The Evolution of Digital Forensics:

Civilizing the Cyber Frontier

Ian Charters
with contributions by Mike Smith and Graydon McKee
1 January 2009

<This page left intentionally blank>

Table of Contents

License

<This page left intentionally blank>

Introduction

This document is based on a paper that I wrote and delivered at NIST's Techno Forensics 2008 Conference, on 27 October, 2008 (http://www.thetrainingco.com/html/AgendaForensic08.html). That paper was called, Digital Forensics: The "No Escape" Zone.

Some time ago I was thinking about the evolution of various aspects of computer security. One of the ideas that occurred to me was that by looking at the way forensics evolved in the past, with an eye to the pressures that guided its evolution, we could get a better understanding of how forensics would evolve in the near future.

So, I decided to apply this idea to my experiences in the arena of computer forensics. The way I see it, computer forensics has undergone three stages of evolution:

1. The Ad Hoc Phase
2. The Structured Phase
3. The Enterprise Phase

Phase 1: The Ad Hoc Phase

The Ad Hoc phase was characterized by a lack of structure, a lack of clear goals, and a lack of adequate tools, processes and procedures. This first phase can almost be called the pre-forensics or proto-forensics period. In application, the greatest weakness most organizations did not understand the importance of an Acceptable Use policy and procedures.

At this time it was not uncommon to see an organization's management carefully collect evidence that IT equipment was being used "inappropriately" by an individual, only to find that HR and Corporate Counsel would refuse to act citing the lack of a published appropriate use policy.

Moreover, these policies needed to be backed up by a set of well conceived and coordinated procedures. These procedures needed to contain information concerning what actions could and should be taken if "inappropriate" use was reported or suspected. Both elements, policy and procedure are critical to being able to enforce sanctions against an individual(s) caught using IT equipment inappropriately. I've seen many cases in which the evidence of inappropriate use was clearly in violation of a well written, published and distributed policy. The lack of established procedures however drove Corporate Counsel into apoplexy over the issue of due process.

Just when things couldn't get worse, the way that evidence was collected and handled began to be challenged in court cases involving inappropriate use. This was a two-pronged attack. The first challenged the accuracy of the forensic tools and the second focused on procedural or

chain-of-custody issues. Could it be proved that the tools captured the data accurately? How can you prove that the evidence is accurate and untampered with if the floppy containing the data was left to sit in an unsecured desk drawer? Imagine being sued by the an employee who was terminated for surfing porn on company equipment during company time and losing the case because your poor processes and procedures provided their attorney the opportunity to argue that their rights had been violated. If this sounds unbelievable then let me assure you that it has in fact happened on many an occasion.

The most unfortunate thing about the Ad Hoc Phase is that we had to live through it twice. The first time occurred during the time of the mainframe and centralized computing. It happened again during the microprocessor age with the proliferation of workstations. Within the centralized model of the mainframe era, information was segregated to a single location. It was relatively easy to physically secure the system and therefore limit access to the information contained therein. As we moved away from the use of centralized computing resources toward the distributed model that we enjoy today the issues of appropriate use, data security, and data integrity have become more pronounced. With the acceptance of this model all of the hard learned lessons learned of the past with regard to the need for digital forensics policy, process,

procedures and tools were largely forgotten. Ironically, the microprocessor also provided the tool to make computer hacking, and abuse more accessible and wide-spread.

Some speculate that quantum computing is the next major stage in computing. If that is the case, let's just hope that we do not have to relive the Ad Hoc Phase yet again.

All of this turmoil created the pressure for the move to the second stage of evolution in digital forensics; the Structured Phase.

Phase 2: The Structured Phase

The Structured Phase evolved out of the confusion surrounding the use of digital forensics. Questions concerning appropriate use, the surveillance of employees and potential intruders, the need for policy and procedure alignment and various legal issues lead to a period in which a great deal of structure was imposed upon the practice of digital forensics. This structure was expressed in three primary areas.

