A reprint from
American Scientist
the magazine of Sigma Xi, The Scientific Research Society

This reprint is provided for personal and noncommercial use. For any other use, please send a request to Permissions, American Scientist, P.O. Box 13975, Research Triangle Park, NC, 27709, U.S.A., or by electronic mail to perms@amsci.org.
©Sigma Xi, The Scientific Research Society and other rightsholders

NOTE: The views expressed in this article are those of the author and do not reflect the official policy or position of the Naval Postgraduate School, the Department of the Navy, the Department of Defense or the U.S. Government

Digital Forensics

Modern crime often leaves an electronic trail. Finding and preserving that evidence requires careful methods as well as technical skill.

Simson L. Garfinkel

Since the 1980s, computers have had increasing roles in all aspects of human life—including an involvement in criminal acts. This development has led to the rise of *digital forensics,* the uncovering and examination of evidence located on all things electronic with digital storage, including computers, cell phones, and networks. Digital forensics researchers and practitioners stand at the forefront of some of the most challenging problems in computer science, including "big data" analysis, natural language processing, data visualizations, and cybersecurity.

Compared with traditional forensic science, digital forensics poses significant challenges. Information on a computer system can be changed without a trace, the scale of data that must be analyzed is vast, and the variety of data types is enormous. Just as a traditional forensic investigator must be prepared to analyze any kind of smear or fragment, no matter the source, a digital investigator must be able to make sense of any data that might be found on any device anywhere on the planet—a very difficult proposition.

From its inception, digital forensics has served two different purposes, each with its own difficulties. First, in many cases computers contain evidence of a crime that took place in the physical world. The computer was all but incidental—except that computerization has made the evidence harder for investigators to analyze than paper records. For example, financial scam artist Bernard Madoff kept track of his victims' accounts by using an IBM AS/400 minicomputer from the 1980s. The age of the computer helped perpetuate his crime, because few people on Wall Street have experience with 25-year-old technology, and it created an added complication after Madoff was arrested, because investigators had few tools with which to make sense of his data.

Today personal computers are so ubiquitous that the collection and use of digital evidence has become a common part of many criminal and civil investigations. Suspects in murder cases routinely have their laptops and cell phones examined for corroborating evidence. Corporate litigation is also dominated by electronic discovery of incriminating material.

Information on a computer system can be changed without a trace, the scale of data to be analyzed is vast, and the variety of data types is enormous.

The second class of digital forensics cases are those in which the crime was inherently one involving computer systems, such as hacking. In these instances, investigators are often hampered by the technical sophistication of the systems and the massive amount of evidence to analyze.

Digital forensics is powerful because computer systems are windows into the past. Many retain vast quantities of information—either intentionally, in the form of log files and archives, or inadvertently, as a result of software that does not cleanly erase memory and files. As a result, investigators can frequently recover old email messages, chat logs, Google search terms, and other kinds of data that were created weeks, months or even years before. Such contemporaneous records can reveal an individual's state of mind or intent at the time the crime was committed.

But whereas pre-computer evidence, such as handwritten letters and photographs, could be reproduced and given to attorneys, judges, and juries, computerized evidence requires special handling and analysis. Electronic data are easily changed, damaged, or erased if handled improperly. Simply turning on a consumer GPS may cause the device to delete critical evidence. Additionally, computers frequently harbor hidden evidence that may be revealed only when specialized tools are used—for example, a digital camera may appear to have 30 photos, but expert examination may show another 300 deleted photos that can be recovered. (When a device "erases" a file, it doesn't clear the memory space, but notes that the space is available; the file may not be really deleted until a new one is written over it.)

Because they can look into the past and uncover hidden data, digital forensics tools are increasingly employed beyond the courtroom. Security professionals routinely use such tools to analyze network intrusions—not to convict the attacker but to understand how the perpetrator gained access and to plug the hole. Data recovery firms rely on similar tools to resurrect files

Simson L. Garfinkel is an associate professor at the Naval Postgraduate School. He holds six U.S. patents for computer-related research and has written 14 books, including Database Nation: The Death of Privacy in the 21st Century (O'Reilley, 2000). Internet: http://simson.net

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

Craig Cunningham

High-tech crime fighting is now needed everywhere, as shown in this lab belonging to the West Virginia State Police Digital Forensics Unit. Police these days typically use powerful computers and software to copy, analyze, and decrypt data from a suspect's electronic devices.

from drives that have been inadvertently reformatted or damaged. Forensic tools can also detect the unintentional disclosures of personal information. In 2009 the Inspector General of the U.S. Department of Defense issued a report stating that many hard drives were not properly wiped of data before leaving government service.

