

# Structured Investigation of Digital Incidents in Complex Computing Environments

Peter Stephenson  
Oxford Brookes University,  
School of Computing and Mathematical Sciences, Oxford, UK

***Abstract:** Critics, unimpressed by the rigor of the forensic digital examination process, have taken the position that forensic digital analysis is, more rightly, simply little more than ad hoc data collection and analysis. The reality is that forensic digital analysis as a whole, in its relative infancy, is the unwilling victim of the rapid advancement of computer technology.*

*Struggling to keep pace with the giant steps taken in recent years by the computing industry, forensic digital analysis is, as are many defensive computing technologies, at the mercy of ever more new and complex computing approaches. Chief among these new paradigms is the need to analyze forensic materials over complex chains of evidence that may range around the globe over a wide variety of heterogeneous computing platforms, environments and transports.*

*This paper discusses a formalized approach to the forensic collection, management and analysis of digital evidence involved in complex cases occurring over complex networks. Its objective is to begin the processes of instilling the same rigor in the practice of forensic digital analysis that exists in many other branches of forensic science.*

---

## 1.0 Background and Problem Statement

In August of 2001 over 50 university researchers, computer forensic examiners and analysts attended the first Digital Forensic Research Workshop (DFRWS) in Utica, New York. Although the objective of the workshop was the forming of the beginnings of a community to engage in research involving digital forensics, much more actually came out of the effort. Perhaps the most important contribution to the near term goals of incorporating appropriate rigor into the science of forensic digital analysis was the definition of what the state of the practice must be to approach the status of an accepted science. [DFRWS01]

The workshop defined a “generic investigative process that can be applied to all (or the majority of) investigations involving digital systems and networks.” The generic process, as defined at the time is:

- Identification
- Preservation
- Collection
- Examination
- Analysis
- Presentation
- Decision

For our purposes we will accept this top level process and develop our approach around it.

## 1.1 PROBLEM STATEMENT

Clearly the workshop attendees, as well as a number of other writers, have come to the conclusion that to be considered a “real” science, digital forensics must undergo a maturing process. The state of the practice today is tool-centered, not process-centered. Processes, where they exist, are associated with tools and with individual platforms.

In considering the maturing process, we may, once again, turn to the DFRWS for a definition of digital forensic science:

*“The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations.”*

This definition gives us a number of benchmarks for determining the state of the practice at any particular time. In particular, today, the digital investigative process (DIP) lacks those *scientifically derived and proven methods*. The beginning of meeting the Workshop’s definition is the formalization of a process for digital investigation. A formalized DIP would allow the verification or validation of a particular investigation’s process and, thus, would move the drawing of conclusions from the digital evidence forward with credibility.

In this paper we show a method for adding structure to the DIP using a process language developed explicitly for the purpose of describing or specifying the DIP for a particular investigation. The core of the language (Digital Forensic Process Language – DFPL) is the Common Intrusion Specification Language (CISL) developed for the purpose of describing

attacks or intrusions over a digital network. DFPL, then, is a superset of CISL designed explicitly for the description or specification of an investigation’s digital investigative process.

The use of DFPL to plan a digital investigation helps to assure that appropriate rigor is applied to the DIP. The use of DFPL to describe a completed investigation helps to validate (or invalidate) the DIP used in the investigation.

While DFPL is not, strictly speaking, a formal language (i.e., mathematically provable), it is the first step towards true formalization of the digital investigative process. Derived from CISL which, in turn, was derived from LISP (a formal language proved using the Lambda Calculus), DFPL, at least, has a pedigree that suggests credibility. In the section of this paper on Future Work, we discuss possible next steps for taking the DIP process to a mathematically provable state.

---

## 2.0 Prior Work

Most of the prior work in the area of structuring the DIP has been done by the Digital Forensic Research Work Shop. A paper by Mark Reith, Clint Carr, and Greg Gunsch of the Air Force Institute of Technology [RCG02] offers some ideas and extends the notions developed by the DFRWS.

Marcus Ranum, developer of Network Flight Recorder, a network intrusion detection system with a strong network forensic data gathering capability, has made significant progress in the area of network forensics. Several other writers have focused upon the difficulties of and potential solutions to network backtracing, despoofing and dealing with large zombie networks such as are found in certain distributed denial of service attacks.