- Policy-based programs
- Defined and coordinated processes closely aligned with Policy
- Requirement for forensically sound tools

It has often been said that armies always prepare for the last war. The truth behind this is that the experiences of the last war are the freshest and most salient. In this respect, it isn't uncommon for security program managers to be caught in the same thought process as generals and admirals. Both sets of decision makers are trying to create effective responses to threats, both are facing a mixture of known and unknown threats, both have limited resources for the effort, and both operating in a real-time threat environment.

So, in the structured phase of development of digital forensics tools and techniques, the most significant developments focused on creating effective responses to the difficult problems faced in the Ad Hoc stage of development.

So, let's start our examination of this phase by looking at the enabling criminal legislation (and just to keep things simple we will only be looking at legislation at the Federal level). The following statutes are usually cited as the core cybercrime statutes:

- 18 U.S.C. § 1029. Fraud and Related Activity in Connection with Access Devices
    - Was intended to criminalize hacking systems while using a model; it also included hacking via remote terminals
- 18 U.S.C. § 1030. Fraud and Related Activity in Connection with Computers

    - Was intended to criminalize hacking systems via direct access (i.e. the insider threat)
- 18 U.S.C. § 1362. Communication Lines, Stations, or Systems
    - Was intended to criminalize hacking a system via a telephone line and telecommunication infrastructure.
    - Closely related to the federal wire fraud statutes
- 18 U.S.C. § 2510 et seq. Wire and Electronic Communications Interception and Interception of Oral Communications
    - Important because it outlines what constitutes illegal bugging which is important if you monitor employees
    - Surprisingly this statute does not apply to law enforcement or intelligence authorities.
    - Most states have their own statutes that deal with bugging so they should be consulted because they vary considerably. The rule of thumb – if the transmission crosses state lines then the federal statutes apply and the individual state statutes may also apply.
- 18 U.S.C. § 2701 et seq. Stored Wire and Electronic Communications and Transactional Records Access
    - Covers the capture or intercept of "communication" including "data", not just voice communications that have been stored electronically.
- 18 U.S.C. § 3121 et seq. Recording of Dialing, Routing, Addressing, and Signaling Information
    - Closely related to 18 U.S.C. § 2701 et seq.
    - Focuses on the capture of routing data

Basically, this legislation criminalizes hacking either on systems or over a wire. Moreover, it makes hacking, data theft, and system disruption a Federal crime. However, this legislation also generally requires that you be able to place a value on the act. That is to say that if you cannot demonstrate monetary loss or potential monetary loss associated with an act of hacking, data theft or system disruption, you may not have a crime. In the past this has led to some very creative cost accounting.

I personally think equal consideration should also be given to various privacy statutes and intellectual property statues. Others would also suggest the importance of anti-spam, identity theft, and data breach legislation -- however, I contend that state legislation is quickly acquiring prominence in these areas. Reviewing the state statutes is largely pointless due to the fact that each month some state legislature passes some significant legislation on the topic. Maintaining currency on the topic requires almost full-time professional attention.

Of course, none of this addresses some of the most contentious issues associated with, 'appropriate use' policies such as the impact inappropriate use may have on other individuals within the environment. Let us take look at an example of one employee surfing porn while at work. If other employees become aware of this fact they may feel that they are working in a hostile work environment. This puts the company in a situation where it may be liable and this leads us into the area of employment law and that again violates my scope rule.

So, without trying to appear glib, you can quickly see that developing a clear understanding of the legal issues surrounding "appropriate use" is a complex undertaking. You must rely heavily on your organizations legal counsel as they are experts in this area. Rely upon their guidance and advice for help with both "appropriate use" policy as well as procedure. As important, but not necessarily as obvious, similar guidance should also be sought with respect to developing and implementing appropriate processes and procedures to take in response to seemingly externally-based hacking of the network. Such policies can provide incident response teams with the critical guidance they'll need in critical situations.

