



A COMPREHENSIVE APPROACH TO DIGITAL INCIDENT INVESTIGATION

An Article Appearing in *Elsevier Information Security Technical Report*

By Peter Stephenson, CISM, CISSP, CIFI, FICAF
Copyright © 2003 Elsevier Advanced Technology

A COMPREHENSIVE APPROACH TO DIGITAL INCIDENT INVESTIGATION

By Peter Stephenson, CPE, CISM, CISSP, CIFI, FICAF

Abstract. *The investigation of digital incidents and computer-related crimes has, over the past 18 months, become increasingly complex. Although the majority of digital incidents investigated by law enforcement still comprise child pornography, on-line frauds and other common crimes, two disturbing trends have emerged. First, digital incidents are becoming more complex and, second, they are becoming more expensive to investigate.*

Attacks such as the recent SQLSlammer worm affect tens of thousands of corporate computer users. Attackers are bouncing attacks off of home computers connected to the Internet by DSL and cable modem. Many of these home computers have no firewall and are left turned on most of the time. Attackers, recognizing the importance of anonymity, use these unprotected PCs to launch and further attacks against corporate targets.

The need for a comprehensive incident investigation technique is becoming obvious. This paper discusses one such technique: End-to-End Digital Investigation or "EEDI". EEDI is a very structured approach to conducting complex digital investigations using the investigation framework developed by the Digital Forensics Research Workshop¹ (DFRWS). The EEDI process allows investigators to use a very structured investigation technique that mixes computer technology with traditional investigative methods. In trials during actual investigations as well as in presentations to law enforcement and civilian practitioners EEDI has received a positive response.

A primary advantage of EED is its ability to feed a formal modeling program. By using a specialized process language such as the Digital Investigation Process Language (DIPL), investigators can model investigations and compare their models with standardized benchmark models of idealized investigations. EEDI and DIPL are not intended for use in simple digital forensic examinations. Rather, they are most useful in complex investigations where other sophisticated tools such as link analyzers are also in use.

In this paper we will describe the underlying background for EEDI, the EEDI process and the top level approach using the DIPL. We begin with some background.

¹ Digital Forensics Research Workshop. "A Road Map for Digital Forensics Research 2001." Digital Forensics Research Workshop 6 November (2001)

INTRODUCTION

The Underlying Issues

As pointed out in the abstract to this paper, the state of digital crime is that it is increasing in both complexity and quantity. The recent Computer Security Institute/Federal Bureau of Investigation computer crime and security survey reported almost half a billion dollars (USD) in quantified financial loss with 74% reporting their Internet connection as the key point of attack.

The SQLSlammer worm, as an example, shut down entire sections of the Internet (Korea, as an example) for up to six hours. To the date of this writing, the exact source of the worm has not been traced. This writer participated in an incident post mortem for a large multi-national organization and found that, once the worm was inside the organization's enterprise, it became very difficult to locate its precise entry point.

Attacks using multiple sources, such as compromised computers on the Internet ("zombies") used in some types of distributed denial of service, are extremely difficult to trace and, in some cases, impossible given available information and trace back techniques. Penetrations resulting in massive data loss may be equally difficult. Earlier this year a penetration into a US credit card processing service resulted in the theft of approximately 8 million credit card numbers. The exact source of the attack has yet to be located. It is conceivable that such an attack could wreck such financial havoc on the victim that the organization would collapse and go out of business.

A big part of the reason that such attacks are difficult to trace is that the targets, effectively, "ask for it" by not preparing their systems to sustain an investigation. In both of the instances cited above the victim organizations were not prepared to defend their computing infrastructure from a determined attack and they were not prepared to investigate the attack once it was successful. Standard information protection countermeasures were not in place in either case and there were no effective logs, monitoring or other forensic information sources available with which to investigate the attack.

