Chapter 1

Digital Forensics and Analyzing Data

Solutions in this chapter:

- **The Evolution of Computer Forensics**
- **The Phases of Digital Forensics**

☑ **Summary**

☑ **Solutions Fast Track**

☑ **Frequently Asked Questions**

Introduction

Digital forensics is probably the most intricate step of the cybercrime investigation process, and often yields the strongest evidence in terms of prosecutable cases. *Digital forensics* is the scientific acquisition, analysis, and preservation of data contained in electronic media whose information can be used as evidence in a court of law. The practice of digital forensics can be a career unto itself, and often is. Other times it is a subset of skills for a more general security practitioner. Although the corporate digital forensic practitioner is not a law enforcement officer, it is a wise practice to follow the same procedures as law enforcement does when performing digital forensics. Even in a corporate environment, the work one performs can quickly make it to a courtroom. Regardless of whether the case is civil or criminal, the evidence will still be presented in the same fashion.

The Evolution of Computer Forensics

Traditionally, the practice of digital forensics encompassed seizure of a computer or some other form of media, followed by bit-by-bit duplication of the drives and media in a forensically sound manner; then the forensic practitioner would comb through the duplication using a hex or disk editor application. Over time, forensic applications and suites evolved and automated or streamlined some of these processes. The forensic practitioner would undelete files, search for temporary files, recover e-mail, and perform other functions to try to find the evidence contained on the media.

Today, user-friendly software programs are available that present data in a GUI and that automate much of the highly technical work which used to require in-depth knowledge of and expertise with a hex editor. In addition, a wealth of hardware is now available to make the practice of digital forensics even more conducive. However, the reality is that the processes used in digital forensics have not changed that much over time.

What has emerged, though, is the following set of *best practices* which provide a foundation for digital forensic work:

- Do not alter the original media in any way.
- Always work on a duplicate copy, not on the original.
- The examination media must be sterile to ensure that no residual data will interfere with the investigation data.
- The investigator must remain impartial and report the facts.

In this chapter, we will discuss these best practices as they relate to the four main phases of the digital forensic process: collection, examination, analysis, and reporting. We will also address some of the technical and procedural challenges a digital forensic examiner faces today.

The Phases of Digital Forensics

You can break down the digital forensic process into four main phases. Some of the work performed may overlap into multiple phases, but the phases themselves are different:

- Collection
- Examination

- Analysis
- Reporting

Collection is the preservation of evidence for analysis. Current best practices state that digital evidence needs to be an exact copy—normally a bitstream copy or bit-for-bit duplication—of the original media. The bitstream copy is then run through a cryptographic hashing algorithm to ensure that it is unaltered. In modern digital forensics, often this is done by physically removing the hard drive from the device, connecting it to a write-blocking unit, and using forensic software that makes forensic duplicates of the data. *Examination* is the methodical combing of the data to find evidence. This includes extracting documents and e-mails, searching for suspicious binaries, and data carving. *Analysis* is the process of using the evidence you recovered to help solve the crime. The analysis pulls together all the bits and pieces and deciphers them into a story of what happened. *Reporting* is the phase where all the other phases are documented and explained. The report should contain documentation of the hardware, the tools used, the techniques used, and the findings. All the individual phases have their own issues and challenges.

TIP

Here are some great resources on computer incident handling and digital forensics:

- NIST "Computer Security Incident Handling Guide," SP800–61: http://csrc.nist.gov/publications/nistpubs/800–61/sp800–61.pdf
- NIST "Guide to Integrating Forensic Techniques into Incident Response," SP800–96: http://csrc.nist.gov/publications/nistpubs/800–86/SP800–86.pdf
- National Institute of Justice, "Forensic Examination of Digital Evidence: A Guide for Law Enforcement": www.ojp.usdoj.gov/nij/pubs-sum/199408.htm
- RFC Guidelines for Evidence Collection and Archiving: www.faqs.org/rfcs/rfc3227.html

Collection

It is a digital forensics best practice to make a full bitstream copy of the physical volume. This usually entails physically removing the hard drives from the suspect system and attaching the drives to another system for forensic duplication. A forensic image is a bit-by-bit copy of the original media. It is a copy of all the data on the storage device, including unused portions, deleted files, and anything else that may have been on the device. The suspect hard drive should be protected from alteration by a hardware solution, a software solution, or both. The hardware solution is normally either a write-blocker or a hardware imaging device. A write-blocker blocks the write commands from the examination system that some operating systems would normally perform. Software solutions entail mounting the suspect drive or device as read-only by the operating system.

The data must be unaltered and the chain of custody must be maintained. Where practical, all the work should be performed on a copy; the originals need to be preserved and archived. To ensure that the data is unaltered, the original drive and the imaged drive are hashed and the hashes are compared to make sure an exact, bit-by-bit copy has been acquired.

NOTE

Hashes use cryptographic algorithms to create a message digest of the data and represent it as a relatively small piece of data. You can compare the hash of the original data to that of the forensic copy. When the hashes match, it is accepted as proof that the data is an exact copy. Although it has not been challenged yet, the traditional hashes of CRC, MD5, and SHA-1 have been cracked. Also, there are limitations in the sheer volume of 128-bit hashing algorithms such as MD5. There are only 2^{128} possible MD5 hashes. If the large, multiterabyte file server being analyzed stores $2^{128} + 1$ files, there will be two different files with unique data with the same hash. Now it is understood that 2^{128} is about 340 billion billion billion billion, which would be an extremely large storage array of tiny files, but this fact opens the door for doubt, which could ruin a criminal prosecution. Although 2^{128} is still a huge number, as storage grows it is not unrealistic to believe that 128-bit hashes will become an increasing issue. It will probably be an issue on large storage systems long before it becomes as big an issue on single workstations. The future appears to be the use of the SHA-256 algorithm and other 256-bit hashes. For now, the National Software Reference Library Hashes use the SHA-1 and MD5 algorithms.

Once you've collected the digital evidence, you must ensure that it meets the following requirements:

- **Admissible** The evidence must conform to certain legal rules before it can be presented in court.
- **Authentic** The data must be proven to relate to the incident. This is where additional documentation is important.
- **Complete** It must be impartial and tell the entire account.
- **Reliable** Nothing relative to the collection and handling of the evidence can create any doubt. Chain of custody procedures are crucial in this regard.
- **Believable** The reports and documentation must present everything so that it is believable and understandable by a judge or jury.

