<table>
<thead>
<tr>
<th>Title</th>
<th>Network Forensics Readiness and Security Awareness Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors(s)</td>
<td>Al-Mahrouqi, Aadil; Abdalla, Sameh; Kechadi, Tahar</td>
</tr>
<tr>
<td>Publication date</td>
<td>2014-10-29</td>
</tr>
<tr>
<td>Link to online version</td>
<td><a href="http://www.univ-annaba.dz/relations-exterieures/manifestations-scientifiques/internationales/item/131-icesti-14.html">http://www.univ-annaba.dz/relations-exterieures/manifestations-scientifiques/internationales/item/131-icesti-14.html</a></td>
</tr>
<tr>
<td>Item record/more information</td>
<td><a href="http://hdl.handle.net/10197/6498">http://hdl.handle.net/10197/6498</a></td>
</tr>
</tbody>
</table>
Abstract—The goal of reaching a high level of security in wireless and wired communication networks is continuously proven difficult to achieve. The speed at which both keepers and violators of secure networks is evolving is relatively close. Nowadays network infrastructures contain a large number of event logs captured by Firewalls and Domain Controllers (DCs). However, these logs are increasingly becoming an obstacle for network administrators in analyzing networks for malicious activities. Forensic investigators mission to detect malicious activities and reconstruct incident scenarios is very complex considering the number as well as the quality of these event logs. In this paper, we present the building blocks of a framework for automated network readiness and awareness. The idea of this framework is to utilize the current network security outputs to construct forensically comprehensive evidence. In the proposed framework, we cover the three vital phases of the cybercrime management chain, which are: 1) Forensics Readiness, 2) Active Forensics, and 3) Forensics Awareness. Keywords: Network Forensics, Forensics Readiness, Network Security, Active Forensics, Reactive Forensics, Forensics Awareness and Network Security Framework.

I. INTRODUCTION

The cybercrime landscape has increased dramatically with the use of more sophisticated techniques and greater knowledge of cybercrime. There are many challenges faced by today’s digital forensics. The lack of both funding and qualified professionals as well as cross-jurisdictional legal struggles are just a sample of the main body of issues [1]. The first Digital Forensics Research Workshop (DFRWS) [15] was held in Utica, N.Y., in 2001. DFRWS provided the first proper framework and presented guidelines for conducting a technical digital investigation.

Since that first workshop, many scholars have worked to make digital evidence easier to demonstrate by establishing many types of graphs in order to represent evidence and attack scenarios. The scholars utilized a mathematical formula and algorithms to construct these graphs to recognize the patterns of the attack [9]. Unfortunately, most of these graphs provide a high-level, abstract view of the complex attack [8]. Examples of investigation graphs consist primarily of scenario graphs, forensic graphs, logic exploitation graphs, attack graphs, and evidence graphs [12]. The digital systems can be described mathematically as a finite state machine and we can represent this information in the form of a graph (nodes and arrows) [7].

This work will take into consideration important factors during forensics investigation, for example, cost, time, low incident impacts, facilities network investigation procedures, high quality outcomes, organization reputation and business activities disruption. Furthermore, the aim to propose attack and evidences integration graph is to increase the efficiency of our investigation results. In addition, the data flow in the proposed framework is designed based on network OSI model.

In other hand, this paper will present forensics awareness model and designed to generate a best practices for system administrations and forensics investigators about the learning security vulnerabilities from previous happened cases in the network infrastructure as well as different sources.

The remaining part of this paper is organized as follows. Section 2 outlines previous work. Section 3 describe the proposal framework. Section 4 establish a case study with the aims to give an idea of how to create criminal graph. Finally, section 5 deals with the conclusion and some perspective as future works.

II. RELATED WORK

Wang et al. [13], in their proposed evidence graph model, seek to facilitate the presentation and manipulation of intrusion evidence. This model aims to reduce the redundancy in firewall output intrusion alerts. The proposed architecture facilitates the evidence presentation process and provides automated intrusion evidence analysis. The evidence module is considered the most important module in the Wang et al. proposed architecture because it plays an important role in analysis visualization of capture evidence.

