International Journal of Information Technology and Knowledge Management
July-December 2010, Volume 2, No. 2, pp. 561-565

A Digital Investigation Tool based on Data Fusion in Management of Cyber Security Systems

Suneeta Satpathy[1], Sateesh K. Pradhan[2] & B.B. Ray[3]

With overwhelming use of Internet, security in Cyberspace has become a prime concern. The forensic digital analysis as a whole, in its relative infancy, is the unwilling victim of the rapid advancement of computer technology, so it is at the mercy of ever more new and complex computing approaches. Forensic digital analysis is unique among the forensic sciences in that it is inherently mathematical and generally comprises more data from an investigation than is present in other types of forensics. The digital investigation process can be driven using numerous forensic investigation models. Among these is the need to analyze forensic materials over complex chains of evidence in a wide variety of heterogeneous computing platforms, environments and transports. This paper compares and contrasts different forensic investigation models and highlights the main components of forensic investigation model. It also proposes a fusion based investigation tool by grouping and merging the same activities or processes that provide the same output into an appropriate phase and mapping them into the domain of data fusion. This grouping process of the activities will balance the investigation process and mapping them into data fusion domain will produce more quality data for analysis and can produce potential legal digital evidence as an expert testimony in the court of law.

Keywords: Information Technology, Digital Investigation, Digital Evidence, Data Fusion.

1. Introduction

The unprecedented growth of Internet has spawned new businesses, opportunities, ideas, and unfortunately, new problems. Though the concept of availing services electronically cutting across geographical boundaries is exciting, there is a great deal of cynicism about the security aspects of vital information resources. The versatility of information technology used to commit sophisticated crimes is steadily increasing: which are being exploited by unscrupulous elements in the society for disrupting peace and causing mayhem. The society is moving from a paper based to a paperless scenario, from centralization to decentralization, from controlled access to totally independent access and so on. In such a scenario it becomes possible for anti-social elements [7] to cause havoc with minimum retribution from the existing criminal justice system. Legal support, criminal justice delivery system and international cooperation have not kept pace with the technological advancements, which have taken place with the advent of information technology. Computer crime [5][6] is a new form of transnational crime, which requires concerted efforts by the Law Enforcement agencies, as the perpetrators of computer crimes can be more elusive than ever before. To effectively combat the computer crime, it is not sufficient to successfully investigate the crime and nab the criminal, but more important is to prosecute and administer justice, according to the law of the land. This requires an effective investigation tool, which fully supports the detection and prosecution of cyber criminals.

The objective is to present an investigation tool, methodology and technology for design and deployment of data fusion applications in digital investigation. Section 2 gives a brief description on digital investigation. Section 3 compares different forensic investigation models and highlights the main components of forensic investigation model. Section 3 outlines the theory of data fusion and its application in digital investigation. And a fusion based investigation tool is proposed by grouping and merging the same activities that provide the same output into an appropriate phase and mapping them into the domain of data fusion in section 5. Section 6 outlines the feasibility of the proposed investigation tool.

2. Digital Investigation and Problem Statement

"Digital investigation is a process that uses science and technology to examine digital evidence and that develops and tests theories, which can be entered into a court of law, to answer questions about events that occur"[6] [9].

Digital Investigation faces several problems. Some of them are:

- Digital investigations are becoming more time consuming and complex as the volumes of data requiring analysis continue to grow.

[1]P.G Department of Computer Application, CEB, BPUT, Bhubaneswar.

[2]Department of Computer Engineering College of Computer Science, King Khalid University, Abha

[3]P.G Department of Computer Application, Utkal University, Bhubaneswar, INDIA

Email: suneetasatpathy@rediffmail.com, sateeshind@yahoo.com

- Digital investigators are finding it increasingly difficult to use current tools to locate vital evidence within the massive volumes of data.
- Log files are often large in size and multi-dimensional, which makes the digital investigation and search for supporting evidence more complex.
- Digital evidence [6] [9] by definition is information of probative value stored or transmitted in digital form. It is fragile in nature and can easily be altered or destroyed. It is unique when compared to other forms of documentary evidence.
- Forensic investigation tools available are unable to analyze all the data found on computer system to reveal the overall pattern of the data set, which can help digital investigators decide what steps to take next in their search. Also the data offered by computer forensic tools can often be misleading due to the dimensionality, complexity and amount of the data presented.