In such instances the investigator needs the ability to pinpoint exactly what is lacking in

investigative resources and, if possible, find ways around the gaps. One approach is to analyze what is available in minute detail. This can be problematic. Digital investigations are distinguished from other types of investigations in two very important ways. First, they may be remote crimes. That means that the attack was initiated at some indeterminate distance from the target. The attacker may have used any of a number of techniques to obfuscate his or her true location. The crime scene, literally, could extend around the world to a cyber café in a third world country.

The second distinguishing factor is the amount of data available to analyze. In a serious digital incident there can be terabytes of data that may (or may not) contain bytes of evidence. Analyzing digital attack data can be like looking for a one-inch needle in a haystack the size of North America. This problem is exacerbated by the fact that the investigator does not even know that the needle exists or, if it does, what it looks like. Structuring a complex digital investigation requires that some very important factors be addressed:

- What is the nature of the incident?
- How can we be sure that there even was an incident?
- What was the entry point into the target system? Was there only one?
- What would evidence of an attack look like? What are we looking for?
- What legal issues need to be addressed (policies, privacy, subpoenas, warrants, etc.)?
- Who was in a position to cause/allow the incident to occur?
- What security measures were in place at the time of the incident?
- What non-technical (business) issues may have impacted the success or failure of the attack?
- Who knew what about the attack and when did they know it?
- Etc.

In a complex attack these questions comprise, often, several parallel investigative threads. Sometimes these threads converge and sometimes they don't. Investigating digital crime is not like investigating any other crime, at least not exactly. However, it has much in

common with certain aspects of traditional investigation.

The overriding difference is complexity. There is no other type of crime with the potential for complexity that digital crime has. It is that very complexity that makes the digital world ideal as a venue for conducting traditional crimes such as fraud, theft and extortion. In order to investigate complex digital incidents successfully, the investigator needs tools, techniques and methods that far surpass the tools, techniques and methods required by investigators of non-digital crime.

The Genesis of a Solution

There are two ongoing debates within the digital investigative community that have relevance. The first is whether it is better to try to make an investigator out of a technologist or to teach technology to an investigator. The second debate is whether digital investigation, and, by extension, digital forensics, is art, technology or science. The only response to the first debate appropriate to this paper is that the writer has seen both approaches work and it is unlikely that the debate ever will be settled to everyone's satisfaction.

It is the second issue that impacts the techniques we discuss here. First, the writer would not disagree that there is, indeed, art in the investigation of cyber or any other type of crime. It is that art that provides the unexplainable intuition plentiful in talented investigators, regardless of their level of investigation. Forensic examiners are as much investigators as the Sherlock Holmes style detective.

Digital investigation, perhaps more than any other type of detection, involves technology. The landscape for digital crime is, potentially, highly technical. The tools and techniques used can be highly technical. In short, we see a technical landscape for technical crimes requiring a technical solution. Mix that with the art and we begin to see how a combination of art and technology defines the cyber gumshoe.

Arguably, technology is the practical output of science. However, when we think of science in the context of digital investigation and forensics, things become a bit more murky – a bit less simplistic. Part of the debate about the scientific

nature of digital forensics, for example, centers upon the fear by practitioners that by characterizing digital forensics as science, they will be required to become scientists. Nothing could (or, probably, should) be farther from the truth, at least not in the stereotypical sense.

While, to be sure, there may be a place for the stereotypical scientist in the digital forensic world, that place probably is not in the field conducting routine investigations. That does not mean that those in the field do not need to use some basic scientific principles as the basis for their work. This may be the main difference between digital forensic science and other forensic sciences. There are other forensic sciences, forensic pathology for example, that are the pure domain of scientists. Digital forensic science is, however, a mixed bag of investigators, digital forensic examiners and digital forensic scientists. They have a common ground however: reliable methods of inquiry.

Jon Nordby, writing in Forensic Science - An Introduction to Scientific and Investigative Techniques² tells us that the common ground between theoretical and forensic science is reliable methods of inquiry that possess characteristics of integrity, competence, defensible technique and relevant experience.

