;login:
THE MAGAZINE OF USENIX & SAGE
August 2002 volume 27 • number 4

inside:

THE LAW
KENNEALLY: COMPUTER FORENSICS

USENIX & SAGE
The Advanced Computing Systems Association &
The System Administrators Guild

computer forensics

by Erin Kenneally

Erin Kenneally is a Forensic Analyst with the Pacific Institute for Computer Security (PICS), San Diego Supercomputer Center. She is a licensed Attorney who holds Juris Doctorate and Master of Forensic Sciences degrees.

erin@sdsc.edu

Beyond the Buzzword

What do the Chandra Levy disappearance, Enron/Arthur Anderson collapse, Danielle Van Dam murder case, Microsoft antitrust trial, former President Clinton sex scandal, and tracking of al Qaeda terrorists all have in common? In each instance, computer forensics figured prominently in investigating the questions at hand. Simply put, computer forensics has reached prime time. It is no longer the stuff of back-office geeks and techno-wizards but has been embraced by both law enforcement and the private sector as a technique to reconstruct crimes, conduct digital discovery, and, generally, uncover the electronic traces that help prove or disprove accounts of historical events.

The negative corollary to the consciousness-raising effect of being in the limelight is that "computer forensics" has become something of a buzzword among profit-savvy businesses seeking to market their "advanced capabilities." This has resulted in a dilution and misrepresentation of the evolving discipline of computer forensics. At the risk of having no agenda save for battling ignorance, this article is meant as a primer on the essence of computer forensics so that one can better appreciate and expect accurate, reliable, and scientifically based standards when encountering digital evidence issues.

Computer Forensics Defined

Odontology, structural engineering, pathology, serology, or analysis of computer systems are all methods used in forensic science. Since forensic science is the application of a scientific discipline to the law, the essence of all forensic disciplines concerns the principles applied to the detection, collection, preservation, and analysis of evidence to ensure its admissibility in legal proceedings. Computer forensics refers to the tools and techniques to recover, preserve, and examine data stored or transmitted in binary form. The application of forensic techniques to digital analysis, therefore, can be viewed as the new kid on the block more commonly populated by the likes of DNA fingerprinting and hair and fiber analysis. And instead of Quincy, ME, examining a corpse to determine cause of death, we're dealing with digital examiners conducting machine autopsies to recover evidence of a crime.

Analogizing Computer Forensics to Traditional Forensic Sciences

The fundamental principles of computer forensics are the same as that of traditional forensic disciplines. All start with intense variability among a large number of attributes and advances are aimed at enhancing the identifying, characterizing and correlative properties of the evidence source. Whereas an MD-5 hash may identify a digital document to the exclusion of all others, the remnants of a deleted Netbus application found in unallocated space may help correlate a suspect to victim's firewall log data of scans on port 12345 coming from the suspect's IP address.

Forensic techniques are designed to uncover these identifying, characterizing and correlative properties more precisely, more accurately, faster and with less evidence.

For instance, a comparison of analysis development between digital data versus biological data (blood) would illustrate how A/B/O typing gave way to Rh factors, which was supplanted by DNA typing via RFLP (restriction fragment length polymorphism)

and PCR (polymerase chain reaction) – which resulted in the same evidence source being used to characterize and then positively identify persons to the exclusion of all others. Similarly, forensic analysis techniques for digital evidence has yielded hash libraries (to identify data files), file signatures (to characterize files by matching filename and file type), and mirror imaging software (to copy larger amounts of evidence without altering the original evidence).

Regardless of whether the discipline is computer forensics or fingerprinting, the driving question is not whether evidence exists but, rather, can investigators uncover and contextualize the evidence. Therefore, the challenges are: Where to look? What techniques will make the evidence apparent? And is the evidence admissible?

Regardless of whether the discipline is computer forensics or fingerprinting, the driving question is not whether evidence exists but, rather, can investigators uncover and contextualize the evidence.