Later, Wang et al. [14] proposed diffusion and graph spectral methods. These proposed methods aimed to establish a systematic forensics investigation process framework. Moreover, through these proposals, Wang et al. attempted to provide high-performance computation methods to be used in the forensics analysis field as a form of well-utilized mathematical science.

In 2004, Gladyshev et al. [6] proposed a formalized approach for Event Reconstruction. This approach was based on the terms of the finite state machine model of computation. The finite state machine model was used to define all possible attack scenarios in the computer network incidents. Furthermore, Gladyshev defined Event Reconstruction as a process of finding all potential computations of the machine that agree with the digital evidence of the incident. The scholar proposed an algorithm for the Event Reconstruction process that consists of three phases. The first phase calls for obtaining the finite state model of the computer system that is under the forensics
investigation. In the second phase, all potential attack scenarios of the computer system incidents are defined by using the back trace method from the point in which the cybercriminal was discovered. The third phase calls for rejecting attack scenarios that conflict with the obtainable evidence [3].

Liu et al. [4] proposed merging sub-evidence graphs with an integrated evidence graph for network forensics analysis. This paper show how to integrate different evidence graphs with or without the help of a corresponding attack graph. The proposal model assumes that an integrated evidence graph shows all attacks using global reasoning. Consequently, the research provided two algorithms that help integrate evidence graphs with a probabilistic evidence graph. Swiler et al. [11] proposed an approach for network risk analysis based on an attack graph that defines the set of attack paths that have a high probability of success for the attacker. This approach requires a predefined data-set as input information before starting to use the system. As a result, the system will generate an attack graph based on predefined information.

Sheyner et al. [10] proposed automated techniques in order to establish and generate the attack graphs. The techniques are based on a set of algorithms that are used to reconstruct attack scenarios automatically. After that, the reconstructed attack scenario is represented in the attack graphs. The visual representation of attack graphs allows forensics investigators to easily understand the attack scenario in an efficient manner. The authors implemented a network forensics tool based on the proposed algorithms, testing it in a small Local Area Network (LAN) that consists of an intrusion detection system and firewalls.

Bruaschi et al. [2] proposed a model that can be organize digital forensics knowledge in reusable way. In other word, this model can reuse the gathering techniques and some hypotheses in order to find the best guideline for hypotheses formulation.

III. NETWORK FORENSICS READINESS AND SECURITY AWARENESS FRAMEWORK

In figure 2 we show the overall view of the proposed network forensic readiness and security awareness framework.

**Logs classification processes submodel** Basically, the operating system in the network firewalls and domain controllers (DCs) are able to classify the computer network and system events logs into predefined groups. This model was designed to increase the filtering process of the output events logs. It will classify the output logs into different groups, namely alerts and information.

**Alert logs collection model** This model is designed to collect only the alert logs. These logs will be stored in the alert logs warehouse.

**Alerts prepossessing model** The stored alert logs contain redundancy data and irrelevant information [11]. The alerts prepossessing model is used to filter out all redundancy data and irrelevant information from the alert logs. The alerts prepossessing model has two stages: format standardization and redundancy management. The format standardization process aims to convert the different event logs formats into one unified, common syntax format while the redundancy management process aims to reduce the duplication of the single event.

**Assets Knowledge warehouse** Assets knowledge warehouse designed to store basics information of all assets available in the network infrastructures.

**Attack knowledge warehouse** The assets knowledge warehouse is designed to store basic information of all assets available in the network infrastructures.

**Attack path Retrieval** An attack graph provides a visual representation of the attack paths as well as evidence for each node (host) in each path (see figure 4). The attack paths describe all exploited network assets. The attack graphs will be generated based on databases, namely asset knowledge and attack knowledge. The nodes indicate the exploit hosts while the edges indicate the security vulnerability used to hack the host. The information shown in this graph is based on a chain of custody manner.

