Challenges to Digital Forensic Evidence

2nd Ed.

by Fred Cohen, Ph.D.

Table of Contents

1 Introduction and background.....5
A good background.....6
Questions.....7
2 Overview.....8
Basics.....8
Identifying Evidence.....11
Evidence Collection.....14
Seizure errors and related issues.....22
Transport of evidence.....27
Storage of evidence.....28
Evidence analysis.....30
Overall summary.....44
Questions.....45
3 Mechanics of writing expert rebuttals.....47
An outline of a report.....47
In the first person.....49
Time is of the essence.....51
We all make mistakes – don't compound them.....52
Hedge in an honest way.....53
Present the basis with the conclusions.....55
Do simple experiments to disprove things.....56
The report is the draft I am working on.....58
Use analogies within limits.....60
Why should the standard be lower?.....61
When in doubt, check it out.....63
Questions.....64
4 Case studies.....65
Don't forget the effects of time.....67
The WayBack Machine and friends.....70
Calibrating your own tools.....71
Issues related to link analysis.....73
Internet operations between companies.....78
Malicious exploitation of computers.....86
The single actor theory.....89
An early case involving audit trails.....91

The case of the video of the case.................................... 102
Mock trial and the MD5.. 105
The virus ate my case.. 106
Spoliation of evidence and chain of custody.....................107
Questions..110
5 Testifying.. 111
Prepare.. 112
Tell the simple truth as you know it....................................113
Don't stretch it - be open...114
Take your time and think...115
Questions...116
6 How to avoid being challenged..117
Questions...119
Endnotes.. 120

Front matter

Challenges to Digital Forensic Evidence - 2nd Ed.
is
Copyright © 2007-10 by Fred Cohen - All Rights Reserved.
ISBN # 1-878109-41-3
Published by Fred Cohen & Associates out of Livermore, CA.
2010-04-01

We are not liable for anything you do with this book or use this book for, including but not limited to choking people with its pages, using the book to trip folks, the kid leaving it on the steps, the pollution it creates when you burn it, or any side effects you may have in the afterlife from having sworn about certain things I may have said. All of these things as well as any other thing that happens as either a direct or indirect consequence of my having written this book is not my fault, and I take no responsibility and MAKE NO WARRANTY EITHER EXPRESSED OR IMPLIED ABOUT ANYTHING IN THIS BOOK. The law says I have to use large block letters to say that so you are amply warned.

If you are offended by anything I have written, or if you feel your children or parents have been corrupted by it, remember, this warning was given to you and you could have looked at it before reading the book or buying it.

WARNING! READ THIS PAGE BEFORE BUYING THE BOOK!

That should do it I think...

1 Introduction and background

This book is about digital forensic evidence and how it can be effectively challenged in legal settings. Many may complain about the notion that anyone should write a book about challenges to evidence because, they might claim, it only supports those who are trying to escape the law and their just punishment. To this I say, those who put this position forth are not supporting the law or justice, they are only supporting their own position.

I wrote this book because people tend to believe whatever a computer tells them and because I know that computers are not oracles, but just tools. I have a lifetime of experience with the way computers fail and the way people who trust computers make mistakes. I spend untold hours fixing problems created by the misuse of computers and a lack of care in their use, and I know just how easy it is to lie with a computer or make it appear to say something that is just plain untrue. I have spent years dealing with computer-based deceptions, including both generating these deceptions and countering them.

I also wrote this book because I believe that in order to have a just society, we need to have an honest and carefully considered scientific basis for things we claim to be facts and use to decide who lives, who dies, who spends years in jail, and who goes free.

Digital systems are increasingly fundamental to the prosecution of crimes and civil actions, and more and more of the evidence is based solely on digital systems and interpretation by so-called experts, many of whom have very little expertise. Yet we have little precedent and even less scientific basis for most of the computer-related evidence produced in legal proceedings.

This book is my best hope for changing the landscape and creating the conditions required for more serious study of the issues of digital forensic evidence, the creation of precedent that makes sense and reflects the scientific realities of our times, and ultimately for bringing about a more just society for my children and their children, and your children to live in.

A good background

I teach students in graduate school about digital forensic evidence and how to deal with it, and many people ask me what sort of background they should have to take my courses and to go into this field. Of course I think that my personal background is particularly good for the work that I do, but it is asking quite a bit to have all of my students come to class with 30 years of experience, a Ph.D. in electrical engineering with a computer engineering emphasis, a lot of computer programming experience, simulation and modeling experience, large-scale network architecture design and analysis, having authored hundreds of peer reviewed articles, reviewed many papers, and so forth. So if you have that kind of background and you want to attend my classes, I would be glad to have you. But most students are just learning, and for the most part, they have none of this experience. So what makes a good background?

A good background should include some amount of computer programming, at least hobby level experience with fixing digital systems, putting computers together and taking them apart, building your own local area networks, running some computers for others, working in environments similar to the matter at hand, and above all, a healthy skepticism about what computers do and what people using computers use those computers to do.

Added bonus background includes strong mathematical and analytical skills, the ability to read technical reports quickly and spot inconsistencies and fallacies, a knowledge of the history of the sorts of systems at issue, and experience in the particular issues involved in the particular case but not with the particular case itself.

Also, there is no such thing as a generic expert in digital forensics. The field is simply too broad for anyone to know everything there is to know about all of the different facets of all of the different cases that might arise. But there are generic principles associated with challenging digital forensic evidence, and that is what this book is about.

As a good rule of thumb, when trying to become an expert, it may be best to start by becoming a specialist in a particular subfield – like evidence handling or fraud investigation. [21]

Questions

1. What background do you have that makes you suited to do digital forensics?
2. What background do you hope to have by the end of this book that will better qualify you for doing digital forensics?
3. In working for a prosecutor, is your job to find evidence for the defense and help them understand it? If not, how is this a job associated with seeking truth and justice?
4. In working for a defendant, is your job to find evidence that would tend to support the prosecution? If not, how is this a job associated with seeking truth and justice?
5. Write a resume of your current qualifications for doing digital forensics, with no exaggeration whatsoever, and with at most a paragraph on each item included. How do you think you would stack up against a professional police investigator with 20 years of experience and training in digital forensics?
6. Describe any special expertise you have and how that expertise would allow you to be better as a digital forensics expert in a particular area of concentration than others without similar background.
7. Do you subscribe to a particular code of professional ethics? If so, what is it? If not, why not?
8. Examine codes of professional ethics from professional societies and identify how they relate to the notions of working to challenge digital forensic evidence.
9. Suppose you are working professionally in this field and you make a major mistake on a case that is now underway. What should you do to notify whom of the mistake you made, or should you just let it go?

2 Overview

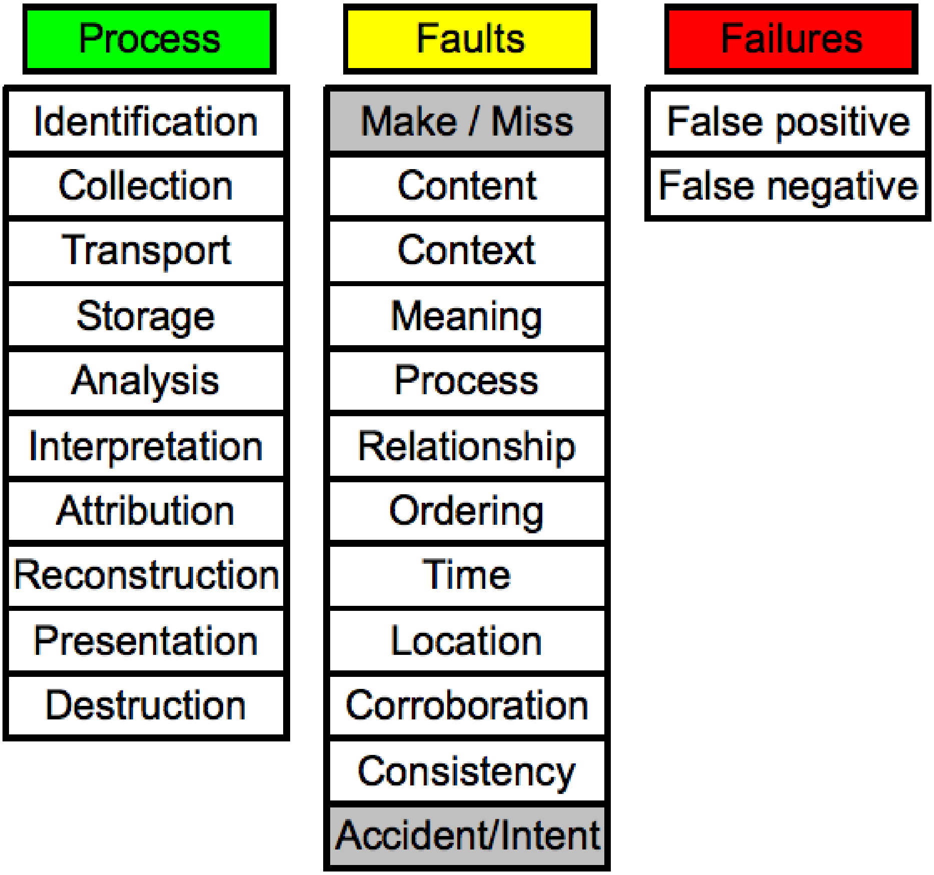

Figure 1 - Challenges overview

Basics

Digital forensic evidence is identified, collected, transported, stored, analyzed, interpreted, attributed, reconstructed, presented, and destroyed through a set of processes. Challenges to this evidence come through challenges to the elements of this process. This process, like all other processes and the people and systems that carry them out, is imperfect. That means that there are certain types of faults that occur in these processes.

Faults and Failures

Faults consist of intentional or accidental making or missing of content, contextual information, the meaning of content, process elements, relationships, ordering, timing, location, corroborating content, consistencies, and inconsistencies.[1]

Not all faults produce failures, but some do. While it may be possible to challenge faults, this generally does not work and is unethical if there is no corresponding failure in the process.

Certain things turn faults into failures, and it is these failures that legitimately should be and can be challenged in legal matters. Failures consist of false positives and false negatives. False negatives are items that should have been found and dealt with in the process but were not, while false positives are things that should have been discarded or discredited in the process but were not.

Legal Issues

In the United States, at the Federal level, evidence in legal cases is admitted or not based on the relative weights of its probative and prejudicial value. Other standards apply in different jurisdictions, but this standard is fairly common worldwide. Probative value is the extent to which the evidence leads to deeper understanding of the issues in the case. Prejudicial value is the extent to which it leads the finder of fact to believe one thing or another about the matter at hand. If the increased understanding from the evidence is greater than the increase in belief, the evidence is admissible. [2]

Part of the issue of probative value is the quality of the evidence. If the process that created the evidence as presented is flawed, this reduces the probative value. Impure evidence, evidence presented by an expert who is shown to lack expertise in the subject at hand, evidence that has not been retained in a proper chain of custody, evidence that fails to take into account the context, or evidence falling under any of the other fault categories described in Figure 1, all lead to reduced probative value. If the result of these faults produces wrong answers, the probative value goes to nearly zero in many cases.

The Latent and Hearsay Nature of the Evidence

In order to deal with digital evidence, it must be presented in court. Since digital data is not directly observable by the finder of fact, it must be presented through expert witnesses using tools to reveal its existence, content, and meaning to the fact finders. This puts it into the category of latent evidence. Thus, it depends on the quality and unbiased opinion of the experts for each side.

In addition, digital evidence is hearsay evidence, in that it is presented by an expert who asserts facts or conclusions based on what the computer recorded, not what they themselves have directly observed. In order for hearsay evidence to be admitted, it normally has to come in under the normal business records exemption to the hearsay evidence prohibition.

Notions Underlying "Good Practice"

One of the results of diverse approaches to collection and analysis of digital forensic evidence is that it became increasingly difficult to show why the process used in any particular case is reliable, trustworthy, and accurate. As a result, sets of 'good practices' were developed by law enforcement in the UK, US, and elsewhere. The use of the term 'good practices' is specifically designed to avoid the use of terms like 'standards' or 'best practices’. This is because of a desire to prevent challenges to evidence based on not following these practices.

The real situation is that there are no best practices or standards for what makes one approach to digital forensic evidence better or worse than another. In the end, what works is what counts. Since the law and the technology are not settled, many things may work in different situations and to choose one over another would only muddy the waters.

Throughout this chapter I will comment on “good practice”, how and why deviations occur, and their implications. It is important in challenging evidence to seek out deviations from good practice, but it is also important to seek out reasons that these deviations are meaningful in terms of the basis of the challenge.

The nature of legal systems and refuting challenges

In some legal systems, there are great rewards to those who challenge everything. The idea is to spread the seeds of doubt in the minds of the finders of fact. In presenting and characterizing evidence, care should be taken to not mischaracterize, over-characterize, or under-characterize the value and meaning of evidence.

There are valid and reasonable challenges to digital evidence and those challenges must be addressed by those presenting it, but in many cases, the challenges performed by court-recognized, but inadequately knowledgeable, experts are just plain wrong. In my experience, such challenges are easily refuted and should be refuted. While some things are indeed opinion, other things are not.

Refuting clearly invalid challenges is often straightforward. In most such cases, ground truth can be clearly shown. As an example, when claims are made that the presence of a file indicates something unrelated to that file, a combination of manufacturers' manuals and demonstrations readily destroy the credibility of the evidence and the person giving it. In one such case, a court-appointed special master made assertions that were wrong. The combination of documentation and demonstration showed this expert for what he was and made a compelling case.

Overview

The rest of this chapter will focus on identifying sources of faults that occur in and between elements of the forensic process and ways that those faults turn into failures. The failures are then used to challenge the process. Challenges can be couched in terms of the process, the fault, and the resulting failure and this makes for an effective presentation of the challenge.

Identifying Evidence

The first step in gathering evidence is identifying possible sources of evidence for collection. It is fairly common that identified evidence includes too little or too much information. If too much is identified, then court mandated search and seizure limitations may be exceeded. If too little is identified, exculpatory or inculpatory evidence may be missed. The most commonly missed evidence comes in the form of network logs from related network components.

Common Misses

There is a great deal of corroborating evidence that can be sought from connected systems that produce log files that can confirm or refute the use of a system by a suspect. If the evidence is not sought and the actions are in question, either in terms of taking place or in terms of their source, path, or content, the lack of intermediate audit trails may complicate the ability to definitively show what took place.

Other evidence that is commonly missed includes storage devices, networked computer contents, deleted file areas from disks, secondary storage, backups, and other similar information. Properly identifying information to be collected often fails because of missed relationships between computers and evidence in those computers. This evidence is often time sensitive and is lost if not identified and gathered within a short time frame.

Relevant information is often located in places not immediately evident from the original crime scene. In cases where evidence is stored for long periods and can be identified as missing in a timely fashion, the fault can usually be mitigated by additional collection. The time frame for much of this information is very limited, particularly in the case of server logs, connection logs, and similar network-related information. The chain of custody issues for such evidence can also be quite complex and involve a large number of participants from multiple jurisdictions.

Information not sought

In some cases evidence is not sought. For example, when one side or another looks for evidence in a case, they may decide to follow-up or not follow-up on different facets of the case, pursue or not pursue various lines of enquiry, or limit the level of detail or sort of evidence they collect. These represent intentional non-identification of evidence. On the other hand, there are plenty of good investigators who miss all sorts of evidence for one reason or another. Evidence is often concealed and not found by investigators. Sometimes it is stored somewhere the investigators are unaware of or cannot gain access to. Sometimes the evidence is destroyed or no longer exists by the time it becomes apparent that it might be of value. These sorts of faults occur in every case, but they rarely rise to the level of a failure causing a substantial error in a case. People do their best or focus their attention on what they think is important, and sometimes they miss things. Time and resources are limited, so certain lines are not always pursued. That's just how the world is.

False evidence

On the other hand, there are also cases, rare as they may be, when evidence is made up "from whole cloth". While this is increasingly difficult to do in all areas, such evidence in the digital arena is exceedingly rare. Indeed, it is very hard to make up digital evidence and have it survive expert challenges and I am aware of no case when this has been done. There are cases when the defense makes such a claim, and there are even cases when digital evidence has been found to not be adequately tied to the party involved, altered, or spoiled. But no cases I am aware of have been successfully challenged on the basis that the evidence was simply constructed. Every claim of construction of this sort that I am aware of to date has been successfully refuted.

Non-stored transient Information

Data that is not stored in a storage media cannot be seized; it can only be collected in real time by placing sensors into the real-time environment. Such evidence must be identified in a different manner than evidence sitting on a desk or within a disk. This sort of evidence must be identified by an intelligence process and special legal means must be applied in many cases to collect this evidence.

Good practice

The general plan for good practice is to find the computer(s) and/or other sources of content to be seized. To the extent that some source of evidence is not found, good practice is not followed. As a challenge, the sources of evidence not discovered may contain exculpatory content or other relevant material. It may seem obvious that anyone doing a search for digital evidence will try to find anything they can, but the technology of today leads to an enormous number of different devices that can be concealed in a wide variety of ways. Small digital cameras are commonly concealed in sprinkler systems, pictures, and similar places. A memory stick or flash RAM drive may contain gigabytes of information and be the size of a fingernail or smaller. It is hard to find every piece of digital evidence, and harder still if it is intentionally concealed.

It is good practice to seize the main system box, monitor, keyboard, mouse, leads and cables, power supplies, connectors, modems, floppy disks, DATs, tapes, Jazz and Zip disks and drives, CDs, DVDs, hard disks, manuals, software, papers, circuit boards, keys, USBs, printers, printouts, and printer paper. Seize mobile phones, pagers, organizers, PDAs, land line telephones, answering machines, audio tapes and recorders, digital cameras, PCMCIA cards, integrated circuits, credit cards, smart cards, facsimile machines, and dictating machines. All of these items may contain digital forensic evidence and may be useful in getting the system to operate again. A good rule of thumb is 'If in doubt, seize it.'

Evidence Collection

Most evidence is collected electronically. In other words, the process by which it is gathered is through the collection of electromagnetic signals. In order to trust evidence there needs to be some basis for the manner in which it was collected. For example, it would be important to establish that it comes from a particular system at which the user sits. This implies some sort of evidence of presence in front of the computer at a given time.

Establishing Presence

Records of activity are often used to establish presence. For example, users may have passwords that are used to authenticate their identity. These may be stored locally or remotely, and typically provide dates and times associated with the start of access, as well as with subsequent accesses. The verification process provides evidence of the presence of the individual at a time and place; however, such validations can be forged, stolen, and lent. In some environments common passwords and user IDs are used, making these identifications less reliable.

Chain of custody

Digital forensic evidence comes in a wide range of forms from a wide range of sources. For example, in a recent terrorism case a computer asserted to be from a defendant was provided to the FBI by someone who purchased the computer at a 'swap meet'. These are generally outdoor small vendor sales of used equipment of all sorts - from old guns to old electronic equipment - sold over folding tables and from the backs of cars. Some of it is stolen, some of it resold by people who bought new versions, some wholesale, some damaged goods, and some made by those who sell them. This computer is asserted to contain evidence, but establishing a chain of custody will be a very difficult proposition, considering that the defendant will likely claim to never have had such a computer.

How the evidence was created

The information that becomes evidence may be generated for various purposes, most of which are not for the purpose of presentation in court. While the business records exception to the hearsay rule applies to normal business records, many other sorts of records may not be allowed in, depending on how they are created, collected, maintained, and presented, and by whom. In most cases when information is gathered from systems as they operate, the systems under scrutiny are altered during the gathering process. While this does not necessarily taint the evidence, it provides an opportunity for tainting that should not be overlooked if there is a reason to believe that tainting may have taken place.

Typical Audit Trails

Typical audit trails include the date and time of creation, last use, and/or modification; identification information like program names, function performed, user names, owners, groups, and IP addresses, and other information, such as port numbers, protocol types, portions or all of the content, and protection settings. If this sort of information exists it should be consistent to a reasonable extent across different elements of the system under scrutiny.

Consistency of Evidence

For example, if a program is asserted to generate a file that was not otherwise altered, then the program must have been running at the time the file was created, must have had the necessary permissions to create the file, must have the capacity to create such a file in such a format, and must have been invoked by a user or the system using another program capable of invoking it. There is a lot of information that should all link together cleanly. If it doesn't, there are reasons to question it.

This is not to say that all of these records always exist in the proper order on all systems. For various reasons, some records get lost, others end up out of order, and times fluctuate to some extent; however, these are all within reasonable expected tolerances and substantial deviations are often detectable. Such deviations are indicators that things are not what they seem, and in such cases alternative explanations are available and should be pursued.

Proper Handling During Collection

In most police-driven investigations, standard evidence handling processes are used for digital forensic evidence, with a few enhancements and exceptions. Photographs and labels are commonly used and an inventory sheet is typically made of all seized evidence. Suspects and others at the location under investigation are interviewed, passwords and similar information is retrieved, and in some cases this is used on site to gain access to computer systems. If proper procedures are not followed, then the evidence arising from this process may be invalidated. For example, if a suspect is arrested, not Mirandized, and asked for a password to a computer system, then all of the evidence from that system may become unavailable for prosecution if the password is used to gain access to that information.

Selective collection and presentation

In some cases prosecution teams have opted to not do a thorough job of collecting or presenting evidence. They prefer to seek out anything that makes the defendant look guilty and stop as soon as they reach a threshold required to bring the case to court. Many US prosecution teams try to prevent the defense from getting the evidence, provide only paper copies of digital evidence, and so forth. In such cases the defense should vigorously challenge the courts to require that the prosecution present all of the evidence gathered in the same form as it was made available to them and for a similar amount of time. On the other hand, most defense teams also fail to present evidence that would tend to convict their clients, and they certainly don’t try to help the prosecution find more evidence against their clients. Other countries are less adversarial.

A good example of such an attempt involved the prosecution providing a printout listing the files on a disk. The printout was hundreds of pages long, contained no useful information, and could not be processed automatically. The defense in this case brought forth arguments that this was unfair and that the evidence should not be admitted at all unless the defense had adequate access to it. The issue of best evidence was also brought up. A paper copy of an extract from original electronic media is not best evidence and should not be allowed to be used when the original and more accurate copies are available and can be provided. This discussion is not intended to indicate that such behavior is limited to prosecution teams. Defense teams also do everything they can to limit discovery and make it as ineffective as possible for the other side. But because the prosecution is the predominant gatherer of digital forensic evidence in most criminal cases it ends up being the prosecution that conceals and the defense that tries to reveal.

Forensic imaging

In order to address decay and corruption of original evidence, common practice is to image the contents of digital evidence and work with the image instead of the original. Imaging must be done in such a way as to accurately reflect the original content and there are now studies done by the United States National Institute of Standards & Technology (NIST) to understand the limitations of imaging hardware and software, as well as standards for forensic imaging.[3] If these standards are not met there may be a challenge to the evidence; however, such challenges can often be defeated if proper experts are properly applied.

In at least one instance a 'disk dump' (dd) image of a disk was thrown out because some versions of the 'dd' program operating on some disks failed to capture the last block of some disks. This is rare and in the particular case no finding was made to indicate that this had happened, yet the people who were trying to get the evidence excluded won, on that issue. Here are some counters:

(1) The original evidence disk is seized and retained. If it is still there, it can be re-imaged and the full content examined.

(2) The image with dd is only inaccurate on disks of certain odd sizes, and since the disk in this case has not been shown to be such a disk, the image can be shown to be accurate.

(3) The image taken with 'dd' is accurate except for that last sector, so all of the evidence provided using it is still accurate. If the other side wants to assert that there is some evidence in the last sector that makes a difference, they can feel free to, but nothing there invalidates the evidence that does exist.