Just as a pathologist may deduce by observing the lack of water in a person's lungs that he was already dead when his car sank to the depths of a lake, computer forensic examiners may analyze file modification/access/creation times to determine if intellectual property was transferred after an employee was fired. And in the same way that sources of biological evidence may be blood, saliva or hair shafts found on clothing, cigarette butts, and weapons, digital evidence can be found on any number of media sources (hard drive, floppy disk, CD-ROM, PDA) and in locations such as print spooler files, hidden partitions, registries, system logs, bad clusters, and/or metafiles. In the biological realm, techniques such as PCR, RFLP, and STR (short tandem repeats) exist to identify DNA in a drop of dried blood that is not visible to the naked eye. In computer forensics, techniques exist to recover deleted data; recover passwords; analyze file slack, unallocated space, and swap files; reconstruct user and application activity on a system; and search email for source and content information.

Finally, in terms of admissibility hurdles, the technology to recover deleted data has been accepted, but what is contested is the inclusiveness of the software that undertakes to recover it – in other words, are there measurable error rates for the software that address the likelihood of missing potentially exculpatory evidence? Likewise, insofar as DNA fingerprinting technology has been accepted in the courtroom, certain techniques (like STR) remain open to challenge.

Contrasting Digital Evidence with Physical Evidence

Despite core similarities, the differences between computer forensic analysis and the more traditional forensic sciences bear reflection. From a historical perspective, computer forensics is a burgeoning discipline compared to traditional forensic sciences, many of which are rooted as far back as the early 20th century. One prominent difference lies in the diametric evolution of computer forensics as compared to traditional forensic sciences. Computer forensics originated in "cop shops" rather than clinical laboratory settings. Electronic tools and techniques have been developed to solve specific problems on known platforms within given parameters, rather than the more traditional application of scientific rigor and controlled testing to derive facts for crime solving to investigations for legal proof[1]. DNA analysis, for instance, was developed for non-forensic purposes and was only later applied for judicial purposes, unlike the forensic analysis software employed by digital technicians today.

To be sure, a grounding in scientific rigor is increasingly being recognized and applied to computer forensic tools and techniques to ensure the reliability and admissibility of digital evidence. However, unlike physical evidence, digital evidence poses novel challenges to computer forensic analysis. The mutable, fleeting, and intangible nature of digital evidence stands in contrast to persistent physical features used in other disci-

1. See generally, David Goodstein, "How Science Works" *Reference Manual on Scientific Evidence* 2nd Edition, Federal Judicial Center (2000) *http://air.fjc.gov/public/fjcweb.nsf/pages/74*

plines – i.e., ridge patterns for fingerprinting, polymarkers for DNA analysis, and bone characteristics for forensic anthropology. This fosters the advantage of conducting comparisons with known exemplars to uncover those identifying, characterizing, or correlative properties. However, the variables involved in complex computer network activity and software/hardware that produces digital evidence are dynamic and not as conducive to reproduction.[2]

Digital Evidence – Search and Seizure Challenges

Digital evidence has shifted paradigms in collecting, preserving, and analyzing evidence, as illustrated by the unique legal challenges facing computer forensic professionals. Specifically, these shifting paradigms can be appreciated by understanding the resource challenges, attempting to define "reasonableness," and paying heed to modification challenges presented by digital evidence.

RESOURCE ISSUES

The traditional approach when investigators would encounter a crime scene with a computer was to seize everything. That approach may have worked at a time when the ratio of computers to employees was 1:1 or greater, or when there was a stand-alone computer at a domestic crime scene. However, this approach is no longer feasible in a society where computers and their appendages dominate the landscape and information is increasingly being stored, transmitted, and created in digital form. Insofar as our ability to collect far outweighs our ability to analyze, the "seize everything" mentality is simply economically infeasible, both for budget-constrained public law enforcement as well as for private sector responders whose work is not part of the corporate profit center. Indeed, the cost of storage media has declined appreciably, yet the man-hour resources and capabilities to image and cull through hundreds of gigabytes worth of data on a compromised network is no small task. Relief does not appear imminent, as technologies such as FMD-ROMs, which store 140GB, may soon supplant CD and DVD media, and consumer grade hard drives are shipping at 75GB and up.

