Using Digital Evidence to Solve Computer Crime

Internet Forensics

easy computing

Robert Jones
Internet Forensics

The Internet is riddled with spammers, con artists, and identity thieves. Everywhere you turn, there are websites selling fake Viagra, touting get-rich-quick schemes, or trying to trick you out of your credit card number. More than a nuisance, these are real crimes targeting the vulnerable members of society as their victims.

Conventional wisdom says that you can never track down the people behind the scams, since the Internet is so large and so unregulated that it is easy for someone to hide his or her identity. But that’s not true. In every spam message, phishing email, or web page, there are all sorts of clues that reveal something about the author.

*Internet Forensics* shows you how to find the clues left behind at an Internet crime scene. The IP address of a server or the layout of files on a web site are the online equivalent of a fingerprint on a door handle or a tire track in the mud. Just like the forensics dramas we watch on TV, the unlikeliest piece of evidence can turn out to be the most important.

*Internet Forensics* is a practical guide to this fascinating and topical field. Targeted at programmers, systems administrators, and power users, this book shows the reader how to extract clues from a range of sources. It describes how the bad guys try to cover their tracks, and the tricks we can use to see through their disguises. This book is packed with real-world examples, explained in detail, that show how much you can find out with ingenuity and a little work.

An informed and motivated developer community has led to tremendous improvements in firewall technology, with practical benefits to millions of users. That same type of community is emerging to fight back against spam and Internet fraud. This book and its readers can help catalyze these efforts and help reclaim the Internet.
Internet Forensics

Robert Jones
# Table of Contents

**Preface** .................................................................................................................. vii  

1. **Introduction** ....................................................................................................... 1  
   What Is Internet Forensics? ................................................................. 1  
   The Seamy Underbelly of the Internet .................................................. 2  
   Pulling Back the Curtain ......................................................................... 4  
   Taking Back Our Internet ....................................................................... 5  
   Protecting Your Privacy ........................................................................... 6  
   Before You Begin ....................................................................................... 6  
   A Network Neighborhood Watch .......................................................... 8  

2. **Names and Numbers** .......................................................................................... 10  
   Addresses on the Internet ....................................................................... 10  
   Internet Address Tools .......................................................................... 14  
   DNS Record Manipulation ..................................................................... 28  
   An Example—Dissecting a Spam Network ........................................... 29  

3. **Email** ................................................................................................................... 34  
   Message Headers ....................................................................................... 34  
   Forged Headers ......................................................................................... 38  
   Forging Your Own Headers ...................................................................... 39  
   Tracking the Spammer ............................................................................. 41  
   Viruses, Worms, and Spam .................................................................... 42  
   Message Attachments ................................................................................. 44  
   Message Content ...................................................................................... 49  
   Is It Really Spam? ...................................................................................... 50
# Table of Contents

4. **Obfuscation** .......................... 52  
- Anatomy of a URL .......................... 53  
- IP Addresses in URLs ......................... 56  
- Usernames in URLs ......................... 57  
- Encoding the Entire Message ................. 59  
- Similar Domain Names ....................... 62  
- Making a Form Look Like a URL ............... 63  
- Bait and Switch—URLRedirection ............. 63  
- JavaScript ................................ 67  
- Browsers and Obfuscation ..................... 68  

5. **Web Sites** ............................. 69  
- Capturing Web Pages ......................... 69  
- Viewing HTML Source ....................... 72  
- Comparing Pages ............................ 75  
- Non-Interactive Downloads Using wget ......... 76  
- Mapping Out the Entire Web Site .............. 82  
- Hidden Directories ......................... 84  
- In-Depth Example—Directory Listings .......... 87  
- Dynamic Web Pages ......................... 89  
- Filling Out Forms ........................... 90  
- In-Depth Example—Server-Side Database ....... 93  
- Opening the Black Box ...................... 97  

6. **Web Servers** .......................... 102  
- Viewing HTTP Headers ....................... 102  
- What Can Headers Tell Us? .................... 104  
- Cookies .................................. 107  
- Redirection ................................. 110  
- Web Server Statistics ....................... 113  
- Controlling HTTP Headers ................... 115  
- A Little Bit of Everything .................... 115  