3. EXISTING DIGITAL INVESTIGATION MODELS

Over the years, several forensic investigation models have been proposed, they include:

- Kruse and Heiser;
- Lee's model;
- Casey's model;
- DFRWS frame work meta-model;
- The Reith, Carr and Gunsch model;
- The Ciardhuain model.

Kruse and Heiser model [4]- Kruser and Heiser stated that forensic investigation consists of 3 basic components: Acquire evidence, Authenticate evidence, Analyzing data. Lee model [13]- Lee proposed a model that consists of 4 steps, they are: Recognition, Identification, Individualization, Reconstruction. The steps proposed by Lee refer to only a part of the forensic investigation process i.e the investigation stage (no preparation or presentation). Casey's model [6]- It is similar to that proposed by Lee, the 1st and last stages are the same. It focuses on processing and examining digital evidence (focuses on investigation). The steps also include: Recognition, Preservation, Classification, Reconstruction. Digital forensic Research working group (DFRW) model [13] [12]-The DFRW model includes crucial stages of the investigation and also includes the Presentation stage. It consists of the following stages: Identification, Preservation, Collection, Examination, Analysis, Presentation, Decision. The Reith, Carr and Gunsch model [11] - The Reith, Carr and Gunsch model included other components not found in the above mentioned frameworks. It consists of: Identification, Preparation, Approach, Strategy, Preservation, Collection, Examination, Analysis, Presentation, Returning evidence. Ciardhuain model [13]- Ciardhuain model consists of the following: Awareness, Authorization, Planning, Notification, Search and identify evidence, Collection, Transportation, Storage, Examination, Hypothesis, Presentation, Proof/ Defense, Dissemenation.

After examination of the above mentioned investigation models, it was noted that: [13] [1]

- Each preceding model modifies the previous.
- Some of the models have very similar approaches.
- Some of the models concentrate on different areas of the investigation
- In all the above models no specific theory has been defined which can be applied for processing the digital data, reducing the data while retaining the useful data, analysis of the data, keeping a record of criminal profiles and digital evidence for its future reference, presenting the evidence documentation as an expert testimony in the court of law.

Since the goals of the evidence preservation principles are [3] [6]:

1. To maximize evidence availability and quality;
2. Maintain the integrity of the evidence during the digital investigation process.

While designing a forensic investigation model, the model should not concentrate on one particular type of case investigation or on a certain stage. But it should provide the general guidelines, approach and keep a balance in the processes identified by these models. It should incorporate the basic components of forensic investigation which are: [10]

- Preparation;
- Investigation;
- Presentation.

4. NEED FOR APPLICATION OF DATA FUSION IN DIGITAL INVESTIGATION

The Law enforcement agencies are facing a novel challenge in terms of jurisdiction and identification of crimes perpetrated on the Internet, which do not know geographical boundaries, need a multi-lateral approach of investigation and prosecution. They have to analyze enormous amount of data and collection and analysis of such large amounts of data may degrade performance to unacceptable levels. Data fusion gains power and relevance for the counter cyber terrorist mission because computer technology enables large volumes of information to be processed in short times. Multi-

sensor data fusion is an evolving technology, concerning the problem of how to fuse data from multiple sensors in order to make a more accurate estimation of the environment and to generate information of a superior quality [2][4][8]. It is a formal framework in which the means and tools for the alliance of data originating from different sources are expressed. The first data fusion methods were primarily applied in the military domain, in recent years these methods have also been applied to problems in the civilian domain and various non-military applications (e.g., air traffic controls, robotics, image processing, remote sensing, hazardous wastes tracking, environmental data fusion, etc.). It also provides an important functional framework for building next generation security systems. A more recent idea is the application of data fusion techniques to the area of information security [14, 15]. Tim Bass presented a Data Fusion model, based on the Joint Directors of Laboratories (JDL) Functional Data Fusion Process Model [14].

The main goal of data fusion system is [2]

- Reliability - to obtain better results than the best individual data source is capable of providing.
- Completeness - no direct way of measuring required property.
- Improvement - the need to take more factors, or influencing quantities, into account.
- Comprehension - the need to reduce information overload.