In many cases it can be demonstrated that the last sector of a disk does not have any relevant evidence because many operating systems use it for redundant copies of other data, in which case the contents can be accurately reconstructed. It turns out that every current imaging product has been shown to have similar flaws under some circumstances. None of them are more accurate than 'dd' according to NIST, so unless all such evidence is to be ignored, this evidence must be allowed. But in this case it seems clear that justice was not served with regard to that disk image.[4]

Proper technique in forensic imaging starts with a clean media for the results of the image. To assure that no evidence is left over from previous content of the media, the media is first cleared of data through a forensically sound erasure process. This is often not done. After clearing (overwriting) the information, it is best to put a known, but unlikely to appear in normal evidence, pattern on the media, to later detect failures to properly image the media. After verifying this content is correct, the image is then taken. The original media is cryptographically checksummed, either in parts or as a whole, the image is made, then the result is verified using the cryptographic checksum(s). The result can be tested for the presence of the identifiable cleared content and the start and end of the evidence can be clearly verified by these patterns. [5]

While failures to do these steps does not invalidate the image, they do bring into question the potential for contamination. Similarly, cryptographic checksums can be questioned as can the validity of the mechanisms for extracting and storing data on the media, but these challenges are unlikely to succeed against a competent forensic imaging expert because the processes are so effective and hard to refute. Perhaps the most promising area for technical challenges in cases where proper technique was used lies in the potential for disk content, as reflected at the normal interface, to fail to reflect accurately the content of the physical media. This is because of the electronics that mediate between the interface and the media. Such challenges have never been successfully made and would require a very high degree of expertise and great expense; however, it is a potential that has not yet been explored. [6]

Non-Stored Transient Information

The collection of non-stored transient evidence typically involves a technical collection mechanism and often requires minimization in a law enforcement context. In the case of analog telephony, tape recorders and special electronics are used, while in the case of digital traffic, the typical tool is a packet sniffer. Most packet sniffers have limitations in the form of collection rates, storage capacity, and ability to capture all packets. These limitations may be the basis for challenges involving missed information. Made data[7] is far more difficult to deal with in a packet sniffing technology. These technologies typically record what is sent through the media, but attribution to a source is more problematic.

Typical Ethernet interfaces use MAC addresses associated with packets to identify the hardware device associated with a transmission. While these include manufacturer specified serial numbers that should be unique to a physical device, they can often be forged with software. If the environment was not examined for the presence of such software and if other hardware is present, it is a reasonable challenge to assert that the data may not have come from the identified computer. In the absence of corroborating evidence, tying traffic to a computer is not directly evident.[8]

In order to assert that such data is legitimate as evidence there is a requirement that the manner in which it was gathered be demonstrated to be reliable. In cases involving communications media there may also be requirements to follow wire tap laws as opposed to other laws. Other examples, such as radar and infrared imaging, tape recording, and so forth, may all involve digital forensic evidence as well.

This evidence is typically stored by the collection mechanisms in some media with specific formats and characteristics, and can potentially be altered. Again the evidence has many characteristics that allow it to be examined by experts to determine if any obvious alteration has taken place. Many examples of this sort of content now exist because of the ready availability of computers with image and sound manipulation programs. In some cases people alter voices by combining recordings or reordering portions of them, pictures can be merged or altered to create false backgrounds and contexts, so-called 'morphing' can be used to make characteristics seem similar, and digital artists can be quite skilled at creating digital renderings. Tools exist for creating shadows and similar realistic patterns, and are relatively easy to use and low cost.

Secret science and countermeasures

This is another similar line of pursuit that has been used to prevent criminal defense teams from gaining access to key evidence and methods of gathering and analyzing evidence. In essence, the prosecution says that they have an expert who used a secret technique to determine that the defendant did this or that. The defense asks for access to the means and detailed evidence so that they can try to refute the evidence, and the prosecution claims that this information is a government secret, classified at a level so that the defense team cannot see it. In addition, because of the way classifications work and because of a concerted effort by those who represent the government, anyone who works for a defendant is prevented from access to many of the people and techniques used by prosecution teams. Secret science presented by secret scientists is presented as objective fact that cannot be challenged. Even worse, some trial judges have let such evidence in.[9]

There are some methods for countering such abuses, and they should be tried with great vigor. One method is to have a digital forensic evidence expert on your team who has security clearances. This may be hard to do since very few such experts exist; however, there are some available. With such a person available, the secrecy argument can largely be eliminated from the process of examining the evidence, but the problem remains of how to try the case. Is the defense expert going to say that the prosecution expert is wrong and provide no details? This would seem ridiculous, and yet it may be the only alternative. Another alternative is to have this part of the case tried before the judge with a resulting stipulation. Yet another alternative is to ask revealing questions that don't violate the secrecy requirements while still getting at the fundamental issues in the case.

For example, a lawyer might go through each of the known technologies for gathering and analyzing the evidence and ask whether each of them was used. Even if there is no response, each can be pursued for its potential flaws with questions that the expert may or may not be able to answer. The opposing expert can then address the flaws associated with the technologies and indicate whether any of them may have been present in this matter without revealing which ones are relevant and thus without revealing the secrets. The best course would be to have the evidence thrown out. For example, (1) it is not best evidence, (2) it is hearsay, (3) it is highly prejudicial, (4) it was not made available to the defense; thus, preventing the defendant from a fair trial, (5) its scientific validity has not been established (as will be discussed later), (6) the expert has not been shown to be an expert in the particular type of evidence under scrutiny, and (7) the defense has been denied the right to challenge the scientific evidence. If the case can be made that it is more prejudicial than probative, the case stands a chance, if not in court, then on appeal.

Seizure errors and related issues

The evidence seizure process has the potential of producing a wide range of errors that may lead to challenges. Search and seizure laws may mandate Title III[10] searches for live capture of digital data passing through telecommunications channels, permission for searches may be removed at any time during a permission search and continued searching at that time may violate rules of evidence, search warrants must be adequately specific to avoid becoming fishing expeditions, and the searches must be limited to meet the requirements of the warrant if a warrant exists. Hot pursuit laws rarely apply to digital evidence, but the laws regarding 'plain sight' are far more complex. Many other problems are seen over networks.[22] Cases have gone both ways in digital media searches.

Warrant scope excess

In one case, a warrant for a search for pornographic images was found to be exceeded when the officer making the search looked in directories with names that were indicative of other legitimate use. Of course this is patently ridiculous because plenty of criminals have been found to store pornographic images under false names, in hidden files, in directories that hold legitimate images, and so forth, but the challenge worked because the judge was convinced. This is the combined result of a poorly written search warrant and a inadequately computer-literate judiciary, and this case was one of the earliest ones in this area. Such errors are likely on all sides.

There are legitimate cases on both sides, not all judges will rule the same way on the same information, and not all experts do the same things on the same cases. This sort of variation is in the nature of the work and is to be expected. Challenges should be undertaken when the evidence search and seizure process used in a non-permission search fails to meet “reasonability” requirements, when search warrant bounds are exceeded, when minimization is not adequately applied, and whenever evidence is found in a search that does not meet the original warrant and the search is not immediately stopped pending a new search warrant for the new sorts of information. In permission searches, there is normally a scope of permission, and unless it is unlimited, it may have the same problems as a warrant search in terms of admissibility.

Acting for law enforcement and private investigations

Similar limitations exist for situations in which a non-law enforcement person is acting on behalf of law enforcement or the government. In most cases when a private individual undertakes a search of their own volition and reports results to law enforcement there is no problem associated with illegal search and seizure, although the purity of the evidence may of course come into question. But in cases where there was a pre-existing relationship with law enforcement, when the specific case was under discussion between law enforcement and the person doing the search, or when the search was ordered by someone who was in touch with law enforcement on the matter, there may be an issue of admissibility under this provision. In addition, in many states, only law enforcement officials, licensed private investigators, and lawyers or their workers can investigate people legally. Evidence gathered by others may be inadmissible and there may be legal recourse against those others for their illegal acts.

Wiretap limitations and Title 3

In some cases where a wire tap or network tap is used, there may also be issues associated with the legality of such a wiretap. There are many states in which all parties to a communication must give permission for a recording. Without such permission, the recording may be inadmissible and the person doing the recording may be legally liable for a criminal act. The laws on real-time collection are not very clear and inadequate case law exists at this time; however, this makes an ideal case for attempting to challenge evidence. The expertise of the person gathering the evidence is important to examine. In addition, if minimization is done, then an argument can sometimes be made the exculpatory evidence[11] was excluded in the gathering phase. This depends heavily on the specifics of the situation.

Detecting alteration

Detecting alteration is very similar to the field of questioned documents, except that in this case the documents are digital rather than analog. There is a lot of tradecraft involved in trying to figure out whether there are forgeries and what is real and fake in such digital images of real-world events. For example, the locations of light sources and their makeup generate complex patterns of shadows that can often be traced to specifics. Specific imaging technologies leave specific headers and other indicators in the image files that result from their use. Aliasing properties of digitizers, start and stop transients, scratches and dust on lenses, frequency characteristics of pickups, and similar information may make forgery and digital alteration readily detectable by sufficiently skilled experts. As in other forms of digital forensic evidence, the cases where this sort of examination is relevant to the matter at hand are rare, but there are times when such analysis yields a valid challenge.

This particular sort of challenge has a tendency to appear more often in civil suits than in criminal cases because it is rare to find an instance where a member of the prosecution team creates such a forgery and defense teams haven’t the desire, time, or money to create such a forgery.

In civil matters; however, these sorts of situations are far more common. For example there are many cases where a celebrity's head is places on a naked body for the purposes of increased sales of pornographic material, and these forgeries tend to be readily detectable.

Another interesting example was the case of a digital image purported to be one of the airplanes photographed from the top of one of the World Trade Center towers as it was about to hit the building. This was asserted to be evidence that the Massad (the Israeli secret police) was a co-conspirator in the September 11, 2001 attacks. The forgery was rapidly detected by an examiner who found many errors in the rendering, including wrong light direction for the time of day, improper scaling for the angle and distance, shadowing errors, edge line aliasing errors, and so forth.

Collection limits

Because all collection methods are physical, there are inherent physical limits in the collection of digital evidence. The digitization process further introduces sources of low-level errors because of the rounding effects associated with clocks relative to time bases and voltages or currents with respect to bit values. The challenges to evidence collected based on signals approaching these limits is typically based on the inability of the mechanism used to gather the evidence to accurately represent and collect the underlying reality it is intended to reflect. Error correction mechanisms often imply changes to underlying physical data to regain consistency and they also produce a probabilistic and increasing chance of error as the physical signals approach these limits

Good practice

According to many of the “good practice” documents, the following advice is offered.

It is good practice to secure the scene and move people away from computers and power. This is a basic safety issue and assures that the people, investigators, and equipment are protected. Failure to do so is not likely to produce any false evidence, but it may result in the loss of otherwise valid evidence.

The investigator should not turn on the computer if it is off, not touch the keyboard if the computer is on, and not take advice from owner or user. Clearly a computer that is not turned on should not be turned on because this is very likely to produce alterations to the computer which may destroy its evidentiary value. Not touching the keyboard is somewhat more problematic. Many systems use a screen saver to lock out users after inactivity, but in most cases, it is better to leave the keyboard and mouse alone. In terms of taking advice from the owner, more care may be necessary. Many systems are interconnected via the Internet, and if the user asserts some potential for harm associated with actions and that harm takes place, liability may be accrued. Such information should be passed through an experienced investigator and not taken out of hand. This is a place where good judgment may be important, and of course judgments always have the potential of being challenged.

The screen should be photographed or its content noted, the printer or similar output processes should be allowed to finish, and the equipment should be powered off by pulling out all plugs. Taking a photograph or making notes of what is on the screen is certainly a reasonable step and cannot reasonably be expected to destroy or create evidence. Indeed it is likely to assure purity and consistency. Allowing output to finish may leave time for other undesirable alterations to the system. However, it may also provide additional evidence. If the system is networked this becomes a more complex issue. For example, a remote user might alter the system after finding out that the normal user is not there, or even observe what is happening via an electronic video link. Powering off systems may also create problems, particularly if these systems act as part of the infrastructure of an enterprise. For example, this could cause all

Internet access to many domains to fail, or cause loss of substantial amounts of data. In some data centers with large numbers of computers this is simply infeasible or so destructive as to be imprudent. [12]

The investigator should label and photograph or video tape all components, remove and label all connection cables, remove all equipment, label, and record details, and note serial numbers and other identifying information associated with each component. The area should be searched for diaries, notebooks, papers, and for passwords or other similar notes. The user should be asked for any passwords and these should be recorded. This process is clearly prudent and to the extent that something is not photographed or labeled it may lead to challenges. Wrong serial numbers, missing serial numbers, and similar errors may destroy the chain of custody argument or lead to challenges about what was really present. At a minimum, these sorts of misses create unnecessary problems later in the case.

Fault type review

Faults in collection are most commonly misses of content, process failures or inaccuracies, missed content caused by inadequate collection technology or skill, missed relationships, missed timing or location information, missed corroborating content, and missed consistencies.

Packaging for transport

Packaging for transport of digital forensic evidence has requirements similar to those of other evidence, and it should be transported in a timely fashion to a facility where it can be logged into an evidence locker, chain of custody requirements must be met throughout the process, and it has to be kept in a suitable environment to the preservation of the evidence.

If a claim of evidence tampering is to be made this will have to be shown to have taken place when it was in the custody of the person who is asserted to have taken this action. In one case I was able to show that the amount of time available to an individual accused of tampering was inadequate to have planted the evidence asserted to have been planted. Tampering is not an easy thing to do without detection. Because of all of the inherent redundancy associated with digital forensic evidence as described earlier, tampering can often be detected by a detailed enough examination.

Transport of evidence

When digital evidence is taken into custody appropriate measures should be taken to assure that it is not damaged or destroyed, that it is properly labeled and kept together, and that it is not mixed up or otherwise tainted. If these precautions are not taken, the results can be effectively challenged.

Possession and chain of custody

It is common practice in some venues to videotape the evidence collection process, and this has been invaluable in meeting subsequent challenges in many cases. In one example, a challenge was made based on the presence of a floppy disk in a floppy disk drive, however, the videotape clearly showed that no floppy disk was present and this defeated the assertion. In any case, a sound chain of custody should be maintained, and things like emailing then deleting evidence not covered by message authentication, transmission of the original evidence in plaintext or without cryptographic checksums through the Internet, or similar acts may break the chain of custody.

Due care takes time

Based on the requirement for a speedy trial and high workloads in most forensic laboratories, time constraints are often placed on storage and analysis of evidence. A job done quickly, normally translates into a job done less thoroughly than it might otherwise have been done. The more time spent, the more detailed an examination can be made and the more of the overall mosaic will be pieced together. In practice, most cases are made with a minimum of time and effort on such evidence and this opens the opportunity for errors and the resulting opportunity for challenges.

Good practice

Handle everything with care. Keep it away from magnetic sources like loudspeakers, heated seats, radios, etc., place boards and disks in anti-static bags, transport monitors face down buckled into seats, place organizers and palmtops in envelopes, and keyboards, leads, mice, and modems in aerated bags. Avoid electron beams, X-rays, and other similar invasive search technologies.

Storage of evidence

Evidence must be stored in a safe, secure environment to assure it is safe from alteration or destruction. Access must be controlled and logged in most cases. But this is not enough for most digital evidence. Special precautions are needed to protect this evidence just as special precautions are needed for some sorts of biological and chemical evidence.

Decay with time

All media decays with time. Decay of media produces errors. Typically, tape, CDs, and disks last 1-3 years if kept well, but can fail in minutes from excessive heat (e.g., in a car on a sunny day, on a radiator, or in a fire). Electromagnetic effects can cause damage in seconds, as can high impulses or overwriting of content. Non-acid paper can last for hundreds of years but can also fail in minutes from excessive heat or in seconds from shredding or being eaten. Audit trails are another thing that tend to decay with time. Some are never stored, while others last minutes, hours, days, weeks, months, or years.

Evidence of integrity

Evidence of integrity is normally used to assert that digital forensic evidence is what it should be. This is generally assured using a combination of notes taken while the data was extracted, a well understood and well tested process of collection, the ability to reproduce results, which is a scientific validity requirement in any case, chain of custody records and procedures, and proper imaging techniques associated with the specific media under examination.

The establishment of purity of evidence is generally better if done earlier in the process. The media being imaged or analyzed should be write-protected so that accidental overwrite cannot happen. A cryptographic checksum should be taken as soon as feasible to allow content to be verified as free from alteration at a later date. It may also be wise to do cryptographic checksums on a block-by-block and file-by-file basis to assure that even if part of the evidence becomes corrupt or loses integrity with time the specific evidence is covered by additional codes. This allows us to assure the freedom from alteration of portions of a large media even if the overall media becomes corrupt. Keeping the original pure by only using it to generate an initial image and working only from images from then on is a wise move when feasible. Validating purity over time also helps to assure that time is not wasted and that no alteration occurs in the analytical process.

Nobody knows for certain that any evidence is completely pure and free from alteration, and likely nobody ever will. But this does not mean that all evidence can be successfully challenged or should be. Just because people are not perfect, doesn't mean they are not good enough. Just because evidence is not perfect doesn't mean it is not good enough.

Principles of best practices

Some principles of best practice have been stated by various parties. This is a list of one of the accepted sets of principles. While there are many such lists, many are quite similar in that they assert some fundamental ideals that are realistically achievable. In this case they regard assuring the integrity and utility of the evidence.

- Principle 1: No action should change data held on a computer or other media.
- Principle 2: In exceptional circumstances where examination of original evidence is required, the examiner must be competent to do it and explain relevance and implications.
- Principle 3: Audit records or other records of all processes applied to digital evidence should be created and preserved. An independent third party should be able to reproduce those actions with similar results.
- Principle 4: Some individual person would be responsible for adhering to these principles.

Evidence analysis

Evidence analysis is perhaps the most complex and error-prone aspect of dealing with digital forensic evidence. It is also the most subjectively applied in many cases. But in almost all cases it should not and need not be subjective. It is subjective largely because of the failure of those undertaking analysis to spend the time and effort to be careful in what they do.

Content

Making content typically involves processing errors. For example, uncleaned media is used in the analysis process and the analysis finds evidence that was left over from a previous case. This was addressed under imaging, above. The challenge to this can come in many forms, and if original evidence or cryptographic checksums are not used, such challenges have a good chance of success because of the inability to independently verify results. If originals are present and checksums can be shown to match, then such challenges will likely only succeed in the presence of an actual and material analysis error, because the purity of the evidence can be usually properly established.

Missing content typically results from limited time or excessive focus of attention. Limited time is almost always an issue because there is usually an enormous amount of evidence present, most of which can only be peripherally examined with simple tools. Examining every bit pattern from every possible perspective is simply too time consuming to be feasible and is almost never necessary to get to the heart of the evidence. Excessive focus, on the other hand, is easier to avoid. By simply taking an open view of what could be meaningful evidence and being thorough in the evaluation process, such misses are avoidable. The challenge is simple. Did you look at everything? Is there any exculpatory evidence? Where did you look? Why did you not look in the other places? What technique did you use? Why did you not use a more definitive technique? Is there a more definitive technique? The questions can be nearly endless.

Contextual information

Information only has meaning in context. Analysis can make context by making assumptions that are invalid or cannot be demonstrated. Context is missed when assumptions that are valid and can be demonstrated are not made. The challenge to made context starts with questioning the basis for assumptions. If assumptions cannot be adequately demonstrated, the context becomes dubious, the assumptions fall away, and the conclusions are not demonstrable. If an alternative context can be demonstrated with the same or better basis, that context can be substituted and the interpretation of the evidence altered. Missed context can be challenged with the introduction of alternative contexts. It then becomes the challenge of the other side to disprove these contexts.

An excellent example of this was a case where the defense challenged the validity of the evidence by introducing the potential that an attack on the computer system being examined caused the illicit effects rather than the user at the console. In the end, the prosecution could not convincingly demonstrate the validity of its assumption that the user that owned the computer carried out the behavior in question, and the case was dismissed. While many computer security experts assert that there was no evidence of the presence of such an attack, the absence of evidence of an attack by the defense is not the same as evidence of absence of such an attack by the prosecution.

Meaning

The meaning of things that are found is obviously the basis for interpretation. Meaning that is missed leads to a failure to interpret, and meaning that is made is an interpretation without adequate support.

For example, the presence of a file with a name associated with a particular program might indicate that the program was present at some time in the past, but not necessarily. The filename could be a coincidence or it could have been placed there by other means such as part of a backup or restoration process. In most cases there are a variety of different meanings that can be applied to content, and determining the most likely meaning typically involves reviewing the different possible meanings relative to the rest of the environment. The same applies to the content, presence, or absence of files, directories, packets, or any other things found in material under examination.

Process elements

Content does not come to exist through magic. It comes to exist through a process. The notion that a sequence of bits appears on a system without the notion of how that sequence came to exist there makes for a very weak case. If the bits were created within the system, the means for their creation should be there unless it was somehow removed. If the bits were obtained from somewhere else, the process by which they got there should be identified. If there are alternative explanations for the arrival of the bit sequence, why is one interpretation better than the other?

Processes normally generate audit records of some sort somewhere. Files have times associated with them in most file systems. If a file was retrieved from a network, audit records from the location it came from and the connection to the network may be recorded. If some image was deleted and the residual information from it is gone, there must be some process by which that event sequence occurred. Without evidence of the process, alternative explanations may be offered with as much credibility as the explanation preferred by the other side. The plausibility of these explanations is key to the meaning of the content they refer to.[13]

Relationships

Just as sequences of events produce content, relationships between event sequences and content produce content. The presence or absence of related content causes differences in the content generated by related processes. The presence or absence of a directory prior to running a program that uses or creates it produces a difference in the time associated with the creation of the directory. Similarly, the placement of the directory in the linked lists associated with the file system relative to the placement of files within that directory may indicate the differences in these relationships. There are many such relationships within systems and those relationships can be explored to challenge the assertions of those who make claims about them.

Ordering or timing

Sequences are a special case of orderings. More generally, orderings can involve things that cannot be differentiated from being simultaneous, while sequences are completely ordered. Timing often cannot be established with perfection, but partial orderings can be derived. The possible orders of events can make an enormous difference in some cases. One obvious reason for this is that ordering is a precondition for cause and effect. To assert that one thing caused another, it must be demonstrable that the cause preceded the effect. If this cannot be established by timing there is the potential to challenge based on the lack of a causality.

While this may seem like a highly theoretical argument, many cases have been made or broken by the ability to show time sequences. If an accusation of theft and disclosure of trade secret information is made and it cannot be shown that the theft preceded the disclosure, then the basis for the claim is invalid. If an attacker was supposed to be present at a computer to commit a crime at a particular time and the accused can show that at that time they were not present at that computer, then the alibi will refute the claim. If the time is uncertain in a computer system, the entire process becomes suspect because computers usually keep time very well. If ordering or timing is missed, the lack of the ordering or timing information leads to challenges.

If timing assumptions are made and not validated, they can be questioned.

The most common challenge to computer-related times stems from the potential difference between a computer clock and the real-world time. Even accounting for time zone variations this is an all too common problem that has to be addressed in the forensic process. If the time reference for the computer is not established at the time the evidence is collected, timing can sometimes be obtained by relating the timing of events within the computer to externally timed network events. Missed time can sometimes be made up for by correlation with outside events, while made time can often be demonstrated wrong by similar correlation. The lack of correlating information represents sloppiness in the collection and analysis process that may itself lead to the inability to determine timing.

A less used challenge stems from the ability to determine ordering of events in the absence of other timing information. For example, there are cases when times and exact sequences can not be determined but orderings can, because of overwrite patterns on disks. In one such case it was shown that a file transfer happened before an erasure was done because the area of disk where the file would have existed had the erasure happened first was covered with the pattern associated with the erasure. What happened between these events and the precise times they occurred could not be determined, but the ordering could be, and it was one of the pieces in a larger puzzle that determined the outcome of the case.