Nordby goes on to discuss the scientific method. Using his approach we begin to see where art, technology and science converge in digital forensics and digital investigation.

Basically, and simplistically, we know that the scientific method requires that we form hypotheses and the test those hypotheses with evidence. This is the approach a good investigator uses, whether he or she is investigating a crime or performing a forensic examination of evidence. However, Nordby is a bit more specific in his discussion of the scientific method. He forms the framework for his definition by clarifying his view of investigative science:

“Whatever the scientific investigation at issue, how one’s scientific opinion is constructed mirrors the certainty of the result. Certainty, in the medical and

² Forensic Science – An Introduction to Scientific and Investigative Techniques ed. James and Nordby, pub CRC Press, 2003

scientific sense, remains determined by the method of derivation applied in the investigation. Medical and scientific certainty remains distinctly independent from either absolute certainty or mere mathematical probability.”

He then goes on to detail “...some of the many features reliable methods implement, enabling the productive scientific investigation of facts before the court.”

“Reliable methods

- Help distinguish evidence from coincidence without ambiguity.
- Allow alternative results to be ranked by some principle basic to the sciences applied.
- Allow for certainty considerations wherever appropriate through the ranking of relevant available alternatives.
- Disallow hypotheses more extraordinary than the facts themselves.
- Pursue general impressions to the level of specific details.
- Pursue testing by breaking hypotheses (alternative explanations) into their smallest logical components, risking one part at a time.
- Allow tests either to prove or disprove alternative explanations (hypotheses).”

Thus we probably can say with a fair amount of confidence that a competent digital investigator or forensic examiner will apply “science” in his or her work if that work is to be successful.

This scientific approach effectively ends the debate because we are not, necessarily, referring to stereotypical scientists with PhDs, dressed in white lab coats carrying out their duties in sterile clean-room laboratories as the definition of the digital forensic practitioner. Rather, we are referring to individuals who exhibit integrity and competence in their investigative approach. They use defensible techniques and possess relevant experience. In short, they are professionals who use reliable methods of inquiry. Digital forensics, then, takes its rightful place with the other forensic sciences.

Adding the Legal Dimension

Nordby and others have defined forensic science as the application of natural science to matters of law. That clearly frames the context for forensics of any kind. However, digital forensics and digital investigation do not derive from the natural sciences. Rather they derive from computer science and mathematics. Thus we might extend the common definition of forensic science, in the case of digital forensic science, as the application of computer science and mathematics to matters of law. The clear common ground here, of course, is that we applying science to the law.

So, we must look to the law for help in how we practice our science, and, most especially, the tools, methods and techniques that we use. The clear guidance here comes from *Daubert v Merrell Dow Pharmaceuticals*. This landmark case has, along with a few others not quite as important, defined what it means to present scientific evidence in a court of law in the United States. Out of that case we have the four “Daubert Tests” for scientific method and the evidence gathered therewith:

1. Whether the theory or technique in question can be and has been tested.
2. Whether it has been subjected to peer review and publication.
3. Its known potential rate of error along with the existence and maintenance of standards controlling the technique’s operation.
4. The degree of acceptance within the relevant scientific community.

These four tests help us determine how we apply reliable methods in the context of digital forensics. Certainly, if we wish digital forensics to be considered scientifically valid, we must show that our tools methods and techniques are defensible, both from a technical and scientific perspective and from the perspective of the law. This is where constructs such as EEDI and DIPL are of significant value.

