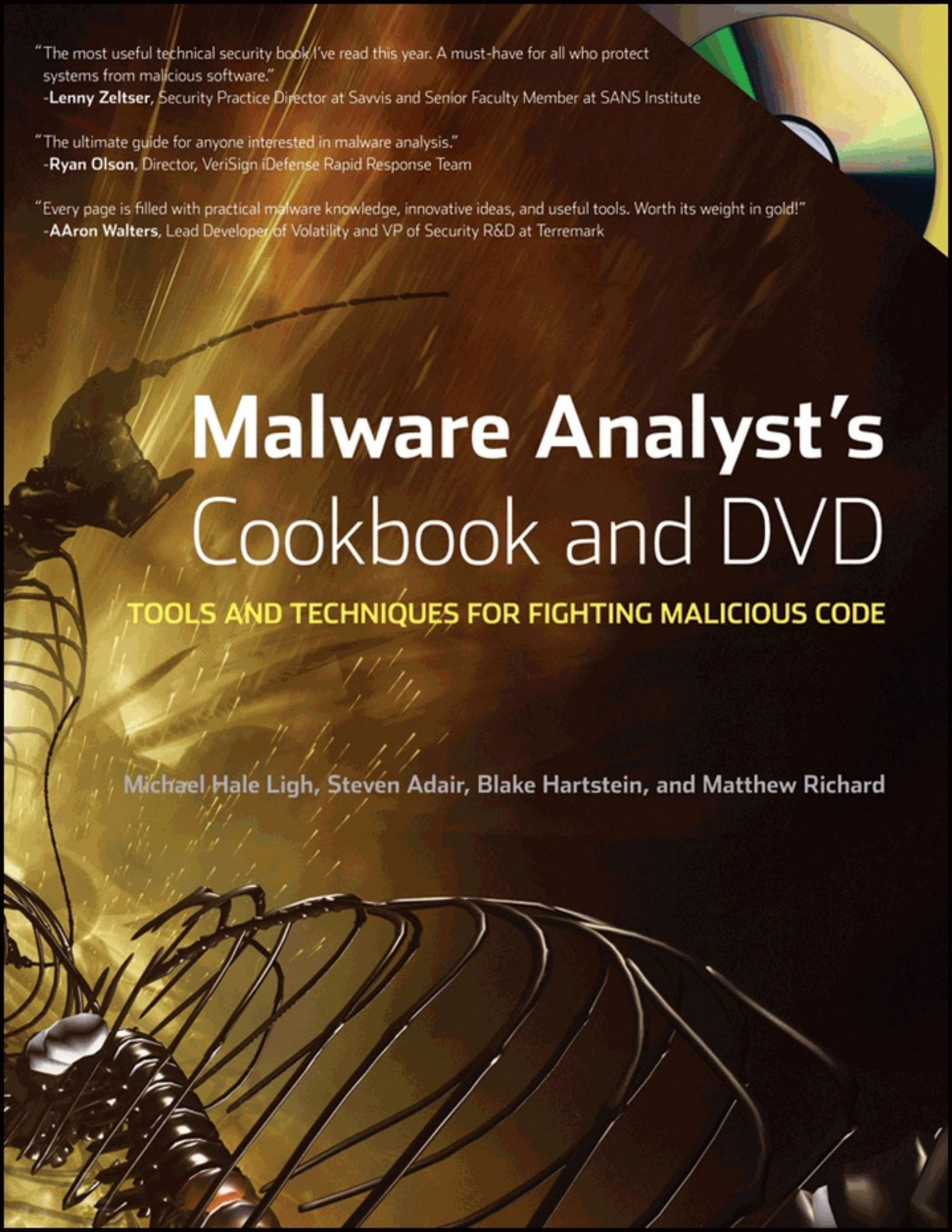
"The most useful technical security book I've read this year. A must-have for all who protect systems from malicious software."
-**Lenny Zeltser**, Security Practice Director at Savvis and Senior Faculty Member at SANS Institute

"The ultimate guide for anyone interested in malware analysis."
-**Ryan Olson**, Director, VeriSign iDefense Rapid Response Team

"Every page is filled with practical malware knowledge, innovative ideas, and useful tools. Worth its weight in gold!"
-**AAron Walters**, Lead Developer of Volatility and VP of Security R&D at Terremark



Malware Analyst's Cookbook and DVD

TOOLS AND TECHNIQUES FOR FIGHTING MALICIOUS CODE

Michael Hale Ligh, Steven Adair, Blake Hartstein, and Matthew Richard

Table of Contents

Wiley Publishing, Inc. End-User License Agreement

Malware Analyst's Cookbook and DVD

Tools and Techniques for Fighting Malicious Code

Michael Hale Ligh
Steven Adair
Blake Hartstein
Matthew Richard

Wiley Publishing, Inc.

