Design for A Network Centric Enterprise Forensic System

Hongye Zhong & Jitian Xiao

Abstract
Increased profitability and exposure of enterprise’s information incite more attackers to attempt exploitation on enterprise network, while striving not to leave any evidences. Although the area of digital forensic analysis is evolving to become more mature in the modern criminology, the scope of network and computer forensics in the large-scale commercial environment is still vague. The conventional forensic techniques, consisting of large proportion of manual operations and isolated processes, are not adequately compatible in modern enterprise context. Data volume of enterprise is usually overwhelming and the interference to business operation during the investigation is unwelcomed. To evidence and monitor these increasing and evolving cyber offences and criminals, forensic investigators are calling for more comprehensive forensic methodology. For comprehension of current insufficiencies, this paper starts from the probes for the weaknesses of various preliminary forensic techniques. Then it proposes an approach to design an enhanced forensic system that integrates the network distributed system concept and information fusion theory as a remedy to the drawbacks of existing forensic techniques.

Keywords: Network, Forensic, Information Security, Enterprise.

1. INTRODUCTION
With the evolution of networking technology and mobile computing, portable devices and communication vehicles such as mobile phones, laptops email and social networks are pervasively participated in our daily lives and production environment. When individuals and enterprises happen to encounter legal or corporate issues, evidences are required to be collected from relevant electronic devices to support legal or business decisions [1]. Moreover, auditing and examining digital trails are usually enabled on computer devices to discover or assure whether the information is secure or has been tempted. Apart from these scenarios, various cases entail computer forensics including collecting reliable digital evidences for a law case in a court of justice. The increasing requirements and complexity of investigation becomes a spur for the study and application of computer forensic science.

The primary purpose of computer forensics is to dig up data to expose or assure what and when something has been done, and by whom, where, why and how. With the improvement of forensic theory and the accumulation of forensic practices, various forensic process, workflow and techniques have been introduced to ensure this fundamental purpose of computer forensics can be accomplished, while the essential procedures of forensic process almost remain unchanged.

Forensic process is a mechanism that uses scientific methods to discover digital evidences. The evidences found will be utilized to support or disprove a hypothesis or reveal or verify works done by others. Forensic process generally is consisted of the five linear procedures in terms of Plan,
Acquire, Extract, Analyze, and Report [2]. In practice, the forensic process can be organized and interpreted into a Model of Digital Forensic Analysis (MDFA, Fig. 1), which eases the implementation of the process and enhances its controllability in a clear and systematic manner.

![Model of Digital Forensic Analysis](image)

**FIGURE 1:** Model of Digital Forensic Analysis.

2. PRELIMINARY FORENSIC TECHNIQUES

Forensic techniques are the means employed by the investigators with the goal to discover and retrieve the evidences stored in devices. Although the investigation methods might vary from case to case, general methods to achieve the goal of digital forensics could be summarized in two categories: Locating Data and Capturing Data [3].

Locating Data is the process of discovering sensible or relevant data from the storage component of devices, which usually involves investigations on components, such as file systems and disk cluster, memory and process image, and history and temporary files. Capturing Data is a more active way for digital investigation. Instead of searching evidences left on the devices, it creates evidences. When the investigator estimates that a typical device might contain data closely related to an incident, he/she might covertly implant some monitoring mechanism onto the targeted device/s to intercept the data processed on the device/s, with the hope to attain useful information that can become digital evidences. The techniques usually used in Capturing Data include keystroke logger, wiretapping, and spyware [3].

In recent years, large number of electronic devices has been adopted by enterprise, both in its operation and production. Meanwhile, more vulnerabilities and security issues have also been introduced, which increases the occurrence of crimes and internal incidents related to information security. These security issues sometimes come from the negligence or fraudulence of its staff, or the attacks from outsiders. Most of the cases require investigations on the electronic devices to collect evidences for making decisions. In some other cases, when the enterprise is under some security agreements, digital investigations are also needed to assure the enterprise’s compliance. For example, if an enterprise has adopted the ISO 27001 Security Standard, evaluation needs to be taken annually to assure the enterprise has been obeying the codes in the standard.
Investigators usually find the difficulty is severely increased in collecting evidences with the conventional forensic techniques in enterprise context, rather than in civil cases. In the modern enterprise environment, especially in large organizations, the amount of data and information is usually overwhelming. In some cases, the data is even geographical distributed [1]. Frequently, the suspected users who are targeting the enterprise’s sensitive information are usually trained and skilled. In contrast, the inadequate computer training of employees may damage the evidence before the arrival of the investigator/s. The difficulty might be increased, if the enterprise might not want the intervention to production while the data is being collected for investigation. All of these hurdles arouse the need to design a more sophisticated and centralized process for computer forensics in enterprise.