IDENTIFICATION	PRESERVATION	COLLECTION	EXAMINATION	ANALYSIS	PRESENTATION
Event/Crime Detection	Case Management	Preservation	Preservation	Preservation	Documentation
Resolve Signature	Imaging Technologies	Approved Methods	Traceability	Traceability	Expert Testimony
Profile Detection	Chain of Custody	Approved Software	Validation Techniques	Statistical	Clarification
Anomalous Detection	Time Synch.	Approved Hardware	Filtering Techniques	Protocols	Mission Impact Statement
Complaints		Legal Authority	Pattern Matching	Data Mining	Recommended Countermeasure
System Monitoring		Lossless Compression	Hidden Data Discovery	Timeline	Statistical Interpretation
Audit Analysis		Sampling	Hidden Data Extraction	Link	
		Data Reduction		Spatial	
		Recovery Techniques			

Figure 1- The DFRWS Digital Investigation Framework

Thus, we now have set the stage, both scientifically and legally, for the introduction of tools and techniques that allow us to bring digital forensic science into the courtroom regardless of the complexity of the investigation or incident being investigated.

ESTABLISHING A PLATFORM FOR EEDI – AN INVESTIGATIVE FRAMEWORK

The Digital Forensics Research Workshop (DFRWS) in the document cited earlier lays out a framework for the digital investigative process. This framework, shown in Figure 1 above, consists of six *classes* of tasks that the investigator must complete in the conduct of a digital investigation. Within those six classes exist individual tasks called *elements*. At least one element in each class is required and in some cases the investigator will apply multiple elements to the investigation. The DFRWS positions the tasks of the digital forensic examiner within the context of a digital investigation. In current practice that may or may not mean that the examiner and investigator are the same individual.

The DFRWS Framework classes contain key elements that are under constant review by the digital forensics community. However, there is a continuity between the classes that is important. For example, we note that the Preservation class continues as an element of the Collection, Examination and Analysis classes. This indicates that *preservation* of evidence, as characterized by case management, imaging technologies, chain of custody and time synchronization, is an ongoing requirement throughout the digital investigative process. Thus *preservation* is “...a guarded principle across ‘forensic’ categories³.” *Traceability*, likewise, is a guarded principle, but not across all forensic categories.

While space does not permit us to provide a detailed discussion of each of the elements of the Framework, a description of the classes is important⁴.

The Identification Class

The purpose of the identification class describes the method by which the investigator is notified of a possible incident. Since about 50% of all reported incidents have benign explanations⁵,

³ See footnote 1

⁴ The descriptions of DFRWS classes are taken from “Structured Investigation of Digital Incidents in Complex Computing Environments” PhD thesis (unpublished) by Peter Stephenson, Oxford Brookes University, Oxford, UK

⁵ Author’s experience over 20 years of conducting incident response

processing evidence in this class is critical to the rest of the investigation. Likewise, as it is the first step in the EEDI process, it is the only primary evidence⁶ not corroborated directly by other primary evidence. Therefore, a more significant amount of secondary evidence is needed to validate the existence of an actual event.

The Preservation Class

The Preservation Class deals with those elements that relate to the management of items of evidence. The DFRWS describes this class as "...a guarded principle across 'forensic' categories." The requirement for proper evidence handling is basic to the digital investigative process as it relates to legal actions.

The Collection Class

The Collection Class is concerned with the specific methods and products used by the investigator and forensic examiner to acquire evidence in a digital environment. As has been noted, the Preservation Class continues as an element of this Class. With the exception of the Legal Authority element, the elements of this class are largely technical.

The Examination Class

The Examination Class deals with the tools and techniques used to examine evidence. It is concerned with evidence discovery and extraction rather than the conclusions to be drawn from the evidence (Analysis Class). While the Collection Class deals with gross procedures to collect data that may contain evidence (such as imaging of computer media), the Examination Class is concerned with the examination of that data and the identification and extraction of possible evidence from it. Note that the Preservation Class continues to be pervasive in this class.

The Analysis Class

The Analysis Class refers to those elements that are involved in the analysis of evidence collected, identified and extracted from a gross

⁶ The concepts of primary and secondary evidence are discussed later

data collection. The validity of techniques used in analysis of potential evidence impact directly the validity of the conclusions drawn from the evidence and the credibility of the evidence chain constructed therefrom. The Analysis Class contains, and is dependent upon, the Preservation Class and the Traceability element of the Examination Class.