7. **Web Browsers** .......................... 118  
- What Your Browser Reveals ................... 118  
- Apache Web Server Logging .................. 121  
- Server Log Analysis ........................ 125  
- Protecting Your Privacy ...................... 129
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. File Contents</td>
<td>136</td>
</tr>
<tr>
<td>Word Document Metadata</td>
<td>136</td>
</tr>
<tr>
<td>U.K. Government Dossier on Iraq</td>
<td>140</td>
</tr>
<tr>
<td>Document Forgery</td>
<td>149</td>
</tr>
<tr>
<td>Redaction of Sensitive Information</td>
<td>150</td>
</tr>
<tr>
<td>9. People and Places</td>
<td>156</td>
</tr>
<tr>
<td>Geographic Location</td>
<td>156</td>
</tr>
<tr>
<td>Time Zone</td>
<td>158</td>
</tr>
<tr>
<td>Language</td>
<td>160</td>
</tr>
<tr>
<td>Expertise</td>
<td>163</td>
</tr>
<tr>
<td>Criminal or Victim?</td>
<td>164</td>
</tr>
<tr>
<td>Hardware and Software</td>
<td>164</td>
</tr>
<tr>
<td>10. Patterns of Activity</td>
<td>166</td>
</tr>
<tr>
<td>Signatures</td>
<td>166</td>
</tr>
<tr>
<td>Searching with Signatures</td>
<td>167</td>
</tr>
<tr>
<td>Problems with Simple Signatures</td>
<td>169</td>
</tr>
<tr>
<td>Full Text Comparison</td>
<td>170</td>
</tr>
<tr>
<td>Using Internet Search Engines for Patterns</td>
<td>175</td>
</tr>
<tr>
<td>11. Case Studies</td>
<td>177</td>
</tr>
<tr>
<td>Case Study 1: Tidball</td>
<td>177</td>
</tr>
<tr>
<td>Case Study 2: Spam Networks</td>
<td>192</td>
</tr>
<tr>
<td>12. Taking Action</td>
<td>202</td>
</tr>
<tr>
<td>What Is Being Done to Tackle Internet Fraud?</td>
<td>202</td>
</tr>
<tr>
<td>What You Can Do to Help</td>
<td>207</td>
</tr>
<tr>
<td>Getting in Over Your Head</td>
<td>211</td>
</tr>
<tr>
<td>Vision of a Community Response</td>
<td>214</td>
</tr>
<tr>
<td>Index</td>
<td>217</td>
</tr>
</tbody>
</table>
Today’s Internet is riddled with spammers, con artists, and identity thieves. Everywhere you turn are web sites selling fake Viagra, touting get-rich-quick schemes, or trying to trick you out of your credit card number.

You and I may see through all the scams, but plenty of people do not. More than a nuisance, these are real crimes that target the vulnerable members of society such as the elderly and the naïve—people like your parents and grandparents.

Conventional wisdom says that you can never track down the people behind the scams, and that the Internet is so large and so unregulated that it is easy for someone to hide their identity. But that’s not true. In every spam message, phishing email, or web page, there are all sorts of clues that reveal something about the author. The Internet address of a server and the layout of files on a web site are the online equivalents of a fingerprint on a door handle or a tire track in the mud.

None of these details, by themselves, tell you a great deal. But when viewed as a whole, and, especially, when compared between cases, clear patterns and connections become apparent. As in real criminal investigations, the unlikeliest piece of evidence can turn out to be the most important.

Internet Forensics shows you how to find the clues left behind at an Internet crime scene. You will learn how to uncover information that lies hidden in every email message, web page, and web server on the Internet. You will gain an understanding of how the Internet functions—what really goes on when you request a web page, for example. You will see how the bad guys take advantage of these protocols and the lengths that they go to in order to hide their tracks.

My own interest in this field has been motivated by several factors. First is the daily frustration of dealing with spam, viruses, and all sorts of scams. With it is the growing unease that our Internet is being taken away from us by these abusers and that, unless we band together and do something about it, the problem is going to get much worse. Collectively, it is getting more difficult for them to operate in secrecy, we can push back against the bad guys and take back the network.
In looking into this sort of scam, you are forced to learn more about the way the Internet and its core protocols function. You see where their shortcomings lie and you start to think of ways they could be made better. It is a great way to learn a lot about Internet technologies without having to become an expert in the details of any one of them.

Last, but not least, is the fun to be had from playing amateur detective and solving Internet mysteries. At every stage of the game you are challenged to uncover information hidden in email message headers or web transactions. What appear to be minor details can become significant when combined with clues revealed by another technique. A passing observation in one study may link it into a much larger network of scams.