The Structured Phase is also responsible for creating a forensic tool industry that is driven by the need for tools that can withstand courtroom challenges as well as collecting data in a forensically correct manner. Since this is such an important concept, let's take a moment to define what "forensically correct" means. I've always believed that in order for data to be collected in a forensically correct manner it has to be:

1. Collected and maintained in accordance with a defined procedure
2. It must be verifiable as authentic
3. It must be verified as relevant
4. It must be collected in a reliable manner
5. It must preserve the original evidence to the extent possible

These general rules also closely follow the guidelines established by the courts that I have dealt with. However, the criteria may be different in your specific jurisdiction, so as always when dealing with legal matters, get professional local council.

If you have not defined and documented your procedure you cannot reproduce it. If you can't reproduce your procedure, its effectiveness and the authenticity of the data collected using the procedure cannot be assessed. This is an important point because under the law each

individual accused of wrongdoing has a right to be treated fairly and equally. If different processes are used to process each individual case, the individuals involved can rightly claim that they were not treated equally and therefore not fairly. I believe procedures also address issues such as chain-of-custody issues – that is to say, addressing how you can verify that the data collected have not been tampered with.

Authenticity is nothing more than being able to prove that the evidence or data is what it is purported to be. This also can be addressed with procedural documentation, and can be part of a chain-of-custody discussion. Hashing of data, such as employing an MD5 hash algorithm to mathematically define and verify a data set, is a great technique to verify that the collected data and the analyzed data set are the same.

Relevance means that the data must relate to the issue at hand.

Reliability must pass the test of "beyond a reasonable doubt". "Beyond a reasonable doubt" is a legal definition. Being convinced "beyond a reasonable doubt" means that you're convinced a majority of (theoretical) reasonable men. This is often a matter of documentation. So, for example if a collection or data copying algorithm is subjected to expert review, and presumably after several iterations of review and revision is documented to be sound and correct, that body of documentation weighs heavily in court as to it reliability.

Also a tool or technique must preserve the original evidence whenever possible. In part this is why when duplicating original disk, write blockers are employed. Write blockers insure that data is not added to the original drive in the copy process. Likewise it is common practice to work off of a copy of a copy of the original disk. This insures that if a mistake is made, you can simply make another copy from the initial copy without harm.

So, even though information security programs became policy based and spawned processes and procedures in alignment with these policies Phase 2 wasn't always effective. In part this is because the tools being developed and utilized were difficult to use and expensive to employ, that is to say that they didn't work well in real-world environments.

Allow me to provide a few examples. The first situation can perhaps be described as not having the right tool for the right job. If you look at hardware-based keystroke recorders that were typical of the first generation of tools, they were very obvious when installed. I think we can all agree employing one in most circumstances would lead the individual who the target of an investigation to conclude that they were either under investigation, or that a hacker was targeting them. Even "modern" USB-based hardware keystroke loggers look pretty much like old school keystroke loggers.

Figure 1: PS/2 Keystroke Logger

Figure 2: USB Keystroke Logger

The second situation is an even more typical, costly and alerting situation. That is the situation in which an organization agrees that there is sufficient cause to warrant an investigation of an individual or group of individuals. The next day the individual or group comes to work and discovers that their computers have been removed from their workspace "for maintenance". In fact, the equipment has all been hauled off to the lab for imaging and analysis because the portable imaging equipment that would allow in situ data collection of data is either unavailable or unaffordable[1].

This kind of thinking provided the impetus for the intellectual and technological development that led to the third phase in the history of digital forensics.

Phase 3: The Enterprise Phase

Once the tools and techniques of digital forensics were accepted as legitimate and effective, the market began to drive digital forensics from a point-based solution into an enterprise-based solution. This imposed some significant technical demands. It also created significant market opportunities.

In general, the Enterprise phase of digital forensics can be characterized by:

- Real-time collection
- Field collections tools tailored to the need of the collectors
- Forensics as a service

Real-time collection requires a central location to store and analyze the data collected. This also requires that the collection takes place over the wire or the network infrastructure.

While real-time collection of data is preferred, some of the devices that contain data needing collection reside on the network either directly or continuously. This requires that tools and techniques be engineered to meet the real world requirements of field or in situ collection. These tools and techniques need to be discrete and should be employed in such a way as to not alert the subject of the investigation that they are being investigated. In these cases speed, portability, accuracy, and a low-profile nature are important.