**Scenario reconstruction submodel** After generating the attack and evidence graphs, this model is used to reconstruct the attack scenario. This process will reprocess the criminal
graph with the help of criminology sciences and hypothesis expert knowledge.

A. Information prepossessing model

**Information collection sub model** The information collection submodel will collect all output of information logs from event log classification processes and forward it to the information logs warehouse.

**Data mining Engine** Because there are so many information logs in the information logs warehouse, it is very difficult to check all of them and update information security awareness. This step is used to convert information logs into an easier format that will be useful for security information awareness.

The data mining engine consists of two types of processes: short classification types and host characteristics associations. First, the host classification process will be used to classify all existing assets in the network infrastructure into certain groups based on host types, such as router, switches, domains controllers, firewalls, etc. Second, host characteristics associations will be used to associate each log to the appropriate predefined group. Using the association process, it will be necessary to analyze the logs header format to be able to know the appropriate predefined group.

**Calculating attack probability** The calculating attack probability process will be used to process the output results of the attack decision tree. This process examines the attack probability for each asset (for example, the file server probability affected by DOS) based on previous experiments knowledge through the data-set analysis.

**Awareness DB** The awareness DB is used to store the attack probability for each asset in the network infrastructure. This database feeds the internal awareness Web page through security awareness and vulnerabilities for network assets.

B. The normalization process of alerts and information logs

As mentioned earlier, the normalization process will be used to convert the event logs formats into a unified format. This process will help to aggregate the logs and reduce redundancy and noise information. Table I I shows a proposed unified structure field of event logs.

C. The relationship between the evidence and attack

It is very important to know the relationship between the evidence and attacks. This relationship helps us in the investigation process as well as increases the admissibility of the investigation case in court. Moreover, the increased amount of evidence related to a specific attack case will increase the background information about the attacker. There are different types of relationships between the detected evidence and the attacks, including one-to-one, one-to-many, and many-to-one.

IV. Example of Criminal Graph

In this paper, we aim to present a picture of the proposed graph that integrates the attack and evidence graphs. Section IV-A establishes a criminal scenario.

A. Example Attack Scenario

A university network has several equipment records of different types of event logs from different network sources. Bob is a student and he hacked a file server (see figure 3). Moreover, several types of network devices in the university infrastructure used by Bob and the victim server records Bob’s activities in different formats and styles, depending on the operating system in each device.

In figure 3 shows the attack scenario for the university network. The red color line indicates the trace path of Bobs steps to reach the target server. Figure 4 shows the integration between the attack graph and the evidence graph. Moreover,
nodes indicate compromised hosts while edges refer to security vulnerabilities used by the attacker. Moreover, under each compromised host, there is another graph that shows the series of actions carried out by the forensics investigators. Furthermore, the actions carried out by the forensics investigation are linked to another graph called the evidence graph. This graph will show the output evidence as a result of each forensics investigative action.

V. CONCLUSIONS AND FUTURE WORK

In summary, we proposed the forensics readiness and awareness framework. This framework contains around fifteen different models. The proposed models work as a single unit in order to process and normalize the captured network logs. The main point of designing the framework is to find a way to forensically visualize the evidence and attack scenario in a computer system. Moreover, this paper listed some methods and approaches proposed by scholars to construct the attack scenario. We demonstrated that the graph representation is one of the best approaches and that scholars have proposed several types of graphs, including scenario graphs, logic exploitation graphs, forensics graphs, attack graphs, and evidence graphs.

Since the attempt to reconstruct scenarios of network attacks from collected data (i.e., alarms, alerts and logs) requires brain-like reasoning to understand these events. Therefore, Bio-inspired approaches [5] to self-organizing network events and creating the linkage between them are of relevance to our studies. In our future work we will examine the possibility to replace the traditional database approach to storing events with a bio-inspired mechanisms and, we will study the affect of that on the quality of the scenarios produced.

This framework acts as a first step toward network logs analysis. For future work, we will focus on involving mathematic and algorithm science for each proposed model in the framework to help us validate the frameworks. Furthermore, we will try to utilize criminology science to enhance any future proposed models or approaches.

REFERENCES