Digital evidence can even be examined to show that something did not happen. Here they are less powerful, for the well-known reason that the absence of evidence is not the evidence of absence. In May 2006 a laptop and external hard drive containing sensitive personal information of 26.5 million veterans and military personnel was stolen from an employee at the U.S. Department of Veterans Affairs. After the laptop was recovered in June 2006, forensic investigators analyzed the media and determined that the sensitive files probably had not been viewed.

One way to make such a judgment is by examining the access and modification times associated with each file on the hard drive. But someone taking advantage of the same forensic techniques could have viewed the laptop files without modifying those timestamps, so the investigators really determined only that the files had not been opened by conventional means.

These examples emphasize that the possibilites of digital forensics are bounded not by technology but by what is cost-effective for a particular case. Convictions are frequently the measure of success. In practice there is a considerable gap between what is theoretically possible and what is necessary; even though there may be an intellectual desire to analyze every last byte, there is rarely a reason to do so.

Following Procedures

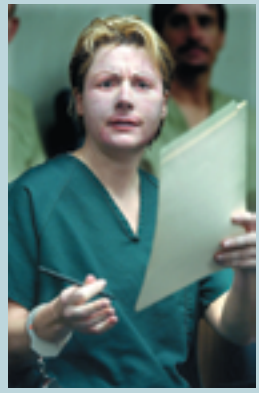
Digital forensics relies on a kit of tools and techniques that can be applied equally to suspects, victims, and bystanders. A cell phone found on a dead body without identification would almost certainly be subjected to analysis, but so would a phone dropped during a house burglary. How the analysis is performed is therefore more a matter of legal issues than technological ones. As the field has grown, practitioners have tried to create a consistent but flexible approach for performing investigations, despite policy variations. Several such *digital forensic models* have been proposed, but most have common elements.

Before data can be analyzed, they are collected from the field (the "scene of the crime"), stabilized, and preserved to create a lasting record. Understanding the inner workings of how computers store data is key to accurate extraction and retention. Although computers are based entirely on computations involving the binary digits 0 and 1, more commonly known as *bits*, modern computers do most of their work on groups of eight bits called *bytes*. A byte can represent the sequences 00000000, 00000001, 00000010, through 11111111, which corresponds to the decimal numbers 0 through 255 (there are two options, 0 and 1, with eight combinations, so $2^8 = 256$). One common use for bytes inside the computer is to store written text, where each letter is represented by a specific binary code. UTF-8, a common representation, uses the binary sequence 00100001 to represent the letter A, 00100010 for the letter B, and so on. (Computers often use hexadecimal codes in memory as well; see figure on page 373.)

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

Major Convictions, and a Few Gaffes, with Digital Data

Famous criminal cases show the power of digital forensics, but a few also highlight the need for careful handling of data and devices.



AP Images

• On December 17, 2000, John Diamond shot and killed Air Force Captain Marty Theer. The victim's wife, Michelle Theer *(right)*, was implicated in the crime, but there was no eyewitness evidence. What prosecutors did have was 88,000 emails and instant messages on her computer, including clear evidence of a sexual relationship between Theer and Diamond, and messages documenting the conspiracy to murder her husband. Theer was found guilty on December 3, 2004, of murder and conspiracy and sentenced to life in prison.

Getty Images News

• In 2005, after eluding police for more than 30 years, Dennis Rader *(above)*, a serial killer in Kansas, reemerged, took another victim, and then sent police a floppy disk with a letter on it. On the disk forensic investigators found a deleted Microsoft Word file. That file's metadata contained the name "Dennis" as the last person to modify the deleted file and a link to a Lutheran church where Rader was a deacon. (Ironically, Rader had sent a floppy disk because he had been previously told, by the police themselves, that letters on floppy disks could not be traced.)