[1] In situ and perhaps even after hours collection, provides the option of not alerting the subject of an investigation – and all of their coworkers in the area -- that they are under investigation. This can be an especially useful option if the investigative subject is in fact innocent. It can avoid all of the embarrassment and ill-feeling associated publically investigating an innocent party.

In some cases, the requirements of real-time data collection can be disruptive to the Enterprise environment so consideration must be given to removing, or outsourcing forensic data collection.

So, let me give you a few examples of this evolution….

Referring back to my previous example of a keystroke logger, in the Enterprise Phase two primarily solutions soon evolved for the desktop and the laptop. The first and least common is an embedded hardware keystroke logger. While these devices have to be installed before collection is considered, they can operate both on-line – that is to say while connected to the Enterprise infrastructure – or off-line. While connected to the network they can be activated, deactivated, and the data that they have collected can be downloaded, providing many collection options, all of which are transparent to the user.

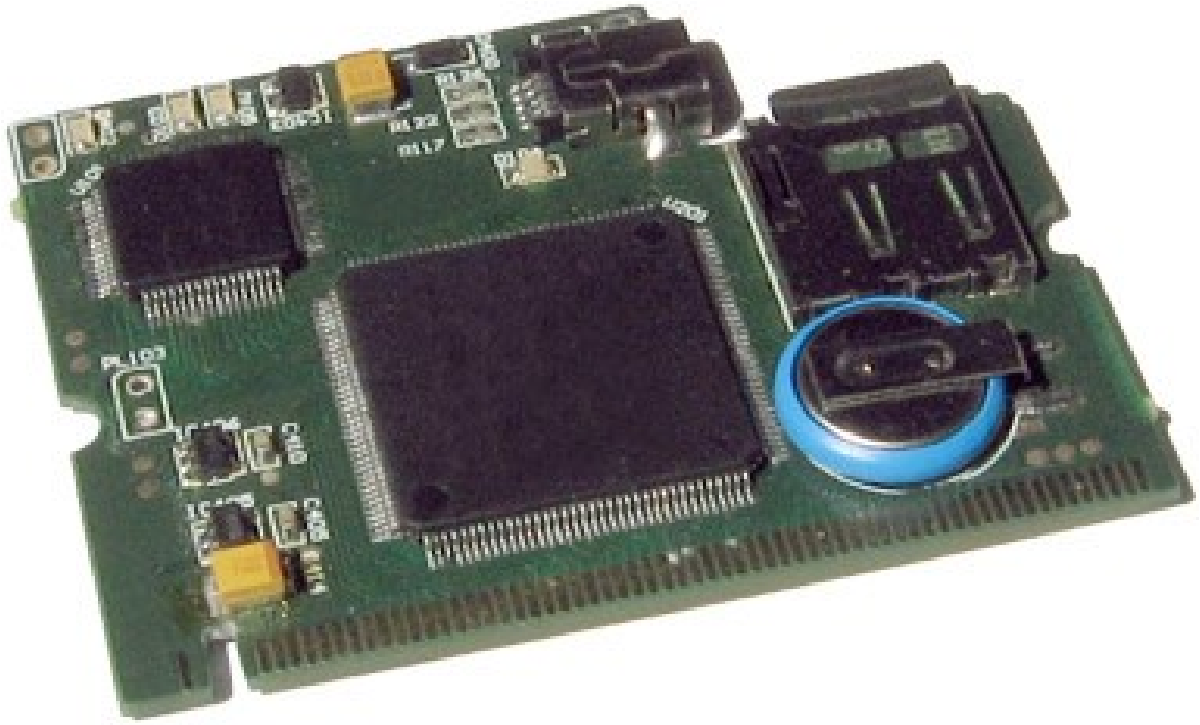

Figure 3: Internal Desktop/Laptop Keystroke Logger

And despite the trade-offs between hardware and software keystroke loggers, Enterprise Phase solutions are usually software-based and can also be centrally administered across the network without the knowledge of the individual end-user. Usually, such solutions are part of a larger more comprehensive forensic solution. Let me take a moment to elaborate on the trade-offs between hardware and software based keystroke logging solutions. It is commonly believed

that hardware based keystroke loggers, if properly implemented, are more accurate than software-based solutions. Even the best software based keystroke loggers are known to drop data under specific circumstances. On the other hand, software based solutions are seen as being much more flexible in that they can be installed and uninstalled over a network connection.