One challenge that is surfacing concerns admissibility. Traditional rules and best practices concentrate on data from static or powered-down systems. Sometimes this approach is difficult or impossible to follow, or leaves large amounts of data behind. Challenges to collecting data for analysis can include getting the files off the systems and then accessing the files. If you cannot physically access the files, how long will it take to move the data off the systems to work with it? An option may be to work with the data on the systems, but do you have enough storage capability to be able to duplicate and analyze the data? If the systems were compromised, can you trust the use of the utilities and binaries on the systems? Most likely, the answer is no.

So, your next option is to move the data off the systems via a network connection. How large is the network link? If you cannot work with the data on-site, do you have the storage capacity to transport it? Do you have the storage capacity to work with it later? Are your systems powerful enough to comb and query through all the data? Are all the systems in the same data center or do you have to travel or have multiple teams working simultaneously? As you can see, you need to be able to answer a multitude of questions, and some preplanning can be essential.

Incidents at a large business or other large network can aggravate these issues, and can be extremely complex. The cybercrime responder will almost surely find a variety of systems running a multitude of operating systems. The devices can encompass nearly everything and anything. The most important step when responding to a large cybercrime incident is to take a few minutes to figure out what kinds of systems you are dealing with. It's worth the time to gather any available documentation, such as network diagrams and system configurations.

The key early on is to avoid tunnel vision. You may need to recover data from a multitude of systems, and in a variety of ways. It is easy to fall into the trap of centering on the first system found to be compromised or involved, when that system may be only the tip of the iceberg. If you concentrate all your efforts on the first system, you may miss all the other evidence initially. Or if the retention times of logs or volatile data are too short, the data may be gone forever. Just like a lost hiker searching for the path out of the woods, you should work in circles, moving outward from the point of discovery. From that initial machine of interest, look outward and concentrate on access paths that lead to the machine of interest. Do not forget the physical paths to a system—access controls and video surveillance are present in most data centers or offices, so you should definitely review physical access logs.

Preparation

To conduct a proper digital forensic examination, you need an assortment of tools, both hardware and software. You should try to get as much information as possible before you start focusing on the suspect systems. If it is in your native environment, preplan what is required for a normal engagement as well as for contingencies. A few extra phone calls or extra minutes to gather extra tools can save hours later trying other acquisition methods or struggling with inadequate hand tools. It can also help you determine whether you need additional resources, or whether the examination is over your head. If you are in a corporate environment you should have the specifications for the critical systems available to assist law enforcement in working with your systems if you are not going to do the acquisitions in-house. Most likely, this information should be available for disaster recovery or hardware failure issues.

Be sure to have enough drives or storage to hold all the forensic images that you will collect. The drives should be prepared beforehand. Preparation should entail wiping the drives to eliminate any existing data and to avoid contaminating the data you collect. This also eliminates any allegations that data was planted or that the evidence collected was tainted. You should also keep a log that documents preparation of the storage media. Considering that many middle-of-the-road PCs today are shipping with 400 GB or larger drives today, making a full bitstream copy or image can be a hardware and time commitment.

A federal law enforcement officer appears at a data center to assist in a cybercrime investigation. He states to the corporate forensics person handling the case, "I'm here to pick up the server." The corporate forensics person stares at him blankly, and then asks, "Did you bring a box truck and a few more men and maybe a few small boys to help?" "Why?" asks the officer. "Because the 'server' is seven racks if you include the storage array!"

When it comes to being prepared for responding to an incident, a Linux machine is a must-have. The Apple Mac will work well in this situation also. A system that can perform a Server Message Block (SMB) and Network File System (NFS) mount, and that can run netcat, ftp, and scp, can be invaluable. Windows systems can do these things also, but they need far more third-party software to do so. A *nix-based system will also be able to mount a wider variety of file systems. Once the data is recovered, all of the native *nix tools will be available to search and manipulate the data.

Notes from the Underground…

Suggested Tool Kit Contents

Your tool kit should contain the following components:

- **Hardware** Target hard drives, write-blocker, and cables (network, IDE, and SCSI)
- **Software** Boot disks and drivers for your forensic system and for any system you may encounter, especially for network cards
- **Tools** Allen keys, large and small screwdrivers (standard, Phillips, and Torx)
- **Other content** Labels, antistatic bags, pens and markers, blank media (CDs, DVDs), and a camera

A final consideration is that you may need to preserve the data in order of volatility. You should preserve the most volatile data first. This applies to running systems for the most part, but the way in which we approach live systems will become more important in the future. An example of an order of recovery of system data according to volatility looks like this:

- **Live system information** Includes memory, the routing table, the Address Resolution Protocol (ARP) cache, and a process list. The concern with live system information is that it is difficult or impossible to image the system memory or other live data without altering the original data.
- **Virtual memory** Swap space or paging files.
- **Physical disks** The physical hard disks of a system.

- **Backups** Offline backup media such as magnetic tape or other media. It is possible that the data you are looking for may not be on the system today, but it was there yesterday and is on last night's backup.

In short, the multitude of systems and devices you may encounter during a cybercrime investigation means you need a large and flexible tool kit that includes not only the hardware and software for dealing with a variety of devices, but also your tricks and procedures for dealing with them. You should also include resources to turn to if you find yourself in a situation that is beyond your skills.

Hardware Documentation Difficulties

Documenting hardware configuration is a tedious but essential part of the forensic process. The magnitude of documentation is in direct correlation to the number and types of devices being acquired. What we as examiners cannot afford to forget are the various aspects to documenting hardware.

Within the documentation process itself, you need to document all the system configurations, including the installed hardware and BIOS settings such as the boot device. Another essential aspect of hardware documentation is the time settings of the system and the system clock of each device. You must document the system time and compare it to the actual time. The time zone setting may also be crucial when creating timelines or performing other analyses. You should note the presence of a Network Time Protocol (NTP) time server. Remember, a system on a Windows domain will sync its time with the domain controller, but the time by default can be off by 20 seconds and still function properly.