• On January 1, 2002, Scott Tyree kidnapped and imprisoned 13-year-old Alicia Kozakiewicz. He sent an instant message of a photograph showing Kozakiewicz to another man, who contacted the FBI and provided the Yahoo! screen name of the person who had sent the message: "masterforteenslavegirls." FBI investigators contacted Yahoo! to obtain the IP address for the person who had used the screen name, then contacted Verizon to learn the name and physical address of the subscriber to whom that IP address had been assigned, leading them to Tyree.

• On July 12, 2008, the strangled body of Nancy Cooper was found. Her husband, Brad Cooper *(below)*, was charged with the crime. This time the role of digital data was more complicated. Local police, who did not seek expert assistance, accidentally erased the victim's phone while attempting to access it. Brad Cooper maintains he went to the grocery store on the morning of his wife's death, and that she called him during that time, but prosecutors charged that he had the technical expertise and access to the necessary equipment to fake such a call. Investigators searching Brad Cooper's computer also found zoomed-in satellite images of the area where his wife's body was discovered, downloaded one day before she was reported missing. Defense attorneys countered that searches done on that and surrounding days contained inaccurate timestamps. Brad Cooper was convicted of murder, although appeals are ongoing.

MCT via Getty Images

When recorded on a hard drive or memory card, these bytes are grouped in blocks called *sectors* that are typically 512 or 4,096 bytes in length. A sector is the smallest block of data that a drive can read or write. Each sector on the disk has a unique identifying number, called the sector's *logical block address*. An email message might require 10 or 20 sectors to store; a movie might require hundreds of thousands. A cell phone advertised as having "8 GB" of storage has 8 billion bytes or roughly 15 million sectors.

Depending on the arrangement of other files on the device, the sectors can be stored as a single sequential stream or fragmented into many different locations. Other sectors contain information that the computer uses to find the stored data; such bookkeeping material is called *file system metadata* (literally "data about data").

To preserve the data on a computer or phone, each of these sectors must be individually copied and stored on another computer in a single file called a *disk image* or *physical image*. This file, which contains every byte from the target device, naturally includes every visible file. But the physical image also records invisible files, as well as portions of files that have been deleted but not yet overwritten by the operating system.

In cases involving networks instead of individual machines, the actual data sent over the network connection are preserved. Thus network forensics is equivalent to a wiretap—and, indeed, law enforcement is increasingly putting network forensics equipment to this use.

The random access memory (RAM) associated with computer systems is also subject to forensic investigation. RAM gets its name because the data it stores can be accessed in any order. This fast access makes RAM particularly useful as temporary storage and working space for a computer's operating systems and programs. But RAM is difficult to work with, because its contents change very quickly and are lost when a computer is turned off. RAM must be captured with a dedicated program (a *memory imager)* and is stored in its own special kind of file, called a *memory dump*. Although data in RAM can be extracted from all kinds of electronic systems—not just desktops, laptops, and cell phones but also network communications equipment such as wireless routers—each of these systems uses different kinds of internal software structures, so programs de-

eproduction
i.org.

signed to analyze one may not work on another. And even though forensics researchers have developed approaches for ensuring the forensic integrity of drive copies, currently there is no widely accepted approach for mathematically ensuring a RAM dump.

Preserving the data is only the first step in the process. Next, an examiner has to explore for information that might be relevant to the investigation. Most examinations are performed with tools that can extract user files from the disk image, search for files that contain a specific word or phrase in a variety of human languages, and even detect the presence of encrypted data. Relevant data are then extracted from the preserved system so they are easier to analyze.

Testable Results

Just two decades ago, there were no digital forensics tools as we know them today. Instead, practitioners had to repurpose tools that had been developed elsewhere. For example, disk backup software was used for collection and preservation, and data recovery tools were used for media analysis. Although these approaches worked, they lacked control, repeatability, and known error rates.