Location

Everything that happens in computers has physicality despite any efforts to portray it as somehow ephemeral. Physicality tends to leave forensic evidence in one form or another. For example, when a person uses a keyboard, particles from hair and skin fall into the keyboard and tend to get stuck there. In a similar fashion, data in computers tends to be placed on the disk and tends to get stuck there. Computer systems have physical characteristics as well, and sometimes they are dead giveaways to location.

In one example of a missed and made location, an attack against a government computer system had an Internet Protocol (IP) address associated with a location in Russia. But when observing the traffic patterns shown in log files, it was determined that the jitter associated with packet arrival times was very small.[14] Packet arrival time jitter tends to occur when packets are mixed in delivery queues across infrastructure. More infrastructure tends to lead to more jitter. The lack of jitter meant that the arriving packets were not being mixed much with other packets and that the computer responsible was therefore close, in terms of network hops, to the observation point. The attack was traced to a point only a few network links away from the surveillance point. Evidence such as this can make a compelling case, and it’s easy to miss the real location and make a false location when such analysis is not thoroughly undertaken.

Inadequate expertise

We also face many low quality 'experts' and people with an axe to grind who are put up as experts. An excellent example of this was a case involving copyright infringement in which a court appointed special master made claims including; (1) the accuser may have altered data, (2) date and time information was unreliable, (3) the system never worked when returned to its owner after it was forensically investigated, (4) programs were destroyed, and (5) pre-existing data was no longer present. In this case, all of these claims were refuted by a better expert who used recorded statements, a videotape of the return process, some details of the physics of writing to disks that dismissed the possibility of forgery, and correlation with other records. It is important to note that under some circumstances all of the things asserted by this special master could have been true, so the claims were not outrageous in the general sense. It is only that they were not in fact true in this case.

Unreliable sources

There are a lot of unreliable sources of digital content. For example, the Internet is full of the widest possible range of different content, only a portion of which is really accurate, and a significant portion of which consists of just plain lies. There is a tendency for people to believe some of these things, and once these things are believed, the belief transcends the original source. For example, when looking up information about the function of a hardware device or software program, at a detailed level, much of the information on the Internet is not accurate. It might reflect a different version, it may reflect a mistake by the author, it may reflect a simplification by the author for reader convenience, and it might be an intentional or malicious misstatement by a disgruntled ex-employee. While this information may be convenient or readily available and useful in many circumstances, it is not generally suited to the level of trust required for digital forensic evidence purposes.

As a good example, the question sometimes comes up of the list of all circumstances under which a file access date and time will be altered by a Windows operating system under normal use. It turns out that this is not an easy question to answer. The answer is different on different versions of Windows, different applications may use different system calls for the same outcome with different side effects, and programs that deal with forensic processes typically do things differently than other operating system programs.[15] Even such simple questions do not have easy answers and the Internet answers are not consistent or accurate in many cases.

Simulated reconstruction

In some recent court cases, computer-based reconstructions of physical events have been used in presentations to juries. In some sense this is no different than the use of story boards to show how a crime is purported to have happened, but in another sense it can be too realistic in that it can give the appearance of certainty about many things that there is no certainty about. In such a case there are a number of approaches to reducing the impact of such fabricated evidence.

One of the keys to countering this sort of evidence is the use of the terms “fabricated”, “fabrication”, and “fabricating”. When referring to this evidence it should always be identified as a fabrication. Underlying this question is the issue of the prejudicial value as opposed to the probative value. The question is one of differentiating the part of the fabrication that is more probative than prejudicial from the part that is not. Is the use of continuous video more probative or prejudicial than a set of storyboards? The enhancement generated by motion is, in almost all cases, more prejudicial than probative because the intervening fabrication of motion is highly specific while the real knowledge of the details is almost always very limited. Did the perpetrator use their right or left hand? Did they really bend their elbow like that? Did they trip over an obstruction on the floor? Did their shirt really wrinkle like that?

If the information provided by the fabrication is not demonstrably accurate, it is not relevant, and provided without a proper foundation. If the coloring of the face is similar to that of a defendant this is prejudicial and, unless there is evidence as to the coloring of the face of the perpetrator, green might be a better choice. It is valid to zoom in on the parts of the presentation and ask whether the information at that detailed level is accurate. If the depicted gun is a different sort than the one used in the crime, the gun type should be shown to the finder of fact and the question should be asked about whether this is evidence that the defendant did not do this crime. The answer may be “no”, and this gives the opportunity to again point out the fabricated nature of the display and its gross inaccuracy as to the facts. What in fact is real about this fabrication? Can you tell us whether the colors in this fabrication are real? How about the shadows? Is the time on the fabricated clock real? Are the footfalls real? Does this fabrication show any trace evidence being left on the site? What evidence do you have of the height of the person in this fabrication?

Reconstructing elements of digital crime scenes

Digital crime scenes can also be reconstructed, and this is a critical area for scientific evidence. While experts may assert any number of things about what might be within a computer, the ultimate test can often be made through a reconstruction. But even reconstruction of a digital crime scene has its limits. While similar circumstances can be created, identical ones often cannot. As a rule of thumb, simple questions can often be answered by digital reconstructions, but complex sequences of events are far harder to confirm or refute.

A simple example where reconstruction is very effective is determining whether or not a file is created by an application in the normal course of operation. To test this it is a simple matter to install the application on a system, operate it normally, and see if this file is created. It is far harder to make this determination through reconstruction in abnormal operation. An excellent case example of a reconstruction that refuted evidence was in a case where the prosecution asserted that a particular program could be used to extract deleted file content from a floppy disk. The prosecution knew that a utility program by a particular name was present and that this program was commonly used for this sort of operation. The defense did a simple reconstruction. They took the actual program on the defendant’s system and tried to do what the prosecution claimed could be done. This failed. It turned out that the particular version of this utility program on the defendant’s computer did not have the capability the prosecution claimed it had. Earlier and later versions had this capability, but not the one on the defendant’s system. Case dismissed.

Digital reconstruction can be a powerful tool, but it cuts both ways. There are plenty of cases where reconstruction confirms rather than refutes the evidence. Indeed, this is one of the great values of doing reconstruction. It tends to get at the truth. The problems with such reconstructions come when they are interpreted too generally or when they are used to try to make claims about complex situations. A good example of the limitations of such reconstructions is any case where many possible sequences of events could have taken place and these events involve complex interactions between components. The larger the number of possible sequences, the more reconstruction runs are necessary to exhaust the space of possibilities. In cases where exhaustion is not feasible statistical samples can be taken, but the nature of digital systems makes such statistics highly questionable. The more intertwined the elements are, the more complex the potential sequences become.

In a simple case where the ability of a program to perform a function through the normal user interface is at question, reconstruction is simple and effective. Simply install the specific software on the specific system in question and try to use the interface to generate the desired results. In some cases complex sequences are required in order to generate specific outputs, but manufacturers and manuals are usually adequate to generate things that are meant to be generated.

If the goal is to prove that a program could have generated an output, it may be far more complex, depending on the specifics. While some outputs are generated often or predictably, other outputs may be very situation-specific and may involve complex interactions with the environment. Creating the entire environment may be problematic if it involves things like network events that are usually out of the control of those doing reconstruction. There are forensic technologies that can largely accomplish these sorts of reconstructions, but they are rarely used and difficult to properly implement.

If the goal is to show that it was impossible for a perpetrator to have entered a computer and performed a function without leaving any evidence, the task may be very difficult. More generally, proving an assertion about something with unlimited numbers of possibilities or disproving something under the assumption of the perfect opponent is very hard and sometimes impossible. But in almost all such cases a demonstration can be constructed to show some subset of the scenarios. If this is done, the challenge can be made on the basis that the space was inadequately covered.

Good practice in analysis

It turns out that nobody has yet compiled a widely accepted collection of good practices for analysis of digital evidence. In fact, parts of this book may be as close to such a compilation as you are likely to find. As a result, all of analysis is subject to expert interpretation and challenges of all sorts, and each case will be judged on its merits without appeal to some standard. But there are some time-tested analysis techniques that should be covered despite the lack of any widely published good practices.

The process of elimination

It is generally considered good practice to use the process of elimination. In this process, a list of the possibilities is made and items on the list are eliminated one at a time or in groups for specific reasons that can be backed up by facts.

The challenges to the process of elimination start with the initial list. Is the list comprehensive? How do you know it is comprehensive? If I find one thing missing does this invalidate the list as being less than comprehensive? What if I find 2? Are there implicit assumptions in the list? What are they? Are they demonstrably true? What if they are wrong?

The next challenge comes in the application of the process. Was the test of each item on the list definitive? Was it properly done? What are the cases where this test would fail to be revealing? Does a positive or negative result in your test environment definitively show that the same result would be seen in the real environment? In other words, what are the possible differences between the real world and the test environment?

And finally, almost all such tests make assumptions about the independence of the items on the list or the elements within the computer systems involved. Suppose these things are not independent - would that invalidate this test? In many cases these assumptions can be demonstrated to be incorrect under certain circumstances.

The scientific method

The basis of the scientific method is that the truth can be only be confirmed by the failure of experiments that attempt to disprove an assertion. But even one refutation implies that the hypothesis under test or that the testing method was incorrect in some respect.

The scientific method, and more generally, a scientific approach, is fundamental to success in digital forensics as elsewhere, and failure to take this approach seriously will likely result in errors in your analysis and successful challenges of it by competent legal or technical professionals.

The Daubert guidelines

Case law in the United States has led to the Daubert Guidelines for the admissibility and validity of scientific evidence. These almost always apply to digital forensics in US Federal cases and most international cases.[16] The tests of scientific evidence include four basic issues; (1) Has the procedure or technique been published? (2) Is the procedure generally accepted? (3) Can and has the procedure been tested? (4) What is the error rate of the procedure?

Most digital forensic analysis methods in use today have not been widely published. Those that are published are rarely published in refereed scientific journals. There are several books on this subject and more such books are being written. In addition to these publications, there are manuals from products and books on special purpose topics. Finally, hardware and software design and source code information is often available to those properly trained to understand it. These then form the literature in this area. The lack of published material leads to many challenges to the admissibility or validity of this evidence. Perhaps most importantly, most of the forensic examination and analysis tools that are made for sale include trade secrets and unpublished content that forms the intellectual property basis which creates competitive advantages and barriers to entry in the market. As a result, the most commonly used tools do not include the information required to determine precisely what they do. Their use and their results can be strongly challenged on this basis. Similarly, file formats, hardware device operations, and similar functions of components are often not known at the most detailed level and their operation is not published. Open source has selectively improved this situation.

In terms of being generally accepted, there are few generally accepted analysis techniques for digital forensic evidence. In some sense, without publication, general acceptance is impossible, but almost any presentation of analysis of digital forensic evidence is challengeable on this basis. For example, if a forensic analyst asserts that a file contains some specific data and was created at some particular date and time, in addition to the technical limitations of this assertion, the methods used to derive this information can be questioned in great detail to try to shake confidence in the validity of the technique undertaken or the credibility of the witness. The problem with such a challenge is that it risks alienating the finder of fact and is likely to come down to things that users do every day but that are not documented as forensic processes. This approach is more applicable in cases where the technique is believed to have produced wrong results. In such cases, wrong results are usually easily demonstrated by following the specific steps taken by the examiner.

Public testing of analysis techniques has not been done to date. While the United States National Institute of Standards and Technology (NIST) is undertaking tests of forensic imaging processes, analysis techniques are essentially only tested today by the individuals doing work in this area. The tests undertaken can usually be described by the forensic examiner, but they are likely to be very limited. It is reasonable in most cases to challenge the tests done to validate the technique used by the examiner, but the import of this depends heavily on the evidence being presented. If the person presenting the evidence has not tested their own tools they will probably be hard pressed to show that it has been tested elsewhere and the credibility of the technique and the person applying it can be challenged. [17]

The error rate of forensic analysis tools is even harder to attest to because in almost all commercial products no error rate can be established without independent tests. As a result, while error rates for things like cryptographic checksums on forensic images can be estimated, error rates on disk searches and depictions of images based on file content are far harder to ascertain. The problem with this approach to challenging evidence is that the underlying digital technology is very good at nearly error-free operation. While there are almost always errors in the programs implementing any digital forensic process, these errors are not apparent and similar processes can be undertaken with independent software to mitigate against these challenges. Again, challenges here should only be made in cases where there is a good reason to believe that the resulting facts are in error. Such results should be tested to confirm that the expert on the other side didn't make an error in their use of the methods or technologies, which is quite common.

Digital data is only a part of the overall picture

Almost every analysis of digital forensic evidence ultimately involves ties to the real world. In order for analysis to produce meaningful results it must tie those results to the matter at hand. The analysis process, as a technical matter, can often be resolved and some set of resulting bit sequences with some time sequencing can be revealed to the finder of fact without significant disagreement. In fact, this is often done by stipulation, subject to an adequate presentation. However, the interpretation of those results is subject to far more variation than the setting of the bits.

The attribution of actions to actors is often hard to pin down. While there are cases where there are films of people using their computers and typing the material of interest to a legal matter, this is certainly the exception rather than the rule. Attribution has been and continues to be studied and while there are many indications that behavioral and biometric indicators can be used to attribute actions to actors, the amount of scientific study in this area to date is limited and the results are far from definitive. Furthermore, the characteristics used in attribution are usually tied to time sequences and interaction sequences. For example different keyboards produce different error rates and error types for typists with different training and experience, but if all we have is a spell corrected end document with no data entry sequences, these errors will not appear in the data used for analysis. This is often a basis for challenge and in some cases it is highly successful.

Physical evidence can often be tied to digital evidence. For example, an online credit card theft may be challenged, but if the credit cards stolen resulted in purchases delivered to the defendant’s address and the defendant did not question these items, try to return them, and is using them, the computer evidence of the theft may be hard to discredit. On the other hand, the lack of a nexus with real-world events should lead to a serious question about validity. Just because someone wrote about a credit card fraud scheme does not mean they perpetrated one. Even if their computer was involved in one that is similar to their writings, the lack of a physical nexus is a potentially fatal flaw.

Means, motive, and opportunity apply to the digital world, as well as the analog one. If the evidence shows a level of expertise in developing and hatching a scheme, and the individual on trial does not have the necessary expertise, the means does not exist. Digital systems are complex and a great deal about the knowledge of individuals is often revealed by the audit records and software present on the system under examination. Opportunity in computers does not always require presence. Because of the potential for telepresence and programmed interactions in networked computers, analysis of events in a computer do not always tie the individual to the events or the events to the presence of any individual. Making or missing these times is commonplace.

Just because a computer says so...

... doesn't make it so.

This is perhaps my favorite basis for challenging the analysis of digital forensic evidence. The seeming infallibility of computers leads many to believe that if a computer says so it must be so. People are increasingly realizing that this is not so, but the point still must be made in many cases. The sources of errors in computers are wide ranging. From computer viruses that leave pornographic content in computers to remote control via Trojan Horses that allow external users to take over a computer from over the Internet, to just plain lies typed in by human beings, computers are full of wrong information. In some studies of data entry errors, rates on the order of 10-20% are common. That is, one-in-five to one-in-ten data entries in a typical commercial database are not complete, accurate, and reflective of reality.

Overall summary

The number of ways that digital forensic evidence can be challenged is staggering and many of these challenges can be successful in the proper circumstances. But a competent digital forensics process and examiner can avoid successful challenges by diligent efforts and thorough consideration of these issues.

Avoiding all faults is impossible, but almost all failures can be avoided by prudent efforts. When faults occur, they may or may not produce failures. And some failures are recoverable while others are not. At steps in the process where faults lead to failures that are not recoverable, special care should be taken to avoid these faults.

Limits of budget, training, tools, and simple human errors have many effects on the challenges to digital forensic evidence. Those who wish to use this evidence should take note of the need for overcoming these limits within their organizations.

The legal process is complex and combines rules and human judgments to make determinations about admissibility and weighting of digital forensic evidence. This is not a legal text, but I have endeavored to include some background in the text and in the endnotes to help address the issues of admissibility and weight. [20]

Those who seek to challenge such evidence have an equally daunting task because of the widespread perception of computers as perfect which leads to excessive belief in the content within them. But as more and more people have identities stolen, computerized records create financial problems, computer failures cause missed flights, and fraudulent spam emails are sent to them, this will change. Planting the seeds of distrust of computers and computer evidence is the basis of any challenge, and these seeds exist for those who seek to find them.

When those engaged in the forensic process miss or make, through accident or intent, it is the job of those who see these faults to point them out and act to correct any failures that may result. A healthy forensic process seeks to poke holes in its own system in order to improve it and seeks ways to compensate for holes it identifies. Hopefully these challenges will be met by those in the digital forensics community so that these challenges never have to rear their ugly heads again.

Questions

Here are some questions you might like to try to answer to get a sense of how well you understand the issues in this chapter

1. How many of what sorts of faults can be made without resulting in a failure?
2. How can an expert witness clarify what constitutes probative vs. prejudicial?
3. Are cryptographic checksums necessary and sufficient to demonstrate the integrity of evidence, and if not, why not?
4. If you fail to identify evidence and it is never seen by the judge or jury, does it matter to the case?
5. How can you tell if you have enough expertise to be able to perform a particular activity with respect to digital evidence?
6. If you have a new technique, how do you make it suitable for use in a legal setting?
7. What do you do to calibrate the tools you use for digital forensic evidence collection and analysis, and what methods apply to which tools?
8. Given that you identify a fault in your evidence that has not been realized in a failure, what should you do about it if you cannot undo the fault?
9. In doing a digital crime scene reconstruction, how do you identify and document the things that your model is accurate and inaccurate about, and how far do you go in trying to make it accurate in which areas?
10. What is the difference between a make and a miss? Give an example of a make that generates a false negative.
11. Give an example of a miss of identification of ordering that produces a false positive.
12. Write a computer program that does an analysis of timing with location to allow the ordering of events from audit trails

in two systems not using a common time base to be seen, and explain why it works, its limitations, and the sorts of faults in process and resulting failures produced by it.

13. If an intentional consistency fault is found in an audit trail, and the only people with access to that audit trail were not otherwise implicated, when would this produce a false positive or negative?

3 Mechanics of writing expert rebuttals

Lawyers have different views on what they want to see in expert reports, but the formats and content are often similar. I have a preferred approach and, while I don't argue over it, I generally try to use it whenever possible. I do this because I feel that it brings advantages to the reader trying to understand what most people would consider dry technical information. After all, if they can't read and understand it, it has little or no value.

An outline of a report

Given the choice, my reports are outlined approximately as follows, with the italicized portions more or less standard across all reports:

- Cover page (this is mandated by the legal system)
- My background relative to the matter at hand
    - *My name is Fred Cohen and I have been asked to write a report discussing some of the issues involved in this matter. Specifically, within the limits of available time, I have been asked to provide opinions related to the processes and procedures used in this matter, identify some of the potential alternative explanations for asserted observed events, evaluate the assumptions, analysis, and conclusions put forth by other experts, discuss standards of care with regard to the issue at hand, discuss chain of custody, due diligence, preservation, and forensic soundness issues associated with digital forensic evidence and investigations related to this matter, discuss [...] issues related to this matter, and review details associated with [...] and other related specifics.*
    - Case-specific background is given a few sentences per area of expertise expressed in the report
    - Education is listed
    - Publications and products of special interest to this matter are given
    - Recent positions are listed
    - Editorial boards are listed
    - Honors and awards are listed

    - I generally add “additional details” stating:
        - *More specific details of my work as it relates to the issues in this matter are provided in the addendum to this report and in specific parts of the report where they are applicable.*
- Summary of my opinions in this matter
    - This generally starts by outlining the matter at hand in a few sentences along with the specific issues addressed in my report. Then it summarizes my opinions in strong but accurate terms ordered as in the detailed section of the report, and provides a summary and/or conclusions based on these opinions.
- Details of my opinions in this matter
    - This has all of the details of my opinions along with their basis provided with as detailed a set of facts as I have to support them.
- Compensation and prior testimony
    - *Legal matters I have testified to in the last X years*
        - The number of years and details of what have to be provided are dictated by the particulars of the court you are in and the type of case.
    - Compensation
        - *In this matter I am being compensated at a rate of $XXX/hr and my compensation is in no way related to any specifics of what I may testify to, my views on any of the issues in this matter, or contingent in any way on the outcome of this or any other case.*
- Materials considered
    - Case specific materials are listed here, usually in bulleted form. I sometimes use sentences to describe things not given to simple titles.
    - Personal knowledge, education, and experience
        - *I relied on my personal knowledge, training, education, and experience regarding [...], and other related areas.*

- Signature
- Addenda

- Publications
- Patents
- Software products developed
- Other details that are relevant

If additional reports are required, I tend to include information from the previous report by reference. The initial paragraphs start something like this:

> *My name is Fred Cohen and I have been asked to write a supplemental report reviewing new information provided to me on or about [...] and relating to some of the issues involved in this matter. Specifically, within the limits of available time, I have been asked to provide opinions related to [...] in the context of my previous report in this matter.*
>
> *The background provided in my previous report is also applicable to this report and the additional details provided in the previous report are also applicable to this report. In addition, the following background information not provided in the previous report is also relevant.*
>
> *Since my last report, I have also... [include new papers published, new positions, etc. as appropriate]*

In the first person

In the scientific and research community, there is a general principle of writing in the third person impersonal style. As a result, many authors have perfected a skill at stating things so as to eliminate themselves and their contributions from their work. There is an old saying (due to Isaac Newton I think):

> *If I have seen farther, it is by standing on the shoulders of giants.*

In writing reports and journal articles for scientific publications, this is not only common practice, it is largely mandated as part of the style requirements. The same is true whether the paper is written in a journal of psychology or an article for the Institute of Electrical and Electronics Engineers (IEEE).

Academics use citations, often provided within the text as only a number within square brackets (e.g., [1]) or as a partial name with a year associated to it (e.g., [Cleo 98]).

While these methods are appropriate and mandatory in the academic journals of the world, in the legal world, things are somewhat different, and many scientists and others who work between the two worlds have difficulty in switching back and forth.

I personally prefer writing in the first person. While this may be viewed as egotism in the scientific community, in the legal world, when I present something I have personally observed as a fact, stating it any other way is problematic at best. Consider these two pieces of testimony:

> *I saw the man in the gold jacket shoot the woman in black.*

compared to:

> *The woman in black was shot by the man in gold.*

The former leaves no doubt and no question about the facts being asserted and who is asserting them. The latter leaves many such questions.

In the former statement, the basis of the statement is clear. The person making the statement personally observed the events as they took place. I saw this – period – no doubt whatsoever.

In the third person, there is all sorts of room for doubt. How do you know this? Did you hear it from someone? It is a conclusion based on some set of indirect indicators? How reliable is this? What is the basis?

Now I don't claim that everyone will agree with me on this point, and perhaps it is just a matter of style, but in legal documents, when I know something, when I do something, or when I see something, I write it in the first person personal form so as to leave no doubt as to where my view comes from and what the basis for my opinion is.

Time is of the essence

In legal matters, time is often short. For digital forensics experts, this means that they have to be quick and careful. There are various legal and case-specific reasons for time crunches. One of them is that all legal work has deadlines associated with court dates, the mandate for a speedy trial, and legal deadlines on how long sides have to do discovery. In many cases, an expert ends up engaged late in the process, and in cases where digital forensic evidence is being challenged, as a rebuttal witness, you can't really get started until there is something to rebut.