Malware Analyst's Cookbook and DVD: Tools and Techniques for Fighting Malicious Code

Published by
Wiley Publishing, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Published simultaneously in Canada

ISBN: 978-0-470-61303-0

ISBN: 978-1-118-00336-7 (ebk)

ISBN: 978-1-118-00829-4 (ebk)

ISBN: 978-1-118-00830-0 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Control Number: 2010933462

To my family for helping me shape my life and to my wife Suzanne for always giving me something to look forward to.

—Michael Hale Ligh

To my new wife and love of my life Irene and my family. Without your support over the many years, I would not be where I am or who I am today.

—Steven Adair

About the Authors

Michael Hale Ligh is a Malicious Code Analyst at Verisign iDefense, where he specializes in developing tools to detect, decrypt, and investigate malware. In the past few years, he has taught malware analysis courses and trained hundreds of students in Rio De Janeiro, Shanghai, Kuala Lumpur, London, Washington D.C., and New York City. Before iDefense, Michael worked as a vulnerability researcher, providing ethical hacking services to one of the nation's largest healthcare providers. Due to this position, he gained a strong background in reverse-engineering and operating system internals. Before that, Michael defended networks and performed forensic investigations for financial institutions throughout New England. He is currently Chief of Special Projects at MNIN Security LLC.

Steven Adair is a security researcher with The Shadowserver Foundation and a Principal Architect at eTouch Federal Systems. At Shadowserver, Steven analyzes malware, tracks botnets, and investigates cyber-attacks of all kinds with an emphasis on those linked to cyber-espionage. Steven frequently presents on these topics at international conferences and co-authored the paper "Shadows in the Cloud: Investigating Cyber Espionage 2.0." In his day job, he leads the Cyber Threat operations for a Federal Agency, proactively detecting, mitigating and preventing cyber-intrusions. He has successfully

implemented enterprise-wide anti-malware solutions across global networks by marrying best practices with new and innovative techniques. Steven is knee deep in malware daily, whether it be supporting his company’s customer or spending his free time with Shadowserver.

Blake Hartstein is a Rapid Response Engineer at Verisign iDefense. He is responsible for analyzing and reporting on suspicious activity and malware. He is the author of the Jsunpack tool that aims to automatically analyze and detect web-based exploits, which he presented at Shmoocon 2009 and 2010. Blake has also authored and contributed Snort rules to the Emerging Threats project.

Matthew Richard is Malicious Code Operations Lead at Raytheon Corporation, where he is responsible for analyzing and reporting on malicious code. Matthew was previously Director of Rapid Response at iDefense. For 7 years before that, Matthew created and ran a managed security service used by 130 banks and credit unions. In addition, he has done independent forensic consulting for a number of national and global companies. Matthew currently holds the CISSP, GCIA, GCFA, and GREM certifications.

Acknowledgments

Michael would like to thank his current and past employers for providing an environment that encourages and stimulates creativity. He would like to thank his coworkers and everyone who has shared knowledge in the past. In particular, AAron Walters and Ryan Smith for never hesitating to engage and debate interesting new ideas and techniques. A special thanks goes out to the guys who took time out of the busy days to review our book: Lenny Zeltser, Tyler Hudak, and Ryan Olson.

Steven would like to extend his gratitude to those who spend countless hours behind the scenes investigating malware and fighting cyber-crime. He would also like to thank his fellow members of the Shadowserver Foundation for their hard work and dedication towards making the Internet a safer place for us all.

We would also like to thank the following:

- Maureen Spears and Carol A. Long from Wiley Publishing, for helping us get through our first book.
- Ilfak Guilfanov (and the team at Hex-Rays) and Halvar Flake (and the team at Zynamics) for allowing us to use some of their really neat tools.
- All the developers of the tools that we referenced throughout the book. In particular, Frank Boldewin, Mario Vilas, Harlan Carvey, and Jesse

Kornblum, who also helped review some recipes in their realm of expertise.
- The authors of other books, blogs, and websites that contribute to the collective knowledge of the community.