3. ENHANCEMENT ON FORENSIC PROCESS

The logical and geographical scattering of data and devices increases the difficulty of evidence collection and hinders the investigation. As most enterprises have different running patterns, we need to consider a strategy that can fit most situations. For digital investigation, the procedures of forensic process are basically mandatory for all investigations. We should break down the forensic process to several operational tasks so as to be more easily adapted them into the information systems of different enterprises [4].

3.1 Enhanced Forensic Workflow

In accordance with enterprise forensic convention, MDFA can be interpreted into more executable and measurable steps. These steps are shown and linked up in the Enhanced Forensic Workflow (EFW, Fig. 2), which generally consists of the following stages.

Collection Planning - targeting device, execution time and search strategy should be determined for the data collection process.

Physical Media Identification - the targeted device are identified and labelled, so that they can be easily located and identified later.

Media Identifier Creation - the media is tagged with date/time and the disk images are acquired for future reference and research purpose.

**FIGURE 2: Enhanced Forensic Workflow.**
Digital Media Photograph - the photo of the device should be recorded and kept in digital form, which helps to identify whether the device has been physically altered. It also helps to easier identify the device.

Forensic Media Capture - the file contained in the disk images acquired above will be extracted and properly tagged.

File Package Copy - captured disk images, metadata information and extracted files will be transferred to a server computer where analysis will be performed.

Forensic Analysis - captured disk images, metadata information and extracted files will be resorted, indexed and organized into an understandable hierarchy, so that the relationship between the evidences can be easily identified.

Archival Analysis - analysis will be performed by searching the collected information and file with selected searching heuristics. The analysis is aimed to identify any breaches to the enterprise security regulations.

Finding Aid Creation - the analysis results will be sorted and linked with original data. Hints and evidences discovered should be annotated so that the evidences are more comprehensible.

Digital Repository Preservation - the original data and findings are preserved and indexed in an isolated location where the evidences cannot be contaminated. These findings are likely to be referred in reporting and assist decision makings.

Forensic Report Composition - by drawing links between original data and the findings and explaining the relationships between them and what they are leading to, an investigation report will be composed to give an overall picture of the investigation.

According to the workflow, the initial tasks involve acquisition of the information from all other information systems in the enterprise. The collected information is assembled into a database for further analysis. In most cases, analysis and report will necessarily be performed based on the data stored in the evidence database. The forensic tasks are planned to be executed periodically, so that it can cover all incidents and detect them in a timely manner. This forensic methodology ensures the efficiency and adequate coverage of the information collection. To ensure the recorded evidences can reflect the most detailed and honest facts about the incidents, the tasks of the workflow are expected to be executed adequately in investigations.

3.2 Automation of the Forensic Workflow

With the objective of reducing the process mistakes and shortening the process duration, the investigation should be executed in a systematic and automatic manner. Since the forensic process has been broken down to smaller executable units of tasks, the automation becomes easier to implement. In the workflow, certain manual processes are unavoidable such as labelling or photographing the device. To maximize the automation processing of the workflow, such processes are required to be handled beforehand. Due to the nature of the processes, the original forensic workflow can be divided into two corresponding workflows: Preliminary Forensic Workflow (PWF), and Automated Forensic Workflow (AWF).

The PFW (Fig. 3) is consisted of a group of processes to be performed to ensure that the devices of the enterprise are prepared for automated data collection. This workflow needs to be executed once for each device only when the device is procured into the enterprise.
The main purpose of PFW is to register the physical and logical identity information of the device for future referencing purpose. This kind of processes have to be done manually. For asset management purposes, normally enterprise has already registered every new device when they were procured. Hence it will be convenient to simply embed the original device register procedure into our workflow. For such purpose, whenever a new device is procured, the following steps must be followed.

Label Device - when a new device is procured by the enterprise, a unique identity number will be generated and labeled on to the device. This identity number is going to be the key to uniquely distinguish between devices.

Register Device - after the device is labelled, its relevant information will be recorded into the forensic system. The register information will be stored in the database of the forensic system along with the forensic information collected in future, and indexed by its unique identity number.

Physical Scan - after the register information of the device has been created, the 3D model of device should be scanned and saved as a part of the register information. Unless the device suffers physical damage, this property of the device does not need to be modified.

Logical Scan - this step is to make an initial record of the disk image of the imported device. In future, disk image might be acquired multiple times for forensic purposes. For example, the record of the disk image will be acquired periodically for comparison analysis. The previous image data will not be removed. The serials of disk image records will be maintained in the forensic database and kept in a hierarchy pattern, and labelled with time stamps. As the result, the image records can be compared and referenced for forensic analysis.