5. PROPOSED FUSION BASED INVESTIGATION TOOL

This paper proposes a fusion based investigation tool (fig-1) by grouping and merging the digital investigation activities or processes that provide the same output into an appropriate phase and mapping them into the domain of data fusion (Table-2). This grouping process of the activities will balance the investigation process and mapping them into data fusion domain will produce more quality data for analysis and can produce potential legal digital evidence to be presented as an expert testimony in the court of law. It is motivated by Data fusion model proposed by the JDL [2] that fuses data from various heterogeneous sources in order to attain low false alarm rates and high threat detection rates.

The data fusion process at different progressions is further explained in (Table-1).

- Preprocessing;
- Processing of events in various levels of fusion;
- Decision making;
- Evidence accumulation;
- Data transformation (converts the raw data into structured information);
- Data reduction (reduces the representation of the dataset into a smaller volume to make analysis more practical and feasible).

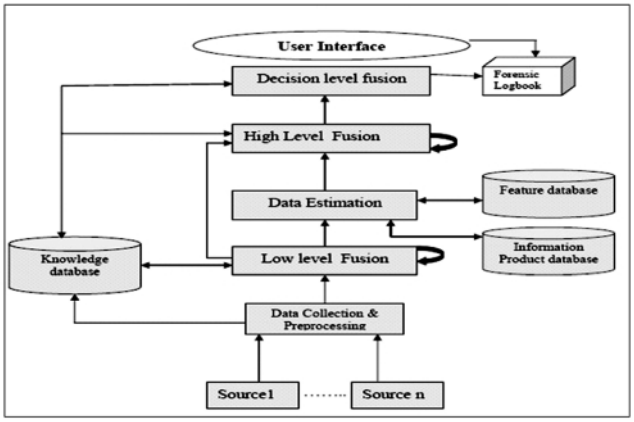

Fig. 1: Fusion based Forensic Investigation Tool

Table. 1
Activities at Different Levels in
Fusion based Investigation Tool

Data Fusion Levels	Activities
Source	Events of the crime scene. Sources are identified only when crime has been reported and authorization is given to the Investigating agencies.
Data Collection and Pre-Processing [2,4]	The first step where data collected from various sources are fused and processed to produce data specifying semantically understandable and interpretable attributes of objects. The collected data are aligned in time, space or measurement units and the extracted information during processing phase is saved to the knowledge database or knowledgebase.
Low level fusion [2,4]	Concerned with data cleaning (removes irrelevant information), data transformation (converts the raw data into structured information), data reduction (reduces the representation of the dataset into a smaller volume to make analysis more practical and feasible). It reduces a search space into smaller, more easily managed parts which can save valuable time during digital investigation.
Data estimation	It is based on a model of the system behavior stored in the feature database and the knowledge acquired by the knowledgebase. It estimates the state of the event. After extracting features from the structured datasets, fusion based

Contd...

Contd...

	investigation tool will save them to an information product database.
High level fusion[2,4]	Develops a background description of relations between entities. It consists of event and activity interpretation and eventually contextual interpretation. Its results are indicative of destructive behavior patterns. It effectively extends and enhances the completeness, consistency, and level of abstraction of the situation description produced by refinement. It involves the use of data mining functionalities such as classification and clustering to extract useful patterns among the data. The results obtained would be indicative of destructive behavior patterns.
Decision level fusion[2,4]	Analyzes the current situation and projects it into the future to draw inferences about possible outcomes. It identifies intent, lethality, and opportunity and finally decision of the fusion result is taken in this level. Result can be stored in the log book in a predefined format from which evidence report can be generated. The same can be stored for future reference.
User interface	It is a means of communicating results to a human operator. Evidence Report prepared and generated is represented as evidence to the problem solved by using the tool.
Forensic Log Book[5,6,9]	The digital information are recorded with a pre-defined format like date and time of the event, type of event, and success or failure of the event, origin of request for authentication data and name of object for object introduction and deletion. A time stamp is added to all data logged. The time line can be seen as a recording of the event. The log book can be used as an expert opinion or legal digital evidence.