Murder mysteries and forensics crime dramas in books and on television are popular for a good reason. People like the challenge of finding clues, putting them together, and solving the puzzle. I think this is particularly true among those of us in the software development community. Alongside the more noble motivations, I hope that you will enjoy the challenge of Internet forensics in its own right.

**Who This Book Is For**

I have written this book with two types of reader in mind. The first are those of you with a professional interest in computer security. The traditional focus of this field has been on preventing attacks on private machines and networks from people and viruses, and using computer forensics to reveal their activities.

But today’s threats require that we go beyond this localized, internal focus and look outward to the Internet. Some of the viruses we see are used to set up email relay servers that are used by international spam operations. Computers are attacked, hijacked, and used to host fake bank web sites that are used for identity theft. Those of you who are computer security professionals will learn the core techniques you need in order to address this evolving type of threat.

The larger, less defined, audience consists of software developers and systems administrators who take a broad interest in the Internet and how it works. Many of us feel a deep frustration with the epidemics of spam, phishing, and viruses and want to fight back against it in some way. The book will show you how the bad guys are able to abuse the technology of the Internet. It will show you how to uncover information about their operations and, in doing so, make their lives a lot more difficult.

To get the most out of this book, it helps to have a working knowledge of Unix and some experience with Perl. But that is not an absolute requirement. If you can use a web browser and an email client, then you can follow along with most of the material in the book. You already have the most important thing you need: an inquisitive mind.
I have tried to avoid complex software as far as possible. Most of the tools come standard with recent versions of Linux and those that don’t are easily downloaded and installed. I have included Perl scripts throughout the book where these can help automate common tasks or help display information more concisely, and I have made a conscious effort to keep these scripts short and simple. My hope is that they are easy to understand and can serve as starting points for your own scripts.

Contents of This Book

The book is organized around the core technologies of the Internet—email, web sites, servers, and browsers. Chapters describe how these are used and abused and show you how information hidden in each of them can be uncovered. Short examples illustrate all the major techniques that are discussed. Two in-depth examples in Chapter 11 show how they are used in concert in real-world case studies.

Dealing with Internet fraud and abuse is not merely a technical matter and so throughout the book, I discuss the ethical and legal issues that arise in this field of work.

The contents of specific chapters are as follows:

Chapter 1, Introduction
An overview of spam, phishing, and the other threats to today’s Internet.

Chapter 2, Names and Numbers
Tools and techniques to retrieve information about Internet addresses and domain names.

Chapter 3, Email
The structure of email messages, how spammers forge message headers, and what you can uncover in spite of their efforts to hide.

Chapter 4, Obfuscation
A review of the many ways that con artists conceal their identities and how you can see through their disguises.

Chapter 5, Web Sites
Dissecting the operation of Internet scams by studying the pages and directories that make up a web site.

Chapter 6, Web Servers
Ways to uncover information about web servers and their operation by looking at the headers records of standard web transactions.

Chapter 7, Web Browsers
What you reveal about yourself every time you visit a web site and some ways in which you can protect your personal information.
Chapter 8, *File Contents*
Techniques to extract information that lies hidden within PDF and Word documents.

Chapter 9, *People and Places*
A collection of miscellaneous techniques.

Chapter 10, *Patterns of Activity*
Ways to search for similar features across multiple files, allowing you to link together different Internet scams.

Chapter 11, *Case Studies*
Two in-depth examples of Internet forensics at work. The first is a detailed study of a phishing scam. The second is a broader analysis of spam networks.

Chapter 12, *Taking Action*
A discussion on combating Internet fraud and how you can play a part.

**Conventions Used in This Book**

The following typographical conventions are used in this book:

Plain text
 Indicates menu titles, menu options, menu buttons, and keyboard accelerators (such as Alt and Ctrl).

*Italic*
 Indicates new terms, URLs, email addresses, filenames, file extensions, pathnames, directories, and Unix utilities.

Constant width
 Indicates commands, options, switches, variables, attributes, keys, functions, types, classes, namespaces, methods, modules, properties, parameters, values, objects, events, event handlers, XML tags, HTML tags, macros, the contents of files, or the output from commands.

**Constant width bold**
 Shows commands or other text that should be typed literally by the user.

**Constant width italic**
 Shows text that should be replaced with user-supplied values.

This icon signifies a tip, suggestion, or general note.