To put this resource challenge into context, imagine that a 1.44MB floppy disk holds the equivalent of a novel. Now, a standard 20GB home computer would produce the paper equivalent of a stack of books roughly as high as a fifteen-story building. Placed in the context of paper-based evidence, it is easier to appreciate how the nature of digital information and the relevant data contained therein strains the resources of forensic professionals who must uncover, collect, preserve, and analyze these electronic haystacks.

DEFINING REASONABLENESS

A fundamental right guaranteed by the Constitution is protection from unreasonable searches and seizures by the government. Courts have applied this protection by ensuring that search warrants are only issued upon a showing of probable cause, which is grounded in "reasonableness" and defined in terms of narrowness and particularity of scope. However, the time and scope variables (narrowness and particularity) that affect the reasonableness of the search and seizure take on a different dynamic. For instance, judges oftentimes authorize a search warrant with narrow time limits to minimize business disruption. In doing so, there is a faulty assumption that the scope of the search will be narrowed by decreasing the time allotted to conduct the search and seizure. In actuality, a narrow time frame will result in a wider scope of data seizure – thus increasing the chances of capturing irrelevant and overbroad data.

2. See Randolph Johnkait, "Forensic Science: The Need for Regulation," *Harv.J.L. & Tech.*, vol. 4, no. 109 (1991), 133–34.

Furthermore, search warrants often authorize authorities to search anywhere that the evidence in question can "reasonably" exist. So, a warrant for a gun would preclude investigators looking in a cell phone case. Digital evidence, however, is not bound by those same physical limits, so notions of what is reasonable must be put into a new context. For instance, large amounts of data can be hidden or compressed in a very small area, and file extension labels do not necessarily reflect the underlying data type (i.e., a strategic diagram in the form of a .jpg can be named "anything.txt").

EVIDENCE MODIFICATION CHALLENGES

Finally, the mutability of digital evidence facilitates legal challenges grounded in chain-of-custody and evidence-tampering arguments. Whereas DNA analysis is performed on the original blood evidence, maintaining the sanctity of original evidence is a tenet of computer forensics, and analysis must be conducted on a copy of the original media (with a few, notable exceptions where circumstances preclude a copy being made). Perhaps because the 33rd copy of a file is indistinguishable from the original and it is trivial to change bytes without leaving a trace, benign actions when handling digital evidence may have probative consequences upon which guilt or innocence hinges. Setting aside the wholesale substitution of blood evidence, if a serologist contaminates a blood sample, there is little risk that the DNA of another person will be created. Rather, the blood sample will not conclusively identify the culprit. With digital evidence, for example, merely turning on a Win95 system opens roughly 8% of the files on the hard drive just to boot the system. The consequence is that 417 access dates, some of which may have been crucial to proving guilt or innocence, have been altered .

So what? The criticality of timestamp data associated with file modification, access, or creation can make or break a case. For example, take the case where the digital evidence found on a defendant's computer was a large collection of adult porn (legal) and a smattering of kiddie porn images (illegal). Now, the defendant may claim that he downloads adult porn via IRC, and the kiddie porn must have been unintentionally downloaded at the same time, unbeknownst to the defendant. If this were true, computer forensic analysis might reveal access dates on the adult images well after the creation dates (initial download), but the child images had creation and access times that matched the creation times for the adult pictures. This would support the defense that he did not view or distribute the child porn. However, if the seizing officer booted the suspect machine and started rifling through the images, he would have changed the timestamps and quashed potentially exculpatory information.

The criticality of timestamp data associated with file modification, access, or creation can make or break a case.

Conclusion

In our increasingly electronic society, digital evidence promises to continue to permeate crime scenes and civil disputes, thus rendering computer forensics an increasingly vital discipline in the resolution of disputes. The danger is that computer forensics will be driven by industry and market forces that lose sight of the need for scientific underpinnings regarding computer forensic tools and techniques. Hopefully, this primer has served to raise awareness about the similarity in principles between computer forensics and the traditional forensic sciences, as well as highlighting the unique nature of digital evidence, so that the collection, preservation, and analysis of digital evidence will advance the search for truth.