Traditional forensics dictates that you document all identifying labels and numbers. Often, an examiner will take pictures of all sides of the system as well as labels on the system as part of the documentation process. This can also be extremely difficult with large systems. It could take a day to unrack and photograph all the systems in a rack. Depending on the approach you take to acquire data from a system, you may need to conduct complete and detailed hardware documentation after acquiring the system. If the system is live, it most likely will not be desirable to shut it down to document it and then to restart it to perform the acquisition. If possible, take no more than a day to analyze a blade server enclosure and the servers in a data center. Consider how to document each blade as you would a typical PC. Then think about the fact that a typical rack can often hold six enclosures holding 16 blade servers. The IT staff at the client company may have decent documentation for you to work from; if you can verify from their existing documentation instead of working from scratch, you can save a lot of time.

A large storage system is probably another example of an instance where you will need to document the devices after you acquire them unless you use the physical option. This is because it may not be practical to image each drive individually. Once the storage system's logical image is complete, you can remove the drives from the enclosure and document them. The documentation of rack after rack of hard drives can be even more daunting than blade servers.

You also should document the network topology and any systems that directly interface with the system, such as through NFS or SMB mounts. If the investigation expands, it may be necessary to increase the documentation of the surrounding network to encompass the switches, routers, and any other network equipment. In the case of an intrusion, any of these paths could be the source of the compromise.

A final item to document is the console location, if one exists. Even today, not all unauthorized access happens through a network connection.

Complete and clear documentation is the key to a successful investigation. If the incident leads to litigation, the report created from the documentation will be a valuable reference for the examiner. Complete documentation will help to remove any doubt cast by the defense or other party in a civil matter.

Difficulties When Collecting Data from RAID Arrays, SANs, and NAS Devices

Enter the corporate or government arena and now the 500 GB hard drive becomes multiterabyte or petabyte storage systems. Faced with a 20 TB storage area network (SAN), the complexity of obtaining a forensic image of physical drives and reassembling the logical volume is considerable. Add the logistics of storing the forensic images or owning the storage hardware "just in case" is not always very practical, due to cost and size of the equipment.

For the sake of argument, let's say you were able to image and hold a 20 TB SAN array, and reassemble it into a logical volume; how much computing power and time would it take to search that volume of data?

The era is approaching where a better triage process needs to occur so that the evidence that is pertinent to the investigation is collected first. The adoption of more parallel operations needs to occur. The examination and analysis phases need to begin as the systems that have been triaged as being less important continue to be acquired and imaged. In time, this will make the examination and analysis processes more efficient, and will allow investigations to be completed in a timelier manner.

Depending on the goals of the investigation, you may not have to collect data from the entire system. If a single individual is under investigation for financial fraud, it may not be of value or necessary to image 20 TB of storage on a file server that affects 200 other employees. It would be more efficient to triage the area where the individual had access and to start with that data.

RAID

A Redundant Array of Inexpensive Disks and Network Attached Storage (NAS) are used to hold large volumes of data and often provide some level of redundancy. A RAID uses multiple disks to provide redundancy or performance enhancements over a single disk. As it applies to forensics, the RAID appears as one logical disk, but spans multiple physical disks. If you remove the individual physical disks and image them separately, you must reassemble the RAID using the forensic software to get the useful data. It is often much simpler to just acquire the logical drive. If your organization policies require it, you can physically acquire the drives after you've performed a logical acquisition. A note about RAID array reassembly: Be sure to get the RAID controller configuration. It can save you a tremendous amount of time later if you need to assemble the physical images.

SANs

SANs, like NAS, are challenging, not only because of their size but also because of the technology involved. The two predominant SAN types are Fibre Channel and iSCSI. The positive thing about SANs is that they are divided into logical unit numbers (LUNs). If the data that is relevant to the investigation is restricted to a single system, the LUN allocated to that system may be the only part of the SAN that you need to acquire. Linux tends to be the logical choice to use as an imaging platform since there are still not many Fibre Channel write-blocks at the time of this writing, but they are appearing more and more. An important point is to make sure the host bus adapter (HBA) is supported. Generally, you can attach iSCSI SANs via the network adapter. If time is more of an issue than budget, iSCSI HBAs with Linux support are available to offload some of the processing from the CPU. The HBAs have an on-board SCSI ASIC, which would provide a considerable performance gain.

The greatest challenge when working with a SAN is the sheer storage required to copy the data. Vendors are building great solutions such as multiterabyte portable RAID enclosures to assist with this issue. Another option is to use software that allows the spanning of target media during an acquisition.

The hardware to deal with large storage systems can be expensive. A multiterabyte portable RAID and a Fibre Channel write-block can run well over $10,000.

NAS Devices

NAS devices are appliances with the sole purpose of providing data storage. It can be challenging to obtain a forensic image from a NAS device since they run limited services and protocols. If you can acquire the image forensically through an attached system, that may be the preferred option. Otherwise, you may need to disassemble the NAS device and image it drive by drive. Many NAS devices are designed and marketed for the home or small business user. They are no longer just in the realm of enterprises.

So, how do we follow the traditional best practices again when there is no practical way to access the drives directly and take physical images? The other very real consideration with large storage systems is that the necessary hardware requires a large investment. Therefore, it would be logical to assume that the system is attached to a system that is at least marginally important. For a business that needs its systems running to generate revenue, it may again become a business decision to limit the scope of work to limit the downtime.

Difficulties When Collecting Data from Virtual Machines

Virtual machines residing on a host system are commonplace for a variety reasons, from enterprise virtual servers to nefarious purposes on a blackhat's machine. Virtualization applications have matured to the extent that reliable systems can be built for production machines, not just for development and testing work, as was the case in the past. What can make virtual machines interesting is that they could comprise one operating system hosting multiple virtualization platforms, each with multiple virtual machines of different operating systems. The forensic practitioner is therefore faced with the specter of multiple OSs, and the complexity of each virtualization application on a single system. Add a RAID or external storage and one may desire a change of profession.

Luckily, most of the major forensic suites today support the most popular virtual disk formats, making acquisition a bit easier. Virtual machines can also be imaged live just like a physical system if a live system is encountered.