There is, however, virtually no prior work on the synthesis of formal modeling and the DIP. Current work focuses upon a structured process for digital investigation without offering a

method for proving that the process has been followed.

---

### 3.0 Hypothesis

If digital investigators are to ensure themselves and the courts that their investigations are valid, they need some level of benchmark to test their processes against. On the surface that sounds fairly easy. Certainly the steps suggested by the DFRWS appear to fit that requirement. However, in reality, adherence to an abstract set of generalized rules allows disagreement over whether and to what degree a process was followed. In a true science, such disagreement should be very difficult, if not impossible.

Additionally, digital investigations that involve the tracing of a penetration across a complex computing environment are very difficult to conduct and prosecute. The approach presented here involves two important concepts.

First, we discuss the notion of end-to-end forensic digital analysis. This is the idea that a complex attack begins with the attacker and ends with the victim. It includes every computing device, platform and transport mechanism along the way, each of which may contain useful evidence. In order to support an end-to-end investigation, we need to provide a corroborated or *linked* chain of evidence over the entire path of the attack. An end-to-end investigation, properly supported and corroborated can offer a solid case that can be presented in a fashion understandable by lay finders of fact.

Second, we discuss first steps toward formalizing the end-to-end process. Since the DIP, in its entirety, implies a full understanding of the attack methods as well as the detection methods used to discover them, a formally supported end-to-end digital investigation must incorporate the same elements. The use of DFPL allows both the investigative process and

the embedded attack(s) to be described fully and with appropriate structure.

These two elements (end-to-end DIP and DFPL) do not, and should not, address conclusions drawn from the investigation. They simply describe the process and the attack(s) with sufficient rigor to allow courts to conclude that the DIP was conducted properly. This reduces the arguments in a digital case to the validity of the conclusions since a validated DIP is no longer the subject of debate. Similarly, if one side can show, using DFPL and the end-to-end process, that the DIP was not conducted properly by the opposing side, evidence collected by the opposition in the investigation may be discredited based upon a flawed investigative process.

---

## 4.0 Objectives and Approach

There are three clear objectives that we must address. First, we must conduct a thorough and complete investigation, systematically covering the entire chain of events, corroborating our findings appropriately, and ensuring that our process is documented completely. Second, we must be able to establish, credibly, that we have done so. Finally, the use of the Digital Forensics Process Language is intended to provide a first step in validating the DIP formally, consistent with the DFRWS goals of attaching it to a specific disciplinary practice considered to be scientifically rigorous.

### 4.1 APPROACH

We begin with a discussion of the end-to-end digital investigative process, and follow up with a description of the DFPL. The end-to-end process deals, largely with two kinds of evidence: evidence that fits directly in the chain and corroborates other evidence and is itself corroborated directly, and evidence whose only purpose is corroboration of other evidence without itself being corroborated. We will term

the former *primary evidence* and the latter *secondary evidence*.

Where secondary evidence applies, usually a single piece of such evidence is insufficient to achieve credibility and avoid challenge. This is especially true when we are depending upon a piece of secondary evidence to support a piece of primary evidence within a chain.

This leads to the first rule of end-to-end forensic digital analysis:

*Primary evidence must always 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.*

Secondary evidence, in appropriate quantity may be very useful and may, in fact, serve to tell a very compelling story of actual events. Secondary evidence also may be, on occasion, referred to as circumstantial evidence. However, not all secondary evidence is only circumstantial and care should be exercised in interchanging the two terms. [PRS21]

In the development of the techniques described in this paper, we used a three-phase approach. Phase one was the development, refinement and testing of the end-to-end process. End-to-end was documented first in 2000 and first used early in 2001. Early in the process of applying it to ongoing investigations we refined it and began demonstrating it to groups of corporate and law enforcement investigators. End-to-end received generally favorable reviews along with some suggestions for improvement, many of which we incorporated.

Phase two was the extending of the Common Intrusion Specification Language (CISL) to accommodate forensic tasks. This was made easier by the development of the general model

for the digital investigative process by the DFRWS. As the DFRWS model was a consensus model, we adopted it for the purpose of beginning with an acceptable investigative structure. To enhance the CISL we added Semantic Identifiers (SIDs) in the areas of investigative and forensic verb SIDs, investigative and forensic atom SIDs, and investigative and forensic role SIDs. We will discuss the structure of CISL in Section 4.1.2 below.

