CERIAS Tech Report 2004-37

DEFINING EVENT RECONSTRUCTION OF DIGITAL CRIME SCENES

by Brian D. Carrier and Eugene H. Spafford

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

J Forensic Sci, Nov. 2004, Vol. 49, No. 6
Paper ID JFS2004127
Available online at: www.astm.org

TECHNICAL NOTE

Brian D. Carrier,[1] *M.S. and Eugene H. Spafford,*[1] *Ph.D.*

Defining Event Reconstruction of Digital Crime Scenes*

ABSTRACT: Event reconstruction plays a critical role in solving physical crimes by explaining why a piece of physical evidence has certain characteristics. With digital crimes, the current focus has been on the recognition and identification of digital evidence using an object's characteristics, but not on the identification of the events that caused the characteristics. This paper examines digital event reconstruction and proposes a process model and procedure that can be used for a digital crime scene. The model has been designed so that it can apply to physical crime scenes, can support the unique aspects of a digital crime scene, and can be implemented in software to automate part of the process. We also examine the differences between physical event reconstruction and digital event reconstruction.

KEYWORDS: forensic science, computer forensics, digital evidence, crime scene event reconstruction, digital investigation

Currently, most of the attention in digital investigations has focused on the search for and collection of digital evidence. The "computer forensic" tools on the market preserve the state of a system or examine a system to find evidence, but they do not try to identify why an object may be evidence. Collecting an object and examining its properties is interesting, but for the evidence to be useful, we must identify what caused the object to have those properties. In the physical world, this is equivalent to recognizing and collecting blood from a crime scene, but not using scientific methods to identify from whence it came.

Event reconstruction, or event analysis, examines the evidence to identify why it has its characteristics. Many events occur at a crime scene, including the ones that are considered a crime or policy violation. The events that occurred prior to the incident may need to be understood to fully explain the incident. The reconstruction phase identifies the events for which evidence exists to support their occurrence. Conceptually, this phase adds an additional dimension to the evidence. Instead of having information about only the final state of an object, this phase attempts to deduce the previous states by examining the events in which an object may have been involved.

In this paper, we will show a procedure that can be used to reconstruct the events at a digital crime scene. Event reconstruction is different than a recreation of the crime, where the entire crime or incident is re-enacted. Event reconstruction will question why an object has properties, where they could have come from, and when they were created. The approach of this work is based on the concept that examining a computer is analogous to examining a physical crime scene (1). The preservation, survey, documentation, search, and reconstruction phases can be applied to the digital world as they are to the physical world.

[1] Purdue University—CERIAS, Recitation Building, 656 Oval Drive, West Lafayette, IN.

* Presented at the AAFS 56th Annual Meeting, 2003 Chicago, IL.

Received 20 March 2004; and in revised form 5 June 2004; accepted 5 June 2004; published xxxx.

The work presented in this paper is more formal than other procedures for event reconstruction in the physical world because our goal is to create a model that can be used to develop software tools that can automate parts of the reconstruction process. We do not claim that an investigator must conduct a reconstruction of events with the same level of formalism. In many simple incidents, the reconstruction process occurs unintentionally in the investigator's mind. It remains to be seen how technically feasible it is to reconstruct incidents with the level of detail shown here, but this framework can be used to direct future research and compare different techniques.

In section 1 of this paper, we provide background material on digital crime scene investigation and physical crime scene event reconstruction. Section 2 describes our process model in an abstract sense and Section 3 applies it to a digital crime scene. Section 4 discusses some of the challenges of digital crime scene reconstruction and Section 5 concludes this paper.

Background

Digital Crime Scene Investigation

A computer being investigated can be considered a digital crime scene and investigated as a subset of the physical crime scene where it is located (1). Physical evidence may exist around a server that was attacked by an employee and usage evidence may exist around a home computer that contains contraband. Furthermore, the end goal of most digital investigations is to identify a person who is responsible and therefore the digital investigation needs to be tied to a physical investigation.

We can describe the investigation process using the same general phases that are used in physical crime scene investigations. The physical preservation phase secures the physical crime scene and detains witnesses and suspects. During the physical survey phase, the investigator walks around the crime scene and recognizes obvious evidence so that she can get the big picture of what occurred at

the crime scene. After the survey, the crime scene is documented in the documentation phase. The search phase examines the physical crime scene and includes a thorough search for any physical evidence of the crime. The reconstruction phase examines the physical and digital evidence that was collected and tries to determine what events took place at the crime scene and use those to test hypotheses about the crime.

We suggest that when a physical computer is recognized at the crime scene, then the digital investigation begins, which uses the same five major phases. The preservation phase reduces the amount of data that is overwritten on the system and a common procedure in this phase is to duplicate the data on the system and conduct the investigation in a special environment that does not modify the copy. The survey phase examines the obvious locations for evidence and develops a strategy for how to search the system for additional evidence. The system is documented and a full search is done. Most computer forensic tools help the investigator perform the survey, documentation, and search phases. The final phase, reconstruction, examines the evidence to identify what events may have occurred in the system. It is in this phase where the hypotheses about the incident will be formally developed and tested. We focus on the reconstruction process in this paper.