A static or dead acquisition depends on the tool choice. One option is to export the virtual disk file from the host machine's image and to mount the virtual disk file as a drive. Another option is to use a tool such as the VMware Disk Mount Utility, which allows the virtual disk to appear as a drive attached to the system; you can then image it with the tool of choice if not natively supported. The reality is that the virtual disk is very similar to a *dd* image with some additional data.

Difficulties When Conducting Memory Acquisition and Analysis

Memory analysis is becoming more necessary and common on running systems. Especially as systems can be compromised without ever accessing the disk, the only artifact may be in memory. Commercial products such as Core Impact do it, so it is conceivable that the product or its technology can be used for nefarious purposes.

Multiple examples of malware, such as the Witty Worm, are memory-resident only. This and other potentially valuable pieces of investigative data will be missed if we continue to examine only systems that have been shut down. The volume of data that is memory-resident today is more than 100 times larger than an entire hard drive from the 1980s. It's another example where the accepted procedures and best practices are lagging behind the technology curve.

TIP

An excellent paper on memory acquisition and analysis, by Mariusz Burdach, is available on his Web site, at www.blackhat.com/presentations/bh-federal-06/BH-Fed-06-Burdach/bh-fed-06-burdach-up.pdf.

Avoid calling a memory acquisition an "image." It is not a true image in the traditional forensic sense. This is because without specialized hardware, it is not really possible to create a bit-by-bit image of the system memory without affecting some part of it. In a way, it is similar in concept to the Heisenberg uncertainty principle: When an electron's location is measured, it is moved. When memory is acquired, it is normally changed.

Most *nixes allow the acquisition of memory fairly easily, because the system sees memory as a file like everything else. You can use *dd* or any of its forensic variants, such as *dcfldd*, to create a memory acquisition. Windows allows access to the physical memory object, but requires administrative privileges to access it. Tools are available that allow the memory to be acquired; the versions of *dd* compiled for Windows are the most common. Tools and scripts are also available to assist in analyzing the dump.

NOTE

Windows XP 64-bit, Windows 2003 Server SP 1, and Windows Vista feature a number of security enhancements. These versions of the Windows operating system block all user mode access to the physical memory.

The future appears to be on hardware-based devices such as dedicated PCI cards or through the IEEE 1394 FireWire interface, but even though the concepts and prototypes have existed for years there are no readily available commercial products.[1, 2] The apparent advantage of hardware solutions is the decreased impact on the running system. For this reason, hardware solutions will most likely emerge as the favored method. There is currently a debate, and there will continue to be so for some time, regarding the practice of memory acquisition. IT is seen by many as contaminating the evidence. Others see it as obtaining all the data and evidence that is available. The often-used defensive analogy is that of a physical crime scene in which the crime scene unit enters the area to recover fibers and fingerprints. Their actions and movements are documented to prove they contaminated the scene as little as possible. In the digital realm, many feel that if the same care is used to document all the actions taken, contamination is controlled and documented.

Examination

Examination consists of the methodical sifting and combing of data. It may consist of examining dates, metadata, images, document content, or anything else. Many forensic practitioners use the same step-by-step process for their examinations: Conduct a keyword search, obtain Web histories, search unallocated space, and search file slack. Your examination method depends on the goal of your investigation, internal audit, criminal prosecution, or civil lawsuit. Remember that forensics is just an aspect of the larger investigation. Since the needs of the examination may change with the investigation, we believe the traditional forensic menu used by many is becoming impractical. Your company should develop checklists to follow in incident investigation that focuses on the type of systems that will be examined.

Larger volumes of data require better triage methods while streamlining the process to allow for deeper inspection of key areas such as the Windows Registry. The increased use of tools such as hashes to filter known files along with other tools to sort the files for focused examination can help to speed the examination process when facing a huge amount of data.

NOTE

Many tools can assist with forensic examination. You can base your tool selection on personal preference, the strengths of the individual application, or your budget. Some forensic packages cost thousands of dollars whereas others are freeware. Regardless of the tools you choose, it is a best practice, when possible, to use multiple tools. The primary reason is so that you do not miss a piece of evidence due to an issue that is inherent to the tool—when multiple tools agree on a finding, it helps to remove any doubts surrounding the reliability of the tool.

Utility of Hash Sets

Hash sets are precompiled lists or databases of known file hashes. For instance, all the files associated with an application install or a series of illegal images are hashed with a cryptographic algorithm and the hashes that result are put into an indexed collection. During an examination, the hashes of the application set are compared to all the hashes of the files found on the system. A matching hash mathematically nearly guarantees that the file is associated with the application, regardless of its name. Hashes traditionally have been used to find known suspicious files such as malware, cracker tools, and illegal images.

Just as you can use hash sets to look for known bad files, through the same process you can use them to locate known good or benign files. You can use hash sets to locate files that are not related to the investigation or are unchanged operating system files, for example, thereby filtering out noise. Depending on the triage of a case, a hash set of known operating system files can quickly filter out a quantity of files that in all likelihood do not need to be examined. The use of hashes to filter out files known to be unaltered from the hardware vendor can greatly reduce the volume of information to be examined and, in turn, the time required to examine a system. The files left behind are either altered or in user space that will probably be where the real evidence or information lies.

TIP

It can save time in the long run if you create personal hash sets as part of your preparation. Creating hash sets of all of an organization's gold or standard images of workstations and servers used for new installs means that only altered or added files need to be analyzed. The files of internal applications can also be hashed and sets can be created to help filter out files that would not be included in more mainstream hash sets.

Difficulties Associated with Examining a System with Full Disk Encryption

An increasingly common issue in digital forensics is full disk encryption. As the issue of lost and stolen laptops continues to impact organizations, many IT departments are turning to full- or partial-disk encryption to protect data. For the forensic practitioner, this usually means the data of interest will be in the encrypted portions of the drive.

If all the data of interest is encrypted, traditional forensic practices will be useless. Your choices are to perform a live image of the system with the encrypted storage mounted, if possible, or to unencrypt the drive after acquisition.

As is the case with many other issues in contemporary digital forensics, this is another area where best practices and procedures are trailing behind the available technology. You should evaluate the solution you use and create your own procedures. In a crunch, the live system image will almost always be faster than a mirrored image.