The situation began to change in 1993. That year, the U.S. Supreme Court held in the case of *Daubert v. Merrell Pharmaceutical* that any scientific testimony presented in court must be based on a theory that is testable, that has been scrutinized and found favorable by the scientific community, that has a known or potential error rate, and that is generally accepted. Although the case didn't directly kick off the demand for digital forensics tools, it gave practitioners grounds for arguing that validated tools were needed not just for good science and procedure but as a matter of law. Since then there has been a steady development of techniques for what has come to be called *technical exploitation*.

Probably the single most transformative technical innovation in the field has been the introduction of *hash functions*, first as a means for ensuring the integrity of forensic data, and later as a way to recognize specific files.

In computer science a hash function maps a sequence of characters (called a *string)* to a binary number of a specific size—that is, a fixed number of bits. A 16-bit hash function can produce 2^{16} = 65,536 different values, whereas a 32-bit hash function can produce 2^{32} = 4,294,967,296 possible values. Hash functions are designed so that changing a single character in the input results in a completely different output. Although many different strings will have the same hash value—something called a *hash collision*—the more bits in the hash, the smaller the chance of such an outcome. (The name *hash* comes from the way hash functions are typically implemented as a two-step process that first chops and then mixes the data, much the same way one might make hash in the kitchen.)

Colors on a web page are indicated using hexadecimal representation, a base 16 system in which the numerals for 10 to 15 represented by A to F so they are not confused with single-digit numbers). Such computer notation is also commonly of use in digital forensics. The number of digits in a hexadecimal code indicates the power to which the base 16 is raised, so for example, 3BF9 means (3 x 16^3) + (11 x 16^2) + (15 x 16^1) + (9 x 16^0) or 15,353. Hexadecimal represents the numbers 0 to 255 as the codes 00 through FF (15 × 16 + 15 = 255). JPEG files generated by digital cameras typically begin with the sequence FF D8 FF E0 and end with FF D9. (Image courtesy of VisiBone; colors in this reproduction are approximated.)

Each system uses a different kind of internal memory structure, so programs designed to analyze one may not work on another.

Hashing was invented by Hans Peter Luhn and first described in a 1953 IBM technical memo; it's been widely used for computerized text processing since the 1960s. For example, because

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

Fragments of files that have been compressed can still be recovered, even when significant reassembly data are missing. One approach creates a model of the ways that a document might be decompressed based on the underlying mathematics, then chooses between the options, using a human language model. (Text courtesy of Ralf Brown, Carnegie Mellon University.)

every sentence in a document can be treated as a string, hashing makes it possible to rapidly see if the same paragraph ever repeats in a long document: Just compute the hash value for each paragraph, put all of the hashes into a list, sort the list, and see if any number occurs two or more times.

If there is no repeat, then no paragraph is duplicated. If a number does repeat, then it's necessary to look at the corresponding paragraphs to determine whether the text really does appear twice, or the duplicate is a the result of a hash collision. Using hashes in this manner is quicker than working directly with the paragraphs because it is much faster for computers to compare numbers than sequences of words—even when you account for the time to perform the hashing.

In 1979, Ralph Merkle, then a Stanford University doctoral student, invented a way to use hashing for computer security. Merkle's idea was to use a hash function that produced more than 100 bits of output and additionally had the property of being *one-way*. That is, it was relatively easy to compute the hash of a string, but it was nearly impossible, given a hash, to find a corresponding string. The essence of Merkle's idea was to use a document's 100-bit one-way hash as a stand-in for the document itself. Instead of digitally certifying a 50-page document, for example, the document could be reduced to a 100-bit hash, which could then be certified. Because there are so many different possible hash values (2^{100} is about 10^{30} combinations), Merkle reasoned that an attacker could not take the digital signature from one document and use it to certify a second document—because to do so would require that both documents had the same hash value.

Merkle got his degree, and today digital signatures applied to hashes are the basis of many cybersecurity systems. They protect credit card numbers sent over the Internet, certify the authenticity and integrity of code run on iPhones, and validate keys used play digital music.

The idea of hashing has been applied to other areas as well—in particular, forensics. One of the field's first and continuing uses of hashing was to establish *chain of custody* for forensic data. Instead of hashing a document or a file, the hash function is applied to the entire disk image. Many law enforcement organizations will create two disk images of a drive and then compute the hash of each image. If the values match, then the copies are assumed to each be a true copy of the data that were on the drive. Any investigator with a later copy of the data can calculate the hash and see if it matches the original reported value. Hashing is so important that many digital forensics tools automatically perform this comparison.

