Journal of Information Security, 2013, 4, 19-32
http://dx.doi.org/10.4236/jis.2013.41004 Published Online January 2013 (http://www.scirp.org/journal/jis)

A Forensic Traceability Index in Digital Forensic Investigation

Siti Rahayu Selamat*, Shahrin Sahib, Nor Hafeizah, Robiah Yusof, Mohd Faizal Abdollah
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka City, Malaysia
Email: *sitirahayu@utem.edu.my, shahrinsahib@utem.edu.my, nor_hafeizah@utem.edu.my, robiah@utem.edu.my, faizalabdollah@utem.edu.my

Received September 14, 2012; revised October 12, 2012; accepted October 26, 2012

ABSTRACT

Digital crime inflicts immense damage to users and systems and now it has reached a level of sophistication that makes it difficult to track its sources or origins especially with the advancements in modern computers, networks and the availability of diverse digital devices. Forensic has an important role to facilitate investigations of illegal activities and inappropriate behaviors using scientific methodologies, techniques and investigation frameworks. Digital forensic is developed to investigate any digital devices in the detection of crime. This paper emphasized on the research of traceability aspects in digital forensic investigation process. This includes discovering of complex and huge volume of evidence and connecting meaningful relationships between them. The aim of this paper is to derive a traceability index as a useful indicator in measuring the accuracy and completeness of discovering the evidence. This index is demonstrated through a model (*TraceMap*) to facilitate the investigator in tracing and mapping the evidence in order to identify the origin of the crime or incident. In this paper, tracing rate, mapping rate and offender identification rate are used to present the level of tracing ability, mapping ability and identifying the offender ability respectively. This research has a high potential of being expanded into other research areas such as in digital evidence presentation.

Keywords: Digital Forensic Investigation; Traceability; Tracing Rate; Mapping Rate; Offender Identification Rate; Forensic Traceability Index; Trace Pattern

1. Introduction

Forensic or forensic science is the term given to an investigation of a crime using scientific means or used to describe crime detection in general. It is the application of a broad spectrum of sciences to answer questions of interest to a legal system. The emergence of forensic comes from the incidence of criminal, illegal and inappropriate behaviors.

The field of forensic science is vast. The majority of the public are probably exposed to only a few different types of forensic. A sub-discipline of forensic known as digital forensic is developed to investigate any digital devices in the detection of crime. With the development of modern computers, network and the internet, computer-related crimes have become a threat to society because of the immense damage it can inflict while at the same time it has reached a level of sophistication. This sophistication makes tracking the sources difficult.

This paper highlights the tracking issues or also known as traceability aspects due to the complexity of the crime in digital forensic investigation process. The current traceability of cybercrimes basically allows two consequences. First is to identify the scope of an attack instead of the actual attacker and second is to assess the liability of an organization [1]. However, according to [1-3], there was a critical need to deal with issue of origin identification and cross referencing in investigation process. Hence, traceability is not only important to avoid misleading in decision making but also to ensure the valuable information collected is complete and accurate. A methodology to overcome the traceability issue in digital forensic investigation process is developed by introducing the evidence tracing and mapping procedures. These procedures were later used to formulate the traceability index. The ability to trace and map the evidence complete and accurately could assist in practitioner decision making.

2. Related Work

2.1. Digital Forensic Investigation Framework

A common definition of digital forensic is the use of scientifically derived and proven methods toward the process of preservation, collection, validation, identification,

*Corresponding author.

analysis, interpretation, documentation, and presentation of digital evidence which is derived from the digital sources [4]. The purpose of these processes is to facilitate the reconstruction of events or to help anticipate illegal actions. [5] had simplified the definition as the process of preservation, identification, extraction, documentation and interpretation of computer media for evidentiary and (or) root cause analysis.

Though to some researchers the digital forensic is inclusive of computer forensic, network forensic, software forensic and information forensic, but it is largely used interchangeably with computer forensic [3]. Computer forensic implies a connection between computers, the scientific method, and crime detection. It includes devices other than general-purpose computer systems such as network devices, cell phones, and other devices with embedded systems. There are over hundreds of digital forensic investigation procedures developed in digital forensic investigation practices. An organization tends to develop its own procedures and some focused on the technology aspects such as data acquisition or data analysis [6]. Most of these procedures were developed in 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. However, [7,8] stated that the process of the investigation should be incorporated with the basic procedures in forensic investigation which are preparation, investigation and presentation. A categorization of investigation process was done in [9] to group and merge the similar activities or processes in five phases that provide the same output. The phases are: Phase 1 (Preparation), Phase 2 (Collection and Preservation), Phase 3 (Examination and Analysis), Phase 4 (Presentation and Reporting), and Phase 5 (Disseminating the case). The researcher also proposed a mapping process of digital forensic investigation process model to eliminate the redundancy of the process involved in the model and standardize the terms used in achieving the investigation goal.

The analysis emphasized that most of the frameworks consist of Phase 2 (Collection and Preservation), Phase 3 (Examination and Analysis), and Phase 4 (Presentation and Reporting) except Phase 1 and Phase 5. The analysis also propose that even though, Phase 1 and Phase 5 are not included in some of the framework, the study of [7,10-17] indicate the needs of both phases to confirm the completeness of the investigation. The purpose of Phase 1 comes in two objectives: 1) to approve that an investigation process can start and run in a proper procedure, and 2) to protect the chain of the evidence. The purpose of Phase 5 is to avoid the possibility of the incomplete investigation and lack of improvement in investigation procedures. From the analysis, it shows that an appropriate digital forensic investigation framework should at least consist of: Preparation Phase, Collection and Preservation Phase, Examination and Analysis Phase, Presentation and Reporting, and Disseminating the case. From the work done in [9], this paper focused on the Collection and Preservation Phase which has been identified as one of the critical phases of the digital forensic investigation process model. The Collection and Preservation Phase is where the digital evidence is identified, collected and preserved which then is analyzed and extracted to be presented in a court of law. However, to make it acceptable in court, there are two issues to be considered: the digital evidence itself and the collection process.

2.2. Characteristic Issues of Digital Evidence

[18] addressed four issues of digital evidence itself. First, the digital evidence is in a disorganized form and as such it can be very difficult to handle and not all of them is obviously readable by human. For example, a hard drive platter contains messy pieces of information mixed together and layered on top of each other over time. Because of that, only a small portion of the information is relevant to the case which makes it necessary to extract useful pieces, fit them together and translate them into a form that can be interpreted. Second, digital evidence generally is an abstraction of some event or digital object and can be seen as residual data that give a partial view of what occurred in the incident being investigated. Third, digital evidence can be maliciously altered or changed during collection without leaving any obvious trace indicating that alteration has taken place. This is due to the fact that computer data can be easily manipulated. Lastly, traditional evidences are created and retrieved as a single record but in a great majority of modern cases, it involves computerized system where evidence is created or retrieved from different records and sources.

2.3. Managing Issues in the Collection Process of Digital Evidence

In the collection process, the issues are the approach on collecting, analyzing and presenting the evidence. During collection process, the evidence is related to the aspect on how the evidence is searched, collected, analyzed, presented and documented without tampering the evidence and preserving the chain of evidence. In analyzing the evidence, the issue is about the aspect on the process of analysis. These cover all aspects such as the tools that are used for the analysis, the person responsible for the analysis and the integrity of the evidence. During the analysis process, the analysis tools used must be legally accepted, performed by experts or qualified person, and the evidence is not tampered. The issue on presenting the evidence is concerned with the approach of presenting

and documenting the evidence in an understandable manner to non-technical person such as jury and judge.

Another problem during the collection process is the diversity of devices. In network, these devices generate a huge volume of evidence [19-21]. This situation leads to difficulty in identifying sources of the potential evidence or in tracing the evidence as stated in [22]. As it is important to obtain acceptable evidence in the court of law, the investigation must be successfully performed without tampering the evidence and also able to prove the evidence is legitimate. To solve the problem mentioned above, the ability to track, link and preserve the chain of evidence in huge volume of evidence is crucial. Hence, the traceability is one of the important elements during the digital forensic investigation process in identifying the origin and become the first challenge in the investigation as mentioned in [1-3].

