        [--use_markdown] [--no_dependencies_check]
                        [--logfile FILENAME] [--status_view TYPE] [-t TEXT]
                        [--buffer_size BUFFER_SIZE] [--queue_size QUEUE_SIZE]
                        [--single_process] [--temporary_directory DIRECTORY]
                        [--worker_memory_limit SIZE] [--workers WORKERS]
                        [--sigsegv_handler] [--profilers PROFILERS_LIST]
                        [--profiling_directory DIRECTORY]
                        [--profiling_sample_rate SAMPLE_RATE]
                        [--storage_format FORMAT]
                        [--task_storage_format FORMAT]
                        [STORAGE_FILE] [SOURCE]
```

Figure 5.7: log2timeline

Try using the `info` modifier, as follows:

```
c:\tools\plaso>log2timeline.exe --info
```

You will get a list of all of the supported plugins, parsers, and output modules:

```
****************************** Parser Presets *******************************
      Name : Description
------------------------------------------------------------------------------
   android : android_app_usage, chrome_cache, filestat, sqlite/android_calls,
             sqlite/android_sms, sqlite/android_webview,
             sqlite/android_webviewcache, sqlite/chrome_27_history,
             sqlite/chrome_8_history, sqlite/chrome_cookies, sqlite/skype
     linux : bash_history, bencode, czip/oxml, dockerjson, dpkg, filestat,
             gdrive_synclog, olecf, pls_recall, popularity_contest, selinux,
             sqlite/google_drive, sqlite/skype, sqlite/zeitgeist, syslog,
             systemd_journal, utmp, webhist, xchatlog, xchatscrollback,
             zsh_extended_history
     macos : asl_log, bash_history, bencode, bsm_log, cups_ipp, czip/oxml,
             filestat, fseventsd, gdrive_synclog, mac_appfirewall_log,
             mac_keychain, mac_securityd, macwifi, olecf, plist,
             sqlite/appusage, sqlite/google_drive, sqlite/imessage,
             sqlite/ls_quarantine, sqlite/mac_document_versions,
             sqlite/mac_notes, sqlite/mackeeper_cache, sqlite/mac_knowledgec,
             sqlite/skype, syslog, utmpx, webhist, zsh_extended_history
   webhist : binary_cookies, chrome_cache, chrome_preferences,
             esedb/msie_webcache, firefox_cache, java_idx, msiecf,
             opera_global, opera_typed_history, plist/safari_history,
             sqlite/chrome_27_history, sqlite/chrome_8_history,
             sqlite/chrome_autofill, sqlite/chrome_cookies,
             sqlite/chrome_extension_activity, sqlite/firefox_cookies,
             sqlite/firefox_downloads, sqlite/firefox_history
      win7 : amcache, custom_destinations, esedb/file_history,
             olecf/olecf_automatic_destinations, recycle_bin, winevtx, win_gen
 win7_slow : mft, win7
   win_gen : bencode, czip/oxml, esedb, filestat, gdrive_synclog, lnk,
             mcafee_protection, olecf, pe, prefetch, sccm, skydrive_log,
             skydrive_log_old, sqlite/google_drive, sqlite/skype,
             symantec_scanlog, usnjrnl, webhist, winfirewall, winjob, winreg
     winxp : recycle_bin_info2, rplog, win_gen, winevt
winxp_slow : mft, winxp
------------------------------------------------------------------------------
```

Figure 5.8: Results of the info modifier

From the preceding output, you can see that some of the presets include collecting artifacts from many filesystems.

At a very basic level, you can use the following command structure:

```
log2timeline OUTPUT INPUT
```

One idiosyncrasy of `log2timeline` is that the output file is the first modifier to the executable and then you specify the input:

```
log2timeline C:\tools\plaso\export\files\jean.plaso C:\tools\plaso\image\
jean.001
```

When the command executes, you should see the following output on the screen:

```
c:\tools\plaso>log2timeline C:\tools\plaso\export\files\jean.plaso C:\tools\plaso\image\jean.001
2022-02-02 13:00:51,847 [INFO] (MainProcess) PID:15324 <data_location> Determined data location: c:\tools\plaso\data
2022-02-02 13:00:51,862 [INFO] (MainProcess) PID:15324 <artifact_definitions> Determined artifact definitions path: c:\tools\plaso\artifacts
Checking availability and versions of dependencies.
[OPTIONAL]      missing: lz4.
[OK]
```

Figure 5.9: Output

As the command executes, it locates the data folder that contains the dependencies for the executable, and then it searches for the files that contain the information about the artifacts that may be stored within the system. This is a default folder and is installed when you install `plaso`.

We now have a `.plaso` file that we can find in the `files` folder. In some cases, you might not want to create the database file with every option, that is, the kitchen sink. Rather, you may wish to do a targeted examination of the timeline, in which case you would need to employ filters. Using the `-f` modifier will allow you to do that.

Note

If you want to download some premade filters, you can do so at `https://github.com/mark-hallman/plaso_filters`.

I downloaded the premade filters and created a folder, named `filter`, within the path of the `plaso` installation. As you see from the following screenshot, I have installed `plaso` in a folder called `tools` at the root of my `C` drive:

```
log2timeline -f filter_windows.txt C:\tools\plaso\export\files\jeanfilter.
plaso C:\tools\plaso\image\jean.001
```

And, as you can see in the following screenshot, the tool was able to locate my filter within the `artifacts` folder and created a new Plaso database file:

```
c:\tools\plaso>log2timeline -f filter_windows.txt C:\tools\plaso\export\files\jeanfilter.plaso C:\tools\plaso\image\jean.001
2022-02-02 13:10:51,785 [INFO] (MainProcess) PID:7896 <data_location> Determined data location: c:\tools\plaso\data
2022-02-02 13:10:51,799 [INFO] (MainProcess) PID:7896 <artifact_definitions> Determined artifact definitions path: c:\tools\plaso\artifacts
Checking availability and versions of dependencies.
[OPTIONAL]      missing: lz4.
[OK]
```

Figure 5.10: Filter

So far, we have covered several commands; however, we still have more to cover. The next command in the framework is `pinfo`.

pinfo

`pinfo` is a command line that is used to display information about the Plaso database file (`.plaso`).

The `plaso` database file will contain the following information:

- When the user executed the tool
- What options were used when the tool was run
- What information was obtained by the tool during the pre-processing stage
- The database metadata
- What was parsed and the parameters that were used
- The number of events extracted
- Tagged events

To learn more about the preceding options, execute the command with the `-h` modifier. While the options are similar, you will have a far smaller selection than with the other tools, as shown in the following screenshot:

```
c:\tools\plaso>pinfo -h
usage: pinfo [-h] [--troubles] [-V] [--compare STORAGE_FILE]
             [--output_format FORMAT] [-v] [-w OUTPUTFILE]
             [STORAGE_FILE]

Shows information about a Plaso storage file, for example how it was collected, what information was extracted from a source, etc.

positional arguments:
  STORAGE_FILE          Path to a storage file.

optional arguments:
  -h, --help            Show this help message and exit.
  --troubles            Show troubleshooting information.
  -V, --version         Show the version information.
  --compare STORAGE_FILE
                        The path of the storage file to compare against.
  --output_format FORMAT, --output-format FORMAT
                        Format of the output, the default is: text. Supported
                        options: json, text.
  -v, --verbose         Print verbose output.
  -w OUTPUTFILE, --write OUTPUTFILE
                        Output filename.
```

Figure 5.11: pinfo

When you use the `pinfo` command in its simplest form, you will get the following results:

```
  ------------------------------------------------------------
  ********** Plaso Storage Information **********
  Filename: jeanfilter.plaso
  Format version: 20190309
  Serialization format: JSON
  ------------------------------------------------------------
  ********** Sessions **********
  276a7520-999e-428b-a6b4-11fcf9cf987d : 2019-07-19T22:19:36.092703Z
  ------------------------------------------------------------
```

As you can see in the preceding output, you get the storage information about the file and how many sessions were used to create it.

You can send the results to the standard output, that is, the monitor, or you can use the `-w` modifier to create a text file with the results. The use of the additional tools on the `.plaso` file will create the GUID and the date timestamp of when the analysis was conducted.

The tool can also provide system information about the source system you are now examining:

```
------------------------------------------------------------------------
******* System configuration: 276a7520-999e-428b-a6b4-11fcf9cf987d *******
Hostname: N/A
Operating system: Windows NT
Operating system product: Microsoft Windows XP
Operating system version: 5.1
Code page : cp1252
Keyboard layout: N/A
Time zone: GMT
------------------------------------------------------------------------
```

After verifying the information in the database file, you can move on to the next command.

psort

`psort` is a CLI tool that allows you to filter, sort, and conduct analysis on the contents of the `plaso` database file. Just like with the previous commands, the `-h` modifier will show you all the options for the command. In the following `psort` screenshot, you can see the available options, and you should be able to recognize the commonality of the options between all of the commands in the `plaso` architecture:

```
c:\tools\plaso>psort -h
usage: psort [-h] [--troubles] [-V] [--analysis PLUGIN_LIST]
             [--temporary_directory DIRECTORY] [--worker-memory-limit SIZE]
             [--logfile FILENAME] [-d] [-q] [--status_view TYPE]
             [--slice DATE] [--slice_size SLICE_SIZE] [--slicer] [--data PATH]
             [-a] [--language LANGUAGE] [-z TIMEZONE] [-o FORMAT]
             [-w OUTPUT_FILE] [--fields FIELDS]
             [--additional_fields ADDITIONAL_FIELDS]
             [--profilers PROFILERS_LIST] [--profiling_directory DIRECTORY]
             [--profiling_sample_rate SAMPLE_RATE]
             [STORAGE_FILE] [FILTER]
```

Figure 5.12: psort

Let's discuss some of the new options:

```
-o FORMAT, --output_format FORMAT, --output-format FORMAT
```

Below is a list of the available output formats:

Name	Description
`dynamic`	Output events to a delimiter (comma by default) separated value output format, that supports a dynamic selection of fields.
`elastic`	Output events to an ElasticSearch database. Requires elasticsearch-py.
`elastic_ts`	Output events to an ElasticSearch database for use with Timesketch. Requires elasticsearch-py. Solely intended to be used by the Timesketch backend.
`json`	Output events to JSON format.
`json_line`	Output events to JSON line format.
`kml`	Output events with geography data into a KML format.
`l2tcsv`	Output events to `log2timeline.pl` legacy CSV format, with 17 fixed fields.
`l2ttln`	Output events to `log2timeline.pl` extended TLN format, with 7 fixed fields.
`null`	Do not output events.
`rawpy`	Output events in "raw" (or native) Python format.
`tln`	Output events to TLN format, with 5 fixed fields.
`xlsx`	Output events to an Excel spreadsheet (XLSX).

As you are processing with `psort`, you can export your findings outside of the `plaso` database. There are a wide variety of options that you can use to export the data for analysis. One of the more common formats for exporting is `l2tcsv`, which is the legacy format for `log2timeline` and is a `.csv` worksheet.

A potential issue you may run into when creating the `.csv` worksheet is that if the file you create is too large, some tools may not analyze it, nor will you be able to open it with your favorite spreadsheet program.

`--analysis list`: `psort` comes with analysis plugins installed by default (you can still create your own custom plugins) to allow you to go through the database file and extract and analyze the contents.

You can use the `--analysis list` modifier to view the complete list of plugins:

```
****************************** Analysis Plugins ******************************
                  Name : Description
------------------------------------------------------------------------------
        browser_search : Analyze browser search entries from events.
                         [Summary/Report plugin]
     chrome_extension : Convert Chrome extension IDs into names, requires
                         Internet connection. [Summary/Report plugin]
          file_hashes : A plugin for generating a list of file paths and
                         corresponding hashes. [Summary/Report plugin]
              nsrlsvr : Analysis plugin for looking up hashes in nsrlsvr.
                         [Summary/Report plugin]
           sessionize : Analysis plugin that labels events by session.
                         [Summary/Report plugin]
              tagging : Analysis plugin that tags events according to rules
                         in a tagging file. [Summary/Report plugin]
unique_domains_visited : A plugin to generate a list all domains visited.
                         [Summary/Report plugin]
               viper : An analysis plugin for looking up SHA256 hashes in
                         Viper. [Summary/Report plugin]
           virustotal : An analysis plugin for looking up hashes in
                         VirusTotal. [Summary/Report plugin]
     windows_services : Provides a single list of for Windows services found
                         in the Registry. [Summary/Report plugin]
------------------------------------------------------------------------------
```

Figure 5.13: List of analysis plugins

If we run the command, it will go through the `plaso` database file, tagging the specific events that have been identified in the `tag_windows.txt` file (which is part of the default installation and can be found in the `data` directory):

```
psort -o null --analysis tagging --tagging-file tag_windows.txt c:/tools/
plaso/export/files/jean.plaso
```

On completion of the process, it will show you how many tags were applied to the database:

```
************************* Analysis report: 0 *************************
String: Report generated from tagging
Generated on:2019-07-20T20:04:46.000000Z
Report text: Tagging plugin produced 9754 tags
----------------------------------------------------------------
```

Additionally, you can filter out extraneous data using the `--slice` modifier.

Note

5 minutes is the default value. If you want a longer or shorter time slice, you can add the amount after `DATE TIME` with `--slice_size <VALUE>`.

If you find the `GET` event, you may want to place that event into context by observing what occurred before and afterward:

```
psort -q --slice '2008-07-20 01:26:17' c:/tools/plaso/export/files/jean.
plaso -w c:/tools/plaso/export/files/jeansliceoutput.csv
```

The command will create a `csv` file, which contains events 5 minutes before and 5 minutes after the timestamp placed in the CLI.

The final tool in the framework is `psteal`, which we will discuss next.

psteal

`psteal` is the final CLI command in the plaso framework. It combines the `log2timeline` and `psort` commands to extract and process events in a single step. It is very much the kitchen sink approach, otherwise known as "I want it ALLLLLL", and it has a limited selection of modifiers when compared to the other CLI commands within the framework.

Once again, `-h` will provide you with a list of options for the command, which are displayed in the following screenshot:

```
c:\tools\plaso>psteal -h
usage: psteal [-h] [--troubles] [-V] [--preferred_year YEAR]
              [--process_archives] [--skip_compressed_streams]
              [--storage_file PATH] [--partitions PARTITIONS]
              [--volumes VOLUMES] [--credential TYPE:DATA]
              [--status_view TYPE] [--source SOURCE] [--data PATH]
              [--language LANGUAGE] [-z TIMEZONE] [-o FORMAT] [-w OUTPUT_FILE]
              [--fields FIELDS] [--additional_fields ADDITIONAL_FIELDS]
              [--buffer_size BUFFER_SIZE] [--queue_size QUEUE_SIZE]
              [--single_process] [--temporary_directory DIRECTORY]
              [--worker_memory_limit SIZE] [--workers WORKERS]

psteal is a command line tool to extract events from individual
files, recursing a directory (e.g. mount point) or storage media
image or device. The output events will be stored in a storage file.
This tool will then read the output and process the events into a CSV
file.
```

Figure 5.14: psteal

At a minimum, specify the source and the output. The process will create the plaso database file and place it in the root of the plaso installation. This location allows you to perform additional tagging, filtering, or analysis after the command completes. The naming convention for the database file created is `<timestamp>-<source>.plaso`.

Here's the command. It creates a `.csv` file that is almost 1 GB in size. However, if I change the output to `.xlsx`, it reduces the size to 35 MB. So, keep in mind that you are processing and analyzing your datasets:

```
psteal --source C:/tools/plaso/image/jean.001 -o l2tcsv -w c:/tools/plaso/
export/files/jean.csv
```

I am using a relatively small forensic image of a 20 GB hard drive. Just imagine if you were using a 500 GB or a 1 TB hard drive and it had been active for an extended period.

Now that we have created our database file and have exported the datasets we find relevant to the investigation, what do we do now? It is time to analyze the datasets to find the evidence that will either prove or disprove the allegation. The tools you use for analysis can simply be the spreadsheet reader of your favorite Office suite or a commercial open-source tool designed for that specific purpose.

It is not possible to cover all the tool options that are available to an examiner in this book. I will highlight several options that are available and summarize the tools for you. Ultimately, the analysis of the data is where the examiner eyeballs the dataset and reviews the findings. Once again, it comes back to the verification/validation of your forensic tools to ensure they are providing accurate results.

Here are a few tools:

- **ELK stack**: This can be found at `https://www.elastic.co`. It is an acronym for three open-source projects: Elasticsearch, Logstash, and Kibana. Elasticsearch is the search and analytical engine. Logstash is the data processor and ingest engine, while Kibana is the visualizer. You have the option to download the three engines and install them in the operating system of your choice. You have options for macOS, Windows, and Linux. There is also the option to pay for the cloud environment if you do not wish to host the systems within your environment.
- **TimelineMaker Pro**: This can be found at `www.timelinemaker.com`. It is a commercial product specifically designed for creating timeline charts. With this tool, you can import the CSV files created with the plaso framework.
- **TimeSketch**: This can be found at `https://github.com/google/timesketch`. It is an open-source forensic timeline-analysis tool. It is Linux-based. I have installed it in a virtual environment so that I can use it as needed. It can also be worked on collaboratively by different members of your team. You can also import from a variety of plaso framework output options.

- **Aeon Timeline**: This can be found at `www.aeontimeline.com`. It is a commercial product specifically designed for creating visual timelines. It will allow you to view relationships among events. It was initially designed for authors, but it can also be used to analyze super timelines. You can import the CSV files created using the plaso framework.
- **Timeline Explorer**: This can be found at `ericzimmerman.github.io/#!index.md`. Timeline Explorer is an open-source platform created by Eric Zimmerman, who wanted a tool to read MAC time and plaso-generated CSV files without the need to use Microsoft Excel. It is not designed to examine very large CSV files; in fact, Zimmerman recommends explicitly that it is best to open smaller, targeted timelines than one giant one.

Media analysis

You can use timeline analysis on several vectors, such as network analysis, media analysis, software analysis, and hardware analysis. Network analysis is where you analyze log files, trace files, and the communication content between users and their devices. Media analysis is analyzing physical storage devices such as hard drives, SSD drives, thumb drives, or optical storage disks. You will examine the content, allocated space, and slack space. Finally, when performing software analysis, you reverse-engineer malicious code and analyze the protection code for potential exports.

So, let's look at media analysis. The primary source for your digital investigation will be the forensic images of storage devices such as hard drives, SSDs, USB devices, optical disks, and mobile devices such as smartphones. Depending on your organization, you may be the person responsible for creating the forensic image, or the forensic image may be provided to you from another part of the organization. Remember, the forensic image is a bit-for-bit copy of the source device. In most cases, you do not want to use a backup as the source of your digital forensic investigation because a backup will not contain all of the information on the storage device.

The storage device may contain four different data types that you want to examine:

- **Allocated space**: This is the space on the storage device that a file occupies. The filesystem recognizes the storage space as being used.
- **Unallocated space**: This is the space on the storage device that is not occupied by a file. The filesystem recognizes the storage space as being available for use.
- **Slack space**: When the data is stored in a cluster, if the file does not completely fill a cluster, the remaining space not used by the file is referred to as slack space.

- **Bad blocks/sectors/clusters**: This is the space on the disk that has been marked bad by the filesystem because of a defect. It can also be used by a user to hide data from a casual inspection.

Brian Carrier describes the progression of media analysis in his paper "Defining Digital Forensic Examination and Analysis Tools" as follows:

- **Disk**: Physical storage devices such as a hard disk drive, SSD, or flash media.
- **Volume**: A container comprising a single disk or multiple disks. You may find numerous volumes on a single disk or a volume may span across multiple discs. You may see the term "volume" used interchangeably with the term "partition." Brian Carrier defines a partition as being restricted to a single physical disk, whereas a volume is a collection of one or more partitions.
- **Filesystem**: This is used within the boundaries of a volume and tracks file allocation and cluster use.
- **Data unit**: The smallest allocation unit available to the filesystem. In most cases, this will be clusters, or, in a UNIX-based system, it will be blocks.
- **Metadata**: This is the data about data. This includes the modified, accessed, and created date-time stamps, as well as any other information about the file that the filesystem and some applications track.

The goal of media analysis in your digital forensic investigation is to find relevant artifacts that will either prove or disprove the allegations you are investigating. In addition, as you conduct the digital forensic investigation, you may find artifacts that will direct your focus to other locations.

We will now discuss some different analysis techniques that you might use during your digital forensic investigation.

String search

A search method you might use during your digital forensic investigation is a string or byte search. This search technique is utilized when you have a keyword list of specific terms that you wish to search for. Most commercial and open-source forensic tools allow for string searches and will search the allocated, unallocated, and file slack spaces. You can use specific words, symbols, or strings of letters as the search criteria. Generally, you will want to have some predefined keyword lists before you start your digital forensic investigation.

Your keyword lists will fall into one of the following categories:

- **Generic keyword list**: This is a keyword list that you will use in every case. This list can also be further categorized by the subject of the investigation. For example, you may have one keyword list for digital forensic investigations into fraudulent activity and a different keyword list for digital forensic investigations into illicit images.
- **Case-specific keyword list**: This is a keyword list that you will use for a specific digital forensic investigation. As you prepare to conduct your digital forensic investigation, you will identify keywords based on the participants, locations, and, sometimes, the slang used by the participants. For example, you could have keywords based on usernames, email addresses, physical addresses, phone numbers, credit card numbers, and more.

Note

You should avoid keyword terms that are generic or have additional meanings. For example, if you were investigating a homicide, the word "kill" seems to be a valid term to search for. Unfortunately, "kill" is also a term used in the programming language(s) you will find in a computer system. This will leave you with a large number of false positives. Ideally, the goal is to have the keyword list to help filter out non-pertinent data so that you can focus your efforts efficiently.

You may encounter different encoding schemes as you are conducting your searches on forensic images, such as the following:

- **American Standard Code for Information Interchange (ASCII)** is a character-encoding scheme based initially on U.S. English and is limited to 256-character codes.
- **Unicode** was developed to overcome the limitations of ASCII. Each character has a unique 2-byte value resulting in the ability to define over 65,000 characters.

While keyword searching can be very powerful, there is a downside to this, as it is very literal when searching for content based on the keyword. For example, if you search for a word, it will not find an alternative spelling; that is, if you are searching for `ally`, the filter will not find `alley`. Luckily, there is an alternative search methodology known as pattern matching/regular expressions.

A regular expression uses character strings to create a search pattern, and it will find all instances that match the pattern. Here are some common symbols and their meanings when used to create a regular expression:

- **The asterisk symbol (*)**: Match the preceding character(s) for `X` amount of times. For example, `ca*t` will cause positive hits for `ct`, `cat`, `caat`, and `caaat`.
- **The pound sign (#)**: This will match a number (`0-9`).
- **The backslash (\)**: The following character will be interpreted literally. `\.` will be construed as a period.
- **Caret (^)**: Match the start of the text. For example, `^123` will cause the positive hits to start with `123`.
- **The dollar sign ($)**: Match the end of the text. For example, `123$` will cause positive hits to end with `123`.
- **Plus symbol (+)**: Repeat the preceding character(s) for one or more times. For example, `ca+t` will cause positive hits for `cat`, `caat`, and `caaat`.
- **Curly brackets** {…}: Repeat the preceding character(s) for `X` times (depending on the value in the bracket).
- **Brackets** `[...]`: This will match a single character in the brackets. For example, `[b,c,d]` will match on b, c, or d.
- **Brackets w/ ^ [^...]**: This will match any single character not in the brackets. For example, `[^b,c,d]` will match any character other than b, c, or d.
- **Brackets (range) [..-..]**: This will match any character within the given range. `[0-9]` will match any character from 0 to 9.
- **Dot** (.): The dot can take the place of any character.
- **Question mark (?)**: The preceding character may/may not be present. For example, `.e01?` will return `.e0(x)` values. x shows it may find any value after `.e0`.
- **Pipe (|)**: This matches any one-character set separated by the pipe (|) character. For example, `br(ead|ake|east)` will return matches for bread or brake or breast.

The following are some common examples of pattern matching that you may find helpful.

To search for an IP address, you can use the following regular expression:

```
\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}
```

The `\d` specifies that the following will match on a digit (number). The curly brackets, `{1,3}`, indicate the number can be from one to three digits. `\.` specifies a search for the `.` character. The `\d{1,3}` pattern then repeats an additional three times until it has the value for an IPv4 address.

To search for a US phone number, you can use the following expression:

```
((\(\d{3}\) ?)|(\d{3}-))?\d{3}-\d{4}
```

The `\(` will match the open bracket. `\d{3}` will match on a three-digit number. `\)` will match on the closed bracket. This pattern will give you the area code, `(###)`, in this format. The remaining regular expression will give you the first three digits, `\d{3}`, the dash, `-`, and the final four digits, `\d{4}`, of the US phone number. If the phone number is not formatted as `(###) ###-####` or `###-###-####`, you will not get a hit.

Regular expressions are a powerful tool, but they can also be very complicated to craft. I like to use the regular expression library (which can be found at `http://regexlib.com/Default.aspx`) to help me with my regular expression skills.

So, what happens when the user deletes a file or folder from the media? Let's discuss what happens when a file or folder is deleted next.

Recovering deleted data

When a file is deleted in the FAT filesystem, the data itself does not get changed. The first character of the directory entry will have it changed to a `xE5` and the file allocation table entries are reset to `x00`. When the filesystem reads the directory entries, and it encounters the `xE5`, it will skip that entry and start reading from the subsequent entries.

To recover deleted files, we need to reverse the process the filesystem used to delete the files. Remember, it has not changed the file contents; they still physically reside in their assigned clusters. We now need to reverse-engineer the deletion and recreate the file entry and the entries in the file allocation table. To do this, we need to find the first cluster of the file, the size of the file, and the size of the clusters in the volume:

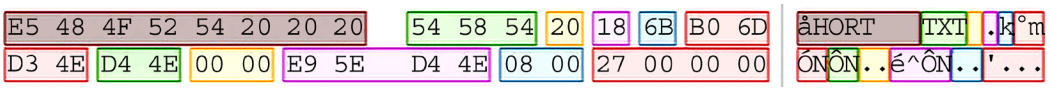


Figure 5.15: Deleted entry

In the preceding screenshot, we have a directory entry showing that a file has been deleted. We see the `xE5` at the start of the directory entry. (This will require the use of a hex editor to make the changes.) Then, we have to determine the starting cluster, `x00 x08` (which is shown as `x08 x00`), which is cluster number 8. To determine the file size, look at the last four bytes (remember that the FAT filesystem stores data in little-endian format, which means that the least-significant byte is on the left, so we would read that value as `x00 x00 x00 x27`, not as it is displayed, `x27 x00 x00 x00`), and when we convert the hexadecimal value to a decimal, we get the value of 39 bytes for the file size.

Now we have to determine how many sectors make up a cluster and what the sector size is. You will need to go to the boot record to get that information. The boot record shows that there are 512 bytes per sector, and there are 8 sectors per cluster, which gives us a cluster size of 4,096 bytes:

Bytes per sector	011	512	15
Sectors per cluster	013	8	114

Figure 5.16: Boot record

This means that our file will only occupy a single cluster. We then go to the file allocation table and look at the entry for cluster 8 and see that it is zeroed out:

```
00 00 00 00 FF FF FF 0F   0B 00 00 00 0C 00 00 00
0D 00 00 00 0E 00 00 00   0F 00 00 00 10 00 00 00
```

Figure 5.17: Deleted FAT

To recover the deleted file, perform the following steps:

1. You need to change the entry in the file allocation table from `x0000 x0000` to `xFFFF FFF8` or `xFFFF FF0F`. If this were a larger file, you would need to change the file allocation table entry to point to the next cluster until you reach the cluster that contains the end of the file. Should you find an entry marked as allocated before you reach the end of the file, you may be dealing with a fragmented file. Another possibility is when the clusters were made available for use when the file was deleted, the data from a new file was placed in the available space. This would cause the old data to be overwritten with the data from the new file.
2. The next step is to go back to the directory entry and replace `xE5` with another character. When replacing the `xE5` character of the filename in the directory entry, be careful not to guess what the character is. If you select an incorrect character, you could change the meaning or create a bias with the new filename, and that would be improper.

> I recommend that when you recover a deleted file, you replace that first character with an underscore or a dash so there is no misunderstanding about the filename.

When recovering a file with a long filename, it is important to relink the long filename to the short filename. This is because when the additional directories are created to accommodate the long filename, the system creates a checksum based on the data of the short filename. When you changed the `xE5` value on the short filename entry, you also want to use the same replacement character for the subsequent `xE5` entries for the long filename directory entries. The reason for linking the long filename to the short filename is that the short filename directory entry contains information such as the date and times, the starting cluster, and the file size.

As we discussed in *Chapter 4*, *Computer Systems,* when a file/directory is created on an NTFS volume, the system creates an entry in the `$MFT` file. The MFT record will contain the metadata about the file/directory; if the contents of the file are non-resident, then the `$Bitmap` file will be updated to show the clusters occupied by the file are allocated.

When a file/directory is deleted, then the sequence count in the MFT file record's header is incremented by one digit. The allocation status for the record will change from allocated to unallocated. If the file data is non-resident, the system will update the `$Bitmap` file to show the clusters occupied by the file are now unallocated.

Every MFT file entry will start with the file signature of the file, which you can use as a search term to locate MFT file entries in unallocated space. Until the clusters containing the data on the disk are overwritten, we can recover the data.

If the MFT file record is unused, then you can reverse the steps and recover the file. You can decipher the file record, as we discussed in *Chapter 4*, *Computer Systems*. If the file is resident within the file record, you will recover the data when you retrieve the MFT file record. If the data is non-resident, then you will have to decipher the MFT file record to determine whether the data runs and identify the occupied clusters.

If the system has overwritten the MFT file record, then you cannot recover the deleted MFT file record data or any resident data. You may recover the non-resident data, but that will depend on the size of the files and the fragmentation. Once the MFT record has been overwritten, you will lose any information regarding the data runs and which clusters contain the data.

Summary

In this chapter, we discussed, in detail, timeline creation and timeline analysis with open-source and commercial forensic tools. We took an in-depth look at utilizing the commercial forensic tool, X-Ways Forensics, and the open-source plaso framework for `log2timeline`. We also touched upon using the kitchen sink approach or a targeted examination of the dataset. Remember, we are not analyzing the contents of files, just the timelines associated with the files and other events within the operating system and filesystems.

In the next chapter, we will discuss the contents of files, specifically, Windows artifacts.

Questions

1. It is important for the examiner to know the time zone in which the evidence was collected.
    a. True
    b. False
2. You can do timeline analysis with X-Way Forensics when you create a(n) _____ list.
    a. Timeline
    b. Date/time
    c. Event
    d. Party
3. Plaso is a framework for how many tools?
    a. One
    b. Three
    c. Five
    d. Seven
4. `pinfo` will give you what information?
    a. Information about the examiner
    b. Information about the database file
    c. Information about the forensic machine
    d. Information about the suspect

5. `log2timeline` is a _____ -based tool.

    a. CLI
    b. GUI
    c. VFD
    d. XYZ

6. `psort` will give you the _____.

    a. Ability to sort
    b. Ability to filter
    c. Ability to connect
    d. All of the above

7. You can do a timeline analysis with an Excel spreadsheet.

    a. True
    b. False

Further reading

You can refer to the following links for more information on the topics covered in this chapter:

- *T. P. P. A. (2019, July 8). Plaso Documentation. Retrieved from The Plaso Project*: `https://buildmedia.readthedocs.org/media/pdf/plaso/latest/plaso.pdf`
- *Carvey, H. (2014). Windows forensic analysis toolkit: Advanced analysis techniques for Windows 8; Waltham, MA: Syngress*. Available at: `https://www.abebooks.com/servlet/SearchResults?sts=t&cm_sp=SearchF-_-home-_-Results&an=&tn=Windows+forensic+analysis+toolkit&kn=&isbn=`

Exercise

Data set

`Chapter 5 Emails.xlsx`

`Chapter 5 Carving.dd`

Software needed

Timeline Explorer - `https://ericzimmerman.github.io/#!index.md`

Microsoft .NET 6 or newer is required. You will get errors without at least .NET 6. When in doubt, install it! Make sure you get the **Desktop** runtime if you plan on running any of the GUI programs.

Autopsy - `https://www.autopsy.com/`

Email exercise

An individual outside of m57.biz purchased a laptop from Craigslist. The laptop the individual purchased contained child pornography and they decided to inform the police about it.

Investigators were able to trace the laptop back to m57.biz. When the police contacted the CEO of m57.biz, the CEO reported that the laptop, as well as other items, had been stolen from the m57 inventory.

The m57 CEO gave consent for the police investigators to search m57.biz and image all of the m57.biz computers, company phones, as well as USB drives.

Analyze the emails found in the `Chapter 5 emails.xlsx` spreadsheet and identify potential suspects and a timeline of their activity.

Data carving exercise

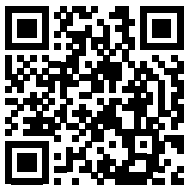
1. Load Autopsy and start a new case.
2. Select **Disk Image** or **VM file** for the data source.
3. Navigate to the folder where you stored the image `Chapter 5 Carving.dd`. Select only the following Ingest Modules:
    - PhotoRec Carver Embedded File Extractor
4. From the drop-down menu, select **All files**, **Directory**, and **Unallocated Space**.

Analyze the results.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`



6

Windows Artifact Analysis

The world runs on the Microsoft Windows operating system, with Microsoft accounting for nearly 90 percent of the operating system market share (`https://netmarketshare.com/`). In my personal experience, I have examined far more Windows operating systems than any other operating system; macOS would be the next most common operating system, with Linux running a distant third. While you have to be prepared to analyze all operating systems, whichever is the most common within the realm you are working in is where you should focus your attention.

This chapter will provide you with an understanding of the Windows operating system and the artifacts you may find. There are entire books written about the Windows operating system; this chapter's goal is to provide you with an understanding of the more common operating system artifacts you may encounter during your investigation. You will start by going through user profiles where an examiner can find most user data. Then, we will look at the Windows Registry to identify the Windows settings. You will also look at artifacts to determine the user's activities and learn how to identify which USB devices were used on the system. Finally, we will cover all of this in the following topics:

- Understanding user profiles
- Understanding Windows Registry
- Determining account usage
- Determining file knowledge
- Identifying physical locations
- Exploring program execution
- Understanding USB/attached devices

An operating system manages the hardware resources and allows the user to run other applications that are essentially programs within the operating system environment. It can be a treasure trove of artifacts to recreate user or system activity at any given moment in time. When we discuss the Windows operating system, this topic could cover multiple versions. At the time of writing, the current version of the Windows operating system is Windows 11, but Windows 10 is still on the majority of systems. That does not mean every system you examine will have Windows 10 installed on it. It is possible that even in a corporate environment, you could still examine a Windows XP client, although Microsoft released it in 2001 and no longer supports it.

I will focus on Windows 7, 8, and 10 for the rest of this chapter. First, however, there may be references to Windows XP because of Microsoft's legacy support for the operating system.

The first item I want to discuss is the different types of user profiles and where the operating system will store the user's data.

Understanding user profiles

When the Windows operating system is installed, it creates a default folder structure to store user and application data. Sometimes, just looking at the folder structure can tell you which version is or isn't installed.

When you are looking for user account profiles, the location can vary depending on the version of the operating system:

- For Windows XP, WinNT, and Win2000
    - `C:\Documents and Settings\%UserName%`
- For Windows Vista, 7, 8, and 10
    - `C:\Users\%UserName%`

When the user first logs on to the system, it will create a user profile. That profile will then be used for any subsequent logins and is now the user's environment for their activity on the system. Microsoft defines the different types of user profiles:

- **Local user profile**: This profile is created when the user logs on to a computer for the first time. You will find the profile stored on the hard disk. When changes are made to the profile, the changes will be specific to the user and stored on the local computer.

- **Roaming user profile**: This profile is an administrator-created, network-based profile. The profile will be downloaded to the localhost when the user logs in to the system. When any changes are made to the profile on the localhost, changes will also be made to the server copy when the user logs off from the localhost. This profile type removes the requirement on the part of the user to create a profile when they log on to different hosts on the network. (You will only find this type of profile in Enterprise environments.)
- **Mandatory user profile**: This profile is a profile created by the network administrators to lock users down to a specific set of settings when they use a host on the network. The user will not be allowed to change the profile without the administrator's approval. Any changes made by the user to the localhost environment will be lost when the user logs off from the localhost.
- **Temporary user profile**: This profile is created when an error occurs when the system loads the user's profile. When the user logs off, the profile is deleted. You will find the use of temporary profiles on computers running Windows 2000 and later.

Each user profile will have a registry hive – `NTUSER.DAT` – and is mapped to the system registry key of **HKEY Current User** when the user logs in. This registry hive contains the user's preferences and configuration settings.

Each user profile contains the following folders:

- `\Users\$USER$\Documents`
- `\Users\$USER$\Music`
- `\Users\$USER$\Pictures`
- `\Users\$USER$\Videos`

The `AppData` folder is a hidden folder that contains user-specific preferences and profile configurations and is further divided into three subfolders:

- `\Users\$USER$\AppData`
- `\Users\$USER$\AppData\Local`
- `\Users\$USER$\AppData\LocalLow`
- `\Users\$USER$\AppData\Roaming`

The `Roaming` folder contains data that can be synced within the server environment. Data such as web browser favorites or bookmarks will travel with the user as they log on to different workstations:

- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Cookies`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Network Shortcuts`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Printer Shortcuts`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Recent`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\SendTo`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Start Menu`
- `\Users\$USER$\AppData\Roaming\Microsoft\Windows\Templates`

The `Local` folder contains data related to the installation of programs. It is workstation specific and will not sync with the server (in a server environment). Temporary files are also stored here:

- `\Users\$USER$\AppData\Local`
- `\Users\$USER$\AppData\Local\Microsoft\Windows\History`
- `\Users\$USER$\AppData\Local\Microsoft\Windows\Temporary Internet Files`

The `LocalLow` folder includes low-level access data, such as the temporary files of your browser when running in protected mode.

That completes our discussion on user accounts, so let's move on to the registry, which is the heart and soul of the Windows operating system.

Understanding Windows Registry

The Windows Registry is the very heart of the Windows operating system and will be the source of many artifacts we will discuss later in the chapter. First, I will provide a high-level view of the registry. Then, suppose you want to dig deeper into the nuts and bolts of the registry. In that case, I highly recommend Harlan Carvey's book *Windows Registry Forensics – Advanced Digital Forensic Analysis of the Windows Registry*. Harlan Carvey is also the developer of the tool RegRipper, which is a tool we will use in this chapter.

What is the registry? Microsoft defines the registry as a central hierarchical database. This database is used to store configuration information about users, hardware devices, and applications.

But what does that mean for the forensic investigator? Windows continually references the information in the registry during operations. Information in the registry will contain profiles for each user, installed applications, different document types, and property settings for folders and application icons. The registry will also contain information about the hardware on the system, including networking information such as the ports used.

Wow. That was a mouthful, but in simple terms, the registry contains information about… almost everything on the computer system.

The components of the registry are found in the `%SystemRoot%\System32\Config` folder and are called hive files. You will find the `SAM`, `SECURITY`, `SOFTWARE`, and `SYSTEM` hives. Below is a brief description of the hives:

- The `SAM` hive is the Security Accounts Manager and contains login information about the users.
- The `SECURITY` hive contains security information and, potentially, password information.
- The `SOFTWARE` hive contains information about application information and the default Windows settings.
- The `SYSTEM` hive includes information on the hardware and system configuration.

There is an additional hive, `NTUser.dat`, which is stored in the root of the user profile. This hive contains information about user behavior and their settings.

Another file in the hive format is the `UsrClass.dat` file, which is found in the `\AppData\Local\Microsoft\Windows` folder of the user account. You will find information concerning **user access control** (**UAC**) configuration and information about the **graphical user interface** (**GUI**) display for the user experience.

The hive comprises subkeys that contain the **Value**, **Type**, and specific **Data** or settings being saved. This will give us a frame of reference as we explore the artifacts contained within the registry.

As you can see in the following screenshot, it is difficult to decipher the meanings of the subkeys and values and what they represent:

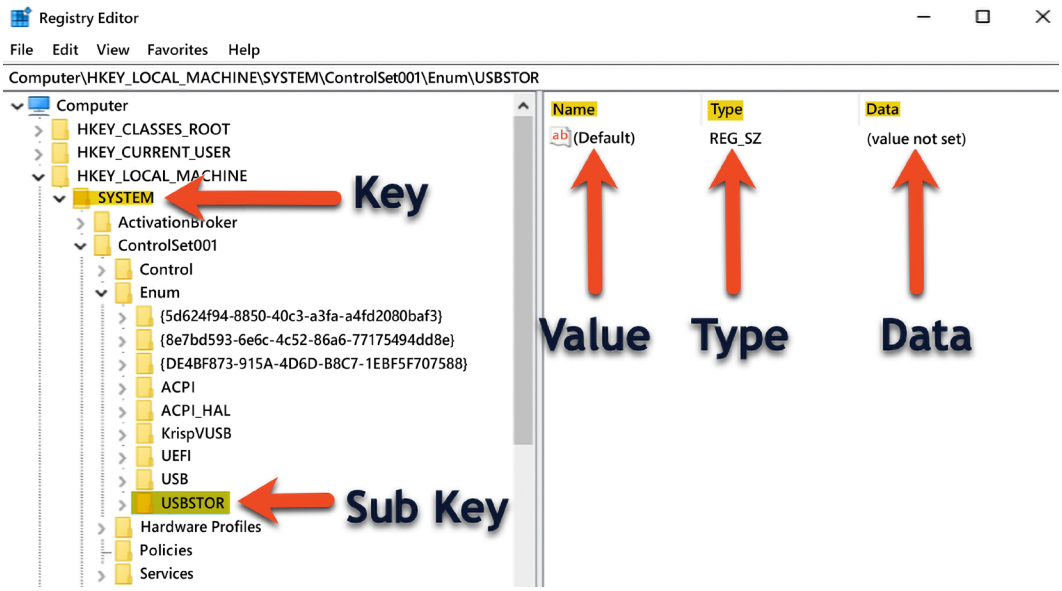


Figure 6.1: Registry Editor showing the USBSTOR registry key

As we go through the artifacts, I will show you the view you will see with the Registry Viewer and the easier-to-read parsed version created by the forensic tools.

We will use some open-source tools during this chapter:

- RegRipper (available for download from `https://github.com/keydet89/RegRipper3.0`), created by Harlan Carvey.
- Eric Zimmerman (whose work is available for download from `https://ericzimmerman.github.io/#!index.md`) has created several open-source utilities to parse Windows artifacts.

There are several categories in which we look for artifacts. I like to use the SANS catalog descriptions of the artifacts, which can be found at `https://digital-forensics.sans.org/community/posters` and are listed as follows:

- Account usage
- File knowledge
- Physical location
- Program execution
- USB/drive usage
- Browser usage (which we will discuss in *Chapter 9*, *Internet Artifacts*)

With this understanding of the user profile, we will now discuss the artifacts that determine what actions are associated with the user accounts.

Determining account usage

Identifying the user behind the keyboard is one of the hardest things you must do when conducting a digital forensic examination. You will have to parse through many artifacts to help make that determination. First, you will want to gather as much information about the user account in question and see whether you can relate it to the physical person. You will want to gain as much information about that user account and its activity related to the matter you are investigating. We will now go over some artifacts from a Windows-based operating system that will help you determine and identify that account activity starting with the user's last login or password change.

Last login/last password change

The following path will contain information about the user accounts on the system:

```
C:\windows\system32\config\SAM\Domains\Account\Users
```

To navigate to the location that contains the user account information, I will use Eric Zimmerman's Registry Explorer. I have exported the registry hive files from the forensic image to run Registry Explorer and RegRipper.

In the following screenshot, we can see that I have already opened the folder path and the subkeys, and within the **Users** subkey, there are folders with hexadecimal names and a folder entitled `Names`. Within the `Names` subkey, you see a listing of the accounts on the machine:

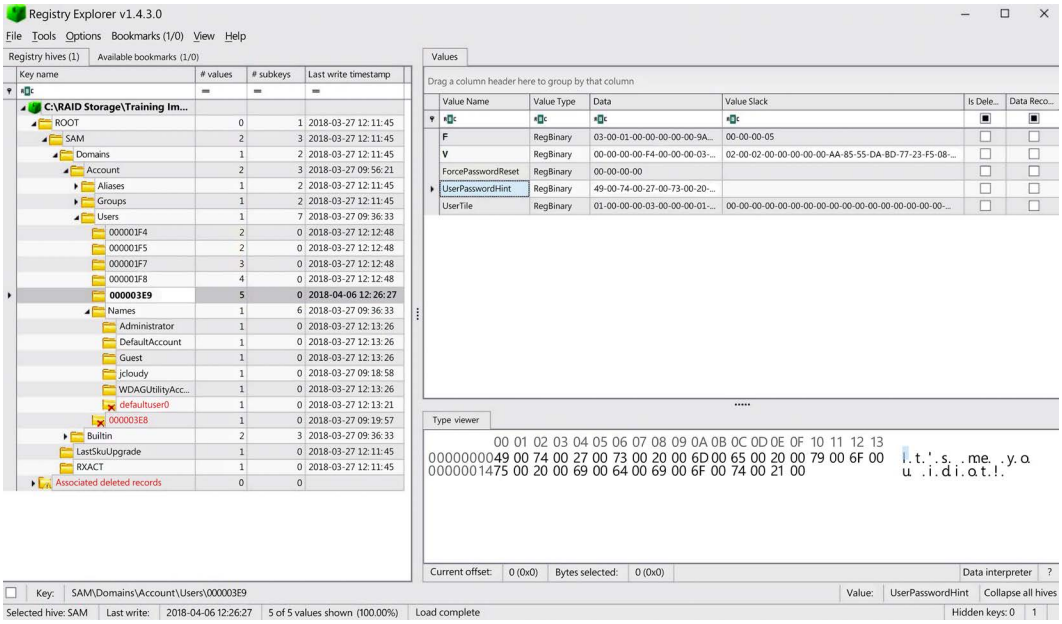


Figure 6.2: Registry Explorer showing the USERS key and subkeys

It lists the names in English, so they are easily readable. Out of the six accounts being shown, one has been deleted (`defaultuser0`), and one has the username of `jcloudy`. The value for the `jcloudy` subkey will point to the subkeys with the hexadecimal values. Here, `jcloudy` points to `x3E9`.

In subkey `x3E9`, as shown in the following screenshot, I see that I have an **F** and a **V** value and below that, I can see information pertaining to the user's passwords:

Drag a column header here to group by that column

Value Name	Value Type	Data	Value Slack	Is Del...	Data Record...
F	RegBinary	03-00-01-00-00-00-00-00-9A...	00-00-00-05	☐	☐
V	RegBinary	00-00-00-00-F4-00-00-00-03...	02-00-02-00-00-00-00-00-00-AA-85-55-DA-BD-77-23-F5-08...	☐	☐
ForcePasswordReset	RegBinary	00-00-00-00		☐	☐
UserPasswordHint	RegBinary	49-00-74-00-27-00-73-00-20...		☐	☐
UserTile	RegBinary	01-00-00-00-03-00-00-00-01...	00-00-00-00-00-00-00-00-00-00-00-00-00-00-00-00-...	☐	☐

Figure 6.3: Registry subkey X3E9

To make it easier, we can run RegRipper and see whether we can get an easier-to-read output. An example of the output for the `jcloudy` account is as follows:

```
Username        : jcloudy [1001]
SID             : S-1-5-21-2734969515-1644526556-1039763013-1001
Full Name       :
User Comment    :
Account Type    :
Account Created : Tue Mar 27 09:18:58 2018 Z
Name            :
Password Hint   : It's me you idiot!
Last Login Date : Fri Apr  6 12:26:27 2018 Z
Pwd Reset Date  : Tue Mar 27 09:18:58 2018 Z
Pwd Fail Date   : Fri Apr  6 03:30:52 2018 Z
Login Count     : 23
  --> Password does not expire
  --> Password not required
  --> Normal user account
```

Figure 6.4: RegRipper output for the jcloudy account

RegRipper parses the data and presents it in an easy-to-read format. And we can see when the account was created, the password hint, the last time the user logged in, and the number of times the user has logged in to the system.

As you look at the username `jcloudy`, you can see the numerals `1001`, and below that, an entry marked `SID`.

SID is the **security identifier** used by the Windows operating system to identify objects within. This is how Windows addresses components internally. At the end of the SID is the **relative identifier** (**RID**), which is the last digits after the SID. For example, if you see `500` as the RID, that will identify the administrator account for that system. The guest account would have an RID of `501`. In this case, as shown in the following diagram, we see the RID of `1001`. This informs me that the `jcloudy` account is user-created, and is not an account created by the system through an automated process:

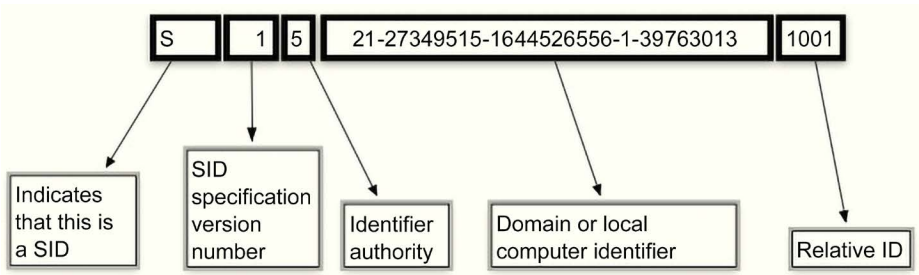


Figure 6.5: Breakdown of the SID

When doing your exam, the most commonly looked at portion of the SID is the RID. We can associate the RID with a specific user account. As the user creates accounts on the system, the RID will increase by one digit. For example, we could have a user, `user X`, with an RID of `1005`, and if I cannot find accounts `1001` through `1004`, it is possible that someone/something deleted those user accounts.

We are going through the registry to find artifacts that support (or do not support) our hypothesis about what occurred. Another source of information to help determine what happened on the system are the event logs.

Windows categorizes events into three different classes:

- **System**: Information generated by the Windows operating system
- **Application**: Information generated by applications on the local machine
- **Security**: Information related to login attempts

In Windows Vista through Windows 10, we can find the event logs at the following path:

```
C:\Windows\System32\winevt\logs
```

A common excuse that users give when they are accused of using the system for criminal or inappropriate reasons is that someone else had access to their system. **Remote Desktop Protocol (RDP)** is a way to access a host from another location. The security log will record any access using the RDP protocol. You will want to look for event ID numbers **4778** and **4779**, which would show you when the service connected/reconnected and when it disconnected.

You can also search for the type of logon into the system. For example, when we examine the security log for event ID **4624**, this will tell us the day, time, username, and the means with which the login was successful. As you can see in the following screenshot of **Event Viewer**, you can use this application to review the exported log files.

Once you have loaded the selected log file you want to examine, you can filter the results only to show the events that are relevant to your investigation:

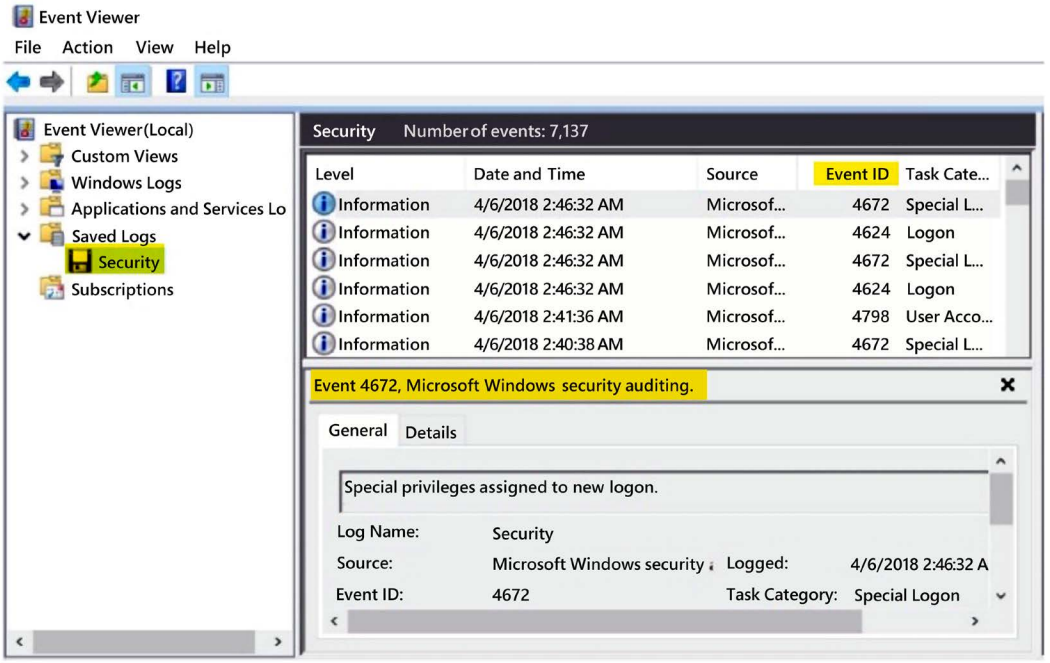


Figure 6.6: Event Viewer displaying event information

The type of logon is also significant. Was the user sitting at the keyboard, or did the user log in from a remote site? Event ID 4624 will identify the login type used by the user. In the following screenshot, you can see the output of Event Viewer showing when the user logged in and the login type.

Here, it shows the user's login was type 2, which is "interactive":

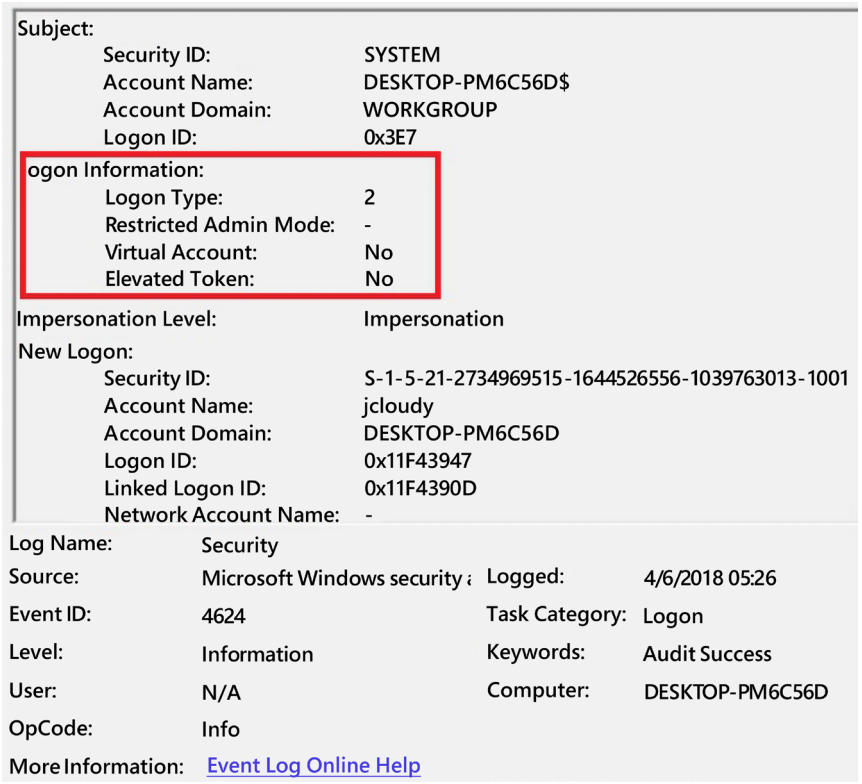


Figure 6.7: Event Viewer showing the logon type

The following is a list from Microsoft of the other logon types you may encounter, together with their descriptions:

Logon Types	Description
Interactive	Logon to the local host by the user.
Network	A network logon to the local host by the user.
Batch	Allows processes to be started without user input.
Service	Automated process. No user input needed.
Unlock	The local host was unlocked via user input.
NetworkCleartext	Network logon to the local host by the user. The password was sent in cleartext to the authentication package. The password was then encrypted before it was sent on the network.
NewCredentials	The user account was duplicated and received new credentials for the network connection leaving the secure network.
RemoteInteractive	A logon to the local host by the user using a remote application.
CachedInteractive	A network logon to the local host by the user, using the network credentials on the local host.

Figure 6.8: Microsoft logon types

You may also want to establish the attempted login events to determine whether an attacker compromised the account. The following event IDs will help you make that determination:

- `4624` - Successful logon
- `4625` - Unsuccessful logon
- `4634` - Logon session terminated
- `4647` - Logon session terminated by the user
- `4648` - User logon was attempted by a user using different credentials
- `4672` - User logon with Admin rights
- `4720` - User account created

A full list of Microsoft Windows Event IDs can be found at:

`https://www.ultimatewindowssecurity.com/securitylog/encyclopedia/`

Suppose you see many failed logins, or a user was granted administrator rights when they usually do not possess superuser rights. In that case, these event ID clues provide you with additional investigative avenues to determine what occurred.

Now that we've examined the user's account activity, next we will discuss the artifacts associated with user account file access.

Determining file knowledge

Some incidents you investigate may deal with contraband images, stolen data, or unlawful access to data. You will have to determine whether the user had knowledge of the file(s) in question, or whether the file(s) existed on the user's system.

We will now talk about some artifacts you can find in the Windows operating system that will help you make that determination.

Exploring the thumbcache

A thumbcache is a database of thumbnail images created when using Windows Explorer in a thumbnail view. Depending on the size of the thumbnail, you may have multiple databases with the same image but with different sizes. It depends on the view the user selected while in Windows Explorer. The existence of an image found in the database is not substantial proof that the user knew the image was on the system. The system can add a thumbnail to the cache without the user's knowledge. The thumbcache can be found in the user's profile at the following path:

```
AppData\Local\Microsoft\Windows\Explorer
```

Your commercial forensic tools will process the thumbcache with no issues. If you want to use an open-source utility, you can use Thumbcache Viewer (which can be downloaded at `https://thumbcacheviewer.github.io/`).

The following is an example of the output of Thumbcache Viewer:

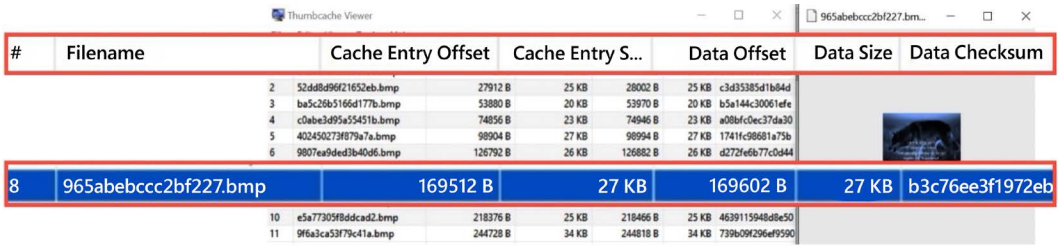


Figure 6.9: Thumbcache Viewer output

As you can see, the thumbnail does not have the same filename as the source image. To identify the original file that was used to create the thumbnail, we need to look in the Windows Search Indexing database, `Windows.edb`, which can be found at the following path:

```
C:\ProgramData\Microsoft\Search\Data\Applications\Windows\Windows.edb
```

You will need an additional tool to find the information about the image used to create the thumbnail. You can use `ESEDatabaseView` (located at `https://www.nirsoft.net/utils/ese_database_view.html`).

The thumbnail name is `96 5a be bc cc 2b f2 27`, which is made up of hexadecimal characters. We need to reverse the values to search the database, so we will want to search for `27 f2 2b cc bc be 5a 96`. The information we are looking for is in different locations depending on the operating system.

- On a Windows 7 system, you want to examine the contents of the table `SystemIndex_0A`.
- On a Windows 8/10 computer, you want to examine the contents of the table `SystemIndex_PropertyStore`.

Once we input the hexadecimal values into the filter, it reduces the data to a single row:

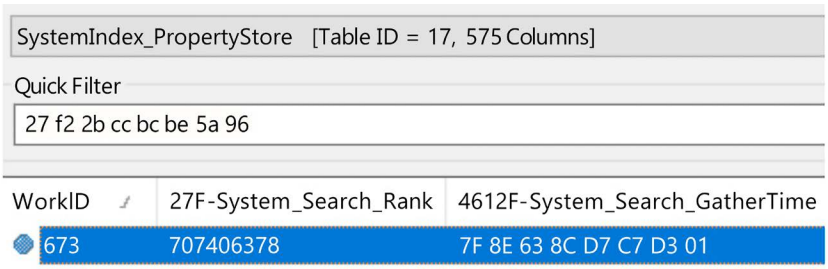


Figure 6.10: Filtered database results

In the following screenshot, we can see that the file came from the desktop of the user `jcloudy`. The name of the image is `MyTiredHead.jpg`:

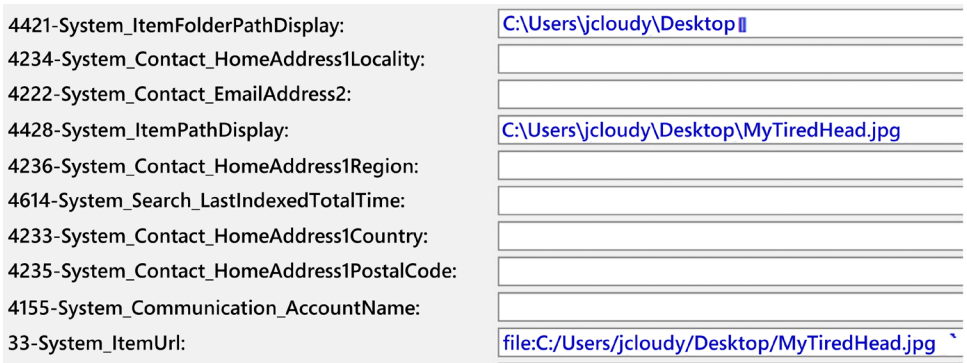


Figure 6.11: Filename display in the database

In the following screenshot, we can verify that this is the correct file when we look in the `System_ThumbnailCacheID` field:

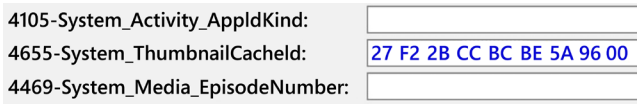


Figure 6.12: Thumbnail name in the database

That completes the discussion on the thumbcache. We will now explore the artifacts created by the Edge/Internet Explorer/File Explorer browsers.

Exploring Microsoft browsers

Microsoft uses the same method to record a user's file activity and internet history with the Internet Explorer/File Explorer/Edge browsers. In addition, it records local and remote file access. Most commercial forensics tools parse these files easily. Depending on the version, the history file will be in the following areas:

```
IE6-7:      %USERPROFILE%\LocalSettings\History\History.IE5

IE8-9:      %USERPROFILE%\AppData\Local\Microsoft\WindowsHistory\History.IE5

IE10-11:    %USERPROFILE%\AppData\Local\Microsoft\Windows\WebCache\WebCacheV*.dat
```

Figure 6.13: IE Locations

In the following screenshot, you can see that the user is using version 10/11 because of the existence of the `WebCacheV01.dat` file:

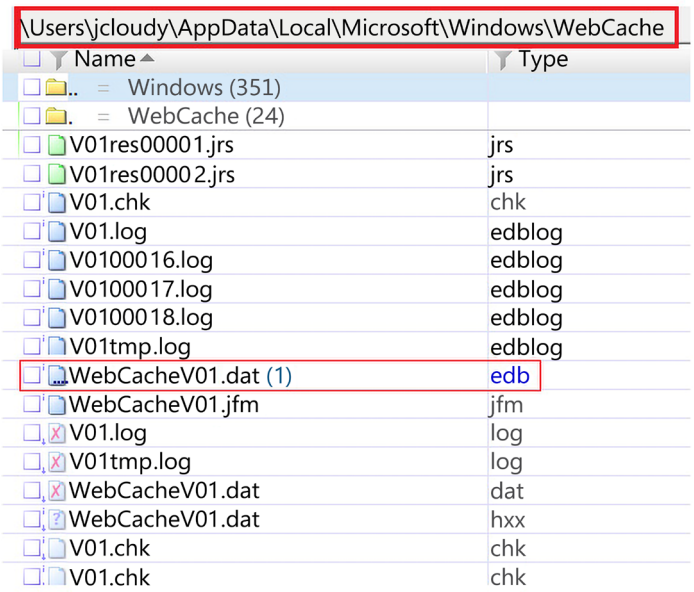


Figure 6.14: File Explorer showing the WebCacheV01.dat file

The `.dat` file is an ESE database. If you want to use a single-use forensic tool, you can export the `.dat` file out of the forensic image and view it with an open-source forensic tool such as `ESEDatabaseView`.

(Located at `https://www.nirsoft.net/utils/ese_database_view.html`)

You will want to navigate to the `Containers` table. The following screenshot is the output from X-Ways Forensics:

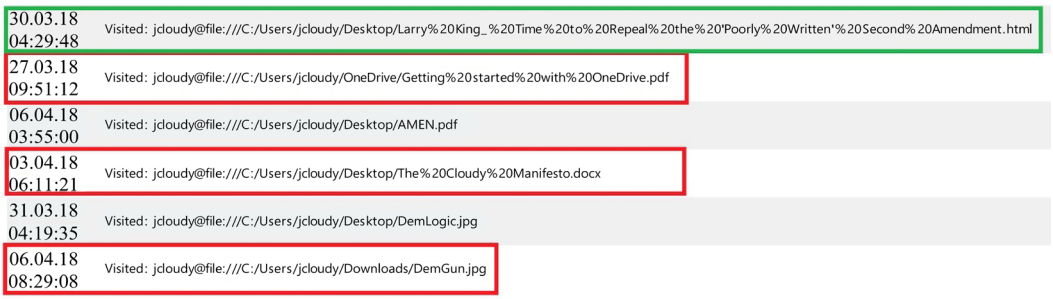


Figure 6.15: X-Ways display of the contents of the WebCache

As you can see, we have a date and timestamp and the file path of the file that was viewed. We have one offline HTML file (the first line), which was located on the user's desktop. We see the user opened two PDF files, two JPEG files, one HTML file, and one DOCX file.

There are additional artifacts that show that a user account accessed a file, which we will discuss next.

Determining most recently used/recently used

An **MRU** (**Most Recently Used**) is a list of recently used files stored in the user's **NTUSER.DAT** hive. When you open an application and see the history list of prior files that the application has used, you are looking at an MRU. There are a lot of MRU lists stored within the registry file. We will go over some more common locations.

OpenSavePidlMRU from the user's **NTUSER.DAT** file tracks the last 20 files opened/saved via the Windows Common Dialogue (these are the commonly encountered **Open/Save As** dialog boxes). In the following example, we can see the last 20 files used by the user:

```
OpenSavePidlMRU\*

LastWrite Time: Fri Apr 6 03:56:31 2018

Note: All value names are listed in MRUListEx order.

My Computer\CLSID_Desktop\LeftUsesBoycotts.pdf
My Computer\CLSID_Desktop\AMEN.pdf
My Computer\CLSID_Desktop\UKknifeBan.pdf
My Computer\CLSID_Desktop\SelfDefenseisMurder.pdf
My Computer\C:\Users\jcloudy\Desktop\Cloudy thoughts (4apr).docx
My Computer\CLSID_Desktop
My Computer\CLSID_Desktop\Operation 2nd Hand Smoke.pptx
My Computer\CLSID_Desktop\The Cloudy Manifesto.docx
My Computer\C:\Users\jcloudy\Desktop\The Cloudy Manifesto.docx
My Computer\CLSID_Desktop\Huckleberry.png
My Computer\CLSID_Desktop\DemLogic.jpg
My Computer\CLSID_Desktop\RedGuns.jpg
```

Figure 6.16: Content of NTUSER.DAT key - OpenSavePidlMRU

Another key to look at is:

```
NTUSER.DAT\Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs
```

This key contains a list of files executed/opened by the user through the Windows Explorer application. You will also have subkeys, based on file extensions, listing those files that were executed/opened. The system will store the entries in chronological order of when the files were executed/opened by the user.

When looking at the last entry/modified time of the key, it will correspond to the last entry in the list. This key will keep track of the previous 150 files that were opened/executed. The following is the output of the key (I am only showing the top-level entries for brevity's sake):

```
recentdocs v.20100405 (NTUSER.DAT)

Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs

LastWrite Time Fri Apr 6 12:27:08 2018 (UTC)

37 = rootkey.csv
36 = Hardware and Sound
10 = DemGun.jpg
34 = LeftUsesBoycotts.pdf
33 = AMEN.pdf
12 = Planning.docx
32 = UKknifeBan.pdf
31 = SelfDefenseisMurder.pdf
30 = Cloudy thoughts (4apr).docx
```

Figure 6.17: Recent Docs entries

This is an example of the file extension subkeys I described earlier, and it shows the recently used CSV files:

```
Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.csv
```

```
LastWrite Time Fri Apr 6 12:27:08 2018 (UTC)

MRUListEx = 0

0 = rootkey.csv
```

Figure 6.18: Content of NTUSER.DAT key - Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs for CSV files

This is an example of the file extension subkeys I described earlier, and it shows the recently used DOCX files:

```
Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.docx
```

```
LastWrite Time Thu Apr 5 08:32:48 2018 (UTC)

MRUListEx = 0,3,1,2

0 = Planning.docx
3 = Cloudy thoughts (4apr).docx
1 = AIRPORT INFORMATION.docx
2 = The Cloudy Manifesto.docx
```

Figure 6.19: Content of NTUSER.DAT key - Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.docx

This is an example of the file extension subkeys I described earlier, and it shows the recently used HTML files:

```
Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.html
```

```
LastWrite Time Fri Mar 30 04:32:26 2018 (UTC)
MRUListEx = 1,0
1 = Cubs' Anthony Rizzo Praises Parkland Kids, Says 'It's too Easy to Get a Gun'.html
0 = Larry King_ Time to Repeal the 'Poorly Written' Second Amendment.html
```

Figure 6.20: Content of NTUSER.DAT key - Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.html

There is also an additional subkey, `\Folder`, that lists when the user opened folders on the system, which is shown as follows:

```
Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\Folder
```

```
LastWrite Time Fri Apr 6 12:27:08 2018 (UTC)
MRUListEx = 4,5,1,3,2,0

4 = Downloads
5 = Hardware and Sound
1 = The Internet
3 = OneDrive
2 = System and Security
0 = CloudLog (D:)
```

Figure 6.21: Content of NTUSER.DAT key - Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\Folders

Entries of potential interest include OneDrive and CloudLog. If I am looking for evidence of specific files, the subject may store the data in cloud storage. When I see artifacts showing the use of cloud storage, it provides additional locations that I will have to locate and acquire the digital evidence to continue my digital forensic investigation.

As you can see, these are great artifacts to see what files were accessed by the user, but what happens when the user deletes a file? That leads us on to our next topic, the Recycle Bin.

Looking into the Recycle Bin

The **Recycle Bin** is Microsoft's effort to protect users from their actions. It provides an intermediary step for when a user deletes a file. Windows will move the file into a holding area known as the **Recycle Bin**.

The **Recycle Bin** is a hidden folder stored in the root directory of every fixed disk on the system. The folder name is `$Recycle.Bin`. On an NTFS formatted disk, there will be sub-folders named with the user's SID. These sub-folders are created whenever a user logs on to the system for the first time:

- `$Recycle.Bin`
    - `S-1-5-21-2734969515-1644526556-1039763013-1001`

When a user deletes a file, the original file gets renamed and becomes part of a set of `Recycle.Bin` files. The system will rename the original file with `$R` and then six random alphanumeric characters for the filename. The file extension will remain the same. The system will create a second file, which will start with `$I` and then have the same six alphanumeric characters that the `$R` file has. The `$I` file will also have the same file extension as the `$R` file.

The `$I` file will track the time of deletion and the path to the original file location:

- **Size**: 4.9 MB
- **Moved to recycle bin**: `04/05/2018 02:20:17 +0 C:\Users\jcloudy\Desktop\Larry King_ Time to Repeal the 'Poorly Written' Second Amendment_files`

As you can see, we have the size of the original file, when the user deleted it, and the original path that includes the filename.

If a user deletes a directory, you will still have the `$R` and `$I` files for the directory. The `$R` file will contain all the subdirectories and all the files with the original names, as shown in the following screenshot:

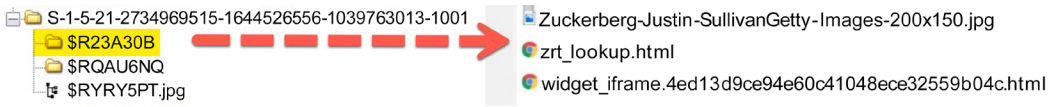


Figure 6.22: Deleted directory

The user can empty the **Recycle Bin**. When that occurs, the filesystem updates that the clusters are now available for use. Until the system overwrites the data, you may recover data from the unallocated clusters. Just be aware that the `$I` (on an NTFS volume) will be resident data in the MFT. NTFS is very efficient in reusing the file entries in the MFT, so it's challenging to recover the information in the `$I` file.

If the **Recycle Bin** is emptied, other artifacts may be referencing the file(s). That brings us to our next topic, link (LNK) files.

Understanding shortcut (LNK) files

A `.lnk` file is used by the Windows operating system as a shortcut or link to files, applications, or resources. It is a simple, easy-to-use method for users to gain access to frequently used documents or applications. The link file will contain useful information for the digital forensic investigator, including the following:

- File MAC times
- File size
- File path
- Volume details

This information will remain even if the destination file has been deleted. The system will create a link file every time the file is double-clicked or when using the **File Open** dialog box. These link files will be stored in the `Recent` folder located at the following path:

```
%Username%\Appdata\Roaming\Microsoft\Windows\
```

Most commercial forensic tools can analyze link files. An open-source option is Eric Zimmerman's LECmd tool (which can be found at `https://ericzimmerman.github.io/`).

When we analyze the contents of the link file, we can see a large amount of information that could be helpful to the digital forensic investigator:

```
Target attributes       A
Target file size        172684
Show Window             SW_NORMAL
Target created          03/30/2018 02:29:57 +0
Last written            04/04/2018 04:59:32 +0
Last accessed           04/04/2018 04:59:32 +0
ID List                 Desktop\AIRPORT INFORMATION.docx
                        C=03/30/2018 02:29:58
                        M=04/04/2018 04:59:34
                        Size=172684
Volume type             Fixed
Volume serial           0xAA920881
Local path              C:\Users\jcloudy\Desktop\AIRPORT INFORMATION.docx
Relative path           ..\..\..\..\..\Desktop\AIRPORT INFORMATION.docx
Working directory       C:\Users\jcloudy\Desktop
Known Folder Tracking   false
Host name               desktop-pm6c56d
Volume ID               {BC7539BE-7B5B-4E04-9F8D-1C0D9B3AFF21}
Object ID               {30D25F11-3208-11E8-9B15-28E347017777}
MAC Address             28 E3 47 01 77 77
Timestamp               03/27/2018 21:45:39 +0, Seq: 6933
PROPERTYSTORAGE         {446D16B1-8DAD-4870-A748-402EA43D788C}
Size                    29
propID                  104
```

Figure 6.23: Link File contents

We can see that the destination file is a Microsoft Word document stored on the user's desktop. When we look at the field ID list, we can also see the file's internal metadata (MAC values). This data can be fundamental when trying to tie knowledge of the file to a specific user. We can also see the date/time when the system created the link file. Additional information is the volume type/serial number and hostname, which allow us to tie this link file to the specific location of the destination file. Be aware that this is an option that the user or systems administrator can turn off. Another artifact similar to LNK files is the JumpList.

Deciphering JumpLists

JumpLists were introduced with Windows 7 and are very similar to the `Recent` folder (which we discussed with LNK files). They allow the user to access frequently used/recently used files from the Windows taskbar. Even if the user clears out the `Recent` folder, it will not clear out the information stored in the JumpLists.

JumpLists can be found at the following paths:

- `%UserProfile%\AppData\Roaming\Microsoft\Windows\Recent\Automaticdestinations`
- `%UserProfile%\AppData\Roaming\Microsoft\Windows\Recent\CustomDestinations`

There are two types of JumpLists:

- Automatic – system-created. Records information about file usage.
- Custom – application-created. Records task-specific information about the application.

In the following screenshot, you can see the `AutomaticDestinations` folder, and inside the folder will be files containing the JumpLists:

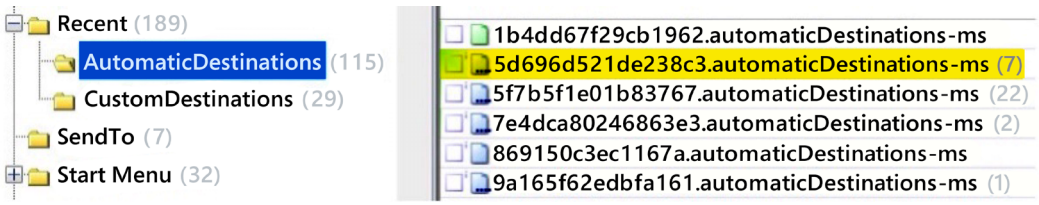


Figure 6.24: JumpList display

The system names the JumpLists based on their JumpLists IDs. For example, in the preceding screenshot, we see `5d696d521de238c3.automaticDestinations-ms`. A search of the JumpLists ID list (which can be found at `https://community.malforensics.com/t/list-of-jump-list-ids/158`) shows that this is the JumpLists ID for the Google Chrome browser.

Most commercial forensic tools will parse out the JumpLists. An open-source option is Eric Zimmerman's JumpList Explorer.

(Located at `https://ericzimmerman.github.io/`)

The following is the information contained in the file. You can see that the user was using Chrome to view PDF files and offline HTML files. It also contains the date/time the user opened the files:

- `7 04/06/2018 03:56:32 +0 C:\Users\jcloudy\Desktop\LeftUsesBoycotts.pdf`
- `6 04/06/2018 03:55:00 +0 C:\Users\jcloudy\Desktop\AMEN.pdf`
- `5 04/05/2018 05:51:41 +0 C:\Users\jcloudy\Desktop\UKknifeBan.pdf`
- `4 04/05/2018 05:48:40 +0 C:\Users\jcloudy\Desktop\SelfDefenseisMurder.pdf`
- `3 03/30/2018 04:32:25 +0 C:\Users\jcloudy\Desktop\Cubs' Anthony Rizzo Praises Parkland Kids, Says 'It's too Easy to Get a Gun'.html`
- `2 03/30/2018 04:29:48 +0 C:\Users\jcloudy\Desktop\Larry King_ Time to Repeal the 'Poorly Written' Second Amendment.html`
- `1 03/27/2018 09:51:18 +0 C:\Users\jcloudy\OneDrive\Getting started with OneDrive.pdf desktop-pm6c56d`

JumpLists are artifacts for files; the next artifact will show which folders the user accessed.

Opening shellbags

Shellbags are a set of registry keys that remember the size and location of the folders and libraries that the user has accessed via the GUI. In addition, you may find artifacts showing user interaction with network devices, removable media, or encrypted containers.

You will find them in a registry hive called `USRCLASS.DAT`, located in the users, `AppData\Local\Microsoft\Windows` folder.

Most commercial forensic tools will parse out the shellbags from the `USRCLASS.DAT` file, but the presentation of the artifact will be different. I like to use Eric Zimmerman's Shellbag Explorer as an open-source alternative.

In the following screenshot, you can see the graphical representation of the folders the user accessed via the Windows GUI. This screenshot is taken from Shellbag Explorer:

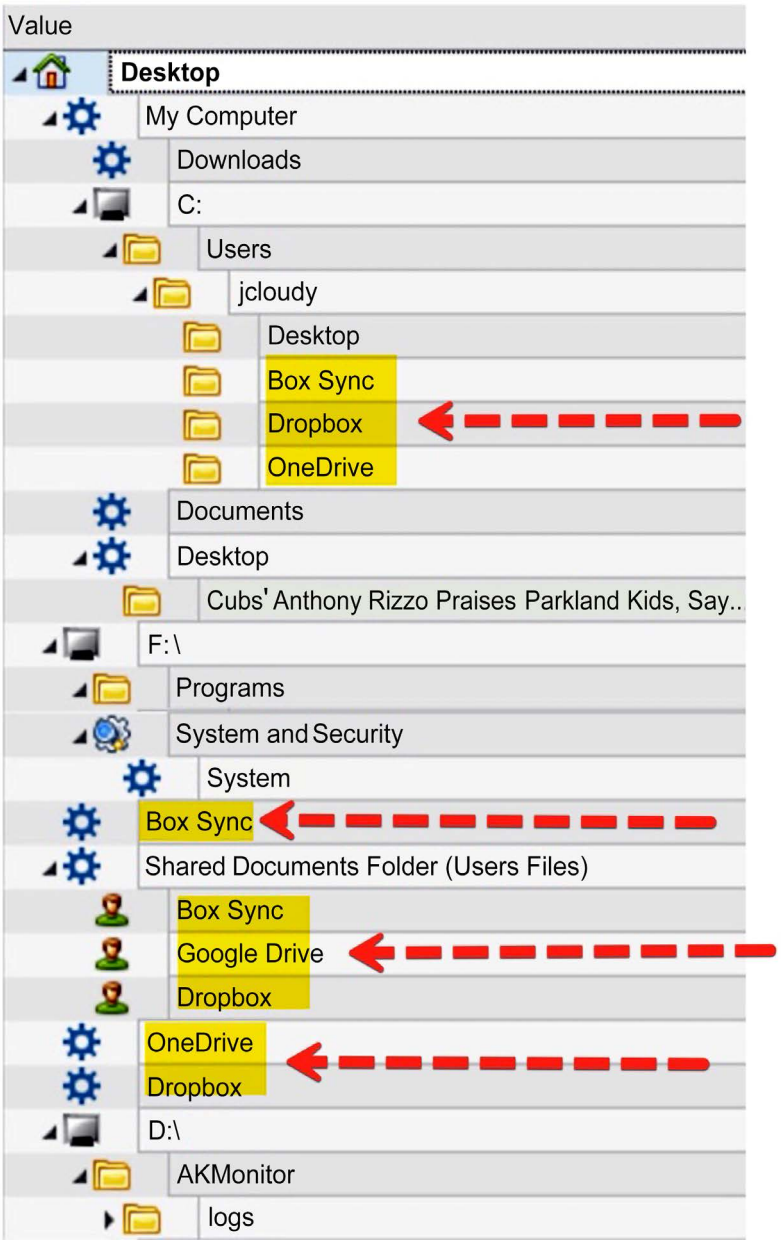


Figure 6.25: Shellbag Explorer: a graphical representation of shellbags

You cannot determine whether the user accessed any files from within the folder through this artifact. What this artifact shows is that the user accessed the folder. As I look at the display, I see that the user was using three cloud storage services. We have seen prior artifacts for Box Sync and Dropbox, but this is the first reference I have seen regarding Google Drive.

In the following output from RegRipper, we can see the access date and timestamps and the date/time of the first access:

```
Name: Google Drive
Absolute path: Desktop\Shared Documents Folder (Users Files)\Google Drive
Key-Value name path: BagMRU\7-1
Registry last write time: 2018-04-05 02:05:13.581

Target timestamps
Created on: 2018-03-28 00:43:24.000
Modified on: 2018-03-28 00:43:24.000
Last accessed on: 2018-03-28 00:43:24.000

Miscellaneous
Shell type: Users Files Folder
Node slot: 14
MRU position: 1
# of child bags: 0

First interacted with: 2018-03-28 00:43:25.373
```

Figure 6.26: RegRipper output Google Drive

This artifact is important if the subject states that they did not know about a file/folder location. This artifact is created by the user's actions. The next artifact can also be used to show user knowledge of a file.

Understanding prefetch

Prefetch is a feature Microsoft introduced to enhance the user experience with the Windows operating system. It allows faster response times by preloading data into the RAM in anticipation of its demand by the user or system. You will find the prefetch files at the following path:

```
%WINDOWS%\PREFETCH
```

The files will have a file extension of `.pf`. In addition, the prefetch file will contain information about the executable file associated with it, such as the list of files used by the executable, the number of times the user ran the executable, and the last date/time when the user ran the executable.

Most commercial forensic tools will parse out the prefetch files. For an open-source option, you can use NirSoft's **WinPrefetchViewtool**.

(Located at `https://www.nirsoft.net/utils/win_prefetch_view.html`)

In the following screenshot, we are looking at the output of `WinPrefetchView`. You can see the date and timestamp and the process path of the executable (be aware that due to the method with which the system monitors the prefetch files, you may have to subtract 10 seconds from the created/modified times to get an accurate time):

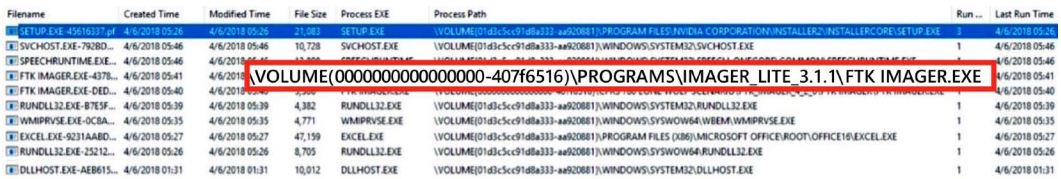


Figure 6.27: Prefetch files displayed by WinPrefetchView

By using these artifacts, you can determine which applications are being used by the user, which may lead to the discovery of hidden partitions, mobile devices, encrypted containers, or cloud storage.

As operating systems change or are updated, the artifacts may move or be removed. You will have to stay current as changes become known. We will now look at artifacts that help us determine the physical location of the system.

Identifying physical locations

Knowing the system's physical location may help you prove or disprove the allegations against the subject you are investigating. For example, there was an investigation into a compromise of the organization's network. A former employee was the suspect in the attack because of their threats during the termination process. When the suspect was interviewed, he denied being in the area and stated he was out of state. A judge authorized a search warrant for the suspect's mobile device and laptop computer. When conducting the forensic analysis of the laptop, the examiner found it to have been recently restored to a new version of the operating system. Artifacts in the unallocated space led us to believe the user had wiped the device. (The user overwrote all available sectors with hexadecimal 00 characters.) The suspect had not tampered with the mobile device, and we could analyze the device. We were able to map out the Wi-Fi hotspots the device had accessed in the immediate neighborhood when the suspect was allegedly out of state. When confronted with the digital evidence, the suspect confessed and admitted he forgot about his mobile device and that it was automatically connecting to Wi-Fi hotspots.

We will now talk about some artifacts you can look at in a system to help determine their physical location at the time of an incident.

Determining time zones

Time zone information on the system allows you to have a starting point with which to correlate activity that is recorded with the date/time that the incident occurred. All the internal dates and timestamps will be based on the time zone information recorded in the registry. We can find the time zone information within the system hive. We can find the key at the following path:

```
SYSTEM\CurrentControlSet\Control\TimeZoneInformation
```

This will give us the following output, courtesy of RegRipper:

```
----------------------------------------
timezone v.20160318
(System) Get TimeZoneInformation key contents

TimeZoneInformation key
ControlSet001\Control\TimeZoneInformation
LastWrite Time Tue Mar 27 09:56:27 2018 (UTC)
  DaylightName   -> @tzres.dll,-111
  StandardName   -> @tzres.dll,-112
  Bias           -> 300 (5 hours)
  ActiveTimeBias -> 240 (4 hours)
  TimeZoneKeyName-> Eastern Standard Time
----------------------------------------
```

Figure 6.28: RegRipper output - SYSTEM\CurrentControlSet\Control\TimeZoneInformation

`Tzres.dll` is the time zone resource `DLL`. You have the fields of `Bias` and `ActiveTimeBias`, which show the values of `300` and `240`, respectively, which is the number of minutes offset from GMT. And then you have the time zone common name, which in this case is `Eastern Standard Time`.

Time zones are not always accurate – the user can set the time zone to the zone of their choice. The next artifact we will examine may help in locating a physical location.

Exploring network history

Knowing which networks, be they wired or wireless, the suspect has connected to might give you location information about their whereabouts at the time in question. You will find the relevant information in the `Software hive` or an XML document managed by the operating system. The Wi-Fi document will be found at the following path:

```
C:\ProgramData\Microsoft\Wlansvc\Profiles\Interfaces
```

This directory contains subfolders (using the GUID naming convention) for each interface. The XML document will contain the **SSID (Service Set Identifier)** of the networks the interface has connected to. The following output is consistent with the information you would find in the XML document:

```
<WLANProfile xmlns='http://www.microsoft.com/networking/WLAN/profile/v1'>
<name>Net 2.4</name><SSIDConfig><SSID><hex>4E657420322E34</hex>
<name>Net 2.4</name><MSM><security><authEncryption><authentication>
WPA2PSK</authentication><encryption>AES</encryption>
```

Figure 6.29: XML Output of WLAN Profile

As you can see, the SSID of the network is `Net 2.4` and it is using `WPA2PSK` authentication.

If you go to the registry location, you will find sub hives that will contain networking information such as the `Profiles` subkey, which gives us additional information about the wireless network(s) the subject connected to:

```
SOFTWARE\Microsoft\Windows NT\CurrentVersion\NetworkList
```

The following is the RegRipper output of the `networklist` sub hive:

```
Launching networklist v.20190128
(Software) Collects network info from Vista+ NetworkList key

Microsoft\Windows NT\CurrentVersion\NetworkList\Profiles
Net 2.4
  DateLastConnected: Fri Mar 30 17:09:01 2018
  DateCreated      : Tue Mar 27 05:15:58 2018
  DefaultGatewayMac: 5C-8F-E0-2A-1C-68
  Type             : wireless
Nla\Wireless
Net 2.4
```

Figure 6.30: RegRipper output - SOFTWARE\Microsoft\Windows NT\CurrentVersion\NetworkList

The registry hive gives us a little bit more information, including the MAC address date and timestamp of when the last connection was made. There is also an additional log file we can examine: the WLAN event log.

Understanding the WLAN event log

Microsoft Windows also keeps an event log of wireless connections. The log can be found at the following path:

```
C:\windows\System32\winevt\Logs\Microsoft-Windows-WLAN-
AutoConfigOperational.evtx
```

This log contains SSID information, MAC addresses, and the date and timestamps of the connection. The following event ID numbers may be pertinent to your investigation:

- `11000` - Wireless network association
- `8001` - Connected to a wireless network
- `8002` - Failed to connect to wireless network
- `8003` - Disconnected from a wireless network
- `6100` - Network diagnostics (System log)

Note

Everything you ever wanted to know about Microsoft Windows can be found at `https://docs.microsoft.com/en-us/`.

The following output is consistent with what you will see in the event log:

```
3/27/2018 12:15:58 +0
Microsoft-Windows-WLAN-AutoConfig
EventID:     11000
Computer:    SYSTEM

Adapter=Broadcom 802.11n Network Adapter DeviceGuid={4B0AE068-B350-4BD4-85AB-77E0E581863}
LocalMac=EC:0E:C4:20:7F:0E
SSID=Net 2.4
BSSType=Infrastructure
Auth=WPA2-Personal Cipher=AES-CCMP OnexEnabled=0
IhvConnectivitySetting= ConnectionId=0000000000000002
```

Figure 6.31: Event log for WIFI access

This is an `11000` event ID, which is the start of a wireless connection. So, based on this specific artifact, you can articulate that a connection was made to the wireless network `Net 2.4` on March 27, 2018, at 12:15:58 (GMT) by the computer `SYSTEM`.

If you know where the wireless network `Net 2.4` is located, you can associate this computer with that physical location.

Next, we will discuss the artifacts that allow us to determine whether the user executed a specific program.

Exploring program execution

Program execution artifacts indicate programs or applications that were run on the system. The user could cause the execution, or an autostart/run event managed by the system. Some categories overlap with the file knowledge category we discussed earlier in the chapter. I am not going to re-examine those specific artifacts in this section. Just be aware that the artifacts from recent apps, JumpLists, an MRU, and prefetch files will also contain information about program/application activity.

Determining UserAssist

`UserAssist` is a registry key in the user's `NTUSER.DAT` file and can be found at the following path:

```
NTUSER.DAT\Software\Microsoft\Windows\Currentversion\Explorer\UserAssist
```

The key tracks the GUI-based applications that were launched in the system. The system encodes the data in the key with ROT 13 encoding. RegRipper will decode the data automatically. The following represents the output you will see from RegRipper:

```
UserAssist
Software\Microsoft\Windows\CurrentVersion\Explorer\UserAssist
LastWrite Time Tue Mar 27 09:19:59 2018 (UTC)

{CEBFF5CD-ACE2-4F4F-9178-9926F41749EA}
Fri Apr  6 12:41:20 2018 Z
  F:\Programs\Imager_Lite_3.1.1\FTK Imager.exe (1)
Fri Apr  6 12:27:04 2018 Z
  {7C5A40EF-A0FB-4BFC-874A-C0F2E0B9FA8E}\Microsoft Office\Root\Office16\EXCEL.EXE (1)
Thu Apr  5 07:02:25 2018 Z
  {7C5A40EF-A0FB-4BFC-874A-C0F2E0B9FA8E}\Microsoft Office\Root\Office16\WINWORD.EXE (1)
Thu Apr  5 06:06:42 2018 Z
  {6D809377-6AF0-444B-8957-A3773F02200E}\S3 Browser\s3browser-win32.exe (4)
Thu Apr  5 02:32:31 2018 Z
  Microsoft.Office.WINWORD.EXE.15 (2)
Thu Apr  5 02:05:01 2018 Z
  {6D809377-6AF0-444B-8957-A3773F02200E}\Box\Box Sync\BoxSync.exe (2)
```

Figure 6.32: Contents of NTUSER.DAT\Software\Microsoft\Windows\Currentversion\Explorer\UserAssist

As shown in the preceding output, you can see the date and timestamp of the last execution and the path of the executable. The number in parentheses at the end indicates the number of times the user/system has activated the executable. Next, we will discuss the Shimcache, which also contains information about executed programs.

Exploring the Shimcache

This is the default location of the Shimcache:

```
SYSTEM\CurrentControlSet\Control\Session Manager\AppCompatCache
```

The Shimcache is used to track compatibility issues with executed programs. Some information that is stored in this cache is as follows:

- File path
- `$Standard` Information Attribute Modify Time
- The update time of the Shimcache

The following represents the output you will see from RegRipper:

```
shimcache v.20190112
(System) Parse file refs from System hive AppCompatCache data

*** ControlSet001 ***
ControlSet001\Control\Session Manager\AppCompatCache
LastWrite Time: Tue Mar 27 21:45:28 2018 Z

C:\Windows\system32\MRT-KB890830.exe  Tue Mar 27 09:38:12 2018 Z
C:\Windows\system32\attrib.exe  Fri Sep 29 13:41:33 2017 Z
C:\Program Files\NVIDIA Corporation\DRS\DBInstaller.exe  Tue Mar 14 14:07:18 2017 Z
C:\Program Files (x86)\Common Files\Microsoft Shared\Source Engine\OSE.EXE  Sat Mar  3 12:03:10 2018 Z
C:\Users\jcloudy\AppData\Local\Microsoft\OneDrive\Update\OneDriveSetup.exe  Tue Mar 27 09:21:57 2018 Z
C:\Windows\system32\OpenWith.exe  Fri Sep 29 13:42:00 2017 Z
C:\Windows\system32\SnippingTool.exe  Fri Sep 29 14:43:00 2017 Z
```

Figure 6.33: Shimcache output

The artifacts found in the Shimcache can provide supporting evidence to the other artifacts found throughout the system, that is, the registry, event logs, filesystem, and so on.

Sometimes, the user will have programs or files contained within a portable device. The next set of artifacts will deal with the use of USB devices.

Understanding USB/attached devices

There are several security risks associated with a USB device. They are small, portable, high-capacity storage devices that can be used to exfiltrate data from an organization, or they can be used to deliver malware to an organization to compromise its security protocols.

As a digital forensic investigator, you will want to know whether there were any USB devices attached to the host you are examining. We will now talk about some Windows system artifacts that will allow you to identify USB device usage on the host.

We will now look at the results for two registry keys. The first key can be found at the following path:

```
SYSTEM\CurrentControlSet\Enum\USB
```

This registry key identifies the USB devices attached to the system, as shown in the following output:

```
usbdevices v.20140416
(System) Parses Enum\USB key for USB & WPD devices

VID_0781&PID_5580
LastWrite: Tue Mar 27 09:22:21 2018
SN : AA010215170355310594
LastWrite: Tue Mar 27 12:13:16 2018

VID_0781&PID_5580
LastWrite: Tue Mar 27 09:22:21 2018
SN : AA010603160707470215
LastWrite: Tue Mar 27 21:45:44 2018
```

Figure 6.34: Content of Registry key - SYSTEM\CurrentControlSet\Enum\USB

We can see there were two USB devices attached to the system at different times. We have different volume serial numbers and the last write times from when the system accessed the devices. The volume serial number found in the registry is not the physical device serial number.

Note

Devices that do not have a unique volume serial number will have an & as the second character of the volume serial number.

The next registry key you want to look at is the following:

```
SYSTEM\CurrentControlSet\Enum\USBSTOR
```

When we look at the values in `USBSTOR`, we get some additional information about the devices, including the commercial name of the device. We also confirm the serial numbers of the devices with these two entries in the `SYSTEM` hive:

```
usbstor v.20141111
(System) Get USBStor key info
USBStor
ControlSet001\Enum\USBStor

Disk&Ven_SanDisk&Prod_Extreme&Rev_0001 [Tue Mar 27 09:22:21 2018]
S/N: AA010215170355310594&0 [Tue Mar 27 12:11:44 2018]
Device Parameters LastWrite: [Tue Mar 27 12:11:42 2018]
Properties LastWrite : [Tue Mar 27 09:16:45 2018]
FriendlyName : SanDisk Extreme USB Device

S/N: AA010603160707470215&0 [Tue Mar 27 09:22:21 2018]
Device Parameters LastWrite: [Tue Mar 27 09:22:21 2018]
Properties LastWrite : [Tue Mar 27 09:23:58 2018]
FriendlyName : SanDisk Extreme USB Device
```

Figure 6.35: Content of Registry key - SYSTEM\CurrentControlSet\Enum\USBSTOR

In the `MountedDevices` key in the `SYSTEM` hive, which can be found in `SYSTEM\MountedDevices`, we can map the USB device(s) via the serial number to a drive letter on the system:

```
mountdev v.20130530

(System) Return contents of System hive MountedDevices key

MountedDevices

LastWrite time = Tue Mar 27 09:22:21 2018Z
Device: _??_USBSTOR#Disk&Ven_SanDisk&Prod_Extreme&Rev_0001#AA010603160707470215&0#
{53f56307-b6bf-11d0-94f2-00a0c91efb8b}\
DosDevices\D:\??\Volume{3869c27a-31b8-11e8-9b12-ecf4bb487fed}

Device: _??_USBSTOR#Disk&Ven_SanDisk&Prod_Extreme&Rev_0001#AA010215170355310594&0#
{53f56307-b6bf-11d0-94f2-00a0c91efb8b}\
DosDevices\E:\??\Volume{5c3108bf-31c0-11e8-9b10-806e6f6e6963}
```

Figure 6.36: Content of Registry key - SYSTEM\MountedDevices

When we analyze the data, we can see that two USB devices (serial numbers `AA010215170355310594` and `AA010603160707470215`) were connected to the system. One was recognized as the `D:` drive and the second device was recognized as the `E:` drive.

Does the question remain as to which user account was responsible for the USB device usage? To determine the answer to that question, we would have to take the GUID from each of the USB devices and compare them to the user's `NTUSER.DAT` file. The GUIDs we are searching for are `3869c27a-31b8-11e8-9b12-ecf4bb487fed` and `5c3108bb-31c0-11e8-9b10-806e6f6e6963`.

RegRipper will also analyze the `NTUSER.DAT` file and give us the information about the devices that were used and associated with the user's account:

```
mp2 v.20120330
(NTUSER.DAT) Gets user's MountPoints2 key contents

MountPoints2
Software\Microsoft\Windows\CurrentVersion\Explorer\MountPoints2
LastWrite Time Fri Apr  6 12:35:08 2018 (UTC)

Remote Drives:

Volumes:
Fri Apr  6 12:35:08 2018 (UTC)
  {76d45981-0000-0000-0000-100000000000}
Tue Mar 27 21:45:54 2018 (UTC)
  {3869c27a-31b8-11e8-9b12-ecf4bb487fed}
Tue Mar 27 09:32:09 2018 (UTC)
  {09931f21-7faf-44a9-81d8-1e73c14b9eaf}
  {5c3108bb-31c0-11e8-9b10-806e6f6e6963}
```

Figure 6.37: Content of Registry key - Software\Microsoft\Windows\CurrentVersion\Explorer\MountPoints2

As you can see, we find both GUIDs in the registry entry, which shows when the devices were last mounted. So we can now say that the user used a specific USB device on the system while the `jcloudy` account was logged in.

Summary

In this chapter, we have discussed how to locate artifacts on a Microsoft Windows-based operating system to determine the subject's culpability in the matter being investigated. You have learned about the different categories of artifacts and what actions of the user/system they represent. Using the knowledge you have gained from this chapter will allow you to quickly determine which accounts were active during the timeframe you are investigating and whether the incident involved a removable storage device. In addition, you have learned about the artifacts to analyze in determining whether a user had knowledge of a specific file or application. Finally, we have used several commercial and open-source forensic tools to access the artifacts. As a result, you now know how to find and analyze digital evidence found on a Microsoft Windows-based operating system.

The next chapter will deal with memory forensics.

Questions

1. Where would you find the registry files?
    a. `%SystemRoot%\System32\Config`
    b. `%SystemRoot%\System32`
    c. `%SystemRoot%\\Config\System32`
    d. `%SystemRoot%\System64\Config`
2. When examining log files, which event ID identifies a successful logon?
    a. 4624
    b. 4625
    c. 4672
    d. 4642
3. A thumbcache is a _____.
    a. Database of toenail images
    b. Database of thumbnail images
    c. Database of deleted thumbnail images
    d. Database of deleted images
4. The user can use Internet Explorer/Edge to view files.
    a. True
    b. False
5. Which of the following will you find in a link (LNK) file?
    a. Volume serial number
    b. Router name
    c. Date of deletion
    d. Volume details
6. Which of the following Microsoft Windows operating systems uses JumpLists?
    a. Windows 98
    b. Windows ME
    c. Windows 7
    d. Windows 2000

7. In which registry hive would we find artifacts relating to USB devices?

    a. NT USER.DAT
    b. SYSTEM
    c. SOFTWARE
    d. SECURITY

The answers can be found at the rear of the book under *Assessments*.

Further reading

Refer to the following links for more information on topics covered in this chapter:

- Altheide, C., Carvey, H. A., and Davidson, R. (2011). *Digital Forensics with Open Source Tools*. Amsterdam: Elsevier/Syngress (available at `https://www.amazon.com/Digital-Forensics-Open-Source-Tools/dp/1597495867`)
- Carvey, H. A. (2005). *Windows forensics and incident recovery*. Boston: Addison-Wesley (available at `https://www.amazon.com/Windows-Forensics-Incident-Recovery-Harlan/dp/0321200985`)
- Bunting, S. (2012). *EnCase Computer Forensics: The Official EnCE: EnCase Certified Examiner; Study Guide*. Indianapolis, IN: Wiley (available at `https://www.amazon.com/EnCase-Computer-Forensics-Official-EnCE/dp/0470901063`)

Exercise

Data set

`Chapter 6 Owl Exercise.e01`

Software needed

Autopsy - `https://www.autopsy.com/`

Scenario

In a jurisdiction where owls are illegal to trade and buy, two users are discussing the illegal trade of owls. A computer is taken into evidence belonging to a user who is attempting to purchase owls illegally. It has been requested that you conduct an analysis of the digital evidence. A forensic image has been obtained and is ready for you. You may use Autopsy or any other tool.

Some artifacts you may want to look for include:

- Web searches
- Shopping searches
- Chat clients
- Email
- Documents
- Social networks
- OS artifacts
- LNK files
- Recycle Bin
- Shellbag

Potential keywords:

- Owl
- Owlet
- Feathers
- Eggs
- Crossbreeding
- Nocturnal
- Nest
- Hoot
- Conservation
- Wingspan

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

7

RAM Memory Forensic Analysis

RAM is a vital source of digital evidence that has been neglected and ignored historically. As our knowledge of digital evidence grew, examiners realized the source of potential digital evidence that existed in RAM. Ultimately, you have an additional multi-gigabyte source of information that needs to be examined and may contain digital artifacts that do not exist in the traditional locations of the system.

In this chapter, we will cover the fundamentals of memory. We will then look at the different sources of memory and learn to capture RAM using RAM capture tools. By the end of this chapter, you will understand the various methods and tools that can process volatile memory.

We'll be covering the following topics in this chapter:

- Fundamentals of memory
- Random access memory
- Identifying sources of memory
- Capturing RAM
- Exploring RAM analyzing tools

Fundamentals of memory

What information does **random access memory** (**RAM**) contain? It will give you information about the current running state of the system before you shut it down. It will contain information about any running programs; these could be legitimate processes, and it could contain running malware processes as well. If attackers have compromised the host, the malware may be a resident in the RAM.

You will also find information related to the host's network connections with other peers. This could be a legitimate use of peer-to-peer file sharing, or it could show a link to the attacker's host. These connections are breadcrumbs for you to follow. The user could also be sharing illicit images. Again, the connection to other computers will allow you to follow and investigate additional users for the same crime.

If the user is using cloud services, we may never find the data they are creating on the physical disk in the system. Instead, we may only see the evidence of the data being hosted in the cloud in the form of RAM.

RAM is the kitchen table of the computer system. Any action the user/system takes will access the RAM. For example, every mouse click and every keyboard button that's pushed will be processed through the RAM, and you can recover entire files, passwords, and the text that was placed into the clipboard. All of these are potential sources of digital evidence. Sometimes, you can recover the encryption keys for closed encrypted containers that the user has created.

In 2004, Rajib K. Mitra was convicted of jamming police radios. The investigation resulted in the seizure of multiple pieces of digital evidence. The lead detective, Cindy Murphy, learned in 2009 that it was possible to recover encryption keys that may have existed only in RAM. Detective Murphy was able to go back and reexamine the evidence and was able to identify the encryption keys Mitra had used to secure his encrypted container. When Detective Murphy opened the encrypted container, she found many illicit images, which led to Mitra being convicted of possessing the images.

How is analyzing RAM different from analyzing a hard drive? RAM is a snapshot of a live running system, whereas a hard drive examination is static. When examining a hard drive, we have shut the system down and we are examining data on the physical device. RAM is much more transient, and if you were to take a forensic image of RAM at two different points of time, you would get different results. Capturing the data in RAM will lead to the loss of potential evidence. You are changing evidence when you collect RAM.

So, let's talk about what RAM is.

Random access memory?

RAM temporarily stores working data/code on an active computer system. Unlike on traditional storage devices such as hard drives, data can be read/written on RAM at extremely fast speeds. Current technology allows the RAM chips to be created around an integrated circuit chip with metal oxide semiconductor cells. The data stored within the RAM chips is volatile.

We lose volatile data when the computer system is no longer powered on. This is a significant reason the *pull the plug* tactic is no longer recommended when responding to a scene involving activated computer systems.

You may run into two different types of RAM: **static RAM (SRAM)** and **dynamic RAM (DRAM)**. SRAM is considered faster and more efficient in terms of energy use, whereas DRAM is cheaper to produce than SRAM. Therefore, you will typically find SRAM used as cache memory for the CPU and DRAM chips used as memory chips for the computer system.

The following is a representation of a DRAM chip you may come across in your investigations:

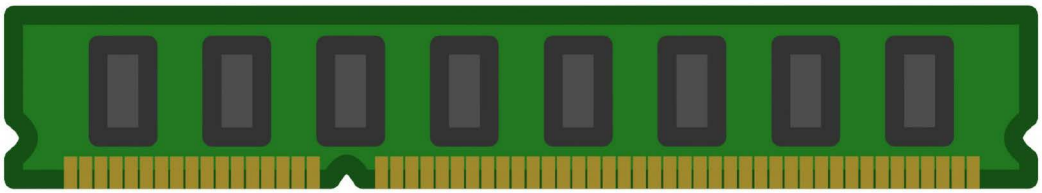


Figure 7.1: DRAM image

Do not confuse RAM with **read-only memory (ROM)**. ROM permanently stores data within the memory chips and is not volatile.

Consider the following: a 32-bit Microsoft Windows-based computer system has a limitation of 4 GB of RAM, while a 64-bit Microsoft Windows-based computer system has a limit of 128 GB of RAM. That is a considerable amount of potential evidence that, historically, has not been analyzed.

For the CPU to access the data/execute code being stored in the memory chips, there must be a unique identifier for the location of that data. When we start examining raw memory dumps, we will be dealing with the physical address, which is an offset of the memory dump.

Data stored in RAM is stored in pages that are 4 kilobytes in size (but can vary in size), and as the system processes add/read data to the pages in RAM, they are utilizing virtual addressing.

All the operating systems access RAM in the same general manner. So let's talk about some concepts that are common to an operating system:

- **Privilege separation**: Privilege determines what a user, user account, and the process are allowed to access. It is a form of access control, and when used by the operating system, it helps provide system stability by isolating users and the CPU kernel's actions. The operating system operates in trusted mode, that is, kernel mode, while the user applications operate in untrusted mode, that is, user mode, when executing commands in the system.

- **System calls**: To access resources controlled by the operating system's kernel, the user application must request access. This is done through a system call to the kernel. It is a bridge between the application and the operating system to allow the untrusted mode to become trusted for a specific instance.
- **Process management**: Program code is executed in memory. The operating system is responsible for managing the processes. Current operating systems operate as multi-programming systems, allowing multiple processes to be executed simultaneously. As we analyze the memory dumps, we are looking at which processes were being executed at the time of capture and analyzing the data stored within RAM.
- **Threads**: A process can have multiple threads. It is the basic unit of using the system's resources, such as the CPU. When we analyze the memory dumps, we are looking for the processes timestamps and starting addresses, which will help identify the code in the process.

The contents of RAM may include artifacts of what is or has occurred on the system. This can include the following:

- Configuration information
- Typed commands
- Passwords
- Encryption keys
- Unencrypted data
- IP addresses
- Internet history
- Chat conversations
- Emails
- Malware

As you can tell, with the collection of RAM, there is great potential to acquire significant evidence. Where do we find data that is stored in RAM? There are several different sources, all of which we will discuss next.

Identifying sources of memory

What happens if you are not the investigator on the scene when the digital evidence is collected in the RAM, and they do not collect volatile data? Is it possible to still access the RAM, despite having the system shut down? While you cannot analyze the RAM, it is possible to examine other sources containing the same data stored in the RAM. This option may not always be viable, depending on the specific set of circumstances surrounding the seizure of the digital evidence.

You need to know that there are potential additional sources containing the same or similar data in RAM. They are as follows:

- **Hibernation file (hiberfil.sys)**: Hibernation is the process of powering down the computer while still maintaining the current state of the system. In Windows, the RAM is compressed and stored in a `hiberfil.sys` file. This will allow the system to power down completely, but when the system is reactivated, the contents of the `hiberfil.sys` file will be placed back into RAM.

Note

If you are examining a laptop, hibernation is usually initiated by closing the laptop. In a desktop, this will be user-initiated. The file header for the `hiberfil.sys` file can be `hibr`, `HIBR`, `wake`, or `WAKE`. When the system is repowered, the header of the file is zeroed out. The `hiberfil.sys` file is a compressed file and will have to be decompressed before you can analyze it.

When analyzing the `hiberfil.sys` file, the last modification date/timestamp will show when the contents of RAM was added to the file.

Note

Another option if you are on scene and cannot do a live capture of the RAM is to place the system into hibernation, which will then create the `hiberfil.sys` file where the current state of the system is saved.

- **Pagefile (pagefile.sys)**: Paging is a method of storing/retrieving data used in the RAM chips with a virtual memory file stored on a traditional storage device. While not as fast as using RAM alone, it allows programs to exceed the physical memory capacity. When using paging, the system will transfer data in pages. The data stored in the page file is typically the least requested data used in memory. When the requests for that data are processed, it places the data back into the physical memory.

Note

In the Windows operating system, the paging file, `pagefile.sys`, is stored at the root of the operating system volume. Be aware that the user can change this location. Typically, the page file can be one to three times larger than the amount of physical memory on the system.

- **Swapfile (Swapfile.sys)**: With Microsoft Windows 8, Microsoft introduced the `Swapfile.sys` file. It is similar to the page file we just discussed but with some differences. The Swapfile was created so that the operating system can use it for paging operations with suspended modern Windows applications. When the application is suspended, the system will write the application data in its entirety into the swapfile. This frees up space in the physical memory, and when the application is resumed, it moves the data back into physical memory.
- **Crash dump (memory.dmp)**: If you have used any version of Microsoft Windows, you might have experienced a system crash or a **blue screen of death (BSoD)**. When that occurs, it may create a dump of memory to store information about the system's state at the time of the crash.

Depending on the settings, you may get one of the following:

- **Complete memory dump**: The data contained within the physical memory. (Not very common because of issues with the capacity of the physical memory chips.)
- **Kernel memory dump**: Will only contain pages of data that were in kernel mode.
- **Small dump files**: Contains information about running processes/loaded drivers at the time of the crash.

The `SYSTEM` hive will contain the key to determine which memory dumps exist on the system you are examining. The key you'll want to explore is as follows:

```
SYSTEM\CurrentControlSet\Control\CrashControl\CrashDumpEnable
```

The dump files will be in a proprietary format and will need a third-party tool to convert them (available at `https://www.comae.com`). So far, we have discussed the locations that will provide sources of RAM. Ultimately, you will want to capture the data within the RAM chips, which is our next topic.

Capturing RAM

When the decision is made to capture the RAM from the system, several factors need to be considered before moving forward. The most significant issue is that you will be changing the system's state when you collect the volatile data.

The **Scientific Workgroup on Digital Evidence** (**SWGDE**) has explored the collection of volatile data and offers the following considerations:

- The application used to collect the data in memory will overwrite some memory contents.
- The larger the tool and associated files are, the more data it overwrites.
- The system may load the USB device driver into memory.
- The system may load the USB device driver into the registry.
- The application used to collect the data in memory will show up in some **Most Recently Used** (**MRUs**).

There is the potential that the collection of RAM may cause a system lockup or instability in the system. Therefore, the digital forensic investigator must know how the tool may affect different operating systems.

After calculating the risk versus the reward, you have decided to go forward and collect the contents of the RAM. What do you need to accomplish this task? It would be best to decide which tool works best in the environment in which you will create the memory dump. One consideration regarding your tool selection is how big a footprint the tool will leave on the system.

Preparing the capturing device

To successfully image the RAM, you will need three things:

- A capturing device (such as a USB device)
- Access to the system
- Administrator privileges

Note

Remember that the amount of RAM installed on the system will dictate the size of your external storage device. If the system has 16 GB of RAM, your external storage device will need to be greater than 16 GB. The memory dump will be the same size as the amount of installed RAM.

You will want to prepare your external storage device before responding to the scene. Your device should be formatted as an NTFS partition. This will alleviate any file size issues you might encounter if you formatted the device in FAT 32.

We will now discuss some tools to create a raw forensic image of the RAM.

Exploring RAM capture tools

I will briefly discuss some tools that you can use to capture RAM. There are additional commercial and open-source tools available. We could write an entire book (and there are some) about some tools used for memory forensics. The goal here is to give you an overview and the skills necessary to accomplish a successful memory dump, but be aware that you can go into much greater detail than I will go into in this chapter.

The following tools are all open source and freely available.

Using DumpIt

DumpIt (available at `https://github.com/thimbleweed/All-In-USB/tree/master/utilities/DumpIt`) was originally developed by MoonSols. Then Comae was maintaining the project. As of 2021, DumpIt appears to have been rolled into the Comae Platform V2.0 Closed Beta. It is a combination of `Win32dd` and `Win64dd` in one executable. The end user isn't given any options to configure. This tool is fast, small, and portable. It leaves the least significant form of footprint on the RAM.

DumpIt is the simplest of all the tools to use. Once you have created your external device and have responded to the scene, you need to follow these steps:

1. Insert your thumb drive into the target host.
2. Type `cmd` (as shown in the following screenshot):

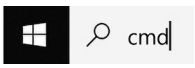


Figure 7.2: Search bar

3. Right-click on **Command Prompt** so that you can run it as an administrator (as shown in the following screenshot):

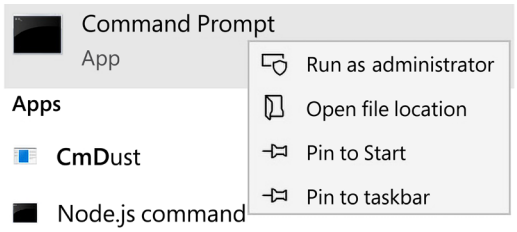


Figure 7.3: Run as administrator

4. Once Command Prompt comes up, navigate to the folder on your USB device that contains the executable. You will then type in the `cmd` command and execute it.
5. The system will then present you with a screen showing the amount of physical memory and the amount of space on the device. It will then ask you if you want to continue. Select `y`, as shown in the following screenshot:

```
DumpIt - v1.3.2.20110401 - One click memory memory dumper
Copyright (c) 2007 - 2011, Matthieu Suiche <http://www.msuiche.net>
Copyright (c) 2010 - 2011, MoonSols <http://www.moonsols.com>


  Address space size:       35945185280 bytes (  34280 Mb)
  Free space size:         369265610752 bytes ( 352159 Mb)

   * Destination = \??\C:\tools\MSI-20220214-221822.raw

   --> Are you sure you want to continue? [y/n]
```

Figure 7.4: DumpIt screen

6. The amount of RAM installed will dictate the amount of time it will take to create the dump of the RAM. Once the process has been completed, the program will notify you that it was successful:

```
* Destination = \??\C:\tools\MSI-20220214-221822.raw

--> Are you sure you want to continue? [y/n] y
+ Processing... Success.
```

Figure 7.5: DumpIt successful

DumpIt is not the only tool available. There are additional open-source alternatives, such as FTK Imager, which we will discuss next.

Using FTK Imager

FTK Imager Lite (available at `http://accessdata.com`) is a GUI-based utility that allows a user to dump the memory of a computer system running either a Windows 32-bit or 64-bit operating system. This tool is easy to use and deployable on a thumb drive. This tool also allows us to mount binary dump files for viewing. Since it is GUI-based, it leaves a significant footprint on the RAM.

FTK Imager is also relatively easy to use. Remember that it is GUI-based, so as you launch the executable from your external storage device, it will overwrite more data in memory than a CLI-based tool.

Once you have responded to the scene, you will need to do the following:

1. Insert your thumb drive into the target host.
2. If **File Explorer** does not automatically launch, use the *Windows* + *E* keyboard shortcut to open it.
3. Launch FTK Imager, left-click on **File**, and select **Capture Memory...**, as shown in the following screenshot:

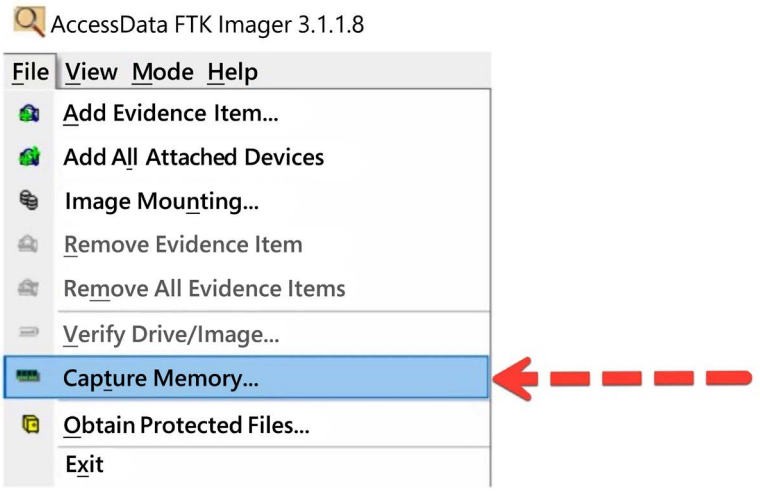


Figure 7.6: FTK Imager menu

4. The **Memory Capture** window will appear as shown in the following screenshot. Here, you can fill in the destination path:

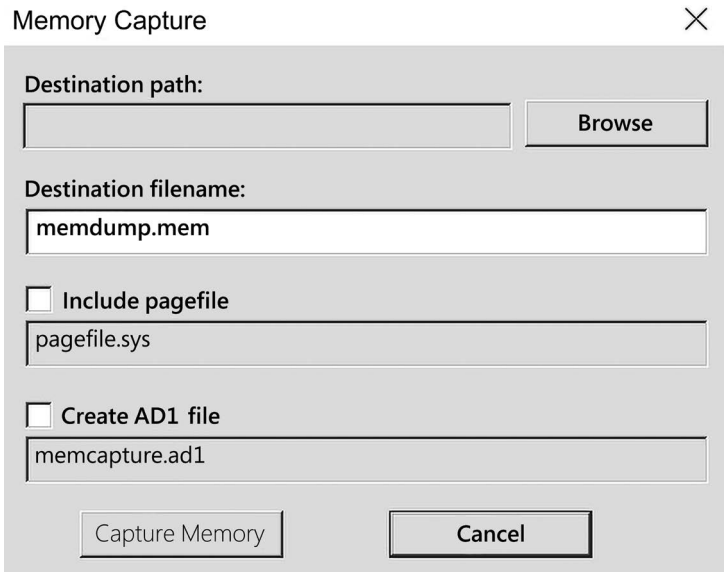


Figure 7.7: FTK Imager memory capture

However, ensure that you select your external storage device.

5. You also have the option to choose the `pagefile`. There is no reason not to. Check that box and then left-click on **Capture Memory**.
6. Once the tool has finished, you will receive a success notification, as shown in the following screenshot. This will store the memory file on your external storage device:

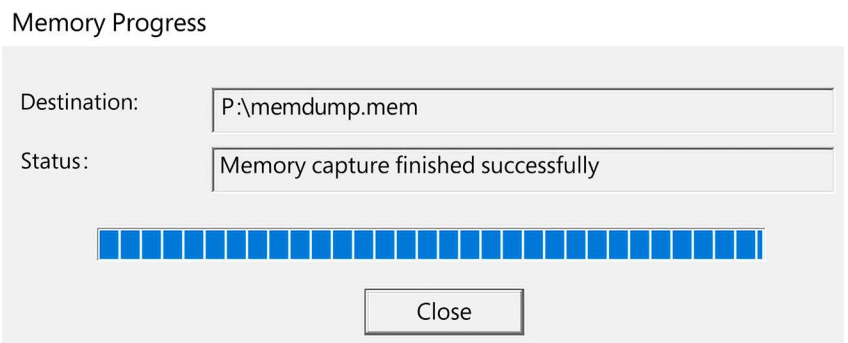


Figure 7.8: FTK Image successful

No matter which tool you used to collect the memory, once you have collected it, you need to get a hash value of the file you just created. You do not want to use the suspect's system because any commands you issue will change the state of the evidence. You will want to use your forensic laptop or your forensic workstation at your laboratory to generate the hash value.

Now that you have created a memory dump of the RAM, what tools will you use to analyze it? Let's talk about some tools that can be used to analyze RAM.

Exploring RAM analyzing tools

Just like when we analyze forensic images created from traditional storage devices, you have the choice of open source or commercial software. It comes down to the examiner's preferences (and sometimes budget) on what tool they wish to use to analyze the dataset. We will go over some available tools, but this is not an all-inclusive list. Most commercial tools will analyze a memory file; we will discuss some open-source options that are available here:

- Bulk Extractor: Bulk Extractor (available at `http://digitalcorpora.org/downloads/bulk_extractor`) scans the target media (disk image, file, directory) and extracts what it believes to be useful information. It ignores the filesystem structure, which allows it to process different parts of the source dataset in parallel. This makes it very fast compared to traditional forensic tools. As Bulk Extractor finds data it believes to be relevant, it creates a histogram of the artifacts.
- Volatility: Volatility (available at `https://www.volatilityfoundation.org`) is an open-source framework for incident response and malware analysis. Volatility supports a wide variety of memory dumps from multiple operating systems. Volatility is very powerful and has numerous plugins.
- VOLIX II v2: VOLIX II (available at `https://www.fh-aachen.de/en/people/schuba/forschung/it-forensik/projekte/volix-en`) is a GUI frontend for Volatility. It allows you to combine commands to enhance usability and speed. It saves you the effort of working with a CLI and enables you to point and click to achieve the same result.

We will discuss the use of some of these open-source options.

Using Bulk Extractor

Let's look at how Bulk Extractor works:

1. Bulk Extractor's documentation lists the following information about its output:

alerts.txt	Processing errors recorded in a text file.
ccn.txt	Processes credit card numbers recorded in a text file.
ccn_track2.txt	Processes credit card "track 2" information, which has been found in some bank card fraud cases recorded in a text file.
domain.txt	Processes Internet domains found on the drive, including dotted-quad addresses found in the text recorded in a text file.
email.txt	Processes email addresses recorded in a text file.
ether.txt	Processes Ethernet MAC addresses found through IP packet carving of swap files and compressed system hibernation files and file fragments recorded in a text file.
exif.txt	Processes EXIFs from JPEGs and video segments. This feature file contains all the EXIF fields, expanded as XML records recorded in a text file.
find.txt	Processes the results of specific regular expression search requests recorded in a text file.
identified_blocks.txt	Processes block hash values that match hash values in a hash database that the scan was run against recorded in a text file.
ip.txt	Processes IP addresses found through IP packet carving recorded in a text file.
rfc822.txt	Processes email message headers including the Date, Subject, and Message- ID: fields recorded in a text file.
tcp.txt	Processes TCP flow information found through IP packet carving recorded in a text file.
telephone.txt	Processes US and international telephone numbers recorded in a text file.
url.txt	Processes URLs, typically found in browser caches, email messages, and pre-compiled into executables recorded in a text file.
url_searches.txt	Processes a histogram of terms used in internet searches from services such as Google, Bing, Yahoo, and others recorded in a text file.
url_services.txt	Processes a histogram of the domain name portion of all the URLs found on the media recorded in a text file.
wordlist.txt	Processes a list of all "words" extracted from the disk, useful for password cracking recorded in a text file.
wordlist_*.text	Processes the wordlist with duplicates, removed, formatted in a form that can be easily imported into a popular password-cracking program recorded in a text file.
zip.txt	Processes information regarding every ZIP file component found on the media. This is exceptionally useful as ZIP files include internal structure and ZIP is increasingly the compound file format of choice for a variety of products such as Microsoft Office recorded in a text file.

Figure 7.9: Bulk Extractor output options

2. You will need to left-click on **Tools** and select **Run bulk_extractor...** to start analyzing your memory dump, as shown in the following screenshot. When you run Bulk Extractor, the viewer will present itself:

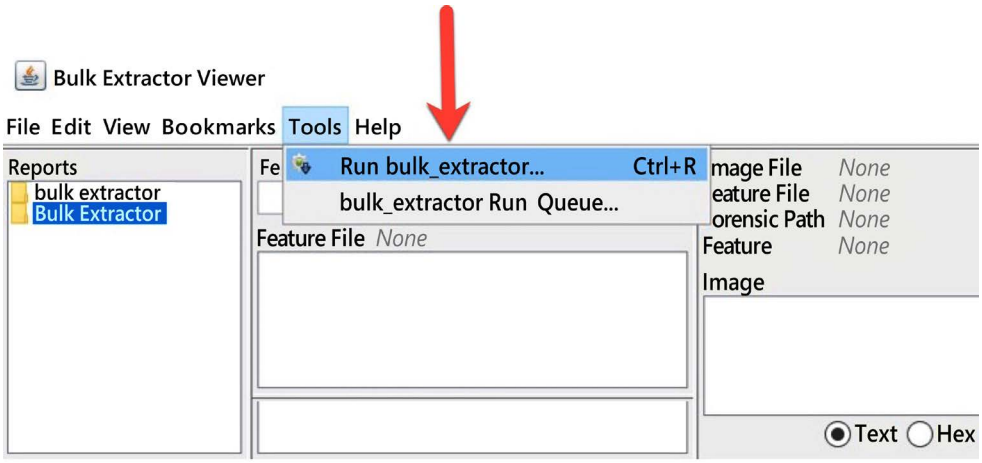


Figure 7.10: Bulk Extractor menu – the run bulk extractor option

It will then present you with the **Run bulk_extractor** menu.

3. Direct the tool to the location of your image file and the output directory you wish to use. As shown in the following screenshot, you can see the numerous scanners that the Bulk Extractor tool uses to look for artifacts within the memory file:

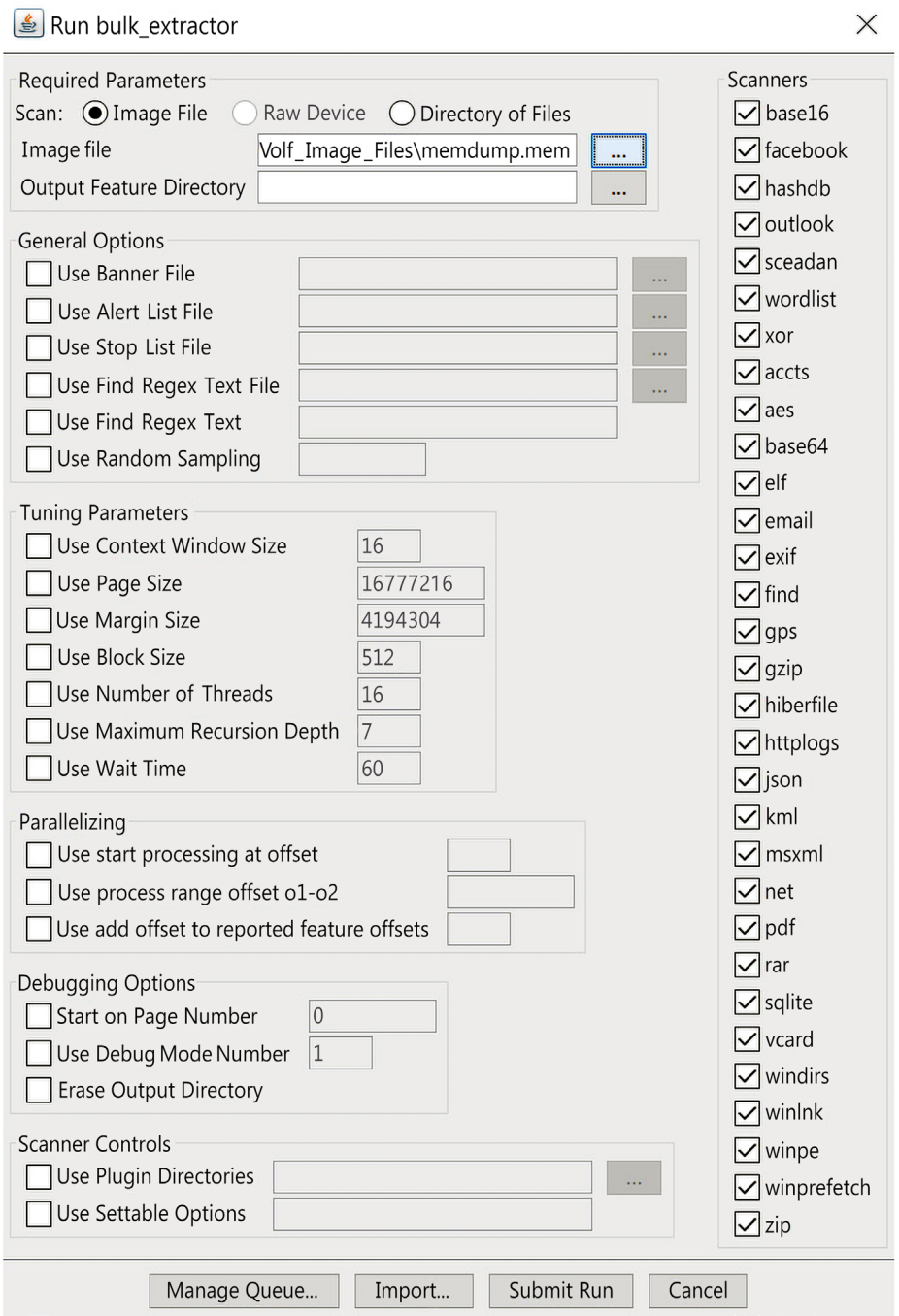


Figure 7.11: Bulk Extractor menu options to run

You can check or uncheck a given specific artifact search as your needs dictate.

4. Once you are satisfied with the setup, left-click on the **Submit Run** button to start the extraction process. Once the extraction has started, it will present you with the extraction window, as shown in the following screenshot:

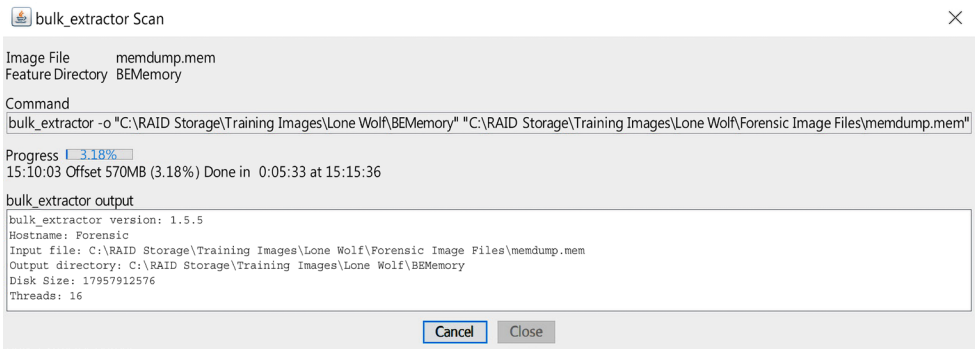


Figure 7.12: Bulk Extraction window

5. Left-click on the **Close** button to go back to the viewer. The following screenshot shows the **Bulk Extractor Viewer**:

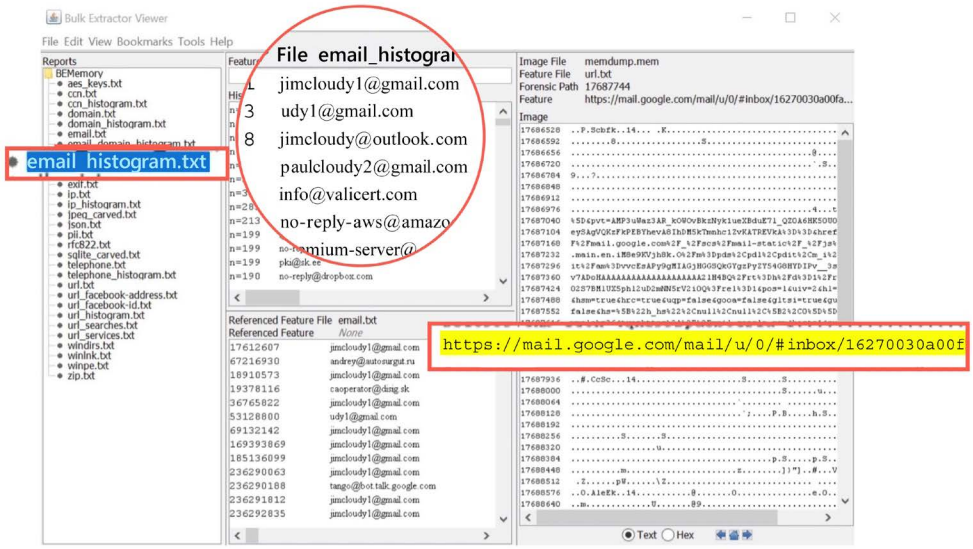


Figure 7.13: Bulk Extractor Viewer – extracted content

On the left-hand side, we can see the specific artifacts recovered by the tool. In the preceding screenshot, I selected the `email_histogram.txt` file, which gives us a list of times it found a particular email address. By looking at the histogram window, I can see that it found the `jcloudy1` email address over 8,000 times. As you go through the email list, you may find emails of evidentiary interest to follow up while using traditional media.

Bulk Extractor is a quick and efficient tool used to extract data strings that you can follow up in your investigation. The following tool we will discuss is Volix II.

Using VOLIX II

Volix is a GUI frontend for the Volatility framework. It makes it a bit easier for those who are not comfortable using a **command-line interface (CLI)**. Once the program has been downloaded and you start it for the first time, it will present you with the following screen:

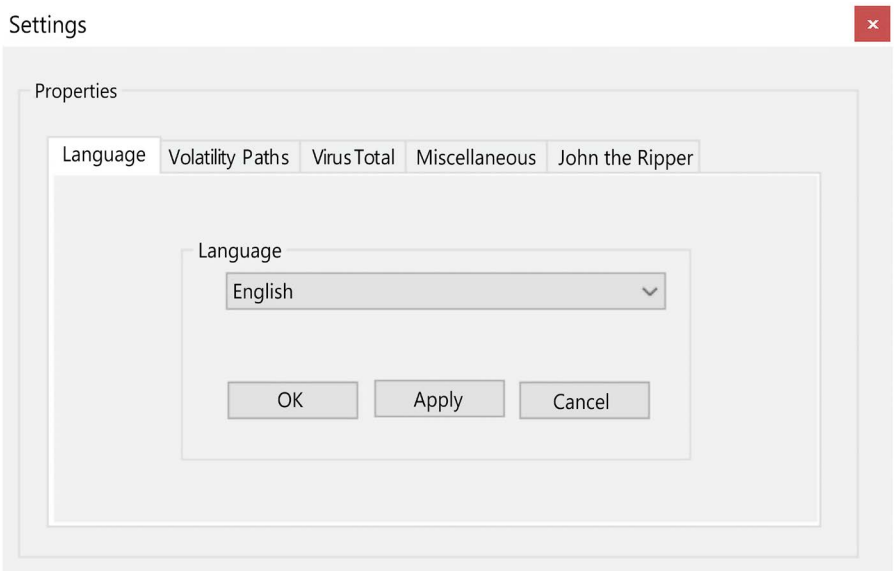


Figure 7.14: Volix settings

Then, you will be pointed to the location of the Volatility framework. You can use either the standalone executable or the binary files to run the Python scripts. Here, I have already downloaded the standalone executable and have pointed Volix to it.

The other options you can select include the language you wish to use Volix in. If you have a VirusTotal API key, you can insert it on that page. This will compare the data captured from RAM and see if it matches any malware being tracked by VirusTotal.

You also have the option of pointing Volix to the John the Ripper executable. If you want to decode/decrypt potential passwords that may be stored in the RAM, follow these steps:

1. Once you've selected **Case**, choose **New**. It will ask you the location of the memory file you wish to analyze, as shown in the following screenshot:

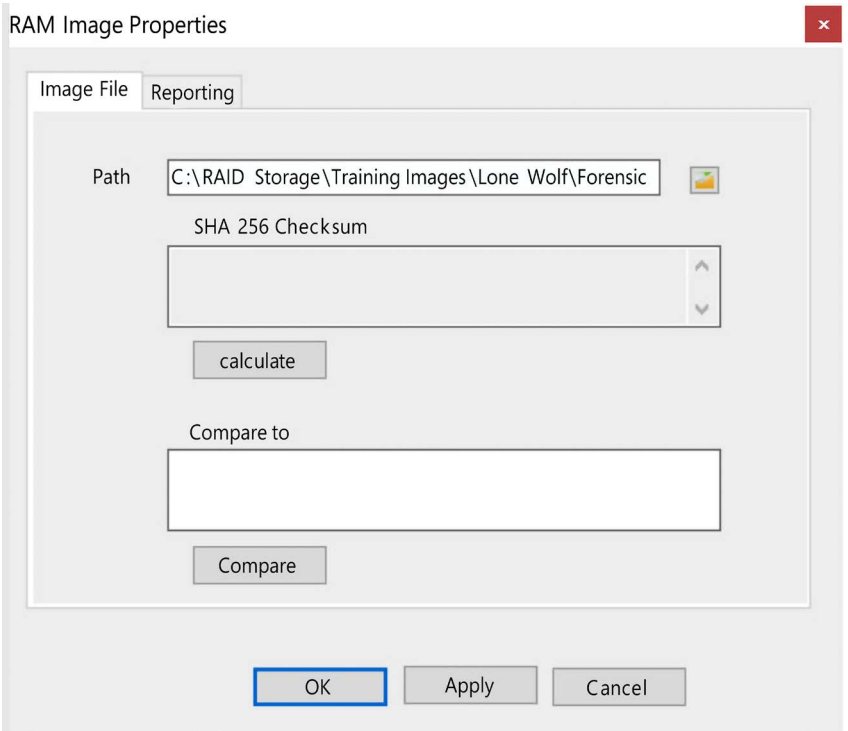


Figure 7.15: Volix RAM location

2. If you click on **Reporting**, you can specify the path for the report file that Volix will generate.
3. Once you select **OK**, the Volix wizard screen will appear. You now have the option of going through a questionnaire to determine what options you wish to run on the memory file. You can also select one of the pre-created scripts to search for that specific artifact, such as **Virus detection** or **Decrypt SAM Hashes**, as shown in the following screenshot:

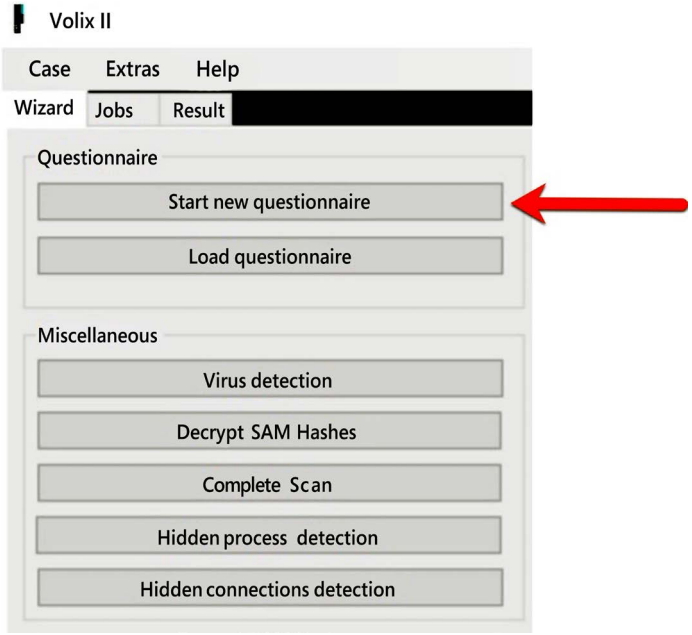


Figure 7.16: VOLIX wizard

4. I selected **Complete Scan**. You can see the results in the following screenshot:

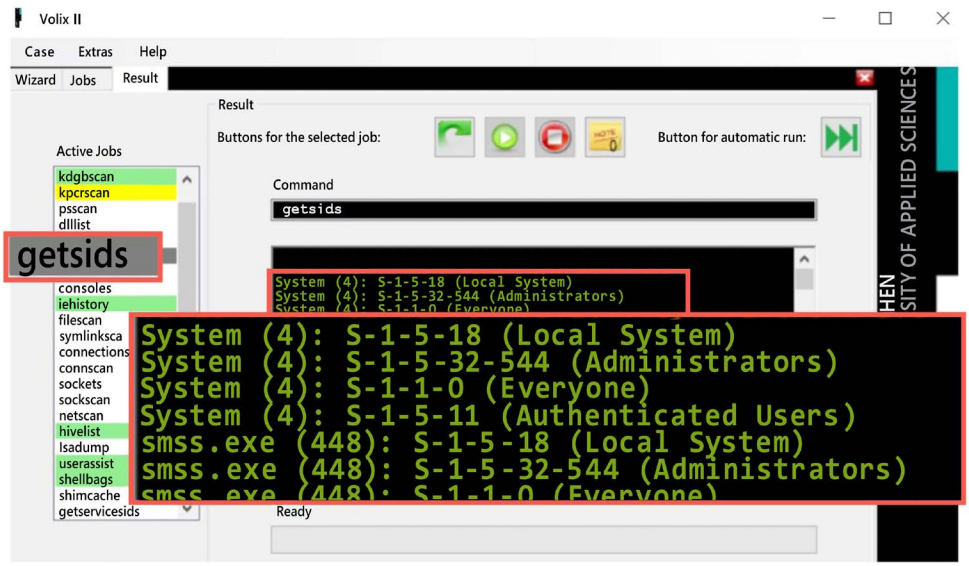


Figure 7.17: Volix scan results

From the preceding screenshot, you can see I have selected `getsids`. Below this on the screen, it has pulled out the SIDs that were in memory at the time of collection.

How many artifacts you ask it to search for at once will dictate the length of time for the application to complete. Overall, it is a relatively quick search compared to other tools.

Summary

In this chapter, you learned about the cornucopia of artifacts you can recover from RAM. You learned about the different tools you can use for the collection process and the tools you can use for analysis. Remember that the tools are constantly changing with the technology, and as new operating systems are released, your primary tool may not be able to collect RAM data. Always have a backup plan in case something like that occurs.

You now have the skills to identify and capture RAM in a manner that conforms to best practices. As you analyze the RAM you have captured, you may find artifacts showing the user's activity on the system, such as social media artifacts and passwords or encryption keys that can be recovered.

You may even find information relating to the user's use of email, which will lead us to our next chapter, which is all about email forensics.

Questions

1. Which of the following are sources of RAM data?
    a. Physical memory
    b. `Pagefile.mem`
    c. `Swapfile.page`
    d. ROM
2. Which file is created when the computer goes to sleep?
    a. `Pagefile.sys`
    b. `Swapfile.sys`
    c. `Hiberfil.sys`
    d. `Hibernation.sys`
3. When should you capture RAM?
    a. Every hour
    b. Every week

c. In every digital forensic investigation

d. When you deem it important

4. In general, how many items do you need in order to collect RAM?

    a. 1

    b. 2

    c. 3

    d. 4

5. DumpIt is a GUI tool.

    a. True

    b. False

6. It is acceptable to install DumpIt on the suspect's computer.

    a. True

    b. False

7. Which of the following are analysis tools?

    a. DumpIt

    b. FTK Imager

    c. Volatility

    d. MD5 hash

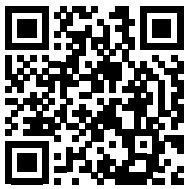
Further reading

Ligh, M. H., Case, A., Levy, J., & Walters, A. (2014). *The Art of Memory Forensics: Detecting Malware and Threats in Windows*, Linu. John Wiley & Sons. (Available at `https://www.amazon.com/Art-Memory-Forensics-Detecting-Malware/dp/1118825098`.)

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`



8

Email Forensics – Investigation Techniques

Email is just one portion of the global internet that has become a daily resource in the consumer and corporate realms. It has become one of the primary communication tools used by nearly every citizen of the industrialized world. Now that email has become part of everyone's everyday lives, criminals will use this vector to commit crimes and collaborate with their co-conspirators.

It can be difficult for the digital forensic investigator to trace an email back to the source from its destination. Therefore, the digital forensic investigator will have to be educated in the methods and delivery systems of the email life cycle. When the digital forensic investigator successfully identifies the source of the email, that will lead to additional forensic investigations of the digital evidence that was found at the source.

Where can you find digital evidence relating to an email investigation? The local machine will have the destination version of the email, the email server(s), the device used to access the email, such as a cell phone, and logs from the internet service provider. The digital forensic investigator will have to know which tools can analyze emails and the compound files of the email box used by some email suites. Knowing how to present this information to a non-technical person will be crucial for conveying the relevance of the data that the examiner recovered. By the end of this chapter, you will understand the protocols used to send and receive emails, how to decode the email headers, and how to analyze client and web-based emails.

We will cover the following topics in this chapter:

- Understanding email protocols
- Decoding emails
- Understanding client-based email analysis
- Understanding WebMail analysis

Understanding email protocols

An email protocol is a standard used to allow two computer hosts to exchange email communication. When an email is sent, it travels from the sender's host to an email server. The email server can forward the email through a series of relays until it arrives at an email server close to the recipient's host. The recipient will receive a notification stating that an email is available; the recipient will then reach out to the email server to get the email.

Users typically use an email client to access emails. An email client can use different protocols to access the email. We will now discuss some email protocols you may encounter when conducting a digital forensics investigation.

Understanding SMTP – Simple Mail Transfer Protocol

SMTP is the protocol for email transmission. It is an internet standard based on RFC 821 but was later updated to RFC 3207, RFC 5321/5322.

Tip

RFC stands for Request for Comments. This is used on internet/communications technology to create standards. An RFC may come from different bodies, such as the Internet Architecture Board/Internet Engineering Task Force or an independent researcher. It was initially designed to track the development of the original ARPANET but has now evolved into a source of official documentation regarding internet specifications and communication protocols.

Mail servers use SMTP to send and receive email messages from all locations accessing the internet. Typically, you will find an SMTP server utilizing **Transmission Control Protocol (TCP)** port **25** on the network. The path from the sender to the recipient is outlined in the following diagram:

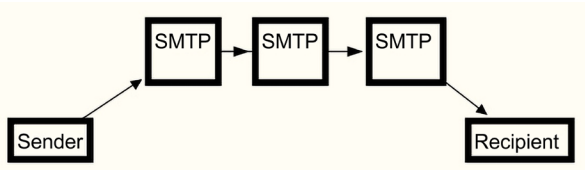


Figure 8.1: Diagram of an email sent by SMTP

When the user sends an email, it will travel from the host to a series of SMTP servers until it reaches the destination SMTP server. The recipient will have to use a different protocol to retrieve the email, which is our next topic.

The next protocol we will discuss is POP3.

Understanding the Post Office Protocol

POP3 is the standardized protocol that allows users to access their inbox and download emails. POP3 is specifically designed only to receive emails; the system does not allow users to send emails. This protocol enables the user to be offline when drafting, reading, or replying and can access the online mailbox on-demand at the user's request. Be aware that the email you are conducting your digital forensic examination on may be the only copy. The user has the option to not leave a copy of the email on the server. Once the email has been downloaded, the system can remove it to reduce storage use on the server.

You will find POP utilizing port **110** on the network.

In the following diagram, you can see the general functionality of the SMTP-POP process:

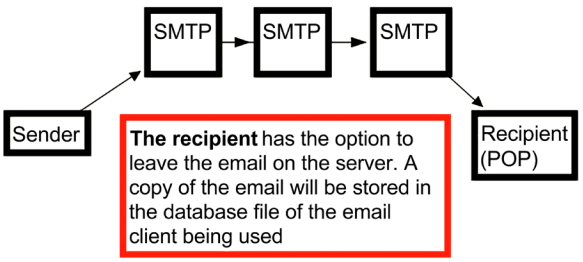


Figure 8.2: SMTP-POP map

Here, you can see the path the email takes, which is as follows:

1. The email originates from the sender.
2. The SMTP server forwards it to the destination.
3. The recipient collects the email from the server. The recipient can decide if a copy of the email stays on the server or whether the email will be deleted when the user downloads the email from the server.

When we look at the next protocol, we will discuss functions similar to SMTP, but with some significant differences. We will discuss these differences in the next section.

IMAP – Internet Message Access Protocol

IMAP is the **Internet Message Access Protocol** and is a standard protocol used by an email client to access emails on an email server. The protocol was designed to allow complete inbox management with multiple clients. In most cases, email messages will be left on the server until the user deletes them. IMAP is a newer protocol than POP, but both are still prevailing email standards in use today. The most significant difference between IMAP and POP is that POP retrieves the mailbox's contents, and IMAP was designed as a remote access mailbox protocol.

In the following diagram, you can see the general functionality of the SMTP-IMAP process:

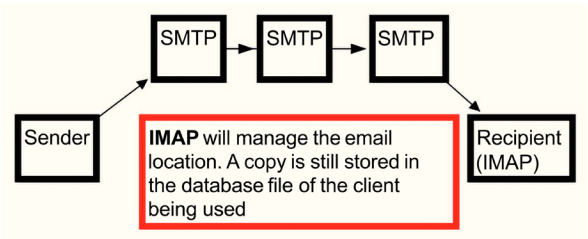


Figure 8.3: IMAP map

Here, you can see the path the email takes:

1. The email originates from the sender.
2. The SMTP server forwards it to the destination.
3. The recipient collects the email from the server. A copy of the email stays on the server until the user explicitly deletes it.

All three protocols we just discussed are typically used in the email client-server relationship. Users also have another option when it comes to accessing emails known as web-based email, which is the topic of the next section.

Understanding web-based email

Web-based email is a service the user accesses with a web browser. Standard webmail providers are Gmail, Yahoo Mail, and Outlook/Hotmail. Some internet service providers also provide an email account that the user can access with a web browser.

User-deleted emails stored on a web-based email server typically remain on the server until the system deletes them. A characteristic feature of web-based email is that when the user deletes an email, it is moved from the inbox into a **Deleted/Trash** folder and can still be accessed. However, after the email remains in the **Deleted** folder for a set timeframe, the system permanently deletes it from the user's inbox.

We have gone over the different methods of how a user may access email services. However, once you have the email dataset available for examination, you may find the contents of the email encoded. So, how do you decode the contents of the email to determine whether a crime/violation has/has not been committed?

In the next topic, we will decode the email header so that you can make an informed choice about your investigative endeavors.

Decoding email

An email has many unique identifiers for a digital forensic investigator to identify and track down. The mailbox and domain name, along with the message ID, will allow a digital forensic investigator to serve judicially approved subpoenas/search warrants on the vendor to follow any investigative leads.

In this section, we will break down the email header one section at a time so that you can decide how to conduct your investigation. First, we will start by discussing the email envelope.

Understanding the email message format

The vast majority of email users are only familiar with basic email information, such as this:

```
Subject background checks
Date 07/19/2008 23:39:57 +0
Sender alison@m57.biz
Recipients jean@m57.biz
```

We are back to dealing with our friend Jean, and by looking at the email, we can see several fields commonly associated with an email. Here, we know the subject, background checks, the date and time when the user sent the email, the sender, and the recipient.

We also have the content of the email, as shown here:

```
Jean,

One of the potential investors that I've been dealing with has asked me to
get a background check of our current employees. Apparently they recently
had some problems at some other company they funded.