—Michael, Steven, Blake, and Matthew

Introduction

Malware Analyst's Cookbook is a collection of solutions and tutorials designed to enhance the skill set and analytical capabilities of anyone who works with, or against, malware. Whether you're performing a forensic investigation, responding to an incident, or reverse-engineering malware for fun or as a profession, this book teaches you creative ways to accomplish your goals. The material for this book was designed with several objectives in mind. The first is that we wanted to convey our many years of experience in dealing with malicious code in a manner friendly enough for non-technical readers to understand, but complex enough so that technical readers won't fall asleep. That being said, malware analysis requires a well-balanced combination of many different skills. We expect that our readers have at least a general familiarity with the following topics:

- Networking and TCP/IP
- Operating system internals (Windows and Unix)
- Computer security
- Forensics and incident response
- Programming (C, C++, Python, and Perl)
- Reverse-engineering
- Vulnerability research
- Malware basics

Our second objective is to teach you how various tools work, rather than just how to use the tools. If you understand what goes on when you click a button (or type a command) as opposed to just knowing which button to click, you'll be better equipped to perform an analysis on the tool's output instead of just collecting the output. We realize that not everyone can or wants to program, so we've included over 50 tools on the DVD that accompanies the book; and we discuss hundreds of others throughout the text. One thing we tried to avoid is providing links to every tool under the sun. We limit our discussions to tools that we're familiar with, and —as much as possible—tools that are freely available.

Lastly, this book is *not* a comprehensive guide to all tasks you should perform during examination of a malware sample or during a forensic investigation. We tried to include solutions to problems that are common enough to be most beneficial to you, but rare enough to not be covered in other books or websites. Furthermore, although malware can target many platforms such as Windows, Linux, Mac OS X, mobile devices, and hardware/firmware components, our book focuses primarily on analyzing Windows malware.

Who Should Read This Book

If you want to learn about malware, you should read this book. We expect our readers to be forensic investigators, incident responders, system

administrators, security engineers, penetration testers, malware analysts (of course), vulnerability researchers, and anyone looking to be more involved in security. If you find yourself in any of the following situations, then you are within our target audience:

- You're a member of your organization's incident handling, incident response, or forensics team and want to learn some new tools and techniques for dealing with malware.
- You work as a systems, security, or network administrator and want to understand how you can protect end users more effectively.
- You're a member of your country's Computer Emergency Response Team (CERT) and need to identify and investigate malware intrusions.
- You work at an antivirus or research company and need practical examples of analyzing and reporting on modern malware.
- You're an aspiring student hoping to learn techniques that colleges and universities just don't teach.
- You work in the IT field and have recently become bored, so you're looking for a new specialty to compliment your technical knowledge.

How This Book Is Organized

This book is organized as a set of recipes that solve specific problems, present new tools, or discuss how to detect and analyze malware in interesting ways. Some of the recipes are standalone, meaning the

problem, discussion, and solution are presented in the same recipe. Other recipes flow together and describe a sequence of actions that you can use to solve a larger problem. The book covers a large array of topics and becomes continually more advanced and specialized as it goes on. Here is a preview of what you can find in each chapter:

- Chapter 1, Anonymizing Your Activities: Describes how you conduct online investigations without exposing your own identity. You'll use this knowledge to stay safe when following along with exercises in the book and when conducting research in the future.
- Chapter 2, Honeypots: Describes how you can use honeypots to collect the malware being distributed by bots and worms. Using these techniques, you can grab new variants of malware families from the wild, share them in real time with other researchers, analyze attack patterns, or build a workflow to automatically analyze the samples.
- Chapter 3, Malware Classification: Shows you how to identify, classify, and organize malware. You'll learn how to detect malicious files using custom antivirus signatures, determine the relationship between samples, and figure out exactly what functionality attackers may have introduced into a new variant.
- Chapter 4, Sandboxes and Multi-AV Scanners: Describes how you can leverage online virus scanners and public sandboxes. You'll learn how

to use scripts to control the behavior of your sample in the target sandbox, how to submit samples on command line with Python scripts, how to store results to a database, and how to scan for malicious artifacts based on sandbox results.

- Chapter 5, Researching Domains and IP Addresses: Shows you how to identify and correlate information regarding domains, hostnames, and IP addresses. You'll learn how to track fast flux domains, determine the alleged owner of a domain, locate other systems owned by the same group of attackers, and create static or interactive maps based on the geographical location of IP addresses.
- Chapter 6, Documents, Shellcode, and URLs: In this chapter, you'll learn to analyze JavaScript, PDFs, Office documents, and packet captures for signs of malicious activity. We discuss how to extract shellcode from exploits and analyze it within a debugger or in an emulated environment.
- Chapter 7, Malware Labs: Shows how to build a safe, flexible, and inexpensive lab in which to execute and monitor malicious code. We discuss solutions involving virtual or physical machines and using real or simulated Internet.
- Chapter 8, Automation: Describes how you can automate the execution of malware in VMware or VirtualBox virtual machines. The chapter introduces several Python scripts to create custom reports about the malware's behavior,

including network traffic logs and artifacts created in physical memory.