This icon indicates a warning or caution.
Safari Enabled

When you see a Safari® Enabled icon on the cover of your favorite technology book, that means the book is available online through the O’Reilly Network Safari Bookshelf.

Safari offers a solution that’s better than e-books. It’s a virtual library that lets you easily search thousands of top technical books, cut and paste code samples, download chapters, and find quick answers when you need the most accurate, current information. Try it for free at http://safari.oreilly.com.

Using Code Examples

This book is here to help you get your job done. In general, you may use the code in this book in your programs and documentation. You do not need to contact us for permission unless you’re reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O’Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product’s documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: “Internet Forensics by Robert Jones. Copyright 2005 O’Reilly Media, Inc., 0-596-10006-X.”

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at permissions@oreilly.com.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O’Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
(800) 998-9938 (in the United States or Canada)
(707) 829-0515 (international or local)
(707) 829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at:

http://www.oreilly.com/catalog/internetforensics
To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the
O'Reilly Network, see our web site at:

http://www.oreilly.com

Acknowledgments

I would like to offer my sincere thanks to Tatiana Apandi, my editor at O'Reilly, for
all her help in guiding this novice author through the process of creating this book.

Allison Randal helped in the editing process, with Matt Messier, Tom Cross, and
Simon Biles providing technical review. Their efforts were of great help to me in
refining the manuscript.

Writing the book has made me appreciate the amount of work that goes on once the
manuscript is complete. I want to thank the production team at O'Reilly for all their
work on this project. I also want to thank the people that have played an unseen role
in creating the volume that you hold in your hands. None of this would happen
without the accountants, managers, printing press operators, and fork lift drivers.

I came to computing by way of molecular biology. Over the years, I have had the
privilege to work with, and learn from, some superb programmers and sysadmins. I
owe them all a debt of gratitude, especially the crews at Thinking Machines and
Darwin Molecular.

Very special thanks are due to my darling Nancy, my wife, for all her encouragement
over the past few months. Writing can be a difficult and time-consuming process.
Her understanding and support were essential in making this book happen.
CHAPTER 1

Introduction

What Is Internet Forensics?

Forensics is the application of scientific methods in criminal investigations. It is a unique field of study that draws from all areas of science, from entomology to genetics, from geology to mathematics, with the single goal of solving a mystery. It holds a great fascination for the general public. Thanks to television dramas, millions of us are familiar with how rifling marks on a bullet can identify a murder weapon and how luminol is used to reveal bloodstains in the bath.

Computer forensics studies how computers are involved in the commission of crimes. In cases ranging from accounting fraud, to blackmail, identity theft, and child pornography, the contents of a hard drive can contain critical evidence of a crime. The analysis of disks and the tracking of emails between individuals have become commonplace tools for law enforcement around the world.

Internet forensics shifts that focus from an individual machine to the Internet at large. With a single massive network that spans the globe, the challenge of identifying criminal activity and the people behind it becomes immense. A con artist in the United States can use a web server in Korea to steal the credit card number of a victim in Germany.

Unfortunately, the underlying protocols that handle Internet traffic were not designed to address the problems of spam, viruses, and so forth. It can be difficult, often impossible, to verify the source of a message or the operator of a web site. In cases like this the minor details become important. The layout of files on a web site or the way that email headers are forged can play the same role as a fingerprint at a physical crime scene.

This book shows you some of the ways in which the bad guys try to conceal their identities. I show you how simple techniques, a knowledge of how the Internet works, and an inquisitive mind can reveal far more about these people than they would like.
The Seamy Underbelly of the Internet

History shows us that any situation that involves people and money will quickly attract crime. That has certainly been the case with the Internet. Online crime is at an all-time high and shows no signs of slowing down, despite the best efforts of the computer security industry.

The Scams

Many forms of criminal activity use the Internet as a means of communication, either using email instead of phone calls or publishing offensive material on a web site instead of hard copy. But the Internet has allowed some types of crime to evolve in new ways so as to exploit the new opportunities that it provides.

Spam is the most widespread of these activities. Unsolicited email places a burden on millions of servers every day. Companies spend huge amounts of money on software and staff to help keep the problem under control. They do so to save their employees from having to deal with all of it on their desktops, which would incur even higher costs in the form of lower productivity.

