Adapting Traceability in Digital Forensic Investigation Process

Siti Rahayu Selamat, Robiah Yusof, Shahrin Sahib, Irdra Roslan, Mohd Faizal Abdollah, Zaki Mas’ud
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia
sitirahayu@utem.edu.my, robiah@utem.edu.my, shahrinsahib@utem.edu.my, irda@utem.edu.my, faizalabdollah@utem.edu.my, zakimasud@utem.edu.my

Abstract

Generally, the goals of digital forensic investigation process in a cyber crime are to identify the origin of the incident reported as well as maintaining the chain of custody so that the legal process can take its option. However, the traceability process has become a key or an important element of the digital investigation process, as it is capable to map the events of an incident from different sources in obtaining evidence of an incident to be used for other auxiliary investigation aspects. Hence, this paper introduces the adaptability of the traceability model to illustrate the relationship in the digital forensic investigation process by integrating the traceability features. The objective of this integration is to provide the capability of trace and map the evidence to the sources and shows the link between the evidence, the entities and the sources involved in the process. Additionally, the proposed model is expected to help the forensic investigator in obtaining accurate and complete evidence that can be further used in a court of law.

Keywords: digital forensic investigation, traceability model, evidence, source of evidence

1. INTRODUCTION

Cyber crimes or digital crime are now serious, widespread, aggressive, growing, and increasingly sophisticated, which poses major implications for national and economic security [1]. Many industries and institutions, public- and private-sector organizations (particularly those within the critical infrastructure) are at significant risk. This statement has been proved by the number of complaints received and processed by the Federal Bureau of Investigation (FBI) in collaboration with Internet Crime Complain Center (IC3). In 2006, 200,481 complaints were received and has been increased in 2010 [2] as illustrated in Fig. 1.

This was a 34.01% increase as compared to 2010 when 303,809 complaints were received. From all complaints received in 2006 and 2010, IC3 referred 86,729 complaints and 121,710 complaints respectively to federal, states and local law enforcement agencies all around the country for further considerations. Although IC3 primarily refers complaints with claims of dollar losses, the top ten referred complaints indicate that 2.8% of complaints are computer intrusion and being increased to 9.1% in 2010.

From the cyber crime or complaints reported, it indicates that the number of crimes involving computers and internet has grown over the last decades and it needs products that can assist law enforcement in using computer-based evidence to determine the who, what, where, when, and how for crimes. As a result, computer and network forensics has evolved to assure proper presentation of computer crime evidentiary data into court and the role of forensic become highly important to get digital evidence.

The purpose of a forensic investigation can be established by either identifying the offender of a case, or establishing an evidence to build a case against the offender [3]. As both situations are common in the law enforcement perspective, the ability to trace the source...
to an evidence or vice versa is essential [4]. Additionally, another limitation is the acceptability of evidence that differs in each of these situations. There was also an issue of origin identification and cross referencing in investigation process [5] [6]. Hence, the traceability information is important to avoid the mislaid of decision and valuable information in collecting and analyzing during the investigation process.

Due to this fact, the goal of this research is to adapt and integrate traceability in the digital forensic investigation process that represents the traceability information in the stage of conceptual and component composition. The purpose of this integration is to help the forensic investigator obtain accurate and complete evidence of the incident. In this paper, the proposed adaptation will be constructed based on the malware intrusion scenario.

The rest of the paper is organised as follows: The next section explains the related work on traceability, traceability models and digital forensic investigation process. Section III further describes the integration traceability in digital forensic investigation. The adaptation of traceability model is proposed in Section IV and a conclusion, together with future works is summarized in the last section.

2. RELATED WORK

2.1 Overview of Traceability

Traceability is the means to identify and follow real or imaginary objects through a process chain [7]. It gives the opportunity to back-track a chain of events, or to predict process outcomes given in the origin of an object. Traceability can be used in different areas. Even though traceability can also be defined in many ways, the meaning is to be able to trace and get information. ISO 8402:1995 defines traceability as the ability to trace the history, application or location of an entity, by means of recorded identifications.