Implant Forensic Agent - in order to transmit forensic data in response to the request of forensic server, a client service needs to be installed and kept running on the device. This service will act as the coordinator between the operating system of the device and the forensic server to perform the task in correspondence with the server requests.

The AFW (Fig. 4) is a set of processes that collect and analyze forensic data from the device, and generates a report to illustrate the findings. AFW assumes the newly procured devices have been properly handled in the preliminary workflow, so that all devices should have been kept
connected with the forensic server via the Forensic Agent implanted in the devices. Through a remote request, forensic server can instruct the devices to carry out the forensic workflow periodically.

The preliminary condition of the automatic workflow is that all devices of the enterprise have been labelled and registered, and the original manual processes of labelling and photographing devices has been removed. AFW continues the incomplete investigations of PFW by working on the recorded devices and corresponding information. Such preliminary condition eliminates the necessity to identify and to label the device again when the forensic investigation is performed on the devices. In most cases, the digital evidences are contained in the logical components of the devices when incidents happen. Unless the incidents involve physical damages to the devices, physical scan on device will not be necessary. Accordingly, the tasks for each stage of the forensic workflow are slightly different, depending on whether there are needs for physical scanning etc.

Collection Planning - in the planning stage, the forensic server schedules and configures the data collection, and dispatches the forensic requests to forensic agents of all devices of the enterprise.

Acquire Information - unlike the original forensic workflow, the Acquire stage is not going to take the whole dump of disk image every time. According the needs and nature of the investigation, the Forensic Agent will selectively acquire the disk image, partial dump, or the log files etc.

Transfer Data - when the Forensic Agent finishes acquiring relevant data, the device will transfer the collected data to the forensic server.

Forensic Analysis - when the forensic server receives the collected data from the devices, it will extract data and form files from the raw data. Then the extracted files will be sorted and indexed in a more sensible and analyzable form.

Comprehensive Analysis - the sorted files will be further analyzed with the customized analysis strategy and the predefined searching heuristics, in order to discover any hints or evidences from the collected information.

Annotate Findings - analysis results from the previous stages will be sorted and indexed, and the reasoning and other interesting information related to the analysis results will also be annotated.

Preservation in Repository - the collected information and the corresponding findings will be sorted and stored in the database of the forensic server.

Compose Report - after finishing the above steps, a summary about the current investigation will be generated. In the summary, the brief of collected information, findings and the investigation process itself will be documented. The summaries will be saved to the forensic database, which can be exported and printed in a more representable form when they are needed to be referenced in future.

4. DESIGN OF NETWORK CENTRIC FORENSIC SYSTEM
The automated forensic workflow introduced above can be implemented in an enterprise as a comprehensive system by combining the existing enterprise networks with modern intelligent computing approaches. This system is called Network Centric Enterprise Forensic System (NCEFS). NCEFS is consisted of two parts: the client side and the server side applications. The client side application is named Forensic Agent (FA), and the forensic server is named Centralized Forensic Processor (CFP).
4.1 Forensic Agent
The FA is a client side application installed on devices of the enterprise. It acts as a coordinator that communicates between CFP and the devices (see Fig. 5). Once FA is installed on a device, it installs itself as a daemon service that starts running once OS starts. Once FA starts, it will establish a TCP connection with the CFP. Since TCP is a stateful network protocol, the connection is always being listened by the server, by which CFP can monitor the running status of the devices and transmit control requests. As FA needs to execute some sensitive processes such as disk imaging, memory imaging and data transmission, special execution privileges should be granted to FA. When FA receives forensic request from CFP, it will interpret the request and ask the hosting OS to complete the request. The requests usually contain various forensic tasks such as collecting certain information about the device or transferring the collected data to CFP.

With FA, the conventional device information such as disk image and files is able to be collected remotely as long as the device is remained connected with CFP. Moreover, as many modern computer devices have been equipped with some advanced input components such as light sensor and camera, they can be used to capture extra evidences remotely when such evidences are needed in certain investigations. Through the operation of FA, the devices are also remotely controllable by CFP. For example, when an incident happens, CFP can lock up the system of a specific device to protect the evidences contained on the device from being contaminated.