- Chapter 9, Dynamic Analysis: One of the best ways to understand malware behavior is to execute it and watch what it does. In this chapter, we cover how to build your own API monitor, how to prevent certain evidence from being destroyed, how to log file system and Registry activity in real time *without* using hooks, how to compare changes to a process's handle table, and how to log commands that attackers send through backdoors.
- Chapter 10, Malware Forensics: Focuses on ways to detect rootkits and stealth malware using forensic tools. We show you how to scan the file system and Registry for hidden data, how to bypass locked file restrictions and remove stubborn malware, how to detect HTML injection and how to investigate a new form of Registry "slack" space.
- Chapter 11, Debugging Malware: Shows how you can use a debugger to analyze, control, and manipulate a malware sample's behaviors. You'll learn how to script debugging sessions with Python and how to create debugger plug-ins that monitor API calls, output HTML behavior reports, and automatically highlight suspicious activity.
- Chapter 12, De-obfuscation: Describes how you can decode, decrypt, and unpack data that attackers intentionally try to hide from you. We walk you through the process of reverse-

engineering a malware sample that encrypts its network traffic so you can recover stolen data. In this chapter, you also learn techniques to crack domain generation algorithms.

- Chapter 13, Working with DLLs: Describes how to analyze malware distributed as Dynamic Link Libraries (DLLs). You'll learn how to enumerate and examine a DLL's exported functions, how to run the DLL in a process of your choice (and bypass host process restrictions), how to execute DLLs as a Windows service, and how to convert DLLs to standalone executables.
- Chapter 14, Kernel Debugging: Some of the most malicious malware operates only in kernel mode. This chapter covers how to debug the kernel of a virtual machine infected with malware to understand its low-level functionality. You learn how to create scripts for WinDbg, unpack kernel drivers, and to leverage IDA Pro's debugger plug-ins.
- Chapter 15, Memory Forensics with Volatility: Shows how to acquire memory samples from physical and virtual machines, how to install the Volatility advanced memory forensics platform and associated plug-ins, and how to begin your analysis by detecting process context tricks and DKOM attacks.
- Chapter 16, Memory Forensics: Code Injection and Extraction: Describes how you can detect and extract code (unlinked DLLs, shellcode, and so on) hiding in process memory. You'll learn to rebuild

binaries, including user mode programs and kernel drivers, from memory samples and how to rebuild the import address tables (IAT) of packed malware based on information in the memory dump.

- Chapter 17, Memory Forensics: Rootkits: Describes how to detect various forms of rootkit activity, including the presence of IAT, EAT, Inline, driver IRP, IDT, and SSDT hooks on a system. You'll learn how to identify malware that hides in kernel memory without a loaded driver, how to locate system-wide notification routines, and how to detect attempts to hide running Windows services.
- Chapter 18, Network and Registry: Shows how to explore the artifacts created on a system due to a malware sample's network activity. You'll learn to detect active connections, listening sockets, and the use of raw sockets and promiscuous mode network cards. This chapter also covers how to extract volatile Registry keys and values from memory.

Setting Up Your Environment

We performed most of the development and testing of Windows tools on 32-bit Windows XP and Windows 7 machines using Microsoft's Visual Studio and Windows Driver Kit. If you need to recompile our tools for any reason (for example to fix a bug), or if you're interested in building your own tools based on source

code that we've provided, then you can download the development environments here:

- The Windows Driver Kit: http://www.microsoft.com/whdc/devtools/WDK/default.mspx
- Visual Studio C++ Express: http://www.microsoft.com/express/Downloads/#2010-Visual-CPP

As for the Python tools, we developed and tested them on Linux (mainly Ubuntu 9.04, 9.10, or 10.04) and Mac OS X 10.4 and 10.5. You'll find that a majority of the Python tools are multi-platform and run wherever Python runs. If you need to install Python, you can get it from the website at http://python.org/download/. We recommend using Python version 2.6 or greater (but not 3.x), because it will be most compatible with the tools on the book's DVD.

Throughout the book, when we discuss how to install various tools on Linux, we assume you're using Ubuntu. As long as you know your way around a Linux system, you're comfortable compiling packages from source, and you know how to solve basic dependency issues, then you shouldn't have a problem using any other Linux distribution. We chose Ubuntu because a majority of the tools (or libraries on which the tools depend) that we reference in the book are either preinstalled, available through the apt-get package manager, or the developers of the tools specifically say that their tools work on Ubuntu.