Could you please put together for me a spreadsheet specifying each of our
employees, their current salary, and their SSN? Please do not mention this
to anybody.

Thanks.

(ps: because of the sensitive nature of this, please do not include the
text of this email in your message to me. Thanks.)
```

As we look at the email, it appears that the email was sent to *Jean* from *Alison*. Alison is requesting a spreadsheet of confidential employee information. Based on the basic examination of this email, there is nothing to contradict what it initially appears to be.

The user has created the information in the *to* and *from* fields and the *subject* and the *content* of the email. The system derives the date and time from the system time, which can be set by the user.

Underneath the typical email information, another layer of data is instrumental when conducting your investigations. This is referred to as the *email header*, and it contains information about the source, transmission, and destination of a specific email.

Most email clients require an additional command to view the email header. For example, Gmail requires you to click **Show original** to see the email header. For example, the following is the email header for the email Jean received from Alison:

```
-----HEADERS-----

Return-Path: simsong@xy.dreamhostps.com

X-Original-To: jean@m57.biz

Delivered-To: x2789967@spunkymail-mx8.g.dreamhost.com
```

```
Received: from smarty.dreamhost.com (sd-green-bigip-81.dreamhost.com
[208.97.132.81]) by spunkymail-mx8.g.dreamhost.com (Postfix) with ESMTP id
E32634D80F for <jean@m57.biz>; Sat, 19 Jul 2008 16:39:57 -0700 (PDT)

Received: from xy.dreamhostps.com (apache2-xy.xy.dreamhostps.com
[208.97.188.9]) by smarty.dreamhost.com (Postfix) with ESMTP id 6E408EE23D
for <jean@m57.biz>; Sat, 19 Jul 2008 16:39:57 -0700 (PDT)

Received: by xy.dreamhostps.com (Postfix, from userid 558838) id
64C683B1DAE; Sat, 19 Jul 2008 16:39:57 -0700 (PDT)

To: jean@m57.biz From: alison@m57.biz

subject: background checks

Message-Id: 20080719233957.64C683B1DAE@xy.dreamhostps.com

Date: Sat, 19 Jul 2008 16:39:57 -0700 (PDT)
```

The email header shows where the email originated from and what servers it touched upon. Starting from the bottom, we can see the `Message-Id` field:

```
Message-Id: <20080719233957.64C683B1DAE@xy.dreamhostps.com>
```

The `Message-Id` field contains a unique ID for every email that has been sent. When a user sends an email, it will receive its message ID from the first email server it touches. The message ID will be globally unique, so there should not be another email with the same message ID. If you find different emails that contain the same message ID, you are dealing with one of two scenarios:

- The email server is not compliant with the standard.
- A user has altered the email.

When you look at the message ID, you will see a string of random alphanumeric characters, as well as the `@` symbol and a domain name. Sometimes, the arbitrary string of alphanumeric characters contains a date/timestamp. Looking at the preceding example, we can see the numbers `20080719233957`, which can be translated to *2008 07 19* – the year, month, and day. *23:39:57* is the time in hours, minutes, and seconds (GMT) when the email touched the first server.

We can see the first **Received** line from the bottom to the top. This email transverses three different email servers. As the email crosses a server on its journey to its destination, each email server will attach a **Received** line on top of the preceding **Received** line. You can follow the email path from source to destination. In the email, we are examining the first server the email touched, which is as follows:

```
Received: by xy.dreamhostps.com (Postfix, from userid 558838) id
64C683B1DAE; Sat, 19 Jul 2008 16:39:57 -0700 (PDT)
```

This is the first server the email touched; we have the domain name `dreamhostps.com` and a user ID. The next logical step would be to subpoena the ISP and identify the subscriber with user ID `558838`. Finally, the term `Postfix` identifies the email server. Postfix is a free, open-source mail transfer agent and could be a commercial email server or an email server maintained by a potential bad actor.

The following two **Received** lines identify the subsequent servers on the path to the destination:

```
Received: from smarty.dreamhost.com (sd-green-bigip-81.dreamhost.com
[208.97.132.81])
by spunkymail-mx8.g.dreamhost.com (Postfix) with ESMTP id E32634D80F for
<jean@m57.biz>;
Sat, 19 Jul 2008 16:39:57 -0700 (PDT)
Received: from xy.dreamhostps.com (apache2-xy.xy.dreamhostps.com
[208.97.188.9])
by smarty.dreamhost.com (Postfix) with ESMTP id 6E408EE23D for <jean@m57.
biz>;
Sat, 19 Jul 2008 16:39:57 -0700 (PDT)
```

In both cases, we now have the IP addresses of the specific servers (and server names) that touched the email.

What's interesting is when we look at the `Return-Path` field:

```
Return-Path: <simsong@xy.dreamhostps.com>
```

The `Return-Path` is the address where undeliverable messages will be sent. `Return-Path` will also override the `From` field that the user will see. You will see this being used in email mailing lists, where you can reply to the user of the post and not to the list.

There are optional fields that you may come across in your investigations. These fields typically start with an `X-`, as shown here:

```
X-Priority: 3
X-Mailer: PHPMailer 5.2.9 (https://github.com/PHPMailer/PHPMailer/)
Message-Id: ff176aaf06e2f6958ada6e2d3c43b095@x3.netcomlearning.com
X-Report-Abuse: Please forward a copy of this message, including all
headers, to abuse@mandrill.com
X-Report-Abuse: You can also report abuse here: http://mandrillapp.com/
contact/abuse?id=30514476.1925a088d66f450cb25a4034f3ec6942 X-Mandrill-
User: md_30514476
```

These fields are not part of the email protocol standard. They can contain information about a virus scan, spam scans, or information about the server. As you can see, it provides information about contact information regarding abuse, such as spam. You may also see an optional field called `X-Originating-IP` that may contain the sender's IP address when the user sent the message. An email provider can strip that information and replace it with a server address, which happens when a message is sent from Gmail.

A note about IP addresses: there are two different types of IPv4 addresses: public and private. You may see both in the email header. However, if you see a private IP address, you cannot identify the provider (unless you are investigating within the organization). Private IPv4 addresses run from the following addressing schemes:

- 10.X.X.X
- 127.X.X.X
- 172.16.X.X
- 192.168.X.X

We will discuss email attachments in the next section.

Email attachments

MIME is the acronym for **Multipurpose Internet Mail Extensions**, the internet standard for allowing emails to accept text other than ASCII, binary attachments, multi-part message bodies, and non-ASCII base header information. When you are viewing the header, you will see MIME indicated with the following:

```
MIME-Version: 1.0
```

An example of this is as follows:

```
MIME-Version: 1.0
Content-Type: text/html; charset=us-ascii
Content-Transfer-Encoding: 7bit
```

Here, we can see the content type, HTML, and with the following line, we see it is using 7-bit coding. We would also see Base64 encoding if there were an attachment, which converts the binary data into ASCII text.

The system will separate the body of the email based on the data type of each segment. For example, a JPEG image will accompany one segment and store ASCII text in a different segment. In addition, each segment will start with a MIME header that includes the keyword `_PART_`.

Now that we have discussed the email and header, we need to look at some of the clients the user may use to access the emails.

Understanding client-based email analysis

A user has access to many email clients to retrieve, read, and send emails. Depending on whether you're in the consumer or commercial environment, you may encounter different email clients. In the consumer market, you will find that Microsoft Outlook/Outlook Express will prevail because it is preinstalled on the system. In addition, Microsoft Outlook comes with the Microsoft Office suite. There are also freeware options available such as the Thunderbird email client.

You can conduct an email examination by exporting the container used by the client and opening it with the email client installed on your forensic computer. Another option is to utilize specialized commercial forensic software created for email examinations. The more common forensic suites will typically analyze the more common email client containers.

We will discuss some more common email clients in the following sections.

Exploring Microsoft Outlook/Outlook Express

Outlook stores email information in several file types, such as `pst`, `.mdb`, and `.ost`. We will find the PST file on the user's hard disk at the following path:

```
\Users\$USER$\AppData\Local\Microsoft\Outlook
```

The OST file is an offline file that may also be stored on the user's hard drive in the same path as the PST file. Finally, you will find the MDB file on the server. Typically, this file is found when you are investigating a corporate environment.

The system will store all the content used with the Outlook client in the PST/OST file. Be aware that the user can change the default location and the naming convention. You do not need a login to access the PST/OST file.

If you need to carve out a PST/OST file from the unallocated space of the storage device, you may have to deal with fragmentation because of the potential size of the PST/OST file.

Microsoft has replaced Outlook Express with Windows Live. The following section will provide details about this client.

Exploring Microsoft Windows Live Mail

Starting with Windows Vista and Windows 7, Windows Live became the default email client shipping with the Windows operating system. (Note that it has been discontinued and that Windows Mail is now included with Windows 10 instead. Windows Mail will not store the emails on the disk, the emails will only be cloud based.) The client stores email messages in the following path:

```
\Users\$USER$\AppData\Local\Microsoft\Windows Live Mail
```

Users can use this client to access their web-based emails as well. Windows Live Mail will download the contents of those accounts and then create the folder structure within the user's path.

The client will store the emails as an `.eml` file in the Windows Live Mail folder, as shown here:

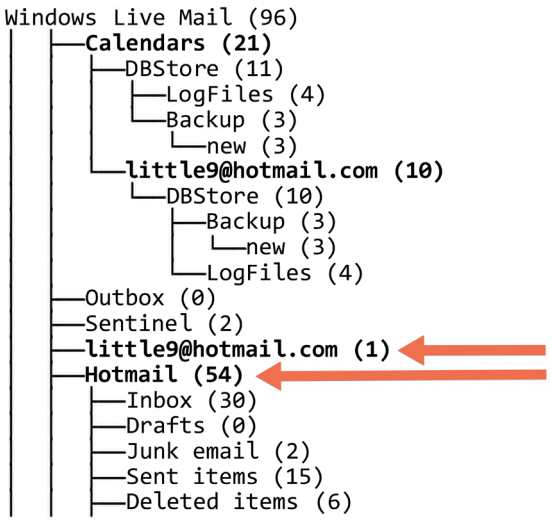


Figure 8.4: Windows Live Mail folder

As you can see, this user was using Hotmail with the Windows Live Mail application. You can see the email address, `little9@hotmail.com`, and see that 54 emails are being stored in the user's folders.

The emails are in the standard text format, `.eml`, which can be read by any forensic tool. Alternatively, you can use a text editor. The next client is also popular and free: Mozilla Thunderbird.

Mozilla Thunderbird

Thunderbird is a free, open-source email client provided by Mozilla. Thunderbird will store emails within a `.MBOX` file. The MBOX format is a generic term for a family of file formats used to store emails. It will keep all the emails from folder into a singular database file. By default, the examiner can find the MBOX file in the following path:

```
$USERNAME$\AppData\Roaming\Thunderbird\Profiles
```

The following is the folder structure you will see when Thunderbird is installed:

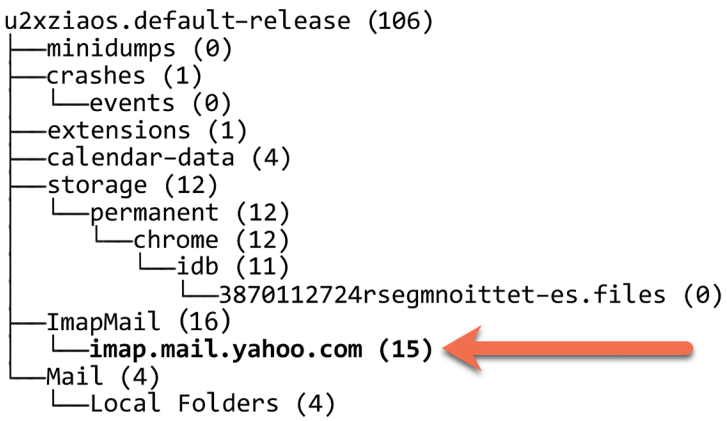


Figure 8.5: Thunderbird folder structure

The profile name is created by Thunderbird. The release version of the software the user has installed can also be seen here. As we analyze the folder structure, we will see that it contains information about crashes and stores data in a minidump when a crash occurs. There may also be calendar data and mailboxes.

Here, the user is using the IMAP protocol to access their Yahoo mail account, and there are 15 items stored within the folder.

When we look in the folder, we will see the following files:

- `Archive.msf`
- `Archives.msf`
- `Bulk Mail.msf`
- `Draft.msf`
- `Drafts.msf`
- `INBOX`
- `INBOX.msf`
- `msgFilterRules.dat`
- `Sent-1.msf`
- `Sent.msf`
- `Templates.msf`
- `Trash.msf`

The **MSF** files are **Mail Summary files**, one part of the email. The email client, Thunderbird, stores the email data in two different parts. The first part is the MBOX file, which does not have a file extension. The MSF files are the index files for Thunderbird and contain email headers and a summary. Thunderbird uses these files as an index to locate the email stored in the MBOX.

In the following screenshot, three emails are being stored in the MBOX. When X-Ways parses out the inbox, the emails will have a `.eml` file extension:

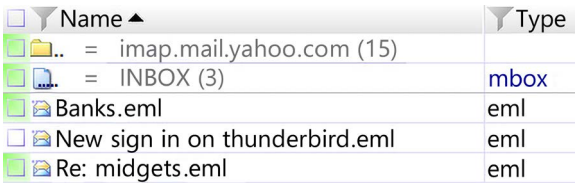


Figure 8.6: Thunderbird inbox

The MBOX format is used by many email clients, including Apple Mail, Opera Mail, and Thunderbird. In addition, most commercial and open-source forensic suites will process the MBOX and provide access to emails.

While the user can access their email from a client, another popular option allows the user to access their email without using a client: WebMail.

Understanding WebMail analysis

Web-based email has become increasingly popular as we transition from the twentieth to the twenty-first century. It's easy to access, requires little to no configuration from the user, and is available from any computer. In the simplest terms, WebMail is just another internet artifact for conducting browser analysis (we will cover internet artifacts in *Chapter 9*, *Internet Artifacts*).

The service provider maintains the user's email and may provide additional services, such as address books and calendars. Users can use a host-based client to access web-based email. Still, I have found that those users are in the minority when content is hosted by the service provider, which provides additional obstacles to the digital forensic investigator. The only artifacts relating to the content may be in the user's internet history, which may be fragmented. If a digital forensic investigator wants to access the content of a user's web-based email, they will have to serve a search warrant (in the United States; your jurisdiction may have different requirements) on the service provider. You may be unable to access or recover any deleted emails from the account. It will depend on the specific set of circumstances for each service provider.

Suppose the digital forensic investigator wants to investigate the user's use of web-based email. In that case, they will have to analyze the temporary internet files or the internet *cache* on the user's system. The temporary internet files/cache contains images, text, or any web page component the user has viewed in their browser.

Their browser saves this information in the temporary internet **files/cache** location to enhance the user experience. It does this by having a faster response time when presenting pages to the user. Instead of continually redownloading the content, you can reach back into the cache and present that information to the user.

Gmail is very popular, and when the service first deployed its web application, it changed how WebMail was presented to users. No longer were static web pages displaying the email content and the user's email folders. Instead, Gmail dynamically created content on the fly for each user. The system no longer saved image files and text to the user's local storage device; instead, Gmail used **Asynchronous JavaScript** (**AJAX**) and XML files. Unfortunately, this new method did not allow a web page to be rebuilt by investigators.

You can still recover artifacts within the internet cache and other potential sources such as RAM or the pagefile on the user's local storage device. You will need to conduct keyword searches for email addresses or keyword searches for terms related to your investigation.

Before I look into the cache, I want to look into the internet history of the installed browser to see if the user has accessed web-based email. For the Chrome browser, you will find the history stored in a SQLite database named `History` at the following path:

```
$USER$\AppData\Local\Google\Chrome\User Data\Default
```

The analysis of the `History` database shows the user accessed the Gmail web-based service, as shown in the following screenshot:

19	08/28/2019 22:19:39 +0	Inbox (2) - badguyneedslove@gmail.com - https://mail.google.com/mail/?pc=topnav-about-n-en Gmail

Figure 8.7: Email – History

We have a date/time stamp, along with the email address. The artifact also shows that the user had two unread emails in the inbox when accessing the service.

I found this in the internet cache for the Google Chrome browser, which the examiner can find at the following location:

```
$USER$\AppData\Local\Google\Chrome\User Data\Default\History Provider
Cache
```

As you can see in the following screenshot showing the Chrome cache, the content is not easily decipherable and does not give us a lot to follow up with:

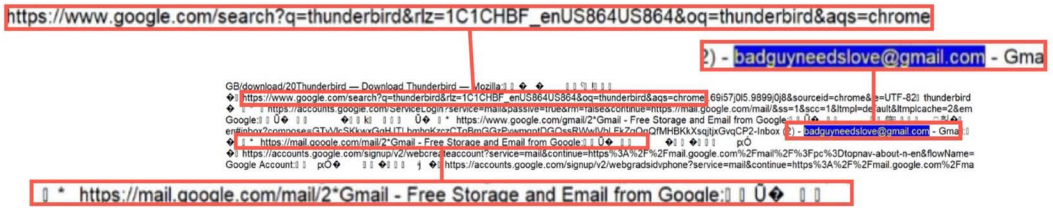


Figure 8.8: Chrome cache displayed

If we keep searching for the email address we found in the cache, `badguyneedslove@gmail.com`, we may find other artifacts, such as the following:

```
{"endpoint_info_list":[{"endpoint":"smtp:badguy27@yahoo.com",
"c_id":"d24c.2d00",
"c_name":"Joe Badguy Smith"},
{"endpoint":"smtp:badguyneedslove@gmail.com",
```

```
"c_id":"e80f.5b71","c_name":"John Badguy Smith"},
{"endpoint":"smtp:yahoo@mail.comms.yahoo.net",
"c_id":"624f.10f0","c_name":"Yahoo! Inc."}]}
```

This artifact, which the examiner can find within the cache, gives us another email address, `badguy27@yahoo.com`, to follow up. Unfortunately, the content of the email remains out of reach.

Let's look at the Firefox cache and see whether it can give us a better look at the cache and history.

The examiner can find the cache and history for the Firefox browser at the following location:

```
$USERS$\AppData\Local\Mozilla\Firefox\Profiles\<profile>\cache2
```

Firefox will store the internet history and cache in the user's profile. The folder structure you will see may look like this:

```
Mozilla (1,505)
└──Firefox (1,505)
   └──Profiles (1,504)
      ├──55abhq00.default-release (1,504)
      │  ├──safebrowsing (50)
      │  │  └──google4 (10)
      │  ├──jumpListCache (5)
      │  ├──startupCache (236)
      │  ├──cache2 (1,162)
      │  │  ├──entries (1,160)
      │  │  └──doomed (0)
      │  ├──thumbnails (0)
      │  ├──OfflineCache (1)
      │  └──safebrowsing-updating (49)
      │     └──google4 (9)
      └──cqr6ioib.default (0)
```

Figure 8.9: Firefox folder structure

It looks like the visual depiction of the content of the Firefox cache is not much better:

Figure 8.10: Firefox Cache

It does not provide a wealth of information, but it does supply breadcrumbs for us to follow up and conduct additional investigative efforts.

In the world of forensics, the artifacts you rely upon can quickly change with new updates to the software or changes in the operating system. Be flexible with your investigative techniques so that you can jump into the latest technology to make your investigation successful. Once you have identified that the subject of your inquiry is using web-based email, your best course of action is to serve the service provider with the appropriate judicial paperwork to freeze the account and get the required content.

Summary

In this chapter, we have gone over standard email protocols: the system uses SMTP to send emails, while POP and IMAP are used to receive emails. IMAP also includes features that the user can use to manage the inbox. We went over the email header and the components that make up the header. We also discussed WebMail and email clients.

You now have the skills necessary to read an email header and determine the servers that the user used to transmit the email and what protocols the system used to send and receive the email. When conducting a digital forensic examination, you can now identify artifacts from typical email clients and web-based emails.

In the next chapter, you will learn about the similarities between web-based email clients.

Questions

1. Which of the following is not an email protocol?
    a. HTML
    b. POP
    c. SMTP
    d. IMAP
2. Which of the following allows the user to manage their inbox?
    a. COC
    b. POP
    c. FreeBSD
    d. IMAP
3. The email header is created by user input information.
    a. True
    b. False
4. Thunderbird stores emails in which file?
    a. Inbox
    b. Outbox
    c. MBOX
    d. Letterbox

5. Which email client uses a PST file?
    a. Thunderbird
    b. Gmail
    c. Yahoo Mail
    d. Outlook
6. Windows Live Mail was replaced with which client?
    a. Outlook Express
    b. Outlook
    c. Windows Mail
    d. Windows Email
7. You will always find the content of a web-based email in the user's cache.
    a. True
    b. False

You will find the answers at the back of this book, under *Assessments*.

Further reading

- *Jones, R. (2006). Internet forensics: Beijing: OReilly* (can be purchased at `http://shop.oreilly.com/product/9780596100063.do`)

Exercise

Data set

`Jean outlook.pst`

Software needed

Autopsy - `https://www.autopsy.com/`

Scenario

A company, XYZ LLC, finds that a spreadsheet containing confidential information was posted as an attachment in the "technical support" forum of a competitor's website.

The spreadsheet came from the CFO of XYZ LLC, Jean's computer.

Interviews

You are tasked with investigating the leak of confidential information. To that end, you conduct interviews with the President and Chief Financial Officer of XYZ LLC, Alison and Jean respectively. Here are excerpts from their interviews:

Alison (President):

- I don't know what Jean is talking about.
- I never asked Jean for the spreadsheet.
- I never received the spreadsheet by email.

Jean (CFO):

- Alison asked me to prepare the spreadsheet as part of a new funding round.
- Alison asked me to send the spreadsheet to her by email.
- That's all I know.

Email accounts

Alison (President):

`alison@m57.biz`

Jean (CFO):

`jean@m57.biz`

Question to answer

Examining Jean's email – How did the documents get on the competitor's website? Use Autopsy (or your tool of choice) to analyze the emails contained in the `.pst` file.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

9

Internet Artifacts

The internet has become a staple in commercial and consumer environments. Digital communication between users is a daily activity. It is uncommon for a household not to have a device connected to the internet in some manner. They give students in elementary school devices to connect to the internet to enhance their education. Email addresses, URLs, social media, and file sharing are all vectors of activities a user can partake in. It is up to the user to decide whether their online activities will meet social norms and be accepted or whether they will cross the line and conduct criminal activity. Your job as a digital forensic investigator will be to investigate their activities in the digital realm.

In this chapter, we will be discussing some of the common browsers and social media sites and the artifacts they provide. We will also discuss P2P file-sharing tools, cloud computing services, and their artifacts.

We'll be covering them in the following topics:

- Understanding browsers
- Social media
- **Peer-to-Peer (P2P)** file sharing
- Cloud computing

Understanding browsers

What is a browser? A user can use a program or an application to access websites via the **World Wide Web (WWW)**. The best browser is an ongoing debate and can be a very personal choice for a user. The user has options to personalize the browser to enhance their experience when accessing the WWW.

As a result, this creates many artifacts that any digital forensic investigator can use to recreate the user's activity. In addition, there will be logs, history files, and cache files that a digital forensic investigator can examine to identify unethical or criminal activity.

Like all technology, browsers are continuously being updated and changed. User experience is typically the key to these changes, but security has also been a driving factor lately. While the security enhancements are not specifically designed to frustrate or hamper digital forensic investigations, they do have that effect.

We will now discuss some common browsers you may encounter during your investigations. This is not an all-inclusive list, and there will be browsers that you may encounter that we will not discuss.

As of July 2019, according to W3Counter, the Chrome browser has a 55 percent market share, followed by Safari at 12 percent, Internet Explorer/Edge at 8 percent, and Firefox at 6 percent. Chrome is the leader in the browser wars, and you can find the Chrome browser in many operating systems. We will now discuss the Chrome browser and the artifacts you may run into during your investigations.

Exploring Google Chrome

Google Chrome was released in 2008 and was very popular with users. It provided a fast and efficient user experience and experienced very few exploits. Chrome stores much of the data within different databases and provides us with the option of syncing data across multiple platforms. This means you may come across artifacts that have been generated with another device. So, let's get into the details of the Chrome browser.

Understanding bookmarks

The first artifact we will look at is the user's bookmarks. The bookmarks allow the user to save web pages they find interesting and may give insight into the user's activity. You can find the bookmarks file at the following path:

```
%USERS%/AppData/Local/Google/Chrome/User Data/Default/Bookmarks
```

The file will *not* have a file extension, and it is a **JSON (JavaScript Object Notation)** formatted file. JSON is an open standard file and data interchange format.

You can open the file in any text reader, and it will show you the contents, as shown in the following screenshot of a JSON BBC bookmark:


```
"date _added": "13105251021405925",
"id": "110",
"meta_info": {
   "last_visited_desktop": "13197567715245509"
},
"name": "BBC News",
"sync_transaction_version": "592"
"type": "url",
"url": "http://news.bbc.co.uk/"
, {
"date_added": "13105251021408611",
"id": "111",
"meta_info": {
   "last_visited_desktop": "13197950930217586"
},
"name": "CNN",
"sync_transaction_version": "592",
"type": "url",
"url": "http://www.cnn.com/"
```


Figure 9.1: JSON BBC bookmark

Here, we can see the following fields:

- Date added
- Last visited desktop
- The name of the bookmark
- The URL

But the presentation of information is not very graphical. An alternative method is to use a free text viewer such as Notepad++ (available at `https://notepad-plus-plus.org/`) and the JSON plugin.

That will make the folder structure easier to read, as depicted in the following screenshot:

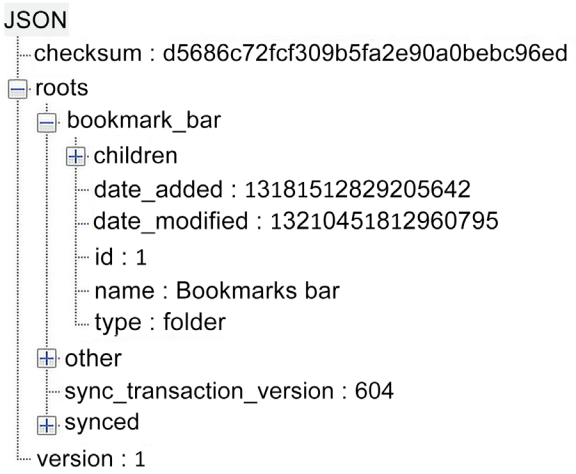


Figure 9.2: JSON root folder

The preceding screenshot shows that there are three folders underneath the root directory: `bookmark_bar`, `other`, and `synced`. When the `bookmark_bar` folder is expanded, it reveals the existence of additional `children` folders, as depicted in the following screenshot:

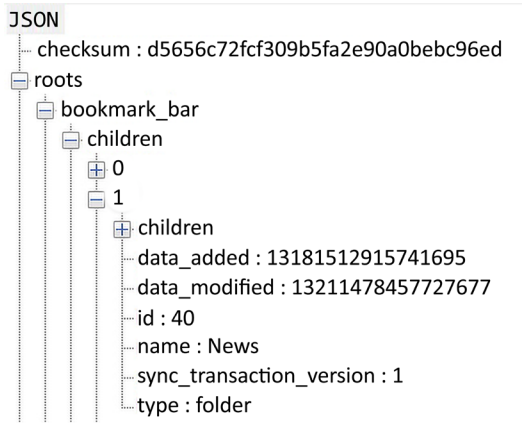


Figure 9.3: JSON children

The first folders are only labeled with numbers starting from zero. When we expand the `1` folder, the folder's name is revealed to be `News`, along with the date/timestamps or when it was added and modified. When the `children` folder is opened, the bookmarks are now available for viewing, as shown in the following screenshot:

Figure 9.4: JSON bookmarks

The date/timestamps are encoded in Google Chrome Value. To decode the date/timestamp, I enjoy using the open-source tool DCode (which is available at `https://www.digital-detective.net/dcode/`). DCode can also be used to decipher numerous different types of date/timestamps.

As shown in the following screenshot, you can see the tool has translated the Google Chrome Value into a more readable format:

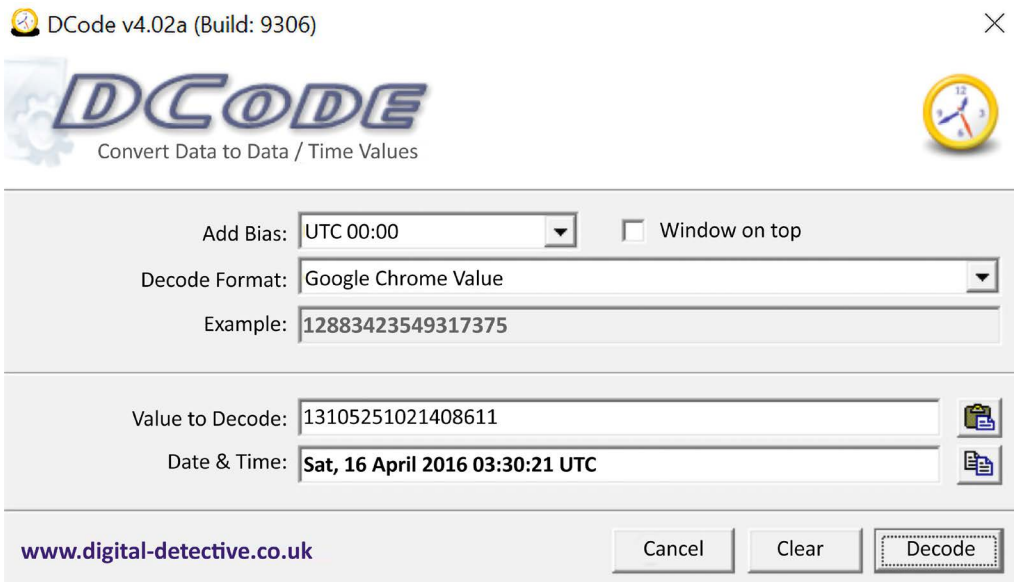


Figure 9.5: DCode tool used for translating the Google Chrome Value

The BBC News bookmark was added on Saturday, April 16, 2016, at 03:30:21 UTC.

The mere presence of an incriminating bookmark may not be enough to act upon. To support the hypothesis, the user visited the web page indicated by the bookmark; you may have to look at additional artifacts. One such artifact is the history file, which we will examine next.

Understanding the Chrome history file

The Google Chrome history file will be found in the following path:

```
%USERS%/AppData/Local/Google/Chrome/User Data/
```

The file will not have a file extension, and it is a SQLite database. Most forensic tools will view the contents of the database. The history database contains quite a lot of information about the user's activity:

- Downloads:

  This will include the path of where the downloaded file was saved, the location the file was downloaded from, the start/stop times of the download, and the size of the file downloaded.

- Keyword search will track the search terms that the user entered using the URL address bar.
- Typed URLs will track the URLs the user typed into the address bar.
- History:

  This will track the URLs visited by the user, the number of times the URL was visited, and the date/time the URL was visited.

As we examine the history file, the following text depicts the user's activity:

```
C:\Users\IEUser\Downloads\Thunderbird Setup 68.0.exe

https://www.thunderbird.net/en-GB/download/

08/29/2019 16:14:38 +0

08/29/2019 16:14:40 +0
```

Figure 9.6: User's Chrome History

The user downloaded a file named `Thunderbird Setup 68.0.exe` and saved it in the `Downloads` folder of the `IEUser`. We also have the start time of 16:14:38 UTC and know that the download was completed at 16:14:40 UTC.

```
gmail

https://www.google.com/search? q=gmail& (REDACTED)

08/28/2019 22:17:04 +0

thunderbird

https://www.google.com/search?q=thunderbird& (REDACTE

08/29/2019 16:14:29 +0
```

Figure 9.7: User's Search History

The user also conducted two keyword searches, as follows:

The user searched for the terms `Gmail` and `Thunderbird`. The search engine Google was used for the search, searching for the term `Gmail`. The search was conducted on August 28, 2019, and the `Thunderbird` search was conducted on August 29, 2019. (I have redacted a portion of the URL(s) to help with formatting.)

The following is a listing of the websites the user visited:

```
08/28/2019 22:22:08 +0

(2 unread) - badguy27@yahoo.com - Yahoo Mail

https://mail.yahoo.com/d/REDACTED)

08/28/2019 22:22:36 +0

Banks - badguyneedslove@gmail.com - Gmail

https://mail.google.com/mail/(REDACTED)

08/29/2019 16:14:29 +0

thunderbird - Google Search

https://www.google.com/search?q=thunderbird&REDACTED)

08/29/2019 16:14:33 +0

Thunderbird — Download Thunderbird — Mozilla

https://www.thunderbird.net/en-GB/download/    1
```

Figure 9.8: User's Internet History

From the history file, we can see that the user searched for the term **Thunderbird** and that seconds later, the user went to the Thunderbird download page. Before the user searched for the term **Thunderbird**, the user visited two different web-based email accounts. In addition, we can see the user visited Yahoo mail and Gmail the night before. Based on this analysis, we have identified distinct and different email addresses the user may be using or, at a minimum, has access to.

Next, in terms of browser artifacts, are cookies.

Cookies

A cookie is a dataset created by a website and stored on the user's system. Cookies are designed to track the user's activity, such as adding an item to a shopping cart or recording the number of pages the user has visited. Be aware that just because a cookie is on the system, it is not conclusive evidence that the user knowingly visited the site. You will need to find other artifacts as supporting evidence of the user's activity.

The Google Chrome cookie file can be found at the following path:

```
%USERS%/AppData/Local/Google/Chrome/User Data/Default
```

The file will not have a file extension, and it is a SQLite database. Most forensic tools will view the contents of the database. The following screenshot is the output of X-Ways Forensics:

creation_utc	host_key	name	value	path	expires_utc	is_secure	is_httponly	last_access_utc	has_expires	is_pe
13211504229653934	.google.com	_ga		/gmail/about	13274576231000000	0	0	13211504229653934	1	1
13211504229654926	.google.com	_gid		/gmail/about	13211590631000000	0	0	13211504229654926	1	1
13211504361869670	.google.com	APISID		/	13274576361869670	0	0	13211568843104513	1	1
13211504373193089	mail.google.com	COMPASS		/mail/u/0	13212368374193089	1	1	13211504554421139		

Figure 9.9: Cookies

While the data is readable, the format leaves a lot to be desired. A third-party application, Chrome Cookies View (available at `http://www.nirsoft.net/utils/chrome_cookies_view.html`), will parse the data in a format that is easy to read, as shown in the following screenshot:

Host Name	Path	Name	Value	Secure	HTTP Only	Last Access...	Created On	Expires
.google.com	/gmail/about	_ga		No	No	8/28/2019 22:17	8/28/2019 22:17	8/27/2019 22:17
mail-ads.google.com	/mail/u/0	COMPASS		Yes	Yes	8/28/2019 22:19	8/28/2019 22:19	9/7/2019 22:19
www.yahoo.com	/	flash_enabled		No	No	8/28/2019 22:21	8/28/2019 22:21	9/27/2019 22:21

Figure 9.10: Cookie View

Now, the columns and the data are lined up and formatted correctly. The tool also converts the date/timestamp from Google Chrome time into UTC time. Cookies are one part of tracking a user's activities, but the cache may also contain artifacts that are useful for your investigation.

Cache

We discussed the cache earlier in *Chapter 8*, *Email Forensics – Investigation Techniques*, and we still have the same issue in that when we examine the content, it is difficult to decipher. There is a third-party tool called Chrome Cache View (available at `www.nirsoft.net/utils/chrome_cache_view.html`) that converts the data into a readable format. The following screenshot is an example of the output you may see:

gmail.html	https://www.google...	text/html	0	8/28/2019 15:17	8/28/2019 15:17			sffe	HTTP/1.1 302			private		172.217.14.100
s2	https://www.google...	text/javascript	14,685	8/28/2019 15:17	8/27/2019 14:17	8/27/2019 14:27	8/26/2020 14:47	sffe	HTTP/1.1 200	br	data_3 [253952]	public, max-age=31536000		172.217.14.100
about.html	https://mail.google...	text/html	0	8/28/2019 15:17	8/27/2019 21:40		8/28/2019 21:40	sffe	HTTP/1.1 301			public, max-age=86400		172.217.11.165
about.html	https://www.google...	text/html	0	8/28/2019 15:17	8/28/2019 04:21		8/29/2019 04:21	sffe	HTTP/1.1 301			public, max-age=86400		172.217.14.100
about.html	https://www.google...	text/html	15,504	8/28/2019 15:17	8/28/2019 15:17	7/19/2019 00:30	8/28/2019 15:17	sffe	HTTP/1.1 200	gzip	data_3 [303104]	private, max-age=3000		172.217.14.100

Figure 9.11: Cache view

When you select the items of interest, Chrome Cache View allows you to export the information into a much easier to read format, as follows:

```
Filename          : gmail.html
URL               : https://www.google.com/gmail
Content Type      : text/html
File Size         : 0
Last Accessed     : 8/28/2019 15:17
Server Time       : 8/28/2019 15:17
Server Last Modified:
Expire Time       :
Server Name       : sffe
Server Response   : HTTP/1.1 302
Content Encoding  :
Cache Name        :
Cache Control     : private
ETag              :
Server IP Address : 172.217.14.100
URL Length        : 28 ================================================
```

This is much easier to read and understand. You are still being shown the same information, but the presentation is much better.

You have the filename and the date/time they visited the web page. Of interest is the server IP address. If you find data within the cache, such as illicit images, this may lead you to the server where the user hosted the original images.

Now, we will discuss passwords and how they are stored within the Chrome browser.

Passwords

Passwords can be key to unlocking different files or encryptions. Most users will reuse the same passwords for multiple different accounts. Your ability to recover the user's previously used passwords can lead to a treasure trove of information. Chrome has the option for a user to save passwords.

You will find the password information in the `Logon Data` file, which can be found at the following path:

```
%USERS%/AppData/Local/Google/Chrome/User Data/Default
```

The file will not have a file extension, and it is a SQLite database. It does not contain the actual password for the user accounts; instead, it stores information about each account, which is used to encrypt the passwords. There is a third-party utility, Chrome Pass (available at `https://www.nirsoft.net/utils/chromepass.html`), which will decrypt the passwords.

The Google Chrome browser is the current number one browser used today. You may encounter multiple browsers on a user's system. The next browser we will talk about is Internet Explorer, the default browser for the Microsoft Windows operating system.

Exploring Internet Explorer/Microsoft Edge (Old Version)

Internet Explorer is the web browser of the Microsoft Windows operating system. Microsoft has included the browser with Windows since 1995. Internet Explorer was the number one browser during the 1990s, but with the release of Firefox in 2004 and Google Chrome in 2008, its popularity has dropped. Internet Explorer is still included with Windows 10 but has been replaced with Microsoft Edge.

The following artifacts are for the older version of Microsoft Edge as the current version is based on Chromium. The Chromium version will have artifacts based on Google Chrome.

The default location of the artifacts for the Chromium-based Edge browser is:

```
C:\Users\%USER%\AppData\Local\Microsoft\Edge\User Data\Default
```

You will find artifacts common to Chromium-based browsers, such as bookmarks (JSON file), cache files (Index, Datablock, and Data files), cookies (SQLite database), and history (SQLite Database). So, let's jump in and look at the artifacts you may recover from the user's activity in the Internet Explorer Edge (Old Version) browser.

Bookmarks

Unlike the Google Chrome browser, Internet Explorer saves bookmarks in a URL format. The default path Internet Explorer keeps the bookmarks in is as follows:

```
%USER%/Favorites
```

All commercial and open-source forensic tools can read the URL format. The following screenshot depicts the typical data structure you will find:

Miniature Schnauzar Dog Breed Information.url	url	2.5 KB	09/02/2019	18:01:11	+0	09/02/2019	18:01:13	+0
schnauzers - Bing images.url	url	1.1 KB	09/02/2019	18:01:30	+0	09/02/2019	18:01:30	+0
Salt and Pepper Miniature Schnauzer - Bing images.url	url	1.3 KB	09/02/2019	18:01:47	+0	09/02/2019	18:01:47	+0
Christen's Miniature Schnauzers - Las Vegas, NV.url	url	180 B	09/02/2019	18:02:11	+0	09/02/2019	18:02:11	+0

Figure 9.12: IE bookmarks

As you examine the bookmarks stored in the `Favorites` folder, you can also view the created and modified timestamps. These are the date/timestamps for when the URL file was created/modified.

When you examine the URL file, you may find content like the following:

```
[DEFAULT]
BASEURL=http://christensminischnauzers.com/
{000214A0-0000-0000-C000-000000000046}
Prop3=19,2
[InternetShortcut]
URL=http://christensminischnauzers.com/IDList=
```

As you can see, it contains the URL, which is the point of interest if the website is dealing with contraband, or the user's activity, which supports your hypothesis about the events you are investigating.

Next, we'll check the user's history and see if we can find any interesting artifacts.

IE history

Internet Explorer will track the user's activity for 20 days. This is the default setting and can be changed by the user. As we discussed in *Chapter 6*, *Windows Artifact Analysis,* Internet Explorer will also track some user activity in the operating system. Internet Explorer is an integral part of the Windows operating system, and even if the user prefers a different browser, there still may be artifacts of interest within the Internet Explorer history.

Edge and Internet Explorer version 10 and higher use an ESE database called `WebCacheV01.dat` that can be found at the following path:

```
%User%\AppData\Local\Microsoft\Windows\WebCache
```

To analyze the `WebCacheV01.dat` file, we will use `ESEDatabaseView` (we first used it in *Chapter 6*, *Windows Artifact Analysis*). Export the database from the forensic image to analyze it.

The first table you will want to look at is the `Containers` table, and you should see something similar to what is depicted in the following screenshot:

ContainerId	LastAccessTime	Name	Directory
1	132119207925900830	Content	C:\Users\IEUser\AppData\Local\Microsoft\Windows\Temporary Internet Files\ Content.IE5\
2	132115040805283385	feedplat	C:\Users\IEUser\AppData\Local\Microsoft\Feeds Cache\
3	131594261121527040	ietld	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\IETldCache\
4	132119207924265464	History	C:\Users\IEUser\AppData\Local\ Microsoft\Windows\ History\ History.1E5\
5	132119207926189424	Cookies	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\Cookies\
6	132119207925419840	iecompat	C:\ Users\IEUser\AppData\Roaming\Microsoft\Windows\IECompatCache\
7	132119207925516038	iecompatua	C:\ Users\IEUser\AppData\Roaming\Microsoft\Windows\iecompatuaCache\
8	132119207913683684	DNTException	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\DNTException\
9	132119207925131246	EmieSiteList	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\ EmieSiteList\
10	132119207925131246	EmieUserList	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\ EmieUserList\
11	132119207944659440	DOMStore	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\ DOMStore\
12	132119207959858724	MSHist012019082820190829	C:\Users\ IEUser\AppData\Local\ Microsoft\Windows\History\ History.1E5\MSHist012019082820190829\
13	132115041147334574	iedownload	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\IEDownloadHistory\
14	132119207959762526	MSHist012019082920190830	C:\Users\IEUser\AppData\Local\Microsoft\Windows\History\History.1E5\MSHist012019082920190830\
15	132119207959762526	MSHist012019082620190902	C:\Users\IEUser\AppData\Local\Microsoft\Windows\History\History.1E5\MSHist012019082620190902
16	132119207959954922	MSH1st012019090220190903	C:\Users\IEUser\AppData\Local\Microsoft\Windows\History\History.1E5\MSHist012019090220190903\

Figure 9.13: ESE database showing the Containers table

There are 16 tables in the database, and the containers of interest are tables 12, 14, 15, and 16. When you look at the names of the tables, they start with `MSHist01`, followed by numbers.

The numbers tell us if the history file is a daily or a weekly file. In the following diagram, you can see the breakdown of the table name:

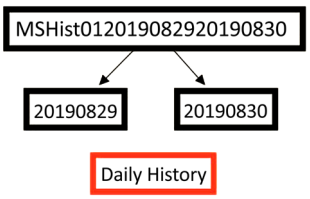


Figure 9.14: The Daily history folder naming convention

The name contains the period of time the history covers. As shown in the preceding diagram, the data it contains spans from August 29, 2019, to August 30, 2019. In the following diagram, we can see the breakdown of a weekly timeframe:

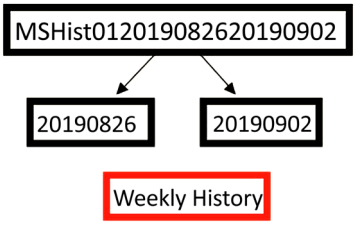


Figure 9.15: Weekly history naming convention

This data spans the time period from August 26, 2019, to September 2, 2019.

For the `MSHist01` files, the file path shown under the `Directory` field is for legacy purposes. If you follow the path shown, you will find a file called `container.dat` that does not contain any information.

Note

For the rest of the entries, the file path will contain data corresponding to the specific table and may be relevant to your investigation.

Let's look at the contents of table `12`, as shown in the following screenshot:

EntryId	SyncTime	ExpiryTime	ModifiedTime	AccessedTime	Url
1	132115734791679663	132138198789360361	132115482789355131	132115734791679663	:20190829201908 30: IEUser@file:///C:Program%20Files/Windows%20Mail/MSOERES.dlll
2	132115734793861687	0	132115482789355131	132115734793861687	:20190829201908 30: IEUser@:Host: Computer
3	132115735348103202	132138195053033732	132115483347995798	132115735348103202	:20190829201908 30: IEUser@file:///C:/Users/IEUser/Downloads/EnableWinMailWin7/msoe_64.zip
4	132115735669898689	132138195374936623	132115483669890000	132115735669898689	:20190829201908 30: IEUser@file:///C:/Program%20Files/Windows%20Mail/msoe.dlll
5	132115736325813786	132138196030768150	132115484325730216	132115736325813786	:20190829201908 30: IEUser@file:///C:/Users/IEUsed/Downloads/EnableWinMailWin7/msoe_32.zip

Figure 9.16: Contents of table 12

Table `12` is a daily history file, and it shows five entries for that day. When we look at the URL, it is not showing up as internet history but is depicting the files the user accessed with File Explorer. The user accessed the `Windows Mail` folder and the `Download` folder. It lists the specific filenames at the end of the path.

The date and time values are a decimal conversion of the hexadecimal Windows 64-bit (Big Endian).

To convert the values, you will need to do the following:

- Take the decimal number, `132115734791679663`
- Convert it into hex, `1D5 5E8F 917F 6EAF`

Then, you need to use DCode to get the date/timestamp:

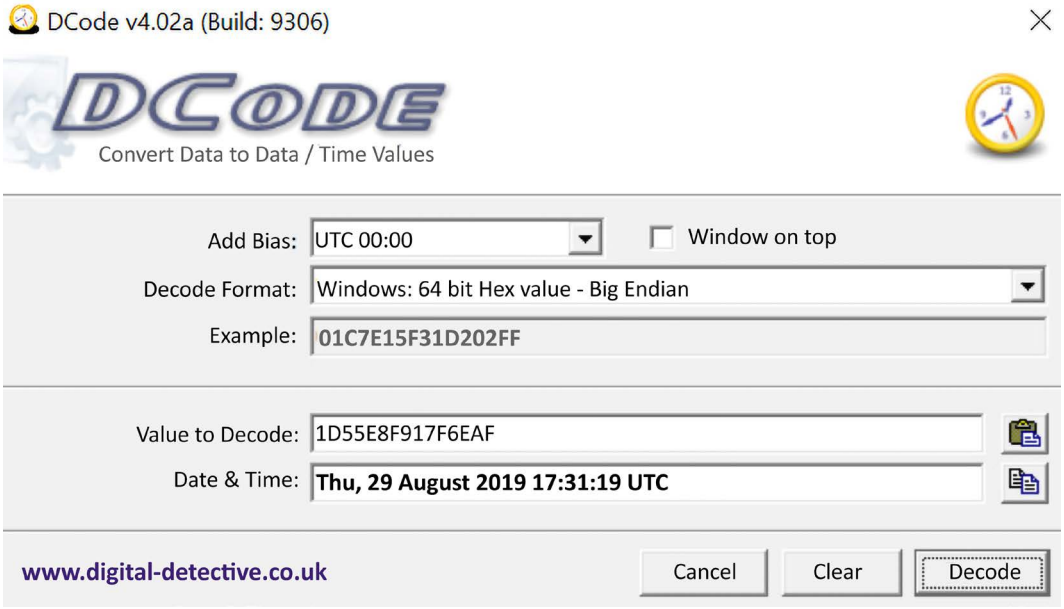


Figure 9.17: DCode tool used to convert the Windows Time value

We have the user's history, but does it show that the user willingly and knowingly visited the sites listed in the history? Is it possible there are references to websites the user never visited? The answer to those questions is yes. With popups and ads, it is possible to have a reference in the history file that the user never visited.

To help show the intent of the user, we want to see an artifact showing the user's actions explicitly. That is the next artifact we will talk about – the typed URL.

Typed URL

When the user types a URL into the address bar, a record is created in the user's `NTUSER.dat` file. The following output is from `RegRipper`:

```
TypedURLs
Software\Microsoft\Internet Explorer\TypedURLs
LastWrite Time Tue Sep 3 17:29:58 2019 (UTC)
url1 -> http://bankrobbery.com/
url2 -> http://yahoo.com/
url3 -> http://gmail.com/
```

The most recent typed URL is `url1`. The system will only list each URL once. If the user enters the same URL, the system will move the URL to the top of the list to become the most recent URL. With Internet Explorer version 10, the maximum number of URLs is 50.

There is a registry key of typed URLs time; see `TypedUrlTime` in the following screenshot:

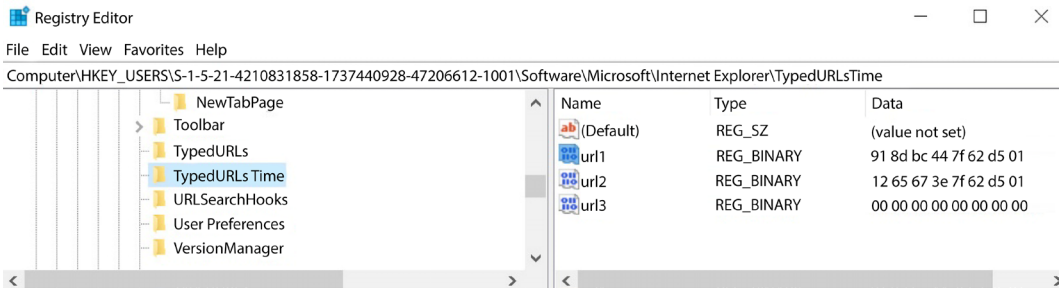


Figure 9.18: TypedURLsTime registry entry

The URL number corresponds to the same value in typed URLs. The hexadecimal value is in the Windows file time format; it represents the date and time when the user entered the URL into the address bar.

Another source of information is the cache, which we will discuss next.

Cache

The `WebCacheV01.dat` file we analyzed in the *IE history* section also handles the cache files. You can use `ESEDatabaseViewer` to analyze the database, but there is another option you can use called Internet Explorer Cache Viewer (available at `https://www.nirsoft.net/utils/ie_cache_viewer.html`).

The following screenshot (cache view) shows the output of the viewer:

Filename	Content Type	URL	Last Accessed	Last Modified	Expiration Time
acquire-80[1].png	image/png	https://f6ef4eacbe624ae1083a-b3d937de523d4a3...	9/2/2019 11:27	12/5/2018 13:05	9/2/2019 11:42
update_2_19_0_1...	text/html	https://f6ef4eacbe624ae1083a-b3d937de523d4a3...	9/2/2019 11:27	8/22/2019 09:59	9/2/2019 11:42
AAGEZpS[1].jpg	image/jpeg	https://static-global-s-msn-com.akamaized.net/i...	9/2/2019 11:23	9/2/2019 04:57	9/7/2019 04:56
AAesHLQ[1].png	image/png	https://static-global-s-msn-com.akamaized.net/i...	9/2/2019 11:23	8/30/2019 00:28	9/4/2019 00:28
AAGHCg4[1].png	image/jpeg	https://static-global-s-msn-com.akamaized.net/i...	9/2/2019 11:23	9/2/2019 10:27	9/7/2019 10:27

Figure 9.19: The output of cache view

The tool will give you the filename and the URL of where the file came from, along with the date/timestamps. The system stores these files in the following path(s):

- For a Windows 7-based system:

```
%USER%/AppData/Local/Microsoft/Windows/Temporary Internet Files\
Content.IE5
```

```
Temporary Internet Files
├──Content.IE5
│   ├──OPDYBC4P
│   ├──S97WTYG7
│   ├──Q67FIXJT
│   ├──4MNQZMD8
│   ├──SCD1EGFC
│   ├──34UZLM61
│   ├──V2I5AL1G
│   └──5S4OGUTD
```

- For a Windows 8/10-based system:

```
%USERS%/AppData/LocalLow/Microsoft/Windows/AppCache
```

```
Windows
    └──AppCache
        └──0Z1ZMDEH
```

```
%USERS%/AppData/Local/Microsoft/Windows/INetCache/IE
```

```
INetCache
│   ├──Low
│   │   ├──IE
│   │   │   ├──4TENJ512
│   │   │   ├──9SKPYC9A
│   │   │   ├──QMIGA2MM
│   │   │   └──EP19S3JV
```

- For the Microsoft Edge browser:

```
%USER%/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/
AC
```

```
MicrosoftEdge
|   |   ├
|   |   ├──Cache
|   |   |   ├──9CG3K1S3
|   |   |   ├──IHCXX8UB
|   |   |   ├──OOW222LO
|   |   |   └──CENY1YGT
```

The system will randomly generate the naming of the subdirectories using alphanumeric characters.

The next artifact we will discuss is cookies.

Cookies

Edge and Internet Explorer save the cookie files as simple text files. `WebCacheV01.dat` also tracks the cookie files, as shown in the following screenshot:

ContainerId	LastAccessTime	Name	PartitionId	Directory
1	132119207925900830	Content	M	C:\Users\IEUser\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\
2	132115040805283385	feedplat	M	C:\Users\IEUser\AppData\Local\Microsoft\Feeds Cache\
3	131594261121527040	ietld	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\IETldCache\
4	132119207924265464	History	M	C:\Users\IEUser\AppData\Local\Microsoft\Windows\History\History.IE5\
5	132119207926189424	Cookies	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\Cookies\
6	132119207925419840	iecompat	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\IECompatCache\
7	132119207925516038	iecompatua	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\iecompatuaCache\
8	132119207913683684	DNTException	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\DNTException\
9	132119207925131246	EmieSiteList	M	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\EmieSiteList\
10	132119207925131246	EmieUserList	M	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\EmieUserList\
11	132119207944659440	DOMStore	M	C:\Users\IEUser\AppData\Local\Microsoft\Internet Explorer\DOMStore\
12	132119207959858724	MSHist012019082820190829	M	C:\Users\IEUser\AppData\Local\Microsoft\Windows\History\History.IE5\MSHist012019082820190829\
13	132115041147334574	iedownload	M	C:\Users\IEUser\AppData\Roaming\Microsoft\Windows\IEDownloadHistory\

Figure 9.20: Content of the Containers table

Table 5 contains information about the cookies and is displayed in the following screenshot:

EntryId	AccessCount	SyncTime	CreationTime	ExpiryTime	ModifiedTime	AccessedTime	Url	Filename
36	18	132119208683123513	132119208683098265	132435432680000000	132119208683098265	132119208683123513	Cookie:ieuser@yahoo.com/	IF0DA7EK.txt
41	2	132115040834588257	132115040834204478	132452000840000000	132115040834204478	132119208375537727	Cookie:ieuser@www2.bingcom/	QZVGRJVN.txt
21	17	132119208820850307	132119208820850307	132460487960000000	132119208820850307	132119222099087060	Cookie:ieuser@www.msn.com/	Q01C9WT2.txt
47	3	132115044434921485	132115044434921485	132115908430000000	132115044434921485	132115044436696825	Cookie:ieuser@www.mozilla.com/	YP2ZL4QD.txt
45	4	132115040938396120	132115040938396120	132192800940000000	132115040938396120	132119208561572465	Cookie:ieuser@www.google.com/	U095JS89.txt
88	1	132119208810386393	132119208810378061	132750792850000000	132119208810378061	132119208810386393	Cookie:ieuser@www.bing.com/images	6MC65DME.txt
38	6	132115040768554247	132115040768554247	132452000760000000	132115040768554247	132119222095615988	Cookie:ieuser@www.bing.com/	IT9CA013.txt
60	20	132119208680379897	132119208680379897	133696008670000000	132119208680379897	132119208686590905	Cookie:ieuser@www.akc.com/	0W6YLVUJ.txt
24	8	132119208661351321	132119208661342905	132461352650000000	132119208661342905	132119208661351321	Cookie:ieuser@w55c.net/	XTF8XNNX.txt
50	1	132119208020635033	132119208020550601	132434568060000000	132119208020550601	132119208020635033	Cookie:ieuser@tvpixel.com/	QLCX0KXB.txt

Figure 9.21: Content of the Cookies table

Just like when we examine the history, the date/timestamps are decimal conversions of the hexadecimal Windows file time. It also contains the URL and the filename being stored on the system.

The cookie files are stored in the following path(s):

- For Internet Explorer:

```
%USER%/AppData/Roaming/Microsoft/Windows/Cookies/
```

- For Microsoft Edge:

```
%USER%/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/
AC/MicrosoftEdge/Cookies
```

When we look at the contents of the folder, we find that it is full of text files, with the filenames of the random alphanumeric character format that Windows uses. The following output is typical of what you may find:

```
Name              Created                  Modified
06PC9CZM.txt    09/03/2019  17:29:48    09/03/2019  17:29:48
09BHTXJM.txt    09/02/2019  18:00:59    09/02/2019  18:00:59
09WSNIHD.txt    09/03/2019  17:29:27    09/03/2019  17:29:27
0W6YLVUJ.txt    09/02/2019  18:01:08    09/02/2019  18:01:08
0WBQAB4E.txt    09/02/2019  18:23:51    09/02/2019  18:23:51
16SUYNBJ.txt    09/03/2019  17:29:51    09/03/2019  17:29:51
1983DVP6.txt    09/02/2019  18:23:46    09/02/2019  18:23:46
28Z2GM8G.txt    09/03/2019  17:29:49    09/03/2019  17:29:49
2CM18GNC.txt    09/03/2019  17:29:38    09/03/2019  17:29:38
```

For an effective analyzation, you will have to work through the `WebCacheV01.dat` file to determine which cookie file is associated with each entry in the database.

The following is an example of the content of a cookie text file:

```
MR
0
c.msn.com/
1024
3308281856
30796639
4095225949
30760429
```

This cookie is from the MSN website, and as we discussed previously, it tracks the user's visits and any of the preferences that may be enabled at the time of the visit.

Examining the cookies of the browser will never be the smoking gun, but you could use it to support your hypothesis regarding what occurred during your investigation.

Exploring Firefox

Firefox is an open-source browser developed by the Mozilla foundation. Mozilla released Firefox in 2004 and is a browser you may encounter during your investigations. We will cover some more common artifacts that you may encounter during your examination.

Profiles

One feature offered by Firefox is the use of multiple profiles. A user has the option to create multiple profiles for the browser to segregate their activity. The path where you can find the profiles is as follows:

```
%USER%/AppData/Local/Mozilla/Firefox
```

I have found three user profiles, and they are displayed as follows:

```
Firefox
└──Profiles
   ├──tszci9zh.Badguy
   │  ├──thumbnails
   │  ├──safebrowsing
   │  │  └──google4
   │  ├──startupCache
   │  ├──cache2
   │  │  ├──entries
   │  │  └──doomed
   │  └──OfflineCache
```

This is the `Badguy` profile:

```
   ├──fd8rnyou.BadGuy Needs Love
    │  ├──startupCache
    │  ├──cache2
    │  │  ├──entries
    │  │  └──doomed
```

```
    │  ├──thumbnails
    │  ├──safebrowsing
    │  │  └──google4
    │  └──OfflineCache
```

The user of this profile is `BadGuy Needs Love`:

```
    ├──30nh3g6c.default-release
    │  ├──startupCache
    │  ├──cache2
    │  │  ├──doomed
    │  │  └──entries
    │  ├──thumbnails
    │  ├──OfflineCache
```

The final profile is the default user.

Firefox creates the profile with a random alphanumeric eight-digit prefix, followed by the username. If the user has not created any additional profiles, you will only see the default-release and default username. Here, the user has created two additional profiles:

- `Badguy`
- `Badguy Needs Love`

Inside each folder structure, Firefox will save data appropriate to each profile.

There is also a `profiles.ini` file, which can be found at the following path:

```
%USER%/AppData/Roaming/Mozilla/Firefox/profiles.ini
```

The contents of the `profile.ini` file are as follows:

```
[Install308046B0AF4A39CB]
Default=Profiles/fd8rnyou.Badguy Needs Love
Locked=1

[Profile2]
Name=Badguy
IsRelative=1
Path=Profiles/tszci9zh.Badguy
Default=1
```

```
[Profile1]
Name=default
IsRelative=1
Path=Profiles/9wofgs9f.default


[Profile0]
Name=default-release
IsRelative=1
Path=Profiles/30nh3g6c.default-release

[General]
Startwithlastprofile=1
Version=2

[Profile3]
Name=Badguy Needs Love
IsRelative=1
Path=Profiles/fd8rnyou.Badguy Needs Love
```

The `Startwithlastprofile` field shows which profile will start when the application starts. Here, it shows that the `BadGuy` profile is the default profile.

We will now move on and look at the cache.

Cache

Firefox stores the cache files under each profile. The file path will remain the same, as we discussed in the previous section:

```
%USER%/AppData/Local/Mozilla/Firefox/Profiles/%Profile%
```

```
Firefox
└──Profiles
   ├──tszci9zh.Badguy
   │  ├──thumbnails
   │  ├──safebrowsing
   │  │  └──google4
   │  ├──startupCache
```

```
    │  ├──cache2
    │  │  ├──entries
    │  │  └──doomed
    │  └──OfflineCache
```

Firefox will store the cache in the `cache2` profiles. You can use the open-source application `MZcacheview` (available at `https://www.nirsoft.net/utils/mozilla_cache_viewer.html`) to view the contents. The results are very similar to what we have seen in the previous browsers.

Cookies

Unlike Internet Explorer, Firefox does not save the browser cookies in single files. Firefox uses a SQLite database to store this information. You can find the cookie database at the following path:

```
%USER%/AppData/Roaming/Mozilla/Firefox/Profiles/%Profile%
```

Note the change in the path. Instead of being stored in the `Local` folder, we are now in the `Roaming` folder.

You can use the third-party open-source application MZCookiesView (available at `https://www.nirsoft.net/utils/mzcv.html`) to view the file or any SQLite database reader.

The next artifact we will discuss is the history file.

History

Mozilla Firefox tracks the browser history in the SQLite database file called `places.sqlite`. Firefox also tracks the users' typed URLs in this database. You can use any SQLite database tool to read the file, or you can use the third-party open-source MZHistoryView (available at `https://www.nirsoft.net/utils/mozilla_history_view.html`). You can find the history database at the following path:

```
%USER%/AppData/Roaming/Mozilla/Firefox/Profiles/%Profile%
```

The following screenshot shows the typical output you would see with this tool:

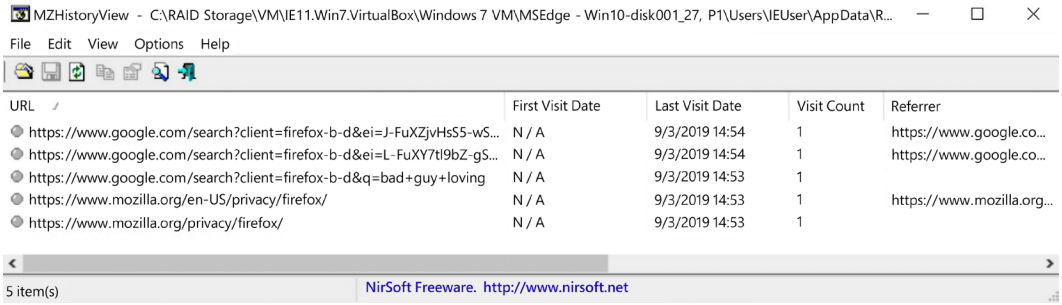


Figure 9.22: Firefox history is shown in MZHistoryView

Note

The preceding screenshot of a Firefox history file only shows a few entries. If this was an actual investigation, depending on how long they used the Firefox browser, you could find thousands of entries in the history file.

An example of the typed URL is in the third record. If you double-click on the entry, it will bring up the `Properties` window, as shown in the following screenshot:

Figure 9.23: Typed URL is shown in record 3

From the preceding screenshot, we can determine that the user did a Google search for the phrase `bad guy loving` on September 3, 2019.

Passwords will be the next artifact we will discuss.

Passwords

Mozilla Firefox offers users the opportunity to save their passwords. Firefox uses two files, `key#.db` (I have seen files named `key3` and `key4`; be aware you may come across additional numbers) and `logins.json`, to store the passwords in an encrypted format. We can decode the passwords using the open-source third-party tool Password Fox (available at `https://www.nirsoft.net/utils/passwordfox.html`).

You can find the files at the following path:

```
%USER%/AppData/Roaming/Mozilla/Firefox/Profiles/%Profile%
```

In the following screenshot, you can see the typical output of Password Fox:

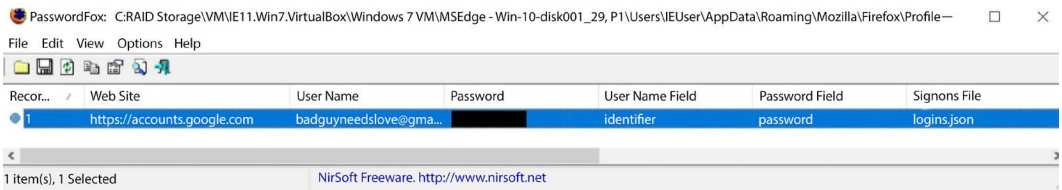


Figure 9.24: Password shown in PasswordFox

I have redacted the password in the following screenshot, but it is beneficial when you gain access to the user's passwords and their accounts (with proper authorization, of course). If you double-click on the record, it will bring up the `Properties` window, as displayed in the following screenshot, showing the password properties:

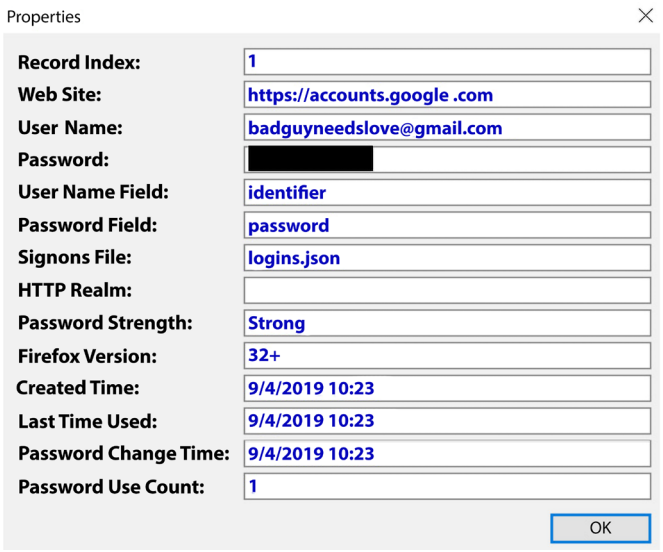


Figure 9.25: Password properties in Password Fox

As we analyze the content, it appears this is the user's Gmail account, and we have the date/timestamps for when the password was created, changed, and used.

Bookmarks

Mozilla Firefox saves the user's bookmarks in a SQLite database file. You can find the database file at the following path:

```
%USER%/AppData/Roaming/Mozilla/Firefox/Profiles/%Profile%
```

You can use the third-party open-source tool FavoritesView (available at `https://www.nirsoft.net/utils/faview.html`) to do this.

The following screenshot shows the output of the FavoritesView application:

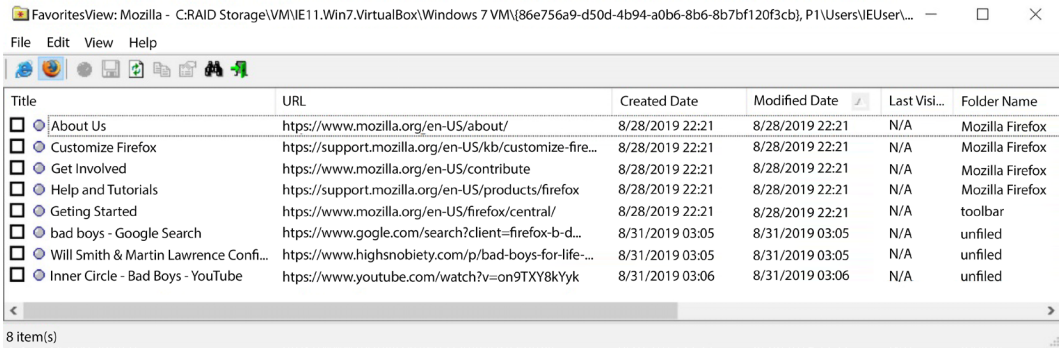


Figure 9.26: Favorites are shown in FavoritesView

You can see the default bookmarks that are included with the Firefox browser. The last three entries are the bookmarks that users have added. We have a Google search for the term `bad boys`, a web page referencing the actors Will Smith and Martin Lawrence, and the final entry is a YouTube video from the music group Inner Circle for their video Bad Boys.

That completes our analysis of the browsers. We will start looking at social media in the next section.

Social media

What is social media? Social media is the use of applications or programs to create and share information, forms of expression, opinions, ideas, and so on through virtual communities on the global internet. Users can access social media through web-based technology such as a mobile device application. In some situations, the user can sync data from one platform to another. These platforms/applications rarely require a fee from the user and are very simple to use.

While most social media users use the services in a manner the service provider intended, some use these new communication media for nefarious purposes. It is a very unusual investigation where social media does not play a part in the investigation.

The user's social media communication leaves a digital trail of breadcrumbs for the investigator to follow. Sometimes, the investigator can determine the user's location at the date/time when the incident being investigated occurred. Alternatively, they may find communications between the suspect and the victim that led to the incident being investigated.

As a digital forensic investigator, you will have to be aware of the existence of social media and what potential artifacts exist in digital evidence. A significant challenge when searching social media is that the majority of the social media artifacts will not be saved to the user system.

The applications store the data in the service provider's cloud, which is a fancy way of saying the data will be on the service provider's servers.

Because of the vast diversity and sheer amount of social media applications, there is not a simple checklist that will cover all situations. The digital forensic investigator must be flexible in their investigation techniques when dealing with new social media technology or changing social media technology.

The goal of this section is to familiarize you with some current social media applications and to provide you with a general plan for your investigation. Remember, when conducting an analysis dealing with social media, there may be two locations for you to find digital evidence relating to your investigation: the user system and the service provider. Do not neglect to serve the appropriate judicial paperwork on the service provider; typically, they will give you extensive information regarding your investigation.

The popularity of specific social media ebbs and flows as the users' demographic changes. Some younger users are disinclined to use Facebook with their friends because their parents and grandparents also use Facebook. Some social media applications may be restricted to geographical locations; for example, KaKaoTalk is very popular in the Republic of Korea but has very few users in the United States. The following is a short description of some popular social media applications that you may come across during your investigations:

- Facebook: The most extensive social media application with nearly 2,000,000,000 users. It combines commercial advertisers and consumer users with their content. It is easy to use. Users can upload images, recorded videos, live videos, and voice/video/text chat via the messenger application.
- Instagram: A photo and video sharing social media application. Users can share photos, recorded videos, and live videos. Users can also chat and comment with other users via the app.
- Snapchat: A video/photo-sharing social media application. Initially, when a user sent a "snap" after the recipient viewed the image, it would be deleted from the system. Now, users have the option of saving "snaps."
- Twitter: A social media application used for news, politics, sports, entertainment, and so on. Twitter allows 280-character tweets, with longer tweets being linked in the following messages.
- WhatsApp: A messaging application that allows users to engage in voice/video chat.
- Tinder: A location-based social media application used as a dating service. Users "swipe right" or "swipe left" on profiles they like/dislike. If both parties "like" each other, then they can chat using the app.

- GroupMe: This is a group messaging application. Users can use their cell phone number or their Facebook or Twitter account to log in to the app. Users can share photos, videos, locations, and text.
- Kik: An instant messaging social media application. The service provider is based in Canada. It allows anonymous communication between individuals. Users can share text, photos, and videos. It is estimated that nearly 40 percent of teenagers in the United States use the application. Kik was acquired by MediaLab in 2019.
- Tumblr: A blogging/social networking application. Users can post pictures/videos to a blog page.
- Reddit: This is a news aggregation and discussion social media application. It contains almost every topic imaginable, including illegal activities. In July 2019, it was the number five most visited website in the United States.
- TikTok: A video-based social networking application. Users can post videos and have "live" streaming events. TikTok was the third fastest-growing social media brand in 2020.

This is not a complete list of social media applications you may encounter, nor would I attempt to create such a list. Social media applications will change as technology changes, and as user choices grow, new social media applications will come to the forefront. As a digital forensic investigator, it will be impossible for you to know all the facets of social media applications. You need to understand what social media applications are being used and how consumers and criminals use them. Some social media applications will not have the means to be accessed via the localhost; the service provider will restrict them to a mobile device only. How can you determine if the subject of your investigation is using a social media platform? The web browser history will be essential; they may access their profile/account via a web browser or email communication between the user and the service provider. One of the more popular social media applications you may come across is Facebook.

Facebook

Facebook is a social media platform that a user can access via the web browser. Analyzing those artifacts may give you the username or the Facebook user ID. Let's say you analyze an URL, such as the one displayed here:

```
https://www.facebook.com/photo.php?fbid=10215539711464494&set =a.1627301761019&t
ype=3&source=11&referrer_profile_id=1190817474
```

As you examine the URL, the portion relevant to our organization is the profile ID, which is as follows:

```
profile_id=1190817474
```

When you add the numbers to the end of `Facebook.com`, it will take you to the user's Facebook page. The profile ID is a unique number and will redirect you to the user's Facebook username, as shown in the following screenshot:

Figure 9.27: Facebook URL

Once you have obtained the user's Facebook profile ID number and username, that information can then be given to the service provider as part of the judicial paperwork. This can be used to access the content being stored on the service provider's servers.

Another tool that you can use is Bulk Extractor (which we discussed in *Chapter 7*, *RAM Memory Forensic Analysis*). When you run Bulk Extractor, it will also find the profile ID numbers, as shown in the following screenshot:

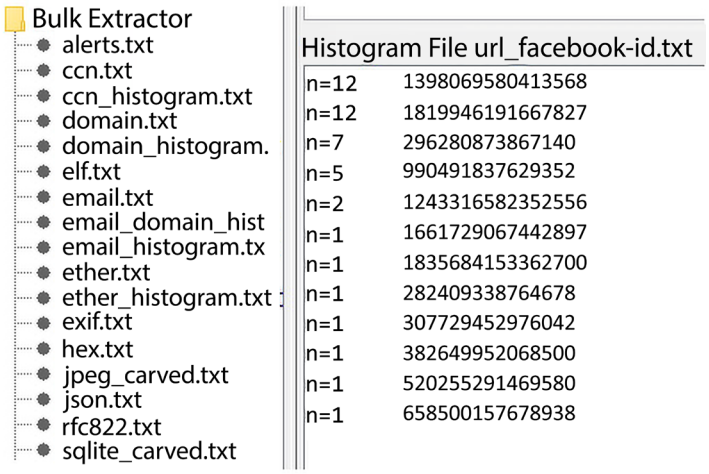


Figure 9.28: Bulk Extractor output for Facebook

You can run Bulk Extractor against the forensic image or a captured memory file to find artifacts relating to social media. Another very popular social media application you may come across is Twitter.

Twitter

We can filter the results from the domain histogram, as shown in the following screenshot. Here, we can see the user accessed the Twitter web page:

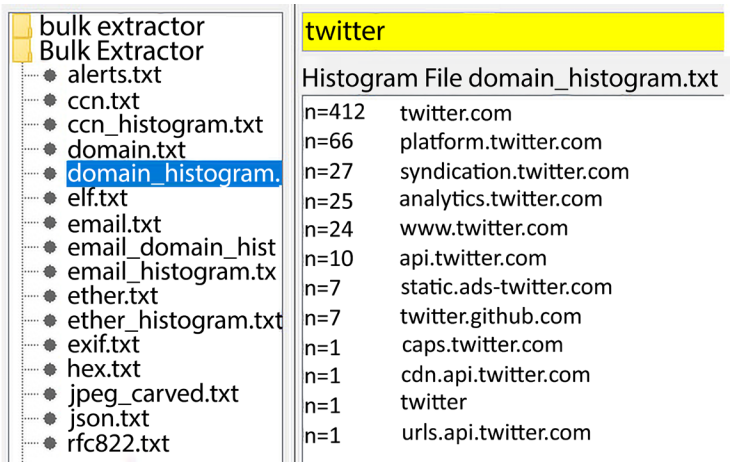


Figure 9.29: Bulk Extractor output for Twitter

Twitter users have a "handle" and a UID (user ID) when they sign up for the service. Users can change their handle whenever they desire, but the UID will remain the same. If we have a user with the handle of `@badguyneedslove` and then they change it to `@badguy27`, the UID will stay the same.

Searching for the term `twid` in the forensic image will identify the UID for the account. You can also take the user's Twitter handle and use the following website: `http://gettwitterid.com`.

This will give you the UID when you input the Twitter handle, as shown in the following screenshot:

Figure 9.30: Twitter ID

We entered the Twitter handle of `badguyneedslove`, and it gave us the Twitter ID of `1170432764291665920`, as well as the full name of the user (which may or may not be accurate; remember this is user-supplied information).

Not all the information you want or need for your investigation will be found on the user's device; you may have to reach out to the service provider to get the information that is being stored on the service provider's servers.

Service provider

Much of the information you need for your investigation will be in the care and custody of the service provider. They will have the subscriber information such as the name, address, age, usage dates/times, and IP addresses. This content is hosted on the service provider's servers. This seized computer system or mobile device may not have all the information you want as you conduct your investigation. Serve the service provider with the appropriate judicial paperwork to get that information. The judicial framework will be based on where the service provider is located. You will have to meet all the requirements of the service provider's judicial system.

Search.org maintains a service provider list and the contact information of the legal department that will receive the completed judicial paperwork. Some service providers also provide that information on their website. For example, the service provider Kik has created a specific web page (located at `https://web.archive.org/web/20201224090043/https://lawenforcement.kik.com/hc/en-us`) that contains all the information needed by the investigator to serve them with additional paperwork.

Investigating a user's activity on social media can be difficult. You may have to use third-party software or manually parse through the hexadecimal data stored on the storage device. Even with all that effort, it may come down to the retention policies of the service provider for you to get the information you need to make your investigation successful. Another aspect of the internet that has the potential for criminal use is P2P file sharing, which is our next topic.

P2P file sharing

P2P file sharing allows users to share files with others in the P2P community. Users will share videos or music files with the community, but you can find almost any file type you can imagine. P2P has legitimate and illegitimate uses, depending on the user's search criteria. It is a popular method of sharing illicit images and videos with other users in the P2P community. There are several P2P applications the user can choose from. I cannot give a detailed analysis of all potential P2P applications, but we will discuss some more common P2P applications you may run into during your investigations.

P2P applications allow the user to become a node on the network. When the user installs the application, they can designate which files/folders they want to make available to the P2P network. The application will then create an index of the shared files/folders to share on the P2P network. When the user searches the P2P network and finds a file they wish to download, the application will identify all the nodes possessing that file. The application will then connect to the nodes and start downloading pieces of the file from all the available nodes.

When the P2P application shares the files/folders, it tracks the filename and the file type and creates a SHA-1 hash value for the file. This will be a variation of the SHA-1 hash value used by the commercial and open-source forensic tools. The P2P version of the SHA-1 creates a hash value using the Base32 numbering system, while the forensic tools use the Base16 numbering system. The Base16 numbering system uses the alphanumeric characters 0–9 and A–F, while the Base32 numbering system uses the alphanumeric characters A–Z and 2–7. Chris Hurst posted how to use Python to convert Base32 values into Base16 (available at `https://github.com/qbittorrent/qBittorrent/wiki/How-to-convert-base32-to-base16-info-hashes`).

The following is the Python code that Chris Hurst provided:

```
>>> import base64
>>> b32Hash = "WRN7ZT6NKMA6SSXYKAFRUGDDIFJUNKI2"
>>> b16Hash = base64.b16encode(base64.b32decode(b32Hash))
>>> b16Hash = b16Hash.lower()
>>> print (b16Hash)
```

After you spend some time working in the field of digital forensics, you will learn the artifacts you may find depend on the operating system, the P2P application, and if the user has modified any of the default settings.

We will now investigate some of the common P2P applications you may encounter during your digital forensic investigations.

Ares

Ares Galaxy is an open-source P2P application utilizing the decentralized network configuration (available at `https://sourceforge.net/projects/aresgalaxy/`). Ares creates entries in the user's local profile path, as shown here:

```
%USER%\AppData\Local\Ares\
```

In the `Data` folder, you will find two files, `ShareH.net` and `ShareL.dat`. These files track the file-name, the hash value, the date/timestamp of when the file was downloaded, and the sharing status of the file. These files are encrypted but can be decrypted using the Magnet Forensics AXIOM forensic tool (available at `https://www.magnetforensics.com`).

Ares creates entries in the user's `NTUSER.dat` file, as shown here:

```
\ntuser (ROOT)\Software\Ares
```

When we run `RegRipper` against the `NTUSER.dat` file, we get the following output:

```
Software\Ares
LastWrite Time Sat Sep  7 21:48:04 2019 (UTC)
Stats.LstConnect: Mon Sep  8 15:51:07 2019 UTC
Personal.Nickname: Badguy27
General.Language: English
PrivateMessage.AwayMessage: This is an automatic away message generated by
Ares program, user isn't here now.
Search Terms: Badguy movies
```

The application stores the last date/timestamp application connected, the nickname of the user (this can also be an autogenerated field based on the account username of the operating system), and the last 25 search terms entered by the user.

Depending on the version of Ares the user has installed, it may change the `location`/information in the `NTUSER.dat` file.

eMule

eMule is an open-source P2P application utilizing the decentralized network configuration and was released in 2002 as an alternative to eDonkey2000 (available at `www.emule-project.net`). When the user installs eMule, it creates an `eMule` folder, containing two subfolders, `incoming` and `temp`, as shown here:

```
Downloads
└──eMule
   ├──Incoming
   └──Temp
```

As files are being downloaded, the file parts are stored in the `temp` folder, and when all the pieces have been downloaded, the completed file is moved into the `incoming` folder. These folders are shared by default and cannot be disabled by the user.

eMule stores its configuration files in the user's local profile, as follows:

```
%USER%\AppData\Local\eMule
```

In the `config` subdirectory, you will find the `preferences.ini` file. Contained within this file will be the user's nickname and the location of the incoming and temporary directories, as follows:

```
AppVersion=0.50a
Nick=http://emule-project.net
IncomingDir=C:\Users\IEUser\Downloads\eMule\Incoming
TempDir=C:\Users\IEUser\Downloads\eMule\Temp
```

If the user does not specify a nickname, the default nickname is the URL email, that is, `project.net`.

Also of interest will be `shareddir.dat` and `sharedfiles.dat`. `shareddir.dat` will contain the user-created shared directories, as follows:

```
%USER%\Downloads\
```

In this case, the user is also sharing their `Downloads` folder. The `sharedfiles.dat` file will contain a list of files being currently shared. The output may be similar to what is shown here:

```
C:\Users\IEUser\Downloads\aresregular246_installer.exe
C:\Users\IEUser\Downloads\bad-guy-pictures-145577-3671477.png
C:\Users\IEUser\Downloads\eMule0.50a-Installer.exe
C:\Users\IEUser\Downloads\Shareaza_2.7.10.2_x64.exe
```

The file indicates the user is sharing three executables and a PNG image.

In the `preferences.dat` file, you will find the unique identification number assigned to each user in the network. It is a 16-byte hexadecimal value, as shown in the following screenshot:

Offset	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	ANSI ASCII
00000000	14	C2	B6	08	E8	AA	0E	5A	EA	26	35	5C	BB	56	F5	6F	Â¶ è ª Zê&5\»Võo
00000016	4C	2C	00	00	00	00	00	00	00	01	00	00	00	FF	FF	FF	L, ÿÿÿ
00000032	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	FF	0A	00	00	ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
00000048	00	0A	00	00	00	1F	03	00	00	58	02	00	00				X

Figure 9.31: eMule User ID

The user identification number is highlighted. Also, note that in every identification number, in the 6th and 15th bytes, you will find the values `x/0E` and `x/6F`.

The `AC_SearchStrings.dat` file will store the last 30 searched terms entered by the user. In the following screenshot, eMule Search Terms shows the user only searched for the term `charlie chaplin`:

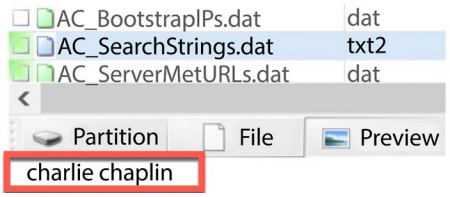


Figure 9.32: eMule Search Terms

The `known.met` file contains a list of files that have been downloaded by the application and files that have been shared by the application. You may find filenames indicative of contraband images that are no longer on the user's system. The application will delete entries as the `.met` file increases in size to prevent the file from becoming too large.

In the following screenshot, you can see the contents of the `.met` file:

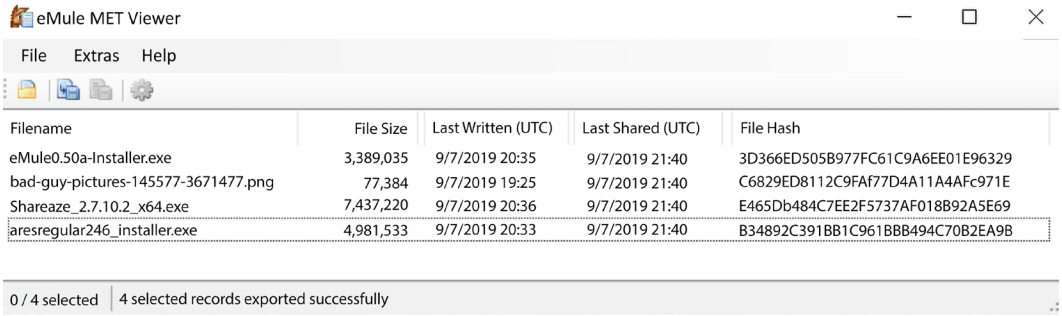


Figure 9.33: MetViewer

The file will contain the filename, file size, and date/timestamps for when this file was shared, along with the hash value of the files. I used the third-party open-source forensic tool eMule MET Viewer (which is available at `https://www.gaijin.at/en/software/emulemetviewer`).

The last P2P application we will look at is Shareaza.

Shareaza

Shareaza is an open-source P2P application that utilizes the decentralized network configuration and was released in 2004 (available at `http://shareaza.sourceforge.net/`). The application will create a `Shareaza` folder in the `Local` and `Roaming` folders of the user's profile. The following is the folder structure(s) you will see in the user profile:

```
%USER%\AppData\Local\Shareaza
%USER%\AppData\Local\Shareaza\Incomplete
%USER%\AppData\Roaming\Shareaza
%USER%\AppData\Roaming\Shareaza\Collections
%USER%\AppData\Roaming\Shareaza\Data
%USER%\AppData\Roaming\Shareaza\Torrents
```

In the `Data` folder, you will find a file called `Profile.xml`, which will contain user-created and application-created artifacts. The user can complete personal information such as their name, location, and gender, which will be included in the XML file.

Shareaza also creates entries in the user's `NTUSER.dat` file. It will create a Shareaza key with many subkeys. In the `Download` subkey, you will find the entries `CollectionPath` and `IncompletePath`. `CollectionPath` is where the completed files will be stored; `IncompletePath` is where the incomplete files are stored.

In the following screenshot, we can see the entries:

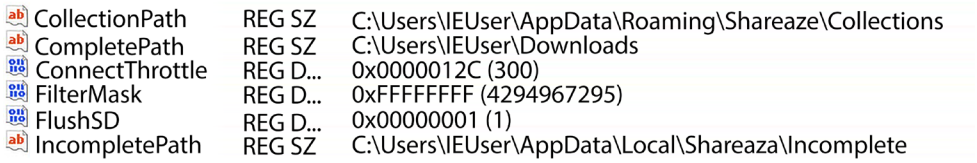


Figure 9.34: Shareaza path

You will also find the search terms inputted by the user in the **Search** subkey, as shown in the following screenshot:

Figure 9.35: Shareaza Search

In the `Data` folder, there is a file called `Library1.dat` that contains a list of shared folders, shared files, and a list of partially downloaded files. There is also a backup of the file, appropriately named `Library2.dat`, which is used if the first file becomes corrupted.

Cloud computing

What is cloud computing? Is it remote storage? Is it a remote server? Is it remote services? The answer to all of the above is yes. Cloud-based services are becoming more popular for businesses and users every day. As a digital forensic investigator, you must know of the potential for cloud-based evidence. We have already discussed some aspects of cloud-based artifacts in this chapter. Now, we will discuss some different elements of cloud-based computing. There are various service models of cloud-based computing you may encounter when conducting your digital forensic investigation. They are as follows:

- **Infrastructure as a Service (IaaS)**: The remote infrastructure is offered to the customer for use, while the provider maintains ownership and control of the hardware. The customer only pays for the hardware/service needed and gives the customer the flexibility to increase/decrease hardware requirements as required.
- **Software as a Service (SaaS)**: Applications are provided to the customer via the network. The customer pays a subscription fee to the vendor to use the software. The content and the user files are stored on the service provider's servers but can be used/shared with other members of the organization.

- **Platform as a Service (PaaS)**: The operating system of the client is provided to the customer via a cloud server. The user can then install their applications and maintain their settings of the software, while the provider manages the hardware and operating system. The client is responsible for the system administration within their network.

Another consideration is the deployment method of cloud resources. There are four to choose from:

- **Public cloud**: A cloud resource that is made available to the public or specific members of an organization. Local government, universities, or a sector of the community can offer a public cloud resource.
- **Private cloud**: A cloud resource that is made available to specific members. The user must have specific rights to access the resource. For example, an organization may maintain a private cloud resource for employees only.
- **Community cloud**: A cloud resource that is similar to a private cloud, where the users comprise multiple organizations with a similar focus. For example, a cloud provider may restrict access to a cloud resource that's used by numerous law enforcement agencies to law enforcement only.
- **Hybrid cloud**: A cloud resource that is made up of two or more different deployment methods.

As you can see, the use of cloud computing can directly affect what artifacts you find or do not find on the local system. It is entirely possible that there will not be any artifacts relating to your investigation on the local system, which then leads to the question, where are the artifacts/evidence?

The answer to that question is anywhere in the world. The data/artifact you may be looking for may be stored on a server located one mile away or several thousand miles away in another jurisdiction. Investigating when the hardware is not physically available creates significant issues for the digital forensic investigator. If you are law enforcement and you have a search warrant, is the search warrant valid for data maintained by the service provider outside of your jurisdiction? For the corporate investigator, a search warrant is not an option, but if the data is stored in a jurisdiction where the privacy expectations differ from where the investigation is based, you may also run into issues with accessing the data.

In the United States, this issue was in dispute until 2018 and would have been decided by the United States Supreme Court. Ultimately, the issue was resolved at the legislative level when Congress clarified the Stored Communications Act and now requires the service provider to provide the requested data if the information "is located within or outside of the United States."

For the corporate investigator, the **service-level agreement (SLA)** should spell out who may access the data and specify if there are to be any limitations when conducting a data acquisition in response to an investigation. The SLA should also address the geographical location of where the data may or may not be stored and address how legal conflicts should be resolved when the data is stored in different jurisdictions.

Different countries provide different protections regarding privacy issues and criminal and/or civil procedures. What may be a crime in one jurisdiction may not be in the jurisdiction where the server containing the data is located. In the **European Union (EU)**, EU citizens must be notified before their personal information is accessed, and they must give their consent.

Once you have been given access to the data that is needed for your digital forensic investigation, you still must deal with the best practices for handling evidence. You may have a chain of custody issues; how do you know if the provider used a forensically sound methodology to collect the evidence? Can you validate the methodology to show that they collected all relevant information? You do not want the opposition in judicial/administrative proceedings to allege that exculpatory evidence was not collected and presented to the fact finder.

When conducting a digital forensic investigation, there are some artifacts you can examine that will show if the user has accessed any cloud-based applications. In this chapter, we have discussed the cache of web browsers and what happens if the user accesses the cloud-based applications with the web browser. In *Chapter 6*, *Windows Artifact Analysis*, we discussed examining the prefetch files. Remember that prefetch files are used to speed up the application's startup and will contain date/timestamps showing when the user last accessed that resource.

Dropbox and Google Drive are two of the most common cloud-based storage options available to the consumer. When the user installs the Dropbox or Google Drive application, the system will create a folder where the user can sync the data to the localhost and the cloud-based storage. As the user changes the local file, the system will then change the file maintained in the cloud-based storage. Alternatively, the user can upload/change a file via a web-based interface using a shared or public computer, and the system will then change the file or the uploaded file available on the user's personal device(s).

When using the Dropbox application, there will be two databases of interest to the digital forensic investigator:

- `config.dbx`: Contains the user ID, account email address, account username, and path for the `dropbox` folder.

- `filecache.dbx`: Contains the `file journal` table, which includes information about the files being synchronized between the localhost and the cloud-based storage. The table will consist of the filename, file path, and file size in the local host ID. The localhost ID is how we identify the host that placed the file into the Dropbox storage.

If you are investigating a user who is using the cloud-based storage Google Drive, the following databases may be of interest:

- `sync_config.db`: This will contain the path of the Google Drive folder on the localhost, show if USB devices are being synced, and show the email account associated with the Google Drive account.
- `snapshot.db`: The `local_entry` table will contain information about the files that have been synced between the localhost and the cloud-based storage. This will include the serial number of the volume, the filename, the modified date/timestamps, the file size, and show if it is a file or a folder.
- The `cloud_entry` table will contain the filename, the modified date/timestamps, the file size, and show if the user shares the file with other users.
- `Device_db.db`: The `external_devices` table will contain the device ID, USB device label, upload date/timestamps, and show if the user has synchronized the device to the cloud-based storage.
- The `device_files` table will contain the device ID of the USB device, the filename of the synced file, the file path, and the date/timestamp of when the file was synchronized to cloud-based storage.

Like most technology, cloud-based computing is rapidly evolving and changing. As the number of consumers who use cloud-based technology grows, so will the number of evidentiary artifacts you will have to examine when conducting your digital forensic investigation. You will always have to be mindful of the potential use of cloud-based computing when conducting your digital forensic investigations.

Summary

In this chapter, we focused on what artifacts may be created by the user as they use a web browser. We covered different web browsers and looked at social media applications. We discussed various social media applications and what artifacts they may leave behind. There may be minimal artifacts left on the user's system, but there may be additional artifacts on the user's mobile device or in the possession of the service provider.

We also went into P2P file sharing and what artifacts this will leave on the user's system. You should now be able to understand and identify the different browsers and artifacts that can be found. You should be able to understand and identify the most popular social media applications and the common locations where artifacts could be found. You should also be able to understand the basic structure and operation of P2P file-sharing networks and be familiar with typical P2P file-sharing client applications and artifacts.

In the next chapter, we will be focusing on report writing. You may be able to find every possible artifact indicating the guilt or innocence of a subject. Still, if you cannot create a report that technical and non-technical readers easily understand, you will lose your audience. If the reader cannot understand your report, they will never know what user activities you uncovered during your investigation.

Questions

1. Google Chrome saves bookmarks in what kind of file?
    a. JSON
    b. Text
    c. URL
    d. XML
2. What file type is the Google Chrome history file?
    a. Word doc
    b. JPEG
    c. SQLite database
    d. XML database
3. Internet Explorer/Edge will save typed URLs in which hive file?
    a. `SOFTWARE`
    b. `SYSTEM`
    c. `SECURITY`
    d. `NTUSER.DAT`

4. What is a cache?
    a. A bunch of files
    b. A bunch of pictures
    c. Files stored by a web browser
    d. Files stored by the user
5. What are cookies?
    a. A tasty afternoon delight
    b. A text file
    c. Something a Girl Scout sells
    d. A small file created when you send an email
6. P2P applications typically use what kind of server scheme?
    a. Centralized
    b. Decentralized
    c. IMAP server
    d. SQL server
7. Which P2P application uses the `ShareH.net` and `ShareL.dat` files?
    a. eMule
    b. Shareaza
    c. Ares
    d. eDonkey

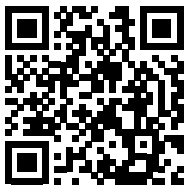
Further reading

Casey, E. (2017). *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. Vancouver, B.C.: Langara College.* This is available at `https://www.amazon.com/Digital-Evidence-Computer-Crime-Computers/dp/0123742684`.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`



10

Online Investigations

What information is available in the online environment? Did you identify a potential suspect? Are there databases that you can use to find information about potential targets of the investigation? Is it possible to conduct a sting operation using online resources? These are valid questions, and in most cases, the answer to these questions will be yes. When you ask someone to find information about an individual, one of the first searches will be conducted via Google. Google will give you a lot of information that may or may not be pertinent to the individual who is the search's target. A Google search will provide you with many different threads that may need to be followed but can also lead you down a rabbit hole that may not provide any helpful information.

Law enforcement and private organizations are conducting online investigations as we speak. Law enforcement is looking for criminal activity, such as adults looking to sexually exploit children. Private organizations are vetting potential new hires, maybe conducting opposition research on a competitor, and institutes of higher learning are researching numbers for the incoming class. I am sure you have read about a job offer or an offer to attend a college that has been rescinded because they discovered historical online information where the user had posted derogatory or inflammatory opinions.

We will cover the following topics in this chapter:

- Undercover Investigations
- Background Searches
- Preserving Online Communications

Undercover investigations

What is an online investigation? Is it a set of resources, or is it a user's activity in the digital world? An online investigation is a systematic search of the internet to identify, preserve, analyze, and report on information relevant to the subject of the investigation. Most of us are familiar with someone being fired for something they posted online. It could be an insensitive tweet, a TikTok video filmed at work, or a Facebook post that could be considered offensive.

Consider, in 2020 and 2021, the following:

- Maryland fired Arthur Love because of his posts on Facebook that supported Kyle Rittenhouse. Love was the Deputy Director for the state of Maryland's Office of Community Initiatives. The firing was because the postings contained "divisive images and statements." Love has since filed a lawsuit stating Maryland violated his rights, and he should not be punished for content created on his private time and personal social media accounts.
- "Eli," a barista for Starbucks, was fired after they posted a video to the social media platform TikTok. The video's subject included a skit on how the baristas wanted to respond to demanding customers. Eli stated that no customers were present for the filming, no company equipment was damaged, and the filming occurred after the store was closed. Starbucks responded and said "Eli" was fired because all employees are expected "to create a respectful, safe, and welcoming environment." Starbucks further stated "Eli" "demonstrated behavior not aligned with Starbucks Mission Values by posting content on TikTok that shows partners mocking customers."
- The Third Circuit Court of Appeals stated that an employer could fire an employee based on social media posts. The case is "Ellis v. Bank of New York Mellon Corp." The employee was fired after several postings advocating violence against protesters on her open Facebook account. Complaints were made to the employer about the postings. As a result, the bank conducted an internal investigation that included an employee interview. After the investigation, the bank terminated the employee for violating the bank's Social Media Policy. The bank felt the posts were "offensive, demonstrated poor judgment, showed a lack of respect for others, and encouraged violent behavior."

Open-source information can be a treasure trove of information relating to your investigation.

Michael Brazzell authored the book "Open-Source Intelligence Techniques," which is an extensive deep dive into using the internet as a resource. I will be providing an overview, but if you want to get deep into the weeds, I recommend his work.

Is information gathering an aspect of an undercover online investigation? Yes, absolutely. Is it possible for a non-law enforcement organization to conduct an undercover investigation, or is it just the purview of an agent of the government? No, it is not, commercial organizations conduct undercover online investigations for various reasons, such as opposition research on a competitor, an investigation into the theft of services, interacting with suspected "threat actor" to identify if the "threat actor" will target their organization, or any other reason, intending to protect the organization.

Undercover platform

Preparation is essential to conduct an online undercover investigation.

Before starting an undercover online investigation, the first consideration is the platform you will use when interacting with other users. Do you want to use your personal desktop/laptop, the organization's equipment, or start fresh with an entirely new desktop/laptop? You will want to make sure, whichever option you choose, that the platform is wiped clean of any prior information that may lead to the true identity of the undercover operative or the organization. Make sure that there isn't any:

- Spyware
- Malware
- Tracking Cookies
- Internet Cache

If you have any of the above-listed items on your platform, when you are using your true identity, it is possible that that information could compromise your undercover identity.

When deciding which platform to use, you will want to ensure that the platform helps support your undercover identity. For example, if your undercover identity is supposed to be an elite threat actor with skills using the **command-line interface** (**CLI**), but you only interact with other users using the **graphical user interface** (**GUI**) of Windows 10, this discrepancy could cause your credibility to be questioned.

Another consideration is that if your undercover online activities result in an arrest or proceedings in an administrative or civil environment, the platform you use will be treated as evidence. If you do not adequately remove all the data relating to your true identity, it is possible that information could be used against you. While in a perfect world, the obvious answer is to use only new equipment, unfortunately, that is not the reality. When you use previously used equipment as your platform, you must reformat and reinstall the operating system and any applications needed for the undercover online investigation.

Once you have completed the process, the undercover investigator must exclusively use the platform for the undercover online investigation.

You also want to ensure that your connection to the network is protected. For example, you will want to deploy an anti-malware/antivirus solution to monitor your platform and your connection to the network. In addition, you will want to use a **virtual private networking (VPN)** connection, which will allow you to obfuscate your location. The VPN will also encrypt your traffic from your platform to the destination, making it unreadable to anyone attempting to eavesdrop on your traffic. There are several free and paid VPN services you can use, along with the option of creating your own VPN.

Another option to protect your online undercover identity is to use the onion network, also known as the Tor network. The Tor project is a non-profit organization whose mission is to "advance human rights and freedoms by creating and deploying free and open-source anonymity and privacy technologies, supporting unrestricted availability and use, and furthering their scientific and popular understanding." The Tor project maintains the Tor network. The Tor network uses open-source software to create an overlay network enabled by volunteers. The Tor network allows users to maintain their anonymity by protecting their online activities. If the undercover operative uses or accesses the "dark web," they will have to use the Tor network.

Once the platform has been created, the next step is to create the online persona of the undercover online investigator.

Online persona

What is needed for an online persona? First, you are pretending to be another person, so you should have a similar online footprint to what the average user might have. You will want to ensure your undercover online persona will stand up to scrutiny.

The first step in creating an undercover online persona is to create an email address. The obvious option is creating a free email account using a provider such as Gmail or Yahoo. Using a free provider will require providing an established email account or a non-voice over IP (VoIP) telephone account. In addition, some providers will allow you to create a disposable email address for verification purposes.

A disposable email address is a method of using a unique email address for online interactions. This method provides the advantage that if the address is compromised or used for illegal activity, the user may quickly cancel it without impacting other online interactions.

You will find options for free and paid disposable accounts. The following is a sampling of disposable email providers:

- Temp Mail `https://temp-mail.org/en/`
- Guerrilla Mail `https://www.guerrillamail.com/`
- Tutanota `https://tutanota.com/`
- ProtonMail `https://protonmail.com/`

The Guerrilla Mail interface is displayed below:

Figure 10.1: Guerrilla Mail interface

Guerrilla Mail is a free service. When you go to the website, it will auto-generate a username, and then you can select the domain for the email account. There is also an option to create a "Scramble Address" to create an alias for the disposable email account.

The undercover online investigator may have to make purchases or transmit funds to the targets of the investigation. You would not want to use any financial resources that can be traced back to your true identity or identify the organization you are working for. Cryptocurrency is perfect to use in this environment and under these restrictions. Cryptocurrency uses an encrypted data string and is administrated through a blockchain. The blockchain will also serve as the "ledger, "tracking the transactions to include when the cryptocurrency is transferred to another party. You will not find a centralized infrastructure that maintains the cryptocurrency; it is the decentralized process that allows for its anonymity.

Bitcoin is a decentralized, open-source, peer-to-peer virtual currency that started in 2009. Some countries have accepted Bitcoin as legal tender, but mostly you will find a small amount of local and national governments using it in some capacity. Bitcoin has come to be recognized by the average consumer. And as a result, it is now being accepted by more and more organizations. A Bitcoin wallet is required to store the user's Bitcoins. Several vendors provide digital wallets for users' digital currencies. There are many digital currencies to choose from that are like Bitcoin. For example, in 2013, Billy Marcus and Jackson Palmer released Dogecoin. This is considered the first "meme coin" because they created it as a joke, making fun of the speculation by investors into cryptocurrency. The official logo of Dogecoin is the face of the Shiba Inu.

Also available to the undercover online investigator is the world of **person-to-person** (**P2P**) transactions. These are typically money transfers through an application on a mobile device or home computer system. Some applications include Cash App, Venmo, and Zelle. These applications are usually associated with the user's credit card or bank account.

Another option is prepaid credit/debit cards, which do not require personal identifiers.

To create the particular details of the undercover online persona, I like to use the website Fake Name Generator (`https://www.fakenamegenerator.com`). This website provides a free service that can provide the name, address, email address, telephone number, mother's maiden name, credit card numbers, national identification numbers, and customize the content to fit any region, nationality, or gender. In the screenshot below, you can see the interface for the fake name generator.

The drop-down menus have been expanded to choose the gender, the name set, and country. You could create a name type if you so desire, such as American, German, hobbit, Klingon, and the ever-popular ninja. In the main body of the screenshot, you can see the generated information.

The name Paul D. Walker has been created with an address in the state of Kentucky. We also have Walker's phone number, birthday, zodiac sign, and some online artifacts. They created an actual email address for the persona and username/passwords.

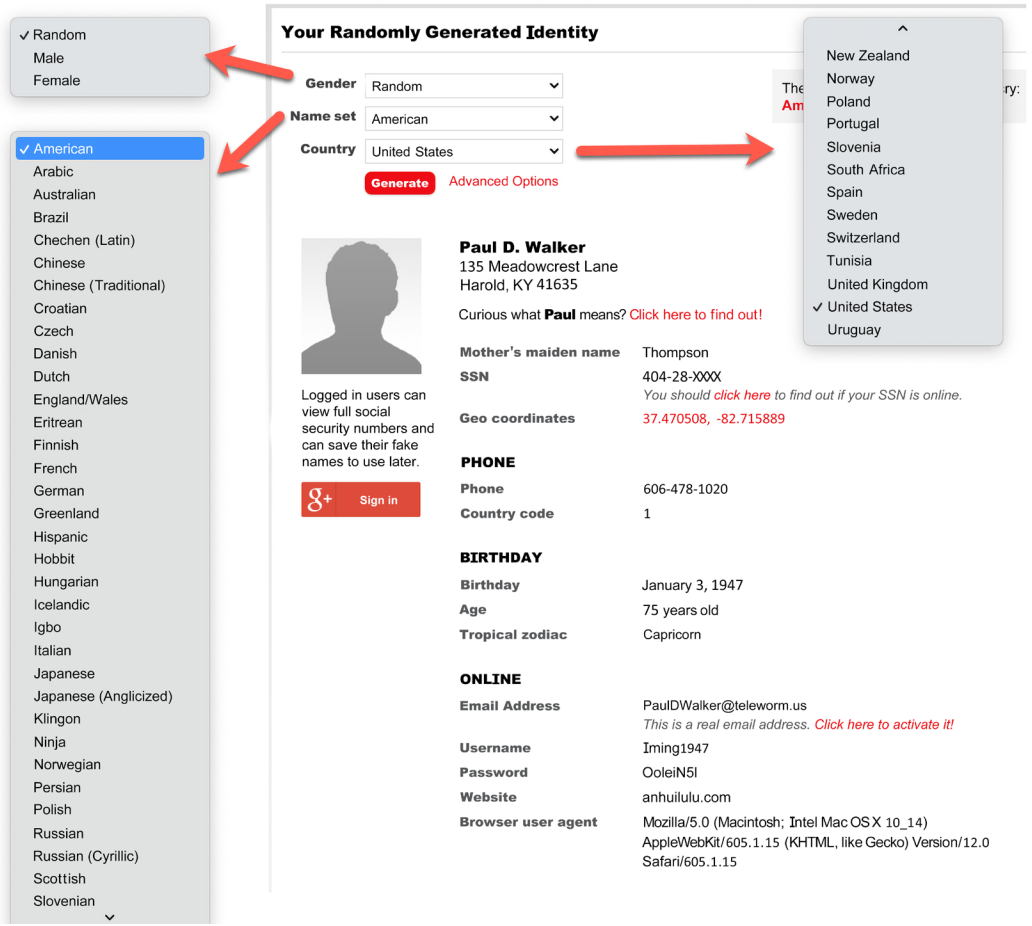


Figure 10.2: Fake Name Generator persona creation

The following screenshot shows the persona's physical characteristics, employment information, and credit card number. The website states they use a credit card number generator that creates a syntax valid credit card number.

However, the expiration date is randomly generated, and the credit card prefix is not a valid entry to help prevent fraud.

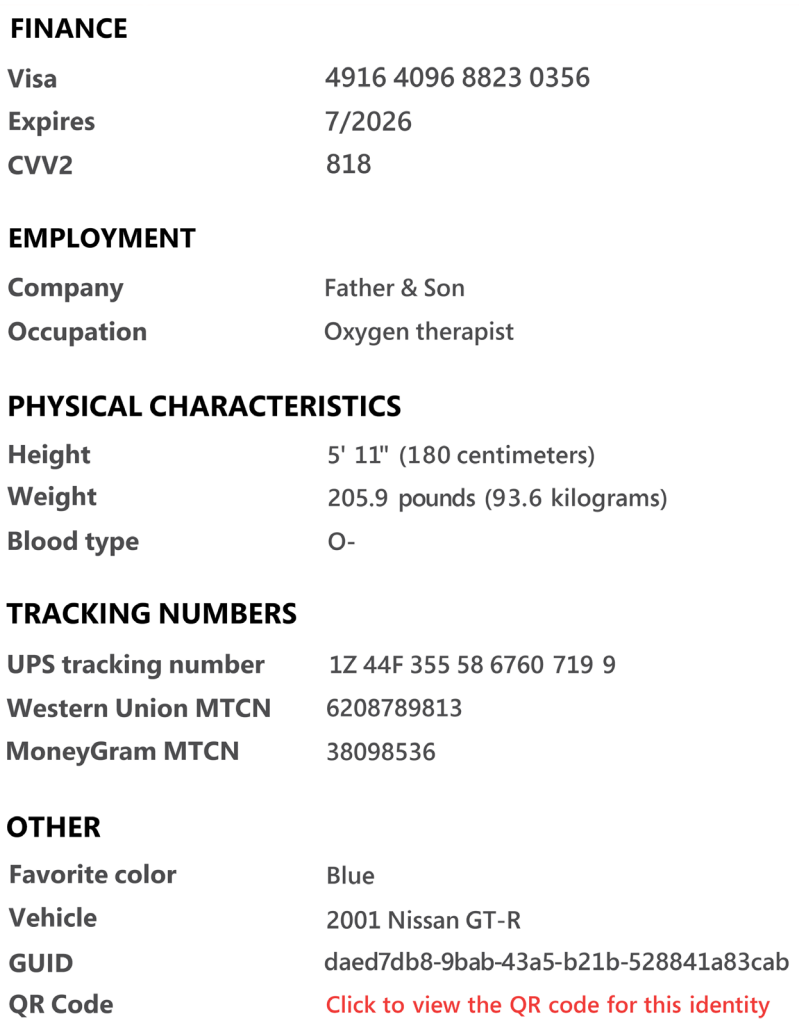


Figure 10.3: Remainder of the persona's information

The undercover online persona may also need a picture to help the believability of the persona. The website This Person Does Not Exist (`https://thispersondoesnotexist.com`) creates a random computer-generated image, and you can refresh the page to generate different images.

The following screenshot shows a selection of images created by the website.

Figure: 10.4: Random set of images from https://thispersondoesnotexist.com

Mobile communications are next on the list to be anonymized. You can deploy several physical and digital methods to protect the true identity of the undercover investigator. Many pay-as-you-go cellular phone vendors offer a 1 to 3 months service plan, which sometimes will also include the mobile device. The communications company will usually bundle together voice and data services. For example, Mint Mobile (`https://www.mintmobile.com`) offers services for $15 a month. If a mobile device is needed, the monthly cost increases from $20 to $44 a month. You will have the option of receiving a OnePlus n200 phone, a Samsung Galaxy A02, or an Apple iPhone SE.

A digital option is Fake Caller ID (`https://fakecallerid.io`), an application that you can use on either Apple or Android devices. The pricing for the service runs from $9.95 (60 credits) to $49.95 (350 credits). It allows the user to create a fake caller identification, change the user's voice, record a call, and send a call directly to voicemail and international calling.

To record voice communications, law enforcement must get approval from a judicial official to intercept wire, oral, and electronic communications. Suppose you are not a member of law enforcement. In that case, you still have to determine whether the state (this is for the United States, your local jurisdiction may have different rules and regulations for recording voice communications) is a one-party or a two-party consent state. A one-party consent state only requires permission from a single party in the communication channel to consent to be recorded; a two-party consent state requires the consent of both parties in the communication channel to consent to be recorded. There are only 12 two-party consent states in the United States. There may be some specific requirements, depending on the state. For example, Oregon only requires one party to consent for electronic communications but requires two-party consent for in-person communications.

If you have ever called a vendor or organization and received a message that they will record the call for training purposes, they consider your action of staying on the line to be consent. If you tell the organization's representative that you decline to be recorded, they will quickly end the call.

Background searches

After identifying a potential target, the online investigator should start their reconnaissance of the subject. The investigator will need to be familiar with what resources are available online and offline. Some of the information that you may uncover may consist of the following:

- Personal identifiers
- Physical location
- Social media activity
- Professional memberships/activity
- Participation in online groups

You may not have all the information to identify the subject's online persona and connect the online persona with the physical persona. For example, if the subject has a common last name, it will be challenging to differentiate one John Smith from another John Smith. However, identifying the target's email address may allow you to distinguish between the potential millions of hits for the common name "John Smith."

Even with an email address, you may not get any responses to your query. For example, suppose the subject of investigation is using the email address "`badguy27@yahoo.com`," and you do not get any responses. It could be because there is an error in the email address, a missing character, or a word was misspelled.

The next step should be to validate the email address to determine if it is valid. Many online services will help you validate email addresses. Below are some of the services. Some are fee-based, while others only require you to create an account:

- Email Hippo (`https://tools.emailhippo.com/`)
- Hunter (`https://hunter.io/`)
- Verify Email (`https://verify-email.org/`)
- DeBounce (`https://debounce.io/`)
- Emailable (`https://emailable.com/`)
- Reacher (`https://reacher.email/`)
- WhoisXML API (`https://geekflare.com/email-verification-api/`)

In the following example, I am using the WhoisXML API service. I entered the email address as shown below:

Figure 10.5: WhoisXML API input badguy27@yahoo.com

WhoisXML API provided a response very quickly. The response is shown below:

badguy27@yahoo.com verification details

Check email by syntax	Valid
SMTP check	The email address exists and can receive email over SMTP.
Domain name system check	The domain in the email address has passed DNS check.
Free email address check	The email address is free.
Check email provider for abuse	The email address isn't disposable.
Catch all emails address	The mail server has a “catch-all” address.

Figure 10.6: WhoisXML API response for badguy27@yahoo.com

As you can see, the email address “`badguy27@yahoo.com`" is valid. Next, you may have to determine if that email account has been compromised in the past. Pastebin (`https://pastebin.com/`) is a website where users are allowed to create public posts, “pastes,” which contain content in plain text. It is prevalent for users to share code or any other text-based content such as:

- Content that exceeds Twitter’s character limit (users can include a link to the full content on Pastebin)
- A Google Docs alternative
- Site promotion
- Sharing source code
- Reposting banned material

- Sharing data sets obtained from a network breach
- Sharing content/links from the dark web

Content from the Sony Pictures attack, the InfraGard attack, and the Ring intrusion were all posted to Pastebin.

There are several online sources where you could check to determine if an email address is a part of a compromised data set. The following is not an all-inclusive list but just a small sample of what can be found.

- PSBDMP (`https://psbdmp.ws/`)
- have i been pwned? (`https://haveibeenpwned.com/`)
- SpyCloud (`https://spycloud.com/`)

Using "have i been pwned?" I have entered the email address "`badguy27@yahoo.com`" to determine if this email account has been involved in any network attack/compromise.

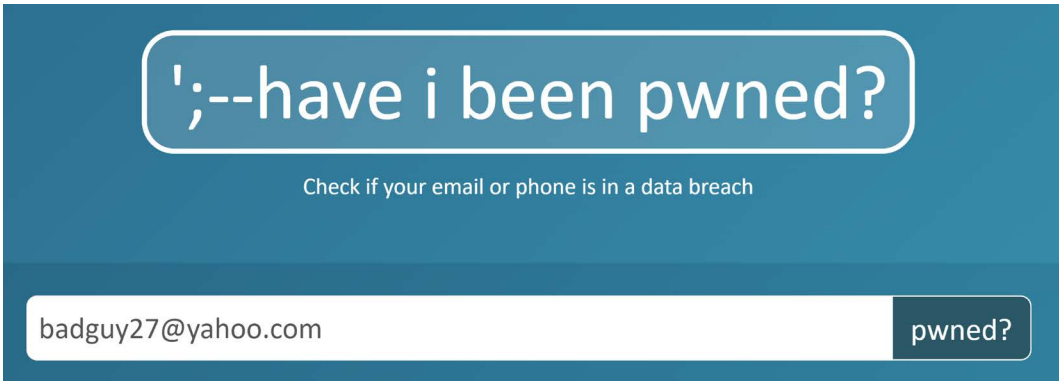


Figure 10.7: have i been pwned? search for badguy27@yahoo.com

In the below screenshot, you will see that the email address "`badguy27@yahoo.com`" has been involved in two data breaches, the MySpace data breach in 2008 and the Gravatar data breach in 2020.

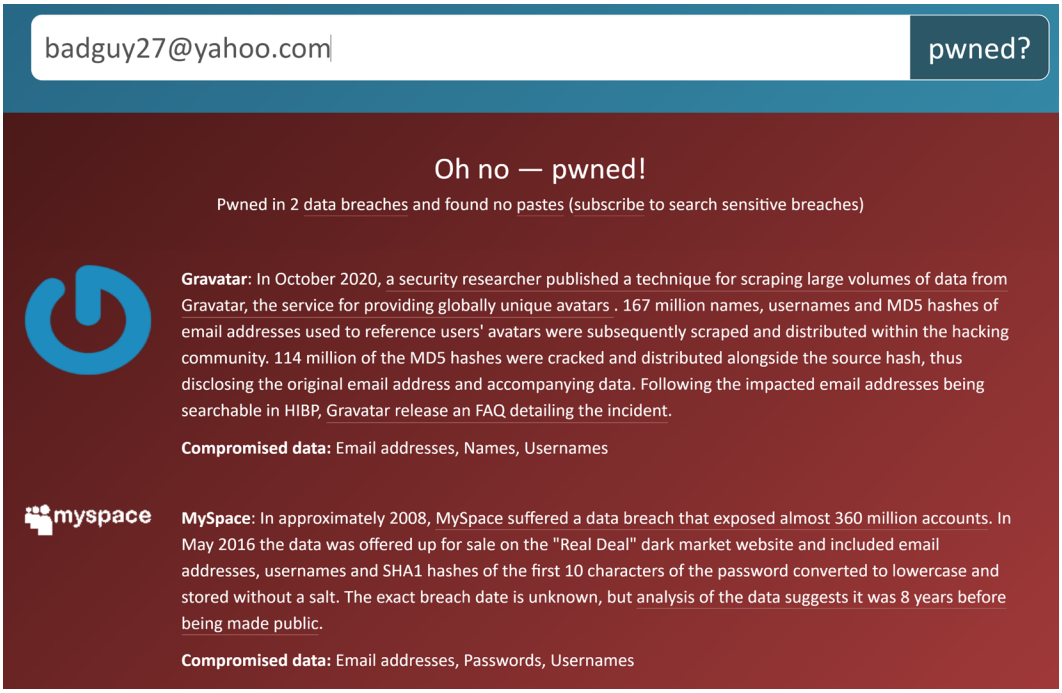


Figure 10.8: have i been pwned? results for badguy27@yahoo.com

So far, we have determined that the email address is valid, but it may be compromised. If you are involved in a criminal investigation and the matter is going before a judicial hearing or adjudication, the opposing counsel may try to use a “threat actor” defense. An email address was involved with data breaches, but it does not automatically equate to being compromised. If you gain access to the subject’s digital devices, you would want to make sure that the devices have not been compromised.

You will want to be proactive, rule out the “threat actor” defense before the opposition, can start to use it as a defense.

An online investigator can also use usernames to help identify the potential target. It is not uncommon for users to use the same username across different accounts. For example, when using the email address `badguy27@yahoo.com`, the username "bad guy 27" may be used for various email providers such as Gmail or AOL. The user may also use it for social media such as Facebook, Instagram, or TikTok. Therefore, you will want to use the term "bad guy 27" or variations of that username when conducting a background check of the target.

Knowem (`knowem.com`) is a website that allows users to search for usernames across many different platforms. The stated purpose of the website is to enable users to discover if their trademark, copyright, or brand name is being used without their permission.

Figure 10.9: Knowem search for badguyneedslove

Once you enter the username and start the search, the service will start searching through social media and other websites and will tell you if the username is "available," or if the term "available" is crossed and grayed out, it is an indication that the username is being used on that site.

Figure 10.10: Knowem search results for badguyneedslove

The below screenshot shows that the username "badguyneedslove" is available on several websites. The results that we are looking for are the websites where the word "available" is grayed out, such as the entry for Facebook.com.

When we add the username URL `www.facebook.com/badguyneedslove`, you now get the Facebook page for the potential target.

Figure 10.11: Facebook and Twitter results

It is certainly allowed to conduct a person search via the Google search engine. Just be prepared to have to sort through the search results hoping to get a hit on your target. There are search engines that are designed to search for people specifically. These people search engines generally allow you to retrieve basic information about your subject and then request a fee before providing any in-depth results. Below are some of the more common people search engines, and the list is not exhaustive. I recommend using multiple people search engines to ensure that you get the best available information. When I have used the search engines, the accuracy I get has varied from search engine to search engine, but I compile all of the results into a single document.

I have discovered current and past addresses, current and past landline and cellular telephone numbers, immediate and extended family members, current and past accounts, and birthdays. In some cases, I was able to determine their voter registration and neighbors.

The people search engine I use first is True People Search (`https://truepeoplesearch.com/`).

Figure 10.12: True People Search – search screen

When searching True People Search, you have the option of searching by name, telephone number, or address.

When you select the search results, you will then be presented with the subject's first and last name, their age and year of birth, their current address, and phone numbers. You can have a map created showing the current address, as well as obtaining all phone numbers associated with that subject. (As shown below).

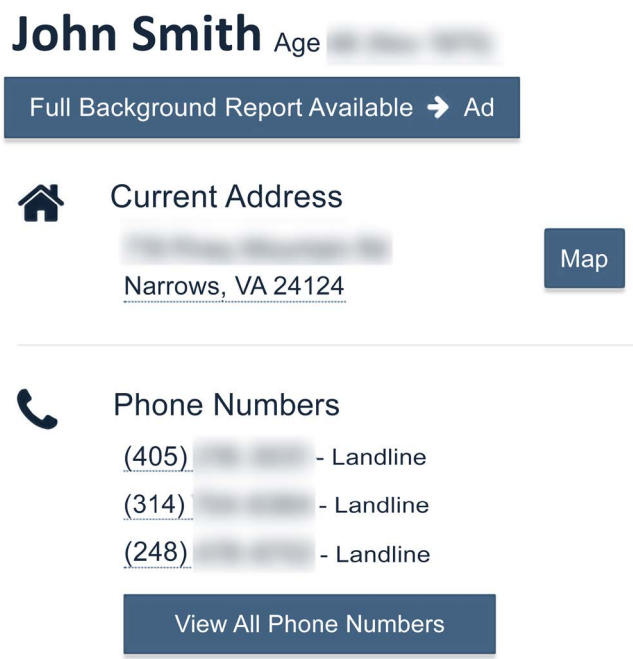


Figure 10.13: True People Search – results (name)

Next, you will see the previous addresses associated with the subject. The addresses will include the street address, the city, state, and ZIP code. Also included is the month and year that the address was associated with this individual. (As shown below).

Figure 10.14: True People Search – results (address)

The following field will then contain the current email addresses associated with the subject. The search engine does not date the email addresses, making it more challenging to determine if the user is still utilizing the email accounts. The user can select a dialogue button to view additional email addresses that may be associated with the subject. Remember, it is up to you to validate these results; there is no guarantee that these results return to the physical person conducting your reconnaissance. There are times when you may get false positives while conducting your reconnaissance on the subject.

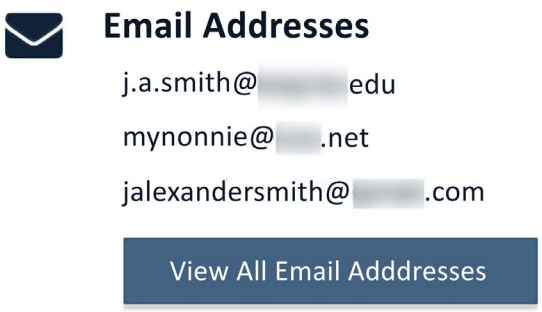


Figure 10.15: True People Search – results (email)

Towards the end of the report, (as shown below) you will find a listing of possible relatives, associates, and businesses. This information provides you with additional investigative avenues to either confirm or deny that this information is related to the subject of your investigation.

Figure 10.16: True People Search – results (possible)

There are many services available for you to conduct people searches. Far too many for me to include all of them in this book. As you become aware of which sites are available for people searches, you can start creating your list of people search engines that you prefer.

As I stated before, True People Search is one of the first people search engines I use. The following also provides information on the subject of your searches. There will be some overlap with the results; you'll have to analyze the results to determine if the information that is being provided applies to the subject you're investigating. Below are some of the additional people search engines that I have used in the past:

- Whitepages – `https://www.whitepages.com/`
- ZabaSearch – `https://zabasearch.com/`
- People Search Now – `https://peoplesearchnow.com/`
- Spokeo – `https://www.spokeo.com/`

Next, we will discuss how to preserve the aspects of your online investigation so that you may properly document your efforts and present your evidence in an administrative or judicial proceeding.

Preserving online communications

Did an investigation happen if you conducted the investigation and failed to document your efforts? Your ability to record your investigative efforts is as critical as the investigative states themselves. You have to document your efforts in the results. You must include those results which are positive as well as negative. When the online investigator is performing investigative activities, there must be documentation. This requirement requires you to either capture your efforts retroactively or in real time as it occurs.

Screen captures are an effective method of capturing what is viewed on the display. However, the screenshot must focus on the artifact or behavior you capture for screen capture to be effective. What I mean is that when a screen capture is taken, the capture needs to focus on what is important. For example, capturing a desktop with six or seven windows open does not effectively communicate to a third party what is pertinent. Both macOS and Windows OS can natively capture what is displayed on the screen.

With macOS, you can do a Spotlight Search by pressing the *command* key and *spacebar* (<Command> + <Spacebar>) at the same time (as shown below.)

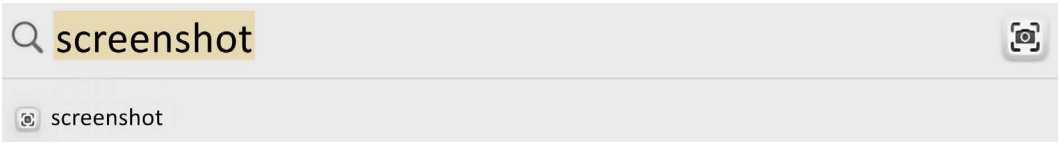


Figure 10.17: Spotlight Search – "screenshot"

With the Windows operating system, you have the Snipping Tool, which is being migrated to the new Snip and Sketch tool. Both tools provide the same functionality. Below is an example of the interface you may experience when using the Snipping Tool or the Snip and Sketch tool.

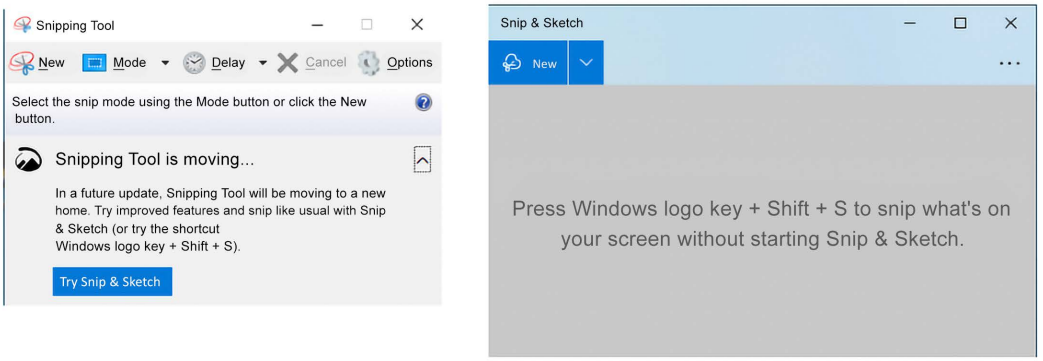


Figure 10.18: MS Windows – Snipping Tool and Snip and Sketch

Using video to record the screen as the events occur is also an option. Using a macOS-based system, you will have access to the free utility QuickTime Player. QuickTime Player can record video using a USB camera and create a screen recording of what is displayed by the monitor. If you are recording audio, make sure if you have any colleagues around that they know you are recording. It can be very frustrating 30 minutes into the recording when one of your coworkers lets out a loud expletive in the middle of your court exhibit.

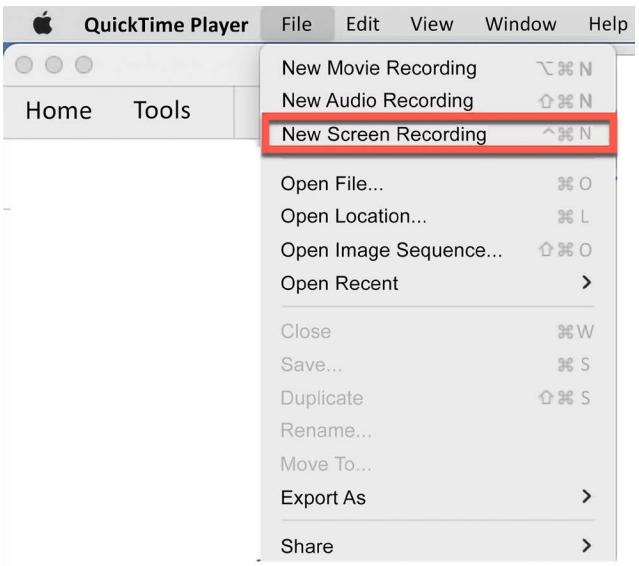


Figure 10.19: QuickTime Player menu

When using Microsoft Windows, you can use the *Windows Key + Alt + R* to start screen recording. A small recording widget will appear. Repeat using the *Windows Key + Alt + R* to stop screen recording or left-click on the stop button on the widget. A notification will appear, *Game clip recorded*. The system will save the video in the user's `Videos/Captures` folder.

The Edge browser can capture webpages. Using the keyboard, press on the *Control*, *Shift*, *+S* key to bring up the interface.

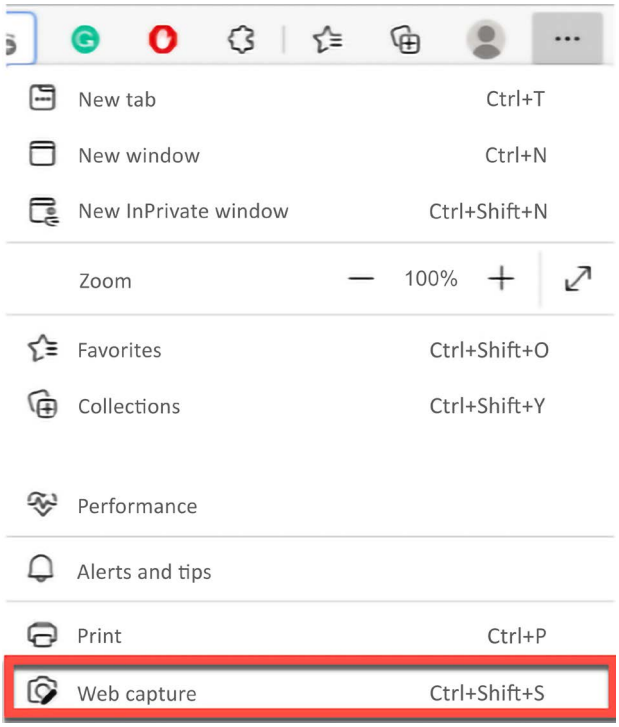


Figure 10.20: Edge browser capture menu

Another option to save the data being displayed via the browser is the tool Hunchly (`https://www.hunch.ly/`). It is not uncommon for an online investigator to start with one tab open on their browser while working through the investigation. As the investigator finds additional investigative leads, additional tabs are then opened. Soon there are many tabs open from following the threads of the investigation. If the investigator did not take appropriate notes/documentation, it could be challenging to go back and accurately preserve what the investigator has seen. The tool Hunchly helps avoid that dilemma. The investigator will have to use a Chrome-based browser and install the Hunchly application.

The application will also install an extension to the browser.

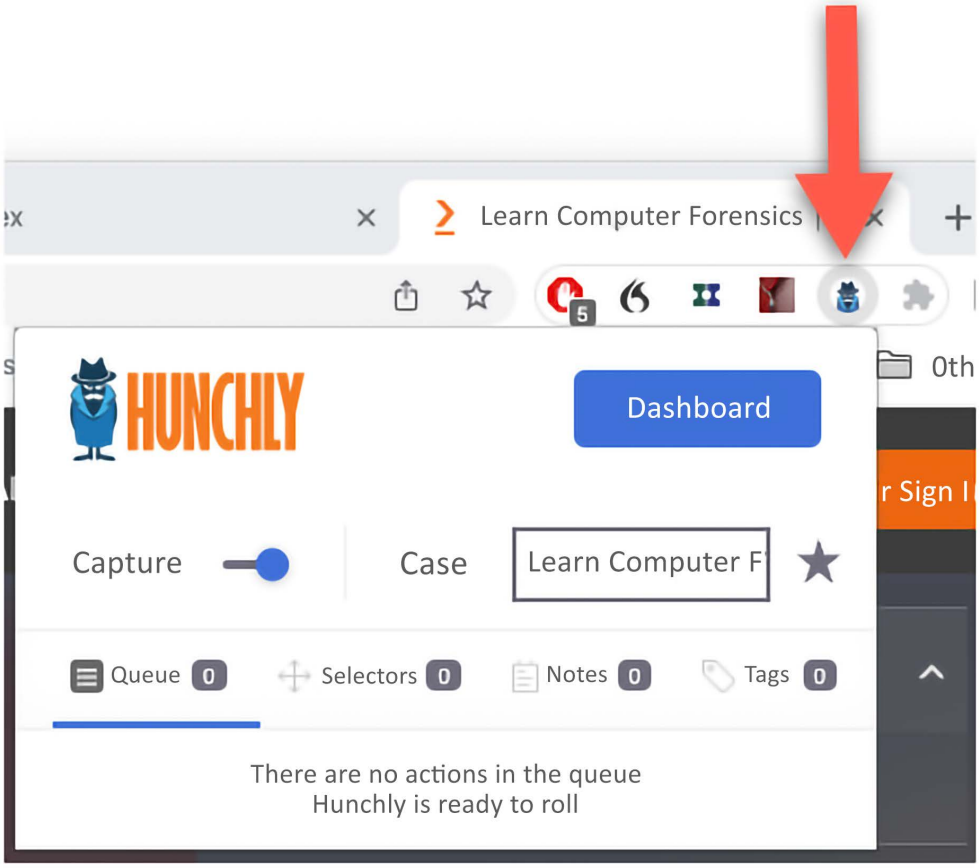


Figure 10.21: Hunchly extension menu

Hunchly does provide a free 30-day trial, after which you can purchase a year-long license for a very reasonable cost.

Once you activate Hunchly via the extension, it will monitor your progress as you switch from webpage to webpage. If you open additional tabs, Hunchly will also record those websites. For example, in the below screenshot, you can see the active **Case** and next, you will see an indicator for **Pages viewed** and **Searches** with a number below the headings. I have visited ten web pages in this trial run and conducted six searches.

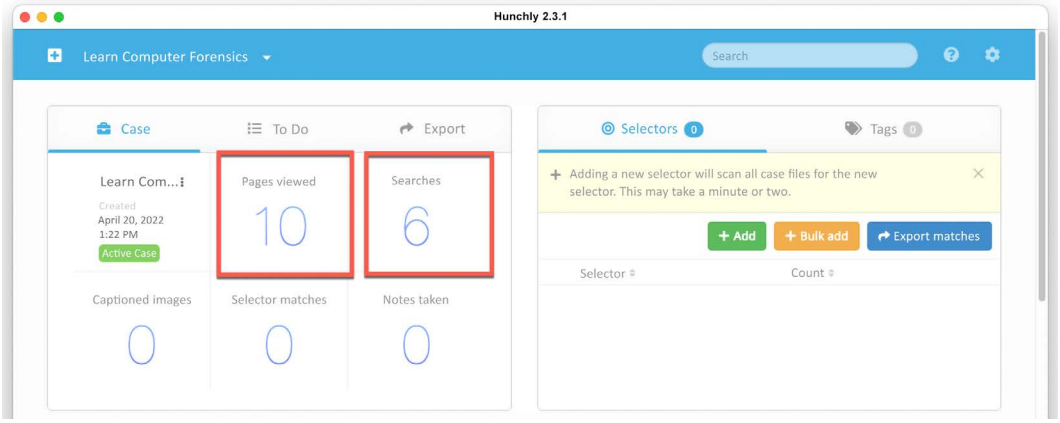


Figure 10.22: Hunchly desktop

One of the websites I visited was the **American Kennel Club** (**AKC**) website, and I looked at the breed information for Miniature Schnauzer. I also did some Google image searches for Miniature Schnauzers. In the below screenshot, the dashboard shows what pages I visited, along with a date and time stamp.

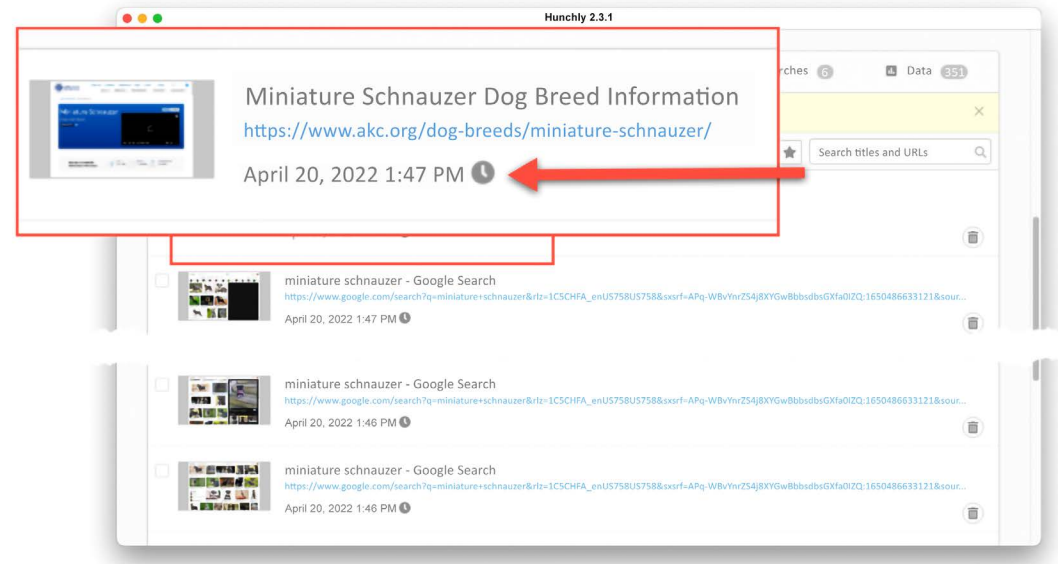


Figure 10.23: Hunchly history

Hunchly does not only collect the visual aspect of the webpage, but it also collects embedded data such as email addresses, IP addresses, Google analytics, and in some cases, GPS coordinates. For example, in the below screenshot, you can see the page preview for the American Kennel Club, and directly above and centered in the preview is the **Data** button.

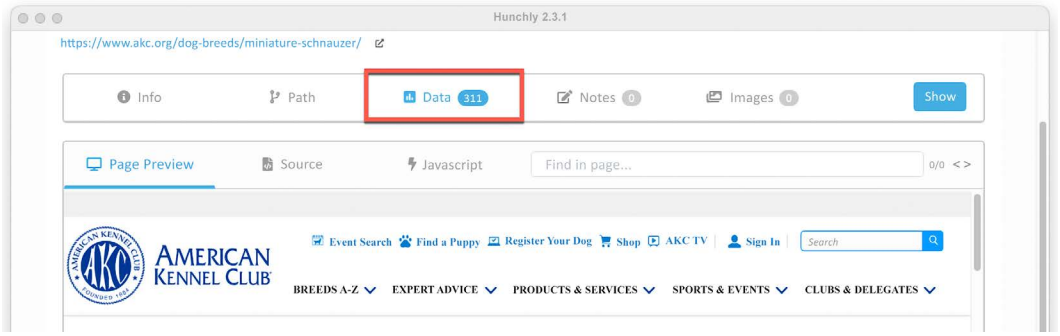


Figure 10.24: Hunchly preview

In this case, 311 data entries have been recovered from the website. In the screenshot below, you can see entries for Google Analytics, Facebook tracking, email addresses, and IP addresses. All of this information was embedded within the website.

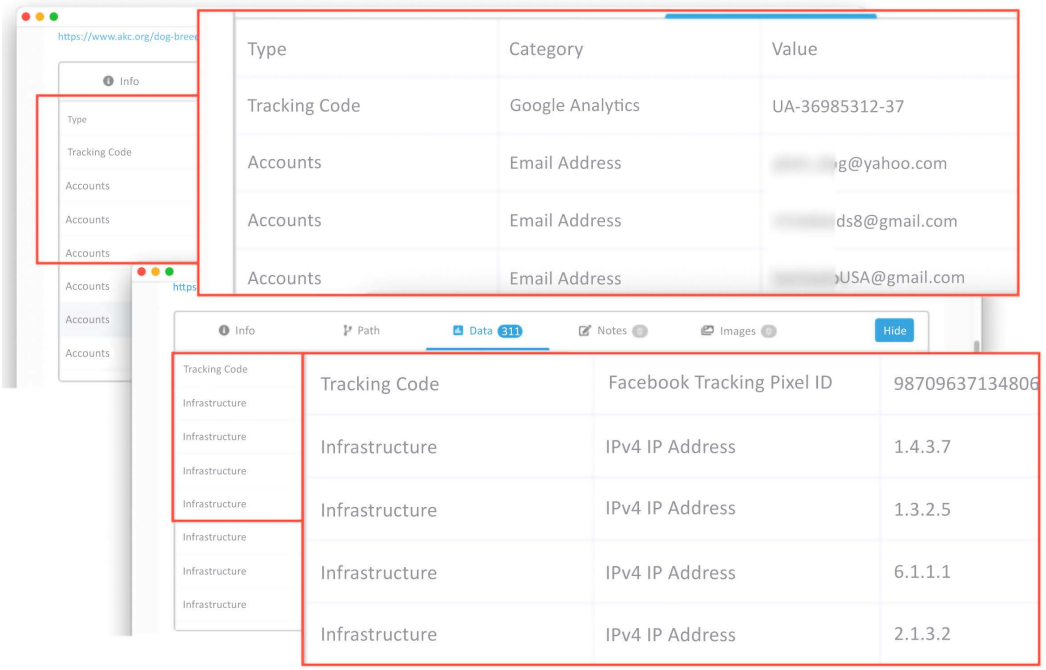


Figure 10.25 Hunchly Data view

Hunchly also allows the user to export the captured data to a PDF file. This will enable you to create a physical copy if you need to disseminate your investigative actions.

This is just one option that you can use to capture webpages; you can also use one of the following options:

- FireShot `https://getfireshot.com/`
- HTTrack `https://www.httrack.com/`
- Web2Disk `http://www.web2disk.com/`
- SiteSucker `https://ricks-apps.com/osx/sitesucker/index.html`
- X1 Social Discovery `https://www.x1.com/products/x1-social-discovery/`
- EyeWitness `https://github.com/FortyNorthSecurity/EyeWitness`
- FAW `https://en.fawproject.com`

This list is not exhaustive; there may be other utilities that will perform the same functionality. Some utilities are free to use, some are reasonably priced, and some are very expensive. Your specific situation will determine which utilities you will be able to use. It does not matter which utility you use; some paid utilities have the same functionality as a free utility. However, you are paying for the customer support that comes with the expense of the utility, while with a free utility, you may have to do your research to solve a problem you may be having. Ultimately the goal is to be proficient in the use of the utilities that you use. You also must understand what is occurring underneath the visual interface of the utility.

Summary

In this chapter, we have discussed an online investigation. We have looked at undercover operations and how to create a persona that the online investigator can use. You never want to use your personal online identity to investigate. If the target of your investigation can determine your true identity and you have been using your personal online persona, you could put yourself in danger. We also looked at the various options that investigators can use to identify a subject to conduct background investigations. A person may have no digital footprint, but I would find that very unlikely. By using open-source information, you may be able to collect that data set without alerting the target. Finally, you will want to document all of your investigative endeavors.

You can do this by using screenshots, videos, and website capture. You should now be able to:

- Determine ways to collect personal information about an individual from online sources
- Determine how investigations are performed online
- Determine how to preserve internet communications, video, photos, and other content that is important to a case

In the next chapter, we will cover networking basics. Understanding how data is transmitted across the internet will be crucial to understanding where to find the artifacts that will help you prove or disprove the allegations you are investigating.

Questions

1. You must be in law enforcement to conduct an online investigation.
    a. True
    b. False
2. What is the first consideration to conduct an online undercover investigation?
    a. RAM
    b. Operating system
    c. Tablet
    d. Encryption tools
3. Which of the following should be removed from the undercover platform?
    a. Spyware
    b. Malware
    c. Facebook
    d. Cookies
4. What can be used to obfuscate your location?
    a. Starbuck's WIFI
    b. Guest account
    c. Virtual Private Network
    d. Mosaic browser

5. Which of the following is a disposable email provider?

    a. Temp Mail
    b. Tal Shiar Mail
    c. Secret Mail
    d. Section 31 Mail

6. Which of the following is a peer-to-peer virtual currency that started in 2009?

    a. Trekcoin
    b. Dogecoin
    c. Bytecoin
    d. Bitcoin

7. Who can authorize law enforcement to intercept wire, oral, and electronic communications?

    a. Judge
    b. Admiral
    c. Governor
    d. No authorization is needed

8. What service does Email Hippo provide?

    a. Verify email address
    b. Verify contents of an email
    c. Verify the sender of an email
    d. Verify the location of the inbox

9. Which is not a reason a user may use Pastebin?

    a. Content that exceeds Facebook's character limit
    b. Site promotion
    c. Sharing source code
    d. Reposting banned material

10. You will need authorization before searching on True People Search?

    a. True
    b. False

Further reading

Bazzell, M. (2018). *Open source intelligence techniques: Resources for searching and analyzing online information*. United States: Inteltechniques.com.

Troia, V. (2020). *Hunting cyber criminals: A hacker's guide to online intelligence gathering tools and techniques*. Indianapolis, Indiana: John Wiley & Sons Inc.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

11

Networking Basics

Very rarely in your career will you have to investigate a single computer that has not connected to a network. When I say "computer," this includes mobile devices. Desktops, laptops, and mobile devices are designed to share information through a network we refer to as the "Internet." There are legitimate uses for such devices as well as illegitimate uses. For example, users of these devices have been known to record audio and video of themselves committing criminal acts. These same users will also use the Internet to research activities they plan on executing in the future; these activities will include criminal acts.

There may also come a time when you are called to investigate a crime where a device has been the target of a crime: a server may have been hacked, a user's mobile device may have been compromised, or a user's online storage may have been compromised. These acts require you to understand how devices communicate on the network. This chapter will discuss some of the essential information you need to understand about networks, from theoretical concepts right through to network topology, addressing, and some basic protocols, to increase your networking knowledge. I recommend that you study for the CompTIA Network+ exam and the Security+ exam.

We will cover the following topics in this chapter:

- The OSI and DOD models
- Network hardware
- Common ports and protocols

A network is built on a set of standards and models. If the clients on the network do not adhere to a set standard, there will be a lack of communication. This ensures that all clients on the network can communicate with each other.

The primary reference model used for the configuration of networks is the **Open Source Interconnection** (**OSI**) model. Developers use the OSI model to break communication into different layers; each layer has a specific role to play in how data is transmitted and received. (If you wish to read about the OSI standard in much greater detail, you can go to `https://www.iso.org/standard/20269.html`.)

The Open Source Interconnection (OSI) model

This reference model is just a reference for developers to utilize while they create the protocols and devices that users will use to connect to the network. This layered model has some advantages, such as these:

- The communication process comprises components that allow for easier development and troubleshooting
- By following standards, different types of equipment and software can communicate with each other
- A change in one layer does not affect other layers
- The standardization of the components allows multiple organizations to develop and deploy components

By using the OSI model, you can have hosts from Microsoft, Apple, UNIX, and Linux connected and communicating to the same network. As shown in *Figure 11.1*, you can see that the OSI model is made up of seven different layers. The layers are as follows:

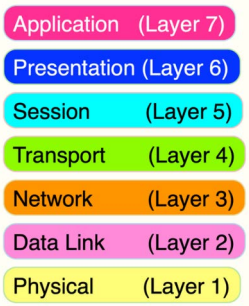


Figure 11.1: OSI model

You can use a mnemonic to remember the order from layer 1 to layer 7, such as “Please do not throw sausage pizza away.” From layer 7 to layer 1, use the mnemonic “All people seem to need data processing.”

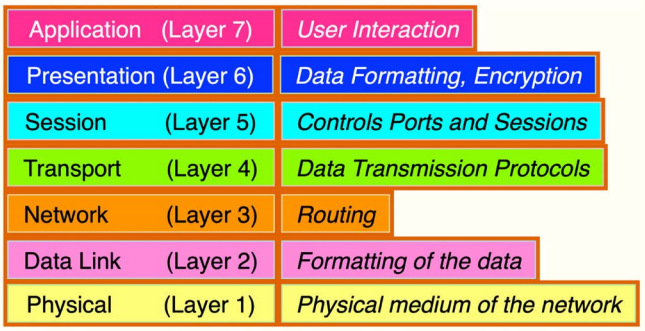


Figure 11.2: OSI model – layer functions

Physical (Layer 1)

The physical layer (Layer 1) is responsible for creating the path between the different hosts on the network. This can be the physical topology for the transmission of the data. The physical topology consists of an Ethernet network, where Ethernet cables are connected to the hosts and network devices; an example of data transmission could be a Wi-Fi network where data is transmitted via radio signal and not a physical medium. This layer takes responsibility for transmitting data and receiving data. Also controlled at this layer are the specifications for the physical medium and the connections used between the host and the physical/transmission medium.

Some of the devices that you may encounter at this layer include:

- Repeaters – electronic devices that receive and re-transmit data
- Hubs – networking devices used to connect multiple hosts and network devices
- Modems – networking devices used to encode and decode signals containing network data

Data link (Layer 2)

The data link layer (Layer 2) is responsible for data transfer. It provides for flow control, error notification, and network topology requirements between hosts and network devices connected to the same logical segment of the network. In addition, this layer is responsible for delivering data to the correct destination. This is done by using a physical address of the network interface device, which is also known as the MAC (Media Access Control) address. The data link layer contains two sub-layers:

- Logical Link Control (LLC) – responsible for identifying network-layer protocols
- Media Access Control (MAC) – responsible for the placement of the packets and physical addressing

At this layer, the data is formatted into frames.

Some of the devices you may encounter at this layer include:

- Network adapters – allow host(s) to connect to the network medium
- Bridge – a networking device used to connect two network segments
- Wireless access point – a networking device that is used to create a bridge between a wired and wireless network

Network (Layer 3)

The network layer (Layer 3) is responsible for the logical addressing of the hosts and network devices, determining the best route for data transmission, and mapping the locations of all the devices connected to the network. The logical addressing used at this layer will depend on the network. The most common address you may be familiar with is the IP address assigned to your local host. There are currently two different versions of IP addressing available: IPv4 and IPv6. At this layer, the data is formatted into packets.

Some devices you may encounter at this layer include:

- Routers – a network device responsible for the routing of data based on logical addressing (IP address)
- Switch – a network device that can forward data based on logical addressing (typically, switches are considered to be a Layer 2 device, but some do have the functionality to work at the network layer)

Transport (Layer 4)

The transport layer (Layer 4) is responsible for reliable data delivery, including establishing/closing virtual connections such as an HTTP connection between a host and a server. Multiplexing data streams are also addressed at this layer. The use of port numbers and network applications is a Layer 4 function; an example of this is web traffic being directed to port 80(HTTP) while email traffic is directed to port 25 (SMTP). At this layer, the data is formatted into segments.

Some devices you may encounter at this layer include:

- Gateway – a network device that is responsible for connecting two networks that are using different protocols. The gateway will act as a translator between two dissimilar networks.
- Firewall – a networking device that is used to prevent unauthorized access to resources on the network.

Session (Layer 5)

The session layer (Layer 5) is responsible for creating, maintaining, and closing data sessions between hosts using presentation layer protocols. This layer will utilize the following modes to coordinate the data streams between hosts/servers on the network:

- Simplex – data can only be transmitted in one direction
- Half duplex – data can be transmitted in both directions, but only in one direction at a time
- Full duplex – data can be transmitted in both directions at the same time

Presentation (Layer 6)

The presentation layer (Layer 6) is responsible for presenting data to applications and converting data from the format used by the application to the transmission format used by the network. This can include encoding schemes such as ASCII (American Standard Code for Information Interchange) and Unicode.

Application (Layer 7)

The application layer (Layer 7) provides an interface for software applications and hosts on a network. An example of this is the transfer of files between systems. This allows applications using different operating systems and file systems to communicate and provide services to the user.

Encapsulation

As data traverses the different layers of the stack, each layer adds information that the corresponding layer at the destination will read. An example of this is shown in *Figure 11.3*. The host on the left is sending a request to the host on the right. As the data is transmitted and travels down the stack, each layer may add a "header" to the data, encapsulating the header and the data added by previous layers. When the data reaches the destination host, the data will now be de-encapsulated, like peeling the layers of an onion, until it reaches the destination. For example, if the request were an HTTP request, the application header would be added to the data. At the transport layer, a TCP header will be added, the network layer will add an IP header, while the data link layer will then create an Ethernet header turning the data into a frame. The frame would then be transmitted along the physical layer. At the destination, the process will be reversed, each layer opening and reading the information added by the corresponding layer.

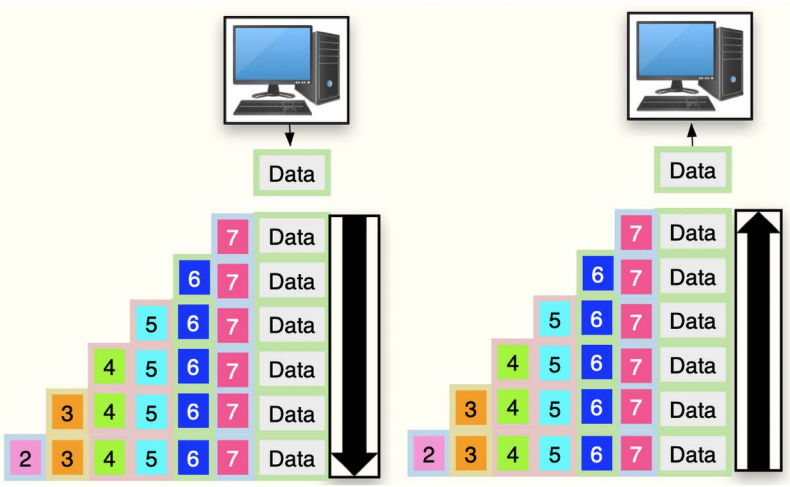


Figure 11.3: Encapsulation

TCP/IP

While the OSI model is theoretical, TCP/IP is a practical implementation of a set of protocols that dominates the current networking environment. TCP/IP was created by the **United States Department of Defense (DoD)** with the intention of creating a protocol that would ensure the integrity of communications in the event of a catastrophe. Therefore, the TCP/IP is also referred to as the DoD model and is made up of the following four layers:

- Application
- Transport
- Internet
- Link/network access

It can be mapped to the different OSI model layers as shown in *Figure 11.4*.

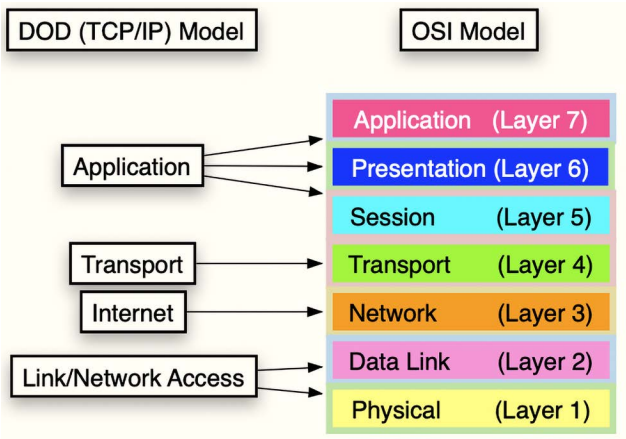


Figure 11.4: The TCP/IP model compared to the OSI model

The link/network access layer corresponds with the OSI model's physical and data link layers. This layer defines how the networking device will connect to the network.

The Internet layer corresponds with the network layer of the OSI model. This layer is responsible for the logical addressing and routing of the data.

The transport layer corresponds with the transport layer of the OSI model. This layer is responsible for establishing connections between the source and destination.

The application layer corresponds with the OSI model's session, presentation, and application layers. This layer is responsible for administrating the data flow of the different applications on the system. Some of the services you may encounter include FTP and HTTP.

TCP/IP is a suite of protocols used by the host and network devices when connected. The main protocols of TCP/IP include TCP (Transmission Control Protocol), IP (Internet Protocol), and UDP (User Datagram Protocol).

TCP is a connection-orientated protocol, which means a connection between the host and server must be established before data can be sent/received. This is achieved through the use of the three-way handshake. When a host wants to communicate with the server, they will initiate the connection by sending an SYN packet to the server. If the server is available, it will respond with an SYN/ACK packet back to the host. The host will then respond with an ACK packet, and the connection can be established. This process is shown in *Figure 11.5*.

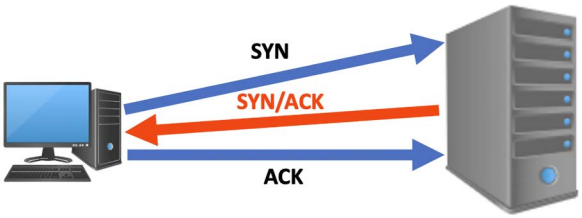


Figure 11.5: TCP three-way handshake

TCP is a transport layer protocol and is used for the accurate delivery of data. Each packet must be accounted for as data is sent from the source to the destination. If a packet gets lost in transmission, the destination will notify the source that the packet is missing and needs to be resent. As a result, TCP communication may experience delays and is not suitable as a communication channel where speed overrides the need for accuracy.

UDP is a connectionless oriented protocol. Whereas TCP requires a three-way handshake before the communication channel can be initiated, UDP does not. Therefore, when using UDP as a communication channel, the source will never know if there is packet loss because the destination does not have the means to track which data packets are being sent. An example of this can be video streaming; if there is packet loss, the user may never notice because it does not affect the quality of the video.

IPv4

IP is responsible for the transmission and routing of data across a network. Currently, **IP version 4 (IPv4)** is the predominant version of the protocol, but IPv6 is slowly coming out and being used by more and more organizations. IP is also responsible for the logical addressing used by hosts when connected to the network. IPv4 uses a 32-bit addressing scheme consisting of 4 octets separated by a period (`.`).

For example, `8.8.8.8` is a valid IPv4 address. IP addresses are assigned to a specific class. The class of the IP address will dictate how many hosts you can have on the network.

Class	Address Range	Max Number of Hosts	Private IP Range
Class A	`1.0.0.1` - `126.255.255.254`	16,777,214 Hosts	`10.0.0.0` – `10.255.255.255`
Class B	`128.1.0.1` - `191.255.255.254`	65,532 Hosts	`176.16.0.0` – `172.31.255.255`
Class C	`192.0.1.1` - `223.255.254.254`	256 Hosts	`192.168.0.0` – `192.168.255.255`
Class D	`224.0.0.0` - `239.255.255.255`	Reserved for Multicast	
Class E	`240.0.0.0` - `254.255.255.254`	Reserved for Research	

Table 11.1: IPv4 IP address classes

A Class A network will allocate the first octet to networks, with the remaining three octets being used for hosts on the network (`NNN.HHH.HHH.HHH`). A Class B network will reserve the first two octets for networks and the remaining two octets for hosts (`NNN.NNN.HHH.HHH`). A Class C network will have the first three octets for networks and the last octet for hosts (`NNN.NNN.NNN.HHH`). If you are running a Class C network, you can only have 256 hosts on the network at one time.

Each class also has a set of private IP addresses that can be used in an internal network but will not be able to send traffic through an external router. This range of IP addresses is to be used in a closed local network. You cannot communicate outside the local network using an IP address in these ranges. You will typically find that the local networks using this range of IP addresses are behind a router and that they are using **Network Address Translation** (**NAT**).

NAT maps a private IP address(es) to a public IP address. The technique is used to conserve address space and avoid IPv4 address exhaustion. NAT allows the router to convert traffic from the hosts on the private network and then relays the traffic using a public IP address. When the traffic is returned, the router will then send the traffic back to the host. The only IP address that is exposed is the public IP address.

One public IP address using a NAT gateway can be used for an entire private network, as shown in *Figure 11.6*.

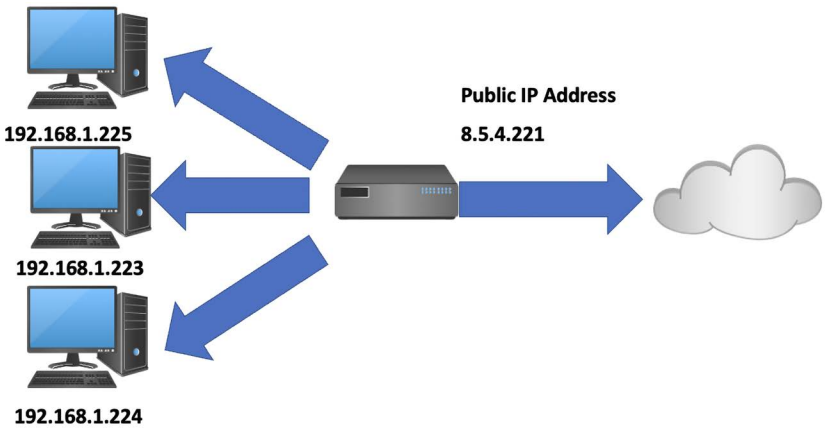


Figure 11.6: Example of NAT

Port numbers

A port is a 16-bit number attached to an IP address in computer networking. A port number is specified for each service (or protocol). The port number completes the data stream's destination or origination network address. Some port numbers are reserved for specific services. Port numbers lower than 1024 are commonly referred to as "well-known" port numbers. Port numbers greater than 1024 are called ephemeral ports. The **Transmission Control Protocol** (**TCP**) and **User Datagram Protocol** (**UDP**) use port numbers.

IPv6

IPv6 was created to address the lack of available IP addresses when using IPv4. When IPv4 was developed, the designers could not imagine the interconnected world we now occupy in the 21st century. While IPv4 can provide more than 4 billion IP addresses, eventually, that will not be enough due to the increasing demand and availability of network devices. IPv6 is a 128-bit addressing scheme and can provide each person on the planet with approximately 4,000 logical addresses. The designers of IPv6 anticipated the world of wireless connectivity between mobile devices and the **Internet of Things** (**IoT**). IPv6 is slowly being deployed to devices connected to the network but is still not the primary addressing scheme in the consumer world. IPv6 is based on the hexadecimal numbering system and consists of eight groups of four hexadecimal characters; a colon separates each group.

Each group consists of four digits, with each digit representing four bits. Each segment can have a hexadecimal value of zero through FFFF. An example of an IPv6 address is shown in *Figure 11.7*.

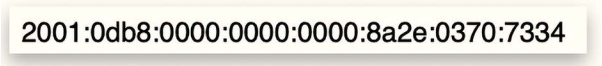


Figure 11.7: IPv6 address

An IPv6 address is divided into two 64-bit segments. The first segment is the network component, and the second segment is the node component, as shown in *Figure 11.8*.

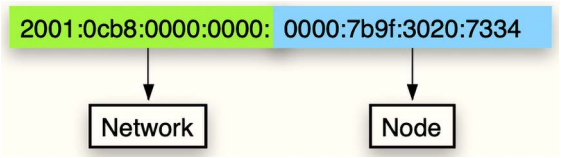


Figure 11.8: IPv6 address separated

The network component of the IPv6 address is used for the routing of data on the network. The node component identifies the network interface (node) of the host and is derived from the network adapter's physical address (MAC address) that is connected to the network.

The network component of the IPv6 address is separated into a 48-bit Global Unicast Address and a 16-bit Subnet ID, as shown in *Figure 11.9*.

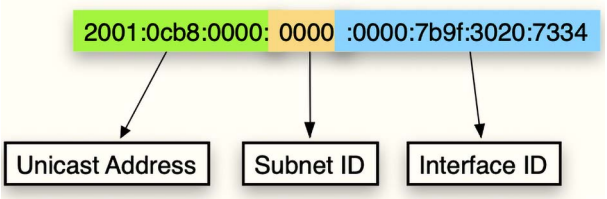


Figure 11.9: IPv6

There are three different classes of IPv6 addresses:

- Global Unicast Address – this address is routable on the Internet (starts with 2001)
- Unique Local – used for internal networks, not routable on the Internet
- Link Local – not routable locally or on the Internet

IPv6 uses the following types of addressing schemes:

- Unique Local Addresses – these addresses are used for non-routing purposes. They are nearly globally unique, so it's unlikely you'll ever have one of them overlap with any other address. In addition, they designed unique local addresses to replace site-local addresses; this allows communication throughout a site while being routable to multiple local networks. The difference between link local and unique local is that you can route unique local within your organization or company.
- Multicast – packets addressed to a multicast address are delivered to all interfaces identified by the multicast address. Sometimes people call them one-to-many addresses. Multicast addresses in IPv6 always start with FF.
- Anycast – like multicast addresses, an anycast address identifies multiple interfaces. An anycast packet is delivered to only one address.

IPv6 address can also be shortened. In IPv6, a group can consist of all zeros. You do not need to write all the zeros in the address in such a case. Instead, you can write one zero, which translates to four zeros in the group (as shown in *Figure 11.10*).

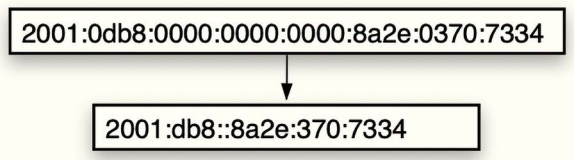


Figure 11.10: IPv6 shorthand example

The following are some protocols you may come across during your investigations. The protocols are sorted by layer in the TCP/IP suite.

Application layer protocols

File Transfer Protocol (FTP) (TCP ports 20 and 21)

This protocol allows the user to send and receive files across an IP-based network. FTP is also a program. When FTP is used as a protocol, applications are used to initiate the FTP process; when used as a program, the user initiates the FTP process.

Secure Shell (SSH) (TCP port 22)

This protocol allows the creation of a secure Telnet session between hosts on an IP-based network. Once an SSH connection has been made, the user can operate and issue commands as if they were physically present at the remote host. SSH uses encryption to create a secure tunnel between the source and destination hosts.

Telnet (TCP port 23)

This protocol connects to a remote host, that is, a terminal. The user can then operate and issue commands as if they were physically present at the remote host. As a result, a Telnet connection is not secure and is vulnerable to a man-in-the-middle attack.

Simple Mail Transfer Protocol (SMTP) (TCP port 25)

This protocol is a communication protocol for e-mail transmission. Mail servers use SMTP to send and receive mail.

Domain Name Service (DNS) (TCP port 53, UDP port 53)

This protocol is used to resolve a hostname to an IP address. For example, it is much easier for a user to type `https://www.packtpub.com` than an IP address such as `104.22.1.175`.

Dynamic Host Configuration Protocol (DHCP) (UDP port 67)

This protocol automatically assigns an IP address when a host joins the network. This allows for the automatic administration of the network as hosts enter and leave the network. Some of the information that the DHCP server will provide includes the IP address, subnet mask, default gateway, and the IP address for the DNS server.

Hypertext Transfer Protocol (HTTP) (TCP port 80)

This protocol is used for communication between a browser and a Web server.

Hypertext Transfer Protocol Secure (HTTPS) (TCP port 443)

This protocol is a secure version of HTTP. It creates a secure communication channel between the browser and server.

Remote Desktop Protocol (RDP) (TCP port 3389)

This protocol is used to create a secure connection to another host. It is similar to Telnet and SSH, but the user has access to a GUI interface instead of the command line. This is a proprietary Microsoft protocol.

Transport layer protocols

Transmission Control Protocol (TCP)

A connection-oriented protocol used for the reliable delivery of data.

User Datagram Protocol (UDP)

A connectionless-oriented protocol used for the fast delivery of data; it does not require the reliable delivery of data.

Internet layer protocols

Internet Protocol (IP)

The Internet protocol is the Internet layer of the TCP/IP stack, with the remaining Internet layer protocols supporting it. This protocol is responsible for the delivery of data packets. This is accomplished solely based on the logical address, also known as the IP address. IP is a connectionless-oriented protocol. There are currently two versions, IPv6 and IPv4.

Internet Control Message Protocol (ICMP)

ICMP is the management protocol and messaging service for IP. ICMP can provide information about issues within the network, such as:

- Destination unreachable – this error is generated when the datagram is unable to reach the final destination.
- Echo request/reply – this is generated with the Ping command. It is used to test the network connection.
- Time exceeded (TTL) – when the packet fails to reach its destination within the specified number of hops. A hop is when the packet is sent from one router to another.
- Redirect – if the router is able to determine a more efficient route, a redirect message will be generated and sent to the host to update its routing table.

Ping is a utility that is used to identify issues with the network. Ping will send a packet to the destination and then track how long it takes the destination server to respond. In the following example, I used the `ping` command to send packets to IP address `104.22.1.175`, and you can see that the server responded, which took anywhere between nine and 14 milliseconds.

