Reconstruction of Events in Digital Forensics

M.A.Rama Prasad1 Y.Naga Satish2
1,2Department of Computer Science and Engineering,
1Vignan’s Lara Institute of Technology and Science, Vadlamudi
2BVC Institute of Technology and Science, Amalapuram

ABSTRACT
Detecting and gaining clues from the crime scene plays a major role in the process of investigation. Nowadays with increase of cybercrime and fraud on the internet, digital forensics gaining importance. It needs to collect information the events happened at crime place. This paper deals with event reconstruction process which is useful for digital crime scene. The process of reconstruction is based on crime event and event characteristics. It also furnishes crime rate guess using some semi-formal techniques.

Keywords: digital investigation, digital forensics, events, attack trees, reconstruction, evidence.

1. INTRODUCTION
Digital forensics is gaining importance with the increase of cybercrime and fraud on the Internet. Tools and methodologies for digital forensics with the soundness necessary for presentation in court are in high demand.

Computer forensics can be summarized as the process of identifying, collecting, preserving, analyzing and presenting the computer-related evidence in a manner that is legally acceptable by court (Mc Kemmish, 1999; Noblett et al., 2000; Robbins, 2000; Borck, 2001; Garber, 2001; Patzakis, 2003; Yasinsac, 2003; Slade, 2004; Bitpipe, 2005). [2]

The goal of this section is to introduce the terms Computer Crime, Digital Evidence and Digital Forensic Analysis and show their basic concepts. There are two aspects of Computer Crime, which are defined by the world’s leading computer forensic equipment company DIBS [DIB] as follows:

- A criminal act in which a computer is essential to the perpetration of the crime.
- A criminal act where a computer, non-essential to perpetration of the crime, acts as a store of information, concerning the crime.

This is to say that not only the crimes committed directly with a computer belong to this field of offense but also crimes where evidence could be found on computers or networks without necessarily using those devices to actually commit the crime.

CASE EXAMPLE
If the police has a certain suspect in a homicide, the investigation of his computer may reveal certain details about his contact to the victim (e.g. emails and chats), or even research about how to get rid of a corpse. The analysis of the suspects cell phone may refute his alibi or also give details about contact between suspect and victim.

With this definition it becomes obvious which important role digital crime analysis has nowadays and that its role in crime investigation will not decrease but more likely increase in the upcoming years. There is nearly no imaginable crime in which no computer or network can be involved. Therefore, besides the increasing number of electronic fraud or crimes committed directly with a computer, the branch of digital investigation will become more important for classical evidence collection and crime investigation.

The term Digital Evidence describes all the information that can be gained from electronic devices. This can be storage media like hard disks, network logs, cell phone logs, emails and so on.

Digital Forensic Analysis is divided into two main branches. The first one is Physical Storage Media Analysis and the second Network Analysis.[1]

One of the standard works on Digital Forensic Analysis is Brian Carrier’s File System Forensic Analysis [15].

Figure 1.1 shows the global price tag of cybercrime.
Figure 1.1. Norton cybercrime 2012 report.

Figure 1.2 shows the three major phases of digital crime investigation. These are the System Preservation-, Evidence Searching- and Event Reconstruction Phase. As Figure 1.2 indicates these three phases do not need to occur one after another but there are trackbacks from every phase to the previous.

- **System Preservation**
  - This phase is always the first thing to do once a digital crime is detected or even assumed. As with classic crimes the first act of investigation is to preserve the crime scene. This is the main aspect of this phase. In classic crime investigation you can close of the crime scene, e.g. a house or at, but it is more difficult to follow this approach in digital investigation. Here it is difficult to shut down a network or computers without altering data. As with classic crime scenes it should be tried to avoid every change of the evidences. It has to be tried to copy and save all information contained in the network or on physical storage devices without changing them. It is important to have a proof that the data was not changed during the investigation process. One approach to achieve this is to compute a cryptographic hash sum of the data, which would indicate a change of them later.