In one case I worked on, I was told about the matter at hand on a Thursday morning, got a Federal Express package at 8AM the next day, and had to provide information to help in a deposition that was underway three time zones away as I received the package. After that information was provided, we had all weekend to figure out what happened. The following Monday, a phone call provided the attorney with two possibilities; either the evidence provided was not forensically sound, or the dates of events were nothing like the dates they had believed to be the case. After a short exchange on the phone, the matter was cleared up. The evidence was not forensically sound; it was an inaccurate reconstruction with misleading information. The information revealed then changed the whole theory of the case.

My record for the longest report done in the least amount of time was a combination of an original and supplemental report, totaling about 240 pages, and written in a span of ten days. It included reading about 500 pages of depositions, reports, and other paperwork, extracting quotes out of scanned documents, doing analysis and commentary, and turning it into a formal report. A few days later, I was deposed in the case, and the opposing counsel, asked me to describe the totality of my opinions in the matter, figuring he would test my memory. About 20 minutes later he tried to interrupt my description, and was told that I had to be able to answer the question. After about an hour, I finished summarizing the 240 pages of content and indicated that there were many particulars that I had left out, but as I recalled more of them throughout the day I would bring them up.

We all make mistakes – don't compound them

Nobody is perfect, and I certainly am not. In the course of doing supplemental reports or preparing for depositions or trial, we (the lawyers, my associates, and I) sometimes find mistakes in previous reports. When we do, I am not shy about them. In my view, basic honesty demands that you identify your own mistakes when you find them and try to undo any harm that might result from them as soon as possible. I try to put these up front in the start of the summary report of the supplemental and within the report at the place where it is relevant. Here is an example from a recent report:

> ***An error in my previous disclosure***
>
> *In my previous disclosure, I made an error in reading page 8 of 26 in Exhibit A of the response to court ordered interrogatories with regard to the 127.0.0.1 address. As a result, in that disclosure, I incorrectly indicated that this may have implied that some of these emails may have originated from a user at XXX. In reviewing the material in this deposition, I again looked at page 8 and determined that I missed another address between the XXX receipt header and the 127.0.0.1 IP address header. I have reiterated this correction in the body of this disclosure where applicable.*

Of course it is far better to never make a mistake, but don't count on it. And don't assume that the lawyers will ignore past mistakes. If the case is important and they have the resources to investigate, they will find every mistake they can from your past and discuss it in court, in front of the jury, to point out that you were wrong then and could be wrong now.

In a well publicized case involving a well known scientific expert, a medical doctor and forensic examiner, previous testimony about a defendant who was found guilty at trial and subsequently proven innocent by DNA testing was brought up in front of the jury. It was devastating to watch. In the first O.J. Simpson trial, a mathematics professor made a calculation error about a probability and the defense lawyers shredded his credibility with the jury as a result. Try to consider the possibility that you could be wrong.

Hedge in an honest way

In some sense, when you write 250 pages of reports within a few weeks, there is no realistic possibility that you will get everything right. But the biggest mistake you can make as an expert is to assume that you know everything or that you are always right. That is why you should always hedge your views, in an honest way.

I use a set of commonly understood phrases in my writings, and many people view these as hedges, which is exactly what they are. I seem to recall that it was Sam Clemens who is quoted by Chester L. Karrass as saying:

Unless you're damned sure, say "I reckon"

I tend to say things like "To the best of my recollection..." and "Based on the information available to me at this time...". Yes, this is a hedge, but it's not a way of sneaking around ever being wrong. It is a simple recognition that the nature of the legal and analysis process, as the nature of science and life, is not absolute and certain in every way. The expert who believes that they know everything is not an expert at all. Or to quite quote another famous author (I think it was Arthur C. Clarke's First Law):

"When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong."

I have seen case after case where people took entrenched positions claiming that they had absolute and definitive knowledge of things when they had no such knowledge at all. This is even worse when done in conclusions based on speculative facts. As a researcher and scientist, I find it offensive when someone who is claiming to be an expert presents things that are not facts as if they were facts to a trier of fact who doesn't know any better. I feel a special responsibility to help clarify the truth of the matter and the limits of these opinions to the lay audience who is making judgments that will affect other peoples' lives.

It is simply unacceptable to claim to have perfect knowledge when you don't in fact have it, or to claim to never make mistakes when we all know that people make mistakes. And to avoid doing this myself, I present my uncertainties and the basis of my views to those I present the information to.

Back to the method of hedging for a moment. I recently completed a forensic conference paper on floppy disk recovery and analysis of the recovery methodology and results of a recovery. In doing so, I got back a review from the peer review process that said something like:

> *"Put a stake in the ground and make a definitive statement"*

To me, this indicated a reviewer that hasn't ever had to go to court and testify about a matter that affects peoples' lives. I would love to have perfect knowledge, but I don't. I only hedge when there is a reason to hedge, but when I do, there is likely a good reason, like that I don't know about something that might be relevant.

A good example of a case when I hedged to a positive advantage was in a case where there were many elements of an expert report that I was not commenting on at the time. I put in a fairly common sort of response – something like this:

> *"As far as I am aware at this time, this is not an issue in dispute."*

Later on in the case, it turned out that this became an issue that was disputed. I also later encountered additional evidence under a subsequent discovery response that indicated that, not only was it in dispute, it appeared to greatly favor my client's situation. As a result of my previous hedge, I was able to put forth the new information without having to correct a previous over-reach.

A final note on hedging and ego. Most people are egotistical, and, like most people, I take pride in doing my job well. But in legal matters in general, and in challenging digital forensic evidence in particular, it is vital to try to put your ego aside for formal documents and formal testimony. Remember that you are, in essence, claiming that the expert on the other side was imperfect in some way. They will likely be very frustrated by the things you say, and perhaps have their egos bruised by it. They will want to get even and will point out anything they find fault with. If you overstep, they will, and should, severely malign you for it, in public.

Present the basis with the conclusions

I have seen quite a few expert reports where the experts simply claimed to know something and didn't share the reasons that they claimed this knowledge in their reports. This is a variation on what I call "secret science", and I don't believe it has any place in legal matters. On a factual or logical basis, it is hard to challenge an expert opinion when the expert doesn't present the reason for the opinion. While they might think that this is advantageous in that it precludes anyone countering their opinion or provides them with a tactical advantage when they spring the reason on you later, I strongly disagree. My personal opinion is that such reports and opinions should never be allowed to see the light of day. And when I see them, I have a fairly standard approach to countering them. Here are some typical responses that I give:

> *These claims were without basis in fact and they failed to perform the activities required to support such claims with such facts.*
>
> *There is no basis for asserting that ...*
>
> *I cannot tell which of these are correct, and neither can ...*
>
> *We are asked to simply believe that whatever it indicates is accurate without any basis for that belief.*
>
> *The lack of a basis in fact or clear demonstration of the link between the Defendant and the evidence...*

And here is a direct quote from one report:

> *The efforts Plaintiff undertook appear to have little or no scientific validity or basis to support the conclusions they drew or the charts and other results they produced. They cannot tell us how to reproduce these results, which seems to me to violate the basic principles of the scientific method. They provide no error rate information for their procedures, and since they have not described the procedures, they cannot be said to be generally accepted. In fact, their results have been contradicted by examples.*

The list goes on and on, but the basic principle remains the same. In presenting expert testimony about scientific matters, the basis has to be presented to a level of clarity that would allow an independent expert with proper background and knowledge in the scientific field to reproduce those results.

The structure of my report format is designed to allow me to present both the basis for my conclusions and the conclusions, and to do so in a manner that allows the reader to read the conclusions in a summary form and then correlate them to the basis by looking in the detailed coverage. To me, this provides the best of both worlds. First they get the results, then they can read further to get the full details of why I express each opinion I express. In most cases, the detailed portion of the report includes exactly the same things stated in the overview of my opinions and in more or less the same order. This allows them to be tracked against one another so that it is easy to verify what I am saying and why I am saying it if you want to, but you can still read the summary without getting mired in the details.

Do simple experiments to disprove things

I am a big fan of experiments, and an even bigger fan of simple ones. When I see a report that makes a statement that I think is too broad, I have a tendency to devise a simple counterexample and demonstrate the error on the spot. As an example, when someone states as a definitive result that they looked up an IP address, associated with a domain name, and that this proves ownership of the source of content, I have a pat response. I look up one of my old IP addresses, 204.7.229.1, which even now after not being used to host my domain for something in excess of ten years, still reports “all.net” using common techniques:

```
> host 204.7.229.1 68.87.76.178
Using domain server: 68.87.76.178#53
1.229.7.204.in-addr.arpa domain name pointer all.net.
```

This is slightly abbreviated for space. In an actual report I would have the exact command typed and response given and then look up all.net to show that they mismatch. This sort of simple experiment clearly shows that the assumptions made by the expert on the other side are not correct.

In fact, there are many such simple experiments that disprove many such overly broad sets of assumptions and conclusions based on them. And in large part, that is what challenging digital forensic evidence is all about. If the challenge is to be successful it has to show that the assumptions, methods, or conclusions of the other side are wrong in some way. The best way to disprove a theory is to show a counterexample, because refutations prove that a theory is wrong while confirmations only show that it was right in one more instance and not that it is always right.

As a rule of thumb, the best disproof is to use the very tool used by the other side and show that the conclusions they are drawing with respect to that tool are wrong. The next best disproof is to use a different tool and show that the tools disagree and then show why your tool is right and theirs is wrong. You should ask how they used the tool, how it was configured, what system is was running on, and so forth in order to be able to reproduce their experiments and identify errors to the extent that this is possible. If they don't provide information on the tool they used or how they did it, make that the first issue and then choose your own tool to do your experiments.

Tools, in order to work right, have to be calibrated and tested. For example, if they use Wikopedia for a definition and you don't agree with it, explain how unreliable and non-authoritative that tool is. Go to a more authoritative source and compare results. The most authoritative source is the original source. The more distant from there, the less likely the source is to be accurate. I use myself and my sites as a basis for comparison whenever possible because I can be more authoritative about myself than anything else. These experiments allow you to refute things that can be refuted and have little risk in terms of not refuting claims made by the other side.

Simple experiments can often be done quickly and they can go a long way toward breaking down the credibility of experts who make excessive claims or use assumptions that are not valid in their reports and testimony. They are compelling demonstrations that, if done well, bring clarity.

The report is the draft I am working on

In legal matters, information exchanged between a lawyer and an expert who ends up testifying is subject to discovery. That means that anything you write down along the way could be used as evidence. Many technical people have complex jobs to do and imperfect memories, and this means that they have to write things down in order to get them right and be systematic in their approach to getting at the issues. When these notes are taken at the time of the activities, they are contemporaneous notes and are considered by most people to be more accurate and reflective of the events as they took place than after-the-fact reminiscences created out of the fog of human memory. This brings a seeming contradiction. How can you do your job well and accurately reflect what was done while not keeping contemporaneous notes that might inaccurately reflect on the results you ultimately generate?

Fortunately, modern technology saves us again. Or it may... Over the years, I have done a lot of work on computers, and there are different people who work in different ways. I have no personal preference for how other people work, but in my case, I prefer to keep only the document I am working on and to not do anything special to retain prior versions, other than to keep backups of the current version as I go. I do this by copying the files I am working on to a backup disk drive and a USB drive, and I often do this several times a day. This is not something special I do for digital forensics cases. I do it for all of my work, including the source to this book and software that I write. I do these backups often because I don't like to lose work, and I don't keep old copies of reports around because I end up confused about which is which. When I print drafts, I read and mark them up, and as I make the changes, I save them to backups and shred the earlier drafts. Again, I do this because, when I don't, I end up with piles of paper on my desk and I end up making mistakes and unchanging things I have already corrected, sending out the wrong version, and re-editing what I have already edited. It turns out that this works well for doing forensic work as well as other work that I do, and as a result, that's what I stick with. I find it a good habit and one that allows me to do my work well.

With that solution in place for retention of old drafts, we still have the issue of contemporaneous notes. I don't know about you, but I cannot remember many things for very long without writing them down. I used to write everything down in paper notebooks, but I found that this is often even worse than not writing them down, because I often cannot find the right notebook or the particulars of a particular activity in the notebook. That's because the notebook is arranged by time and I may work on many things at once. I also have to keep different issues separate because of client confidentiality on the other work I do. Again, technology comes to the rescue. I take contemporaneous notes in the report documents I work in, and keep copies of programs and other similar things done on each case in a directory created for the case. I also use a spiral binder to create a “Case book” that has all of the inbound paper associated with the case, time records of my work, and correspondences. These are put in a standard legal pouch for each case and labeled as such. If there is a significant amount of data, I burn it on CD-ROMS and put them in the folder as well. This allows me to have everything in one place (one on the computer and one in physical form), provides enough redundancy to assure that I don't permanently lose anything significant, and lets me take any contemporaneous notes right in the report as I write it. I can document the precise time that I did something, the exact thing I did, and the exact results that it produced without fear of loss or the creation of multiple copies of handwritten notes strewn all over my office. I can bring it all with me when I go to meetings without problems, I can lock it in a media safe for safe keeping, and if copies are required, I can easily make them in an orderly manner.

By following this process and methodology for doing my work, both in regard to legal cases and other work that I do, I have the best of all worlds. I have definitive copies of the information I need, made at the time I did the work, and without any problems associated with retention and disposition. I can produce them at any time, systematically destroy them if ordered to do so by the courts, as is often the case in civil matters, and retain client confidentiality, which is fundamental to my consulting practice, while retaining reliability, and do so at a reasonably low cost.[18]

Use analogies within limits

I like to use analogies in reports when I can find good ones, but I try to be careful in that analogies are rarely perfect. If you use an analogy without being careful, you could risk being seen as a fool. As a result, when I use analogies I tend to explain the extent to which they really apply. I also tend to be somewhat careful in their use because I don't want to seem like I am talking down to the reader.

A good example of an analogy that is sometimes useful is the analogy of computerized documents to physical documents.

Computerized documents are very similar to physical documents in many ways. Directories are similar to file folders, drawers, and cabinets. But the analogy only goes so far. A set of directories 20 levels deep containing a mix of programs, video files, data files, and pointers to outside URLs is really quite different from a file within a file drawer within a file cabinet within an office within a building within a campus within a city, and so forth.

The key to using analogies is to be careful in drawing those analogies to keep them relevant to the matters at hand and not to extend the analogy to places it does not belong. If a paper file is like a computer file, what is a file pointer, a bad block, a file allocation table, and a disk head? When I "steal" a computer file, how come you still have a copy? When I talk about finding a needle in a haystack, how big does the haystack have to be to be the equivalent of searching for the contents of a file within a terabyte disk, and how fast are the people looking compared to a modern computer?

Analogies bring relevance but they also break down. So when using them, you have to be careful. And of course, as someone challenging evidence by another party, you should know how to take advantage of false analogies. The best technique I know is to show a case in which the analogy breaks down, use it to conclude something ridiculous, and then tell the court that you don't mean to imply the ridiculous thing, but rather that the analogy is invalid.

Why should the standard be lower?

The last thing I want to bring up in report writing is a technique that I like to use to compare the standard used in digital forensic evidence to the standard used in other evidence. The sad truth is that the standard often applied to digital forensic evidence is lower than the standard used for other sorts of evidence, and this is because the experts testifying fail to point out the inadequacies and imperfections of the field. Here is an example from a recent case:

> *Suppose Plaintiff was suing over a faulty traffic light in front of her store, and that she had a closed circuit television (CCTV) camera that pointed toward the intersection and took pictures once every few seconds, and had a contractor with a CCTV camera across the street pointing at the same*

> *intersection that also took pictures every few seconds. When she knew she was going to sue, suppose she started crossing the street time and again and collecting pictures of post-accident injuries she received, but she failed to secure copies of the pictures from either of the cameras, even when she started the legal action and had an obligation to secure such evidence. Just because the evidence in this case is digital evidence doesn't eliminate the need to secure and produce known available evidence or the obligation to do so.*

This example combines a physical analogy to digital surveillance systems and audit trails within computers and a standard of care with regard to spoliation of evidence that is widely known and accepted in the physical domain. It asks the relatively simple legal question of whether the standard of care for securing digital evidence that is known to exist is lower than it would be for physical evidence. If the asserted position is right, the answer is no and therefore a lack of diligence in securing the potentially exculpatory digital forensic evidence will be considered spoliation and subject the plaintiff to sanctions.

Here's another example that focuses in on an investigator or expert for the other side who has, in the view of the expert writing the report, failed to meet the requirements of due care.

> *Suppose this was a blood test. It was performed by a technician with no training in the techniques being used and who did not know how the tests worked, how samples could be contaminated, how results could be misinterpreted, or how to calibrate or test the equipment being used. They did not clean the equipment before or after use, did not take notes on the procedures they used, did not have defined procedures, and could not tell us what tests they performed or how they performed those tests. I don't see how results of such tests or procedures could be allowed to be presented or how the results could be meaningfully presented to a trier of fact. Just because this is digital evidence does not mean we should ignore standards of due care.*

This analogy and appeal to a standard of care is designed to clarify things in terms of exert testimony and technician behaviors that would be expected in any case where similar technical evidence was being considered relative to a medical claim.

The problem is that there are no real standards of care for the protection of digital forensic evidence, because the courts have not yet ruled on definitive cases, laws are not yet definitive, and the field is changing at a rapid pace. I am not rallying for definitive standards, and neither is anyone else I am aware of, but it would be nice to get a better sense of the requirements both for civil and criminal cases.

The differences between civil and criminal are fairly complex, because of the differences in standards of proof. In a criminal case, proof beyond a reasonable doubt (which you might also think of as to within a reasonable certainty) is required. In a civil matter, the preponderance of the evidence is the standard. To me that sounds like there has to be more evidence on one side than the other. This then becomes a contest in throwing lots of theories up against the wall and seeing which stick for the plaintiff and tossing doubt around here and there by the defendant to see what can be knocked off of the wall. The approach of challenges to a wide range of questionable evidence may tend to become one of volume over quality unless the courts take a hand in controlling the situation more clearly.

As a result, it is incumbent on the challenger of digital forensic evidence to make all potential claims and provide guidance as to the basis and strength of each of them, preferably in terms that are understandable by others.

When in doubt, check it out

Another technique I use when I write reports is to check things out with other people I know, who have expertise, and whom I trust. There are a few folks that I work with on a regular basis; fellow professors in my graduate program, consultants I have worked with for many years, and “uber-geeks” who, like me, spend a lot of their life in specialized areas. I often don't share any of the context with them and read or discuss issues with them over the phone. And they do the same with me when they are working on a complex issue that they think they know but want to be certain of.

I also review certain terms that I use and seem to remember from way back when. For example, in one report I was discussing issues of flux density in electromagnetic media and how reductions in flux density could reduce the current to below the threshold required to trigger a transition. Now I have a Ph.D. in electrical engineering, and I should well know what flux density means. And in fact it turns out that I do. But even though I thought I remembered it, I also know that there are people in the world who know more about electromagnetics than I do. Maybe it was out of fear of a better expert on the other side, but I decided to check it out just to make absolutely certain that I didn't use a wrong word or use a word incorrectly. I went back and read my text books from 30 years earlier, called folks that I know who know more about it than I do, and checked it against refereed journal articles. I didn't charge the client for this extra time because I think that is my responsibility if I am going to claim to be an expert. I have subsequently written a refereed paper on it for a scientific conference, and this helps to authenticate it in the scientific community. If I were called to testify about it today, you had better believe I would look it all up again so that, if the best expert in the world was the lawyer asking me questions about it, I wouldn't know as much as them, but I would not say anything that wasn't strictly correct and accurate.

Questions

1. In writing your report, will you take the same approach outlined in this book, or do you have a different approach? If different, what is it?

2. To what extent does the adversarial nature of challenges to digital forensic evidence create ethical problems and how do you address those problems in writing your report?

3. If you have to rush, how does this cause you to make mistakes, and how should you keep yourself from making mistakes associated with the lack of available time?

4. If the lawyer you are working for does not tell you everything about a matter and you give an opinion that turns out to be wrong because of what you were not told, how do you correct this problem? How do you write your report so that this does not destroy your future credibility?

5. In correcting an error in a previous report that you made, discuss the virtues and problems with opening your next report with the correction and describe how you would go about making such a correction, if you would make it at all.
6. How much does it sound like egoism to write in the first person? Use this book to judge it and explain how you will do better in your reports than I have done in mine?
7. If a report makes claims that are not substantiated, is it better to prove those claims wrong, assert that there is no basis for them, or do both?
8. In doing simple experiments, if those experiments support the contentions of the other side rather than refute them, to what extent is it your responsibility to include those results in your report? Should you do experiments only on things that you know to be problematic? Should you do experiments on things you think are likely to be right? How do you decide what to experiment on?
9. Who do you check things out with? How do you do this?

4 Case studies

There are many particularly challenging issues in the field of digital forensics, and they require specific expertise to address. It is quite common for those who act as experts or testify about matters related to this evidence to lack the underlying knowledge required to authoritatively make the case for the things they ultimately claim.

In these case studies I hope to clarify particulars by showing how challenges are made, examples of the sorts of claims people make that are over-reaching, and the sorts of things that seem logical but are actually fallacies. The case studies and examples presented here are extracted from real cases I have worked on and include some of the particulars associated with these challenges. I won't identify the specific cases and I will often remove or alter the specifics to protect the parties involved, but this should not reduce the value in terms of understanding how these matters go and how false and misleading claims may be challenged.

Of course this is not quite the whole story. Everybody makes mistakes, and to the extent that challenges are made, they may or may not be valid. Those who testify as experts may or may not have access to all of the information they need to make accurate claims. For example, it is common for those who are on one side or another of a case to limit the information available to experts to assure that they will have opinions that meet with their need to support their clients. That is the nature of the legal system as it exists today, and as such, those who seek justice have to be careful about what they say as well as how they say it.

The form that challenges to evidence usually take are reports. These reports are normally rebuttal reports. As such, they don't act to make claims about evidence in most cases, but rather to challenge claims made by others. At the same time, they can sometimes draw conclusions that the expert on the other side missed. But in large part, by their nature, they are made up of item by item identification of flaws in what another expert asserts. They are normally best presented in detail and with the statements of previous experts explicitly aligned to specific challenges.

Challenges are typically told as organized stories challenging the elements of the case that are at issue and that can be reasonably challenged by digital forensic evidence. They may follow the logical flow of the case, the sequence of events that took place, or another structure such as the structure used to describe challenges in this book. This selection is a matter of presentation and, to be effective, it should be selected for the specific case based on the situation.

The lawyers involved in challenges to evidence presented by the other side are usually trying to challenge very specific issues, and to the extent feasible, they may want to limit what their expert challenges and what areas they get into. This is because legal cases are oppositional in nature and because legal precedent and doctrine mandate that if one party opens a door by addressing an issue, the other party can “walk through” that door. If the other side has missed something and the forensic expert in rebuttal finds it, they cannot necessarily present this, because it does not rebut the claims of the other side. Strategically, the lawyer may want to avoid specific issues because they open the door to the other side going after issues that were not permitted by previous legal rulings. This is the nature of the legal system, and if you are going to challenge digital forensic evidence in legal matters, you have to accept these limitations and the fact that the legal world is imperfect. You can only do what you can do; tell the truth as you see it, based on the best available scientific evidence, experiments, analysis, and the facts available to you.

On the other hand, when you write reports, you should take special care to not jump to conclusions. After all, you are challenging others and their work, largely based on the fact that they have drawn conclusions that were too strong for, not strong enough for, or just plain at odds with, the available evidence. You must understand and assume that there is more evidence out there and that your conclusions are only based on what you have seen so far. Indeed, to do your job correctly, you must consider that there could be more evidence out there that will directly contradict your current opinion, that could add information that would clarify the situation, and that could bring into doubt the information you already have. Case studies will help to point this out.