Previous Reconstruction Work

Miller in James et al. (2) and Lee et al. (3) describes a five-phase process for event reconstruction at a physical crime scene that reflects the scientific method. The phases are based on the process of formulating and refining hypotheses and theories about the crime and are more conceptual than an actual process. The first phase deals with the collection of evidence from the crime scene and the second phase creates an initial conjecture about events at the crime scene. The third phase formulates hypotheses about the incident as the evidence is examined and it is during this phase where, for example, blood spatter is examined and objects are analyzed for traces and impression patterns. The fourth phase tests the hypotheses about the incident and the fifth phase formulates the final theory.

Rynearson describes a "common sense" method of reconstruction (4). He focuses on evaluating a crime scene and recognizing the "individual objects, relationships between objects, or environmental observations"(4). Then "common sense reasoning" is applied to determine how the objects got there. Each object is interpreted to reveal observational clues, including relational, functional, and temporal. Relational clues come from an object's location and orientation relative to other objects. Functional clues come from the operational condition of the object and temporal clues come from "the interaction of time and environment upon the evidence," such as body temperature or body decay (4).

Rynearson's procedure is to get an initial impression of the crime scene and then begin to reconstruct each of the major events that may have occurred and develop a hypothesis. If evidence is found to refute the hypothesis, then the reconstruction needs to back up to account for the contradiction. Chisum has a similar approach as Rynearson, but does not describe an actual process in Turvey's Criminal Profiling book (5).

Bevel and Gardner (6) provide one of the only formal procedures for conducting a crime scene reconstruction. They use the term *event* to describe an occurrence at the macro level and the term *event segments* to describe the micro level events that make up an event. Their conceptual information analysis model starts when the information, or evidence, is first collected. The second phase, evaluation, examines the reliability and credibility of the information to determine if it is staged or could have been caused by the first responders. The evidence then undergoes assessment where an investigator starts to look for evidence of events and identifies the "basic nature of the segment and evidence," the "relational aspects to other segments and evidence," and the "time and sequencing aspects" (6). The final phase of the procedure is integration where everything is combined to sequence the events and break the crime into groups of events.

The first step in Bevel and Gardner's reconstruction procedure is to collect and examine the evidence. Event segments are then created from the evidence and the event segments are sequenced and grouped into larger events. Some types of incidents have known events, such as an entry and exit from the building where a crime occurred, and they can be used to sequence the event segments. After the sequence of events and event segments have been determined, a flow chart of the incident can be created.

Reconstruction has also been examined in the digital world. Casey and Turvey apply the principles of temporal, relational, and functional information to digital evidence (7). For temporal analysis they use the time information from files, logs, and witness interviews to develop timelines or histograms. They discuss using the relational aspects of the suspect, victim, and other devices to identify which attacks could have occurred and where additional evidence may exist. They use functional analysis to determine if a computer or user could have performed the events that are believed to have occurred during the crime.

Stephenson has developed a Petri net model for testing an incident hypothesis (8). The model uses "event correlation," "normalizing," and "deconfliction" techniques to transform the collected evidence into a format for the model, which will show if there is evidence to support the hypothesis. The details of the correlation, normalizing, and deconfliction processes have not been published and that is part of the process that we are examining in this paper.

A Role-based Event Reconstruction Model

The phases and procedures for a physical crime scene reconstruction can be applied to digital crime scenes, but the results present some difficulties because there is evidence at a digital crime scene that is not typically used as evidence in a physical crime. For example, the laws of Newtonian physics do not have to be considered evidence in a case that involves a physical attack and the investigator does not need to measure gravity at each crime scene so that he can prove the trajectory path of an object. With a digital environment though, the laws that are equivalent to gravity and forces are the instructions that make up the operating system and software. These instructions can be unique to every computer and may need to be used as evidence because an attacker may have modified them.

One of the goals of this work is to formally define the reconstruction process so that requirements and tools can be developed. Unlike physical evidence, all digital evidence requires tools to be used when examining it. This can make analysis more difficult, but it also has the advantage that some procedures can be more easily automated. With a formal model, we can develop requirements for the process. This section will describe our abstract model that can be applied to both physical and digital crimes.

Events

We will now examine events in more detail and start with definitions. We define *digital evidence of an incident* as any digital data that contains reliable information that supports or refutes a hypothesis about the incident. Digital and physical objects have

characteristics that help to identify them. The *state* of an object is the value of its characteristics, or the data it contains.

An *event* is an occurrence that changes the state of one or more objects. A *crime or incident* is an event that violates a law or policy. From our previous definition of evidence, we can state that an object is evidence of an incident if its state was used to cause an event related to the incident or if its state was changed by an event that was related to the incident. *Event reconstruction* is the process of determining the events that occurred at a crime scene by using evidence characteristics.

In a continuous process, which occurs in the physical world, we cannot naturally discuss individual events. Instead, we must transform the continuous process into an approximate discrete process so that distinct events can be determined and examined (9). In a computer, events can occur only at each processor cycle and therefore the code that a computer executes is already a discrete process.