4.2 Centralized Forensic Processor
The Centralized Forensic Processor (CFP, Fig. 6) is a dedicated server that automatically organizes and controls the processing of digital forensics on the devices of the enterprise. CFP is closely corresponding with the forensic workflow. The role of CFP is to guarantee the steps in forensic workflow to be executed in a timely manner. As the workflow is performed automatically, the collected evidences and generated reports can be stored and managed more appropriately with labelling and sorting processes, so that the retrieved information are analyzable and can be easily referenced when they are needed in the future. CFP is mainly consisted of five components: Planner, Collector, Storer, Analyzer, and Reporter.
CFP Planner - it is the component that systematically schedules and organizes information collecting in accordance with the devices status, network situation, and investigation requirements. As CFP is able to query the status of the devices, Planner can schedule the devices to transfer the collected information when the devices are idle, with the purpose of avoiding interference to the normal usage of the devices. As there might be considerable number of devices connected with CFP, to avoid network conjunction, the Planner attempts to request the devices to transfer data when the network traffic is idle. The devices under investigation should not be required to upload full dump of disk image every time the information is collected. Instead, Planner considers the natures and aims of the investigations, and requests the device to collect only necessary data accordingly. For instance, for investigating whether there are devices in the enterprise breached the Application Installation Agreement of Apple Inc., CFP only needs to request all the iOS devices to submit the lists of their installed applications. By doing so, the side effects of forensic investigation to the enterprise normal operation can be greatly reduced.

CFP Collector - the information collection schedules made by CFP Planner will be passed to Collector to be executed. When Collector receives an information collection request, it will interpret and dispatch the request to FA of the targeted device via the TCP tunnel established. When the collection request has been completed, or failed, FA will inform Collector to incept the collected information.

CFP Storer - it is functionally an internal database that stores the collected data, discovered findings and summarized reports. The information stored in the database should be indexed with the unique identity number of the devices, so that they can be rapidly referenced and queried. The database should be encrypted and the creation time and modification of all the data should be time stamped and recorded to ensure the collected evidences will not be contaminated. Moreover, for security purpose, the database is only accessible to the other components of CFP, and it must not be visited from the outside of the server.

CFP Analyzer - it is an intelligent facility that can refine the raw data into more understandable form and conduct customized searching and heuristic searching on the collected information to discovery the hints and patterns hidden in the data. After analysis, the findings will be stored into the database and forwarded to Reporter to generate a summary about the current investigation.

Reporter - all the collected information, discovered findings, and the investigation itself such as process duration, involved devices etc. will be assembled and summarized into a brief and comprehensible conclusion. The generated summaries will then be saved in Storer for future references and analysis. The saved reports can be exported from the database and print out into a representable form.

4.3 Collection and Analysis Strategy

Generally, there might be hundreds or even thousands of computer devices in a signal enterprise (as Fig. 7). The CFP Collector needs to simultaneously perform the information collections for many agents. This causes a multi-agent planning problem, in which the task performed to combine information from several sources, called Information Integration [5]. When multiple
agents working in the environment to collect information, it raises a need to ensure each agent can achieve its own goal with the help or hindrance of others [6]. When communication constrains exist, although the planning phase is a centralized process, the execution phase may need to be decentralized or at least partial decoupled. In such a case, explicit communicative instructions are needed for organizing the operations of multiple agents to achieve the goal.

For the security purpose that the evidence of an agent should not be contaminated by another, the communication requirement between the agents are prohibited. In such a case, the communication between CFP and the agents form a Star network topology. However, as the enterprise will recruit new computer devices from time to time, changing of the edges causes the topology to be dynamic. The edge set of the network is time varying in which edges may disappear and reappear in accordance with the changing state of network agents [7]. Such situation can be represented as:

\[ V = \{V_1, V_2, ..., V_n\} \]

In the representation, \( V \) denotes vertex set in underlying context and the set is consisted of \( n \) elements. An analysis strategy is needed to ensure the information collecting can go through without interruption when new elements join the set. Lyapunov theory is an intuitive framework for the analyzing asymptotic properties of dynamical systems, which provides a viable solution to such kind of problems [7]. The theory treats the system as a graph. When the edges of the system change, it rebuilds the graph without interrupting analysis of the graph. The Edge Agreement Protocol proves that a connected graph with changed edges can steer the edge states to the origin. Such graph has the following relationship:

\[ X_e(t) = -L_e(G)X_e(t) \]

In above representation, \( G \) denotes a graph with \( n \) nodes and \( m \) edges, \( X_e(t) \in R^m \) represents the internode states, and \( L_e(G) \) is a lieu of the vertex-to-edge transformation induced by the incidence matrix of \( G \). As our system uses a design of Star topology, we do not need to worry about that circles in the graph will prevent the agreement state to be reached [7]. The integrated state information for all the network nodes can be calculated by repeatedly applying the Lyapunov equation on each element node. When there is a new node appended to the network, the state information for the network can be updated by applying the Lyapunov equation on the additional node.