A second use for hashing is to identify specific files. This approach takes advantage of the property that it is extraordinarily unlikely for two files to have the same hash value, so they can label files in much the same way a person can be recognized by their fingerprints.

Today forensic practitioners distribute databases containing file hashes. These data sets can be used to identify *known goods*, such as programs distributed as part of operating systems, or *known bads*, such as computer viruses, stolen docu-

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

ments, or child pornography. Recently, several groups, including my team at the Naval Postgraduate School, have applied cryptographic hashing to blocks of data smaller than files, taking advantage of the fact that even relatively short 512-byte and 4,096-byte segments of files can be highly identifying.

File and sector identification with hashing means that a hard drive containing millions of files can be automatically searched against a database containing the hashes of hundreds of millions of files in a relatively short amount of time, perhaps just a few hours. The search can be done without any human intervention.

Finding Lost Files

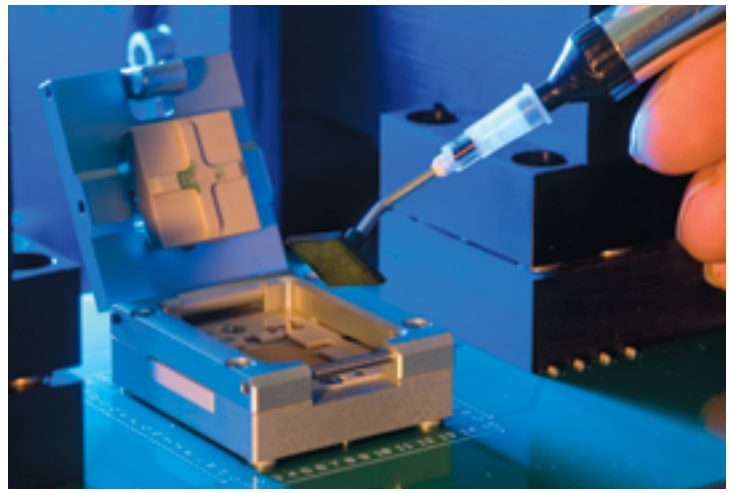
Many forensic investigations start with the examiner looking for files belonging to the computer's previous user.

Allocated files are ones that can be viewed through the file system and whose contents under normal circumstances will not be inadvertently overwritten by the operating system. The word *allocated* refers to the disk sectors in which the file's content is stored, which are dedicated to this particular file and cannot be assigned to others. Many digital forensics tools allow the examiner to see allocated files present in a disk image without having to use the computer's native operating system, which maintains forensic integrity of the evidence.

One of the major technical digital forensics innovations of the past 15 years has been approaches for recovering a file after it is deleted. These files are not simply in a computer's "trash can" or "recycle bin," but have been removed by emptying the trash. File names can be hidden, and the storage associated with the files is *deallocated*. But a file's contents sometimes can remain on the hard drive, in memory, or on external media, even though the metadata that could be used to locate it are lost. Recovering these kinds of data requires a technique called *file carving*, invented around 1999 by independent security researcher Dan Farmer, and now widely used.

The first file carvers took advantage of the fact that many file types contain characteristic sequences of bytes at the beginning and end of each file. Such sequences are called *file headers* and *footers*. The file carver scans the disk image for these headers and footers. When ones are found, the two sequences of bytes and all of the data between them are saved in a new file.

Modern carvers can validate the data that they are carving (for example, to make sure that the bytes between the JPEG header and footer can be actually displayed as a digital photograph) and can even reassemble files that are broken into multiple pieces. Such *fragment recovery carving* is computationally challenging because the number of ways that fragments can be realigned; the result is a combinatorial explosion as the size of the media increases. Missing fragments further complicate the problem.



Forensics tools allow investigators to directly access memory chips removed from devices such as mobile phones, satellite navigation devices, car electronics, and USB flash drives. This technique can be used to recover data from devices that have been physically damaged or are password protected. (Image courtesy of the Netherlands Forensic Institute.)