The roles of objects in events have been examined in many fields. In physics, objects can be cause and effects. Artificial intelligence uses the same concepts, but sometimes uses the term preconditions instead of cause (10). Regardless of terms, we too can classify objects with respect to their roles in events. At the highest level, we can use the following roles:

- *Cause*: An object plays the role of a cause if its characteristics were used in the event. A test for this role is to identify if the same effect would have occurred if the object were to not exist. A cause object has an influence on the effect.
- *Effect*: An object plays the role of an effect if its state was changed by a cause object in the event.

Objects that are causes may be passive. That is, they are used in the event, but they are not changed by the event. If a cause object is changed by the event, then it is both a cause and an effect object. From this it follows that if an object is an effect but not a cause, then it must have been created as a result of the event. The changes to an effect object's characteristics are related to the characteristics of one or more cause objects.

The cause objects can be thought of as the tools and scientific laws that determine how an event will occur. We can redefine an event as an occurrence that uses the characteristics from one or more objects and changes the characteristics of one or more objects.

We can graphically represent this definition of an event as shown in Fig. 1(*A*). Each circle represents a state of an object and each box represents an event (9). In this graph, objects X, Y, and Z were causes of the event E and object X is also an effect, with its new state noted by X'.

In some cases, it may be possible to identify a cause object that initiated the event. The *initiator* of the event is the object or event that began the event. In many cases, it will be difficult or impossible to identify the initiator, especially for a continuous process in the physical world. For example, finding the initiator

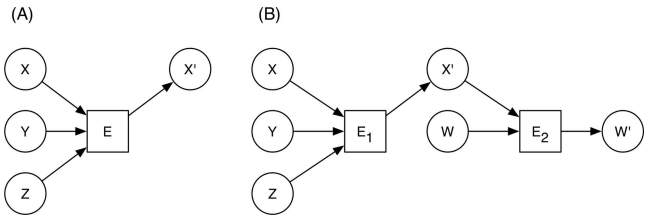

FIG. 1—*Graphical representation of (A) an event with three cause objects and one effect object and (B) an event chain with two events.*

of a head-on collision involving two cars in the middle of the road would be difficult and may not exist in the form of physical evidence. Identifying the source of an event in a discrete process may be easier than in a continuous process because there are distinct event starting times and there are a finite number of events that can occur at the same time.

An *event chain* is a sequence of events $\langle e_0, e_1, \ldots, e_k \rangle$ such that an effect of event e_i is a cause of event e_{i+1} for $i = 1, 2, \ldots, k-1$. We can see this in Fig. 1(*B*), where event E_1 causes event E_2 because of its effect on object X'.

Event Reconstruction Process

With our definition of an event and its roles, we can examine the process that occurs in the event reconstruction phase. Recall that when the reconstruction phase begins, the evidence has already been recognized at the crime scene and collected. There are five phases in the reconstruction process:

1. Evidence Examination
2. Role Classification
3. Event Construction and Testing
4. Event Sequencing
5. Hypothesis Testing

We will now discuss each of the phases in more detail. As one of our goals is to develop a model that can be used to build software tools, we also provide metrics for each phase. These can be used to compare different techniques and procedures that are implemented to perform event reconstruction.

Evidence Examination

The first step in the reconstruction process is to fully examine the evidence. A cursory examination of most evidence occurs at the crime scene so that an investigator can recognize it as evidence. The goal of this phase is to identify all of the object's relevant information and identify what characteristics it has. At the end of this phase, we will have a list of characteristics for each piece of evidence.

In this phase an object will be *identified* using its class characteristics and *individualized* using its individual characteristics. The *class characteristics* of an object are those that "can be associated only with a group and never with a single source" and the *individual characteristics* of an object are those that "can be associated with a common source with an extremely high degree of probability" (11). The reliability and credibility of the object's characteristics will also be evaluated in this phase. It needs to be considered if the information exists because of the incident, the response to the incident, or if the attacker staged it (6).

In our graph representation of the process, we have a set of object vertices with no edges or event vertices. We can see this in Fig. 2(*A*), where we have four objects from the crime scene, X, Y, Z, and W.

The properties, or metrics, for this phase include the error rates for the identification and individualization of an object. The time and storage space complexities of the object examination process are also important in this phase.