People who are computer savvy tend to focus on the nuisance factor of spam because that is what directly affects us. We tend to overlook the content of those messages because we already know them to be scams. We would never dream of clicking on URLs for web sites that promise us cheap Viagra, great rates on mortgages, or the chance to meet lonely singles in our neighborhoods. But other people do! If they didn’t, then the people running the web sites would not waste their money hiring the spammers to distribute their emails.

Most of these are traditional scams that have been updated to entice Internet-savvy victims. Their goal is to get you to hand over your credit card number. Being able to reach millions of potential victims through the power of spam is what makes it so attractive.

Phishing is the name we give to frauds involving fake web sites that look like those of banks or credit card companies. A phishing email is sent out like most other spam, but it attempts to entice victims by appearing to come from a well-known, legitimate business like Citibank or eBay. The message asks you to click on a URL that takes you to a web site. That web page, at first glance, looks just like the site of the genuine financial institution. The users are prompted to enter their online account information along with other personal details like their date of birth, credit card information, and so forth.

Computer viruses and worms were initially regarded as the malevolent creations of people who wanted to show off their programming skills and wanted to “get in the face” of computer users around the world. The immediate damage they caused ranged from negligible to minor. They were comparable to a graffiti tag spray painted...
on a wall. Their real impact lay in the effort it took to deal with infected computers and in preventing future attacks. But these threats have become more serious over time. Today’s viruses will actively disrupt the function of antivirus software and prevent such tools from being installed on an already infected system.

Perhaps the most significant development in this field is the convergence of viruses and spam, with certain recent viruses existing solely for the purpose of installing clandestine email servers on the desktop systems they infect. These servers are later employed as relays through which spam emails are sent, and which block the identification of the original sender.

**The Numbers**

The statistics on these threats are amazing. MessageLabs, a company that provides email security services, tracks their occurrence in the billions of messages that flow through their servers. Their Annual Email Security Report for 2004 paints a discouraging picture ([http://www.messagelabs.com/intelligence/2004report](http://www.messagelabs.com/intelligence/2004report)).

They report that spam made up 73% of all emails in 2004, with monthly fluctuations peaking at 94% in July of that year. That sounds like an incredibly high percentage, and I was skeptical when I first read it, but a quick, unscientific survey of my Inbox puts my percentage of junk mail into the same range.

Computer viruses were identified in 6% of all emails. Unlike previous years where a range of distinct viruses were rampant, 2004 saw the emergence of variations on a limited set of known viruses. Whether this reflects better anti-virus software or a shift in the approach taken by their creators is a hotly debated issue.

Phishing experienced the most dramatic growth in 2004. MessageLabs saw a monthly average of around 250,000 phishing emails in the first half of the year. But that ramped up rapidly in the second half to reach around 4,500,000 by year-end, an 18-fold increase in 6 months.

Bear in mind that all this activity on the part of the bad guys is taking place in spite of the widespread use of excellent anti-virus software and spam filters. Collectively, we are working really hard on this problem, but we seem to be losing ground.

**Why Is It Getting Worse?**

Several factors lie behind this seemingly unstoppable growth:

- Internet scams don’t cost much to set up.
- The potential audience is huge.
- The chance of getting caught is low.
- The chance of getting prosecuted is minimal.
- People are making money doing it.
The cost involved in setting up a phishing scam is almost negligible. You need a web server that you control, a little programming experience, and some way to send a lot of email messages. That is an investment of a few hundred dollars at most. All you need is one victim to give up their credit data number and you will have turned a profit.

Creating a large spam operation is a more expensive endeavor, as you need a pool of mail servers that can send out the messages. Using commercial servers, the costs are still low relative to the potential rewards, but that expense can be dispensed with entirely if you are able to commandeer the computers of unsuspecting users. That has been the rationale behind the recent computer viruses, which have installed email relay servers on their infected hosts.

The key to reaching the largest possible audience lies in automating the generation and distribution of email messages. Writing good scripts to do this is easy enough, but in the face of rapidly improving spam filters, increasingly more effort is being applied to the automated generation of messages that can evade these defenses. A form of intellectual arms race is starting to take shape between us and them. I hope that this book and the efforts of its readers will help tip the balance in our favor.

The risk of getting caught and convicted should serve as a strong deterrent to crime. Unfortunately the chances of either of these happening on the Internet are slim. The conviction rate for spamming remains so low that any individual case still attracts significant attention in the press. I discuss this more in Chapter 12.

Above all, the number one reason why Internet crime is growing so rapidly is that people are making money doing it. As long as that remains the case, criminals will find the resources they need to make it happen.