Trusted Platform Module (TPM)

The Trusted Platform Module is another emerging technology that will enhance existing encryption schemes. The TPM is a chipset being installed in newer machines that stores keys, passwords, and certificates. The chipset provides for hardware-based encryption functionality that may prove to be a challenge in gathering forensic information due to the native encryption technology.

A suggested methodology for dealing with drives that have been encrypted with full disk encryption follows:

1. Image in state traditionally.
2. Restore the acquired image back to a sanitized target disk.
3. Decrypt the target disk.
4. Acquire the decrypted target disk.
5. Analyze the decrypted disk as normal.

This methodology—although significantly increasing the time required and doubling the required storage—leaves the original data unaltered and maintains a forensic image of the original. It sounds simple, but the challenge is in step 3. Decrypting the drive may take a few Cray supercomputers and the code breakers of the NSA if the encryption is strong and the key is unavailable. In lieu of those resources, you can use the normal tricks of password cracking. The requirement for complex passwords and the volume of passwords the average user must remember have rekindled the trend of written down passwords. When searching for passwords look for hiding places within an arm's length. Remember to check for passwords during the incident response and seizure phases. Another trick is to use the other evidence found to create a dictionary to use for a brute force attack. Remember that the hash of the original encrypted drive will not match the unencrypted drive. They are different data sets and you need to document them as such.

Alternative Forensic Processes

A newer concept, at least in name, is *fast forensics*. Fast forensics is defined as "those investigative processes that are conducted within the first few hours of an investigation, that provides information used during the suspect interview phase. Due to the need for information to be obtained in a relatively short time frame, fast forensics usually involves an on site/field analysis of the computer system in question."[3] The implementation of fast forensics creates a need for some additional resources and procedures to perform some examination and initial analysis functions outside the lab. The focus is to provide important intelligence to give investigators key pieces of evidence or leads to use in interviews or other searches.

Some fast forensic techniques utilize Linux or other forensic boot disks to perform on-scene searches or document extraction. The boot disks run in memory only and mount the hard drives as read-only so as not to corrupt the evidence.

Analysis

Every cybercrime incident will involve at least some analysis of data retrieved from systems, whether it's only a few small files from a system or two or terabytes from many machines. The core of an investigation could consist of a single piece of media or it may consist of thousands of hard drives. The trick to success lies in the analysis that will put all the pieces together. The analysis of an entire cybercrime event can be far more complex than the analysis of any of the systems themselves; the sum of the parts is truly greater than the whole. It can be likened to a symphony. Any single instrument may be difficult to play, but to bring all the pieces together is far more complex. The cybercrime investigator needs to build a toolbox of utilities to analyze the data from myriad systems and be able to correlate the data into a complete, coherent picture.

In the analysis phase of the digital forensic process is where we look deep into the data. The analysis is the sum of all the data applied toward the resolution of the incident.

An example of an analysis follows.

An intellectual property theft case didn't yield much until the data from a bunch of systems was pulled together. The file server audit logs were reviewed and the user list it provided was used to query the proxy server logs. When the log files for those uses were reviewed a short list was created by focusing on webmail and forum traffic. The short list was used to triage and prioritize the exams of the user workstations. The exams of the workstations quickly revealed the individual when the webmail messages were pulled from the Internet cache, and re-created.

During the analysis phase, it is imperative that you tie in any other investigation intelligence that has been gathered. In this phase, the data from multiple systems or sources is pulled together to create as complete a picture and event reconstruction as possible. The evidence used in court is different from the evidence used to find the next piece evidence for an investigation. A piece of evidence may not be strong enough to stand on its own, but may be the item that provides the next lead.

Another challenging factor is that analysis of large amounts of data takes time. In the heat of an incident or a large, high-profile investigation, it is often difficult to manage the expectation of management. It can take huge amounts of time to import logs into various applications, and to move and copy data between storage systems. Be prepared to explain to management why it could take weeks or months to comb through all the data you've collected, especially if the incident affects customer data and has reporting requirements.

Notes from the Underground…

Anti-Forensics

Anti-forensics is the movement to exploit weaknesses in the forensic process or tools. It can also be the act of hiding data from forensic examiners. Old anti-forensic techniques were as simple as running a script to perform a *touch* command on every file to alter the files' dates and timestamps, and included log and temporary file deletion. Other tools and techniques have emerged that are far more sophisticated, including the following:

- **Metasploit** Well known for the well-integrated suite of penetration testing tools, the Metasploit Framework had branched out into a suite of anti-forensic tools.
- **Timestomp** This tool allows you to modify all four NTFS timestamp values: modified, accessed, created, and entry modified.
- **Slacker** This tool allows you to hide files within the slack space of the NTFS file system.

Continued

- **Transmogrify** This tool defeats forensic tools' file signature capabilities by masking and unmasking your files as any file type.

In addition, the following may not be considered as "anti-forensic" as those in the preceding list, but you should know about them nonetheless:

- **Sam Juicer** This is a Meterpreter module that dumps the hashes from the SAM, but does so without ever hitting the disk. Tools such as pwdump access the disk and potentially leave more footprints (www.metasploit.com/research/projects/antiforensics/).
- **The Defiler's Tool Kit** This consists of a pair of tools that allow for more secure deletion of files on UNIX systems. The tool kit includes Necrofile and Klismafile, which alter the file system to remove evidence of the files that once existed. Necrofile overwrites or basically wipes the inodes that no longer have a file name associated to it. Klismafile does the same to the directory table. In theory, you can detect the use of Klismafile by noticing the blank space in the directory table, but you would have to explicitly look for it. More information about the Defiler's Tool kit is available at www.networkintrusion.co.uk/index.php/products/Forensic-Solutions/Anti-Forensic-Tools/The-Defiler's-Toolkit/details.html.
- **Commercial tools** Anti-forensic tools are no longer only in the realm of überhacker. With the availability of commercial tools to perform secure deletion, even novice computer users can work to hide their electronic footprints. Two examples of commercial tools are Robin Hood Software's Evidence Eliminator (www.evidence-eliminator.com/) and Webroot Software's Window Washer (www.webroot.com/En_US/consumer-products-windowwasher.html?rc=4929&s_kwcid=window%20washer|2584030850).

Although these tools are not foolproof, they can make the forensic task much more difficult.

Just as the investigation of a cybercrime event can involve a variety of systems or devices, it can also involve a single machine or thousands. The addition of multiple systems complicates the analysis process as the data from the many examinations is pulled together.