Traceability is a tool to achieve different objectives and can never be completed. Therefore, [8] described the definition of traceability can be broad, because in most of the time the processes are very complex. On the other hand, [9] defined traceability as the ability to map events in cyberspace, particularly on the Internet, back to real-world instigators, often with a view to holding them accountable for their actions. In contrast, [10] define traceability in the networks perspective as how difficult it is to establish the source and destination of communications on computers and communication networks, such as the Internet. Therefore, based on the definition reviewed in this research, this paper summarized the definition of traceability as the ability to trace and map the events of an incident from difference sources in order to obtain evidence of an incident for further process of investigation.

In order to trace the requirement, the traceability approach is needed. Hence, [11] suggested that tracing the requirement can be performed in several ways based on the direction of tracing activities that are forward traceability and backward traceability as depicted in Fig.2.

![Fig. 2 Basic Traceability Approach](image)

Forward traceability approach defined in [12] is the ability to trace a requirement to components such as a design or implementation whereas backward traceability is the ability to trace requirement to its sources such as a person, institution, and argument. This basic concept of these traceability approaches can be represented as in Fig. 3.

![Fig. 3 Concept of Backward and Forward Traceability](image)
necessary to trace a requirement to the artifacts that implement it as well as tracing from an artifact to the requirement that the artifact itself implements. This circumstances create an idea on tracing in both ways, forward and backward or called as bidirectional approach as discussed in [16]. Hence in order to provide an accurate and complete evidence to prosecute the offender, this research will use the traceability approach discussed in [16]. To demonstrate the approach, knowledge of organizing the procedures, techniques and tools are needed. Identified as a traceability model, this knowledge is discussed in the next sub-section.

2.2 Traceability Model

A traceability model is a central component of a traceability environment around where the tracing procedures, techniques or methods, and tools are organized. A traceability model not only defines what entities and traces are, and which traces should be captured, but also represents traceability information in the stage of conceptual design, component composition, deployment and runtime [17] [18]. However, based on [18] [17] [11], the traceability model is used to represent the traceability information which demonstrate the relationship between the traces, entities and sources involved in a process or system.

In a traceability model, the conceptual explanation is covered by three features, namely the definition, the production and the extraction of traces [11]. The definition feature is concerned with the specification of the traces and traceable objects. It is within this feature that traceability model should define its traces, attributes and represented method. The definition of traces and traceable objects should promote a uniform understanding in order to avoid any errors caused by different interpretation during the tracing activities.

The production feature is concerned with the capture of traces that is usually by the means of an explicit registration of the objects and their relationships. The trace production deals with the actual occurrence of traces that roughly corresponds to the pragmatics of a traceability model in order to get a constructive traceability model.

The extraction feature of the traceability model is concerned with the actual process of tracing such as the retrieval of registered traces. A traceability model should provide diverse and flexible ways to retrieve (extract) the information registered in it as discussed in [19].

Consequently, a traceability model should provide a representation for traces and trace attributes as discussed in [17] in which the trace model provide two significant guidelines; relationship guideline and tracing guideline. The former guideline describes the relationship guidelines that explain what traces should be established and the later guideline describes how traces determined by the relationship guidelines should be documented. Both guidelines establish the structures containing the elements and the relations used in tracing, specifying their type as well as the constraints under which elements of the model can be related.

Hence, this research will employ all three features of the traceability model by adapting and integrating the features into the forensic investigation process. The goal of this adaptation and integration is to acquire accurate and complete evidence traces to help the forensic investigator on investigation process especially on collecting the evidence and the evidence sources of an incident.

2.3 Digital Forensic Investigation Process

In the digital forensics investigation practices, there are over hundreds of digital forensics investigation procedures developed all over the world. Each organization tends to develop its own procedures and some focused on the technology aspects such as data acquisition or data analysis [20]. Most of these procedures were developed for tackling different technology used in the inspected device. As a result, when underlying technology of the target device changes, new procedures have to be developed.

A research done in [21] introduced a mapping process which occurs inside digital forensic investigation process model. The mapping is formulated by grouping and merging the same activities or processes in five phases that provide the same output into an appropriate phase. From the analysis, most of the models consist of the critical phases which are Phase 2 – Collection and Preservation, Phase 3 – Examination and Analysis, and Phase 4 – Presentation and Reporting except Phase 1 and Phase 5. Even though, Phase 1 and Phase 5 are not included in some of the model reviewed, the study done by [22] [23] [24] [25] [26] [27] [28] [29] [30] indicate that both phases are important to ensure the completeness of the investigation. Phases 1 is to ensure the investigation process can start and run in the proper procedure, and protect the chain of evidence. While by eliminating Phase 5, it will lead to the possibility of the incomplete investigation and no improvement in investigation procedures or policies. Therefore, a good model should consist of all important phases; Preparation Phase, Collection and Preservation Phase, Examination and Analysis Phase, Presentation and Reporting, and Disseminating the case.