**Pulling Back the Curtain**

Who exactly is involved in Internet crime? The popular media seem to have settled on two very different profiles. The first is the Russian mob that has enlisted physicists, displaced from Cold War era government programs, to help them with their plans. The second is the American teenage boy nerd, seated in the dark isolation of his bedroom, working on the next great computer virus. Neither of these is really representative, although both contain substantial elements of truth. The fact is that the opportunities for this kind of fraud are so broad that someone can find a niche regardless of their technical background.

The advance fee scam, the so-called Nigerian 419 scam, requires nothing more than a good cover story, a list of email addresses, and the gall to carry it out. Creating a computer virus, or operating a professional spam distribution network, requires significant technical expertise. Some scams are so complex that multiple individuals must be involved. For an interesting perspective on a few individuals from the world of spam, I refer you to the book *Spam Kings* by Brian S. McWilliams (O’Reilly). In it,
he describes how two well-known spammers got involved in the trade and how techniques like those described here were used to reveal them.

One thing common to everyone involved in Internet fraud is the desire to remain anonymous and thereby safe from prosecution. The bad guys go to great lengths to hang a curtain of disguise behind which they can operate. The forensic skills that you will learn from this book will help you pull back that curtain.

Just like traditional criminal forensics, you will use your skills to find the clues left behind at a crime scene. The only difference being that our crime scene takes the form of a web site, server, or email message. You are unlikely to uncover the name and address of the culprit, but you will be able to build up a picture of their operation, which can contain a surprising amount of detail.

**Taking Back Our Internet**

Over and above the immediate desire to identify the bad guys, I think a lot of us feel a deeper unease about their activities.

The developers and systems administrators among us talk about the Open Source Community, the informal collection of people responsible for creating and using Linux, Perl, and all the other tools that we use every day in our work. The word “community” is not just a convenient buzzword. Many of us feel a real sense of belonging to this global movement that has made the Internet what it is today.

No one can truly claim ownership of the Internet, but the Open Source Community can rightfully claim to be its stewards and guardians. As such, we feel betrayed by those who have crossed over to the Dark Side who are responsible for the nuisances and threats that all users now have to deal with.

Many developers have already stepped up to the challenge of taking back the Internet. Spam-filtering tools, firewalls, secure browsers, such as Firefox and Mozilla, along with a host of security patches, have been developed by open source developers for the good of the community. With the forensic techniques described in this book, I want to help advance another approach in this ongoing battle. By identifying the people responsible for these threats, we can put them under a great deal of pressure and force them to work much harder to achieve their goals.

I want this book to show you how easy it can be to uncover clues about Internet scams. You don’t need to be a computer security expert to apply these skills. In fact the key to their success lies in having hundreds and thousands of people like you pushing back and putting pressure on the bad guys. Collectively, we can be a very powerful force.
Protecting Your Privacy

Disclosure and privacy are two sides of the same coin. The same forensic techniques that you use to investigate a phishing web site can be used against you by someone else. The techniques do not discriminate. Privacy is a major concern for some people, less so for others. Regardless of where you fall on that scale, you should always be aware of what others can learn about you. Throughout the book, I will play for both teams. I will show you how to, for example, mine a web site for useful data and then show how, as the operator of a site, you can limit that disclosure.

You can make the argument that, by taking this approach, this book may actually help the scammers evade detection. In some cases, this may happen. However, this same issue has been raised many times in the field of conventional computer security. The counter argument, that I think has prevailed in that field, is that most of the bad guys already know how to improve their operations if they choose to. Either they are just lazy, or they don’t think the chance of being identified is high enough to warrant the effort.

By providing a full disclosure of the ways that scammers use to conceal themselves, and showing how you can still uncover identifying information, Internet forensics forces the bad guys further into a corner. There are many more of us than them, and our collective attention forces them to either work harder to practice their trade or, I hope, decide that it’s not worth the effort.

That is exactly what we have seen with other aspects of computer security. In the Linux community, new security problems are disclosed for all to see as soon as they are discovered. That prompts developers to fix the issues in a timely manner. In the early days, some of the vulnerabilities were serious and undoubtedly their disclosure led to some systems being attacked. But overall the approach has been a resounding success. Vulnerabilities are still being discovered, but their impact is typically much reduced and often they are fixed before any real-world exploit has been created. Full disclosure of the ways scammers work has made life increasingly difficult for system attackers and has undoubtedly led many to focus their attentions elsewhere.