The various elements of the Analysis Class refer to the means by which a forensic examiner or investigator might develop a set of conclusions regarding evidence presented from the other five classes. As with all elements of the Framework a clear understanding of the applicable process is required. Wherever possible, adherence to standard tools, technologies and techniques is critical.

Finally, when mapping this class to the DIPL or when performing model checking, we are concerned solely with the process, not the results of the analysis or the detailing of the contents of evidentiary items.

The Link element is the key element used to form a chain of evidence. It is related to traceability and, as such, is a required element.

The Presentation Class

This class refers to the tools and techniques used to present the conclusions of the investigator and the digital forensic examiner to a court of enquiry or other finder of fact. Each of these techniques has its own elements and a discussion of expert witnessing is beyond the scope of this thesis. However, for our purposes we will stipulate that the EEDI process emphasizes the use of timelines as an embodiment of the Clarification element of this class.

SOME IMPORTANT EEDI DEFINITIONS

Before we continue, we need to define some key EEDI concepts. These definitions are taken from the writer's PhD thesis (see footnote 4).

Definition 1: Primary evidence

Primary evidence is evidence that is corroborated by other pieces of primary evidence and, in turn, corroborates additional primary evidence in a

chain of evidence. Primary evidence makes up the evidence chain in a digital investigation. Primary evidence may, in turn, be corroborated additionally by *secondary evidence*. In special circumstances, such as the first piece of evidence in a chain, sufficiently clear and obvious evidence (such as evidence that a computer has been the victim of an attack) may be considered primary evidence if it is corroborated by a significant body of secondary evidence and, in turn, corroborates other primary evidence.

Definition 2: Secondary evidence

Secondary evidence is evidence that is not, itself, corroborated but may serve to corroborate primary evidence. Secondary evidence rarely stands alone credibly since it does not have anything to support it directly. Secondary evidence may be circumstantial, for example. The presence of secondary evidence in sufficient quantity and of sufficient quality may, however, serve to tell a compelling story of how a series of digital events occurred.

These first two definitions lead to the First Rule of End-to-End forensic digital analysis:

Primary evidence should be corroborated by at least one other piece of relevant primary evidence to be considered a valid part of the evidence chain. Evidence that does not fit this description, but does serve to corroborate some other piece of evidence without itself being corroborated, is considered to be secondary evidence.

Definition 3: Forensic digital analysis

Forensic digital analysis refers to the use of the techniques of digital forensic science to perform analysis of digital events or data, whether on computer networks or on computer media.

Definition 4: Forensic digital evidence collection

The use of approved tools and techniques by trained technicians to obtain digital evidence

from computer devices, networks and media. By “approved” we mean those tools and techniques generally accepted by the discipline and the courts where collected evidence will be presented.

Definition 5: Digital forensic correlation

The comparison of evidentiary information from a variety of sources with the objective of discovering information that stands alone, in concert with other information, or corroborates or is corroborated by other evidentiary information.

Definition 6: Digital forensic normalization

The combining of evidentiary data of the same type from different sources with different vocabularies into a single, integrated terminology that can be used effectively in the correlation process.

Definition 7: Digital forensic deconfliction

The combining of multiple reportings of the same evidentiary event by the same or different reporting sources, into a single, reported, normalized evidentiary event.

Definition 8: Digital forensic data fusion

The process by which all of the available evidentiary data is analyzed and correlated into a single consistent representational model such as a timeline.

THE EEDI PROCESS

The End-to-End Digital Investigation process is a collection of generalized steps to be taken in conjunction with the DFRWS Framework. While the Framework gives a roadmap for addressing those issues comprising a formal investigation, the EEDI process provides a set of steps the investigator must perform in order to

preserve, collect, examine and analyze digital evidence.