2.4. The Tracing and Mapping in Digital Forensic Traceability

Traceability gives meaningful information through the study of the related links. The collected digital evidence must give appropriate information or meanings to the collector. The information cannot be attained via single digital evidence as it is meaningless [23-26]. Therefore, to avoid meaningless information, the link between the collected digital evidence must be identified. The objective of traceability is to identify and track real or imaginary objects through a process chain [28]. Given the origin of an object, traceability provides the opportunity to track a chain of events, or to predict process outcomes. The definition of traceability can be broad due to the complexity of it processes and the way it implemented [29,30]. For example, in networks, traceability refers to how difficult it is to establish the source and destination of communications on computers and communication networks.

In this paper, traceability is defined as the ability to trace and map the events of an incident from different sources in order to obtain useful evidence and well managed. In order to have the evidence well managed, the works of [31-33] suggested that traceability can be established from the source evidence to its lower level evidence and from the lower level evidence back to their sources. This situation brings the concept of forward and backward traceability or called as bidirectional traceability approach as discussed by [33,34]. The concept was used and further extend as an enhanced traceability model discussed in [35]. The discussed traceability model consists of definition, production and extraction. This model established the concepts of trace and map within the traceability: The process of establishing the structures is referred as tracing the digital evidence, whereas the process of putting the structure according to the hypothesis/scenario is referred as mapping the digital evidence. Next, the construction of tracing and mapping procedures from the model are explained.

3. Method

In this section, the experimental design to establish the trace and map concepts in traceability from the perspective of digital forensic investigation is presented. It is note that while the traceability model is adapted from other domain, the model structure is able to be implemented in digital forensic investigation due to its compatibility and capability. We present the data collection and data analysis through a controlled experiment for data scenario using malware intrusion. The findings from the analysis will be used as the primary guideline to establish the generic incident trace pattern. The components of the trace pattern are used to formulate the traceability index for digital forensic investigation process.

3.1. Inquisition of Incident Scenario

In this research, a controlled experiment is designed to run the worm intrusion, to collect logs from each of the devices involved and to design the intrusion scenario. The design is motivated by [36]. In Rahmani *et al*. research, the experiment is focusing on DDoS attack that involves few processes which are declare attack setup, run selected attack, collect logged MIB variables and analyze result. It consists of four processes: Network Environment Setup, Attack Activation, Incident Log Collection and Incident Log Analysis as described in [37, 38].

In this experiment, the worm intrusion is launched and the intrusion activities are captured in the selected logs which are personal firewall log, security log, system log, application log, IDS log, tcpdump and Wireshark log. The researchers have collected all logs generated during the experiment and nine intrusion scenarios are derived based on the log analysis are identified. For the purpose of this paper, Scenario A as depicted in **Figure 1** is selected as the example. **Figure 1** illustrates the incident scenarios of Blaster A (Scenario A). In this incident scenario, the attack is activated in host *Mohd* and this host is successfully exploited all hosts except for hosts *Ramly*, *Abdollah*, *Roslan* and *Sahib*. Subsequently, one of the infected hosts, *Selamat* has organized an attack on host *Ramly*. In this incident scenario, the hosts that are marked with 1,354,444 and 69 are indicated as have been successfully being exploited by the attacker and have been infected. The hosts that are marked with 135 and 4444 demonstrated the attacker has already opened the backdoor but has not successfully transferred the exploit codes through *port* 69.

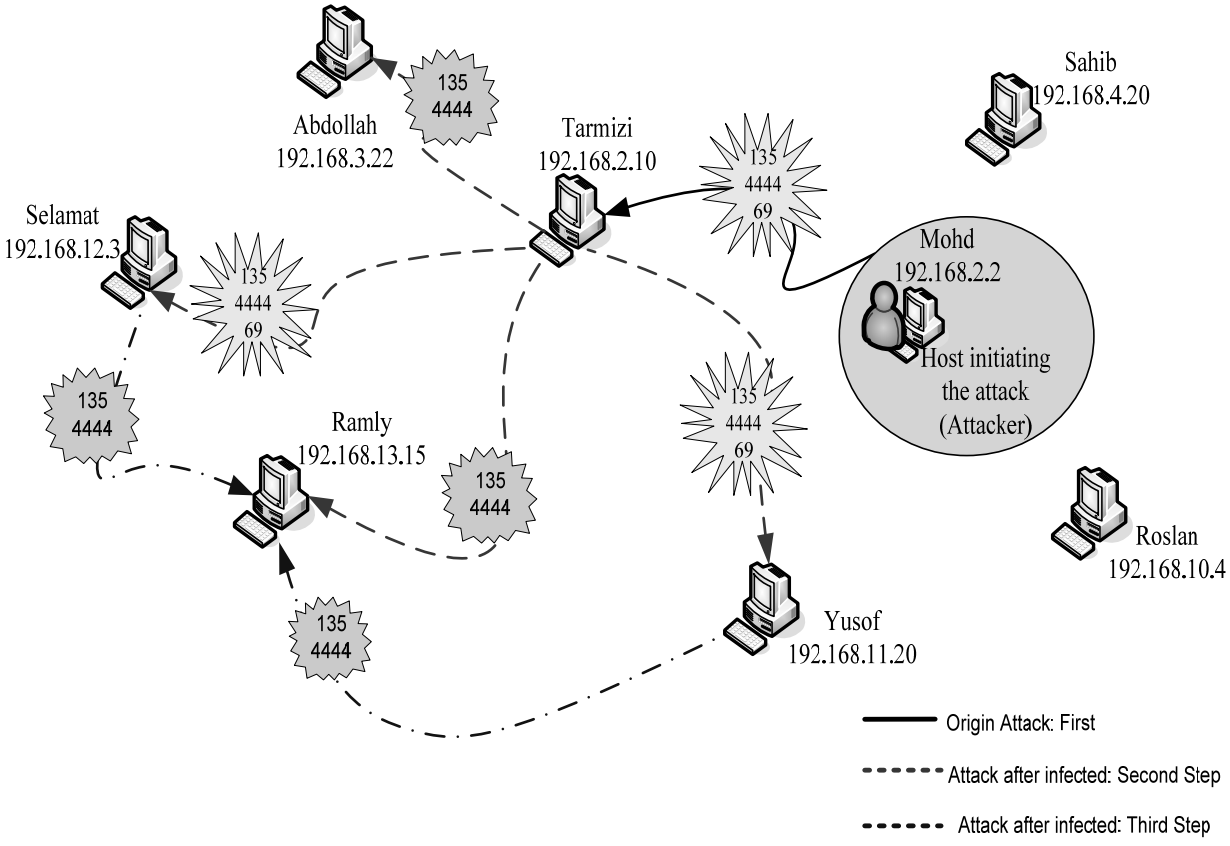

Figure 1. Blaster A incident-Scenario A.

3.2. Identification of Incident Trace Pattern

An attack pattern is defined as a mechanism to capture and communicate at the attacker perspective that shows the common methods for exploiting software, system or network [39-41]. It describes how an attack is performed, how the security pattern is enumerates to defeat the attack, and how to trace an attack once it has occurred [42,43]. It provides a systematic description of the attack goals and attack strategies to defend and trace the attack. Attack patterns can guide forensic investigators in searching the evidence. This also helps them at the data collection phase to determine and identify all the components to be collected, decide the priority of the data, find the location of the components and collect data from each of the component during the investigation process [43]. In general, attack pattern is very important in providing a way to protect the system from any potential attack. The existing researches done by [39-43] reveal an attack pattern is a type of pattern that focuses on the attacker perspective while victim perspective is omitted. In forensic view, both perspectives are important. A victim or attacker can be identified based on the traces data found in the attack pattern analysis. In forensic, these traces data are represent in the form of trace pattern to determine how a crime is being committed.

Trace pattern is defined as a regular way of process discovering the origin or starting point of a scenario that has happened. It is an essential element in helping investtigator in finding evidence of crime or incident. For example, in a digital crime the evidence can be found in any digital devices. The objective of these patterns is to provide clear view on how an attack is performed and its impact.

Based on the previous work done in [37,38,44,45], three generic malware incident trace pattern for victim perspective, attacker perspective and multistep attacker perspective are established by observing the traces leave on the selected logs. For example, the generic malware incident trace pattern for victim perspective is depicted in **Figure 2**.