The analogy of an arms race is appropriate. It may be an inefficient way to defeat an enemy but it can be very effective way to control their activities.

Before You Begin

I need to offer a few words of caution before you begin poking around some of the more dubious corners of the Internet.

Viruses, Worms, and Other Threats

Computer viruses and spyware are everyday threats on the Internet. But in actively seeking out and examining dubious web sites, you may be exposing your systems to
higher than normal risks. As I describe in Chapter 3, the worlds of spam distribution and computer viruses have already merged in the form of the Sobig virus. This type of threat should not be a problem as long as you take suitable, simple precautions.

A Unix-based operating system, such as Linux or Mac OS X, is the preferred platform from which to investigate dubious web sites and email messages. The Unix environment is less susceptible to computer viruses, with control mechanisms that make it difficult for rogue executables to be installed simply by downloading them.

If you do use a Windows system to follow the techniques and examples given in this book, then you need to take several important precautions. It goes without saying that you need to have good antivirus software installed and running on the system. Not only that, it needs to be kept up to date with current virus definitions. If you are actively exploring web sites, then make sure you scan your system frequently.

The same goes for spyware, which is perhaps even more a problem in the context of visiting web sites. There are some excellent free tools available for finding and eradicating this on Windows computers—for example:

Ad-Aware
   www.lavasoftusa.com/software/adaware/
Spybot - Search & Destroy
Microsoft AntiSpyware
   www.microsoft.com/athome/security/spyware/software/default.mspx

Again, you should scan your system frequently with these tools.

Historically, a major vulnerability on Windows systems has been Internet Explorer itself. A series of vulnerabilities have been exposed, exploited, and then patched over the past few years, giving this browser a poor security reputation. Hopefully those problems are a thing of the past, but if that is a concern, then you might want to use Mozilla Firefox (www.mozilla.org/products/firefox/) as an alternate browser.

**Ethics**

All of the techniques that I describe in the book make use of information that people disclose in the emails that they send and the web sites that they host. That information is readily accessible by anyone who knows where to look.

None of the techniques involve breaking into computers or probing them for vulnerabilities. That crosses the line from legitimate investigation into computer cracking, which in most instances is illegal. I do not, in any way, shape, or form, condone that activity.

But, as with most aspects of life, between these black-and-white extremes lies a gray area where things are not so clear-cut. For example, I have no problem mining a fake
bank web site for every piece of information about its creators that I can find. But I would not dream of using those same skills to identify the people involved in, say, a support group for recovering addicts. To me, one target is legitimate and the other is not.

As you work your way through the book and apply the techniques to real emails and web sites, take a moment to consider the ethical implications of what you are doing. Use your powers wisely and stay away from the Dark Side!

**Innocent Until Proven Guilty**

Whenever they show a telephone number on television, they include 555 after the area code. This is a reserved block of numbers that don’t work, which the film companies use to prevent prank calls to regular phone lines. I have taken a similar approach by masking some of the Internet and email addresses that are used in this book.

Throughout the book, you will find many examples of email messages, domain names, URLs, and web pages. These are used to illustrate different techniques, and most are real examples from my Inbox or real sites that I have visited. Most were examples of spam, phishing, or some other dodgy operation, at that point in time. It is important to realize that most web sites that are involved in a scam are short-lived. The chance that any of these sites will still be operational by the time you read this book is minimal. In many of those cases, the specific Internet addresses will have been reassigned to other sites and most will be completely legitimate. Others may represent innocent sites that had been hijacked in order to host a phishing attempt.

You should not make assumptions about the current usage of any specific numeric addresses, hostnames, or web servers that are included in this book.

**A Network Neighborhood Watch**

Taking back the Internet from the con artists will require more than the efforts of computer security professionals. If it were that easy, then the problem would already have been solved. Educating consumers has undoubtedly helped, but people still fall victim to these scams every day.

I view myself as part of the global community of programmers and systems administrators, the power users of computing and the Internet. I suspect most readers of this book would feel the same affiliation. Given the technical skills that we possess, I feel that we have a collective responsibility to guide the development of the Internet and ensure that the values of freedom and openness are preserved as it continues to evolve.