Role Classification

After each object has been examined and we know what information it has, we can begin to examine why it has the information. This phase starts the process of translating the state of an object

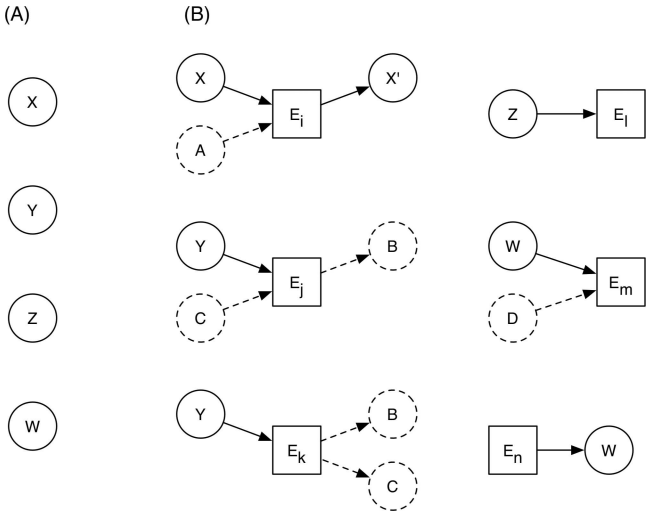

FIG. 2—*Graphical representation of (A) the Evidence Examination Phase with four objects that were recognized during the search phase and (B) the Role Classification Phase where the four objects are assigned to six events.*

to the possible events it was involved in. The goal of the role classification phase is to identify the characteristics that are related to the incident and are the effect of an event or the cause of an event. The result of this phase will be a list of roles and events that each object may have been involved in. This phase is where an imagination and open mind are needed because the investigator will have to identify possible events that occurred to and because of an object.

When identifying roles, we know that every object is the effect of at least one event. After all, it had to be created and get to its final location. To identify the effect roles that it may have played, we can examine each of the characteristics and question how that characteristic got there. By definition, the individual characteristics will be the most useful for reconstruction because they contain unique information. It is in this phase that we look for characteristics that provide temporal, relational, or functional information.

Each characteristic of an object may be the cause of one or more events. These are the objects that work as rules or laws of an event. For example, a loaded gun can be an effect of an event that loaded bullets into it (its functional state) and it can be a cause of an event that shoots a bullet.

When a potential role is identified, the investigator must identify characteristics of the other cause and effect objects in the event. These can be thought of as restrictions. Ultimately, the restrictions are not needed because when it comes time to test the event only those that meet the restrictions will pass, but the restrictions allow the investigator to reduce the number of tests that must be performed.

Chisum uses an interesting example of a clay ball with one area that is flat and the goal is to reconstruct the event that led to that state (5). In our process, we examine the clay object and identify its material as a class characteristic and any patterns and markings as individual characteristics. One explanation for the flat area may be that the object was created with it and therefore it is the effect of an event that created the object from a clump of clay. Alternatively, the object could have started as a ball and the flat area was because it was thrown at another object, such as a wall. In that scenario, the ball, the wall (or similar object), the earth's gravity, and the thing that forced the ball to fly are causes of the event and the flattened ball and wall are effects of the event. The flat area could be because the round part was cut off, in which case a cutting object such as a blade or laser would be a cause.

In all of these possible scenarios, there may be individual characteristics on the object that will help determine which event occurred. For example, any patterns that were on the wall may also exist on the object if it was thrown and any patterns from a blade may exist if it was cut. Each of these scenarios will be evaluated and tested in the next phase.

In our graph representation, we are adding edges and event vertices to the graph. For each role that an object played, a new object and event vertex will be added to the graph. If the object is a cause of the event, then there is an edge from the object to the event. If the object is an effect of the event, then there is an edge from the event to the object. Each event can be labeled with its type. If the event requires certain cause or effect objects to exist, then the objects should be placed in the graph. We represent them with dotted lines. Figure 2(*B*) shows examples of the graph at this stage using the objects from Fig. 2(*A*). In the first one, we see that *X* could be a cause and effect of an event, the second and third show that *Y* could be a cause of an event with two effects or an event with one effect, the fourth shows that *Z* could be a cause of an event with unknown effects, the fifth shows *W* could be a cause of an event with another object and an unknown effect, and the sixth shows *W* could be the effect of an event with an unknown cause.

The metrics for this phase include the error rates for identifying the characteristic of an object as being involved in the incident, the error rates associated with creating roles that did not occur (false positives), and the error rates associated with not creating roles that did occur (false negatives). The time and space complexities for the analysis of each characteristic are also a factor.

Event Construction and Testing

After we have classified the evidence with respect to its event roles, we can construct events using other objects and test if they are possible. This phase will begin with a collection of objects and their roles and characteristics and end with a collection of unordered or partially ordered events that may have occurred. During this process, we may end up searching the crime scene for additional objects. If additional objects are collected, they will undergo the procedures for the Evidence Examination and Role Classification Phases.

Every event must have at least one cause and one effect object and this phase tries to find matching pairs. The cause and effect correlation techniques are unique to each event, but the changes to the characteristics of an effect object must have occurred because of a characteristic from a cause object. Therefore, if we know which characteristics were changed by an event, we can do a backwards search for a cause object. Similarly, we can search for effect objects using the characteristics of a cause object. During the role classification, some roles were given restrictions and they will be used to create events.

In some cases, the characteristic that a cause object used in an event may not exist when the investigation occurs, in which case a direct link with an effect cannot be determined. For example, consider a power drill that was used to make a hole so that a perpetrator could gain access to a building. Later, the perpetrator changed the bit to a larger size, put the original bit in his pocket, and left the drill behind. The drill was a cause in the event to create the hole, but it does not have the characteristic needed to show it. Similarly, a search for an object may identify several possible objects because the search used class characteristics

instead of individual. Each possible event must be tested and a confidence value can be assigned to each with respect to how much information exists at the crime scene to show that it occurred.