The final phase, reserved for future work, is the extension of the Digital Forensics Process Language (DFPL) to a process algebra such as CSP (Communicating Sequential Processes) or Colored Petri Nets. This final formalization would enable the machine analysis of an investigative process, including the incident itself, potentially pinpointing flaws in the process and incident description. While the process does not, and should not, draw conclusions regarding the perpetrator of the incident, it could help ensure that those conclusions are not drawn from flawed assumptions.

#### **4.1.1 THE END-TO-END PROCESS**

In practice, the end-to-end process can be very complex. The tracing of an attack from the victim back to the attacker often is very difficult and may, under certain circumstances, be impossible using only network back tracing techniques. Therefore, the end-to-end process allows the introduction of evidence collected in traditional investigations. Such evidence usually is secondary evidence, however, even though it may be instrumental in leading investigators to a particular intermediate site for which network back tracing provided no useful results. Once a suspected intermediate site is identified, traditional forensic digital examination should be carried out to extract appropriate primary digital evidence that the site was used in the attack.

The end-to-end DIP is based upon the creation of a corroborated chain of digital evidence. It is a generalized approach in that it is not platform or transport dependent for its success. It allows the use of whatever digital analysis technique is appropriate at each point along the chain as long as the analysis technique used is, itself, accepted for the purpose. Thus, evidence along the chain may include evidence extracted from individual devices using forensic examination tools, system logs, sniffer traces, intrusion detection logs, firewall logs, configuration and rules files, etc.

Evidence collected along the chain is analyzed first as discrete evidence, that is, as individual pieces of evidence standing alone. However, in complex systems, the same events may be collected in a variety of ways. For example, a probe of a firewall may elude the intrusion detection system (IDS), but be picked up by the firewall's system logs. Or, the same probe may be seen both by the firewall and the IDS. However, if the latter is the case we note two points. First, we have two instances of the same probe. And, second, those two instances appear with different formats or names due to the different ways of detecting them.

As it happens, we have need of the evidence in a variety of ways. First, we need to know that the event occurred and, although it was seen twice, only occurred once. Second, we need to know that, although the two sensors (system logs and IDS) reported the probe differently, it is actually the same probe. Finally, we need to know that the probe occurred at some particular time. To reconcile the second issue we perform a function called *normalizing*. This technique views the event as a single event even though it was reported with different descriptive names by two different sensors. To reconcile the first, we *deconflict* the data. That means that we recognize that although the probe was reported twice, really it happened only once.

In presenting the end-to-end analysis, we use the deconflicted data to create a timeline of events and the normalized data to describe the

event. We use the raw data as corroborating evidence since we can see the event in several different (and independent) sensors. The timeline analysis becomes the chain of primary evidence laid out on a time scale for easy interpretation by lay finders of fact. The process of connecting the links in the chain and corroborating or linking each one is referred to as *correlation*.

Clearly, this process, due largely to its complexity, is most appropriate for use in investigating digital events that result in penetration and occur over a complex computing environment. The end-to-end process must include, to be credible, each and every step in the attack process, laid out in temporal order and should include pre-attack (probes, scans, searches for information about the victim, etc.) and post-attack (back dooring, execution of Trojan horses, etc.) evidence.

#### **4.1.2 THE DIGITAL FORENSICS PROCESS LANGUAGE**

DFPL is an enhancement of the CISL. CISL was, most recently, updated in 1999 and was intended to provide a "language that can be used to disseminate event records, analysis results, and countermeasure directives amongst intrusion detection and response components." [RF99] It was found to be unsuitable for that purpose by Doyle at MIT [JD99]. However, CISL still offers a rich environment for characterizing the digital forensic investigative process. It adds the useful notion of attack description as well. Thus, the attack and the investigative process may be analyzed and described formally from the perspective of the investigator.

CISL was derived from the LISP language. LISP is underpinned by the Lambda Calculus, however, CISL evolved from LISP, not the Lambda Calculus. The basic construct of CISL is the S-Expression, first defined by Rivest in 1997. An S-expression is a data structure that is "...suitable for representing arbitrary complex

data structures.” [RR97] S-expressions may be byte strings or lists of simpler S-expressions.