Closely related to file carving is the problem of reconstructing compressed data. *Compression* is a technique that is widely used on computer systems to squeeze data so that it takes up less space. The technique exploits redundancy; for example, if asked to compress the character sequence "humble humbleness," a computer might replace the six characters of the second instance of "humble" with a pointer to the first occurrence. English text typically compresses to one-sixth its original size.

Text must be compressed with *lossless algorithms*, programs that faithfully restore the original text when the data are decompressed. However, photographs and videos are typically compressed with *lossy* systems that exploit deficiencies in the human perceptual system. For example, a few dozen pixels of slightly different colors might be replaced by a single rectangle of uniform hue. The resulting savings can be immense. Without compression an hour of full-screen video might require 99 gigabytes but with compression the same video might take up only 500 megabytes—roughly 1/200th the original size.

The primary challenge posed by compression is recovering data when the compressed file is corrupted or partially missing. Just five years ago such corruption frequently made it impossible to recover anything of use, but lately there have been dramatic advances in this area.

In 2009 Husrev Sencar of TOBB University of Economics and Technology in Turkey and Nasir Memon of the Polytechnic Institute of New York University developed an approach that can show a fragment of a JPEG digital photograph even if the beginning and end of the file are missing. In 2011 Ralf Brown of Carnegie Mellon University developed an approach for recovering data from fragments of files compressed with the common ZIP or DEFLATE algorithms, even when critical information needed for reassembly is missing. Brown's approach creates a model of the many different ways that a document might be decompressed based on the underlying mathematics of compression, and then chooses between the different possible documents based on a second model of the human language in which the document is written *(see figure on page 374).*

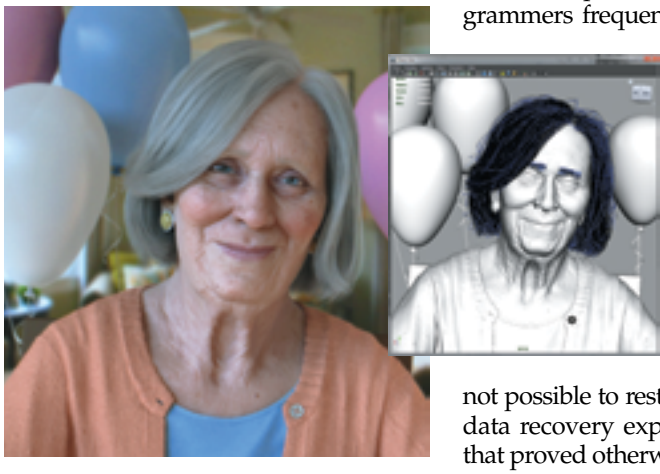
Recovering files in temporary computer memory can also be illuminating for digital evidence. The RAM of a desktop, laptop, or cell phone is a mosaic of 4,096-byte blocks that variously contain running program code, remnants of programs that recently ran and have closed, portions of the operating system, fragments of what was sent and received over the network, pieces of windows displayed on the screen, the copy-and-paste buffer, and other kinds of information. Memory changes

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

rapidly—typical memory systems support several billion changes per second—so it is nearly impossible to make a copy that is internally consistent without halting the machine. An added complication is that the very specific manner by which programs store information in memory is rarely documented and changes between one version of a program and another. As a result, each version may need to be painstakingly reverse-engineered by computer forensics researchers. Thus, memory analysis is time consuming, very difficult, and necessarily incomplete.

A Faraday cage shields a cell phone from any outside radio signals, minimizing possible changes to the phone's memory that might come from its connecting to the cell phone network; synchronizing data with the Internet; or receiving SMS messages, phone calls, or even a remote erase command. The analyst would normally keep the box closed for protection and work on the phone by using the internal camera. (Image courtesy of the Netherlands Forensic Institute.)

Despite these challenges, recent years have seen the development of techniques for acquiring and analyzing the contents of a running computer system, a process called *memory parsing*. Today there are open-source and proprietary tools that can report the system time when a memory dump was captured, display a list of running processes, and even show the contents of the computer's clipboard and screen. Such tools are widely used for reverse-engineering *malware*, such as computer viruses and worms, as well as understanding an attacker's actions in computer intrusion cases. Memory parsing can be combined with file carving to recover digital photographs and video.