Here, we present that collection of steps and some top level descriptions. Space does not permit us more detail, however, the following should be ample to explain the approach and the basic methods used to translate the DFRWS Framework into a practical investigative process.

We apply the EEDI process to each class of the Framework as appropriate. Note that this application of process is not slavish in that we do not map, one-for-one, process onto framework. Rather, we apply the Framework to the process in the context of ensuring that we do not miss important elements of the investigation as defined by the Framework. The Framework, then, does not represent a series of investigative steps. Instead, it represents critical elements of the digital investigation. EEDI represents the application of process to those elements.

The basic End-to-End process consists of:

- Collecting evidence
- Analysis of individual events
- Preliminary correlation
- Event normalizing
- Event deconfliction
- Second level correlation (consider both normalized and non-normalized events)
- Timeline analysis
- Chain of evidence construction
- Corroboration (consider only non-normalized events)

Collecting Evidence

The collection of evidence in a computer security incident is very time sensitive. When an event occurs we have the first warning of a potential incident. An event may not be, by itself, particularly noteworthy. However, taken in the context of other events, it may become extremely important. From the forensic perspective we want to consider all relevant events whether they appear to have been tied to an incident or not. From the definitive point of view, then, events are the most granular elements (at the “atomic” level) of an incident.

An incident is defined as a collection of events that lead to, or could lead to, a compromise of

some sort. That compromise may include unauthorized disclosure or modification of a system or its data or destruction of the system or its data. An incident becomes a crime when a law or laws is/are violated.

As soon as possible, in the context of an incident, collecting evidence from all possible locations where it may reside must begin. The methods vary according to the type of evidence (forensic, logs, indirect, traditionally developed, etc.). It is important to emphasize that EEDI is concerned not only with digital evidence. Gathering witness information should be accomplished as early in the evidence collection process as possible. Witness impressions and information play a crucial role in determining the steps the forensic examiner must take to uncover digital evidence.

Critical in this process are:

- Images of effected computers
- Logs of intermediate devices, especially those on the Internet
- Logs of effected computers
- Logs and data from intrusion detection systems, firewalls, etc.

Analysis of Individual Events

An alert or incident is made up of one or more individual events. These events may be duplicates reported in different logs from different devices. These events and duplications have value both as they appear and “normalized” (see below). The first analysis effort should be to examine these isolated events and assess what value they may have to the overall investigation and how they may tie into each other.

Preliminary Correlation

The first correlation step is to examine the individual events and see how they may correlate into a chain of evidence. The main purpose is to understand in broad terms what happened, what systems or devices were involved and when the events occurred.

Event Normalizing

There may be some events that are reported from multiple sources using different syntaxes. During part of the analysis (timeline analysis for example) these duplications must be eliminated. This process is known as normalizing. EEDI uses, eventually, both normalized and non-normalized events.

Event Deconfliction

Sometimes events are reported multiple times from the same source. An example is a denial of service attack where multiple packets are directed against a target and each one is reported by a reporting resource. The EEDI process should not count each of those packets as a separate event. The process of viewing the packets as a single event instead of multiple events is called deconfliction.

Second Level Correlation

This is just an extension of earlier correlation efforts. However, at this point views of various events have been refined through normalization or deconfliction.

Timeline Analysis

In this step normalized and deconflicted events are used to build a timeline. This is an iterative process and should be updated constantly as the investigation continues to develop new evidence. The entire event analysis, correlation, deconfliction and timeline analysis is iterative.

Chain of Evidence Construction

Once there is a preliminary timeline of events, the process of developing a coherent chain of evidence begins. Ideally each link in the chain, supported by one or more pieces of evidence, will lead to the next link. That rarely happens in large-scale network traces, however, because there often are gaps in the evidence-gathering process due to lack of logs or other missing event data.