A general procedure for event construction is:

1. Select an object that was the effect of an event that is the most important to the current needs of the investigation and that has the most unique characteristic change.
2. Conduct a backward search of all cause objects to find an object with one or more characteristics that could have changed the characteristic on the effect object and that can satisfy the restrictions that were placed on the role. If needed and possible, search the crime scene for additional objects.
3. For each possible cause object that is found, examine it to identify other cause roles that would have needed to exist for the event to occur and identify if the object would have had to be in a special state. Add the additional cause roles to the event requirements. The cause objects are searched for any that meet the role requirements and those that are found are added to the event.
4. If one or more cause objects were found during the backward search, then conduct a forward search of the effect objects for all objects with one or more characteristics that could have been an effect of the same event. This may include a new search of the crime scene.
5. If additional effect objects were found, then return to step 2 to perform another round of backward searching.
6. If no additional cause and effect objects can be found for the event, then the event should be tested. If there are missing roles in the event, then hypotheses should be created about what they are and why they are missing.
7. If the test passed, then it should be added as a possible event of the incident with a confidence value that corresponds to both the amount of evidence that exists to support the event and the amount of evidence that does not exist and for which hypotheses had to be created. If the test failed, then we do not use the event.
8. After we have tested the previous event, we start the search process again to find other objects that could have created the same effect. Therefore, we forget about all objects except the original effect object and return to step 2. We choose different objects so that we do not recreate an event that has already been tested.
9. After we have tried and tested the possible events for that effect object, we repeat the process by finding another effect object in step 1.

We present three techniques that can be used to correlate the cause and effect objects:

- The changes to the effect object are related to those of the cause object. For example, dents from a physical impact will have an inverse shape similar to the cause object. In the digital world, data that is written to a file will also exist in the process that wrote it.
- The location of the effect object is relative to the location of the cause object and the rules and laws of the event and the environment. In the physical world, a common example is using blood spatter as an effect of a gunshot. We can use its location to identify the location of the cause. The locations may contain additional evidence.

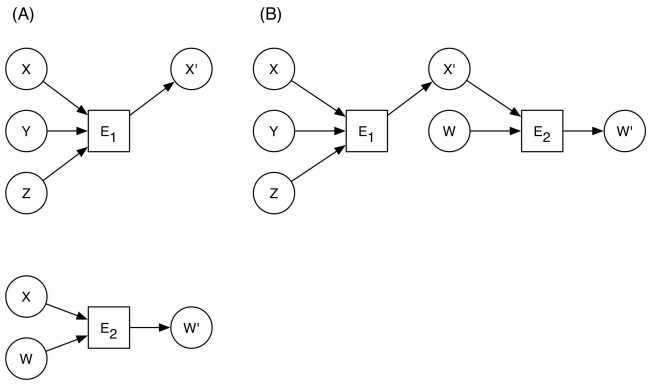

FIG. 3—*Graphical representation of (A) the Event Construction and Testing Phase with the six roles from Fig. 2 reduced to two events and (B) the Event Sequencing Phase where the two events are placed in order.*

- If an event is in progress during the evidence collection, then the time of the cause of the event may be determined using the current state of the event and the rules and laws of the event. A common example of this in the physical world is using different measures of decay to identify when something died. In the digital world, this may involve using process information to identify when the suspect process was started.

If we return to the clay ball example from Chisum, then we can search the crime scene for cutting tools and can search the floor or walls for evidence that the ball was thrown against them. We can test the throwing or dropping theory by making a ball of the same size and material and dropping it from different heights. We can also try and cut the material and see the effects. Based on the location and shape of the ball we may be able to calculate where it could have been dropped or thrown from (5).

In our graphical representation, we are identifying which of the event vertices are for the same event and which objects can fill in the empty roles. When two event vertices are found to be the same, one is removed and the edges and object vertices are moved to the other event. Only events that have been tested should be added to the graph at this point. This can be seen in Fig. 3(*A*), where we see that the roles from Fig. 2(*B*) have been reduced to two events.

The metrics for this phase include the error rate associated with correlating event roles that are not actually from the same event (false positives), the error rate associated with not correlating event roles that are from the same event (false negatives), and the error rate for the hypotheses that were developed about the missing roles. The time and space complexities of the object role search procedure are also a factor in this phase.

Event Sequencing

After we construct the individual events, we can link them together to form event chains. In many cases, we will not be able to make one big event chain for the entire incident or even for all of the collected evidence. Instead, we will have many single events or small event chains that will need to be sequenced.

If an object contains reliable temporal information and the actual time value of an event is known, then it can be easily sequenced relative to the other events whose time is also known. In many cases though, the exact times of the events are not known, so we need to rely on other sequencing techniques.