Similar trends can be seen in the area of digital forensics support for mobile devices. Ad Hoc Phase solutions simply didn't exist and Structured Phase solutions were invariable tied to a forensics laboratory workstation. Enterprise Phase solutions offer in situ or on-site collection opportunities. This is an obvious requirement given the nature of mobile devices. Again, nothing changes the nature of an investigation like showing up in the investigative subject's office and seizing all of their computer equipment, media, phone and PDA so that it all can be imaged back in the lab. A perfect example is the Forensic Mini Digidrive Write-Blocked Memory Reader as seen at http://www.forensicpc.com/products.asp?cat=13. This device is an excellent example of the reduction in form factor over past media copying solutions. Equipment like this allows forensic data collection to move out of the lab and on-site. This trend is continuing and we are seeing tools being developed that move forensic collection from the "mobile lab" to the "back-pack lab" or even the "messenger bag lab."

Figure 4: Portable Media Copier with Write-Blocker

And where highly mobile solutions for media imaging have evolved, imaging solutions for mobile devices have also kept pace. One of the real innovators in this field is Paraben. They have produced the Project-a-Phone device for collecting forensic data from mobile phones (http://www.projectaphone.com/). Prior to this a desktop or laptop computer was required for imaging a mobile phone. While this was effective, the size of these devices placed a burden on the forensic collection team who was already burdened with many other data collection tools. Small, "low-drag" solutions like this are a blessing.

The ultimate in high-performance, low-drag imaging solutions for mobile devices may very well be Paraben's CSI Stick. http://www.csistick.com/. While the name of the product does make me laugh, it certainly creates an effective mental image of what the device is all about. The CSI Stick is also an affordable solution. This is typical of the sort of capabilities and devices that define the Enterprise Phase forensic evolution.

Figure 5: Paraben Project-a-phone

Figure 6: Paraben CSI Stick

In addition to the hardware based solutions, we have seen a similar evolution in software based digital forensic solutions. The following are some of the landmark software packages that have helped define Enterprise Phase software-based solutions:

- Access Data – Known File Filter (KFF)
- National Drug Intelligence Center's - Hashkeeper
- Guidance Software – EnCase Enterprise
- Access Data – Enterprise
- Brian Carrier's – The Sleuth Kit (TSK)
- Mediant - Intelligent Response
- Clearwell – E-discovery Platform
- Athena Archiver
- LogLogic

Known File Filter and Hashkeeper are key tools that have established the foundation for the automation and standards-based forensic data collection and analysis. They incorporate a scientific, reproducible, and disciplined approach allowing the forensic professional to ignore large volumes of known system and application files so that attention can be focused on potentially significant data and in the case of Hashkeeper, the development of know bad file signatures.

It was only recently that I understood the significance of the development of known "bad" file signatures. A discussion with a friend and former FBI Special Agent alerted me to the fact that there is a range of illegal computer activity in which the perpetrators almost invariably have certain key (underground) files on their systems. By searching for these key file signatures, investigators can quickly determine if they exist on the target's computer. This just shows that after working in the computer security and forensics field for many years, there is always something to learn.

Guidance Software really pioneered the move to Enterprise forensic solutions. Their EnCase Enterprise makes the collection of data in real-time of the Enterprise infrastructure possible and manageable.

Mediant's Intelligent Response may very well be establishing the next stage of evolution in digital forensics. Their Intelligent Response product is a rules-based appliance that takes automated forensic data collection to the next level. The evolution of this product and the response by the rest of the industry will be worth watching.