Analysis of a Single Computer

Most cybercrime investigations involve examination of a system or device, and most start with examination of a single computer. The focus of the examination can be as diverse as the tasks for which the computer is used.

Metadata

Metadata is data about data. Examples are the author of a Word document and the creation date of a spreadsheet. A resource for an overview of Microsoft Office metadata is Microsoft KB223396. Depending on the scope or type of investigation, do not discount the importance of metadata.

A case that got its big lead from document metadata was the BTK case. The serial killer known as BTK sent Wichita TV station KSAS a floppy disk with a message contained in a document. A forensic exam of the floppy disk revealed a file and some deleted files. The file metadata of the Test Art.rtf file showed that the file was last saved by user Dennis and listed the name of a church. A search for the church's Web site revealed that the president of the congregation was Dennis Rader, who was eventually convicted of the BTK murders.

Exchangeable Image File Format

Exchangeable Image File Format (EXIF) is metadata contained in an image file, and although it varies among devices, it can provide valuable information such as the make and model of the camera that took the image of a system, as well as whether an image was altered with a graphics program. EXIF data also often will have a date and timestamp of when the image was taken or altered. There are several EXIF formats; therefore, the data can vary slightly. Also be aware that not all devices will propagate all the data.

Binary and Malware Analysis

Some binary and malware analysis ability is a requirement of digital forensic practitioners. The initial step in this type of analysis is to identify any malware that may be present on a system. This is often achieved through either being identified by hash sets, or not filtered by a hash set. Once a file that is suspicious is identified, you can analyze it in one of two ways: statically or dynamically.

Static analysis entails searching the binary for text strings or identifying whether the file was packed. Packing an executable compresses the file, normally to make reverse engineering more difficult.

Dynamic analysis uses behavioral analysis to identify the malware or its actions. The file is placed in a safe environment such as a test network or virtual machine. The file is then executed and its actions are observed in a sandbox, or isolated area, for software. Items such as network traffic generated or files accessed are noted and used to analyze the binary.

Notes from the Underground…

Virtual Machines

Virtual machines are the crash test dummies of forensics. In addition to being useful for malware analysis, they can be useful for documenting the actions of legitimate software or even user actions. When faced with trying to find out where evidence related to certain programs may reside on a system, testing in a virtual machine allows the dynamic monitoring to lead the examiner to the static artifact on the real system.

Deleted Items

The strength of forensic applications is their ability to recover deleted files in their entirety, or at least the artifact that proves the files existed. When an operating system deletes a file it does not remove the data. It only changes the pointer to the file to tell the file system that the file no longer exists and the space is available for new data. Forensic applications can identify the deleted files that still exist or display the artifact that proves they once did exist. Deleted files may affect the culpability of a suspect by demonstrating willful actions to hide his or her transgressions.

Data Carving

Files of different types have pieces of data at the beginning and end that define the file. These pieces of data are called the *headers* and *footers*. Using the signatures of the headers and footers, digital forensic applications and tools can recover or carve files or pieces of files out of the *cruft* that ends up on storage media. Files that contain plain-text characters can have words carved out of their remnants. Data carving can be time-consuming and tedious, but it can also be rewarding because evidence can be recovered that would otherwise have been missed.

E-Mail Analysis

The analysis of e-mail has a burden of legal process in addition to technical challenges. For law enforcement agents, the legal process depends on the state (virtual, physical, or logical) of the data. For the private sector, the proper policies need to be implemented and reviewed by attorneys to address the expectation of privacy issues.

Far more analysis can be performed on e-mail than just header analysis. E-mail analysis can depend on whether the data is stored on the server or on the client. Do not overlook the utilities included in the server or client platform for search and advanced search functions. Import and export functions are usually included that allow the data to be analyzed in other applications. For example, you can export a Microsoft Outlook PST file to Excel for analysis. Then you can run summary reports such as a pivot table count to find trends.

TIP

A powerful commercial tool for analyzing many types of e-mail formats is Paraben Forensics Email Examiner. In addition to working with many e-mail file formats, it can recover deleted e-mail and perform advanced searches on a variety of e-mail formats from multiple vendors. Visit www.paraben-forensics.com/catalog/product_info.php?cPath=25&products_id=393 for more information.

Analysis of an Enterprise Event

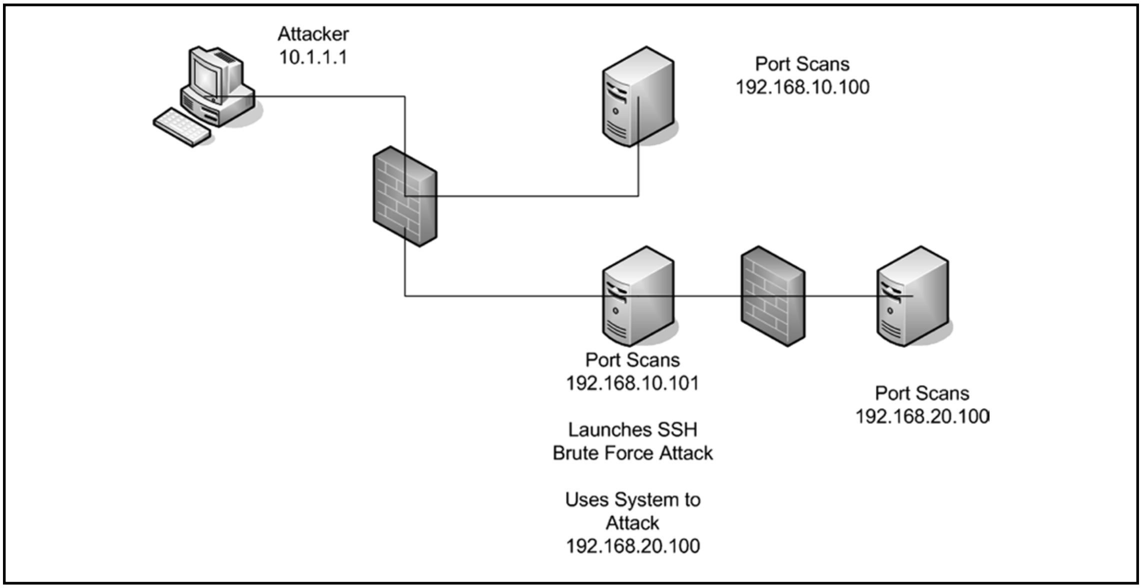
The examination of a single machine can be complex and time-consuming, but it can also be the tip of the iceberg in a digital forensic examination. The complexity of a single workstation exam can be multiplied hundreds or thousands of times over. The likelihood of multiple operating systems and architectures and the additional burden of potentially complex network configurations can task even highly skilled practitioners.