[21] findings also show that the existing models mentioned in each of the proposed models build on the
experience of the previous; and some of the models have similar approaches and some of the models focus on different areas of the investigation. However, all of the models in the output mapping have the same output even though the activity is slightly different on the term used and the order of the steps. On the other hand, all of these frameworks identified in the output mapping show that each framework has their own strength; however until nowadays there is no single framework that can be used as a general guideline for investigating all incident cases.

Therefore, in order to obtain the evidence and for it to be accepted in the court of law, digital forensic investigation must be successfully performed without tampering the evidence. Additionally, the chain of evidence should be presented to prove the evidence is legitimate. Hence, the evidence traceability identification of the origin of the crime scene or the location of the incident or crime originated is one of the important elements during the digital forensic investigation process and become the first challenge in the investigation as mentioned in [6] [31] [32].

3. TRACEABILITY IN DIGITAL FORENSIC INVESTIGATION PROCESS

In digital forensic investigation process, tracing is described as a process of finding or discovering the origin or cause of certain scenario. The tracing activities are able to discover the traces left in digital devices. In the computer crime perspective, trace can be found in any digital devices. These traces consist of activities such as login and logout of the system, visit of pages, accesses documents, create items and affiliation groups found in records of data. These traces data are analysed by identifying their relationship among the attributes such as port, action, protocol, source IP address and destination IP address where this consistent relationship will produce trace pattern of the incident or crime. This trace pattern can be further used on assisting the investigator during the investigation process.

3.1 Incident Trace Pattern

Incident trace pattern is essential in assisting the investigators tracing out the evidence found at crime scenes [33]. In this research, we affirm the definition of trace as any digital evidence in an incident. Meanwhile, tracing is defined as the observation of the moving trace on the various tracks. In addition, pattern is defined as a regular way in which certain scenario happened [34]. Therefore, in order to get a trace pattern, the observed movement of these trace is studied to confirm its regular way, with the help of the acquired hypothesis.

The incident trace pattern is confirmed using two steps. Firstly, the hypothesis which explains the initial scenario of the incident is taken. Secondly, the trace which was recorded in the source of evidence (host and network logs) is formulated. Using these two, the movement is observed which conclude that within the source of evidence, there are three courses of action that occurred. This course of action is referred as an event instead of process due to their focuses. A process merely focuses on progress or series of action toward a particular result; whereas, an event focus on the occurrence of something which not only concern with its action, but also with the attributes associated with it [35].

The extraction of the two steps above derives the three events of incident which are scan, exploit and impact/effect. Scan consists of the inspection activity which are not only to find vulnerability, but also to determine any available services (e.g. port number) on the target system (system being attacked) [36]. In this activity, if the port number responds to a scan, it will indicate the type of service running on the target system and reveal the exploitable services to attackers. Therefore, once the system determined which services are running on it, the vulnerability of the system could be exploited. Eventually, these exploited vulnerability can become a threat, such as unauthorized access (gain access) or unavailable service for intended users (deny service).

Exploit consists of the abuse activity traces that disclose any manipulation activity on the target system services such as attempting on downloading malicious codes to the target system and breaking the target system for opening backdoor on specific port. Meanwhile, the impact/effect event shows the traces on the goal of an attack which shows the goal of an attack as the consequences of the scan and exploits activities of the incident such as the target system is restarted, the services are terminated (expectedly) and new process is forced to be created.
In this research, the combination of the three events discussed previously form different trace patterns in order to identify the offender of the incident; victim, attacker and victim/attacker. As this pattern reflects the complainer or perspective, henceforth, it is named as Perspective Trace Pattern. From the analysis and findings in [33], victim/attacker and victim perspective trace pattern must consists of all three events, and attacker perspective trace pattern must consist of scan and exploit events of incident but it is optional in having the impact/effect event as depicted in Fig. 4. However, the difference between them is the content of the attributes belongs to each of the events such as the number of the destination port open, the type of operation, the protocol of the connection request, the services that are vulnerable and the item transferred during the communication exist.