Figure 2 depicts the *generic malware incident trace pattern* for victim perspective consists of the incident traces (evidence) from host level and network level. In host level, three main events of the incident occur. The events are *scan event*, *exploit event* and *impact*/*effect event*. In this level, the *scan event* and *exploit event* occur in the *personal firewall log.* Whereas the incident traces of the *impact*/*effect event* are found in the *system log*, *application log* and *security log*.

In this research, the *personal firewall log* is identified

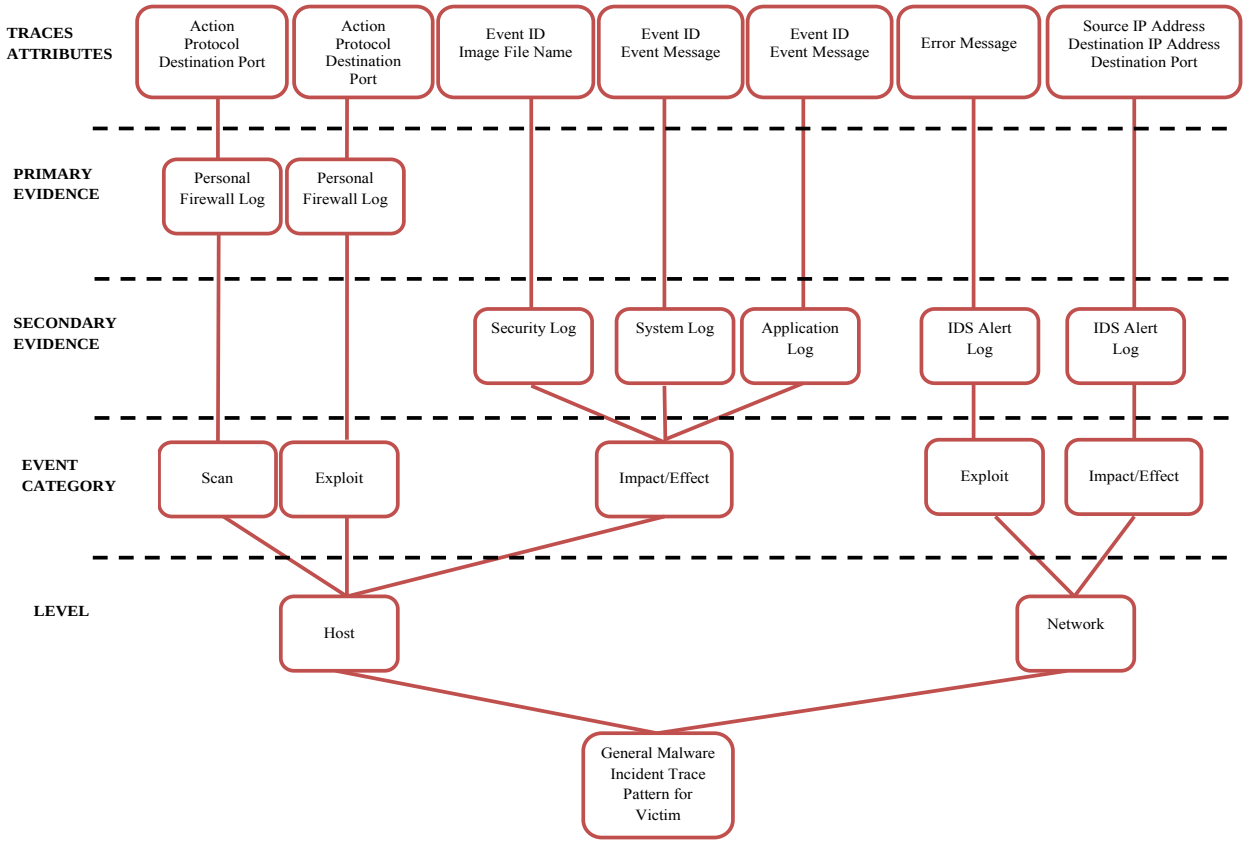

Figure 2. Generic malware incident trace pattern-victim perspective.

as the *primary evidence* and the generic attributes of this log are *action*, *protocol* and *destination port*. The *security log*, *application log* and *system log* are considered as the *secondary evidence* and the generic attributes of these logs are *event id*, *image filename* and *event message*.

In network level, only two main events occurred namely *exploit event* and *impact/effect event*. In this level, the *exploit event* and *impact/effect event* exist in the IDS *alert log*. This IDS *alert log* is considered as *secondary evidence* and the generic attributes of this log are *error message*, *source IP address*, *destination IP address* and *destination port*. The attribute of *error message* indicates the exploiting activities and the attributes of *source IP address*, *destination IP address* and *destination port* indicate the impact of the attack that shows the offender of the incident (*victim*, *attacker* or *multi-step attacker*).

In contrast, the *generic malware incident trace pattern* for attacker perspective consists of incident traces from host level and network level. In these levels, three events (*scan*, *exploit* and *impact/effect*) of the incidents occur at the host level and two events (*scan* and *impact/effect*) occur at the network level. The events at the network level show the difference between the *generic malware incident trace pattern* for victim perspective and *generic malware incident trace pattern* for attacker perspective.

In victim perspective, it consists of *exploit event* and *impact/effect event* as highlighted in **Figure 2** with the traces attributes are *source IP address*, *destination IP address*, *destination port* and *error message* compared to attacker perspective trace pattern, it consists of *scan event* and *impact/effect event* with the traces attributes are *error message* and *source IP address*. The *generic malware incident trace pattern* for multistep attacker perspective is similar to the *generic malware incident trace pattern* for victim perspective except for *impact/effect* in network log. The considered traces attributes are only *destination IP address* and *destination port* for victim; and *error message* and *source IP address* for multi-step attacker incident trace pattern.

3.3. Constructing Tracing and Mapping Procedures

The generic malware *victim*, *attacker* and *multi-step attacker incident trace pattern* then are used in formulating the tracing and mapping evidence procedures to demonstrate the ability of trace and map evidence (traceability) in digital forensic investigation process. Tracing evidence procedures are necessary in order to extract the incident traces from the logs [46]. However, the extracted traces are meaningless without knowing the rela-

tion between those traces; hence identifying the relationships are important. These relationships can be identified by mapping or linking process [47].

In this research, mapping the evidence is connecting or linking all traces discovered from the tracing activities by correlating the traces with the origin of the traces. The purpose of this mapping is to provide evidence that can answer the questions about the incident occurred. The mapping also is an aid to diagnostic the decision regarding to the incident. Due to the important of tracing and mapping evidence in digital forensic investigation process, the tracing and mapping procedures are formulated based on the proposed generic malware incident trace pattern as shown in **Figures 3** and **4** respectively.

As shown in **Figure 3**, the tracing processes are initially started at the *personal firewall log* followed by *security log*, *system log*, *application log* and IDS *alert log*. The aim of these tracing processes is to examine the traces left in the logs by focusing the traces of three main events of an incident which are *scan*, *exploit and impact*/*effect*. *Scan* events can only be found in *personal firewall log*, while *exploit* events can be found in both *personal firewall log* and IDS *alert log*. The *impact*/*effect* events can be found in *system*, *security* and *application logs*.

For example, the tracing procedure for tracing the incident traces from the *personal firewall log* starts with tracing the traces attributes identified in the proposed generic incident trace pattern. The traces attributes are perspective *IP address*, *action*, *protocol* and *destination*

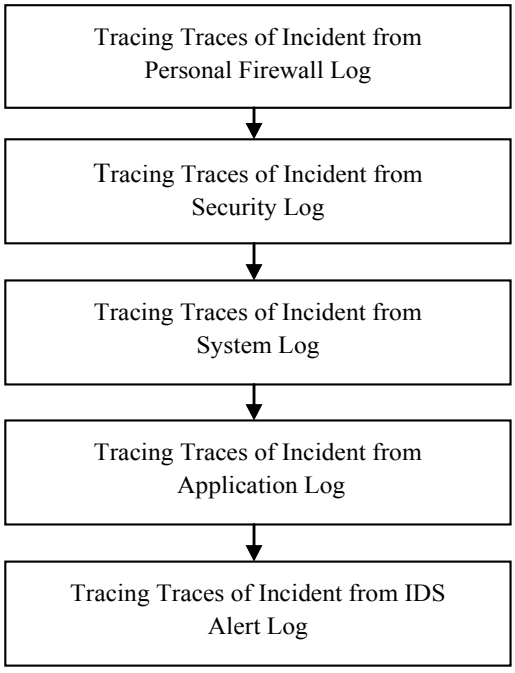

Figure 3. Tracing procedures for tracing evidence of malware incident.