Therefore, you need additional tools to help correlate the data from a number of individual systems and devices into a comprehensive form where it can be digested and analyzed. A series of log files can take on a whole new meaning when presented graphically. Examples of these are system flow charts and event timelines.

System Flow Charts

A flow chart, or other graphical representation of the network, can show which systems were impacted and when based on the analyzed data (see Figure 1.1). The chart would show the data excerpt of an Internet Protocol (IP) address from the firewall log. Next, it could show the snippet of a directory transversal from the Apache logs, and so forth. A system flow chart can be valuable especially when explaining the incident to nontechnical individuals.

Figure 1.1 System Flow Chart

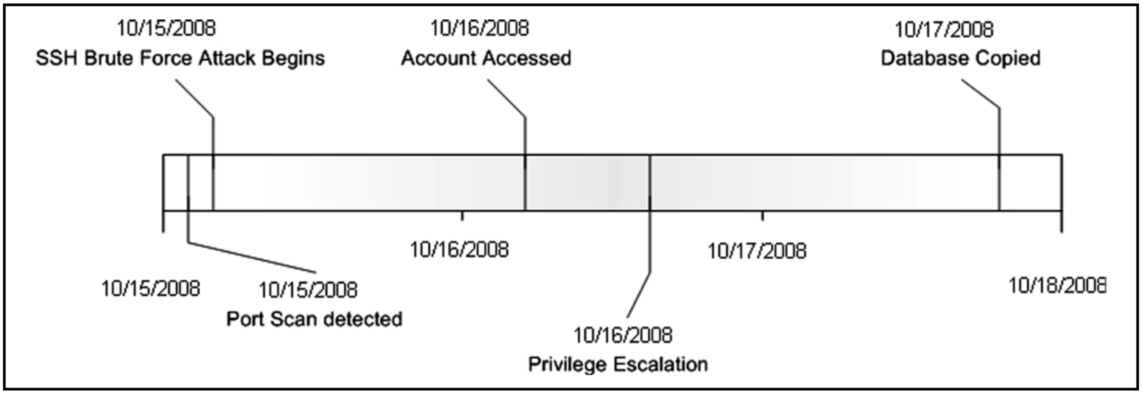


Beyond the usefulness of the graphical representation of the traffic, a system flow chart when compared to a network diagram may help to point out areas that may have been affected but are not yet identified. Graphical documents tend to work well when explaining results to nontechnical management or if the events lead to litigation, attorneys, and juries.

Timelines

A timeline graph of the incident or the analysis can also be valuable. It can help to display the entire progression of what analysis was conducted when and on what system (see Figure 1.2). It is often easier to look at a chart and see the progression of an incident instead of sifting through 100 e-mails after the fact. Also, a timeline could show what systems were impacted and when based on the analysis data. The chart would show the data excerpt of an IP address from the firewall log, as well as the snippet of a directory transversal from the Apache logs, and so forth.

Figure 1.2 Timeline Graph



Timelines are useful for laying out the progression of events as they unfolded. They also are useful for highlighting gaps in activity that contain evidence that was missed or activity that has not yet been uncovered. As mentioned before, graphical documents tend to work well when explaining results to nontechnical management or if the events lead to litigation, attorneys, and juries.

Tools for Data Analysis

There are as many ways to analyze data as there are log files. However, they each have their tradeoffs, whether in terms of cost, performance, or complexity. Often, tools that system administrators use on a daily basis to perform proactive troubleshooting and tuning can be the same tools used for reactive analysis.

Normally, as the tools increase in performance, they also increase in cost and/or complexity. Some of these tools are GREP, PERL scripts, Spreadsheets, Structured Query Language (SQL), and commercial network forensic tools.

GREP

GREP is an indispensable tool and an essential skill for the incident responder or forensic practitioner. The *GREP* command simply searches a file or files for a pattern. The power of GREP is in the flexibility of the patterns that can be created or the ability to recursively search directory structures of files. GREP is licensed under the GPL, so it costs nothing. It exists natively on virtually every *nix operating system, and has been ported to everything else. For the novice, there are many Internet sources on how to craft GREP patterns. An important limitation to remember is that GREP works on text-based files, and will not be able to search every file you may encounter. If you are dealing with large text-based log files, however, GREP is extremely useful.

Spreadsheets

If you are a more visual person, you are more comfortable in a GUI, and your log files are relatively small, a spreadsheet may be an option. Spreadsheets can sort, count, and manipulate your data. Another bonus is the ability to create visual graphs and charts based on your data, to explain to management, law enforcement, the prosecutor, or the jury later. Simple functions can be created to display items such as unique IP addresses or counts of IP addresses. If the log files are fairly small, the uses are limited only by your ability to create formulas or manipulate the data.

Databases

If your log files are large, another available tool is databases. Databases are used on a daily basis to store and report on data, so why not for log files involved in cybercrime incidents? Which database you should use is a matter of budget and expertise. Some issues to keep in mind are the overhead involved in the essential aspects of the database, such as primary keys. This additional data will add to your storage requirements.

One advantage of SQL databases is that the ways to analyze and report the data are limited only by your creativity. Additionally, SQL databases allow you to correlate logs from various systems once you've loaded them into tables. Therefore, you can load all the system logs and query the database to find everywhere an IP address has gone or attempted to go. Finally, since SQL queries are a standard, they can be easily explained to those who are familiar with SQL.

The disadvantages of a SQL database are that they can require huge volumes of storage if you have large log files and want to correlate the data. Complex queries of large databases can also require a lot of processing power or time. Correlation and reporting can take even larger amounts of computing power or time.

The flexibility and power of the SQL database makes it an invaluable tool for crunching through massive numbers of log files and correlating them into a comprehensive report.