The attributes of scan are communication exist, destination port open, operation type and connection request. Conversely, the attributes of exploit event are similar to scan event, with an addition of vulnerable service attribute. Nevertheless, the attributes of impact/effect are communication exists, new process creates, malicious code transferred as well as service terminated.

For the purpose of this paper, in explaining perspective trace pattern, let’s consider an incident that was caused by a worm, Sasser. Based on the logs (host and networks), the traces of scan event shows the attribute of communication exist between the victim and attacker via Destination IP Address and Source IP Address respectively. Next, the destination port open responded is port 445 and the operation type (action) is OPEN (in/out communication is allowed), whereas the connection request (protocol) is TCP (traffic packet is transmitted). The success of this event leads to the next event, exploit. In this event, the action continues with destination port open responded is 9996 and 5554. If the port 9996 is exploited, the operation type (action) is OPEN and the connection request is TCP, then partial exploit is in place. Port 5554 is also exploitable. If port 5554 is exploited, the operation type (action) is OPEN-INBOUND (traffic is allowed in) and the connection request is TCP (file is transmitted) which leads to vulnerable service (service) as FTP (file transfer occurred). We consider the exploit is successful if both ports above are exploited. As the consequences of the scan and exploit event, the impact/effect incident occurred. This event consists of few attributes namely; a) offender identified (who is victim and attacker), b) a process created (traffic action) which reside at %WINDIR%\System32\ftp.exe, c) the service terminated is lsass, and d) malicious code transferred (file transmitted) is ~\*.exe. The above example describes that the traces belong to attacker trace pattern.

The example can also be represented as an algorithm depicted in Table 1.

<table>
<thead>
<tr>
<th>Event Name: Scan</th>
<th>Attribute:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Exist := Source IP Address, Destination IP Address</td>
<td></td>
</tr>
<tr>
<td>Destination Port Open := 445</td>
<td></td>
</tr>
<tr>
<td>Operation Type := OPEN</td>
<td></td>
</tr>
<tr>
<td>Connection Request := TCP</td>
<td></td>
</tr>
<tr>
<td>Action := find_vulnerability(); determine_services();</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Name: Exploit</th>
<th>Attribute:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Exist := Source IP Address, Destination IP Address</td>
<td></td>
</tr>
<tr>
<td>Destination Port Open := 9996</td>
<td></td>
</tr>
<tr>
<td>Operation Type := OPEN</td>
<td></td>
</tr>
<tr>
<td>Connection Request := TCP</td>
<td></td>
</tr>
<tr>
<td>Vulnerable Services := FTP (9996 44 5554)</td>
<td></td>
</tr>
<tr>
<td>Action := exploit(); show_manipulation_activity();</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Name: Impact/effect</th>
<th>Attribute:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offender Identified := Attacker</td>
<td></td>
</tr>
<tr>
<td>New Process Created := lsass(System32)ftp.exe</td>
<td></td>
</tr>
<tr>
<td>Malicious Codes Transferred := ~*.exe</td>
<td></td>
</tr>
<tr>
<td>Service Terminated := lsass</td>
<td></td>
</tr>
<tr>
<td>Action := show_impact();</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Integration of Traceability Features and Digital Forensic Investigation Process

In order to provide the capability of tracing and mapping the accurate and complete evidence in digital forensic investigation process, the relationship between each trace should be identified to form the incident trace pattern. In this research, the ways for identifying this relationship is accomplished using features in traceability approach (definition, production and extraction) discussed previously. The integration of the traceability model’s features (TMF) in digital forensic investigation process (DFIP) is illustrated in Table 2.

In Table 2, TMF in DFIP indicate that there is a potential in implementing traceability features in forensic investigation process. As mentioned by [6] [37], traceability is an important element in forensic investigation process and it is related to the link element which is the key element used to form evidence’s chain of custody. It is impossible to prevent all internet misuse but it is not impossible to identify and trace the evidence, and then take appropriate action.