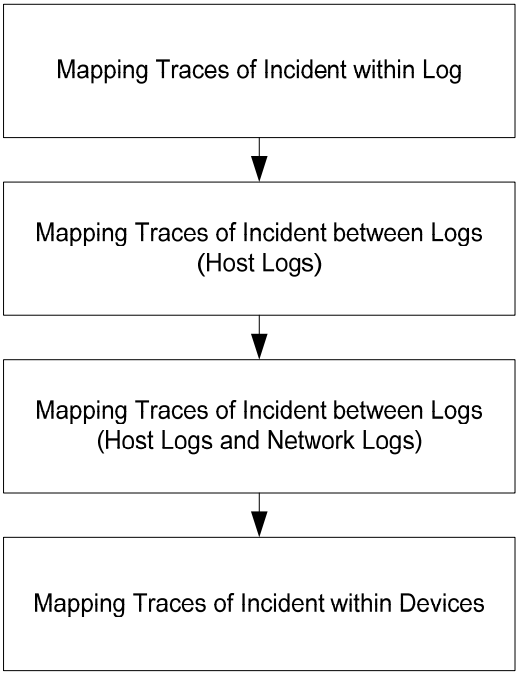

Figure 4. Mapping procedures of incident traces.

port. The perspective *IP address* refers to the *destination IP address* or the *source IP address* of the perspective. *Action* indicates the perspective is trying to open the connection. *Protocol* and *destination port* show an attack is attempted to establish the communication. In this tracing process, any relevant set of traces that consist of *destination port* (*x*), *action* (*y*) and *protocol* (*z*) that are found in the *personal firewall log* is assigned as trace P.

The tracing begin with some assumptions:

Let Dp(*x*) = the set of all vulnerable destination ports that is used by malware in the incident

Let Act(*y*) = the set of all actions that is used by malware in the incident

Let Pr(*z*) = the set of all protocols that is used by malware in the incident

Let $P_n(x,y,z)$ = set of traces in personal firewall log

A predicate is established that defines the traces as below:

$$\forall x \forall y \forall z \left(\left(\mathrm{Dp}(x) \wedge \mathrm{Act}(y) \wedge \mathrm{Pr}(z) \right) \rightarrow \exists \mathrm{P}_n(x,y,z) \right) \quad (1)$$

which can defines that for all vulnerable destination port *x*, action *y* and protocol *z*, there exists trace of an incident from personal firewall log, such that $\mathrm{P}_n(x,y,z)$.

The incident traces found from the tracing procedures are then mapped to show the relationship of the evidence of the incident discovered during the investigation process. The mapping procedures are depicted in **Figure 4**.

Figure 4 illustrates the mapping procedures of the incident traces discovered from the tracing process. Firstly, the traces that are discovered from the tracing process are

mapped within logs. For example, the traces discovered in *personal firewall log* are mapped with the traces discovered in the same log.

Secondly, the traces that are mapped from the first mapping process are further mapped between logs. For example, the traces discovered in *personal firewall log* are mapped to the traces discovered in *security log*.

Thirdly, the traces that are mapped from the second mapping process are further mapped to the logs that are in different level of communication layer. For example, the traces mapped from the host logs (*personal firewall log*, *security log*, *system log and application log*) are further mapped to network log (IDS *alert log*).

Finally, the traces that are mapped from the third mapping process are further mapped within devices. For example, the mapped traces that belong to the same devices will be mapped and merged to the same devices. In this research, the traces will be tagged for each new mapped trace. The objective of this tagging is to show the traces are already mapped between one trace to another trace either belongs to same log (within log), different logs (between logs), same level (host logs), different level (host and network logs) or same devices (within devices).

For example, in the mapping procedures for mapping traces of incident between logs (host log), the identified incident traces from specified host log are mapped within the same host. This mapping is required to identify the relationship of the relevant incident traces at host level. In this process, the traces of incident in *security log* SE_m are firstly mapped to the incident traces in *system log* SY_m. The traces that are mapped are assigned to a new mapped trace $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m)$. These traces, $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m)$ are further mapped with the incident traces in *personal firewall log* P_m to produce a new mapped trace and assigned the trace as $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m,\mathrm{P}_m)$. Finally, the mapped traces $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m,\mathrm{P}_m)$ are further mapped with *application log* AP_m to produce the complete mapped traces from host log and assigned as $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m,\mathrm{P}_m,\mathrm{AP}_m)$.

In this research, the *security log* is identified as one of the important logs for malware incident; therefore the unmapped traces of incident in *security log* also are mapped with the traces of incident in *personal firewall log*. This mapping process produces a new mapped traces and assigned as $\mathrm{TM}(\mathrm{SE}_n,\mathrm{P}_m,\mathrm{AP}_m)$. Hence, these mapping procedures can produce three types of mapped traces that consist of traces of incident from logs selected in this research: $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m,\mathrm{P}_m,\mathrm{AP}_m), \mathrm{TM}(\mathrm{SE}_n,\mathrm{P}_m,\mathrm{AP}_m)$ and $\mathrm{TM}(\mathrm{SE}_m,\mathrm{SY}_m,\mathrm{P}_m,\mathrm{AP}_m)$. These mapping procedures are summarized as illustrated in **Figure 5**.

These tracing and mapping procedures have been implemented on the proposed model in the previous work

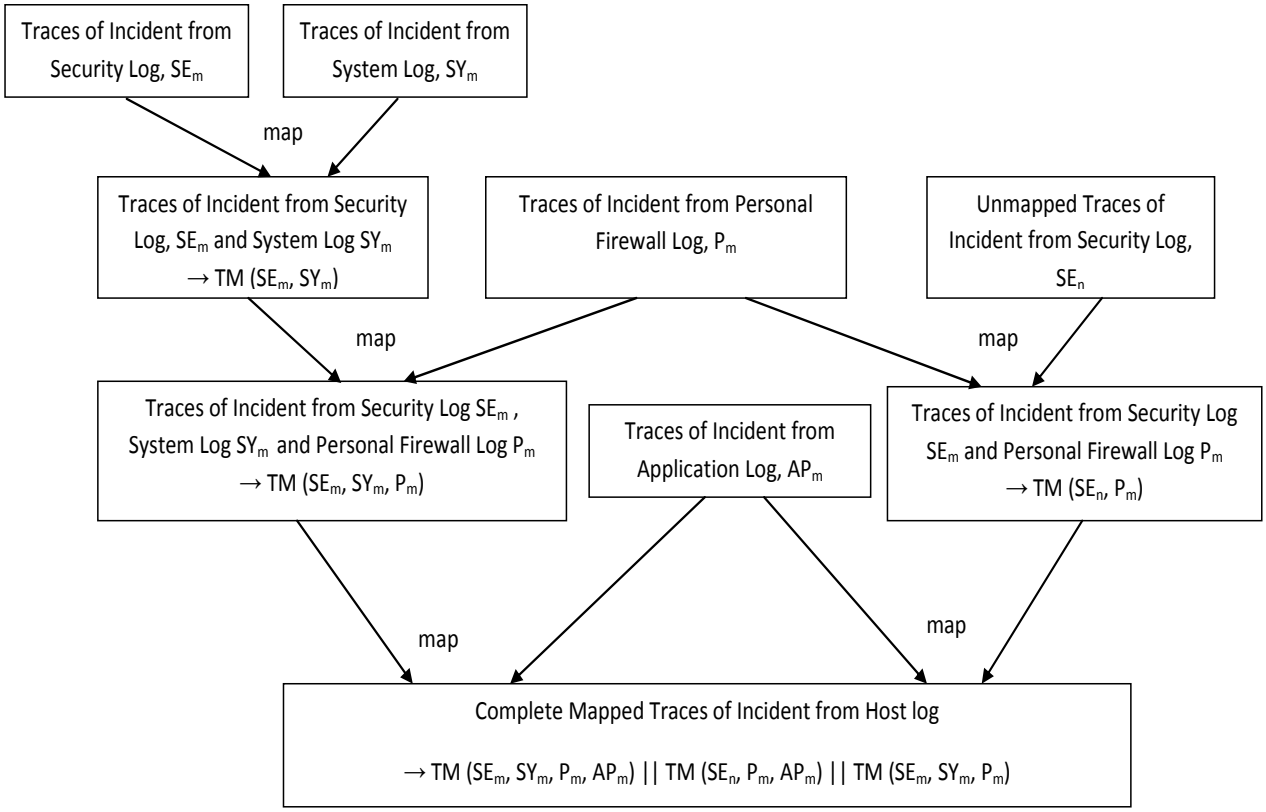

Figure 5. Mapping procedures for host incident traces.

done and discussed in [35,48]. A prototype version is developed in order to demonstrate the proposed model that called as *TraceMap* and can benefit the digital forensic investigation research and community. The *TraceMap* consist of generic trace pattern structure, tracing and mapping procedure, and offender identification procedure. Twelve datasets are seed into the *TraceMap* implementation and the result is generated as an incident report.