Clearwell's E-discovery Platform is another interesting offering that may be a trendsetter. The concept behind E-discovery Platform is that for some Enterprises, the proper management,

archiving, and subsequent forensic data collection of e-mail is so burdensome, difficult, and disruptive that out-sourcing the entire function is attractive. With new Federal guidelines for e-discovery on the books, Clearwell may have gotten themselves in front of increasingly vexing problem with a well thought-out and timed solution.

While not specifically a digital forensic tool, the company LogLogic is providing a ground breaking integrated log collection and analysis capability that is long overdue. LogLogic technology may also lend itself to an out-sourced service. If a standard could be developed for log reporting and annotation, similar to NIST's SCAP standard http://nvd.nist.gov/scap.cfm, we will see fabulous returns on investment on the lowly, long-neglected, and oft forgotten activity of log analysis. Not to over state the situation, but LogLogic's emphasis on automation coupled with an admittedly hypothetical NIST reporting standard for forensics, could very well lead to be THE technologies that can fully integrate all on-line forensic technologies. This integration has been something of the holy grail of the Digital Forensic community.

The Future

I believe the future of digital forensics will be aimed at greater automation and interoperability. In general it will be characterized by:

- Proactive collection and analysis
    - Appliance-based tools
- Government/Commercial partnerships
    - InfraGard
- Standards-based software architectures
- Standards-based reporting
    - Forensics version of SCAP

Automation is a key element to the future of digital forensics. Automation will allow proactive collection and detection. This will translate into reduced cost for detection and mitigation. This will be a welcome capability considering the increasing concern over data leakage and the new rules for e-discovery. Automation will also reduce program costs.

The creation of proactive forensic appliances is a clear move in this direction. If we can characterize forensic events effectively and embed those profiles in an automated appliance, it would allow not only for the real-time collection of forensic data, but also provide real-time prevention of related "events". This will also provide the ability to have forensic appliances work more interactively with more general IT security appliances allowing both programs to be managed centrally and efficiently.

The pooling of experience as manifested in the Government/Commercial InfraGard Initiative http://www.infragard.net/ is another essential forward move. This lays the groundwork for collecting hard data with regards to forensic events that can be analyzed and translated into the establishment of standards for characterizing, reporting, and predicting forensic events. Such standards would accelerate the trend toward interoperability between forensic tools, and wider interoperability with core IT security software and appliances.

I recently attended a launch event for NIST's SCAP program. SCAP is a program designed to create a common reporting protocol or language to describe a broad range of security events. It should be obvious that the wide-scale implementation of such a reporting protocol could lead to much great security hardware and software interoperability and integration. In fact, many major vendors are implementing the SCAP protocol, primarily for the Federal IT market. I think this is such a strong tool that it will quickly find acceptance in the broader commercial market.

The reason why I bring this up is that it seems to me that a similar and perhaps complimentary reporting protocol could be developed to describe and report on digital forensics events. Such a protocol would also provide an opportunity for forensic hardware and software interoperability and integration. And it would also address many of the emerging and cutting edge concerns in the area of digital forensics such as data leakage and document control. Such a development could also create a clear roadmap indicating the most useful near-term development paths for the digital forensics vendor community for the immediate future. In the end, such a technology would also make digital forensics more affordable and supportable. If a forensics version or extension can be developed and accepted, it would probably increase the market penetration of the technology, that in turn would creates a better computing environment for us all.

Summary

Well we started this journey through the evolution of digital forensics with the aim of getting a better understanding how the field might evolve in the future. During the process we tried to pay special attention to the pressures or drivers that sparked the various developments within the evolution. These drivers were useful in breaking up the evolution into distinct phases:

- The Ad Hoc Phase
- The Structured Phase
- The Enterprise Phase

The Ad Hoc Phase was characterized by a lack of structure, goals, and adequate tools, processes, and procedures. The primary issue faced within this phase was that most organizations did not understand the importance of an Acceptable Use Policy and the resulting procedures. Both of these, policy and procedure are critical to being able to enforce sanctions and implement a sound information security program.