- **Evidence Searching**
  - Now, after the crime scene is preserved, the next step is to look for evidences. As a digital crime is assumed the digital investigator creates hypotheses which can either be supported or refuted by evidence found in the data. It is a major aspect not only to look for evidence that supports a specific hypothesis because the hypothesis always could be wrong. The main methods for searching for evidence are:
    - to look at log files, e.g., those of routers or other network components,
    - search for altered data, e.g., again with cryptographic hash sums,
    - looking for root kits, e.g., by checking the low levels of the operating system,
    - search the file system for ominous files.

- **Event Reconstruction**
  - The third and last phase of the digital investigation process is to use the collected evidences to reconstruct what has happened in the system or network. To do this, it is necessary to correlate various evidence, maybe even from different sources, to get a proof of the one hypothesis that stands last. For this phase it is important to have a knowledge of the operating systems and the network basics of the digital components involved in the crime. To understand how an operating system or the network components work is essential to come to a clue what the hints are indicating.
2. EVENT RECONSTRUCTION

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.

2.1. General concepts

2.1.1. Definitions

Digital data are data represented in a numerical form. A digital event is an occurrence that changes the state of one or more digital objects [5]. If the state of an object changes as a result of an event, then it is an effect of the event. A digital object is a discrete collection of digital data, such as a file, a hard disk sector, a network packet, a memory page, or a process. The state of an object is the value of its characteristics. If a letter were changed in an ASCII text document, then the object corresponding to the file would have a new state. Note that because digital objects are stored in a physical form, then their state can be changed by both physical and digital events. An object is evidence of an event if the event changed the object’s state.

Now we 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.
- **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 if an object is an effect but not a cause, then it must have been created as a result of the event. The cause objects can be thought of as the tools and scientific laws that determine how an event will occur. We can graphically represent this definition of an event as shown in Fig. 2.1.1(A). Each circle represents a state of an object and each box represents an event. In this graph, objects P, Q, and R were causes of the event E and object P is also an effect, with its new state noted by P'.

In some cases, it may be possible or may not possible to identify a cause object that initiated the event. The initiator of the event is the object or event that began the event.

![Graphical representation of (A) an event with three cause objects and one effect object and (B) an event chain with two events.](image)

**Figure. 2.1.1** Graphical representation of (A) an event with three cause objects and one effect object and (B) an event chain with two events.

An event chain is a sequence of events \( e_0, e_1, \ldots, e_k \) 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. 2.1.1(B), where event E1 causes event E2 because of its effect on object P.

3. SEMI-FORMAL TECHNIQUES FOR EVENT RECONSTRUCTION PROCESS

3.1. Attack trees

An important part of reconstruction process is identification of possible incident Scenarios. Attack trees described in [7] is a semi-formal approach to discovery, documentation, and analysis of possible incident scenarios. An attack tree is a diagram that describes different scenarios achieving some goal. The goal is represented by the root node. Other nodes represent sub goals that must be achieved - either alone, or in conjunction with other sub goals - to achieve the goal.
Figure 3.1 shows an attack tree of opening a safe without authorization (this example is taken from [8]). The goal of the tree is to open a safe. The goal can be achieved in a number of different ways, such as picking the lock.

![Attack Tree Diagram]

Figure 3.1. Attack trees

Attack tree describing different ways to open a safe cutting through the safe wall, learning the combo, or making a security hole when the safe is installed. Different ways of learning a combo are elaborated under the "learn combo" node. Basic attack tree is built from two types of node: AND nodes, and OR nodes. In this paper, AND node is graphically distinguished from an OR node by an arc that joins arrows coming out of the AND node. Both AND and OR nodes describe sub-goals that need to be fulfilled to fulfill the main goal of attack. To fulfill an OR node, any one of its child nodes must be fulfilled. To fulfill an AND node, all of its child nodes must be fulfilled. A possible scenario corresponds to every group of leaf nodes, whose joint fulfillment results in the fulfillment of the tree's root node.