We have the potential to make life very difficult for those behind Internet scams. With thousands of us working to reveal them, their sense of security will be threatened. I believe that this sense that nobody can touch them is a major reason for the
growth of Internet crime. A community-based effort to uncover these scams has the potential to have a major impact. We need an effort similar to that of ordinary people who take part in a Neighborhood Watch to keep crime away from where they live simply by keeping an eye out for each other. We need a Network Neighborhood Watch.

This book will show you how to uncover information about web sites, servers, and email messages. It was written for anyone with modest computer skills, as opposed to the professional computer security expert. Anyone can apply these techniques. They use the basic tools and protocols of the Internet in creative ways to reveal clues that mostly go unnoticed. I think most readers will be surprised just how much can be revealed.

I encourage you to learn and experiment with the techniques, scripts, and hacks that are described here. If your Inbox is anything like mine, then you already have plenty of targets. I hope that you build upon these ideas and go on to share your own with the rest of this community. And I hope that you will do your part to make the life harder for the bad guys and in doing so, make the Internet a better place for all of us.
CHAPTER 2
Names and Numbers

Hostnames, and the numeric addresses they correspond to, are the way to identify computers on the Internet. Understanding how these names and numbers are managed is therefore a fundamental aspect of Internet forensics. This chapter describes the types of information you can obtain from public databases of Internet addresses and discusses three essential tools that can help you identify machines and the people behind them. I’ll start with a short review of how computers are identified on the Internet.

Addresses on the Internet

Each computer on the Internet has a unique identifier in the form of its Internet Protocol (IP) address. This is a 32-bit integer, which we normally represent as four 8-bit integers separated by periods, such as 208.12.16.5.

Numeric addresses are fine for systems administrators who need to set up networks and who like that sort of thing. But for most people, they are impossible to remember and so we have real names for computers, the hostnames that we are all familiar with, such as www.oreilly.com.

The translation between hostnames and IP addresses is handled by the Domain Name System (DNS). For example, when you type a hostname into a browser as part of a URL, the browser converts the name into the corresponding IP address and then uses that to communicate with the web server. The browser queries a DNS server on the network, which looks up the name in its database and returns the numeric address to the browser.

In its simplest form, a DNS server consists of two tables of data and the software necessary to interrogate them. The first table is a list of hostnames and the IP addresses to which they correspond. The second is a list of IP addresses and the hostnames to which they map. Storing the addresses of all computers on the Internet on every server is not practical, so the data is distributed across many thousands of servers.
About the Author

Dr. Robert Jones runs Craic Computing (http://www.craic.com), a small bioinformatics company in Seattle that provides advanced software and data analysis services to the biotechnology industry.

He worked as a molecular biologist for a number of years before programming got the better of him. He hung up his lab coat in 1989 and has worked in the fields of bioinformatics and high-performance computing ever since.

Colophon

Our look is the result of reader comments, our own experimentation, and feedback from distribution channels. Distinctive covers complement our distinctive approach to technical topics, breathing personality and life into potentially dry subjects.

The image on the cover of Internet Forensics is of a man holding a listening device. Although that particular listening device is decidedly on the large side, hearing aids that are manufactured today are small enough to fit around, or even inside, the ear. The technology for hearing aids has advanced greatly as well. A common complaint among hearing aid users is that the device amplifies all sounds, making it difficult to deal with being in crowds, for example. Although background noise issues still exist, now digital hearing aids are being produced that can help to significantly minimize the problem.

Mary Brady was the production editor and proofreader for Internet Forensics. Nancy Reinhardt was the copyeditor. Mary Anne Weeks Mayo and Claire Cloutier provided quality control. Marlowe Shaeffer provided production assistance. Johnna VanHoose Dinse wrote the index.

Karen Montgomery designed the cover of this book, based on a series design by Edie Freedman. The cover image is a 19th-century engraving from the Dover Pictorial Archive. Karen Montgomery produced the cover layout with Adobe InDesign CS using Adobe’s ITC Garamond font.

David Futato designed the interior layout. This book was converted by Keith Fahlgren to FrameMaker 5.5.6 with a format conversion tool created by Erik Ray, Jason McIntosh, Neil Walls, and Mike Sierra that uses Perl and XML technologies. The text font is Linotype Birka; the heading font is Adobe Myriad Condensed; and the code font is LucasFont’s TheSans Mono Condensed. The illustrations that appear in the book were produced by Robert Romano, Jessamyn Read, and Lesley Borash using Macromedia FreeHand MX and Adobe Photoshop CS. The tip and warning icons were drawn by Christopher Bing. This colophon was written by Mary Brady.