Snort

Snort is an open source application that you can use to analyze captured files, not just real-time traffic. Snort is useful for parsing out attack signatures from captures where an IDS may not have been. An added benefit is that you can use Snort to parse out traffic that may not traditionally be an attack but may be valuable to an investigation, such as login attempts. Since Snort is an open source application, its cost is low. Snort also has a supportive user community, and it is well documented. Plenty of resources are available to assist in creating custom signatures.

Security Event Management Systems

Many organizations have begun to install Security Event Management (SEM) systems to compile and correlate all the logs from the various systems. The SEM systems may well be the future of analysis tools for the network. A SEM system can quickly correlate data from a variety of security appliances and systems.

SEM systems are valuable for analyzing data through correlation and reporting. A caveat to SEM system reporting is that the logs received or displayed often are altered (truncated or normalized), so you will need to retrieve and preserve the original raw logs from the originating system.

Many SEM systems are still plagued by performance issues as they struggle to deal with the deluge of data streaming from systems. The databases often have performance issues in large implementations as well.

If a SEM system is implemented well and is operating in an enterprise, it is an excellent resource for assisting in triaging affected systems early in an incident.

Reporting

Once the examination and analysis are complete the most tedious but arguably the most important phase of the digital forensic process begins: reporting.

The report is a compilation of all the documentation, evidence from examinations, and analysis obtained during an investigation. The report needs to document all the systems analyzed, the tools used, and the discoveries made, needs to include the dates and times of the analysis and detailed results, and should be complete and clear so that the results and content are understood years down the road.

Reporting may be the most important phase of digital forensics. If the report is incomplete, or does not accurately document the tools, process, and methodology used, all the work may be for nothing. Reporting will vary depending on your organization's needs, but in most cases a report must at least include documentation of the devices that were examined, the tools used, and the factual findings. Even if a procedure was used and yielded nothing of value, it should be documented, not only for completeness but also to demonstrate that the examination covered all the bases.

Perhaps the greatest challenge after all the other hurdles of acquisition, examination, and analysis have been met is how to present everything you've collected in a manner that cannot be questioned. There is a very real risk that some newer forensic techniques have not yet been challenged in a courtroom.

WARNING

Document that all the software used was properly licensed. It may not be necessary to go into great detail regarding the licenses, but close that hole early.

In a corporate environment, there is often a need for multiple reports—the forensic analysis report and the report created for executive management, at a minimum. A challenge is that in the midst of an important or high-profile investigation, management will want updates and answers. Often when the incident involves volumes of data, one is being asked for answers when it is premature to give them. A strategy may be to provide a "shiny thing" to distract management long enough to get some results. The shiny thing may be just a statistical report and a high-level overview of the occurrence, such as the acquisition of 10 systems for a total of 7.5 TB of data that is now being examined and analyzed.

Other ways to present the data in reports include timelines and a flow chart of accesses. A timeline report of a forensic examination of a system would display the dates and times of file accesses, and a timeline report of data from disparate systems would show the steps taken during the investigation or analysis. The flow chart would show details of the impact or interaction with a system, such as the traffic through a firewall, and then the access to a server.

Summary

The greatest challenge for forensic practitioners going forward will be to forge ahead without best practices to back them up. Forensic practitioners will need to accomplish the same tasks in a more diverse and volatile environment. It is becoming the norm that devices may not be completely imaged because it is sometimes impossible to take a complete physical image of a device. It may also be impractical to take a physical image of an entire multiterabyte SAN array.

The sheer volume of diverse devices and formats will make it much more difficult for forensic practitioners to be experts on everything. It will also create an increasing need for continuing education. The tool kit required to work in digital forensics is not like the handyman's toolbox; it has become the mechanic's large tool chest.

A refreshing trend is the increasing focus of academia on the research of digital forensics. There also has been an increase in academic programs specifically for digital forensics, bridging the gap between traditional computer science and IT degree programs and criminal justice curricula.

The last piece of wisdom: Know when to ask for help.

Solutions Fast Track

The Evolution of Computer Forensics

☑ The technology is changing faster than forensic best practices.

☑ The volume of data is increasing rapidly.

☑ The drive diversity continues to grow.

☑ Some data is increasingly volatile.

The Phases of Digital Forensics

☑ Data storage diversity requires many tools and procedures.

☑ Increased data storage requires large target storage devices.

☑ The time requirement for collection will continue to increase.

☑ More data collected equates to more data to sift through.

☑ An increased use of techniques to reduce the data of interest should be employed.

☑ The increase in the data available can simplify the final analysis, or it can just create a bigger haystack in which to hide the needle.

☑ Analysis of the entire incident is far more complex than examination of any single system.

☑ Reporting is possibly more important than ever, as techniques and procedures must be more finely documented because of potential impacts on volatile data.

☑ A poor report can make the best cybercrime investigation appear to be a disaster.

Frequently Asked Questions

Q: Is specialized equipment required for proper digital forensics?

A: Yes. The debate continues as to the requirement for formal digital forensic training, but training in the proper processes and methods is required.

Q: What is the most important part of digital forensics?

A: The procedures and methodologies are the foundation. If they are solid, the rest will follow.

Q: Will one piece of forensic software do everything I need?

A: You can never have enough tools in your toolbox. That being said, the major forensic suites should handle most of the functions the average digital forensic practitioner may need. It is also a best practice to back up your findings with a second tool, so more than one may be needed.

Endnotes

[1] Brian D. Carrier and Joe Grand. "A Hardware-Based Memory Acquisition Procedure for Digital Investigations." Brian D. Carrier, www.digital-evidence.org/papers/tribble-preprint.pdf.

[2] Adam Boileau. "Hit by a Bus: Physical Access Attacks with Firewire." Security-Assessment.com, www.security-assessment.com/files/presentations/ab_firewire_rux2k6-final.pdf.

[3] "Fast CyberForensic Triage (FCT)." National White Collar Crime Center, www.nw3c.org/ocr/courses_desc.cfm?cn=FCT.

We fully acknowledge use of Chapter 7, "Digital Forensics and Analyzing Data," from Cyber Crime Investigations: Bridging the Gaps Between Security Professionals, Law Enforcement, and Prosecutors, ISBN 978-1-59749-133-4.