4. Results

The objective of this section is to identify the ability of the tracing and mapping procedures in the *TraceMap* implementation to discover the incident traces. This could facilitate the investigator in identifying the origin of the incident. In this research, three main capabilities need to be measured to identify the effectiveness of the *TraceMap*. The capabilities are tracing capability, mapping capability and offender identification capability. Thus, the metrics for evaluation are formulated and further discussed in next subsection paper.

4.1. Forensic Traceability Measurement

According to [49], there is a need for security metrics in digital forensic that: 1) meet legal requirements for measureable reliability, authenticity, accuracy and precision, 2) based on a sound scientific methodology properly applied, and 3) have a basis provided for independent testing. Unfortunately, the digital forensic metric is not yet formulated and there is no industry consensus that a judge and jury can rely upon as adequate to support a claim and meet legal requirements for measurable reliability, authenticity, accuracy, and precision. These are currently elusive and must be constructed on a case-by-case basis. Due to this reason, it is possible to transfer concepts from another research area to build a metric to be used in the current research as described in [50]. Therefore, this research is proposed to transfer concepts from another research area to build a metric to measure the effectiveness of the tracing evidence (incident traces), mapping evidence (incident traces), and identifying the origin of the incident.

Two metrics from other research domain, namely information retrieval and IDS information theory is explained to be used to build a metric for digital forensic. The aim is to measure the capability of tracing the incident traces and to measure the capability of mapping the incident traces in the *TraceMap* respectively.

In the concept of information retrieval, two metrics are used to measure the information retrieval; recall and precision. In the field of information retrieval, precision is the fraction of retrieved documents that are relevant to the search whereas recall is the fraction of the documents that are relevant to the query that are successfully retrieved [51]. This concept has been used in measuring the performance of the digital forensic investigation tools such as FTK™ and Encase™ by measuring the query precision and query recall [52]. However, the classic precision-recall trade-off dilemma is that as recall increases, the query precision decreases. While digital forensic seeks recall rates at or near 100%, query precision is usually low [53]. Thus, based on the aim of transferring the concept of information retrieval mentioned previously, the recall metric is adapted due to the aim of the tracing evidence (incident traces) in this research is to retrieve all relevant incident traces to the hypothesis (query). This metric is used to measure the capability of tracing evidence (incident traces) and later known as *Tracing Rate* (TC_R).

Tracing Rate (TC_R) is the ratio of relevant traces discovered (output), $N_{relevant\_traces}$ and the total traces (input), N_{total_traces}. This metric is represented as in Equation (2):

$$TC_R = \frac{N_{relevant\_traces}}{N_{total\_traces}} \quad (2)$$

where:

$N_{relevant\_traces}$ is the number of relevant traces discovered (output) from all potential sources of evidence.

N_{total_traces} is the number of the traces (input) from all potential sources of evidence.

Equation (2) is used to evaluate the effectiveness of the tracing process of the *TraceMap*. The result from this evaluation is used to demonstrate the capability of tracing evidence (incident traces) for digital forensic investtigation process.

[54] uses the concepts of information theory to motivate an information theoretic metric for IDS. This metric is used to measure the capability of detecting the intrusion which is known as Intrusion Detection Capability (C_{ID}). C_{ID} is the ratio of mutual information between IDS input and output $I(X,Y)$ and the entropy $H(X)$ of the input that can be expressed as $C_{ID} = I(X,Y)/H(X)$. The mutual information $I(X;Y)$ corresponds to the intersection of the information in X with the information in Y. The entropy $H(X)$ is the sum of the mutual information of X and Y. Thus, based on the aim of transferring the concept of information theory used for IDS mentioned previously, this metric is adapted in this research, in which the mutual information $I(X;Y)$ is referred to the relevant incident traces that are mapped or connected. As a result, this metric is used to measure the capability of mapping the incident traces and is later known as *Mapping Rate* (MP_R).

Mapping Rate (MP_R) is the ratio of the output traces that are relevant and mapped, $N_{relevant\_map\_traces}$ and the

input traces that are relevant, $N_{relevant\_traces}$. This metric is represented as in Equation (3):

$$MP_R = \frac{N_{relevant\_map\_traces}}{N_{relevant\_traces}} \quad (3)$$

where:

$N_{relevant\_map\_traces}$ is the number of traces that is part of incident is relevant and mapped

$N_{relevant\_traces}$ is the number of the traces relevant traces

Equation 3 is used to evaluate the effectiveness of the mapping procedures of the *TraceMap* on mapping the relevant traces of the incident within and between *sources of evidence*. The result from the evaluation is used to demonstrate the capability of mapping the incident traces for digital forensic investigation process.

In order to identify the evidence origin, there is a need to determine the rate of successfully origin found from the mapped traces. To achieve this, another metric is proposed named as *Offender Identification Rate* (OI_R). This metric is significant in identifying the offenders (perspectives) of the incident. The *Offender Identification Rate* (OI_R) is the ratio of mapped incident traces that are matched with the incident trace pattern, $N_{tracepattern\_traces}$ and the number of traces that are relevant and mapped, $N_{relevant\_map\_traces}$. This metric is represented as in Equation (4):

$$OI_R = \frac{N_{tracepattern\_traces}}{N_{relevant\_map\_traces}} \quad (4)$$

where:

$N_{tracepattern\_traces}$ is the number of mapped traces that can match with the perspective incident trace pattern.

$N_{relevant\_map\_traces}$ is the number of the traces that is relevant and mapped.

Equation (4) is used to evaluate the effectiveness of the *TraceMap* and the result from the evaluation is used to demonstrate the capability of identifying the offender or the origin of the incident.

4.2. The Significant Findings of Tracing Process

According to [52,53], the percentage of tracing rate and mapping rate imply the effectiveness level of the *TraceMap.* To be specific, in order to indicate it effectiveness, the range of the rate percentage of TC_R should be at or near 100% [52]. However, to the best of our knowledge, the range of rate percentage of MP_R has not been mentioned specifically in any research. Thus, in this paper, to show the accuracy of the incident traces, the level of MP_R rate should be close to TC_R rate.

The effectiveness of the tracing process, mapping process and offender identification process is measured using the metrics proposed in this research. The metrics are *Tracing Rate* (TC_R), *Mapping Rate* (MP_R) and *Offender Identification Rate* (OI_R). The objective of tracing is to discover the relevant incident traces of the incident. By using the proposed TC_R metric, the results show the *TraceMap* is able to discover the relevant incident traces and could answer the incident from the range of 82.60% to 99.17%. The relevant traces discovered to the irrelevant traces discovered are compared as depicted in **Figure 6**.

The graph in **Figure 6** shows that there is significant result in relevant incident traces especially in DS7, DS8 and DS9. Overall, using the *TraceMap*, the relevant traces are 82.60% to 99.17% while irrelevant traces are 0.83% to 17.40%. It means that out of 100 reported potential evidence for Incident A, the probability of successfully evidence traced as relevant to that particular incident is as maximum as 99%.