Let there be two events, e_i and e_j, and let e_i occur before e_j. If an effect of e_i was also a cause of e_j and e_j uses the characteristics that were changed in e_i, then we may be able to determine the sequence of events using the final state of the objects and the rules and laws about e_i and e_j.

The information from this technique is sometimes called relational because the information added to the object from the second event can be examined in relation to the information from the first event. A good physical world example of this is a broken piece of glass with blood on it (5). If the blood is facing down, then the event that caused the blood likely occurred before the glass was broken (or the glass was flipped over). If the glass is in the middle of a pool of blood and there is no blood under the glass, the glass was likely broken first.

In some cases, this ordering technique uses functional information from an object because the object must have been in a given state for an event to occur. For example, an event to load a gun must occur before an event to shoot it. For a digital system, a network application needs to download an image from the network before it can save it to the local system.

A variation of this sequencing technique is used when phases of a class of incident are known. For example, in an assault incident in a house, the attacker must have entered the house, made contact with the victim, and exited the house. We can use these general phases of the incident and the class of events that could occur in them to sequence the events for which we have evidence (4,6).

After the events have been sequenced, there will be one or more event chains. To account for the event gaps, hypotheses should be created for the events that occurred in between the known event chains.

After the events have been determined, the boundaries of the crime scene can be determined using the location characteristics of the objects in the events. If it is found that additional events may have occurred outside of the original crime scene that was searched, a new search may occur for additional evidence.

In our graphical representation, we are identifying object vertices that are for the same state of the object. If two object vertices are found to represent the same state of an object, one is removed and the edges that it had are moved to the other object. This can be seen in Fig. 3(B), where we were able to sequence the two events because event E_2 could only occur after object X was an effect of event E_1.

The metrics for this phase include the error rate for incorrectly sequencing two events, the error rate associated with not sequencing two events when enough information existed to sequence them, and the error rate associated with the hypotheses that are generated for missing events. The time and space complexities for the sequencing process are a factor in this phase.

Hypothesis Testing

After the events have been sequenced, the hypotheses about the incident can be tested. At this point in the reconstruction process, we will have event chains and one or more hypotheses about the events for which no evidence was found. The final theory, if one is found, must be supported by the evidence and must account for the events where there are missing cause or effect objects. Any confidence values that were assigned during the event construction should be taken into account when evaluating hypotheses.

At this point in the investigation, we are down to the final theory and therefore there are fewer metrics. For this phase, the error rate associated with the final theory can be used as well as the time and space complexities for the process to test a theory and show that it is supported by evidence.

Comparison to Existing Reconstruction Phases

This model is most similar to the procedure that was described by Bevel and Gardner (6). The main concepts in their model can be found in our model, but occur in different phases. Our approach is a more formal and systematic process of developing the roles and requirements of each event. The existing models are more conceptual and designed to give the investigator insight into the process, but not formalize it, as our goal was. This model can be more easily implemented as a reconstruction tool.

Digital Crime Scene Model

We will now examine the five event reconstruction phases with respect to a digital crime scene.

Evidence Examination

The evidence examination phase examines each piece of digital evidence to identify it and individualize it. In the process, the class and individual characteristics will be determined. Examples of the class characteristics of digital data include any general data format values, such as the header signatures ("magic values") and file extensions. Individual characteristics are those that may be unique to that file and will include the actual content of the file outside of the standard format data. Individual characteristics are rare in digital data. This phase also includes network packets and logs from network devices, not only data from a hard disk. The details associated with data characteristics need more research to identify those that are the most useful and provide the most information.

The reliability and credibility of the digital evidence is also examined in this phase. An example of data that could be examined is the times associated with a file and identifying if they can be trusted, if they were updated while responding to the system, or if the attacker modified them. If deleted files were recovered, then the recovery tool should be considered to determine if the recovered file is accurate. If data was taken from a live system, then the procedure relied on software that could have been modified by the attacker and therefore the data should be examined in more detail to find evidence of tampering.

Role Classification

The role classification phase examines each of the objects and identifies what types of information it has. For example, an investigator can use Casey's functional, relational, and temporal analysis techniques (6) to identify the information types an object has. Using the object's information, hypotheses are created about what events the object was a cause of and what events it was an effect of. Every object in the digital crime scene is the effect of an event. For example, a process is the effect of the kernel creating it and data on a disk is an effect of it being written there by the kernel, which was likely an effect of a process causing a system call event.

In a typical computer, there are at least two objects that are causes in every event: the hardware and the operating system. The hardware has an influence on the effects of every event and the operating system dictates what events will occur. If an attacker has modified the operating system, then the effects of some events will be different than if the attacker did not modify the operating system.

Depending on the level of detail that is needed for the investigation, many of the events dealing with data storage, devices, or processes will be caused and initiated by the kernel and the kernel will be an effect of a system call or similar request from a process to initiate the event. It is unlikely that evidence will exist to reconstruct events at this level, especially because it will require that the memory of a system be acquired before the system is powered off.

In a computer, we can reduce all events to reading and writing events. This is similar to being able to reduce the five senses in the physical world to touch at a molecular level. It will not be possible to find evidence of all events at this level though. In general, if data is read from an object for an event, then it is a cause. If data is written to an object from an event, then it is an effect.