After the information has been integrated, the uncertain information needs to be further processed, aiming to analyze uncertainty and derive the meaning of the information. The task to achieve this purpose is called Data Fusion [8]. As the information of collected evidence is subject to the analysis difficulties of incompleteness, imprecision, and uncertainty, the core of the analysis is to find the probabilities behind the information. These output probabilities provide a support to decision makings and court judgments, and thus for such purposes, the generation of these probabilities is required to be representable. Some handling approaches are building probabilities based on belief measures. These approaches cause the analysis results difficult to be represented. For instance, the inference of Bayesian Theory under the assumption that \( a \) and \( b \) are disjoint propositions can be expressed as:

\[ P(a + b) = P(a) + P(b) + P(a \cup b) \]

Such inference cannot distinguish between lack of belief and disbelief, and disallows to withhold belief from a proposition before the negation of the proposition belief [9]. In contrast, the approaches based on plausibility measures can easily represent the analysis result, and hence they are encouraged to be used in the forensic context.
Possibilistic Logic (PL) is a viable approach developed from possibility theory that handles uncertainty in a logical setting. PL measures probability by classical logic formulae associated with weights of necessity degrees and such measuring is inconsistency-tolerant [10]. A first-order PL is basically a pair \((P, \alpha)\) made of a classical first order logic formula \((P)\) and a weight expressing \((\alpha)\) certainty or priority. The inference rules can be expressed as:

\[\neg P \lor Q, \alpha \vdash (Q, \min(\alpha, \beta))\]

Such inference can be improved to deal with inconsistency and be applied to derive implicit relationships from PL knowledge bases [11]. Such knowledge bases are expected to be used to generate reports by the Reporter of CFP and retained as supporting evidences for the reports.

5. USING NETWORK CENTRIC FORENSIC SYSTEM

NCEFS can be utilized for various investigation purposes including security compliance assurance, employee behavior monitoring and incident investigation, etc. The GTMC, a large motor vehicle manufacturer in China, is establishing and expanding its internal information control and forensic infrastructure referring to this concept. Based on the nature of the investigations, the usages of NCEFS can be divided into the following two categories.

5.1 Regular Investigation

Regular Investigation is a circling process needed for assuring certain activities have happened or have not happened in a timely manner. For example, if an enterprise has established some security agreement that prohibits employees from installing social applications on the office computers, regular investigations are required to examine the office computers to assure the employees are complying with the security agreement. The regular investigation usually is only interested in certain aspects of the devices, such as network logs or installed software list etc. However, if all investigations are treated equally and for every single investigation, the devices acquire the full dump of disk and memory, it will heavily increase the burden of the systems and networks of the enterprise, and make the investigations inefficient. For this type of investigations, not all available information of the devices is required to be resented. CFP Planner of NCEFS will request the CFP Collector only to acquire relevant information to avoid redundant processing and shorten the analysis time.

5.2 Incident Response Investigation

When an information security incident happens and have been aware by the enterprise, investigations must be carried out to probe out when, where, why, and with whom did the incident occur [12]. In this situation, as the scale and influence is unknown at the beginning when an incident is just detected, investigator must collect as much information as available to discover the facts and consequences about the incident, then seek for evidences that can prove the crime and the wrecker. Under this circumstance, in addition to the permanent data storage of the victim device, live response in terms of information on current network connections, running processes, open files and other artefacts must be collected immediately. When an incident happens, NCEFS will request Forensic Agent to lock down the victim device and cease its normal operation, and start to collect data on its permanent storage components and information of live response. NCEFS will not release the lock of the device until the data collection has been finished, in order to guarantee the digital evidences not to be dropped or contaminated.

6. CONCLUSION

This research explored the potential value of integrating conventional forensic tools and manual processes into a systematic and automated forensic system in enterprise context. The automated forensic processing is introduced to reduce the operational mistakes and increase the efficiency. An enhanced forensic workflow has been proposed to prevent the negligence and ignorance upon the essential procedures of investigation during the planning, acquiring, analyzing and reporting stages.
This research promotes the utilization of modern network distributed system concept and information fusion theory in the implementation of forensic. By the integration of network distributed system with forensic techniques, the investigation can become more efficient and agile in response to incidents. The idea of implementing forensic system with network distribution and information fusion concept encourages innovative forensic practices in terms of intelligent forensic planning, remote evidence collection, and comprehensive information analysis. Such innovative forensic approaches are expected to overcome the weaknesses of the traditional forensic techniques, reduce redundant processing, and render more robust forensic processes. In future, research effort will be spent on improving the incident analytic algorithm, so that more accurate forensic reports can be produced.

7. REFERENCES