CISL takes S-expressions and enhances them with Semantic Identifiers (SIDs). SIDs are Tags added at the beginning of an S-expression that give a semantic clue to the interpretation of the rest of the S-expression. [RF99] SIDs fall into several broad groups:

- Verb SIDs
- Role SIDs
- Atom SIDs

- Conjunction SIDs
- Referent SIDs

In order to extend the CISL for use describing the DIP, we added specialized verb SIDs, role SIDs and atom SIDs. [PRS22] The process of extending CISL is ongoing at present. Refinement as a result of use of DFPL on test cases continues to suggest additions of new SIDs to the vocabulary.

A fragment typical of CISL is shown in Figure 1 below.

```
(OpenApplicationSession
  (When
    (Time 14:57:36 24 Feb 1998)
  )
  (Initiator
    (HostName 'big.evil.com')
  )
  (Account
    (UserName 'joe')
    (RealName 'Joe Cool')
    (HostName 'ten.ada.net')
  )
  (Receiver
    (standardTCPPort 23)
  )
)
```

Figure 1. Example CISL Listing [RF99]

A simple example of application of the extended DFPL appears in Section 4.2 below. DFPL preserves many of the constructs of the CISL while eliminating those constructs peculiar to the exchange of data between intrusion detection systems. We refer the reader to [RF99] for a detailed description of the CISL.

## 4.2 A SIMPLE EXAMPLE

We begin the example by describing an investigation against which we applied the

DFPL after the fact. The salient points of the investigation were as follows.

A disgruntled system administrator left the employ of a small credit union owned by a larger corporation. Some time later, employed as a contractor to the larger corporation, he attacked the credit union’s NT server he once had administrated. He deleted a mortgage database and an investigation followed ending in the suspect’s admission of guilt. The detailed events of the incident and subsequent investigation, arranged according to the DFRWS model appear in Figure 2 below.

- **Identification**
  - Call received
- **Preservation**
  - Case file opened
  - Server imaged
    - Image in chain of custody
  - Server logs preserved
  - Entry in case file
- **Collection**
  - Policies reviewed for authority to proceed
  - SafeBack used
  - Began interviews
  - Event described
    - Unavailable mortgage database
    - Server checked: db gone
    - Observed action by admin including remote login
    - Restore from backup unsuccessful – data bad
  - Entry in case file
- **Examination**
  - Data recovered from server drive
    - Database deleted and partially overwritten
    - Placed in chain of custody
    - Entry in case file
  - Data recovered from server logs
    - Login by admin from a network connection
      - Gateway address
      - Attack PC address and name
    - Placed in chain of custody
    - Entry in case file
  - Data recovered from gateway logs
    - Time & date of access to gateway by attack PC
    - IP address of attack PC
    - Entry in case file
    - Placed in chain of custody
  - Data recovered from attack PC
    - Policies reviewed for authority to proceed
    - SafeBack used
    - Placed in chain of custody
    - Login info re: victim recovered
    - Authentication data for victim recovered
    - Attack PC username recovered: suspect identified
    - Suspect logged in at time of event
    - Entry in case file
  - Data recovered from floor swipe card access log
  - Placed in chain of custody
  - Entry in case file
  - Witness interviews



```

)
  (ApprovedMethod      [Imaging method is approved]
    (Certification
      (Certifier
        (RealName 'NTI')
        (CertType 'NTI Training')
        (CertNumber 'Course 1-1-95')
        (Observer
          (RealName 'Peter Stephenson')
        )
      )
    )
  )
)
  (Policy              [Policy is examined and OK]
    (PolicyName 'Information Privacy Policy')
    (PolicyDate '1 Jan 1990')
    (Observer
      (RealName 'Peter Stephenson')
    )
  )
)
(Image                [Take image]
  (Initiator
    (RealName 'Peter Stephenson')
  )
  (Tool
    (ProgramName 'SafeBack')
    (VersionNumber '3.0')
  )
  (BeginTime 17:00 1 Jan 1998)
  (EndTime 20:14 1 Jan 1998)
  (Target
    (HostName 'Server1')
  )
  (ReferAs 0x12345678)
  (PreserveCustody
    (Evidence
      (ReferTo 0x12345678)
    )
  )
)
)
(ManageCase          [Entry in the case file]
  (Initiator
    (RealName 'Peter Stephenson')
  )
  (CaseName 'Case123')
  (BeginTime 20:25 1 Jan 1998)
)
(ExtractData         [Log extracted from image]
  (Evidence
    (FileName 'server.log')
    (ReferAs 0x87654321)
  )
  (Target
    (ReferTo 0x12345678)
  )
  (PreserveCustody [Entered in chain of custody]
    (Evidence
      (ReferTo 0x87654321)
      (BeginTime 20:45 1 Jan 1998)
    )
  )
)