Don't forget the effects of time

One of the most common errors that people make in digital forensic analysis is to assume that computers are nearly instantaneous and that everything happens right away. This is a bad assumption because it is very often wrong, and particularly wrong in the context of networked systems and digital forensic evidence. One of the reasons for this problem is that digital systems keep such precise time. But precision, in terms of the number of digits retained, does not always correspond to accuracy, as in the correspondence to objective reality or outside standards. Two systems might each keep time to within microseconds and yet differ from each other by months or years. When time is an issue in the case, time should be carefully studied and identified as problematic when it is indeed problematic.

In one such case I worked on, a matter of a few minutes five years earlier was so important that it could have definitively proven innocence. In such cases, it is well worth chasing the details to ground and making as definitive a call as you can, based on the realities you encounter, and identifying remaining uncertainties.

People who use forensics tools, but don't understand how they work or the internal formats of the files being analyzed by those tools, commonly make mistakes in the way they use the tools or the way they interpret their output. The tools themselves are designed to make interpretation easy and present it as definitive, but interpretation is not in fact easy or definitive in many cases. And the place they make more mistakes than anywhere else is in the interpretation of time-related fields within files.

In one particular case, Microsoft Word document files were stored indicating sets of time for object linking and embedding (OLE) files within the document files. These OLE files each had pairs of date and time stamps that were being interpreted by the automated tools incorrectly in two ways.

1. The interpretation ignored the second time and date stamp that differed from the first by an amount of a few seconds.

2. The interpretation assumed standard time for all events when some of the the events took place in daylight savings time.

Because of the specific times involved, the one hour difference might have resulted in someone who was at a meeting during the time in question not being able to have been on their computer taking the actions in question. It would have been definitive proof of innocence if the tool output was right. The prosecution expert asserted that the time was wrong for some unknown reason and that it was irrelevant, while I was certain that, if accurate, it would have been very relevant and proven innocence. But as I undertook the detailed analysis I delved into the particulars of internal time formats in those documents and wrote my own tool to extract and interpret the raw data. My tool presented results in the time base as stored, which was, as I recall, in microseconds from Greenwich Mean Time (GMT) offset by a starting date of Jan 1, 1401, and not in any way dependent on daylight savings or standard time differences. The actual evidence, when properly interpreted, did not show that the file was handled in the middle of the Atlantic ocean or that the defendant was in a meeting at the time, and could not have done the deed asserted to have been done. Rather, it showed that the expert on the other side and their tool had made an error. My report corrected it and moved on.

A second issue involving time differentials in the data fields of a file was particularly problematic because, if the time indicated a differential between two events, it would have shown that not enough time was available for the events asserted to have taken place to actually have taken place, and the defendant would again have been proven incapable of acting in the manner asserted by the prosecution. Again, experimentation came into play because the information on how the fields were originated was not available. In this case, we did a forensic crime scene reconstruction and created mock ups of the actual environment at the time. Without going into the sordid details, it turned out that the time differentials were due to the difference between the time on a remote sever and the time on the system where the work was done. These time offsets could have been correlated with other audit trails to determine where the content was in fact stored, and could have allowed further evidence to have been sought, but it was too late for that, and we now knew that this was not proof one way or the other of the defendant's innocence or guilt.

The results were reported out and we moved on again. Time was of the essence, but digital forensic evidence would not bring clarity. At the end of the day, this was the only real evidence that anyone had, and the lack of clarity was itself clarifying. No evidence, no case. Or so you would think.

The defendant was found guilty, but on appeal, the case was overturned and a directed verdict was issued finding the defendant not guilty. This is quite unusual, and apparently it was because the digital forensic evidence was given credence by the jury even though the appeals court ruled that it had none. Challenges don't always work, but sometimes things work out anyway.

In another case, a common problem arose in the interpretation of times associated with emails. In this case, a tool that converted from one email format to another was used by the Plaintiff. The resulting evidence was not best evidence, in that it was altered from the original evidence that arose and that was available. This brought out the issue of spoliation. But in addition, because the conversion took sequences of things written by different mechanisms and replaced them with (key, value) pairs in a different format, some of the original evidence was lost forever. In particular, the reception time stamp from the receiving computer was gone and the conversion process replaced it with a different time stamp that was unreliable. This altered the apparent ordering of events. When analyzed in light of the specific tools applied, it became clear which fields arose from which sources, and only in that context could the data be properly understood.

In this particular case things were even more complicated by dates, but not times, of account openings and closings that had to be figured in. Since the computer time stamps were more accurate than the date information, many possibilities were feasible and, if this were critical to the case, each of the many combinations might have had to be analyzed.

The WayBack Machine and friends

I have seen a lot of cases lately involving people who went to Internet sites and viewed what looked to them like historical records, found those records to clearly demonstrate that something nefarious happened, and proceeded on that basis down the long and expensive road to a legal case. Of course for lay people to do this is to be expected, but for anyone who deals with forensics or is acting as an expert witness to do this, they had better have a really good explanation.

In order to demonstrate the limitations of this sort of technology, I have put up an example that should sway any reasonable person away from making these sorts of assumptions about these sorts of sites. To see my demonstration, turn off “JavaScript” in your Web browser and go to http://www.archive.org/. This is the home of the “WayBack Machine”. Enter the URL http://all.net/ into the WayBack machine entry area and tell it to go back in time. Select the first entry for all.net, the entry from way back in 1997.

Look carefully at the image that appears in the right half of the window – about 35% of the way down the page. View it in its own window to get a larger image. Hopefully, by the time you have gotten this far, you will agree with me that the WayBack machine should not always be relied upon as a source for digital forensic evidence.

If you haven't looked yet, look at the WayBack machine's page source. You will see that it uses the original URL of the page as the base page source for all references, specifically including images. That means that the pictures you see are the pictures from the current site, not from the site as it existed at the time when the image of the “html” portion of the site was taken.

Of course the WayBack machine and other similar resources, such as Google and other collections, are very useful tools. But as useful as they are, they are also potentially misleading. When in doubt check it out – which is to say – when involved in a legal matter make sure that you are right before you leap to conclusions or allow your clients to leap to them.

Calibrating your own tools

I check out and calibrate my tools, and I think it is very important to check out your tools. If you are going to assert that the opponent didn't check out their tools, you should check out yours. So how do you do this? ... I'm waiting...

In a recent case, I had to analyze a collection of emails. To do this I decided to purchase a converter that converts between widely used formats. While I can do the same thing in other ways, I have long been looking to get a single tool to automate the process, and since I often don't have the time to do it all myself, I decided to pay a company a few hundred dollars for the privilege of saving myself a few days of work. I used the tool, and it seemed to work just fine. But as I worked my way through the case, I understood that I would need to calibrate and verify its operation in some way, so I started to work to do so.

I engaged one of my long-time co-workers in the activity and we did a variety of tests. He converted known and generated email collections into different formats using different techniques and extracted the results using the tool to determine what it did and how. This included testing out different collections intended to model, as closely as possible, the actions of the tool with respect to the issues at hand. Did it resort the output or provide it in the order it originally arrived? Did it change any content? Did it add or remove anything? Did it change the ordering of headers? Did it use internal dates to replace fields not available in the original email? Did it take multiple headers of the same name and convert them into a single header of the same name? Which one did it use?

Without going into the full details, as I analyzed the evidence in the case, I started to have suspicions about how which outputs were generated, and eventually managed to get the programmer on the phone and asked specific questions. The program, even though it claims to be usable for forensic purposes, has peculiarities that make it more or less forensically sound for different purposes. In this particular case, if the tool's output was believed as presented, I had a clear case for proving that the emails were a complete fabrication. But before I was willing to make that claim, I did enough calibration and testing to show that I was wrong about that result because the tool was providing me with information that, while it was required to be in the output file to meet the output format requirements, was not what it would normally indicate if not processed through the tool. It's not that the tool is bad. It's that you need to know what the source of every output field is, in order to make sense of the output for forensic purposes. It's not in the documentation, and you can't get the source. So it's a combination of experimentation, calibration, testing, and contacting the vendor that gives you the real answer.

All in all, I spend as much or more time in calibration and testing than I do in analysis and report writing. And whenever I write something up, I then go through it. When possible I also have someone else go through it to find the places where I don't have definitive information that I can state as true based on my actual experience or testing. This pays off just about every time, and if you are going to seek the truth, you had better do a good job of this. This ultimately moves into the field of digital crime scene reconstruction, where the idea is to create a good enough reconstruction of the actual crime scene to model the things of import so as to be able to assert with authority that the statements you make are truthful and accurate to the real situation.

Just a bit more on calibration is warranted. I find that most tools are of only limited forensic value and that most tools do not present you with the proper ancestry for the information they provide. That is, in challenging (or presenting) digital forensic evidence, it is vital to be able to trace information you have to its source and through all of the processes undertaken that may have generated and/or altered it. It is impossible to know exactly what everything does in every circumstance, no matter how much of an expert you are in any specific thing. I have to examine my own code to make certain of what it does. If the other side has not done this tracing, it becomes your job to do it in order to show that they were wrong, and your job to point out that they failed to do this tracing or provide the information on this tracing as a basis for their conclusions. This actually segues nicely into link analysis...

Issues related to link analysis

In one case, a plaintiff was trying to make assertions about links between a defendant and a set of computers with similar domain names and IP addresses. The claim was, in essence, that the defendant owned a whole set of domain names and therefore was responsible for a larger overall scheme. The overall scheme in this particular case was related to sending unsolicited commercial emails in violation of US law. The indented portions are extracted from my report in this matter.

> *There are a number of assertions made by Plaintiff and Plaintiff's Investigators both here and in their reports and other documents I have reviewed with regard to this matter, that assert that associations between similar domain names and IP addresses indicate ownership or control of various computers, domain names, IP addresses, or other things by Defendant. I have some experience with the buying, selling, parking, and other matters related to domain names that may help to better understand this issue, and I present that information here to aid in bringing clarity to the general notions being presented and the specific assertions being made.*

The careful use of language in terms of couching everything within the bounds of the information available to me at the time is very important. There may be facts of which I am unaware that could dramatically alter my opinion. In addition, the description and reality of expertise is foundational in asserting claims with regard to expertise.

> *In the early days of the ARPAnet and then the Internet, domain names were not considered very valuable as there were many possible names and few people who owned those names. But in the 1990s, the explosive growth of the Internet changed this. For example, in or about 1994, when I acquired the all.net domain name, we tried to get the all.com domain name, but it was previously owned and there was nothing we could do about it. In the late 1990s, domain names became far more valuable properties, and individuals would purchase large numbers of domain names on the*

> *speculation that they would increase in value. Many of them did increase in value, and people made livings and small fortunes by buying and selling domain names. As industries started to embrace the Internet they also started to buy up domain names that were similar to their company names, trade names, and so forth, and again names became more valuable.*

It is often fairly important to put things in historical context if you know the historical context. This is for two reasons. First, it clearly establishes expertise, not just over the particulars of this case, but also over the subject domain in the larger sense. It also opens you up to being cross-examined on these things, so you must be careful not to say it unless it is true and the lawyers agree that it is not damaging to the matter at hand. The second reason is that, in this case, the history brings life to an otherwise very boring subject. By discussing the ebb and flow of things in history, you tell a story that is worth reading, and in a technical matter such as this, a "just the facts" approach is likely to put even the most technical of readers quickly to sleep.

> *When new areas open up, like mortgage refinancing, many companies that want to get into that business want names that are meaningful so that users can find them easily, so that the domain name seems related to their businesses, to protect their brand names and associations, and for other similar reasons. Companies and individuals who perceive this earlier than others often buy up large numbers of related domain names to protect the space of names or make money selling them to others. When they buy them, they tend to buy many related names, such as cars.com, car.com, carshop.com, carhop.com, carsell.com, and so forth. As they buy them, the providers that they buy them from typically provide initial Web and email hosting services, typically located on one of a small number of servers, and at very low cost. These domain names often share the Internet Protocol (IP) address, and each can be loaded with data as desired. The new owners often sell these domains to others for a profit. The result is that it is common for many similar domain names to be hosted at some point in time on the*

> *same server with the same IP address, even though they may or may not belong to or be populated with data from the same owner. In some cases, the sellers populate these domains with pre-existing data before sale so that the new owners have less work to do to get a usable site. Thus different owners may have similar content on the same servers under similar names without those owners being in any way related to each other.*

This particular case was about mortgage refinancing, so it must be included, but the analogy to cars is a common one used today because it is something that most people can relate to easily. The sharing of IP addresses and similar domain names is the thing that is asserted in this matter as the reason to believe that many sites were owned by the defendant.

> *As domain names fall into disuse, for example as a market shakes out, or as speculation dwindles, the domain names that are not paid for periodically, revert in ownership. There are businesses that buy and take ownership of these out-of-use domains and put them up for sale, typically hosting them on a small number of servers, at little or no cost. The result is that it is common for many similar domain names that fall into disuse to be hosted at some point in time on the same server with the same IP address, even though they may have come from many different owners or sources. The new owners often change the "home page" of the site but don't always clean up any of the other data, and as a result, some old Universal Resource Locators (URLs) may remain functional and others may not.*

Again, the URLs and other similar changes are related to the specific allegations in the case.

> *Any attempt to associate domain names, IP addresses, and other similar information and that fails to take this process into consideration, will almost certainly yield faulty results, create false associations, and miss real associations. It appears that this was done in this case with regard to the charts created by Plaintiff's Investigators and with their conclusions based on this work. While those charts seem to*

> *me to be unusable because of many other flaws in their construction and derivation, even if all of the technical flaws in the manner in which they were created were ignored, and even if they were redone taking proper notes and care to making valid associations between IP addresses and domain names and domain names and IP addresses, this would not solve the problem of identifying the reasons for those associations.*

In legal matters, there are often a wide range of issues involved, and you never know how people will perceive things. Where logically feasible, different aspects of the expert testimony should stand on their own as well as be additive.[19] For that reason, it is important to point out that these particular issues do not depend on other issues and to make it explicit so that someone who doesn't believe one thing or believes someone else is right with regard to one issue has the opportunity to believe you in regard to another issue. Unlike formal proof where a single point wins the day, legal battles are sometimes more like prize fights, which, barring a knock out, are scored by points. You have to make your points with clarity when you can.

> *The notion of guilt by association is clearly problematic in a legal setting, and in the Internet, it is very easy to create or use automated tools to generate recursive associations that rapidly grow to include almost any computer, company, individual, or IP address. Depending on how such tools are used, they can appear to include almost anyone in the wide net they throw. This is particularly problematic when trying to find a link between one person and another, one system and another, or any similar effort focused on finding links rather than exploring all links. As a result, special care must be taken in considering claims of association, statistics related to associations, or other similar methods, that don't do the very substantial research and clerical work to definitively tie specific individuals or systems to specific events.*

This is particularly important for several reasons. One is that there is a right to associate in most countries and the notion of guilt by association is generally viewed in a negative light. Who doesn't know someone who has been arrested? Does that mean that you are subject to being smeared with their crime? Just because you used Microsoft's email that doesn't mean that you are associated with another person who used their email to commit a crime or that you had anything to do with the crime. You can also relate this to the “Seven degrees of separation” notion that asserts that every person knows someone who knows someone, ... who knows anyone else on the planet. If we go too far in this sort of analysis, we will paint everyone with guilt.

> *It is my opinion that the conclusions, results, and information provided by Plaintiff in regard to link analysis are not reliable, not based on any scientific results, and not adequately supported to be relied upon.*

The issue of link analysis is particularly interesting, and I checked out my writing with an independent expert who has taught this subject in graduate classes before putting it in my written report. I did this by reading the relevant paragraphs to him and taking any criticism he may have had into consideration before making changes or deciding not to. If I had three such experts I would ask all three, unless there was no controversy left after the first two.

This link analysis problem is even more extreme when taken in combination with other problems, like the problems with time associated with the WayBack machine described earlier. For example, a tool used in creating an original Web page might have a copyright associated with its presentations. The same tool may have been used for building many Web sites and there may be an apparent link between the Web sites because of the common Copyright notices, URLs, graphic image file names, or other residuals of tools used at different points in their development.

Link analysis, like so many other analytical techniques, also has tools and techniques, and those tools and techniques also have their limitations. As always, traceability and definitive cause and effect determinations are the core issues and challenges.

Internet operations between companies

In the area of digital forensics, a lot of things today happen over the Internet, and a lot of folks who are relatively new to the Internet make a lot of mistakes in understanding how it works. There are tools that appear to be amazingly fast and accurate, but that are in fact, not authoritative, and that are often, just plain wrong. While this level of surety is fine for checking out a new toy you are thinking of buying, it's not so fine for use in court.

One area that seems to come up fairly often in Internet-related cases is how companies operate. Many naïve analysts seem to think that their version of logic and reasoning reflects the realities of the day. Never confuse reason with reality. People do all sorts of things that you may not understand the reasoning behind.

> *Another area in which Plaintiff's Expert seems to make a lot of assumptions is in the operation of the Internet and its use by companies of different sorts and sizes. For example, there have been expressed assumptions about XXX LLC as being too small to be used by Yahoo! as a source for sending emails, and there have been assertions that MSN.com, a Microsoft owned set of domains, could not send email from servers in China without the domain name server in China indicating that the IP addresses were part of MSN.com. The problem with these assumptions is that they are without any real basis and expressed as being based on logical reasoning associated with the behaviors of different companies.*

At this point, I introduced some additional background that was specific to the issue, and as I wrote the report, I updated the background in the front of the report to add a sentence or two introducing this area as part of my expertise. You simply cannot put a lifetime of experience or an autobiography into the introduction to every report, so beyond the basic background, I add the specifics for each case as needed. I like the approach of putting a quick sentence or two in the introduction and then fleshing it out in more depth in the body of the report where appropriate.

> *From personal experience, I know that many such companies have in the past, did in the relevant time frame, and still today, maintain servers that transmit emails from all over the World. This includes mechanisms that allow and support transmission of emails through other countries even*

> *though the messages originate in a different country, computers in other countries that have IP addresses that do not resolve to company domain names when looked up from the Internet, and gateway computers that operate through other companies and do not have the originating company's domain names associated with them. This also includes buying services such as email transmission and reception services through external providers including providers of substantially smaller sizes, the use of providers that are small and/or minority owned businesses to meet legal, contractual, or political requirements, the use of providers in other countries to meet similar obligations, and many other similar things to meet personal interests or obligations of various sorts.*

At this point, there are a lot of worthwhile things to say, and in this particular case, I may have gone overboard, but these were all issues that the other side had made assertions about and the background was relevant to the issues at hand. I also find it useful to describe some of the details about how things are done and why, so that, if questioned about it later, I have the basis in the report. I don't like others to make claims without backing them up, and so I want to back up my claims as well.

> *Some such companies also have global information infrastructures with substantial numbers of systems and mechanisms that have been exploited by malicious attackers so as to create automated distributed computer networks (often called botnets) that operate within, through, or between other countries and are used for purposes ranging from sending unsolicited emails to performing distributed denial of service attacks. Some such companies have insiders that use company resources for similar sorts of things, including selling use of those resources to legitimate and illegitimate outside businesses for their use in all manner of things, including the transmission of high volume emails. In many instances, I have observed systems containing scores, hundreds, and occasionally even thousands of instances of such mechanisms, including mechanisms that search through each others' contents to*

> *find email addresses and use those addresses to send out unsolicited commercial emails.*

Some of the information here was a bit musty in my mind, so as I wrote it, I went through historical records and called upon those I had worked with in these particular instances to make sure that my recollections were accurate. When there is question, I err on the conservative side and write things in such a way that they are true even if my recollection is off a bit. For example, in the previous paragraph, “thousands” is actually tens of thousands as I recall, but because I didn't have the facts in front of me, I didn't want to over-reach. Since tens of thousands is a number of thousands, I reckon it's true and correct.

> *It is very common for mechanisms such as these to gather lists of email addresses from within company computers that they invade and to augment those lists with other lists that they have, creating larger and larger lists. It is common to observe such computers sending out or attempting to send out large numbers of emails, and for those emails to contain very similar content. It is very common for such companies to avoid investigations of such systems and to minimize publicity surrounding their existence. And I have seen cases in which employees and contractors for such companies used information from such systems, including email addresses, to send out their own unsolicited bulk emails or to make additional money by selling the services of those networks to others.*

I find it particularly helpful to testify about things that I actually know about from personal experience and have observed, as opposed to reporting something that someone else reported. Lacking such personal knowledge, it is best to say nothing or cite the source as the basis for your views. At the same time, I have confidentiality issues in my contracts that prevent me from naming names. This is a weakness in my reports that the other side could try to attack. The way I generally avoid this, if it comes up in depositions, is to state that client confidentiality prohibits me from giving out specifics, and then point to several reports from the media from matters I have not been involved in to bolster my position without revealing client confidential information. I haven't had a problem with it yet, but I think it will become one if enough lawyers read this book.

> *It is also very common for large companies such as these to have workers located all over the World, and for the computers of those workers to be used to perform other tasks for those enterprises, including performing those tasks from home computers, small offices, and home offices. Many such companies allow those users to work from wherever they are, including hotels, conference centers, and private homes, and from those locations, they commonly have IP addresses assigned to them temporarily that do not have company reverse DNS information that reflects their employers. Many of these workers and offices send emails directly to recipients without going through corporate email servers and do so under the corporate email name. Many of these facilitates have their own outbound email systems and operate inbound emails through other providers, including but not limited to XXX, the provider used by Plaintiff in this case. In almost all of these cases, the inbound email servers reflected in the mail exchange (MX) records of the DNS system do not reflect the IP addresses from which outbound emails are sent.*

Again, this particular issue is specific to the case at hand.

> *It is also very common for such companies to have workers who perform services for them on an outsourced or other similar basis and who work from all over the world. Many of these users use their own home computers, computers owned and operated by third parties, computers operated from within office buildings, or computers operated from facilities owned and operated by others, and in most of these cases, those systems do not have DNS records that reflect the company that the workers work for, but rather the owners of the network or network providers whose facilities they work from within.*

As I go through this discussion, it is my intent to clarify the extent to which records on the Internet can or cannot be relied upon and for what. At the core of the Internet is the Domain Name System, and because it is a “best effort” service with no real authentication as to the underlying truth of things it asserts, it is problematic for all Internet investigations where the names or addresses it provides are asserted as incontrovertible facts. They are no such things.

> *It is also very common for large enterprises to outsource large volume operations such as call centers to companies that house multiple services for multiple enterprises and that emit emails with source information associated with those enterprises from the call center's network. Most such call centers have DNS records that identify the call center or its provider and not the enterprises that they are working for or sending emails on behalf of. Many of these call centers may also host many outbound email services on a single outbound email server, so that mails from multiple domain names are emitted from the same server. In many cases these servers fail over to other servers so that when emails from one enterprise are otherwise unable to be sent, they are rerouted through the internal company networks to other locations, sometime in different countries, and are transmitted from there.*

This is a test question. Did you spot the over-reaching in that last paragraph? I said something there that I would have corrected if I had noticed it in time, and if questioned on it I would immediately clarify that I didn't have the real data to back it up. It turns out that in the particular case it would not matter significantly, but these are the sorts of errors that are easy to make and easy to challenge.

The more you read of this, the more you may get a sense that the Internet is a gob of stuff that you cannot count on. That is largely true for legal purposes, even if it is fine for buying theater tickets.