4.3. The Comprehensiveness of Mapping Process

In this research, the process of mapping is divided into four stages. The stages are mapping the incident traces within log, mapping the incident traces between logs (host logs), mapping the incident traces between logs (network logs) and mapping the incident traces within devices (hosts). In this process, the result from each stage becomes the input to the next stage. For example, the result obtained from the first stage (mapping incident traces within log) will be used as the input to the second stage (mapping incident traces between logs). For each stage, the mapping rates are measured using the mapping metric as in Equation (4). The summary of the result generated is depicted in **Table 1**.

In **Table 1**, the second column (Stage 1) demonstrates the process of mapping the incident traces within host logs or sensor logs in which the percentage of the mapping rates are in the range of 82.60% to 99.96%. These rates show most of the relevant incident traces discovered has been mapped. It shows the number of traces that are relevant and mapped is consistent to the number of relevant traces discovered. These results indicate the effectiveness of the mapping process for mapping the incident traces within logs in which the most significant result shown in DS4, DS5 and DS6 with the percentage of mapping rates are 99.90%, 99.95% and 99.96% respectively.

The third column (Stage II) demonstrates the mapping process of the incident traces between logs. The result in this column implies that the incident traces from the mapping incident traces within log is the most relevant traces to be mapped. It shows that all relevant traces discovered from different *sources of evidence* are the most relevant traces that can be mapped between the logs belong to the same devices or hosts with the percentage of

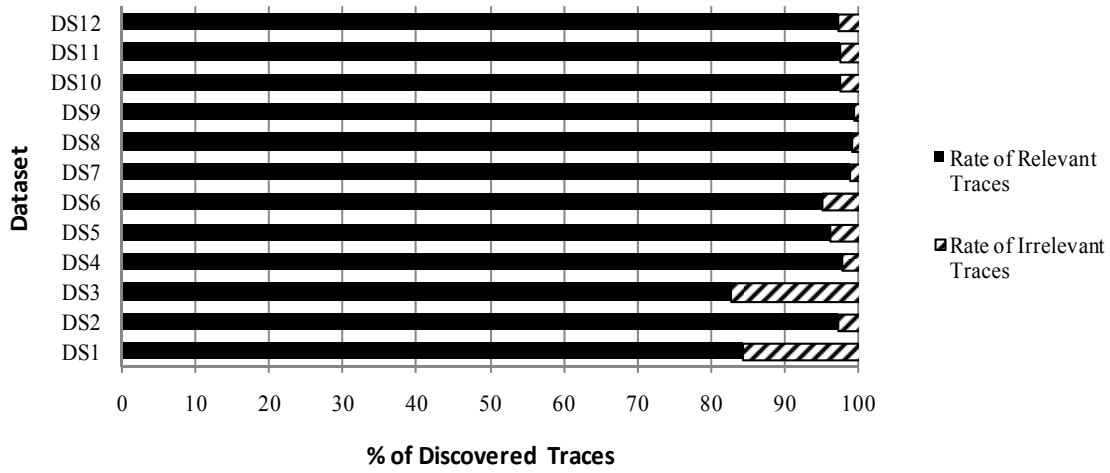

Figure 6. Rates of Relevant and Irrelevant Traces of Incident.

Table 1. Summary of mapping rate (MP_R).

Dataset	Stage 1	Stage II	Stage III	Stage IV
	% of Mapping Rate MP_R	% of Mapping Rate MP_R	% of Mapping Rate, MP_R	% of Mapping Rate, MP_R
DS1	97.61	100.00	69.57	100.00
DS2	97.08	100.00	82.86	100.00
DS3	82.60	100.00	66.67	100.00
DS4	99.90	100.00	81.48	100.00
DS5	99.95	100.00	87.80	100.00
DS6	99.96	100.00	63.89	100.00
DS7	99.31	100.00	77.27	100.00
DS8	99.21	100.00	78.38	100.00
DS9	98.92	100.00	74.29	100.00
DS10	99.29	100.00	93.62	100.00
DS11	99.10	100.00	78.57	100.00
DS12	99.14	100.00	85.71	100.00

the mapping rates are at 100%. These results indicate the process of mapping the incident traces between logs for host logs are effective and capable to map all relevant traces successfully.

The fourth column (Stage III) depicts the percentage of the *Mapping Rate* (MP_R) achieved in this process is in the range of 66.67% to 93.62% with the highest rate is obtained from Dataset 10 (DS10) and the lowest rate is obtained from Dataset 3 (DS3). These rates show the incident traces from network log is not probable to be mapped with the incident traces from host. These also reveal that it is difficult to prove the incident has happened based on the evidence provided from network and this cause the origin identification of the incident become difficult. These findings show that some of the traces from the network log are considered as false alarm which is unrelated with the traces discovered in the host log. Even though the percentage of the mapping rate in this stage indicates a low rate because only one dataset has achieved the high percentage which is 93.62% compared to other dataset (66.67% to 87.80%). But, this mapping process is still effective since the mapping procedure in this stage is able to determine the relationship between traces from host log and traces from network log.

Finally, the last column (Stage IV) shows the percentage of the MP_R is 100% for all dataset evaluated. These rates demonstrate the results from the third mapping process are the most relevant traces of the incident. The

rates also show the mapping process is completed in which all relevant incident traces are mapped.

4.4. The Use of Trace Pattern for Successfulness Offender Identification

The purpose of metric used in this process is to measure the effectiveness of *TraceMap* to assist the investigator in identifying all potential offenders involved in incident. In this research, the offender is named as perspective; *victim*, *attacker* and *multi-step attacker* (*victim*/*attacker*). The *Offender identification Rate* (OI_R) is calculated using Equation 3 and the result generated is shows in **Table 2**.

As shown in **Table 2**, the percentage rate for offender identification calculated from DS1 is 100% which indicates the offender has been successfully identified from all relevant incident traces mapped and matched with the hypothesis formulated at the beginning of the investigation process.

In the case of DS1, seven offenders (perspectives) are identified with four of them are victims, one attacker and two are multi-step attacker (victim/attacker). Although the percentage of the OI_R is 100% for all datasets as depicted in **Table 2**, three (DS2, DS3 and DS12) out of twelve dataset are unsuccessful in identifying the true attacker. These results are gained due to the traces identified are not documented in any logs selected in this research. However, in such cases, the attacker can still be identified by analyzing the multi-step attacker (victim/attacker) traces.

5. Derivation of Traceability Metric in Digital Forensic Investigation Process

As shown in Section IV, the traceability metric in this research, consists of *Tracing Rate* (TC_R), *Mapping Rate* (MP_R) and *Offender Identification Rate* (OI_R).

where:

n as the maximum number of stage and *i* as the current stage

$N_{relevant\_traces}$ is the number of relevant traces discovered (output) from all potential sources of evidence

N_{total_traces} is the number of the traces (input) from all potential sources of evidence

$N_{relevant\_map\_traces}$ is the number of traces that is part of incident that is relevant and mapped

$N_{relevant\_traces\_TC}$ is the number of the traces relevant traces discovered (output) given by TC

$N_{relevant\_map\_traces\_MP\_stage\_of\_n}$ is the number of traces that is part of incident is relevant and mapped given by MP at the stage n

$N_{relevant\_traces\_MP}$ is the number of the traces relevant traces discovered (output) given by MP

$N_{tracepattern\_traces}$ is the number of mapped traces that can match with the perspective incident trace pattern, given by $N_{relevant\_map\_traces\_MP\_stage\_of\_n}$

Hence, the whole metric for traceability in digital forensic investigation process is derived as follow:

Tracing process equation:

$$TC = \frac{N_{relevant\_traces}}{N_{total\_traces}} \tag{5}$$

TC is used as N relevant traces in $MP_{stage\_of\_i}$ for mapping process.