Digital artwork can be sufficiently lifelike that analytical techniques are needed to distinguish real photographs from fake ones. The image above was synthesized using the model shown at right. (Image courtesy of Dan Roarty.)

Reverse engineering is another important part of digital forensics because software and hardware developers generally do not provide the public with details of how their systems work. As a result, considerable effort is needed to backtrack through systems code and understand how data are stored. Today's techniques to extract allocated files from disk images were largely developed through this method.

System analysis is the second leg of forensic research. It's similar to reverse engineering, but the fundamental difference is that the information the analyst seeks may be unknown to the developers themselves. Although this idea may seem strange, remember that computers are complicated systems: Just as programmers frequently put bugs in their code without realizing it, programs invariably have other behaviors that aren't bugs but are equally unforeseen by the original creators. Many system users and quite a few developers assumed that it was not possible to restore deleted files until data recovery experts developed tools that proved otherwise.

Image Integrity

Even when photos and video can be recovered from a subject's computer or cell phone, another question to consider is whether the imagery is real. Photographs were doctored long before the advent of Photoshop. For example, in the era of Soviet Russia, after he was purged by Stalin, Bolshevik official Avel Enukidze was carefully removed from official photographs through a series of skillful manipulations of light and negatives in a Kremlin darkroom. Computer animation now takes such manipulation to a completely new level, with synthesized scenes that are visually indistinguishable from recorded reality.

Image processing advances have made it possible to find some kinds of artifacts that indicate tampering or wholesale synthesis. Light reflections, highlights, and shadows also can be closely examined to reveal that different objects in a single "photograph" actually were assembled from images that were in slightly different physical environments.

In one dramatic demonstration in 2009, computer scientist Hany Farid of Dartmouth College showed that a single "group picture" had been created by pasting in people from different photos because the reflection of the room lights on each person's eyes were inconsistent with were they were standing in the frame.

At the leading edge of digital forensics research are systems that attempt to assist an analyst's reasoning—to find evidence automatically that is out of the ordinary, strange, or inconsistent. Such details can indicate that there is a deeper, hidden story. Inconsistencies can also indicate that evidence has been tampered with or entirely falsified. Ultimately, such automated reasoning systems are likely the only way that today's analysts will be able to keep up with the vast quantities and increasing diversity of data in the coming years. Progress in this area remains tricky, however. Some developments have been made on systems that can find timestamp inconsistencies (a file can't be deleted before it is created), but such rules are invariably complicated by the messiness of the real world (for example, daylight savings time).

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.

Forging Ahead

For all its power, digital forensics faces stark challenges that are likely to grow in the coming years. Today's computers have on average 1,000 times more storage but are only 100 times faster than the high-end workstations of the early 1990s, so there is less computing power available to process each byte of memory.

The number of cases in which digital evidence is collected is rising far faster than the number of forensic investigators available to do the examinations. And police now realize that digital evidence can be used to solve crimes—that is, as part of the investigation process—whereas in the past it was mainly a tool for assisting in convictions.

Cell phones may be equipped with "self-destruct" applications that wipe their data if they receive a particular text, so it is now standard practice to store phones in a shielded metal box, called a *Faraday cage,* which blocks radio waves. But many cell phones will "forget" their stored memory if left off for too long, so the Faraday cages must be equipped with power strips and cell phone chargers. Because many low-end cell phones have proprietary plugs, police must seize chargers as well. However, some phones will wipe their data if they can't call home, whereas others will encrypt their data with algorithms too powerful for law enforcement to decipher.

Further complicating the investigator's job is the emergence of cloud computing and other technologies for storing data on the Internet. As a result of the cloud, there is no way to ensure that a seized cell phone actually holds the suspect's data—the phone might simply be a tool for accessing a remote server. A law enforcement professional who is authorized to search a device may not have legal authority to use information on that device to access remotely stored data. Worse still, the data might be deleted in the meantime by one of the suspect's collaborators.