We will now give some examples of how roles can be determined. If the object is an executable file, then analysis of its system calls can show what events it could have caused. For example, it may open files or network sockets. A more detailed analysis of the executable file may show that it only opens network sockets on a certain port or that it only opens files for reading in a given directory. Files that contain time stamped entries may show that the file was an effect of an event at that time. The Modified, Accessed, and Changed (MAC) times of a file may also show when the object played a role in an event. The modified and changed times show that it was the effect of an event and the accessed time can show that it was either role. Note that the occurrence of an event can be determined even if the attacker forged the actual date or time.

As an example, consider the notepad.exe application. A process that is loaded by this executable can be the cause of an event to write an ASCII file. It can also be the cause and effect of an event to read an ASCII file into memory. Now consider an ASCII text file that contains sensitive data. It can be the effect of an event where a process wrote data to it and it can also be the cause of an event where a process read from it.

Event Construction and Testing

The event construction and testing phase takes the role assignments and correlates the cause and effect objects. This phase can be difficult with digital computers because the process and kernel objects are not always collected from the crime scene and they initiate most events. Furthermore, the process and kernel information will be erased when the system is powered off. In many cases, hypotheses will need to be created about the processes that played a role in events. Executable files on the system can be examined to determine the roles a process may play if it were loaded from the executable.

One of the benefits of most digital investigations is that the investigator always has a copy of the crime scene and can easily search it for new evidence. Therefore, in many cases the search for other objects in an event can be performed on both the evidence that has already been collected and on the digital crime scene. When new evidence is found in this phase, it must be examined and have its roles classified so that it can be fully utilized in the reconstruction process.

When doing a backward search to find cause objects of events, we can look for objects that could have created the data. Consider data that is found on the disk. Either a process or the kernel initiated the event that wrote it there (we will ignore the possibility that the hard disk initiated it). Using the individual or class characteristics of the data, we can find values that are unique to it and the possible effect objects can be searched. For example, consider a JPEG image. It has a format that not every application can process; so only a limited number of applications would be able to successfully initiate a read event for a JPEG file and reasonably process it. Similarly, only a limited number of applications can write a JPEG file format, so we can search for applications that could have initiated a write event.

We can also focus a backward search using access control permissions. Not all users or applications will have permissions to write to or read from a file. Once a possible cause object has been identified, its dependencies need to be identified. For example, an application may have one or more configuration files that are needed for the event to occur and they may contain additional evidence.

A forward search identifies the unique data that was being used in the event and searches for data that may have been written to because of the event. The permissions associated with data can also be used in this search direction to restrict the search to only objects that the cause had access to.

Testing the events can be challenging in a digital environment because it may require the investigator to execute code from the system. This is dangerous because the investigator may not know everything that the program will do. It is typically safer to test the theories in a trusted and safe environment, such as a virtual machine (12) where the system can be easily isolated and rebuilt.

Using our previous example of the ASCII file and notepad.exe, if we wanted to know how the ASCII file was created then we could search the system for all executables that can create an ASCII file. This would result in many applications, including notepad.exe. Tests could be conducted with all identified applications to identify any unique characteristics that may show which created the file.

Event Sequencing

The event sequencing phase orders the events based on when they occurred. Some events will generate a timestamp on a file or in a log file, but another event may change the time. If the execution flow of an executable or process is known, then that information can be used to sequence the events that it caused.

Using low-level file system analysis techniques may also help to show the sequence of events. The location of the data structures and storage locations that a file system allocates to a file may reveal information about other files that were created before it. For example, the order of the file name structures in a directory or the order of clusters in files may show when two files were created relative to each other. Using the shell history file from a Unix system is a common method of sequencing the events on the system. Unfortunately, many attackers will delete or modify the history file contents.

Theoretically, event sequencing can be easier in the digital world versus the physical world because computers are deterministic and events are initiated by code. Therefore, if the programs and operating system can be reverse engineered, then we may be able to better determine what events need to occur first. Many investigators do not have full access to the code of applications and operating systems though, and therefore the investigator is left to testing applications and observing the events that occur.

Hypothesis Testing

Hypothesis testing for digital crime scenes is no different than for physical crime scenes. At this point in the investigation, we will have a series of event chains and hypotheses about missing events. Each hypothesis should have a confidence level attached to it and this phase examines each hypothesis to determine which was most likely and which the evidence can refute. Knowledge of how digital systems work is important to this phase, but there are no procedures

that are unique to digital evidence and not physical evidence. It is in this phase where Stephenson's Petri net hypotheses testing model could be used (8).

Discussion

As we have shown, the reconstruction process for a digital environment is similar to a physical environment. There are some differences that work the digital environment's advantage and disadvantage. We will discuss those in this section.

A difficulty of the digital world is the lack of randomness and therefore a small number of individual characteristics. Computers have been designed to execute a series of instructions and therefore there is little difference between two computers or between two files that were made by similar programs. This is a problem when an investigator is trying to identify the source of data.