<table>
<thead>
<tr>
<th>Feature</th>
<th>TMF</th>
<th>TMF in DFIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition: related to the specification of the traces and traceable objects</td>
<td>identify traces, attributes</td>
<td>identify component in incident</td>
</tr>
<tr>
<td>Production: related to the capture of traces (relationships)</td>
<td>perception, registration and maintenance</td>
<td>hypothesis, identify forward and backward traceability, preservation of evidence</td>
</tr>
<tr>
<td>Extraction: related to the actual process of tracing</td>
<td>trace extraction mechanism</td>
<td>tracing the evidence using selective tracing to promote trace pattern</td>
</tr>
</tbody>
</table>

Therefore, without the traceability information, the investigation decisions and other valuable information for collecting and analysing the evidence could be mislaid. Hence, a traceability approach is necessary and
in this research, the proposed integration is named as trace map model.

3.3 Proposed Adaptation of Traceability Model in Digital Forensic Investigation Process Model

The proposed adaptation of traceability model in digital forensic investigation process model is based on work done in [33] and later known as Trace Map Model. This model uses event-based traceability technique which was motivated from the traceability model discussed in [38]. Ramesh introduced three components: stakeholder, subject and object.

In this model, the stakeholder represents people who have an interest on requirements and on the tracing of requirements, the source represents the origins of a requirement and the artifacts used for documentation purposes, and the object represents the inputs and outputs being traced. In Fig. 5, the model represents what type of information is presented including salient attributes or characteristics of the information which is referred as object. For example, this information can be represented as an attribute of object and the traceability across various object is represented by a link namely traces to. The model also shows the stakeholders are the people who play different roles in the creation, maintenance and use the various objects and traceability links across them. These stakeholders act in different roles or capacities in the establishment and use the various conceptual object and traceability links. The subject represents the location of the documented traceability information i.e. which state that all objects are documented by subjects.

Various dimension of traceability information is discussed in Ramesh’s model such as what kind of information is represented, who are the people that play the role, where and how the traceability information are represented, why and where the object are created, modified and evolved. The compatibility and the capability model also have been discussed in various business areas with different traceability focus. In this research, this model is adapted and integrated within the digital forensic investigation process which consists of three components, namely stakeholder, source of evidence and digital evidence as shown in Fig. 6.

These components map to the components in Ramesh’s model: stakeholder, subject and object respectively. Stakeholders refer to the people involved in the whole process of digital forensic investigation such as the auditor, network administrator, complainer (perspective as discussed in [33]) and forensic expert. In this research, these investigators will manage the source of evidence on the incident reported such as the devices (host and network) and the logs involved in the incident. Meanwhile, the digital evidence is defined as events of incident (see subsection Trace Pattern) that are documented in the source of evidence. This current relationship is further depicted using the diagram in Fig. 7. For the purpose of this research, the domain selected is malware intrusion incident.

Fig. 6 Conceptual Diagram of Digital Forensic Investigation Process

Fig. 7 Trace Map Model
incident. In this model, the traces of the offender are based on the primary events of incident that scan, exploit and impact/effect. In each event of incident, the trace patterns of the perspectives (victim, attacker, multi-step attacker) are established. The model also assists the investigator on identifying the relationship between the source of evidence, the digital evidence and the people involve during the investigation process, and provide a complete and accurate digital evidence of the incident reported.

4. CONCLUSIONS AND FUTURE WORKS

Traceability is an important element in forensic investigation process and related to the link element which is the key element used in forming the chain of evidence. Therefore, this research introduced a Trace Map Model that derived from the adaptation and integration of the traceability in digital forensic investigation process. The proposed model is used to provide the forensic investigation the capability to trace and map the digital evidence and source of evidence during the forensic investigation process.

This capability of the model is based on the preliminary assessment through the case study as presented in this paper. It also shows that the trace pattern enables us to identify the origin of malware intrusion through the traces attributes. These assist the investigator to show the relationship of the incident traces for obtaining the evidence accuracy and completeness that could enable the legal process to take its due course. In future, the effectiveness of the evidence tracing and mapping is evaluated through a validation process. It is foreseeable to develop a prototype that can be used as one of the forensic investigation tool through this proposed model.

ACKNOWLEDGMENT

We thank Universiti Teknikal Malaysia Melaka for the Short Grant Funding for this research project.

REFERENCES