The Structured Phase led us further through the evolution as structure began to appear. Organizations began to institute policy and develop the associated processes to support the use

of digital forensics in response to court losses. These in turn also drove the changes in state and federal law which helped to define what actions are legal and which are illegal.

Tools emerged out of the Structured Phase which, fed by the requirements of the new laws, were designed to withstand the challenges raised against digital forensics within the courtroom. In general, these tools support the collection and analysis of data that is:

- Collected and maintained in accordance with a defined procedure;
- Verifiable as Authentic;
- Verifiable as Relevant;
- Collected in a reliable manner;
- Preserved to the extent possible.

The use of these tools, while beneficial, proved to be costly and time consuming. As a result organizations tended to avoid the use of digital forensics wherever possible. This pressure has pushed the digital forensic industry into the Enterprise Phase.

The Enterprise Phase evolved out of a need to reduce the cost of digital investigations in terms of both time and money. This has sparked a move toward the real-time collection of data, customizable tools for field collections, and the possibility of Forensics as a service. This is where we stand today.

Automation is key to the future of digital forensics. It allows for proactive collection and detection and can be accomplished in a manner that is consistent with the process approved by the courts. In support of this continuing evolution, standards for characterizing, reporting and predicting forensic events should be developed and incorporated into the next generation of automated forensic devices. There would be a considerable benefit to be realized from the development of such standards and the U.S. Government's National Institute of Standards and Technology may be in the best position in which to do so.

All told, digital forensics has had a pretty logical evolution, though be it we had to learn some of these lessons more than once. Let us hope that we don't repeat the mistakes of the past and can move digital forensics into the new age of information management.

End Notes

Trying to understand the legal issues surrounding Digital Forensics can be a difficult task. Below are some resources that I've used to help me navigate through these issues. Keep in mind that these are hardcopy resources and this vulnerable to becoming outdated, especially with the brisk pace with which we sue each other here in the United States. For this and a host of other reasons, when in need you should also consult competent legal counsel as soon as possible.

NetLaw by Lance Rose (http://www.weyrich.com/book_reviews/netlaw.html) – Written by a legal expert. The text is accessible and understandable by the lay leader

Syslaw: The Sysops Legal Manual by Jonathan D. Wallace and Rees W. Morrison (http://tinyurl.com/sysopslegalmanual) – Somewhat dated now but still a good reference

SysLaw by Lance Rose (http://tinyurl.com/syslaw) SysLaw is also written by the author of NetLaw.

If you would like to learn more and to keep up-to-date on groundbreaking security news, subscribe to the guerilla-ciso blog feed at http://www.guerilla-ciso.com/

About the Author

With over 20 years of experience in the field of digital forensics, Ian Charters has a unique perspective on the evolution of digital forensics. His career has taken him from the private sector into government service and back to the private sector.

After successfully starting and running his own networking, software development and systems integration firm, Ian was recruited into the nation's Intelligence Community, including service in both the Defense Intelligence Agency and the Central Intelligence Agency where he proudly served his country for over 20 years.

Upon retiring from Federal service, Ian served as the Security Practice Leader with the Unisys Federal Group based in Reston, Virginia. While being responsible for leading the practices efforts in the development, sales, and delivery of a full range of IT security solutions, he was responsible for the development and introduction of a code application assurance into the Federal market.

Ian is currently a Senior Manager in a Big 4 Accounting firm's information security and risk management practice. His responsibilities include leading engagements to provide security services to both commercial enterprises and government agencies.

Ian holds a Bachelor of Arts in Political Science from Washington State University and a Masters of Arts in Security Policy Studies from George Washington University in addition to completing extensive post-graduate and commercial coursework in Computer Security, Architectures, Networking, Programming, Telecommunications, Computer Simulation and Simulation Theory. He is a frequent seminar speaker with the Potomac Forum Ltd, a non-profit educational foundation (www.potomacforum.org) and serves on the Board of Advisors for Ascension Risk Management LLC (www.ascensionriskmanagement.com).