For example, consider a contraband JPEG picture that is found on a suspect's hard disk. We consider that it is the effect of a write event and we search for applications that could have initiated the event. Our search results on a standard Microsoft Windows system may find Microsoft Internet Explorer, File Explorer (drag and dropping), Paint, Microsoft Outlook, WinZip, and the move.exe and copy.exe command line tools. Each of these will need to be investigated and most systems will have additional applications that can save JPEG files. We may be able to narrow the scope based on records that the applications keep, such as recently opened or saved files. In any case, this is a time-intensive process, but may become necessary to show that a contraband file was intentionally saved.

The kernel and processes initiate many of the events in a computer, but their state is lost when the system is turned off. Therefore, the evidence of the events in which they were causes or effects are gone if the computer is found in a powered off state or is rebooted after the incident. This is similar to a criminal being able to clone himself and sending the clones to commit crimes. After the crime, the clone vanishes and the criminal has no evidence of the crime on him. In the digital world, this forces the investigation to rely on the contents of the executable file that was used to load the process. In the future, more investigations will likely have the memory contents of the system so that some process and kernel evidence can be collected.

There are also few tools that can provide an investigator with quick information about the capabilities of an executable file. Automated executable analysis tools are needed to allow a law enforcement lab to identify the applications that can initiate events. It is also difficult to know what state a closed source application must be in for it to perform an event. Knowing the state of a system may be required to show the reliability of the evidence (13).

On the positive side, digital evidence requires us to use tools to process it. Therefore, it lends itself well to databases of data that can be queried and processed (14, 15). This allows the search for data to occur more efficiently.

Conclusion

In this paper, we have examined the procedures for conducting a physical crime scene event reconstruction and applied them to a new abstract model. That model is then applied to a digital crime scene. Much of the current focus in digital investigations is on recognizing and analyzing pieces of digital evidence, but work also needs to focus on determining why the evidence exists. With this model, we can develop different techniques for each of the phases. The techniques can be compared and examined to identify those that are most accurate. Future research may show that it is technically infeasible to examine a system at this level.

Event reconstruction will become important because investigators must be able to defend their hypotheses about why evidence exists. As more defendants claim that evidence was planted on their systems (16), the investigators will need to identify if the user downloaded a file or if it was planted there by someone else. This model will help to perform the reconstruction task and to develop tools to automate this process. Our model should also help investigators when they encounter an incident that is not supported by well-established tools and procedures because they can use it as a framework for their analysis.

References

1. Carrier B, Spafford EH. Getting physical with the digital investigation process. International Journal of Digital Evidence 2003;2(2).
2. James S, Nordby J, editors. Forensic science: an introduction to scientific and investigative techniques. Boca Raton, FL: CRC Press, 2003.
3. Lee H, Palmbach T, Miller M. Henry Lee's crime scene handbook. London, UK: Academic Press, 2001.
4. Rynearson J. Evidence and crime scene reconstruction. 6th ed. Redding, CA: National Crime Investigation and Training, 2002.
5. Turvey B. Criminal profiling: an introduction to behavioral evidence analysis. 2nd ed. London, UK: Academic Press, 2002.
6. Bevel T, Gardner RM. Bloodstain pattern analysis: with an introduction to crime scene reconstruction. 2nd ed. Boca Raton, FL: CRC Press, 2002.
7. Casey E. Digital evidence and computer crime: forensic science, computers and the internet. 2nd ed. London, UK: Academic Press, 2004.
8. Stephenson P. Modeling of post-incident root cause analysis. International Journal of Digital Evidence 2003;2(2).
9. Sowa JF. Processes and causality, 1999. Available at: http://www.jfsowa.com/ontology/causal.htm.
10. Allen J, Kautz H, Pelavin R, Tenenberg J. Reasoning about plans. San Mateo, CA: Morgan Kaufmann Publishers, 1991.
11. Saferstein R. Criminalistics: An introduction to forensic science. 8th ed. Upper Saddle River, NJ: Pearson, 2003.
12. VMWare GSX Server [computer program]. Palo Alto, CA: VMWare, Inc, 2003.
13. Kenneally E. Gatekeeping out of the box: open source software as a mechanism to assess reliability for digital evidence. Virginia Journal of Law and Technology 2001;6(3).
14. Chen K, Clark A, De Vel O, Mohay G. ECF-event correlation for forensics. In: Proceedings of the 1st Australian Computer, Network & Information Forensics Conference; 2003 Nov 25.
15. Stallard T, Levitt K. Automated analysis for digital forensic science: semantic integrity checking. In: Proceedings of the Annual Computer Security Applications Conference; 2003 Dec 8–12.
16. Mark Rasch. The giant wooden horse did it!. Security Focus 2004 Jan 19. http://www.securityfocus.com/columnists/208.

Additional information and reprint requests:
Brian Carrier, M.S.
CERIAS
Recitation Building
656 Oval Drive
W. Lafayette, IN 47907-2086
E-mail: carrier@cerias.purdue.edu