Mapping process Equation for stage $1 \cdots n$:

$$MP_{stage\_of\_i} = \frac{N_{relevant\_map\_traces}}{N_{relevant\_traces\_TC}} \tag{6}$$

$$MP_{stage\_of\_i+1} = \frac{N_{relevant\_map\_trace\_MP\_stage\_of\_i+1}}{N_{relevant\_traces\_TC} - N_{relevant\_map\_trace\_MP\_stage\_of\_i}} \tag{7}$$

Therefore, the generic mapping process equation:

$$MP = \frac{N_{relevant\_map\_traces}}{N_{relevant\_traces}} \tag{8}$$

In order to complete the traceability of forensic, the identification of the origin is determined by the offender identification equation:

$$OI = \frac{N_{relevant\_map\_traces\_MP\_stage\_of\_n}}{N_{relevant\_map\_traces\_MP} - N_{relevant\_map\_trace\_MP\_Stage\_of\_n}} \tag{9}$$

$$\infty = \frac{N_{tracepattern\_traces}}{N_{relevant\_map\_traces}} \tag{10}$$

Thus, a traceability index for digital forensic investigation process named as *Forensic Traceability Index* is defined as:

$$TraceIndex_{forensic} = \left(TC, \sum_{i=1}^{n} MP, OI \right) \tag{11}$$

where the output of TC is used as the input of MP and the result of MP is used to identify the *OI*.

This $TraceIndex_{forensic}$ is a useful indicator in measuring the accuracy and the completeness of evidence. The results obtained prove that the *TraceMap* is an effective model that supports the tracing and mapping evidence to overcome the traceability problem in digital forensic investigation process.

6. Conclusions

This paper introduces an approach to overcome traceability issues in digital forensic investigation process. The approach which consists of generic trace pattern, and tracing and mapping procedure was embedded in a model named as *TraceMap.* This *TraceMap* is used to provide an effective way to trace and map digital evi-

Table 2. Summary of offender identification rate (OI_R).

Dataset	Total Relevant and Mapped Traces $N_{relevant\_map\_traces}$	Total Trace Pattern Traces $N_{tracepattern\_traces}$	% of Offender Identification OI_R	Offender Identified		
				Victim	Attacker	Multi-step Attacker
DS1	7	7	100.00	4	1	2
DS2	7	7	100.00	3	0	4
DS3	8	8	100.00	5	0	3
DS4	4	4	100.00	2	1	1
DS5	4	4	100.00	1	1	2
DS6	4	4	100.00	2	1	1
DS7	4	4	100.00	1	1	2
DS8	8	8	100.00	2	1	5
DS9	7	7	100.00	2	1	4
DS10	8	8	100.00	4	1	3
DS11	8	8	100.00	5	1	2
DS12	8	8	100.00	5	0	3

dence in digital forensic investigation process specifically in collection and preservation phase. Thus, it facilitates the investigator in formulating the hypothesis, identifying, tracing, mapping and presenting the digital evidence.

In this *TraceMap*, the identification of offender is based on three primary events of incident which are *scan*, *exploit* and *impact*/*effect*. It could help the investigator in providing complete and accurate digital evidence to identify the origin of the incident.

The results obtained in the experiment demonstrates the *TraceMap* is able to discover the evidence, able to connect or map the evidence and able to identify the origin of the incident. These abilities are demonstrated through the result obtained using three metrics: *Tracing Rate* (TC_R) *Mapping Rate* (MP_R) and *Offender Identification Rate* (OI_R).

The *Tracing Rate* (TC_R) are used to demonstrate the ability of discovering the evidence, *Mapping Rate* (MP_R) is used to demonstrate the ability of mapping the evidence and *Offender Identification Rate* (OI_R) is used to demonstrate the ability of identifying the origin of the incident.

Based on these metrics, traceability index ($\text{TraceIndex}_{forensic}$) is derived that can be used in digital forensic investigation process to help the practitioner in measuring the accuracy and completeness of the evidence discovery. This research has a high potential of being expanded into other research areas such as in digital evidence presentation.

7. Acknowledgements

This research was kindly supported by Universiti Teknikal Malaysia Melaka and Ministry of Higher Education Malaysia.

REFERENCES

[1] E. Casey and G. L. Palmer, "Digital Evidence and Computer Crime," 2nd Edition, Elsevier Academic Press, Cambridge, 2004.

[2] D. Birk and C. Wegener, "Technical Issues of Forensic Investigations in Cloud Computing Environments," *Proceedings of the 6th International Workshop on Systematic Approaches to Digital Forensic Engineering*, 26-26 May 2011, Oakland, pp. 1-10.

[3] P. Stephenson, "A Comprehensive Approach to Digital Incident Investigation," *Information Security Technical Report*, Vol. 8, No. 2, 2003, pp. 42-54.

[4] G. Palmer, "A Road Map for Digital Forensic Research," Technical Report (DTR-T001-01) for Digital Forensic Research Workshop (DFRWS), New York, 2001.

[5] W. Kruse and J. Heiser, "Computer Forensics: Incident Response Essentials," Addison Wesley, Indianapolis, 2002.

[6] A. Brill and M. Pollitt, "The Evolution of Computer Forensic Best Practices: An Update on Programs and Publications," *Journal of Digital Forensic Practice*, Vol. 1, No. 1, 2006, pp. 3-11. doi:10.1080/15567280500541488

[7] M. Kohn, J. Eloff and M. Olivier, "Framework for a Digital Forenisc Investigation," *Proceedings of the Information Security South Africa* (*ISSA*) 2006 *from Insight to Foresight Conference*, Sandton, 5-7 July 2006, pp. 1-7.

[8] S. Satpathy, S. K. Pradhan and B. B. Ray, "A Digital

Investigation Tool based on Data Fusion in Management of Cyber Security Systems," *International Journal of Information Technology and Knowledge Management*, Vol. 2, No. 2, 2010, pp. 561-565.

[9] S. S. Rahayu, Y. Robiah and S. Shahrin, "Mapping Process of Digital Forensic Investigation Framework," *International Journal of Computer Science and Network Security*, Vol. 8, No. 10, 2008, pp. 163-169.

[10] V. Baryamureeba and F. Tushabe, "The Enhanced Digital Investigation Process Model," *Proceedings of the Digital Forensic Research Workshop* (*DFRWS*), 11-13 August 2004, Baltimore, pp. 1-9.

[11] B. Carrier and E. Spafford, "Getting Physical with the Digital Investigation Process," *International Journal of Digital Evidence*, Vol. 2, No. 2, 2003, pp. 1-21.

[12] S. Ó. Ciardhuáin, "An Extended Model of Cybercrime Investigations," *International Journal of Digital Evidence*, Vol. 3, No. 1, 2004, pp. 1-22.

[13] M. Roger, "DCSA: Applied Digital Crime Scene Analysis," Handbook of Information Security, New York, 2006.

[14] M. Reith, C. Carr and G. Gunsch, "An Examination of Digital Forensic Models," *International Journal of Digital Evidence*, Vol. 1, No. 3, 2002, pp. 1-12.

[15] N. L. Beebe and J. G. Clark, "A Hierarchical, Objectives-Based Framework for the Digital Investigations Process," *Proceedings of the Digital Forensic Research Workshop* (*DFRWS*), Baltimore, 11-13 August 2004, pp. 146-166.

[16] F. C. Freiling and B. Schwittay, "A Common Process Model for Incident Response and Computer Forensics," *Proceedings of the Conference on IT Incident Management and IT Forensics*, Stuttgart, 11-13 September 2007, pp. 1-21.

[17] S. Perumal, "Digital Forensic Model Based on Malaysian Investigation Process," *International Journal of Computer Science and Network Security*, Vol. 9, No. 8, 2009, pp. 38-44.

[18] S. Rekhis, J. Krichene and N. Boudriga, "Cognitive-Maps Based Investigation of Digital Security Incident," *Proceedings of the Third International Workshop on Systematic Approaches to Digital Forensic Engineering*, Oakland, 22 May 2008, pp. 25-40.

[19] S. L. Garfinkel, "Digital Forensics Research: The next 10 Years," *Journal of Digital Investigation*, Vol. 7, 2010, pp. S64-S73.

[20] T. Lindsey, "Challenges in Digital Forensics," The Digital Forensic Research Workshop (DFRWS), New York, 2006.

[21] K. Nance, B. Hay and M. Bishop, "Digital Forensics: Defining a Research Agenda," *Proceedings of the* 42*nd Hawaii International Conference on System Sciences*, Big Island, 5-8 January 2009, pp. 1-6.