Despite its technical sophistication and reliance on the minutiae of digital systems, the single biggest challenge facing digital forensics practice today has a decidedly human dimension: the lack of qualified people to serve as researchers and practitioners. Not merely the result of the general tech shortage, the very nature of digital forensics makes staffing significantly harder than in other disciplines. Because the field's mission is to understand any data that might be stored, we need individuals who have knowledge of both current and past computer systems, applications, and data formats. We need generalists in a technological society that increasingly rewards experts and specialization.

One way to address this training problem is to look for opportunities to break down forensic problems into modular pieces so that experts in related fields can make meaningful contributions. I believe that another approach is to show how the underlying principles and current tools of digital forensics can be widely applied throughout our society. This relevancy should increase research into tools and hopefully expand the user base of the software.

Many system users and developers assumed that it was not possible to restore deleted files until data recovery experts developed tools that proved otherwise.

Many of the tools of digital forensics can be used for privacy auditing. Instead of finding personal information that might be relevant to a case, businesses and individuals can use the tools to look for the inappropriate presence of personal information left behind because of bugs or oversight. Likewise, individuals can use programs such as file carvers to recover photographs that have been accidentally deleted from digital cameras.

More generally, as our cars, street signs, communication systems, electrical networks, buildings, and even social interactions are increasingly computerized, digital forensics is likely to be one of the only ways of understanding these systems when they misbehave—or when they are subverted.

Without developing fundamentally new tools and capabilities, forensics experts will face increasing difficulty and cost along with ever-expanding data size and system complexity. Thus today's digital detectives are in an arms race not just with criminals, but also with the developers of tomorrow's computer systems.

Bibliography

Brown, R. 2011. Reconstructing corrupt DEFLATEd files. *Digital Investigation* 8:S125–S131.

Farid, H. 2009. Digital doctoring: Can we trust photographs? In *Deception: From Ancient Empires to Internet Dating,* ed. B. Harrington. Stanford, CA: Stanford University Press, 95–107.

Garfinkel, S. 2011. Every last byte. *Journal of Digital Forensics, Security and Law* 6(2):7–8.

Garfinkel, S., A. Nelson, D. White, and V. Roussev. 2010. Using purpose-built functions and block hashes to enable small block and sub-file forensics. *Digital Investigation* 7:S13–S23.

Granetto, P. J. 2009. *Sanitization and Disposal of Excess Information Technology Equipment.* Technical Report D-2009-104 of the Inspector General of the United States Department of Defense. http://www.dodig.mil/Audit/reports/fy09/09-104.pdf

King, D. 1997. *The Commissar Vanishes: The Falsification of Photographs and Art in Stalin's Russia.* New York: Metropolitan Books.

Pal, A., H. T. Sencar, and N. Memon. 2008. Detecting file fragmentation point using sequential hypothesis testing. *Digital Investigation* 5: S2–S13.

Pollitt, M. M. 2007. An ad hoc review of digital forensic models. In *Second International Workshop on Systematic Approaches to Digital Forensic Engineering*, 43–54. Los Alamitos, CA: IEEE Computer Society.

Reith, M., C. Carr, and G. Gunsch. 2002. An examination of digital forensic models. *International Journal of Digital Evidence* 1(3):10–22.

Sencar, H., and N. Memon. 2009. Identification and recovery of JPEG files with missing fragments. *Digital Investigation* 6:S88–S98.

Walls, R. J., B. N. Levine, M. Liberatore, and C. Shields. 2011. Effective digital forensics research is investigator-centric. *Proceedings of the Sixth USENIX Workshop on Hot Topics in Security,* 1–7. https://www.usenix.org/conference/hotsec11/effective-digital-forensics-research-investigator-centric

Walters, A., and N. Petroni. Feb. 2007. Volatools: Integrating volatile memory forensics into the digital investigation process. In *Black Hat DC 2007 Proceedings,* 1–18. http://www.blackhat.com/presentations/bh-dc-07/Walters/Paper/bh-dc-07-Walters-WP.pdf

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/issues/id.104/past.aspx

© 2013 Sigma Xi, The Scientific Research Society. Reproduction with permission only. Contact perms@amsci.org.