Once the attack tree is built, the analyst can calculate various properties of the discovered scenarios. The simplest property to calculate is whether a scenario is possible. Finding possible scenarios is a three-step process. In the first step, the analyst assigns a value “possible” or “impossible” to each leaf node. To decide whether a particular leaf node is possible or impossible, the analyst uses available evidence. In the second step, the analyst decides for every other node whether it is possible or not using the following rules:

- An OR node is possible if at least one of its child nodes is possible;
- An AND node is possible if all of its child nodes are possible.

In the third step, all possible scenarios are identified by tracing them from root to leaves along chains of “possible” nodes. Attack trees permit other types of scenario comparison. In particular, they can be used to calculate probability and cost of different scenarios.

3.2. Visual investigative analysis

Another semi-formal technique used for event reconstruction is Visual Investigative Analysis (VIA). It is a charting technique that uses a network approach to display graphically the sequences of occurrences and the relationships of all the elements of a criminal incident [9].

ISSN: 2231-5381  http://www.ijettjournal.org  Page 3463
3.2.1 Graphical notation
VIA chart is a directed acyclic graph whose arrows represent activities and whose nodes represent states of the world in which activities start and finish. The following graphical notation is used to draw VIA charts.

Solid line arrows:
A solid line arrow represents single activity. The description of the activity is placed above the arrow. Activity always start at one node and ends at another. To improve clarity of the chart, only one solid line arrow is permitted between any two nodes.

Circles and triangles:
Circles and triangles are nodes. A node represents the world state in which one or more activities start or finish. Terminal is the last node in a sequence of activities. There is usually only one terminal on a chart. Terminals are drawn as triangles. All other nodes are drawn as circles.

Dotted lines and arrows:
A dotted line or arrow denotes a dummy activity. Dummy activity consumes no time and does nothing. It simply says that two nodes denote the same world state. Dotted lines are used for drawing parallel activities whose starting and ending world states are the same.

3.3 Multilinear events sequencing
Multilinear event sequencing (MES) is a set of semi-formal techniques for conducting investigations [10]. It is based on the same ideas as visual investigative analysis and attack trees.

3.4 Why-because analysis
Why-Because Analysis (WBA) is a method for determining causes of complex accidents. Like MES, it uses counterfactual reasoning to establish causality, but the approach taken by WBA is more formal.

3.4.1 Why-because graph
The graphical notation used by WBA to represent accidents is Why-because graph (WB-graph). It is a directed acyclic graph whose nodes represent elements of the accident, and whose arrows represent causal relationship. Four kinds of nodes are defined for building WB-graphs. A node can be either of the following
- event --change from one state to another
- process --undifferentiated mix of events and states
- non-event --causally important absence of some event

4. EVENT RECONSTRUCTION PROCESS

![Event Reconstruction Process Diagram]

Figure 4.1. Event reconstruction phases
We will now examine the five event reconstruction phases with respect to a digital crime scene.

4.1. 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.

4.2. 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 [11] 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 the 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.

4.3. 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.[4]

4.4. 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.

4.5. 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 [13].

5. CONCLUSION

In this paper, we have shown the event reconstruction process in digital crime scenes. It also gives introduction to the field of digital forensics and terminology used. In the event reconstruction process, there are 5 phases, which plays vital role in the investigation is discussed in this paper. In addition to that it also covers the some of the semi-formal approaches used in the reconstruction process to construct events.

REFERENCES


BIBLIOGRAPHY OF AUTHORS

M.A.Rama Prasad received B.Tech degree in Computer Science from J.N.T. University and M.Tech from J.N.T. University, Hyderabad. Currently he is working as an Assistant Professor of Computer Science and Engineering in Vignan’s Lara Institute of Technology and Science. He is involved in research of Computer Forensics and Mobile Computing.

Y. N. Satish received B.Tech degree in Computer Science from J. N. T. University and M.E from Satya Bhama University. Currently he is working as an Assistant Professor of Computer Science and Engineering in BVC Institute of Technology and Science. He is involved in research of Computer forensics, and Mobile Computing.