```
User@Server ~ % ping packtpub.com
PING packtpub.com (104.22.1.175): 56 data bytes

64 bytes from 104.22.1.175: icmp_seq=0 ttl=60 time=10.270 ms
64 bytes from 104.22.1.175: icmp_seq=1 ttl=60 time=9.949 ms
64 bytes from 104.22.1.175: icmp_seq=2 ttl=60 time=14.081 ms
64 bytes from 104.22.1.175: icmp_seq=3 ttl=60 time=13.323 ms
64 bytes from 104.22.1.175: icmp_seq=4 ttl=60 time=9.048 ms
64 bytes from 104.22.1.175: icmp_seq=5 ttl=60 time=9.077 ms
64 bytes from 104.22.1.175: icmp_seq=6 ttl=60 time=9.254 ms
```

If the ping command were unsuccessful, there would be a message that the destination host was unreachable or that there was no reply in the request timed out. This does not mean there are conductivity issues; system administrators have the option of blocking ICMP traffic that originates from outside of their local network, as shown in the example below.

```
User@Server ~ % ping 192.168.86.22
PING 192.168.86.22 (192.168.86.22): 56 data bytes
Request timeout for icmp_seq 0
Request timeout for icmp_seq 1
Request timeout for icmp_seq 2
Request timeout for icmp_seq 3
ping: sendto: No route to host
```

Address Resolution Protocol (ARP)

ARP is used to translate logical addresses to physical addresses. This is accomplished when ARP sends a broadcast message to the hosts on the same network segment. The broadcast message asks the network interface device if it has a specific IP address. If the network interface device is assigned that IP address, then the network interface device will reply with the MAC address. ARP is a broadcast protocol that will result in all hosts on the network segment being able to see the request and replies. ARP requests are not routable, which means the ARP packets will not be able to cross router boundaries.

Internet Protocol Security (IPSec)

IPSec is a security protocol that enhances security by providing end-to-end encryption and packet authentication. VPNs (Virtual Private Networks) utilize IPSec when creating their secure communication channels.

Summary

In this chapter, we have discussed the fundamentals of a computer network. First, you learned about the differences between the OSI model and the TCP/IP (DoD) model. The OSI model is only theoretical and does not have a physical implementation. Don't be confused when someone mentions the TCP/IP model when talking about the TCP/IP suite of protocols. We also looked at the different types of hardware you may find in a networking environment to understand the differences between a hub and a router. A router is a much more complex device responsible for sending data to the next hop. In contrast, a hub is only used to extend the network segment by giving additional hosts a location to connect to the network. A hub does not "do" anything; it merely creates an extension of the network.

You should be able to explain the differences between a connection-oriented protocol and a connectionless-oriented protocol and in what situations each protocol should be used. For example, it would not be good to utilize a connection-oriented protocol when the throughput of data is the primary concern. The current IP addressing scheme is still being used, but changes are slowly coming and IPv6 will become the dominant protocol.

An examiner should now be able to:

- Determine the differences between OSI and TCP/IP (DoD) models
- Determine which addressing scheme is being utilized on a network
- Understand the differences between a public IP address and a private IP address
- Understand the differences between a port number and an IP address
- Identify some of the commonly used protocols and ports

In the next chapter, we will cover report writing. The ability to draft a report so that the reader can understand a technical subject while using non-technical language is critical. If you cannot explain in writing how you did the exam file, it will affect your ability to be a successful investigator.

Questions

1. How many layers are in the OSI model?

    a. 5
    b. 6
    c. 7
    d. 8

2. How many layers are in the DoD model?

    a. 4
    b. 5
    c. 6
    d. 7

3. What is layer 1 of the OSI model?

    a. Application
    b. Session
    c. Network
    d. Physical

4. Which of the following is a layer 3 device?

    a. Hub

    b. Router

    c. Switch

    d. Printer

5. How long is an IPv4 address (in bits)?

    a. 16

    b. 32

    c. 64

    d. 128

6. How long is an IPv6 address (in bits)?

    a. 16

    b. 32

    c. 64

    d. 128

7. What is the following "2001:db8::8a2e:370:7334" an example of?

    a. A MAC address

    b. A physical address

    c. An IPv6 address

    d. An encryption key

8. Which of the following is NOT a response for a successful ping command?

    a. 64 bytes from `104.22.1.175`

    b. ttl=60

    c. time=10.270 ms

    d. No route to host

9. What is the management protocol and messaging service for IP?

    a. ICMP
    b. DHCP
    c. TYFMS
    d. ARP

10. What protocol allows the creation of a secure telnet session?

    a. FTP
    b. HTTPS
    c. SSH
    d. IFYKYK

Further reading

Neil, I. (2018). *Comptia Security+ Certification Guide: Master It security essentials and exam topics for Comptia security+ sy0-501 certification*. Packt Publishing Ltd.

Davies, G. (2019). *Networking fundamentals: Develop the networking skills required to pass the Microsoft Mta Networking Fundamentals Exam 98-366*. Packt Publishing Ltd.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

12

Report Writing

I have worked with examiners who loved getting down to the bits and bytes of investigation. No one worked harder as they examined the digital evidence, tracking the digital breadcrumbs until they had the evidence they needed. They were intelligent and brilliant, and if I had committed a digital crime, I would not want them to investigate it. It had nothing to do with their ability to investigate and everything to do with their ability to write a report. To say their report was lacking is a massive understatement.

Report writing is one of the hardest things you can do as a digital forensic investigator. You must take a very technical subject and explain it in a manner that a non-technical person will understand while not making any assumptions about the potential user or the digital evidence.

We will cover the following topics in this chapter:

- Effective note taking
- Writing a valuable report

Effective note taking

Your ability to take notes will directly impact your ability to write an effective report on your digital forensic investigation. Your notes will be the foundation of your reporting. A simple phrase that has impacted me as I conduct my exams is *if you do not write it down, it did not happen*. One of your examinations may take days or months; you will simply not be able to remember what exactly you did on day 14 of your examination.

The fundamental elements of notetaking should include the following:

- When you did something
- What you did
- What you saw
- Why you did something

Your notetaking starts when you get the notification, and you have to respond to the scene. This includes the date/time when you are notified, who notified you, and when you arrived at the scene. Document any actions you take; if you collected volatile data, RAM from the system in question, did you alter the digital evidence? The answer will be *yes*. This is where the *why* is essential. *Why* did you alter the digital evidence? Here, the answer is simple – because the evidence would be lost if it was not collected at that time.

Another example if you are responding to a commercial business is whether the digital evidence is contained in a server environment. You cannot (in most cases) shut down the server to create a whole disk forensic image; you will have to create a logical forensic image of the files in question. Once again, the question of *why* may be asked, and you will have to explain.

If the matter you are investigating goes to trial, the opposing counsel will have access to the same digital evidence and the notes taken during the investigation. They will use your notes and report to recreate your examination of the digital evidence. They are attempting to see whether they can get a different result or reach a different conclusion based on your actions.

How detailed should your notes be? The format of your notes is typically personalized to each digital forensic investigator. The baseline consideration should be, if the matter goes to trial years later, can you remember the details of the investigation?

There is no note standard, but you should include the following information:

- Details of the subject under investigation.
- Details of individuals or entities harmed by the incident under investigation
- Location of the digital evidence at the scene.
- Specifics of the digital evidence, make, model, the serial number of the system, any identifying marks (also include damage – there have been claims made that the system was damaged after it was seized. If you document the condition of the system(s) at the time of seizure, this will remove the effectiveness of those complaints).
- Condition of evidence bags/seals – if there is damage or broken seals.

- Details about the forensic hardware that was used, such as firmware/serial number.
- Details about the forensic software that was used, such as version number.
- Any findings that support or do not support your hypothesis about what occurred.

As a minimum, that is the information you should include in your notetaking. Can you incorporate more? Absolutely!

What medium should you use to take notes? I prefer handwriting at the scene and then transferring the notes into the digital medium. My handwriting is not the easiest to decipher, which is why that is my method. Also, digital photos are a medium for notetaking. When documenting the system's condition or the storage device, it is straightforward to take a digital image and then refer to the image when completing your report.

Each organization/examiner will have their own standards about what information needs to be recorded and the method to use to record information. No matter which method you use, it is crucial that you are organized and consistent throughout the entire investigation.

As you can see, notetaking is the foundation for writing the report, which is our next topic.

Writing the report

The purpose of your report is to document the results of your forensic examination. Your report may be used to support additional investigative endeavors. The report may also be used in criminal court proceedings, civil court proceedings, or administrative proceedings. Others can use your findings to support a probable cause hearing, grand jury proceeding, or as a basis for an administrative sanction in the corporate environment.

Your report will be the first step in providing testimony regarding the matter you are investigating. The opposition will scrutinize your report and if they call you to testify, expect to be questioned about the content of the report you created.

As you prepare to draft a report, identify who will be your audience. Suppose you are writing the report for the Chief of Information, the IT security section, or any technology-based group. In that case, your report should go into much greater technical detail than the report directed toward lawyers, judges, or juries. If you go into minute detail about every artifact you found, you will lose your non-technical audience. While the technical audience will want those types of specific details and may feel insulted if you explain the details in a non-technical manner, it is possible to draft a report that addresses the technical and non-technical audience.

The following is a general template you can follow:

- Administrative information
- Executive summary
- Narrative
- Exhibits/technical details
- Glossary

The administrative section will contain information about your investigation, such as the following:

- The name of the agency, the case number(s), and the participants in the investigation. This will include information about the investigators, the victim(s), and the suspect(s). If the investigation started with another agency, you would also include the administrative information from that organization. Include a brief history of the investigation.
- When was the investigation started and what events transpired before you were assigned to the investigation? This could be who was interviewed or interrogated, or any search warrants prepared and served. You are providing a synopsis of the investigation before your involvement. Include the search authority you have in order to investigate/examine the evidence. Include what you are investigating, that is, the scope of the search, and who authorized the search. If the digital forensic examination is being conducted pursuant to a search warrant, include the search warrant and the affidavit used as an exhibit in the *Exhibit/technical details* section.

The executive summary is a section that summarizes the report. The narrative of the report will go into much greater detail than the executive summary. When the reader is finished reading the executive summary, they should have a high-level view of what occurred in the investigation. The executive summary should follow the following guidelines:

- Should be only 10 percent of the report
- Written in short, clear, concise paragraphs
- Should follow the same timeline as the narrative
- Should not include any information not included in the narrative
- Should contain your findings/conclusions

This allows for the non-technical user to understand what actions you took during the investigation without going into technical detail. For example, if you found illicit images within the user's picture folder on a Windows 10-based operating system, you could report that fact in the executive summary like so:

```
During the computer forensic exam, I found 10 images, whose content depicted
what appears to be a male and female juvenile engaged in illicit activities.
The images were located in the Pictures folder of the user account.
```

Your non-technical audience will understand exactly what you intended. Most consumers are familiar with the Windows operating system and how the `Pictures` folder of the user's account is accessed and used. In the narrative section, you can include a more detailed explanation, such as the following:

```
I conducted an exam on evidence tag 2016 – 001, which is an Asus laptop serial
number ABC 00 DEF. I identified one user account "bad guy 27" which had an
RID of "1005". In the folder labeled "Pictures," I found the following images:
001.jpg, 002.jpg, and 003.jpg , which depicted what appears to be a male and
female juvenile engaged in illicit acts in violation of state law NRS 200.481.
The folder and the pictures had ownership properties associated with RID 1005,
"bad guy 27". Additional technical details about the image(s) are contained in
exhibit #1 in the technical details section of this report.
```

Clarity is one goal you seek to achieve as you draft the narrative. You do not want the reader to have questions or be unclear about your report. This can be difficult as you are combining the technical and non-technical aspects of the investigation. You also do not want to overwhelm the reader with technical details and acronyms. If you are in the criminal justice environment and the prosecuting attorney will read your report, you will most likely educate them on the technical aspect. Define the technical terms and concepts within the detailed narrative. How detailed does the narrative need to be?

There is not an easy answer. You should detail the narrative enough to inform the reader about the investigation, so it should suffice for a judge, jury, or lawyer to understand if you are not available to answer questions. Can your investigation be recreated based on the details in your narrative? The opposing counsel will have the ability to review the evidence and your report.

If there is not enough detail for them to recreate your actions, it gives them the ability to question your results. Remember, it is possible that the judicial proceeding will take place months or years later. Your report will be the official memory of your organization for what occurred during that investigation.

You also want to ensure that the narrative is not biased. Your goal is to report the facts without overstating/understating their importance. During the investigation, one of the hardest things is identifying the physical person behind the keyboard. You will base your identification on the digital identification of the user account. You are correlating the user account with the physical person operating the keyboard, using additional sources of digital evidence.

The narrative should contain various subsections, all of which we will go over now.

Evidence analyzed

In this section, you will include all the evidence you have examined, including the make/model, serial numbers, and so on. If it is a desktop/laptop, you should include the hard drives as a separate but related item.

The following is an example of the evidence that could be examined:

Item Name	Tag Number	Description
Compaq Presario	Tag1	Compaq Presario Laptop Computer
Toshiba HD	Tag1 HD001	256 GB Toshiba SATA Hard Drive from the Compaq Presario Laptop Computer
SanDisk Cruzer	Tag1 TD001	128 GB SanDisk Cruzer Glide Thumb drive

Figure 12.1: Evidence Tag example

In this example, the specific item has been identified and assigned the organizational identification number. I have assigned the Compaq laptop the organizational identification number "Tag1."Any storage devices found in the computer will also contain the same tag number. There was a Toshiba hard drive storage device located inside the laptop, so it has the organizational identification number of "Tag1 HD001." HD is an abbreviation for hard disk. If the laptop had two hard drives, the second drive would have the organizational identification number of "Tag1 HD002." If, when the laptop was seized, a thumb drive was found in a USB port, the thumb drive will have the organizational identification number of "Tag1 TD001." TD stands for thumb drive. You can also include the serial number of the item (if it has one) in the description field.

Acquisition details

This section will describe the acquisition process of creating the forensic image(s) as we discussed in *Chapter 3*, *Acquisition of Evidence*. First, identify the hardware or software used in the process and include the serial/version numbers. You should also include the date the hardware/software was verified. Your narrative should consist of a step-by-step analysis of how you (or a colleague) created the forensic image(s). Include descriptions of what steps were performed as expected and also include what did not function as expected. If the forensic image hash value was not verified, include that fact in your report and what steps you took to troubleshoot the issue. Finally, you must understand if there was an issue with creating the forensic image(s). Failing to identify these issues can question the totality of your investigation and the analysis of the forensic image(s).

Analysis details

This section will comprise a large part of your narrative. Your analysis cannot be a printout of pages of files you deemed pertinent to the investigation. You have to analyze the artifact(s) and explain why it is relevant to the investigation to the reader. Include screenshots to help reference the reader to your explanation. Including a screenshot does not remove the requirement to explain what the screenshot is depicting. Tell the reader why the screenshot is important and explain the relevance to the investigation. Do not assume the reader can determine what information in a screenshot is important. There are several different ways to present your analysis. You can do this chronologically, by device, or by the suspect. There is no right or wrong way to write this section. My preference is to create the report chronologically and by subject. For example, for a storage device that was the system drive for a desktop/laptop, I would establish ownership and usage of the device. Then, I would move on to the specific artifacts of the incident being investigated. Be careful not to get too far into the weeds and be overly technical with your technical descriptions. I would include that information with the specific exhibit you are describing in the Exhibits/technical details section for the technical descriptions.

For example, if the date/timestamp of an artifact is pertinent, in the narrative, you can state that the user accessed the application on X X day at X X time. Then, in the next section, you could go into more detail about the byte offset for the date/timestamps in the file record in the MFT.

Be careful when dealing with absolute statements or using unnecessary adjectives. I once read a report that described the user's Google searches as disturbing. You do not want to categorize the behavior/actions you find while doing your digital forensic exam. Your duty is to provide the facts to the fact finder, the judge/jury, and allow them to make that determination.

At the end of the narrative is the time to present your conclusions/findings. This is where you offer your opinions on the subject's culpability in your digital forensic investigation. Keep it short and straightforward – for example, *based on my examination of the following evidence (list the items you examined in the course of your digital forensic examination), it is my opinion...* – and then lay out the facts based on the artifacts you analyzed. You want to avoid any inflammatory/descriptive language and remain unbiased and professional.

Exhibits/technical details

As you create the narrative in the *Analysis* section, you will reference specific artifacts. You should place the screenshots of those artifacts in the Exhibits/technical details section. This will also include the output reports of the forensic tool(s) you used in the exam process. If you reference the artifact in the narrative, you must include it in the Exhibits/technical detail section; likewise, if you have an exhibit in that section, you must reference it in your narrative. I find it helpful to organize the exhibits and technical details in the same order I referenced them in the narrative. It helps the reader comprehend the report's content if they view the exhibits after reading the narrative and they are in the same order.

In the following example, I have included the owner information of the operating system, along with the install date/timestamps, time zone, and last shutdown date/timestamp:

Compaq Presario (Tag1 HD001)	
Product Name	Windows 10
Computer Name	BadGuy Laptop
Registered Owner	BadGuy27
Install Date	August 13, 2018, 08:52:58 (Local)
Last Shutdown	October 12, 2018, 23:44:11 (Local)
Time Zone	Pacific Standard Time

Figure 12.2: Evidence example

An appropriate narrative of the information would be as follows:

```
The system stores the operating system information in the SYSTEM hive of the
operating system. The data was located in the CurrentVersion sub-key. The fields
"Computer Name" and "Registered Owner" are user inputted values. The operating
system was installed on August 13, 2018, at 08:52:58 PST. The registered owner
information is BadGuy27, and the System name is BadGuy Laptop.
```

The description is concise, factual, and unbiased, which is the goal of the report.

The final portion of this section will be a table of software/hardware used. You want to include the version numbers of the software/firmware so that others can repeat your examination. You also want to make sure that the organization licenses for your software are authentic. This can be a simple list, as shown here:

- FTK Imager 3.0.0.1443
- X-Ways Forensics 19.7
- Paladin 7.05
- Recon 3.14.1.12

There have been issues where an organization has used unlicensed/pirated software in the exam process. This is not recommended and can result in negative sanctions against you and your agency. In addition, the use of unlicensed/pirated software can call the validity of your findings into question.

I cannot stress enough how important it is for you to proofread your report. You will not create a perfect report the first time around (or the second time). You will have grammar, spelling, and content errors in the report. Therefore, you should always have a second person proofread the report after making the first draft. The second person will find mistakes you missed and help determine how the report flows from one section to the next. You already have an idea in your head of what to say; the second proof-reader will help determine if this is effective. The second proof-reader will also bring a different insight to the report's presentation and help ensure that your conclusions are logical and without bias. Whenever I proofread a colleague's report, I would look at it from the opposition's point of view. My goal was to find inconsistencies or gaps that the opposition could exploit if the matter went to trial. Another option is once the report has been finalized, your organization may have a peer-review process. A supervisor or colleague reviews your report and findings to ensure you followed the appropriate policies and procedures and that your conclusions are fact-based.

What format should your report be in when it is disseminated to the stakeholders? My preference was to deliver my reports in a digitally signed PDF. If the report is altered, it will break the digital signature. Some digital forensic investigators will create an HTML-based report and burn it onto an optical disk, while others will use the reporting function of their forensic tool. You can use many options in the report's presentation; you will want to make sure you can authenticate its contents if you have to testify in a judicial/administrative proceeding. A PDF will allow you to authenticate with a digital signature while saving the report and related data to an optical disk, which will enable you to create a hash value to ensure that no one has changed the contents.

Summary

This chapter has discussed notetaking and how important it is to take quality notes. You learned that notetaking is the fundamental building block in creating your report. You understood the makeup of a digital forensic examination report and what information it should include. We also discussed that technical and non-technical readers may read your report and that you have to draft your report with that in mind. With that, you are able to take effective notes and prepare a clear and concise report.

In the next chapter, we will discuss the culmination of your investigation and report writing, taking the stand as a witness.

Questions

1. You should start taking notes _____?
    a. When you receive notification
    b. When you get to the scene
    c. When you start the exam
    d. When you start the report
2. What information should you include in your notes?
    a. What you had for breakfast
    b. The shoe size of the suspect
    c. Location of the digital evidence at the scene
    d. Weather conditions
3. There is a national standard for notetaking.
    a. True
    b. False
4. When drafting the report, who should you keep in mind?
    a. Supervisor
    b. Chief of police
    c. District attorney
    d. The reader

5. What information is not contained in the *Administrative Information* section?
    a. Your birthday
    b. Agency name
    c. Suspect information
    d. Witness information
6. The Executive Summary should not exceed 25 percent of your report.
    a. True
    b. False
7. What should your draft report be?
    a. Detailed
    b. Brief
    c. Clear
    d. Efficient

The answers can be found at the back of this book.

Further reading

For more information, you can refer to *Forensic Examination of Digital Evidence, A Guide for Law Enforcement*, from `https://www.ncjrs.gov/pdffiles1/nij/199408.pdf`.

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

13

Expert Witness Ethics

This is the final step in your digital forensic investigation: you, as the investigator, have received a subpoena to testify in a judicial or administrative hearing. Now, it is time for you to explain your actions and findings to an unbiased third party: the jury. It does not matter how good or strong the evidence you found during your digital forensic investigation is if you cannot testify effectively. You must be able to testify and authenticate the evidence in your actions.

I know some digital forensic investigators who hate to testify. They love the collection of evidence; they love doing the exam and finding the relevant artifacts, but to get them into a judicial/administrative proceeding is very difficult. The first time you walk into a courtroom, it can be intimidating. You may not know the rules or the procedures, and you may be afraid of making mistakes. To overcome those issues, you will need to prepare yourself.

We'll be covering the following topics in this chapter:

- Understanding the types of proceedings
- Beginning the preparation phase
- Understanding the curriculum vitae
- Understanding testimony and evidence
- Understanding the importance of ethical behavior

So, let's jump in and start talking about how to prepare to testify in a judicial/administrative proceeding.

Understanding the types of proceedings

There are a variety of proceedings where you may be called to testify or to present evidence. The following are US-based proceedings; your jurisdiction may differ. We will discuss some of the more common proceedings that you may encounter:

- **Grand jury**: A grand jury is a panel of citizens empowered to investigate potential criminal conduct and to determine whether the conduct requires criminal charges. A grand jury will have subpoena powers that could include compelling testimony or requesting physical evidence.
- **Arraignment**: This is the formal reading of a criminal complaint. The accused is present and informed of the charges. At this hearing, the accused will plead guilty/not guilty.
- **Detention hearing**: This is a proceeding before a judge to determine whether the accused is to be detained/released while the matter is progressing in the criminal justice system.
- **Evidentiary hearing**: This is a hearing before a judge to examine the potential evidence that will be presented to the jury. The judge can exclude or limit the evidence that may be offered.
- **Trial**: This could be a criminal or civil matter. Here, both sides present evidence to the fact-finder (judge/jury), and you may be called to testify during the case in chief, as well as the sentencing portion.
- **Deposition**: This is a sworn testimony that occurs outside of the confines of the court and is commonly used in civil litigation proceedings. Typically, there will not be a judge present, only the attorneys and the witness. The record of the deposition may be presented at a later time to a finder of fact.

As you can see, there are several occasions where you may be called to testify or to take part. Treat every avenue as if you are taking part in a jury trial. Ultimately, it should not matter in which aspect of the system you will be testifying. You will still need to prepare in the same manner.

There are many moving parts within the courtroom. Let's discuss some participants you may encounter as a digital forensic investigator:

- **Judge**: This is the supreme being overseeing the matter. The judge will determine all motions during the trial.
- **Court reporter**: A court officer responsible for creating the official record of proceedings.
- **Court clerk**: A court officer responsible for administrative issues within the courtroom.
- **Bailiff**: A court officer responsible for maintaining order and dignity in the courtroom.

- **Prosecutor**: A representative of the sovereign who will present the government's case against the accused.
- **Defense attorney**: Represents the accused in the matter being presented in the courtroom.
- **Plaintiff**: In civil proceedings, this is the party that is claiming that the actions of the defendant have harmed them.
- **Defendant**: The accused in a criminal/civil matter.
- **Jury**: A panel of citizens who will determine the guilt/innocence of the defendant.
- **Witnesses**: Individuals who have knowledge of the incident in question and can present evidence in the matter.

As a digital forensic investigator, your role may be an expert witness. Being called to testify can be stressful; remember that your ability to reduce this stress lies in your preparation. Part of your preparation will be becoming familiar with the process and its participants. The more knowledge you have of your jurisdiction's criminal and civil procedures, the more comfortable you will be as you navigate through the process.

The process begins when the accused is arrested, or a warrant is issued in a criminal matter. The accused is taken before a judge (after being arrested) and is arraigned. A preliminary hearing is held, where the judge decides whether probable cause exists to go forward with a trial. Alternatively, the matter could go before a grand jury, where they will determine whether an indictment is warranted.

There will be several hearings before the matter is presented at a trial. Once the matter goes to trial, the prosecution representing the sovereign will present their case through the presentation of evidence and witnesses. The defense will be able to cross-examine each witness immediately after the prosecution has conducted its direct examination.

Once the government rests its case, the defense can present a case, or if they feel that the government did not present enough evidence to overcome reasonable doubt, they may rest also.

Once both parties have rested, then the judge will give the jury instructions on how they must proceed in their deliberations.

Before you get to the proceedings where you may testify, you must be ready. This is not something you can walk into without preparing.

Beginning the preparation phase

As a digital forensic investigator, your role in a judicial/administrative proceeding can be defined in two ways:

- **Witness (also referred to as a lay or fact witness)**: You will testify about events you observed. You are just presenting facts that you have personal knowledge of, such as where the evidence was found.
- **Expert witness**: You can testify to everything a lay/fact witness can, but now you may offer your opinion. You form your opinions based on your training and experience as a digital forensic investigator. It is your ability to provide an opinion that makes you an expert witness.

Your preparation starts with your participation in the investigation. You should treat every investigation as if it will go to trial and you will have to testify. No matter which side you are on in the judicial/administrative proceeding, start communicating with the attorney at the very beginning. Discuss what they need for a successful outcome. You want to learn everything you can about the participants, the suspects, victims, and attorneys of the proceeding. Educate yourself about the points of dispute in the proceeding.

For example, suppose the point of dispute is the wilful and knowing possession of illicit images. What artifacts show/do not show that the subject willingly and knowingly possessed the illicit images? As you work to answer this question, you must inform the attorney when you find information that proves or disproves the point of contention.

I can almost guarantee there will be an expert witness on the opposition team. You will want to learn about them. You will want to review their curriculum vitae and learn about their experience, education, and certifications. If possible, review their prior testimony.

I remember one incident where counsel called me to testify as an expert witness in a motion hearing. During the hearing, I was on the stand for over 4 hours, questioned by the prosecution, the defense, and the judge. Once I had completed my testimony, the opposition's expert witness was called to the stand. One of the first questions asked by the judge was, *"You have heard Mr. Oettinger's testimony. Is there anything you disagree with about his opinions on the state of the evidence?"* The opposing expert witness thought for a moment and replied, *"No."* I will say that it was a powerful moment for me. Having another professional in my field validate my findings and opinions during a contested trial is something I strive for.

As you are preparing for your testimony, you are trying to answer the following questions:

- What is the theory of the case?
- Does my theory fit within the facts of the case?
- What facts are central to my testimony?
- What facts can I confirm or cannot confirm?

I cannot emphasize this enough: review your report and your notes before you take the stand and testify in the proceeding. Practice answering questions. Key to the preparation phase is working with the attorney to have a clear understanding of the state of the evidence and your interpretation of the evidence. Before you can be appointed as an expert witness in a matter, you will have to be approved by the judge.

To begin the review process, you will have to submit a curriculum vitae, which is the next topic.

Understanding the curriculum vitae

A **curriculum vitae** (also known as a **CV**) is a document you create that outlines your education and experience, your certifications and membership, and professional organizations. The court and attorneys who determine your qualifications as an expert witness will use your CV to make that determination. The contents of your CV will contain a synopsis of *what* makes you an expert; it will highlight all the experiences that make you an expert in your field.

There is no specific format you have to use when creating your CV, but all of them will contain the same content as it is the history of your professional life.

At the top of the CV will be your name and contact information. This ensures that your name is spelled correctly throughout the proceeding and when added to the witness list. You will also want to identify the field you are an expert in. If the attorney, judge, or court clerk is dealing with multiple experts in a matter, this helps to identify the area of testimony that you may be asked to speak about. You will also want to include a contact number, email address, and physical address. This allows for all parties to contact you. Also, others may share your CV with other attorneys in different matters, and they will use that information to solicit you for additional opportunities.

Note

A note about the address used on your CV: you should not include your home address. You may testify in a matter that could deal with physical violence or the potential for incarceration. It does not matter which side you testify on; someone will likely be unhappy with the results. If you are working for an organization, use the organization's address. I recommend getting a PO Box or a private mailbox if you are working for yourself.

You will then create a summary of your biography. This will include a synopsis of your career, education, and experience.

Next is the bulk of your CV, where you list your formal education and work history. Again, you can use the following categories to organize the information being presented:

- **Formal education**: Degrees and, certificates awarded.
- **Employment history**: As it relates to the field.
- **Teaching experience**: This will cross over with your employment history. Keep it relevant to the field you will testify about in the proceeding.
- **Licensing/professional membership**: List the relevant professional organizations you belong to. If the government requires licensure, be sure to include that as well.
- **Publication**: If you have authored a book, white paper, an article, or a blog, identify the publisher's name and address, as well as when the item was printed.
- **Awards**: If you have received an award for your work in the field, please list it.
- **Previous testimony**: You should list the previous cases where they have appointed you as an expert. This does not have to include a summary of the matter; instead, the use of a simple *US v Smith (2015)* will suffice.

Do not get caught up and overthink the CV by including *everything*. You will want to keep the content pertinent to the specific matter and what subjects will come into play during your testimony. You want to stay focused on the field-specific items; whether you graduated high school or worked at a fast-food restaurant during college isn't relevant. You only provide the information needed for the judge/attorneys to determine whether your education and experiences qualify you as an expert witness.

When drafting your CV, I cannot emphasize enough that you refrain from adding information that is not true and accurate. I understand wanting to *pad* the document to make it appear you are the best candidate. Still, if you lie about your CV and continue to lie after being appointed as an expert and testifying, you could face severe repercussions when the lie is discovered.

Note

In 2016, the government arrested Chester Kwitowski after providing false information about his education, experience, and credentials. At his arrest, defense teams had hired him for five other pending matters. He also provided expert testimony in state and federal courts over 50 times. The prosecutors determined that the educational degrees Kwitowski claimed to have received did not exist, nor was there any record he had completed the professional certifications he claimed. Kwitowski claimed to have worked with NASA, but the organization denied any involvement or work history with Kwitowski. Kwitowski also had a criminal record dating back nearly 20 years that included battery, domestic violence, and aggravated battery with a deadly weapon.

After you are requested to be an expert in the matter, and you have submitted your CV, there may be a hearing to determine your qualifications as an expert. The bailiff will swear you under oath, and you will take your seat on the witness stand. Next, each counsel will ask you questions to assess your qualifications to be an expert in the matter at hand. In some jurisdictions, the judge may also ask you questions. Finally, the judge will make a ruling to either approve or disapprove of you acting as an expert.

If the judge approves you, you will work closely with the attorney that requested you and work with them to determine the pros and cons of the matter. On the day of the trial, the attorney may call you as a witness in the matter, and you will have to testify. We will cover this in the next section.

Understanding testimony and evidence

You are at the point in the trial where you are asked to take an oath and promise to tell the truth. You then take your seat, and the room's focus is on you. You may have the judge sitting next to you at an elevated position. Across from you, you may see two tables. One table will be hosting the prosecution, which could be one or more attorneys. At the next table will be the defense, which can also comprise more than one attorney and the subject of the trial.

There could also be a jury box that could contain 12 or more citizens whose job is to determine the guilt or innocence of the accused. Every single one of them is now watching you. This can be a little stress-inducing. Take a deep breath and focus on the questions being asked of you.

Your testimony will comprise technical details and your expert opinion. The technical information will include you explaining complex technical issues in simple terms. This enables the non-technical audience, the judge and the jury, to understand what occurred and how it is being described.

You will want to speak in a slow, deliberate manner. This ensures your audience, including the jury and the court reporter, can understand the concepts you are relaying. You also want to add analogies to help explain the complex technical subject.

I remember a trial I was part of a few years back. I was the defense's expert in a matter that dealt with digital evidence and the possession of illicit images. While reviewing the reports, there were issues in the method used in the seizure of digital evidence. Based on the information in the reports, the computer systems were not seized in the technique that would conform to best practices. I informed the lead attorney about these issues as he was preparing to cross-examine the lead agent responsible for the seizure. The lead agent did not have a significant amount of experience testifying. As a result, during the cross-examination, the agent was not testifying as effectively as they could have. When the attorney started the subject of the seizure of the digital evidence, the agent admitted to violating the "prime directive" of seizing digital evidence.

I asked myself the same question going through your mind: *What is the prime directive?* My only reference for the prime directive is watching episodes of the TV show Star Trek. What occurred is that the attorney conflated best practices with the prime directive, which the agent agreed he violated.

Once the trial was complete, I had coffee with the agent, and I asked him why he answered that he violated the prime directive. He stated that he had just been worn down by the defense attorney's questions and did not want to appear stupid in front of the jury or his peers. I can understand that. Let's now talk about how to prevent that.

If the lawyers ask you a question you do not understand, it is perfectly acceptable to answer, *"I am not sure what you are asking. Could you rephrase the question"* or *"I do not know."* All are very valid answers. You may also be asked a question outside your expertise. Answer that question as *"that is beyond the scope of my expertise"* or *"that was not part of the investigation."*

Lawyers love to ask exceedingly complex questions; you have the right – if not the duty – to ask for clarification for any question you do not understand. Sometimes, lawyers will ask you a question that requires a narrative answer but they want a firm *yes* or *no* answer. Your answer should be, "that is not a *yes* or *no* question, but one that requires a more detailed response."

Your words are not the only thing your audience uses to grade your credibility. Your physical appearance, tone, and posture convey your attitude to your audience. For example, if you take the stand in a rumpled suit, tie undone, and shirttail untucked, you will not be as effective as wearing a freshly pressed suit, properly tied tie, and looking and answering questions like a professional.

The following are some guidelines to consider when testifying:

- **Do not argue with the attorney**: You are an unbiased professional. You need to answer the questions to the best of your ability. If you get into an argument with the attorney, it does not help the audience understand the evidence. In fact, they may discount your testimony because of the appearance of bias.
- **Speak clearly and slowly**: If your audience cannot understand what you are saying, you are not as effective as a witness.
- **Avoid slang and acronyms**: Remember that you are translating a technical topic for a non-technical audience.
- **Do not be a comedian**: Do not make a joke; this is a serious situation. Someone's freedom could be at stake; it is not the place to be humorous.
- **Listen to the entire question**: Do not interrupt the attorney and try to answer what you think the question is. Only answer the question that was asked.

Remember, you are an unbiased advocate. Your job is to assist the fact-finder in determining what occurred based on the evidence.

Digital evidence caused some issues with the rules of evidence when first used in a judicial proceeding. Therefore, you will want to follow all the best practices in collecting digital evidence to ensure its integrity. By demonstrating your efforts to ensure the integrity of the digital evidence, you will reduce the likelihood of the judge excluding the digital evidence.

All evidence must be authenticated. This means there must be a witness to testify about their knowledge of the evidence being presented. For example, if a photo is being presented as evidence, the photographer must attest that they took the photo.

With digital evidence, the digital forensic investigator must testify that the evidence being presented is based on an exact and true copy of the original. Remember, we do not want to conduct our digital forensic examination on the original evidence. Digital evidence is fragile, so we need to create an exact and true copy that can be validated using hashing.

For the evidence to be admitted in court, it must be reliable and credible, relevant to the facts of the matter, and material to an issue being questioned. If you collected the evidence in a manner the court determines to be illegal, then that evidence is tainted and can be excluded.

When you (or someone in your organization) collect the digital evidence, you want to ensure that you preserved the original evidence in the state in which you found it. If you collected volatile data, explain your reasoning for doing so to the court. The collection of volatile data will cause changes to the system and alter the original state of the evidence.

As you can see, this process can be overwhelming. While conducting your digital forensic investigation, you may find yourself in a situation where you question what the *right* thing to do is. This is an ethical dilemma, which leads us to our next topic.

Understanding the importance of ethical behavior

You are responsible for conducting due diligence, being truthful, and being objective during your digital forensic investigation. Your personal and professional ethics determine the baseline of your behavior. Failure to act ethically during your digital forensic investigation can cause the evidence to be excluded and/or result in you facing professional repercussions.

As a digital forensic investigator, you have specialized knowledge that has the potential for misuse. Failure to follow up on potential leads you discovered during your forensic examination is an ethical lapse that could have repercussions on you, a third party, or your organization.

What is the definition of ethics? It is the moral principles that govern an individual's behavior or activity. It is not a distinct standard; it will depend on your culture to determine what is acceptable and what is not. An organization may declare a professional set of ethics in a professional setting.

The **International Association of Computer Investigative Specialists (IACIS)** is an organization I belong to, and because of my membership, I agree to follow their Code of Ethics.

The following Code of Ethics is taken from:

`https://www.iacis.com/wp-content/uploads/2019/11/IACIS-Code-of-Ethics-and-Professional-Conduct-Ver-1.4.pdf`

- IACIS personnel will advise and provide assistance to other IACIS personnel within the scope of their legal authority.
- IACIS personnel will be honest and ethical when dealing with each other.
- IACIS personnel must respect the rights and authorities of the directors, fellow members, and individuals encountered as a result of their membership in IACIS or in connection with IACIS sponsored or sanctioned activities.
- IACIS personnel's actions, when representing or acting on behalf of IACIS, must be free from discrimination, libel, slander or harassment. Each person must be accorded equal opportunity, regardless of age, race, sex, sexual preference, color, creed, religion, national origin, marital status, veteran's status, handicap or disability.
- IACIS personnel may not misrepresent their credentials, employment, education, training and experience, or membership status; nor may they misrepresent the credentials, employment, education, training and experience, or membership status of any other member of IACIS.
- IACIS personnel may not issue public statements that appear to represent the position of IACIS without specific written authority from the Board of Directors. IACIS personnel must not commit any act of professional dishonesty.
- IACIS personnel may not knowingly submit, aid or abet the submission of plagiarized or any non-uniquely authored piece of work during any phase of an IACIS certification process or test. To do so will be considered to have been a dishonest act.
- IACIS personnel have an obligation to report acts or suspected acts of dishonesty committed by IACIS personnel. Failure to report acts or suspected acts of dishonesty will be considered to have been a dishonest act.
- IACIS personnel's criminal convictions are a serious affront to the ideals of IACIS and as such are not tolerated.
- IACIS personnel have an obligation to fully and honestly cooperate with any investigation or inquiry conducted at the direction of the IACIS Ethics Committee or members of an IACIS Investigative Team.

Does that code of ethics apply to all digital forensic investigators? No. The ethics of an organization apply specifically to that organization. You can take that framework and use it in your professional and personal environments. You will notice that the only portion dealing with a digital forensic investigation is the one that states that IACIS personnel must not commit any act of professional dishonesty.

That is a broad statement. It is not a clear line in the sand of what is allowed or not allowed. Determine what is ethical as you perform your duties as a digital forensic investigator.

The **International Society of Forensic Computer Examiners** (**ISFCE**) has a much more specific code of ethics regarding professional behavior during a digital forensic investigation.

The following Code of Ethics has been taken from `https://www.isfce.com/ethics2.htm` for your reference:

- *Demonstrate commitment and diligence in the performance of assigned duties.*
- *Demonstrate integrity in completing professional assignments.*
- *Maintain the utmost objectivity in all forensic examinations and accurately present findings.*
- *Conduct examinations based on established, validated procedures.*
- *Abide by the highest moral and ethical standards and abide by the Code of the ISFCE.*
- *Testify truthfully in all matters before any board, court, or proceeding.*
- *Avoid any action that would knowingly present a conflict of interest.*
- *Comply with all legal orders of the courts.*
- *Thoroughly examine all evidence within the scope of the engagement.*

This code of ethics contains definitive language about what is allowed or not allowed by members of their organization who have certified as a **Certified Computer Examiner** (**CCE**). Members and non-members alike should use this code of conduct whenever they are conducting a digital forensic examination.

Maintaining a code of ethics in your professional life allows you to keep your objectivity during the investigation. If you cannot be impartial, you should not be a party to the investigation. I recently took part in a motion hearing to determine whether they should appoint me as an expert. After they questioned me about my qualifications, education, and experience, I was asked my opinion about the state of the evidence I had reviewed. On cross-examination, the prosecution told me, "*Your duty as an expert is to find things wrong in the evidence.*" My reply was that as an expert, my job was to see whether I could recreate the examination and achieve the same results and conclusion. If the information I found was detrimental to the theory supported by the defense or the prosecution, I would disclose it no matter which side of the matter I was appointed to represent. With digital forensics, the data is the data; there is not a lot of interpretation about what the data means.

As you gain training and experience, I recommend that you achieve industry-specific certifications. The possession of certifications does not guarantee or make you an exceptional digital forensic investigator.

A certification states that you have met the minimum standards of that organization. It does not mean you cannot make a mistake or come to the wrong conclusion. It also ensures that you are keeping current with the changes in the field. What was acceptable 5 years ago may not be acceptable now because of changes in technology or the law. Your training never stops as you pursue a career in this field.

Ethics is doing the right thing when no one is looking. If you compromise your ethics, you can negatively affect your career and investigation. Remember that your goal is to be unbiased and present the facts of the matter to the fact-finder. You are not an advocate for either side of the matter and you now have the knowledge to accomplish that goal.

Summary

During this chapter, you learned how to prepare to give testimony in an administrative or judicial proceeding. You can now identify the different proceedings and the participants. You can also create a CV and differentiate one from a résumé. You also have the skills to ensure that you conduct your digital forensic investigation and exam while maintaining your objectivity and impartiality through the use of a code of ethics.

Thank you for your efforts and for working through my book! I am confident that you can use the skills you've learned here and apply them to a real-world setting.

Questions

1. An expert witness can offer _____.
    a. Testimony
    b. Facts
    c. Opinion
    d. Hearsay evidence
2. Preparation starts _____.
    a. When you receive a subpoena
    b. When your supervisor tells you to begin
    c. When the judge calls you
    d. When you start the investigation

3. Which court officer represents the sovereign?
    a. The judge
    b. The prosecutor
    c. The court reporter
    d. The bailiff
4. In a trial, the fact-finder will be who?
    a. The jury
    b. The grand jury
    c. The judge
    d. The attorney
5. Which of the following should you NOT include on a CV?
    a. Formal education
    b. Teaching experience
    c. Professional memberships
    d. Salary
6. Which of the following is an appropriate answer to a question you do not understand?
    a. I do not know.
    b. You should try and guess.
    c. Ask to repeat the question.
    d. Look to the judge for help.
7. Why should you adhere to a code of ethics?
    a. To maintain your impartiality
    b. To make sure the correct side wins
    c. To ensure the accused is found guilty
    d. To keep your certification

The answers can be found at the back of this book.

Further reading

Refer to the following sources for more information:

- Smith, F. C., and Bace, R. G. (2003). *A guide to forensic testimony: the art and practice of presenting testimony as an expert technical witness*. Boston, MA: Addison-Wesley (available at `https://www.amazon.com/Guide-Forensic-Testimony-Presenting-Technical/dp/0201752794`)
- Poynter, D. (2012). *Expert witness handbook: tips and techniques for the litigation consultant*. Santa Barbara, CA: Para Pub (available at `https://www.amazon.com/Expert-Witness-Handbook-Techniques-Litigations/dp/1568601522`)

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`

Assessments

Chapter 01

1. False
2. D
3. True
4. D
5. B
6. B
7. A

Chapter 02

1. D
2. False
3. D
4. A
5. A
6. D
7. C

Chapter 03

1. A
2. C
3. True
4. C
5. True
6. B
7. B

Chapter 04

1. B
2. C
3. False
4. False
5. C
6. False
7. B

Chapter 05

1. True
2. C
3. C
4. B
5. A
6. A
7. True

Chapter 06

1. A
2. C
3. B
4. A
5. A
6. C
7. C

Chapter 07

1. A and B
2. C
3. C
4. C

5. B
6. B
7. B

Chapter 08

1. A
2. D
3. B
4. C
5. D
6. C
7. B

Chapter 09

1. A
2. C
3. D
4. C
5. B
6. B
7. C

Chapter 10

1. B
2. B
3. B
4. C
5. A
6. D
7. A
8. A
9. A
10. B

Chapter 11

1. C
2. A
3. D
4. B
5. B
6. D
7. C
8. D
9. A
10. C

Chapter 12

1. A
2. C
3. B
4. D
5. A
6. B
7. C

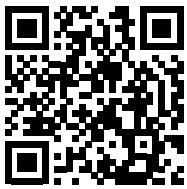
Chapter 13

1. C
2. D
3. B
4. A
5. D
6. A, C
7. A

Join our community on Discord

Join our community's Discord space for discussions with the author and other readers:

`https://packt.link/CyberSec`



packt.com

Subscribe to our online digital library for full access to over 7,000 books and videos, as well as industry leading tools to help you plan your personal development and advance your career. For more information, please visit our website.

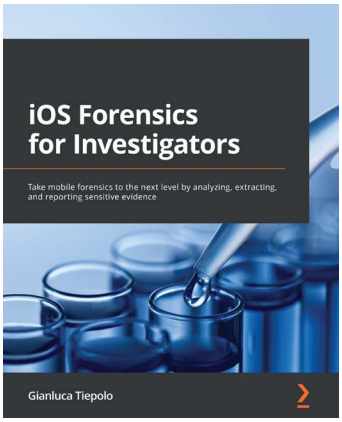
Why subscribe?

- Spend less time learning and more time coding with practical eBooks and Videos from over 4,000 industry professionals
- Improve your learning with Skill Plans built especially for you
- Get a free eBook or video every month
- Fully searchable for easy access to vital information
- Copy and paste, print, and bookmark content

At www.packt.com, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on Packt books and eBooks.

Other Books You May Enjoy

If you enjoyed this book, you may be interested in these other books by Packt:

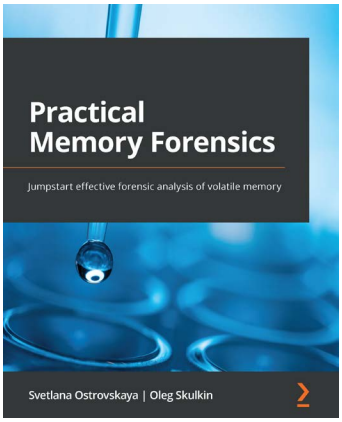


iOS Forensics for Investigators

Gianluca Tiepolo

ISBN: 9781803234083

- Become familiar with the mobile forensics workflow
- Understand how to legally seize iOS devices and preserve their data
- Extract evidence through logical and filesystem acquisitions
- Perform a deep-dive analysis of user data and system data
- Gain insights by analyzing third-party applications
- Get to grips with gathering evidence stored on iCloud



Practical Memory Forensics

Svetlana Ostrovskaya

Oleg Skulkin

ISBN: 9781801070331

- Understand the fundamental concepts of memory organization
- Discover how to perform a forensic investigation of random access memory
- Create full memory dumps as well as dumps of individual processes in Windows, Linux, and macOS
- Analyze hibernation files, swap files, and crash dumps
- Apply various methods to analyze user activities
- Use multiple approaches to search for traces of malicious activity
- Reconstruct threat actor tactics and techniques using random access memory analysis

Packt is searching for authors like you

If you're interested in becoming an author for Packt, please visit `authors.packtpub.com` and apply today. We have worked with thousands of developers and tech professionals, just like you, to help them share their insight with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Share your thoughts

Now you've finished *Learn Computer Forensics, Second Edition*, we'd love to hear your thoughts! If you purchased the book from Amazon, please `click here to go straight to the Amazon review page` for this book and share your feedback or leave a review on the site that you purchased it from.

Your review is important to us and the tech community and will help us make sure we're delivering excellent quality content.

Index

A

active partition 122

addressing schemes, IPv6
- anycast addresses 352
- multicast addresses 352
- unique local addresses 352

Address Resolution Protocol (ARP) 355

Advanced Forensics Format (AFF) 100

Aeon Timeline 176

American Kennel Club (AKC) 335

American Standard Code for Information Interchange (ASCII) 178

analysis process 56, 57
- antivirus 62-66
- dates and time zones 57
- file signature analysis 60-62
- hash analysis 57-60

antivirus 62-66

application layer (Layer 7), OSI model 345

application layer protocols
- Domain Name Service (DNS) (TCP port 53, UDP port 53) 353
- Dynamic Host Configuration Protocol (DHCP) (UDP port 67) 353
- File Transfer Protocol (FTP) (TCP ports 20 and 21) 352
- Hypertext Transfer Protocol (HTTP) (TCP port 80) 353
- Hypertext Transfer Protocol Secure (HTTPS) (TCP port 443) 353
- Remote Desktop Protocol (RDP) (TCP port 3389) 353
- Secure Shell (SSH) (TCP port 22) 352
- Simple Mail Transfer Protocol (SMTP) (TCP port 25) 353
- Telnet (TCP port 23) 353

Ares Galaxy 301
- reference link 301

arraignment 372

Asynchronous JavaScript (AJAX) 262

B

background searches 322-331

badguyneedslove
- reference link 327

bailiff 372

Basic Input/Output System (BIOS) 114

Bind Torture Kill (BTK) 28

blue screen of death (BSoD) 232

bookmarks, Firefox 293

bookmarks, Google Chrome 270-274

bookmarks, Internet Explorer 279

bootable forensic device
creating 117-119
boot process 113-115
forensic boot media 116, 117
GPT partitions 125-130
hard disk drives 119, 120
Master Boot Record (MBR) partitions 122-125
browsers 269, 270
Google Chrome 270
Bulk Extractor 297
download link 238
using 238-243

C

cache, Firefox 289, 290
cache, Google Chrome 277
cache, Internet Explorer 283-285
case information 48-50
Certified Computer Examiner (CCE) 382
cfldd
URL 92
chain of custody 53-56
choices 56
child exploitation material (CEM) 28
child sexual abuse material (CSAM) 28
Chrome Cache View
reference link 277
Chrome Cookies View
reference link 276
Chrome Pass
reference link 278
client-based email
analysis 258
Microsoft Outlook/Outlook Express, exploring 258, 259
Microsoft Windows Live Mail, exploring 259, 260
Mozilla Thunderbird 260
cloud computing 305, 306
command-line interface (CLI) 44, 164, 243, 315
commercial forensic tools
for Windows-based users 46
community cloud 306
computer-based investigations 2, 3
Computer Forensic Reference Dataset
reference link 79
Computer Forensic Tool Testing Project (CFTT) 45
cookie 276
cookies, Firefox 290
cookies, Google Chrome 276
cookies, Internet Explorer 285-287
corporate espionage 19, 20
real-world experience 24, 25
security 20
social engineering 21
threat actor 20
corporate investigations 16
corporate espionage 19, 20
employee misconduct 17-19
insider threat 25-27
court clerk 372
court reporter 372
crash dump (memory.dmp) 232
crime scene technician 5-7
illicit images 7, 8
criminal conspiracy 14, 15

criminal investigations 4
crime scene technician 5, 7
criminal conspiracy 14, 15
first responders 4
investigator 5
curriculum vitae (CV) 375-377
information, organizing 376
cyberbullying 12
cyberstalking 12, 13
Cylinder, Head, Sector (CHS) 122

D

data acquisition 50-52
chain of custody 53-56
data area 136-138
data Attribute 150
Data, hive subkeys 191
data link layer (Layer 2), OSI model 343
dates and time zones 57
dc3dd
features 93
DCode
URL 273
dd command
features 92
DD image 91-93
DeBounce
URL 322
defendant 373
defense attorney 373
deleted data
recovering 180-182
deleted files
recovering 139-141
Dennis Rader 28, 29
Department of Homeland Security (DHS) 30
Department of Justice (DOJ) 32
deposition 372
Description of Evidence field 53
detention hearing 372
Device Configuration Overlay (DCO) 130, 131
Digital Corpora
URL 160
Digital Forensic Examination and Analysis Tools 177
digital forensic investigation
keyword 178
string search 177-180
disposable email address 316
Domain Name Service (DNS) 353
Dropbox 307
Drug Enforcement Administration (DEA) 30
DumpIt
using 234, 235
Dynamic Host Configuration Protocol (DHCP) 353
dynamic RAM (DRAM) 229

E

ELK stack 175
email
attachments 257, 258
decoding 253
message format 253-257
Emailable
URL 322
email-based communications 8

Email Hippo
URL 322
email protocols 250
Internet Message Access Protocol (IMAP) 252
Post Office Protocol (POP3) 251, 252
Simple Mail Transfer Protocol (SMTP) 250, 251
employee misconduct 17-19
eMule 302-304
reference link 302
eMule MET Viewer
reference link 304
EnCase evidence file 93, 94
Eric Zimmerman's JumpList Explorer
reference link 210
Eric Zimmerman's tools
reference link 192, 208
ESEDatabaseView
reference link 201-203
ethical behavior
need for 380-383
event list 160
events classes
application 196
security 196
system 196
evidence 377-379
exploring 73-76
evidentiary hearing 372
Expert Witness Format (EWF) 93
extended boot record (EBR) 125

F

Facebook 295-297
Fake Caller ID
reference link 321
Fake Name Generator
URL 318
FavoritesView
reference link 293
FAW
URL 337
Federal Bureau of Investigation (FBI) 30
File Allocation Table (FAT) filesystem 132
boot record 133, 134
data area 136-138
deleted files, recovering 139-141
file allocation table 134-136
slack space 141
file attribute map 148
file knowledge
determining 200
file signature analysis 60-62
filesystems 131
File Transfer Protocol (FTP) 352
Firefox 287
bookmarks 293
cache 289, 290
cookies 290
history 290-292
passwords 292, 293
profiles 287-289
FireShot
URL 337

first responders 4
forensic boot media 116, 117
forensic copy 90
forensic evidence file 91
forensic examination environment 76, 77
forensic image 91
encoding schemes 178
forensic imaging 90, 91
DD image 91-93
EnCase evidence file 93, 94
SSD device 94
tools 95
forensic imaging, tools
FTK Imager 95-104
PALADIN 104-109
forensic investigator training 47, 48
certifications 47
forensic software 43-47
forensic workstation 38- 40
FTK Imager 95-104
using 236-238

G

globally unique identifier (GUIDs) 126
Google Chrome 270
bookmarks 270-274
cache 277
cookies 276
history file 274-276
passwords 278
Google Drive 307
Gophish tool 22-24
GPT partitions 125-130
grand jury 372
graphical user interface (GUI) 44, 164, 191, 315
Graphics Processing Unit (GPU) 39
GroupMe 296
Guerrilla Mail 317
URL 317
GUID Partition Table (GPT) 115

H

hard disk drive 119-121
geometry 121, 122
interfaces 120
hardware write blocker 88
reference link 88
hash analysis 57
have i been pwned?
URL 324
hibernation file (hiberfil.sys) 231
HIDDEN flag 138
history file, Google Chrome 274, 275
history, Firefox 290-292
history, Internet Explorer 279-282
hive files 191
HKEY Current User 189
Host Protected Area (HPA) 130, 131
HTTrack
URL 337
Hunchly
reference link 333
Hunter
reference link 322
hybrid cloud 306

Hypertext Transfer Protocol (HTTP) 353

Hypertext Transfer Protocol Secure (HTTPS) 353

I

IACIS Code of Ethics and Professional Conduct
reference link 380

illicit images 7, 8
email-based communications 8
newsgroups 9, 10
Peer-to-Peer (P2P) file-sharing 10, 11
USENET 9, 10

image_export 164-166

Image Fragmentation Size 101

Infrastructure as a Service (IaaS) 305

insider threat 25-27

Instagram 295

Integrated Drive Electronics (IDE/EIDE) 120

Internal Revenue Service (IRS) 30

International Association of Computer Investigative Specialists (IACIS) 380
Code of Ethics 380, 381

International Society of Forensic Computer Examiners (ISFCE) 382
Code of Ethics 382

Internet Control Message Protocol (ICMP) 354

Internet Explorer 278
bookmarks 279
cache 283-285
cookies 285-287
history 279-282
typed URL 282, 283

Internet Explorer Cache Viewer
reference link 283

internet layer protocols
Address Resolution Protocol (ARP) 355
Internet Control Message Protocol (ICMP) 354
Internet Protocol (IP) 354
Internet Protocol Security (IPSec) 355

Internet Message Access Protocol (IMAP) 252

Internet of Things (IoT) 15, 350

Internet Protocol (IP) 354

Internet Protocol Security (IPSec) 355

investigation findings
details, including 67, 68
document circumstances 68, 69
document facts 68, 69
report conclusion 70
reporting 66, 67

investigator 5

IP version 4 (IPv4) 348, 349
port numbers 350

IP version 6 (IPv6) 350, 351
addressing schemes 352

ISFCE Code of Ethics
reference link 382

J

judge 372

judicial/administrative proceeding
expert witness 374
witness 374

Jump List IDs
reference link 210

JumpLists 209-211
types 210

jury 373

K

Kik 296

Knowem 326

L

legal issues 48-50

link (LNK) files 208, 209

local user profile 188

log2timeline 166-169

Logical Block Addressing (LBA) 122

logical forensic image 91

logon types 199

long filename (LFN) 132, 138, 139

M

Macintosh-based tools 46

Magnet Forensics AXIOM forensic tool
reference link 301

Mail Summary files (MSF) 261

mandatory user profile 189

Master Boot Record (MBR) 115

Master Boot Record (MBR), partitions 122-125
extended partitions 125

media analysis 176, 177

Message Digest 5 (MD5) 57

Microsoft browsers
exploring 202-204

Microsoft Outlook/Outlook Express
exploring 258, 259

Microsoft technical documentation
URL 217

Microsoft Windows Event IDs
reference link 200

Microsoft Windows Live Mail
exploring 259, 260

Mint Mobile
URL 321

Modified, Accessed, and Created (MAC) 158

Most Recently Used (MRU)
determining 204-207

Mozilla Thunderbird 260

Multipurpose Internet Mail Extensions (MIME) 257

MZcacheview
reference link 290

MZHistoryView
reference link 290

N

National Institute of Standards and Technology (NIST) 45

National Security Agency (NSA) 32

National Software Reference Library (NSRL) 59

Network Address Translation (NAT) 349

network-attached storage (NAS) 101

network history
exploring 215, 216

network layer (Layer 3), OSI model 344

New Technology File System (NTFS) filesystem 141-153

non-sworn 6

Notepad++
URL 271

notetaking 359
fundamental elements 360
information, including 360, 361
medium, usage 361
NTUSER.DAT hive 204

O

onion network 316
online investigation 314
preserving 331-337
OpenSavePidlMRU 204
open-source forensic tools 45
autopsy 45
Computer-Aided Investigative Environment (CAINE) 46
PALADIN Forensic Suite 46
SIFT Workstation 46
open source intelligence techniques (OSINTs) 23
Open Source Interconnection (OSI) model 342
advantages 342
application layer (Layer 7) 345
data link layer (Layer 2) 343
encapsulation 345
network layer (Layer 3) 344
physical layer (Layer 1) 343
presentation layer (Layer 6) 345
session layer (Layer 5) 345
transport layer (Layer 4) 344
using 342
order of volatility 51

P

P2P file sharing 300, 301
pagefile (pagefile.sys) 232
PALADIN 104-109
partition identifiers
reference link 124
Password Fox
reference link 292
passwords, Firefox 292, 293
passwords, Google Chrome 278
Pastebin
URL 323
Peer-to-Peer (P2P) file-sharing 10, 11
People Search Now
URL 331
person-to-person (P2P) transactions 318
physical layer (Layer 1), OSI model 343
physical locations
identifying 214, 215
network history, exploring 215, 216
time zones, determining 215
WLAN event log, examining 217, 218
pinfo 169-171
ping utility 354
plaintiff 373
Plaso (Plaso Langar Að Safna Öllu) 164
download link 164
Platform as a Service (PaaS) 306
Portable Operating System Interface (POSIX) 126
port number 350
Posters & Cheat Sheets
reference link 193
Post Office Protocol (POP3) 251, 252
Power-On Self-Test (POST) 114
prefetch 213, 214

pre-investigation
considerations 38
pre-investigation, considerations
forensic investigator training 47, 48
forensic software 43-47
forensic workstation 38-40
response kit 40-43
premade filters
download link 168
preparation phase
initiating 374, 375
presentation layer (Layer 6), OSI model 345
private cloud 306
proceedings 372, 373
profiles, Firefox 287-289
program execution
exploring 218
Shimcache, exploring 219
UserAssist, determining 218, 219
prosecutor 373
protocols, TCP/IP
application layer protocols 352
internet layer protocols 354
transport layer protocols 353
ProtonMail
URL 317
PSBDMP
reference link 324
psort 171-174
psteal 174, 175
public cloud 306

R

radio frequency identification (RFID) 24
RAM analyzing tools
Bulk Extractor, using 238-243
exploring 238
VOLIX II, using 243-246
random access memory (RAM) 228-230
capturing 233
capturing device, preparing 233
fundamentals 227, 228
operating system 229
sources, identifying 231-233
random access memory (RAM), capture tools
DumpIt, using 234, 235
FTK Imager, using 236-238
random access memory (RAM), capturing device
capture tools, exploring 234
random access memory (RAM), operating system
privilege separation 229
process management 230
system calls 230
threads 230
Reacher
URL 322
READ ONLY flag 138
Read-Only Memory (ROM) 114, 229
real-world experience 24, 25
Recycle Bin 207, 208
Reddit 296
Reference Data Set (RDS) 59
regular expression
symbols, used for creating 179
relative identifier (RID) 195
Remote Desktop Protocol (RDP) 196, 353

report
acquisition details 365
administrative section 362
analysis details 365
evidence analyzed 364
executive summary section 362
exhibits/technical details 366, 367
narrative section 363, 364
template 362
writing 361
Request for Comments (RFC) 250
response kit 40, 43
equipment 40
response kit, equipment
antistatic bags 41
digital camera 40, 41
encryption 42
forensic laptop 42
frequency shielding material 42
latex or nitrile gloves 41
miscellaneous items 42
notepads 41
organizational paperwork 41
paper storage bags 41
software security keys 43
storage media 41
toolkit 42
write blocking devices 42
roaming user profile 189
Rufus
URL 117

S

SAM hive 191
San Bernardino terror attack 31, 32
Scientific Workgroup on Digital Evidence (SWGDE) 233
Secure Hashing Algorithm (SHA-1) 57
Secure Shell (SSH) 352
security 20
SECURITY hive 191
security identifier (SID) 195
Serial Advanced Technology Attachment (SATA) 120
Serial Attached SCSI (SAS) 120
service-level agreement (SLA) 307
service provider 299
Service Set Identifier (SSID)
URL 216
session layer (Layer 5), OSI model 345
Shareaza 304
reference link 304
shellbags 211
opening 211
working with 213
Shimcache
exploring 219
short filename (SFN) 137, 138
Silk Road 29-31
Simple Mail Transfer Protocol (SMTP) 250, 251, 353
SiteSucker
URL 337
slack space 141
Small Computer System Interface (SCSI) 120
SMART 100
reference link 46
Snapchat 295
social engineering 21
Gophish tool 22-24

social media 294
Software as a Service (SaaS) 305
SOFTWARE hive 191
software write blocker 89, 90
solid state drives (SSDs) 121
function 121
Solid-State Storage (SSD) device 94
Spokeo
URL 331
SpyCloud
URL 324
Standard_Information Attribute 147
static RAM (SRAM) 229
sterile media
creating 82-87
storage device
data types 176
string search 177-180
SUMURI
URL 39
Swapfile (Swapfile.sys) 232
sworn 6
SYSTEM hive 191

T

TCP/IP 346-348
IP version 4 (IPv4) 348-350
IP version 6 (IPv6) 350, 351
layers 346
protocols 352
Telnet 353
Temp Mail
URL 317
temporary user profile 189
testimony 377-379
guidelines 379
theft of intellectual property 32-34
This Person Does Not Exist
URL 320
threat actor 20
thumbcache 200
exploring 200-202
Thumbcache Viewer
reference link 200
Thunderbird 276
TikTok 296
timeline analysis 158, 159
Timeline Explorer 176
TimelineMaker Pro 175
TimeSketch 175
time zones
determining 215
Tinder 295
tool validation 77-82
Tor network 316
Transmission Control Protocol (TCP) 250, 350, 353
transport layer (Layer 4), OSI model 344
transport layer protocols
Transmission Control Protocol (TCP) 353
User Datagram Protocol (UDP) 354
trial 372
trim commands 95
True People Search
URL 328
Tumblr 296
Tutanota
URL 317

Twitter 295, 298, 299

typed URL, Internet Explorer 282, 283

Type, hive subkeys 191

U

undercover agent (UC) 75

undercover online investigation 315
- platform consideration 315, 316
- undercover online persona, creating 316-322

undercover online persona
- creating 316-322

Unified Extensible Firmware Interface (UEFI) 114

uniform resource identifier (URI) 74

United States Department of Defense (DoD) 346

universal time (UTC) 57

US-based proceedings
- arraignment 372
- deposition 372
- detention hearing 372
- evidentiary hearing 372
- grand jury 372
- trial 372

USB/attached devices
- usage, identifying on host 219-222
- usage, identify on host 220

user access control (UAC) 191

user accounts
- last login/last password, modifying 193-200
- usage, determining 193

UserAssist
- determining 218

User Datagram Protocol (UDP) 350, 354

user profiles 188
- AppData folder 189
- Local folder 190
- LocalLow folder 190
- Roaming folder 190

user profiles, types
- local user profile 188
- mandatory user profile 189
- roaming user profile 189
- temporary user profile 189

V

Value, hive subkeys 191

Verify Email
- URL 322

Virtual Hard Disk (VHD) 107

VirtualHere
- URL 43

virtual private networking (VPN) 316

VMware Virtual Disk Format (VMDK) 107

Volatility
- reference link 238

VOLIX II
- reference link 238
- using 243-246

Volume Boot Record (VBR) 124

W

Web2Disk
- URL 337

web-based email 253

WebMail
- analysis 262-265

well-known port numbers 350

WhatsApp 295

Whitepages
URL 331

WhoisXML API
reference link 322

Windows artifact
file knowledge, determining 200
JumpLists 209-211
link (LNK) files 208, 209
Microsoft browsers, exploring 202-204
most recently used/recently used, determining 204-207
physical locations, identifying 214, 215
prefetch 213, 214
program execution, exploring 218
Recycle Bin 207, 208
shellbags, opening 211
shellbags, working with 213
thumbcache, exploring 200-202
USB/attached devices usage, identifying 219-222
user accounts usage, determining 193

Windows-based users
commercial forensic tools 46

Windows Forensic Environment (WinFE) 116

Windows Registry 190-193

WinPrefetchViewtool
reference link 214

witnesses 373

WLAN event log
examining 217, 218

World Wide Web (WWW) 269

write blocking 87, 88

X

X1 Social Discovery
URL 337

X-Ways
Plaso (Plaso Langar Að Safna Öllu) 164

X-Ways Forensics 160-163

X-Ways, Plaso (Plaso Langar Að Safna Öllu)
image_export 164-166
log2timeline 166-169
pinfo 169-171
psort 171-174
psteal 174, 175

Z

ZabaSearch
URL 331