> *For a wide variety of reasons, many companies use service provider email names to send some or all of their emails. For example, I use a number of email accounts at “mac.com” for people who work for and with Fred Cohen & Associates. Many people have email accounts at companies like Yahoo!, MSN, and elsewhere, that they use for many legitimate purposes. Many such individuals create multiple account names including account names that are very hard to guess*

> *or are generated by pseudo-random character sequence generators. For example, in order to avoid incoming spam, some companies and individuals create a new email name for every dealing they undertake so that any unsolicited emails or responses can be associated with the initial communication. There are commercial companies that offer services to create individual email addresses for every transaction undertaken, and some of these are seemingly random or unusual. Some companies have their workers create such accounts so they can have anonymity when working on sensitive projects, like mergers and acquisitions, where emails with company names might indicate relationships that are not allowed to be disclosed. And it seems highly likely that some people who send out commercial emails use large numbers of addresses like these for a wide variety of reasons ranging from anonymity to association of orders with senders or campaigns... Email addresses are also assigned to people from all over the world who have different traditions, native languages, and character sets. They also produce email addresses with names of all sorts that might seem strange to others from other places and cultures.*

I like to provide personal details to clarify that I have personal knowledge of this, and then back it up with things that other people have stated in public disclosures. In this case I cited a recent IEEE talk to support my statements.

> *There are also a wide variety of different mechanisms for anonymizing emails and other actions to reduce or prevent tracing of exchanged content. These are commonly put in place by privacy advocates for personal or political reasons, but they are also used by companies for diffusing attempts to do network intelligence studies on them, by individual users who are concerned about their privacy, and by governments and others who wish to provide cover for various operations they perform. Many of these anonymizing systems and many corporate outbound email systems use proxy services and similar functions to remove internal header information from emails so as to reduce the available*

> *information on internal network structure in order to reduce external intelligence attacks and similar sorts of activities. Some such gateways produce unique user identities to prevent or reduce misuse of their email systems. ...*

The problem of anonymity in the Internet and the inability to attribute actions to actors is a fundamental challenge to digital forensics as it has been for some time. It means that in order to track down and prosecute a bad actor, you have to do the hard work of finding them, getting search warrants, and so forth. You simply cannot draw conclusions based on Internet identities or locations without some corroborating information.

> *In order to draw meaningful conclusions with a reasonable level of reliability regarding the origin, mechanism, source, or other similar things, about email, the method applied must avoid all of these pitfalls and more. The only approach that has proven reliable over time, is for competent investigators to track emails to their sources using legal means such as warrants, subpoenas, and similar methods... Once a definitive link has been made to a suspect, forensics is then used to demonstrate related information, intent, and other similar elements required to make a legal case.*

In this particular case, I was asked to provide information on how these processes work so that the judge could make a clear decision about discovery issues. In civil cases, there are various factors that have to be weighed in order to share discovery costs or determine who bears what burden, and in many cases, requests for searches are refused because there is inadequate basis for further search and the costs are substantial relative to the matter under consideration. This prevents so-called fishing expeditions.

For example, in one such case, defamatory emails were being sent through a Web site. In order to trace the sender, a subpoena had to be issued to get the records from that Web site on the email address that originally posted the defamatory messages. From there the perpetrator could be traced through credit cards used to pay for the account with their ISP. Once an individual is identified, a search warrant or subpoena allows searches of their computer, and proof may be found, if it exists.

This or similar processes happen in case after case and is the approach that has been used for many years to make many such cases. Just like any other legal matter, it involves judges, paperwork, sound forensics and investigations, and it takes time and effort.

While deceptions and other tricks, like entering false information into Web sites, can be effective at helping to get an investigation going or providing leads, that is only the beginning of the process that is necessary in order to successfully solve most such cases and make certain that the right perpetrator is identified and prosecuted.

Clearly, some process is required in order to get authoritative information that can be reasonably presented in court. It can't be rumor or innuendo, and it can't be hearsay. A sound investigation is necessary to make a legal case, and digital forensic evidence without such basis should be challenged at every turn.

In cases involving multiple companies at different locations, these processes are expensive and time consuming. This, in practice, keeps smaller companies and people without substantial resources, out of most law suits. Those that try to fight David and Goliath battles sometimes win, but they are usually unable to do the work required to win. In most such cases that I see, they start with inadequate expertise and end up with untenable cases.

Malicious exploitation of computers

Again, this is an example where background provides meaningful insight into credibility and introduces an area of challenge.

> *As someone who has investigated computer and network-related incidents for companies, reviewed security incidents, performed security assessments, and worked with law enforcement and government agencies to seek out and prosecute criminals, I have seen, investigated, or been briefed on many instances where servers used for one purpose by one party are also used for other purposes by another party at the same time. Some of these may have direct bearing on the matter at hand.*

One of my worst writing flaws is the use of run-on sentences. When I don't spend a lot of time in copy editing, I tend to write things that are hard to read, even if they are accurate.

> *In some instances this is done by external attack methods, while in other cases this is done by or with an insider. It is fairly common for insiders to use computer systems for purposes not understood by their owners. As an example, sales leads, customer lists, and profiles are often considered very valuable. I have seen cases where leads were provided to a competitor for fees, making money for the malicious actor without interfering with the normal operations of the system, from the point of view of the owner. In one such instance, the activity ultimately ended up in the company losing half of its business. In another such case, the FBI prosecuted insiders within a company for duplicating emails of private individuals in response to advertisements, and selling these leads to other companies. Industrial espionage such as this is more common than most people believe.*

I find it useful to discuss prior cases as foundation for the claims I make, and using the FBI and Secret Service as examples is especially useful in criminal cases. They have both made great strides in chasing down and prosecuting computer criminals in recent years.

> *In the Internet environment it is also fairly common for malicious and/or criminal competitors to invade each others systems and those of trading partners for a wide range of reasons. In some cases they do this to take information in an ongoing fashion so as to gain competitive advantage, while in other cases they do so in order to cause malicious damage, including damage to reputation or the redirection of customers. Documented cases are commonly discussed in security forums where such individuals file legal complaints against each other or report each other to the police in order to get even for such things. Name calling and perceived insults or challenges sometimes lead to escalating malicious acts, and many of these acts violate laws.*

In cases where there are claims associated with attributing actions to individuals without basis, I find it useful to characterize the sorts of things that lead to false trails.

> *There are also a very large number of examples of instances involving false claims of different sorts. These range from widespread medical frauds widely documented in the media, to computer-related frauds in which records are forged in order to create a false trail. Many examples of computer-related crimes are provided in the literature on frauds, and several examples of alternative possibilities are presented in my previous report that relate to the specifics of this case.*

A really good source of these sorts of things in the frauds literature are case histories from the "Encyclopedia of Fraud" which is published annually by the Association of Fraud Examiners.

> *I have also done significant work researching and investigating matters related to activists groups and their use of information infrastructure. For example, activists have broken into computers and posted protests on others' Web sites, gained access to computer accounts and threatened to expose officials via the Internet, sent large numbers of emails from many places toward military sites to protest wars, and created global networks of computers to deny services to those they oppose. A culture of "hacktivism" has emerged on the Internet. In some cases the intensity of these "hactivities" reaches the level where it can be considered information warfare. At the extremes, information infrastructure protest-based attacks have been used to disrupt the Estonian economy, to limit Internet communications for the Israeli army, and to attack networks of bulk emailers.*

There are quite a few responsible authors who have looked into these issues, including Dorothy Denning who is currently a professor at the Naval Postgraduate School. In my case, I ended up doing sponsored research related to terrorism and one of the group types we investigated was activist groups. They have done quite a few very hostile things and, more recently, have moved increasingly into the information warfare arena. In the context of the particular case where this discussion took place, one of the parties was a self-declared activist, so this was relevant.

It is important not to “overplay” claims of malice or fraud on the part of actors who are involved in a legal proceeding. In a deposition, I was asked whether assertions of the possibility of a Plaintiff acting intentionally amounted to an accusation. The question went something like:

> *As you sit here today, can you show me one instance where you have found that shows that [my client] ...*

My response was something like:

> *The evidence is unclear on whether or not your client ... just as the evidence that you have presented that [the other side] did ... is unclear. My point is that it is impossible to tell from the evidence presented and your expert's report, whether or not ....*

The point is, don't over-reach. Reach only as far as you are comfortable going based on the facts. There are actually very few civil cases in which malicious attacks or forgeries are in dispute and have been shown to be the real origin of the evidence. That doesn't mean that they are not the source – only that it's rarely proven.

The single actor theory

A major place where many experts make mistakes in the Internet arena is an assumption that a single actor was responsible for all of the elements of a complex event sequence. The Internet has many different things going on at one time, and they often interact in hard to predict ways to produce results.

The Internet is a complex place, and in the case of criminal activities, many of the sorts of things we see today involve what amounts to sets of criminal enterprises that are loosely tied together in coalitions for individual, often mass, computer-related crimes. For example, in the area of frauds, there are often different groups involved in different aspects of the chain of events. One individual or group might have attacked and taken over a set of computers that they then lease time or capabilities on for another individual or group to use to send a set of emails claiming to be from a bank, providing URLs that go to Web sites that are served by a different set of computers taken over by a third individual or group. The results may be the collection of large numbers of credit card numbers that are then sold to a completely different group of criminals organized to perpetrate fraudulent purchases with those credit cards.

Here's an example of how to challenge claims of a single actor that have not been demonstrated:

> *Another presumption on the part of Plaintiff in this matter is that there is a single actor involved in each and every one of the NN [items] involved. There are two major problems here that I see.*
>
> 1. *One problem is that no statistical argument or presentation could reasonably be used to show or conclude that each and every one of the [items] individually violated this law. Rather, each and every one of the [items] will have to be individually shown to have done so. Groupings that are similar but not identical is not sufficient, in my view, to show that they were individually part of the long chain of events involving several different and competing [parties]. There are many automated tools, [items] are often exchanged and sold in the open market, and many unsavory individuals use content and images from other peoples' Web servers without permission and to their own ends. As such, the mere appearance of similarity between [items] or Web pages, nearly identical codes within [items] or Web pages, or even completely identical [items] or content could be used by different actors.*
> 2. *The other closely related problem is that Plaintiff seems to be saying that Defendant was the single actor involved in each of a set of chains of events responsible for each of several different kinds of [items] sent through many hundreds of locations, even though another actor has already defaulted, a different company has been found to have been*

> *responsible for what appears to be many very similar [items], competitors of Defendant also got the one [item] that was actually generated as evidence in this case, and that [item] came from an [item] that was not shown to have been included in Plaintiff's case. Experience indicates that, in the Internet, multiple actors often independently undertake different activities that result in similar outcomes.*

This goes to the problem of tracking things in the Internet and attributing actions to actors. And in this particular case this was problematic, as it often is, because it is time consuming and expensive to actually track things down to their sources.

> *At this point, not even one actual chain of events has been traced from beginning to end to show that:*
>
> 1. *Defendant knew or had reason to believe [item] would be used in violation of law,*
> 2. *any such [item] was actually sent by the party from which the order was placed or their authorized agent,*
> 3. *any such [item] ever arrived at Plaintiff's site.*

An early case involving audit trails

In one of the first cases I was involved in, a defendant was accused of breaking into a computer system. I was asked to provide information to the disciplinary hearing that could have caused him to be kicked out of the university with the possible long term effect of ruining his education and career.

This student employee was accused of breaking into a school computer system. The offered proof was a set of audit trails from the purported target system. The accuser claimed that these audit tails indicated a break-in. A logical trail of reasoning was used to assert that this break-in must have come from the student. While the logical reasoning might have been sound, I was approached by the student to try to bring light to the matter on behalf of the disciplinary board. The disciplinary board agreed that I was to act as an impartial expert and that the issue at hand for me was to determine whether and to what extent the audit trails that were offered indicated that an attack of the identified sort took place.

This sort of “friend of the court” position is far better, in my view, than working for one or the other side in a case, because it allows me, as an expert, to avoid all of the partisan issues and be granted information based on what is available and what I want to see rather than based on what one side or the other decides they want me to see..

The audit trails I was provided with were sent to me in emails. I was not working with originals, and I was told, in effect, to assume that the audit trails presented to me were those found on the system, ignoring any details regarding how they were identified, collected, preserved, and so on. Therefore, no issue of forensic soundness of the processes undertaken or the validity of the evidence provided was available to be considered.

I received a set of audit trails from a Unix-like system, including *syslog* entries, *lastlog* entries, and log files entries from several system programs, and a claim that the defendant had tried to guess passwords, had successfully done so, had entered the system, and had done whatever harm was asserted.

I could find no evidence of password guessing attempts. Based on my knowledge and experience with Unix systems such as these, such entries would normally be found as a series of failed login attempts shown in the *syslog* file. While the lack of these entries did not prove that no attempt had been made, I stated the obvious in my report, which said something to the effect that I saw no proof or evidence that would support the notion that an attack had taken place in the time frame of interest. At that point, one of the people representing the school asserted that the reason there was no evidence of a break-in must have been that the defendant wiped all of the evidence out of the audit trails. This led to me being asked about this issue.

Many people might say that it was possible or not based on a general notion of the issue at hand, but since I had the audit trails, it came to me that there might be some analysis that would shed more clarity on the matter. So I set out to get a more definitive answer and, before long, found an analysis method that shed more light and wrote a paper on the subject that was later published as a refereed journal article.[24] Some extracts are interleaved in the text that follows without further citation.

In a general sense, if an attacker gains unlimited access to a system, if audit trails are not protected by write-only or write-once technology, and if no physical means are used or effective in determining the authenticity of audit trails, it is always possible to create a forged audit trail that is not differentiable from a legitimate audit trail. All an attacker has to do is create a complete set of audit trails for an equivalent system and replace the real audit trails with the replacements while acting as a privileged user. This can be done with direct hardware input and output, thus avoiding audits of the forgery process itself. Even this is non-trivial, but possible.

Experience shows that attackers rarely meet the theoretical limits of how effectively they can carry out an attack. Instead, when they try to cover up attacks, they tend to do one of three things. They (1) attempt to delete all files on a system to remove all trace of their entry, (2) try to modify select audit trails to remove the indications of their use, or (3) try to prevent their attacks from being audited by avoiding the use of audited events. However:

> (1) If they attempt to delete all files on a system to remove all trace of their entry, they will either succeed or fail. If they fail, the audit trail of their attack will remain, while if they succeed, their success will be detectable by the resulting disruption in the audit trails and other system activities. We may or may not be able to recover other details.
>
> (3) If they prevent their attacks from being audited by avoiding the use of audited events, there is little that can be done to detect their tampering within the system.

The middle case, case number 2, was the issue at hand in this particular matter because audit trails weren't missing and no indication of the claimed attack was present. The claim was that the defendant modified the audit trails so as to remove indications of their attack without eliminating or disabling audit trails during their use. The challenge to this theory (i.e., that the absence of evidence was due to the acts of the defendant rather than being evidence of absence), came in the form of analysis. But in order to do the analysis, we generally have to start with some underlying facts and assumptions.

Most audit trails are generated on a record-by-record basis. Whenever an audited event occurs, a record is appended to a file containing an audit trail. There may be several files containing different classes of audited events. Most audit entries include a time and date, an indicator of the object that caused the audit trail, the identity of the subject executing the audited event, and optional information such as file names, error indicators, and so forth.

Audit trails are typically generated by a number of sources. For example, a Web site typically generates audit trails of attempts to access URLs. This audit trail typically covers the same time period on the same system as other audit trails, but includes details of the operation performed and the Web content on which that operation was performed. Another audit trail details every process executed, by giving the program name, special characteristics, the user executing the process, the controlling terminal (if any), the runtime of the process, and the date and time of the event. This audit trail commonly produces several entries per second.

Based on these assumptions, for the second case (2) above, there must be a method by which modifications to the audit trails are carried out. The most common method is to edit a file containing the audit trail by using an on-line editor. Other options might include removing all audit entries over a time frame by using on-line tools, removing entries associated with a particular user over a particular time, removing records associated with a particular program over a particular time frame, or removing records associated with a particular file over a particular time frame.

For example, if an attacker thinks that the records:

```
Oct 23 06:07:12 all sendmail[28011]: AA28011: message-id=
Oct 23 06:07:12 all sendmail[28011]: AA28011: from=, size=2075, class=-60
Oct 23 06:14:15 all sendmail[28822]: AA28011: to=, delay=00:07:05, stat=Sent
```

indicate something about an attack that the attacker wants to cover it up, they might try to remove those records from that audit trail. One method of modification would be to use an on-line editor to edit the file containing the audit trail, remove the undesired lines, and save the resulting file. One way to detect tampering with audit trails such as those displayed above is by looking for expected behavior that is missing. In some cases, it may be nearly impossible to remove all indicators.

In the example here using sendmail, incomplete removal of the items corresponding to message AA28011 might leave a single line of AA28011 without the other lines displayed above. Since these lines always come in sequences (although not always with three entries), it is possible to detect any set of missing lines given that one line is present. Complete removal of these lines leaves us with no entries corresponding to AA28011, but since this is a sequence number, there will be indicators of AA28010 and AA28012, leaving us with clear identification of the missing AA28011 records. In order to tamper with this information without permitting this sort of detection, this audit trail has to be replaced with another similar audit trail which is plausible and maintains properties consistent with the records replaced. Or alternatively, a completely fictitious audit trail could be put in place of the current one without altering the rest of the file, presumably by matching the sizes of the entries.

Sometimes tampering can be detected as a side effect of the process used to tamper. For example, the record of completed processes shown above produces several audit records per second. This is not true on all systems, but the system being examined in this instance ran a series of automated programs with great regularity. Since audit trails are often quite long (1 million bytes per day is not unusual for an active timesharing system) modifying an audit trail may take a significant amount of time. If an editor or a set of programs are used to create a forged audit trail, the overwriting of the original audit trail produces an end-of-file associated with the modified version of the audit trail. Since audit trails are written many times per second, this may result in gaps in the audit record corresponding to the period during which the records were being overwritten. In order to eliminate this sort of detection, audit trails must be modified in place.

Missing information makes stopping and restarting audit trails a problem as well. When audit trails are restarted, they typically indicate the restart, while the lack of audit records over a period of time is an indicator of tampering. In the same way that missing sendmail audit records may indicate tampering, many programs are commonly used in conjunction with other programs in particular sequences and/or proportions. For example, the following sequence is repeated twice within the same minute and is likely to be repeated throughout the audit trail because it is associated with a particular automated process.

```
grep fc ttyp0 0.08 secs Sat Oct 21 07:01
grep fc ttyp0 0.06 secs Sat Oct 21 07:01
sh fc ttyp0 0.03 secs Sat Oct 21 07:01
tail fc ttyp0 0.06 secs Sat Oct 21 07:01
grep fc ttyp0 0.05 secs Sat Oct 21 07:01
...
grep fc ttyp0 0.05 secs Sat Oct 21 07:01
grep fc ttyp0 0.06 secs Sat Oct 21 07:01
sh fc ttyp0 0.02 secs Sat Oct 21 07:01
tail fc ttyp0 0.09 secs Sat Oct 21 07:01
grep fc ttyp0 0.06 secs Sat Oct 21 07:01
```

Another example of missing behavior is missing login or remote access information. In the following audit trail, the user connected using the ftp file transfer protocol twice within 21 minutes and not at all for the rest of the morning.

```
Oct 23 06:22:09 all in.ftpd[29745]: connect from unix
...
Oct 23 06:43:00 all in.ftpd[2245]: connect from unix
```

If one of these were missing and the forensics expert was sufficiently aware of the normal operation to be able to pick up this sort of behavior, this would be a good indicator of tampering. The problem with this sort of detection is that this sort of behavior is rarely noticed and people are rarely reliable enough to remember such minutia. The predictability of automated processes are far better indicators of tampering in most cases.

There are many other types of missing behavior that can result from attempts to tamper with audit trails. Another attack method for this example would be to selectively remove the lines relating to mail message AA28011 by simply deleting all records containing AA28011 from the audit trail using another program. But one way or the other, in a computer, some program has to undertake every action that is taken. A way to detect the possibility of such tampering, or to rule it out, is to determine whether there was any program that was run in a relevant time frame that could have caused such a change. Since different programs do different things, only specific programs or programs with vulnerabilities that can be exploited without producing other audit trails can be used to do this without a trace.

Another way that tampering may be detected is through the introduction of unexpected behavior. For example, when looking at the audit records pertaining to program usage, there were no occurrences of root running any process other than “sync” and “ls”:

```
sync root ttyp1 0.20 secs Sat Oct 21 07:01
...
sync root ttyp1 0.17 secs Sat Oct 21 07:01
sync root ttyp1 0.16 secs Sat Oct 21 07:01
ls root ttyp1 0.06 secs Sat Oct 21 07:01
sync root ttyp1 0.16 secs Sat Oct 21 07:01
sync root ttyp1 0.17 secs Sat Oct 21 07:01
sync root ttyp1 0.17 secs Sat Oct 21 07:01
sync root ttyp1 0.22 secs Sat Oct 21 07:01
sync root ttyp1 0.20 secs Sat Oct 21 07:01
```

The use of an unexpected program by any user could be a side effect of tampering, but tampering with audit trails normally requires privileges, and privileged users usually don't run many of the same programs commonly used by unprivileged users. If we saw an entry with root running vi (the visual editor), it would be highly suspect unless an authorized user was using vi at or about that time. For each kind of system, there are programs that are more and less commonly used by each user. By looking for uncommon usage, we may be able to detect tampering.

Perhaps the most powerful method for detecting tampering with audit trails is using the correlation between different sorts of audit trails. Using examples from another case, we have a sequence of events from the syslog audit:

```
Oct 23 06:03:07 all in.thttpd[27542]: twist tcprelay@gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk tcprelay
...
Oct 23 06:03:31 all in.thttpd[27572]: twist gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk unknown
Oct 23 06:03:35 all in.thttpd[27573]: twist tcprelay@gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk tcprelay
```

and a corresponding sequence of events from the in.thttpd audit:

```
gate.uwe.ac.uk tcprelay 1995/10/23 06:03:08 cat /heaven.html
gate.uwe.ac.uk unknown 1995/10/23 06:03:32 cat /searcher/index.html
gate.uwe.ac.uk tcprelay 1995/10/23 06:03:35 cat /admin/mlist/index.html
```

Each Web service request is recorded in both the syslog and thttpd log files. The redundancy between different audit trails is quite hard to forge. For example, the mechanisms for turning auditing on and off are different for different audit trails. If an audited event occurs between turning off one audit trail and turning off the other, an inconsistency will result. In the case of these programs, there are at least two other audit trails generated by these events that can also be correlated to detect tampering.

Tampering without inconsistencies is harder in environments where there is more activity because there are more audited events, which means that the windows for turning off audit trails without detection are far smaller. It may take a lot of time in the system to find all of the audit trails being generated and tamper with enough of them in consistent ways to remove all of the inconsistencies. The longer the attacker is in the system, the more likely they are to make a mistake, and as they attempt to remove signs of tampering, they may generate still more audit trails, and of course they may be detected as an illicit user and traced to a source.

To assess the difficulty associated with these methods, I wrote a set of analysis programs to correlate these sorts of records. The first program explored the problem of comparing the *twist* records from the *syslog* audit trail with records from the *in.thttpd* audit trail. For the purposes of demonstration, I used a sampling of about 40 records taken from a low-usage period on my server. Our first observation was that the sizes of the log files were different. One had 41 entries while the other had 43. This means that we should get some indication of an inconsistency because there should (in most cases) be a one-to-one correspondence between *twist* entries in the *syslog* audit trail and *cat* entries in the *in.thttpd* audit trail.