Corroboration

In this stage we attempt to corroborate each piece of evidence and each event in our chain with other, independent, evidence or events. For this we use the non-correlated event data as well as any other evidence developed either digitally or traditionally. The best evidence is that which has been developed digitally and corroborated through traditional investigation or vice versa. The final evidence chain consists of primary evidence corroborated by additional secondary evidence. This chain will consist of both digital and traditional evidence.

The overall EEDI process does not differ materially between an investigation and an event post mortem.

TOOLS AND TECHNIQUES USED WITHIN EEDI

Implementation of EEDI in very complex digital environments requires some basic tools and techniques not common in other investigation types. Some of these tools and techniques are in use in other types of investigations such as complex cases of fraud. We describe, briefly, a few of those tools and techniques.

Determining if an Attack Actually Occurred

As mentioned earlier, a significant proportion of apparent attacks are, simply, anomalous network or computer events. Often, the Identification class will present such an event as “security-relevant” (for example, signature resolution, anomalous detection and certain types of system monitoring can give false positives resulting in a reported event).

While there are a number of ways to investigate the actual nature of such a report, there is one way that, often, can act as an important “sanity check”. It is important to note that this, and other EEDI techniques, should never be taken in isolation. No result of a single test is absolute. The need for corroboration is critical. That criticality is nowhere as important as it is in the Identification class because making a wrong determination at that point in the investigation creates a series of investigative steps that

proceed from a false premise: that an attack occurred when it actually did not.

This technique involves taking deconflicted and normalized data for all security-relevant events from some point in time preceding the incident and mapping them onto a standard spreadsheet on a day-by-day basis. Thus, the investigator has a sheet where for each day there are some number of occurrences of each event, recorded by all reporting sources and fully normalized/deconflicted to present the simplest picture of events over time.

The second step is to graph this spreadsheet. Note the peaks of pre-incident events. Investigate those events and determine if there is a benign (from the security perspective) explanation for them. If the investigator can provide plausible explanations for all pre-incident events of consequence, there probably was no attack. Again, be sure to corroborate this finding using other investigative techniques.

Determining Premeditation

In most criminal laws the element of intent is important in determining if the law has been violated. This technique, to help determine intent, is an extension of that for determining if an attack has taken place. There is an important caveat here, however. Not all intentional attacks exhibit pre-incident activity. In these cases both techniques (determining premeditation and determining the existence of an attack) are, of course, invalid.

To determine intent, take the same graph prepared in the technique of determining if an attack occurred. Examine the peaks for two important characteristics: the nature of the event and the source of the event. The nature of the event should represent a logical event preceding an attack. Examples include probes, port scans and other reconnaissance activities.

The source of the event is more difficult because source addresses are easy to spoof. We address spoofing in the next topic. The pre-incident events should be traceable to the same source or group of sources in the case of a cooperative attack (one where more than a single source participates in the attack). If the investigator can achieve the goal of identifying a common source or set of common sources and a set of logical

pre-attack activities, there is a probability of intent. As in all EEDI techniques, it is important to corroborate this finding with secondary evidence.

Determining if a Source Address Has Been Spoofed

A clever attacker will attempt to obfuscate his or her true location by impersonating, or “spoofing” a different source IP address from his or her actual address. This requires one of two things. The first possibility is that the attacker locates an IP address somewhere on the Internet (or within the enterprise if the attack is internal to the organization) that is not in current use. The attacker configures his or her computer for that address. Of course, this usually means that, due to the nature of Internet routing, not data will return to the attacker, but this technique usually is used for “one-way” attacks such as planting Trojan horses.

The second method requires that the attacker disable the computer with the legitimate address before assuming that address for his or her computer. In both cases it is likely that the actual location of the attacker is a different number of hops from the victim than is the location of the spoofed address.