You have a few options for getting access to an Ubuntu machine:

- Download Ubuntu directly: http://www.ubuntu.com/desktop/get-ubuntu/download
- Download Lenny Zeltser's REMnux: http://REMnux.org. REMnux is an Ubuntu system preconfigured with various open source malware analysis tools. REMnux is available as a VMware appliance or ISO image.
- Download Rob Lee's SANS SIFT Workstation: https://computer-forensics2.sans.org/community/siftkit/. SIFT is an Ubuntu system preconfigured with various forensic tools. SIFT is available as a VMware appliance or ISO image.

We always try to provide a URL to the tools we mention in a recipe. However, we use some tools significantly more than others, thus they appear in five to ten recipes. Instead of linking to each tool each time, here is a list of the tools that you should have access to throughout all chapters:

- Sysinternals Suite: http://technet.microsoft.com/en-us/sysinternals/bb842062.aspx
- Wireshark: http://www.wireshark.org/
- IDA Pro and Hex-Rays: http://www.hex-rays.com/idapro/
- Volatility: http://code.google.com/p/volatility/
- WinDbg Debugger: http://www.microsoft.com/whdc/devtools/debuggi

ng/default.mspx

- YARA: http://code.google.com/p/yara-project/
- Process Hacker: http://processhacker.sourceforge.net/

You should note a few final things before you begin working with the material in the book. Many of the tools require administrative privileges to install *and* execute. Typically, mixing malicious code and administrative privileges isn't a good idea, so you must be sure to properly secure your environment (see Chapter 7 for setting up a virtual machine if you do not already have one). You must also be aware of any laws that may prohibit you from collecting, analyzing, sharing, or reporting on malicious code. Just because we discuss a technique in the book does not mean it's legal in the city or country in which you reside.

Conventions

To help you get the most from the text and keep track of what's happening, we've used a number of conventions throughout the book.

Recipe X-X: Recipe Title

Boxes like this contain recipes, which solve specific problems, present new tools, or discuss how to detect and analyze malware in interesting ways. Recipes may contain helpful steps, supporting figures, and notes from the authors. They also may have supporting materials associated with them on the companion

DVD. If they do have supporting DVD materials, you will see a DVD icon and descriptive text, as follows:

You can find supporting material for this recipe on the companion DVD.

For your further reading and research, recipes may also have endnotes[1] that site Internet or other supporting sources. You will find endnote references at the end of the recipe. Endnotes are numbered sequentially throughout a chapter.

[1] This is an endnote. This is the format for a website source

Note Tips, hints, tricks, and asides to the current discussion look like this.

As for other conventions in the text:

- New terms and important words appear in *italics* when first introduced.
- Keyboard combinations are treated like this: Ctrl+R.
- File names are in parafont, (filename.txt), URLs and code (API functions and variable names) within the text are treated like so: `www.site.org`, `LoadLibrary`, `var1`.
- This book uses monofont type with no highlighting for most code examples. Code fragments may be broken into multiple lines or truncated to fit on the page:

```
This is an example of monofont type with a long \
      line of code that needed to be broken.
This truncated line shows how [REMOVED]
```

- This book uses bolding to emphasize code. User input for commands and code that is of particular importance appears in bold:

```
$ date ; typing into a Unix shell
Wed Sep  1 14:30:20 EDT 2010
C:\> date ; typing into a Windows shell
Wed 09/01/2010
```

On The Book's DVD

The book's DVD contains evidence files, videos, source code, and programs that you can use to follow along with recipes or to conduct your own investigations and analysis. It also contains the full-size, original images and figures that you can view, since they appear in black and white in the book. The files are organized on the DVD in folders named according to the chapter and recipe number. Most of the tools on the DVD are written in C, Python, or Perl and carry a GPLv2 or GPLv3 license. You can use a majority of them as-is, but a few may require small modifications depending on your system's configuration. Thus, even if you're not a programmer, you should take a look at the top of the source file to see if there are any notes regarding dependencies, the platforms on which we tested the tools, and any variables that you may need to change according to your environment.

We do not guarantee that all programs are bug free (who does?), thus, we welcome feature requests and bug reports addressed to malwarecookbook@gmail.com. If we do provide updates for the code in the future, you can always find the most recent versions at http://www.malwarecookbook.com.

The following table shows a summary of the tools that you can find on the DVD, including the