Several sources of differences such as this can be accounted for. For example, when examining an audit over a time span, there may be edge conditions where a *twist* happened before the start of the time span in question and the corresponding *cat* happened within the time span. Similarly, a trailing *cat* can be lost at the end of the time span. Another complexity results from the limit behavior of processes in a multiprocessing environment. For example, as the number of available processes is exhausted, there may be a *twist* record corresponding to an http process that couldn't be started. Another cause of such a mismatch would be a case where a process that would result in a *cat* record is terminated before it can produce that record. Another example where a *twist* could be lost would be a case where different disks store different audit trails and the disk storing twist records becomes full or has a write failure. The problem is further complicated when multiple machines are being used in the audit process. For example, many audit systems transmit audit records from one machine to another for storage. If the time stamps are different on the different machines, or if on one machine the time is changed (e.g., by the action of a systems administrator or attacker), the correlation may not be properly ordered (but the audit trails should still be in real-time sequence of arrival, leaving an indication of the change).

Before making judgments about the source of inconsistencies, it is important to examine the various ways in which they can occur and to rule out as many as possible. In this particular example, finding obvious inconsistencies is not very hard. By understanding the way these programs interact, I know that for each *cat* record there should be a proceeding *twist* record with identical user and system fields (e.g., *gate.uwe.ac.uk* is a system and *tcprelay* is a user). From the case:

```
Oct 23 06:03:07 all in.thttpd[27542]: twist tcprelay@gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk tcprelay
...
Oct 23 06:03:31 all in.thttpd[27572]: twist gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk unknown
Oct 23 06:03:35 all in.thttpd[27573]: twist tcprelay@gate.uwe.ac.uk to
/usr/etc/in.thttpd gate.uwe.ac.uk tcprelay
```

For each of the entries above, there is a corresponding subsequent entry below. In this case, the times are quite close to each other. In one case both events were logged within the same second.

```
gate.uwe.ac.uk tcprelay 1995/10/23 06:03:08 cat /heaven.html
gate.uwe.ac.uk unknown 1995/10/23 06:03:32 cat /searcher/index.html
gate.uwe.ac.uk tcprelay 1995/10/23 06:03:35 cat /admin/mlist/index.html
```

Unfortunately, there is no guarantee that the records will be this well ordered. A *cat* could take several minutes, and there could be many intervening *twist* and *cat* records from the same user on the same remote machine. As a result of this analysis, these particular audit sources have now been augmented to provide process identification information that allows these audit records to be correlated very easily. A more traceable audit trail looks like this:

```
128.163.83.60 1995/10/25 12:28:34 25127 25108 cat /browse.html
128.163.83.60 1995/10/25 12:28:47 25136 25135 cat /journal/index.html
128.163.83.60 1995/10/25 12:29:18 25191 25189 cat
/journal/index2.html
```

In this case, the *cat* record relates back to the *twist* record (i.e., 25136 is the child process of 25135 and thus correlates to the *twist* record marked with that number) and the analysis is straight forward. Unfortunately, many audit sources available with systems don't provide this sort of information. The resulting analysis is significantly more complicated and less certain as a result. Analysis in this case showed that the following lines appear in the *twist* log but don't appear in the *cat* log:

```
Oct 25 12:22:12 all in.thttpd 24461 : twist cbpc38.research.aa.wl.com to
 /usr/etc/in.thttpd cbpc38.research.aa.wl.com unknown
Oct 25 12:37:34 all in.thttpd 26016 : twist 128.163.83.60 to
/usr/etc/in.thttpd 128.163.83.60 unknown
```

As the last audit record in the analysis, the second entry is likely a boundary condition where the corresponding *cat* record had not yet been generated. The first of the two records is an inconsistency that likely indicates an uncompleted request in which the *http* program terminated before audit trails were generated.

An efficient algorithm for doing this analysis is easy to come up with. For example, a merge/sort of the records using the process number as the key followed by a linear search for records that don't come in pairs, yields a list of mismatches very quickly. I believe that there is a linear time algorithm for doing this analysis based on filling two tables referenced by process number as the data is read in, and then linearly searching for entries in one table that don't appear in the other table.

As a side note, process identification numbers are not unique. In most systems, they are allocated incrementally from 0 through a maximum in a fixed modulus (e.g., the process after 23 is 24, and the process after 32447 is 0). In exceptionally large audit records, the analysis may be complicated by this, however, in practice, a lot of time passes between successive reuses of the same process identification number.

In the particular case under discussion, before making a declaration about the lack of evidence of tampering, these analytical methods were developed, tested on other systems, and applied to the audit trails in question. The result was that there was no indication of tampering. While this is not a proof that tampering was impossible in this case, I was able to determine that tampering of the sorts consistent with the details of this particular case would require a lot of computation. Given the fact that there was no evidence to support the claim of tampering and evidence of the lack of tampering in the form of consistency across multiple sources of audit trails, the student employee was found not guilty and he continued to finish his degree, graduate, and continue with his career.

This brings me to another issue that I have encountered in a lot of recent cases. For one reason or another, parties to cases are reluctant to provide audit trails and courts are reluctant to enforce requests for them. Audit trails have long been the bread and butter of digital forensic evidence and have formed the foundation of many cases. Maybe this is why parties try to avoid providing them.

When I don't get and somehow can't get these sorts of records provided to me, I am always suspicious that someone is trying to avoid detection relating to something they did. On some systems, these sorts of records are not generated, and in other systems they are not enabled, but if I get configuration information indicating the presence of audit trails and get no audit trails in response to requests for them, I couch this in terms of spoliation.

I tend to challenge the available evidence by finding places where these records would have settled an issue and indicate that the lack of this evidence forced me to consider many possibilities that otherwise could have been easily ruled out or in by the presence of these records. I indicate that exculpatory or inculpatory information might have been present in these records and that they were knowingly not produced even though there is evidence that they existed. I talk about the duty to preserve these records and ask whether they were not preserved in violation of the court's orders or not produced in violation of the court's orders.[25] And I try hard to find ways of identifying information that would otherwise have been identifiable through these audit records, because I hate to be snookered or prevented from finding the truth because someone refused to let me see available evidence.

The case of the video of the case

Digital forensics is not always about the interpretation of the bits. Sometimes it's about the devices that hold the bits.

Many people who sell their services as experts limit the people they are willing to work for in one way or the other. My feelings run along the lines that anybody, no matter who they are or what they are accused of, deserves to have competent representation, including competent experts. I do require that I get paid for the work that I do and get to tell the truth as I see it, but other than that, I have no particular restrictions on who I will work for. In one such case, I ended up working for a religious group that I don't believe in or even agree with in any way I am aware of. It is not my religion, and I don't care the least little bit about whether they win or lose any particular case. In my view, this is fundamental to doing a good job as an expert witness. If you care about the person you are working for, either by liking or disliking them, it can only bias your views, and that makes for poor science. In this case, I was working for the plaintiff and the issues at hand came down to three things:

1. There was evidence that was claimed by the opposing expert to indicate that the plaintiff had done something they should not have done.
2. There was a claim that the plaintiff had altered records on the way from the place they were gathered to the place they were analyzed (i.e., a claim of spoliation).

3. There was a claim that the plaintiff had altered computers that were returned to defendant thus causing them to fail.

Claiming to be an expert, a professor from a local university was accepted as a special master for the court, even though he was in fact biased toward the defendant and not very much of an expert at all in the matters to which he was testifying and reporting. In making these claims, he asserted that a set of time and date stamps within the contents of a particular file indicated that the plaintiff had remotely attempted to gain access to a computer owned by the defendant and tampered with the system in question so as to create false evidence and remove materials in question.

The expert presented the evidence as being authoritative and definitive, but there was a problem. The names of the files being asserted as the source of the evidence were, according to all of the documentation I got hold of, not files from the program that was claimed to have been used. In fact, the files in question were from a completely different sort of program that did not have the capabilities to do the things that the special master was claiming were done. The format of the content that was used to make the claims was in the format of the files from the program that could not do the claimed actions, and consistent with the filenames given. As a result, the challenge was that the special master got it wrong. By simply stating the facts and providing authoritative sources (the manuals and documentation provided by the makers of the different products in question), the special master's case in this regard was certain to quickly evaporate, and worse yet for the other side, his expertise would then be in question and his status as a special master would also come under scrutiny.

The second claim, that of spoliation, was based on the assertion that the plaintiff had introduced all of the records in question in the case to the disk drive of the defendant while the disk and the computer that contained it were on the way from where they were collected to the police lockup where they were held prior to being forensically imaged. There was, of course, no evidence whatsoever of this, but it was a challenge by the other side, and it had to be answered.

The answer turned out to be fairly simple. Using the technology of the day, a simple calculation was made of how long it would take to plant the volume of information in question on a hard disk of a computer. In this particular case, there was a great deal of evidence present, and by simple calculations, I was able to show that there was not enough time between when the evidence was clocked out of the residence where it was confiscated and the time when it was clocked into the police evidence room to have made the number of changes required. This was only a matter of 15 minutes or so, and the time stamps were taken and kept accurately by normal standards of the day. That was claim 2.

So now we come to the namesake of the case, the case of the video of the case. The third claim was far more interesting. The defendant was claiming that the plaintiff had destroyed the content of the computer system in question when removing the contraband that was within the computer in question.

The problem with this claim was that there was a videotape of the return of the computer in question to its owner and, at the time of the return, the computer was turned on and the owner was shown that it worked. The owner indicated verbally that the computer appeared to work and spent a small amount of time testing it out. All of this was done in front of a group of observers and recorded on the videotape.

The defendant asserted, however; that the videotape failed to show that the whole activity was a clever ruse intended to fool the cameras, the observers, the defendant, and the court. They claimed that a floppy disk was placed within the computer and that the floppy disk simulated the normal startup operation of the computer.

The thing that defeated this attempted claim was the videotape of the front of the defendant's computer. While the tape was grainy, there was enough resolution to tell that the floppy drive did not have a floppy disk in it. As it turns out, when you look at a floppy disk drive of that particular type, when the disk is not present, the door flap over the drive is closed, and when a disk is present, the door flap is swung open and the floppy disk trailing edge is visible in the slot. I was able to visually verify that there was no floppy disk present, and that more or less settled the issue.

But the nature of legal cases is not always or even usually one of technical issues winning out over other technical issues. The case was actually ended in favor of the plaintiff when the judge in the case, a friend of the defendant's lawyer for many years, recused himself from the matter and was replaced by another judge whose wife was a friend of the wife of the lawyer for the plaintiff. I never had to testify, the case was settled, and that's just the nature of the legal system.

Mock trial and the MD5

Every year or so, at a conference somewhere in the world, there is a mock trial involving digital forensic evidence. These are abbreviated show trials with a real judge, real lawyers, and real experts. The thing that is not real is the testimony and the circumstances set up for the trial. They are just sort of realistic.

At one of these, I was privileged to play the naïve expert witness who used a cryptographic checksum (the message digest MD-5 checksum in this case) as the basis for matching one thing against another. The scientific literature shows that the MD-5 checksum, and over time most other similar mechanisms, can be forged in the sense that it is possible to generate a different document that has the same checksum as a questioned document. As an overly confident expert, I testified that I used the MD-5 and therefore that there could be no question that two files were identical.

Of course this was a setup, as the opposing lawyer had a trick up their sleeve. They generated a new file that had an identical MD-5 checksum, by prepending a byte sequence that leaves the MD-5 checksum generation mechanism in its initial state, and therefore produces a different file with the original checksum. At that point, when faced with the evidence in front of me, there was little I could do but be unable to answer any further questions and act like I didn't understand what was in front of me. It's good practice if you have never done this.

At the end of the little mock trial segment, someone in the audience asked how to do it right, so we did it again with the original questions on how I validated that the file I was looking at was the same one originally collected at the scene. The exchange was something like this:

Q: Did you rely on the MD-5 checksum to prove that the data you are presenting here is the same as was originally collected?

A: No. I relied on my knowledge and experience, my training and the techniques that I used, and I used the MD-5 checksum to validate that my work was done correctly.

The virus ate my case

In another strange case out of England, a defendant offered the defense that a computer virus was at fault, and not the user. Apparently, they offered no evidence to support the claim that a computer virus had deposited the contraband material on their computer. But on the other hand, the prosecution could not show that such a virus was not present and counter the claim.

This seems to me to represent a lack of care on the part of the prosecution. As challenges go, the claim that a computer virus did or did not do something should be able to be demonstrated pretty clearly. Computer viruses are not the same sorts of mysteries they were 20 years ago. There is enough experience with them to make fairly definitive statements about whether or not any widespread virus deposits particular files in a particular type of operating environment. Furthermore, the files themselves could be examined, along with other related logs, to determine more definitively what actually happened.

I suspect that, in this case, the prosecutors thought there was no possibility that the defendant would be believed and they were simply surprised by the result. On the other hand, the nature of the legal system is that you need to be diligent in all things or one of them might well trip your case up.

I should also note that it is indeed possible for computer viruses to carry contraband onto computers and, as such, it is incumbent on those who search those computers to take the time and effort to not only look for the presence of content, but also identify the path by which it arrived and the manner and time frame in which it was stored. This is not a particularly hard thing to do in most cases, and while computer-related time stamps on files in many file systems can be easily altered or forged, in the vast majority of cases I have encountered, they don't appear to have been modified or altered at all. Traceability is necessary to make a sound forensic case.

That is to say, a level of diligence necessary to withstand scrutiny should be undertaken by those trying to make a case, and the individual challenging that evidence should feel free to take advantage of anything that is missing or not properly done.

Spoliation of evidence and chain of custody

When evidence is spoiled in one way or another through the lack of diligence or intentional acts of one party, the legal term is spoliation (or compromise) of evidence. [23] Why the 'i' is moved for this word is something I don't exactly grasp, but then I am not in charge of the legal language police. Spoliation is one of the major challenges to digital forensic evidence that comes up in cases where the people taking the evidence, holding it, or moving from place to place, don't have adequate training. The other similar challenge is the loss of chain of custody, which basically makes it impossible to guarantee that the evidence provided to the court or the other side is in fact the same evidence gathered in the first place.

In my view, it is important to be careful in claims of spoliation or loss of chain of custody, because there is a tendency to assert malicious intent rather than incompetence. There is an old saying that I like to use:

> *Never claim malice for what can be explained by incompetence.*

A lot more people are incompetent than malicious. To mistake one for the other may lead you down the path of incompetence. Remember, you don't want to be guilty of the things you claim they are doing by over-reaching in your claims.

This is not to say that people don't intentionally alter, fail to identify and collect, or lose track of evidence. They do. But as a matter of form, I think it is better to stay just short of accusing someone of intent if you cannot prove it, while at the same time providing insight to the court of indicators of motive, opportunity, and means.

Here is an example from a recent case where the plaintiff had the capabilities, had expressed the intent, and had the opportunity to forge the evidence in the case.

In my previous disclosure, I indicated that there were many possibilities relating to spoliation of evidence and chain of custody. The additional information provided within this report appears to indicate that some of those concerns were justified and does nothing to alleviate the other concerns. In particular:

- *It appears, based on the deposition of Mr. XX, that the [items] constituting the evidence in this matter was altered by someone else while Mr. XX was working with it. Mr. XX does not know specifically what was removed, added, or altered, however, he appears to indicate that some portion of the evidence was removed.*
- *It appears, based on the deposition of Mr. YY, that certain notes, reports, or other materials subject to preservation requirements in this case were not preserved, and at least some of them may have been shredded.*
- *It appears, based on the deposition of Mr. YY and Mr. XX, that the [items] used for analysis were in the "DBX" format, while the [items] sent by Plaintiff were, according to Plaintiff's deposition, in "ZIP" format. Nothing has been presented to describe how these ZIP format files were turned into DBX files, or whether Defendant ever got the original content provided by Plaintiff.*
- *The entirety of about NN [items] originally identified by Plaintiff still have not been provided to Defendant and no explanation has been provided of how those NN [items] turned into the M [items] identified for this case. Some if not many of these additional [items] may contain exculpatory evidence, and Defendant has not been allowed to see them.*
- *The database containing something like X [items] from Plaintiff and Y [items] from other sources, with which the NN [items] in this case were mixed and then culled, may have contaminated the [items] in*

> *question, and Plaintiff has provided no information to allow me to determine whether or not this is likely or what measures, if any, were taken to protect the evidence during this process. ... The [items] in this case might be contaminated.*

After going through a litany of things that could have gone wrong with the evidence, the lack of chain of custody, and so forth, I came to the issue of the potential for malice:

- *Additional information was provided with respect to Plaintiff's level of knowledge of technical matters and Plaintiff's ability to gather, retain, generate, and manipulate potentially meaningful records related to this matter. Specifically,*
    - *Plaintiff has been shown to know a great deal about the specifics of [items] records... and has shown more knowledge than Plaintiff's Investigators in this matter.*
    - *Plaintiff has been shown to know about the records kept at ZZ and how they are accessed. This appears to indicate that Plaintiff knew of additional data sources that could have been preserved and failed to preserve them. Those records could have refuted the evidence that she presented in this matter.*
    - *She had the capability and opportunity to forge or alter records prior to providing them to counsel ...*
    - *Her background includes computer programming and systems administration and she had systems administration privileges on the computers where the original evidence is asserted to have been collected and stored.*

In the deposition that came after this report, I was asked directly about the issue of spoliation and I indicated that I wasn't accusing anyone of anything. Rather, I indicated, the evidence against the plaintiff for intentional spoliation was better than the evidence against the defendant for what they were accused of.

I clarified that I would not accuse anyone of such a thing based only on this sort of evidence, and that I thought that the defendant should not be accused of anything based on the amount of evidence in this matter and the level of credibility associated with it.

Questions

1. Identify all of the time-related issues that should be checked out for a case involving two computers connected to the Internet that are purported to have taken part in the same criminal act. Which of them could potentially create false positives with respect to identifying criminal acts and how?
2. In doing a link analysis, your opponent's expert failed to do proper investigation of ownerships of like domain names even though they had the same IP addresses during the time in question with respect to the matter. How important is their failure in terms of challenging that evidence and why? Under what circumstances would it be critical?
3. Given the complexity of Internet operations between companies, how problematic is it to use the Internet for investigations? What does the other side have to do to do a good enough job so that you cannot challenge it?
4. Given that wide range of possibilities involved, how helpful is the systematic approach identified in chapter one?
5. Go through this chapter, and for everything in italics, identify any statement that might be excessive or could be challenged. Discuss how it could be challenged, how it is excessive or inaccurate, and back it up with facts, analysis, experiments, or other methods that show that you are right and I was wrong.
6. Given that the time is limited in forensic cases, how much has to be done to be right enough to defeat the challenges identified in this chapter?
7. How would you have done a better job than the experts I opposed in the cases here? Take an example of one or more of the cases discussed and explain how you would have identified, collected, stored, transported, processed, and presented things to avoid the challenges I put forth.

5 Testifying

Any expert work you may do may ultimately end up challenged in a deposition and later in a trial, if things get that far. While most digital evidence never sees the light of a courtroom, most experts face depositions fairly often and testify in court somewhat less so.

In some sense, depositions are simple and straight forward. Someone who is being very pleasant and nice will introduce themselves and act like your best friend, ask you all sorts of questions about yourself, and every once in a while, try to stick a shiv in you.

Your job is to stay calm, listen carefully to what is asked, answer only the things you are supposed to answer, and do so with simple clarity, honestly, openly, and without saying anything that you might come to regret later.

When I say regret later, I do not mean to say that you should be careful to protect your side of the case or anything like that. Remember, you get paid no matter what you say. That's the whole point of acting as a professional and an expert. You only want to speak the truth as you know it in as clear a fashion as you can express it. What I mean to say in terms of regretting things later, is that you need to be careful to make certain that what you say is really right, not an approximation, not a generalization, and not the sort of thing you might say in every day conversation. What you will regret later is any time you say something like “any time” or “always”, if it turns out that you are not quite right in every possible case.

The reason you will regret it later is that someone like me will show up and have hours and hours to read what you said and find a single counterexample that will be portrayed, as it rightly should, as demonstrating that you were wrong and that therefore you should not be given credibility in other things you say.

So, having clarified the gentle art of what happens when a professional wordsmith gets hold of the words stated in anger or a moment of weakness, I will get down to more of the specifics.

Prepare

Preparation for testimony typically involves more than just rereading everything you have written on the matter, which you should most certainly do. I think it is important to read all of the materials you have seen, to go over what you have stated in your reports, review any outside materials you have referenced or looked at, and generally prepare to face any and all questions about what you have said or written.

Then, it is generally a good idea to do a preparatory session with the counsel for the side of the case you are working for. This includes having them prepare you by asking questions that the other side might ask, by having them test your memory and ask things intended to show how confrontational it can get, and to prepare you by reviewing everything you have covered.

Most lawyers know this and schedule it, but if they don't, as a good expert, you should. It is particularly important that you make certain you know all of the details of the protocols, formats, and other related technical things you have covered. As I mentioned earlier, if I was going to testify with regard to forensics related to a disk drive, I would review the electromagnetics and physics as well as the electronics, digital components, formats, use in computers, interfaces, programs that access these devices, drivers, operating systems, and so forth. If you don't know about these things before you start the case, you should not be testifying about issues related to disk drives. But assuming that you do know about them, you should make certain of your facts by reviewing them in detail again just prior to the days where you will be deposed or testify.

I know that I have weaknesses here and there, and whenever I am preparing to testify, I review the weaknesses carefully, more so than the strengths. I consider it a matter of professionalism to know my stuff when I come into a legal situation, and this is all the more important if my role is to criticize others for not knowing their stuff or for making mistakes in their work. How bad will you look if you critique someone else for making the same sorts of mistakes you yourself make?

Tell the simple truth as you know it

In legal settings, I always tell the truth as clearly and completely as I can. I don't work for my client as an advocate. That's the job of the lawyer. I work to seek the truth and to try to help make certain that the legal system works as well as it can. That means that I have an obligation to the truth more than to any loyalty to my client. In fact, I say so before I will take a case. I don't care about my client at all. If I do, I cannot honestly participate in the trial as an independent expert, and I have to either act in some other role or recuse myself.

It is easy to tell the truth. All you have to do is listen carefully to the questions you are asked and answer them honestly and openly. But at the same time, you cannot be so naïve as to think that the lawyers are not advocating for their clients. That is their job. The truth has to be told carefully, not haphazardly, and not in a manner that allows it to be twisted for the purposes of one side or the other.

There are tricks that lawyers use. For example, they will ask you to identify an authoritative source of information in your field. I will tell you right now that I don't know of any such thing in the computer security or digital forensics field. There are some very good works on very specific subjects, like the original paper on the RSA cryptosystem, which I consider to be authoritative in terms of identifying the mathematics of the RSA cryptosystem – and nothing else. I think that my Ph.D. thesis is authoritative in terms of the formal definition of computer viruses that it puts forth, and not authoritative in any other way. I don't think that this monograph is authoritative in its coverage of challenges to digital forensic evidence, it's just the best I can do right now.

The counsel on your side should tell you about issues in the case that affect your testimony. For example, they might tell you that the other side is trying to make the case that your testimony is invalid for some technical reason. When you know this, you can try to be careful about what you say in that regard, avoiding particular terms, and so forth. But at the end of the day, you still have to tell the truth as you see it. And that, as it turns out, is pretty simple to do. Just say what you mean and mean what you say.

Don't stretch it - be open

So having said that, I have to also tell you that people have egos, weaknesses, vulnerabilities, and psychological limitations, and good lawyers know this and try to exploit it. People that call themselves experts have egos and often try to claim expertise that they don't actually have. If you stretch credibility even a little bit, you might get away with it, or you might get caught. Assume you will get caught, and don't stretch it at all.

When questioned in a manner that seems insulting, people have a tendency to become defensive, which is a misnomer for becoming offensive to defend themselves from the ego attack. If you do this, you will almost certainly make a fool of yourself by stretching something too far, acting like a child, or saying something you don't really mean and then having to back down from it.

Being open is very important to this. There is an exercise you might try out in order to get yourself prepared. Repeat after me:

I don't know.