Researchers at the University of Michigan and Cal. Tech. have developed a technique for determining the probability that an address is being spoofed.⁷ The technique is as follows:

- Extract the final TTL from the packet header – call it T
- Extract the source address from that packet – call it S
- Infer the initial TTL from the type of packet, standards, etc. (remembering that a packet can be crafted with a different TTL than the standard) – call that T_0
- Calculate the hop count (Hc) by $Hc = T_0 - T$
- Perform a trace route (try several to ensure that you have a good average

⁷ “Hop-Count Filtering: An Effective Defense Against Spoofed Traffic”, Jin, Chang, Haining Wang, Kang G. Shin, U Michigan & CalTech

- number of hops) to S to get the stored hop count, H_s
- If $H_c \neq H_s$ you may have a spoofed packet.

Knowing that the packet may be spoofed is, at least, a great time saver. If the real address is not in use, of course, your attempt to traceroute will end in failure and the rest of the technique becomes unnecessary.

Using Link Analysis

Link analysis is a technique borrowed from complex fraud investigations. Powerful link analyzers such as I2⁸ are able to collect such important data as IP address pairs and search for correlations between them. We will not expound upon the specifics of the techniques here, but briefly, the idea is to use link analysis to infer solutions to holes in the chain of evidence.

It is not uncommon for a back trace on a complex network such as the Internet to result in points where no evidence is obviously available. Examples are sites without logs, situations where addresses have been spoofed and instances where evidence is, at best, murky and difficult or impossible to corroborate.

What is required to use link analysis effectively to infer solutions to these problems is a very comprehensive set of IP address pairs from as many locations in the verified chain of evidence as possible. Sources for these address pairs are intrusion detection systems, host logs, firewall logs, router routing tables and sniffer traces.

The link analyzer looks for associations in these address pairs, one pair with another, and reports the associations allowing the investigator to infer individual routes within the suspected attack path. While it is true that inference is not evidence, inference does, often, provide substantial leads that, when followed, may result in hard evidence or, at least, corroboration of other evidence.

⁸ <http://www.i2inc.com/>

The Digital Investigation Process Language (DIPL)

DIPL is a formal process language, loosely derived from LISP, that allows the characterization of an investigation in formal terms. Without going into the language in detail, we may describe it as formal (mathematically provable), list-oriented and built upon its own syntax and vocabulary. It is not a computer language and does not contain such constructs as looping or if-then-else. Figure 2 shows a very brief example of a listing.

```
(ManageCase
  (Initiator
    (RealName 'Peter Stephenson')
  )
  (CaseName 'Case123')
  (BeginTime 21:05 1 Jan 1998)
)
(TraceAuthority
  (ApprovedSoftware
    (Tool
      (ProgramName 'SafeBack')
      (VersionNumber '3.0')
    )
    (Citation
      (CaseName 'joe v volcano')
    )
  )
)
)
```

Figure 2 – An Example of a DIPL Listing

In this example we show that the investigator (Peter Stephenson) has opened his case notes for Case Number 123 on 1 January 1998 at 21:05. He then established that the tool being used was SafeBack version 3.0 and that it had been court challenged in a case called joe v. volcano. The language actually is capable of far more complex representations and is acceptably rich for the purpose of characterizing the investigative process fully.

SUMMARY

The EEDI process offers a comprehensive approach to complex digital investigations. Its primary benefit is its ability to allow an investigator to use and document a scientific approach, consistent with the Daubert tests, to solve, document and present a complex digital investigation. Additionally, it lends itself to

future enhancements such as case modeling and simulation.

Author Bio:

Peter Stephenson, Executive Director of the International Institute for Digital Forensic Studies, is a writer, consultant, researcher and lecturer in information protection and forensics on large-scale computer networks. He has spoken extensively on digital forensics and security, and has written or contributed to 14 books and several hundred articles in major national and international trade publications. He has lectured and delivered consulting engagements for the past 17 years in eleven countries plus the United States.

He may be reached at:
pstephenson@iidfs.infoforensics.org.