[22] C. Shields, O. Frieder and M. Maloof, "A System for the Proactive, Continuous and Efficient Collection of Digital Forensic Evidence," *Journal of Digital Investigation*, Vol. 8, 2011, pp. S3-S13. doi:10.1016/j.diin.2011.05.002

[23] A. Ahmad, "The Forensic Chain-of-Evidence Model: Improving the Process of Evidence Collection in Incident Handling Procedures," *Proceedings of the* 6*th Asia Conference on Information Systems* (*PACIS* 2002), 2-4 September 2002, Tokyo pp. 1-5.

[24] V. H. Bhat, P. G. Rao, R. V. Abhilash, P. D. Shenoy, K. R. Venugopal and L. M. Patnaik, "A Data Mining Approach for Data Generation and Analysis for Digital Forensic Application," *IACSIT International Journal of Engineering and Technology*, Vol. 2, No. 3, 2010, pp. 314-319.

[25] G. Giova, "Improving Chain of Custody in Forensic Investigation of Electronic Digital Systems," *International Journal of Computer Science and Network Security*, Vol. 11, No. 1, 2011, pp. 1-9.

[26] J. Herrerias and R. Gomez, "A Log Correlation Model to Support the Evidence Search Process in a Forensic Investigation," *Proceedings of the* 2*nd International Workshop on Systematic Approaches to Digital Forensic Engineering* (*SADFE*'07), Bell Harbor, 10-12 April 2007, pp. 31-42. doi:10.1109/SADFE.2007.1

[27] P. Sommer, "Intrusion Detection Systems as Evidence," *Computer Networks*, Vol. 31, No. 23, 1999, pp. 2477-2487. doi:10.1016/S1389-1286(99)00113-9

[28] P. Oghazi, B. Palsson and K. Tano, "An Attempt to Apply Traceability to Grinding Circuits," *Proceedings of the Conference in Mineral Processing*, Lulea, 6-7 February 2007, pp. 169-183.

[29] R. Clayton, "Anonymity and Traceability in Cyberspace," Ph.D. Thesis, University of Cambridge, Cambridge, 2005.

[30] E. Golan, B. Krissoff, F. Kuchler, L. Calvin, K. Nelson and G. Price, "Traceability in the U.S. Food Supply: Economic Theory and Industry Studies," Department of Agriculture, Washington DC, 2004.

[31] G. Zemont, "Towards Value-Based Requirements Traceability," Master Thesis, DePaul University, Chicago, 2005.

[32] M. Narmanli, "A Business Rule Approach to Requirements Traceability," Master Thesis, Middle East Technical University, Ankara, 2010.

[33] Morckos, M. "Requirements Traceability," Report for School of Computer Science, University of Waterloo, Waterloo, 2011.

[34] L. Westfall, "Bidirectional Requirements Traceability," White Paper, The Westfall Team, Dallas, 2006.

[35] S. R. Selamat, R. Yusof, S. Sahib, M. F. Abdollah, M. Z. Mas'ud and I. Roslan, "Adapting Traceability in Digital Forensic Investigation Process," *Proceedings of the Malaysian Technical Universities International Conference on Engineering and Technology* (*MUiCET* 2011), Johor, 13-15 November 2011, pp. 1-8.

[36] C. Rahmani, M. Sharifi and T. Tafazzoli, "An Exprimental Analysis of Proactive Detection of Distributed Denial of Service Attacks," *Proceedings of the IIT Kanpur Hacker's Workshop* (*IITKHACK*04), Kanpur, 23-24 February 2004, pp. 37-44.

[37] S. R. Selamat, R. Yusof, S. Sahib, M. F. Abdollah, M. Z. Masud and I. Roslan, "Tracing Technique for Blaster Attack," *International Journal of Computer Science and Information Security*, Vol. 4, No. 1, 2009, pp. 1-8.

[38] S. R. Selamat, R. Yusof, S. Sahib, M. Z. Masud, I. Roslan and M. F. Abdollah, "Scenario Based Worm Trace Pattern Identification Technique," *International Journal of Computer Science and Information Security*, Vol. 7, No. 1, 2010, pp. 1-9.

[39] S. R. Selamat, R. Yusof, S. Sahib, M. Z. Masud, M. F. Abdollah and Z. Z. Abidin, "Advanced Trace Pattern for Computer Intrusion Discovery," *Journal of Computing*, Vol. 2, No. 6, 2010, pp. 200-207.

[40] G. Hoglund and G. McGraw, "Exploiting Software: How to Break Code," Addison-Wesley/Pearson, Indianapolis, 2004.

[41] A. Moore, R. Ellison and R. Linger, "Attack Modeling for Information Security and Survability. Technical Note (CMU/SEI-2001-TN-001) for Software," Carnegie Mellon University, Pittsburgh, 2001.

[42] B. Sean and S. Amit, "Introduction to Attack Patterns," 2006. https://buildsecurityin.us-cert.gov/

[43] Fernandez, E., Pelaez, J. and M. Larrondo-Petrie, "Attack Patterns: A New Forensic and Design Tool," *Advances in Digital Forensics III*, *Proceeding of Third Annual IFIP WG* 11.9 *International Conference of Digital Forensics*, 28-31 January 2007, Cozumel, pp. 345-357. doi:10.1007/978-0-387-73742-3_24

[44] K. Kent, S. Chevalier, T. Grance and H. Dang, "Guide to Integrating Forensic Techniques into Incident Response," National Institute of Standards and Technology (NIST), Gaithersburg, 2006,

[45] R. Yusof , S. R. Selamat, S. Sahib, M. F. Abdollah, M. Z. Masud and M. Ramly, "An Improved Traditional Worm Attack Pattern," *Proceedings of the* 4*th International Symposium on Information Technology* 2010 (*ITSIM* 2010), Kuala Lumpur,17 June 2010, pp. 1067-1072.

[46] R. Yusof, S. R. Selamat, S. Sahib, M. F. Abdollah, M. Z. Mas'ud and M. Ramly, "A New Malware Attack Pattern Generalization," *Proceedings of the Malaysian Technical Universities International Conference on Engineering and Technology* (*MUiCET* 2011), Johor, 13-15 November 2011, pp. 20-29.

[47] J. Velasco, "A Guide to Electronic Evidence Collection Methodologies," White Paper, RenewData Corporation, Austin, 2007.

[48] A. Hassanzadeh and B. Sadeghiyan, "A Data Correlation Method for Anomaly Detection Systems using Regression Relations," *Proceedings of the* 1*st International Conference on Future Information Networks*, Beijing, 14-17 October 2009, pp. 242-248.

[49] S. R. Selamat, R. Yusof, S. Sahib, N. H. Hassan, M. Z. Mas'ud, and Z. Z. Abidin, "Traceability in Digital Forensic Investigation Process," *Proceedings of the IEEE Conference on Open Systems*, Langkawi, 25-28 September 2011, pp. 101-106.

[50] F. Cohen, "Metrics for Digital Forensics," *Proceedings of the MiniMetriCon Conference*, 14 February 2011, San Francisco, pp. 1-22.

[51] T. Holz, "Security Measurements and Metrics for Networks," *Dependability Metrics*: *Advanced Lectures Notes in Computer Science* (*LNCS*), Vol. 4909, 2008, pp. 157-165. doi:10.1007/978-3-540-68947-8_13

[52] A. Al-Dallal and R. S. Abdulwahab, "Achieving High Recall and Precision with HTLM Documents: An Innovation Approach in Information Retrieval," *Proceedings of the World Congress on Engineering* (*WCE* 2011), London, 6-8 July 2011, pp. 1-6.

[53] N. L. Beebe and J. G. Clark, "Digital Forensic Text String Searching: Improving Information Retrieval Effectiveness by Thematically Clustering Search Results," *Journal of Digital Investigation*, Vol. 4, 2007, pp. S49-S54. doi:10.1016/j.diin.2007.06.005

[54] G. Peterson, S. Shenoi and N. Beebe, "Digital Forensic Research: The Good, the Bad and the Unaddressed," *Advances in Digital Forensics V*, Vol. 306, 2009, pp. 17-36. doi:10.1007/978-3-642-04155-6

[55] G. Gu, P. Fogla, D. Dagon, W. Lee and B. Skoric, "Towards an Information-theoretic Framework for Analyzing Intrusion Detection Systems," *Proceedings of the* 11*st European Symposium on Research in Computer Security* (*ESORICS*'06), Hamburg, 18-20 September 2006, pp. 1-20.