Say it again and again – in front of others – in situations where you don't know something. As adults, we all become experts in all sorts of things we don't know anything about. Telling our children how to drive. Telling others how to solve particular computer problems. Telling ourselves that we can do this and know what we are doing. In a legal setting, you have to learn to abandon these pretenses. Repeat that phrase again and again and again. When you practice your testimony to yourself before you go into the trial, ask yourself questions that you do not know the answer to and practice answering "*I don't know*".

Prepare to be silent as well. When you get to the end of an answer, don't feel as if you need to ramble on as if you were teaching a class or instructing someone who knows less than you. Just stop when you are done and wait. This is an area where I am really not very good at all. I always feel as if I should go the extra mile to explain things, and when I do, if I am not always very careful, I will say something stupid, stretch beyond my actual expertise, or say something that is not precisely true in the technical sense.

Take your time and think

After a recent deposition that I gave, I asked the lawyer on my side what he thought of my testimony. He answered me in a peculiar way. He told me that I was the first witness he had ever seen that stopped to think, and when asked to continue, stated that he was thinking.

I was amazed by this comment. For some reason, the society we live in seems to reward those who talk more. If you can't make an immediate comment, somehow that means you are not as smart or knowledgeable. That's how the theory goes.

I guess that I'm just not that smart. I don't immediately know the answer to all questions as soon as they are put to me. Often, in fact, I need to pause for a moment and think carefully about what was asked, consider it, and give an answer based on that thought. My advice to you is that you should do the same thing.

I find that I have to listen carefully to questions asked of me in order to figure out what is actually being asked. Unlike a politician, I try to answer the question rather than make some sort of a statement that was prepared in advance after twisting the question into my prepared speech. I don't prepare much in the way of speeches for testimony. I think through things and prepare my thoughts. When I practice with myself, I think through the questions I think I might get. I try to be careful to form proper sentences that meaningfully and carefully provide the answers that reflect my honest views on the subjects. But I get plenty of questions that are worded in ways that give me pause. So I think carefully before I answer them.

Sometimes questions are complex, and because of the legal process, objections may come along and be sustained or overruled. In the case of depositions, you just go ahead and answer anyway and let the judge sort it all out later. I often have to ask the reporter to read back the question so I can review it carefully in my mind and answer it responsively and responsibly.

To quote my SCUBA instructor (and the training manual):

Stop – Think – Act

Questions

1. In preparing for testimony, how important is a good night's sleep?
2. If the lawyer doesn't want to do a prep session, how hard should you push them to do one? How would you prepare if they didn't want to help do such a session?
3. Identify questions that you would ask from one of the example case studies provided in this chapter and explain how you would attack the expert challenges provided.
4. Telling the "simple truth" may not be so easy in a complex case. How would you present an accurate description of the Internet file transfer protocol so that it could be understood simply, in only words, by someone who knows nothing about the Internet?
5. Suppose that in the previous hour of testimony, you made a mistake about something you thought was important. If a related question came up, how would you correct your previous mistake as part of your answer? Give a particular example and see how it works for you.
6. In listening to questions, you made an error in interpreting what was asked and answered the wrong question. The lawyer on the other side was rather harsh in the way they treated you and came back with a very loud and abrasive. You are a bit flustered and pretty stressed at this point. They ask you a yes or no question in a leading manner and there is no objection. What do you do?
7. You get a question that you are not certain of the answer to. What should you answer?
8. Do a mock trial where you and another person press each other with hard questions and practice coming up with good answers that meet all of the criteria for being honest and accurate.

6 How to avoid being challenged

I actually have no inherent desire to present challenges to digital forensic evidence. If I had my way, I would never have anything to challenge. People doing investigations and analysis would be thoughtful and careful, present the basis for what they do, understand their tools and use them properly, not overreach in their claims, and be independent and fair in their approach to every aspect of their work. In a few cases I have been involved in, there were legitimate differences of opinions where reasonable people might disagree, and I think that is fine. But this is not where the challenges usually arise today. More often than not, in such cases, both sides agree to the bits and to what they mean. And when there are differences, they tend to be quite minor.

> *In one such criminal case, there were experts on both sides, and all of the experts and the lawyers agreed to stipulations about the evidence. The opinions were so close to each other that you couldn't tell them apart with a roadmap. When I testified and was asked about why my opinion differed from the other experts, I told the jury that in fact we all agreed to just about everything and that any attempt to claim otherwise was just not right. But the lawyer for the other side still presented their closing arguments as if there was a world of difference between us.*

There is nothing we can do as experts to prevent this sort of thing. That's just how legal systems are. I certainly don't think that the experts on the other side considered my views to be substantially different from theirs, and I think that they told the truth just as I did. In fact, when asked about differences of opinion, I said that very clearly. They did a good job of interpreting almost everything there was to interpret, and I largely agreed with them as they largely agreed with me. I haven't encountered them since, but I doubt that we would have any disagreement about anything in the case, even today after the verdict is in and the issues are all settled.

The best way to avoid challenges to your digital forensic evidence is to do your job carefully and properly, don't allow yourself to get personally involved in the case, study the issues at hand, try hard to be a thoughtful individual who takes their job seriously, and seek the truth regardless of where it leads you. I admire professionals who do their jobs well, and it's always a pleasure to meet them.

When a lawyer contacts me to take part in a case, if the professional on the other side has done their job well, I will tell the lawyer so, and advise them that the digital forensic evidence is sound and should be stipulated to. I still charge them for the time I spend, but the cost is far less.

In many cases, lawyers ask me a few questions and, by the time we are done, they become convinced that I am not the expert for them, usually because I tell them that the other side sounds like they have it right. They might ask me about the ways someone can defeat some mechanism, or whether there are hidden bits that could prove their client innocent, or whatever. And for the most part, if the other side has done their job, there is nothing in particular to challenge and I tell them that.

In most cases I am aware of there is little about the digital forensic evidence to dispute. In most cases, the digital forensics experts never see the witness stand because the cases are settled, plea bargained, or dropped. And in most cases, accused criminals really did commit the crime. But there are cases where this is not true, and if you are in such a case and don't do your job properly, carefully, fairly, and with an eye toward the truth rather than the position of the lawyer who hired you, look out.

So my simple advice for those who don't wish to have their digital forensic evidence challenged, is to learn what you need to know to do your job professionally and accurately, do all of the things that are taught the way they are taught, and don't go beyond your area of true expertise.

Everybody makes mistakes, and most of them are relatively easy to compensate for, are not very important to the issues at hand, and don't make much real difference in the outcome. But the one mistake that seems to always create challenges is when someone claims expertise they don't have, doesn't take the time and effort to make damned sure, and doesn't say “I reckon”.

Questions

1. At what point do faults become failures?

2. What are your areas of expertise?
3. What areas of expertise are you expanding into and how are you going about assuring that you have the expertise you need to be competent to testify about them?
4. How deep does your expertise have to go?
5. In analyzing the overarching structure of challenges to digital forensic evidence, what has this book missed?
6. Do you believe that every person deserves an expert who will evaluate the evidence and advise them on the possible challenges to that evidence? If not, why not? What are the exceptions?
7. Looking over case histories from open court proceedings that involve digital forensic evidence, do you identify faults in every case that could technically be pointed out by someone who is tasked with challenging that evidence? List the faults you identified and explain how these faults could have been avoided.
8. Looking over those same histories, do you find that the digital forensic evidence faults reached the level of failures in terms identified in this book? Identify samples and explain why these faults produced failures and how these failures could be mitigated.
9. Set up a mock trial and plant a series of realistic and forged items of digital forensic evidence. In reviewing these forgeries and non-forgeries, how many of the people involved in analysis of the evidence properly differentiated one from the other? How many did not? Do you expect that forensics examiners in real court cases will be any better at doing this?

Endnotes

1. For a more detailed discussion of the concepts of faults and failures, see: "*Basic Concepts and Taxonomy of Dependable and Secure Computing*" Algirdas Avizzienis, Jean-Claude Laprie, Brian Randell, and Carl Landwehr, IEEE Transactions on Dependable and Secure Computing, V1,#1, Jan-Mar 2004.

2. There are many good discussions of the issues surrounding legal admissibility. For one opinion in a recent case, see "*Lorraine v. Markel American Insurance Company*" (241 F.R.D. 534) starting on page 5. But please note that this is not settled law. Another place to start your search is "A *Bibliography Related to Crime Scene Interpretation with Emphases in Forensic Geotaphonomic and Forensic Archaeological Field Techniques*" United States Department of Justice, Federal Bureau of Investigation, FBI Print Shop, Washington, D.C.

3. The NIST Computer Forensics Tool Testing (CFTT) Project Web site is located at http://www.cftt.nist.gov

4. In 2001 there were a series of situations in which forensic imaging software was found to fail by missing individual blocks on specific disk drives. This was discovered when in-depth testing started to take place by NIST and others. Individual blocks were missed on many popular products and the questions in the on-line forums were increasingly expressing concern about how many cases could be appealed because of such errors. But over time, even though many such tools have known errors, no cases other than one case that faced problems in a pre-trial motion have been determined by such errors.

5. Cryptographic checksums, sometimes also referred to as cryptographic hashes or simply hashes, stem from early work on the creation of hash tables in which content is transformed in a manner that tends to spread it equally over a space (a many to one into mapping) so that the likelihood of two different, even if quite similar, items of content have an equal probability of resulting in any given hash, or checksum value. This was explored for integrity protection in the 1980s in papers including F. Cohen, "*A Complexity Based Integrity Maintenance Mechanism*", Conference on Information Sciences and Systems, Princeton University, March 1986. and a series of papers ultimately leading to F. Cohen, "*Models of Practical Defenses Against Computer Viruses*", IFIP-TC11, ``Computers and Security'', V7#6, December, 1988. Many other related works exist, and standards for the creation of cryptographic checksums have been created, updates over time, and used in many different ways for digital integrity protection.

6. Because of the way that computer basic input output systems (BIOS), hardware, and device mechanisms work, there are often different combinations of physical locations that may contain content. While most forensic imaging uses the BIOS-level mechanisms used by operating systems to gain access to disk content, lower level mechanisms may also exist in some media that could grant still more and more definitive information from physical areas of media.

7. Made data in networks often results from protocol analysis errors and cognitive error mechanisms that produce things that look like valid, meaningful, authentic data when they are no such thing. In the presence of deception mechanisms, made data is commonly achievable by malicious actors because of the lack of integrity and attribution mechanisms in most protocol systems of today. Even more interesting is made data created by analysts who look at one kind of content and interpret it as another kind, thus making something that is not relevant data at all into something that ends up being treated as data for the purposes of analysis.

8. The so-called hardware address of Ethernet and similar interface cards is often forgeable by raw packet interface software. While some hardware devices prevent this, many do not, and forgery of MAC addresses and responses to protocols can lead to many sorts of failures of this sort. Having said this, it is also important to note that in the absence of such mechanisms, hardware addresses of interface cards and other similar information is often very reliable and identifies a computer interface to the individual machine as well as to the class of interface cards. This is often very strong evidence for individualization.

9. See for example, United States v. Scarfo, Criminal No. 00-404 (D.N.J.). In an article at http://epic.org/crypto/scarfo.html the case is described in more detail. In essence, the government was allowed to present as evidence an unclassified explanation of a classified keyboard sniffing software mechanism without the defense being allowed to challenge the real mechanism in use. A plea agreement was ultimately reached so no appeals ruling was made.

10. Title III (or Title 3) refers to Title III of the US Code, also known as the “Wiretap Act”. The federal wiretap law was enacted as Title III of the Omnibus Crime Control and Safe Streets Act of 1968. Separate legislation, the Foreign Intelligence Surveillance Act (FISA), established probable cause requirements for wiretaps in foreign intelligence and international terrorism investigations. There are also laws in each state in the US regarding wire taps, there are maintenance exceptions for common carriers, and there are many other elements of this area of law.

11. Exculpatory evidence is evidence that would tend to show that the party is not responsible for or guilty of the identified act, whether criminal or civil in nature. Inculpatory evidence is evidence that tends to show that the identified act was undertaken in the manner described.

12. This is a particularly complex area because of the potential for booby-traps, or other pre-programmed behaviors, the presence of keys that may not exist or be accessible after power-down, remote and local mechanisms that may interact, and a host of other possibilities. Consider, for example, a remote systems administrator who has a real-time visual of the user based on a Web camera and who sees the entry of a search team and commands destruction of evidence remotely. Unless this is understood in advance, presumably through an intelligence process, any move the team makes could be the wrong one. There can be no set rule that is always right.

13. In addition to the many sorts of data that might exist in different places within and outside a file, there is also so-called “meta-data”, or data about data. This meta-data includes anything from file times stored in file systems and embedded in the file systems that form the object linking and embedding (OLE) storage mechanisms used within many files, to the protection and related settings associated with archive files, header information associated with electronic mails, and a wide array of other things that may be considered similar to tool marks in the physical world. Some programs also store data within files that they call meta-data, such as author information, edit times, last author, and so forth. It is also important to understand that much of this data may be inaccurate and, in many cases, is undocumented, unsupported, and unreliable. The process by which this data comes to be and can be altered is also often uncontrolled so that, while the normal process might be one that is somewhat traceable, the alternative processes might be undetectable in normal operating environments. For example, file system date and time stamps can be altered by a command or system call, meta-data within files can be altered by any program that can alter the file, and there are specific tools for wiping out or altering such fields within files without leaving additional traces. Determining what happened by what process is, in general, very hard, and highly subject to specifics of the circumstance.

14. The notion of packet arrival time jitter was new to many of the experienced people who read this monograph, which is to say, it isn't a standard approach that is well documented. But jitter is a well known phenomena in electronics and timing issues abound in many areas of digital systems. Many papers on network performance discuss things like group delay and jitter because it is important to keeping high fidelity with low delay times in real-time processes such as audio and video signal processing. In many cases, new phenomena are identified that seem to show something of interest, and they fall into the category that is commonly called “Just Doesn't Look Right” (JDLR). When such things are identified, in order to be admitted for legal purposes, additional analysis and experimental work may have to be done to meet the standards for admissibility.

15. The digital forensic expert who is going to testify in court will have to be familiar with operating systems and how they work in order to work on issues such as these. Many good books on operating systems are available and it is highly advised that the potential expert have adequate background in this and related areas if they are going to try to identify what events took place in order to produce the evidence proffered. Of course the same is true of Internet protocols and operations, interface behavior, storage media and formats, and so forth.

16. The Daubert case (Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 125 L. Ed. 2d 469, 113 S. Ct. 2786 (1993) dominates in US Federal cases. Frye (Frye v. United States, 293 F 1013 D.C. Cir, 1923) may apply in many states for non-Federal cases. The Frye standard is basically: (1) whether or not the findings presented are generally accepted within the relevant field; and (2) whether they are beyond the general knowledge of the jurors. Daubert also allows accepted methods of analysis that properly reflect the data they rely on.

17. It is for this reason that it is usually best to take contemporaneous notes, describe the experiments performed, and provide additional details surrounding any such techniques applied. While some assert that it is harder to challenge processes undertaken when not fully disclosed, it is my view that failure to disclose should normally exclude the process or its results from being admissible. This then begs the question of what constitutes full disclosure. I think the answer is that the process should be readily repeatable by the other side so that it can be tested and refuted if refutable. This notion of testability is a fundamental tenant of scientific evidence and part of the notion underlying both Daubert and Frye. It also involves the right to face the accuser, which is to say, the right to have your expert examine the work of their expert to see if there were errors that the other side made. If it is not a scientifically valid and testable analysis or technique, or if the information on how the technique was performed is not available, it is not scientific evidence. It is more like magic and mysticism, neither of which is likely to meet standards for admissibility when it comes to digital forensic evidence presented by expert witnesses.

18. In recent years, internal search capabilities within computer systems have dramatically changed the nature of historical records of drafts and other related material. For example, the OS-X Time Machine, internal Google Search, and the Spotlight features are only a few examples of new mechanisms for generating backups of versions of files over time. Time Machine, for example, does a backup of changes every hour. This introduces the possibility that there may be hundreds of draft versions of every written report on some systems. A similar situation exists for computers being searched, many of which may have numerous copies of drafts, partial documents, and other similar materials in forensically accessible areas of their disks. In cases where this is an issue, the forensic expert must be aware of the mechanisms in use and proper precautions to take against both loss of vital data and accidentally using the wrong version in a report or other work. As a final example in this arena, I recently provided a draft to counsel at his request, and he accidentally sent the draft instead of the final to the counsel for the other side. As soon as this was discovered, counsel for the other side was notified, sent the signed document, and agreed to delay the deposition to allow the signed version to be reviewed, under the condition that the draft would not be used or introduced. After agreeing to these terms, counsel for the other side went right ahead and introduced the unsigned draft in the deposition. In this particular case it was not damaging, but does show the potential dangers in producing and using such drafts. As experts, we work for and with the lawyers, and if they want a draft sent, all we can really do is warn them. Still, it is a good idea to make your drafts as clean as possible before sending them on.

19. Another important reason to be thorough and try to create as many independent and valid pieces of evidence and analysis as possible is that the appeals process may invalidate certain types of analysis or evidence. This in turn may result in old cases being retried or appealed. If there are many threads in the case, a single thread being destroyed by a subsequent ruling will be less likely to result in a successful appeal or the overturning of a verdict.

20. Legal challenges to admissibility under the Federal Rules of Evidence in the US generally go under the following categories. Evidence admitted has to be weighed by the trier of fact in making determinations. Depending on specifics of the circumstances and judicial opinion, evidence may or may not be admitted and weight may be expressed by the judge to the jury in formal admonitions for admitted evidence to go to weight.

Relevance: The tendency for evidence to make a fact of consequence determination of the action more or less probably than it would be without the evidence.

Authenticity: Rules 901-903. There is evidence sufficient to support a finding that the matter in question is what its proponent claims. Many illustrative examples are provided, but they are not exhaustive. They include personal knowledge, non-experts familiar with a unique property such as handwriting, comparisons to known samples by trier or experts, distinctive characteristics, public records, ancient documents, reliable process or system, and methods provided for by statute or rule. Some records may be self-authenticating, such as public documents, certified copies of documents, official publications, and certified records of regularly conducted activity.

Hearsay: Rule 801. An out of court statement offered in evidence to prove the truth of the matter asserted is hearsay, but there are many exceptions; most notably business records taken in the normal course of business and relied on for their accuracy and reliability as a matter of course in carrying out that business.

Original writing (best evidence): Rules 1001-1008. To prove content, the original is required unless certain exceptions apply. Exceptions include: (1) originals lost or destroyed, (2) original is not obtainable, (3) the opponent who holds it refuses to produce it upon judicial demand, (4) the content is not closely related to the matter at hand and is thus collateral. Official records are admitted as duplicates. Voluminous records may be represented by statistical samples when they are representative and subject to examination of the originals out of court. When the admission of other evidence depends on facts in this evidence, the court makes the determination, otherwise it goes to weight. When the issue is whether (a) the asserted content ever existed, (b) another piece of content admitted produced it, (c) the evidence in question accurately represents the original, the trier of fact determines it.

More prejudicial than probative: Rule 403. Evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by the considerations of undue delay, waste of time, or needless presentation of cumulative evidence

Scientific evidence (expert testimony): Rules 701-706, Frye, Daubert. Non-expert testimony is only admitted if it is (a) rationally based on the perception of the witness, and (b) helpful to a clear understanding of the witness' testimony or the determination of a fact in issue, and (c) not based on scientific, technical, or other specialized knowledge within the scope of expert testimony. A witness qualified as an expert by knowledge, skill, experience, training, or education, may testify in the form of an opinion or otherwise, if (1) the testimony is based on sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. If facts are reasonably relied upon by experts in forming opinions or inferences, the facts need not be admissible for the opinion or inference to be admitted however; the expert may in any event be required to disclose the underlying facts or data on cross-examination. Endnote 16 describes the Frye and Daubert standards.

In order to be admitted, digital forensic evidence must survive challenges to relevance, authenticity, its hearsay nature, the original writing requirement, must not be far more prejudicial than it is probative, and must be introduced and analyzed by people who meet standards. It is incumbent on the party introducing evidence to meet these criteria and on the party challenging to oppose based on these criteria and to do so in a timely fashion as part of the legal process. Experts can help make this happen by identifying all lines of challenge and providing expert analysis, advice, knowledge, and skills to help create the conditions for challenges.

It is generally better to make as many such challenges as possible under the theory that if any challenge succeeds it may get evidence disallowed and the more such challenges are presented, the less weight and credibility the evidence will have. Lawyers may not be able to use all of the things that you find as an expert, and time or monetary limits may prevent you from doing as thorough a job as you would like to do, but you can only do what you can do.

It is also important to note that the experts are subject to challenges. If they make too many mistakes, if they are unreliable, if they use techniques that are not in the scientific literature or widely known and used, or if they lack the skill, knowledge, training, experience, or education necessary to qualify them, they can be disqualified along with much of their work.

Last, but by far not least, it is critical to understand that these are not hard and fixed rules that are uniformly applied according to some strict algorithmic formula. Judges and juries are people and subject to all of the human failings and amazing human capabilities that are inherent in the human species. They have beliefs, points of view, they make cognitive errors, and have likes, dislikes, and biases of all sorts. No matter how hard they try, they may not be able to abandon all of these as triers of fact, and they are not supposed to.

There is an old saying used to describe the system of justice as it exists today:

"It is the worst system of justice ever designed – except all the others."

21. One of the reviewers of this book has a lot of experience in this area, and he advises not to describe yourself as an "expert" with specific areas, because there is almost always some other expert in a sub-area that knows more then you do about some facet of the area. The "expert" would then become no such thing.

22. One of the biggest sources of problems today has to do with network-based investigations where searches over networks are illegal or otherwise problematic.

People sometimes do foolish things like breaking into remote systems or using deceptions in ways that are not strictly legal. They may run remote commends that are unauthorized or exploit weaknesses in remote systems to get information from them. They may assume the identity of an offender because they have a valid user identity and password, perhaps gleaned from attack code the attacker placed on their system to retrieve remote malicious code scripts or other similar things. They may engage in illegal wiretaps when they think they have permission of the owner to do network sniffing. In international cases, these may violate national sovereignty issues and even create international incidents. The investigator had better know the laws that affect what they do, or their evidence, competence, and legitimacy may be questioned and they may end up in jail. Here are a few suggested by one of my reviewers (plus one or two of my own):

- 18 USC 1029 Fraud etc. in connection with access devices
- 18 USC 1030 The Computer Fraud and Abuse Act (amended by the 1996 National Infrastructure Protection Act)
- 18 USC 1831-1839 The Economic Espionage Act of 1996
- 18 USC 2700-2710 Stored Wire and electronic communications and transactional messages access
- 18 USC 2000 The Privacy Protection Act
- 18 USC 2511 The Federal Wiretapping Statute
- 18 USC 2510 The Electronic Communications Privacy Act (ref. 2700-2710 above).
- The Patriot Act Amendments to ECPA and other laws above.
- The Digital Millennium Copyright Act.

There are also state laws on private investigations from many states that make it illegal for anyone other than a government investigator (typically the police, special investigators, Federal investigators, etc.), a licensed private investigator, a lawyer, or a full time employee of a lawyer (this is not made clear in all state laws) to investigate or hunt a person. This is particularly relevant to digital forensics investigations because of the tendency to try to figure out who did what and to track a person down through computer networks.

23. In civil cases, it is typically called spoliation, while in criminal cases it might be called compromise.

24. F. Cohen, “A Note on Detecting Tampering with Audit Trails”', IFIP-TC11, Computers and Security, 1996.

25, For good coverage of data retention and disposition, spoliation, and duty to preserve, see: *The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age*, A Project of The Sedona Conference Working Group on Best Practices for Electronic Document Retention & Production, September 2004 Public Comment Draft.