```

```

)
)
)
(Interview                               [Begin interviews]
  (Initiator
    (RealName `Peter Stephenson`)
  )
  (Subject
    (RealName `Jane Sneaker`)
  )
  (BeginTime 08:30 2 Jan 1998)
  (EndTime 10:45 2 Jan 1998)
)
(ManageCase
  (Initiator
    (RealName `Peter Stephenson`)
  )
  (CaseName `Case123`)
  (BeginTime 21:05 1 Jan 1998)
)

===== Listing Continues =====

)

```

Figure 3 – Partial DFPL Listing of Credit Union Investigation

---

## 5.0 Conclusions

The DFPL shows promise for the formal characterization of a digital forensic investigation. The ability to validate an investigation and included incident can be a valuable beginning to the formalization of the digital forensic process. The DFRWS concluded:

“To be considered a discipline, Digital Forensic Science must be characterized by the following associated entities:

- **Theory:** *a body of statements and principles that attempts to explain how things work*
- **Abstractions and models:** *considerations beyond the obvious, factual, or observed*
- **Elements of practice:** *related technologies, tools, and methods*
- **Corpus of literature and professional practice**

- **Confidence and trust in results:** *usefulness and purpose*

The current state of Digital Forensic Science exhibits only some of these characteristics and they are not tied to specific disciplinary practices considered by any group as scientifically rigorous.” [DFRWS01]

It appears that the DFPL provides a beginning for associating digital forensic science with “specific disciplinary practices” that may be considered to be “scientifically rigorous” through the use of accepted formalization techniques. The DFPL preserves the structure of the CISL which, in turn, reflects LISP, a formally proven language.

---

## 6.0 Future Work

Future work on the DFPL is focused upon further formalizing the approach and developing a benchmark for a properly

structured investigation. Some possible tasks include:

1. Fully document the DFPL
2. Extend the DFPL using CSP and/or Petri Nets
3. Formally prove a DFPL process using the Lambda Calculus or similar technique
4. Develop structure “templates” that define “measuring sticks” for a complete and correct investigation

Plans are underway by a third party for accomplishing the CSP portion of the second task and the author will pursue tasks one and three as well as the use of Petri Nets to create a formal model of a digital investigation. The author feels that it would be most appropriate for a body such as the DFRWS to undertake task four since the continuation of the consensus approach is important to the success of the process.

---

## 7.0 Author

Peter Stephenson, CPE, PCE, CISSP, FICAF is a director of research for QinetiQ Trusted Information Management. He is the author of several books and numerous articles in computer trade publications. Stephenson is a PhD candidate at Oxford Brookes University, Oxford, UK. Prior to joining QinetiQ-TIM in 2002, Mr. Stephenson operated a security consulting practice for 15 years and was director of technology for the Global Security Practice of Netigy Corp. He is the developer of the Intrusion Management taxonomy for information protection, a structured process for network vulnerability assessment and the S-TR AIS method for standards-based security requirements engineering.

---

## 8.0 References

[DFRWS01] Report from the First Digital Forensic Research Workshop. *DTR-T001-01 FINAL A Road Map for Digital Forensic Research*. Final version, November 6, 2001. <http://www.dfrws.org>

[PRS21] Stephenson, Peter. *Getting the Whole Picture, Collecting Evidence of a Computer Crime*. “Computer Forensics and Security”, Elsevier, November 2002.

[RF99] Feiertag, Rich et al. *A Common Intrusion Specification Language (CISL)*. Paper published and last revised 11 June 1999.

[JD99] Doyle, Jon. *Some Representational Limitations of the Common Intrusion Specification Language*. Paper sponsored under a grant from DARPA and published (revised version) 5 November 1999.

[RR97] Rivest, Ron. *S-Expressions*. Paper published 4 May 1997.

[PRS22] Stephenson, Peter. *Digital Forensics Process Language (DFPL) Language Definition Document*. Unpublished reference.

[RCG02] Reith, Mark, Carr, Clint and Gunsch, Gregg. *An Examination of Digital Forensic Models*. Published in the “International Journal of Digital Evidence” Fall, 2002, Volume 1, Issue 3