Table. 2
Mapping Phases of Digital Investigation Model into Fusion based investigation Tool

Investigation Phase	Mapping into Fusion based investigation Tool	Output
Preparation	source	Awareness, Authorization, Plan, Warrant, Notification, Confirmation.
collection and preservation	Preprocessing	Potential Evidence Sources, Media, Devices, definition of important events, definition of unimportant events, elimination procedure.
Investigation/ Examination/ Analysis	Low level Fusion	Crime type, common reference format events should be aligned, alignment procedure.
	High level Fusion	Develops a background description of relations between entities. Event and activity interpretation and eventually contextual interpretation. Results are indicative of destructive behavior patterns. Extends and enhances the completeness, consistency, and level of abstraction of the situation description produced by Refinement.
	Decision Level Fusion	Log Files, File, Events log, Data, Information Evidence, Evidence Report.
Presentation in the court of law	Forensic Log Book/ User Interface	Evidence Explanation, Evidence Disposed, New Policies, New investigation procedures, Investigation Closed.
Storage	Database for future reference	Database can be used to store the crime types criminal profiles and intent, lethality which will be helpful for them in solving crime cases in future.

6. FEASIBILITY

The proposed tool can be useful as an evidence acquisition tool for supplying the Offline admissible legal digital evidence for the Investigating agencies including preservation and continuity of evidence, and transparency of the forensic methods. As the admissibility and weight are the two determinants in the legal acceptability of digital evidence [9], the courts deal with issues related to the difference between the novel scientific evidence and the legal evidence. There are three requirements for the evidence to be admissible in the court [6]:

- Authentication (showing a true copy of the original);
- The best evidence rule (presenting the original);
- Exceptions to the hearsay rule. (allowable

exceptions are when confession, business or other official records are involved).

From an evidence perspective, the law enforcement agencies will seek something that they can proof and demonstrate to others long after the event is over.

7. CONCLUSION

Profiling, identifying, tracing, and apprehending cyber suspects are the important issues of research today. Different tools have been developed to detect the misuse and suspicious activities, but very few of them have the provision to help law enforcement agencies. They require adequate evidence in order to penalize the criminal, thus, heavily depending on reports of forensic scientists. To collect the digital evidence is not an easy task. Within a computer system the anonymity afforded by the criminal encourages destructive behavior while making it extremely difficult to prove the identity of the criminal. In this paper we have given the idea of constructing a proprietary fusion based investigation tool for investigative agencies which can work at a stretch, process different types data both syntactically and semantically, filter out the files required for forensic analysis to retrieve the legal digital evidence. The output of fusion-based investigation systems will be the estimates of the identity of a threat source, the malicious activity, taxonomy of the threats, the attack rates, and an assessment of the potential severity of the projected targets. Our future work includes extending the investigation tool to analyze different cyber crime cases.

REFERENCES

[1] Baryamureeba, V., Tushabe, F.: "The Enhanced Digital Investigation Process Model", Makere University Institute of Computer Science, Uganda 2004.

[2] David L. Hall, Sonya A.H. McMullen, "Mathematical Techniques in Multisensor Data Fusion", 2nd edition, Artech House, 2004.

[3] D. Brezinski and T. Killalea, "Guidelines for Evidence Collection and Archiving", RFC3227, February 2002.

[4] E. Waltz and J. Linas, "Multisensor Data Fusion". Artech House, Boston, MA, 1990.

[5] E. Casey (ed.), "Handbook of Computer Crime Investigation", Academic Press, 2001.

[6] E. Casey, "Digital Evidence and Computer Crime", 2nd Edition, Elsevier Academic Press, 2004.

[7] H Lipson, "Tracking and Tracing Cyber Attacks: Technical Challenges and Global Policy Issues (CMU/SEI-2002-SR-009)", CERT Coordination Center, November 2002.

[8] http://www.data-fusion.org.

[9] J. Danielsson, "Project Description a System for Collection and Analysis of Forensic Evidence", Application to NFR, April 2002.

[10] Michael Kohn, Jhp Eloff, Ms Olivier. "Framework for Digital Forensic Investigation: Information and Computer Security Architectures Research Group (ICSA)", University of Pretoria.

[11] Reith, M., Carr, c. and Gunsch, G.: "An Examination of Digital Forensic Model", International of Digital Evidence. Fall 2002, 1, Issue 3, 2002.

[12] Siti Rahayu Selamat, Robiah Yusof, Shahrin Sahib.: "Mapping Process of Digital Forensic Investigation Framework". IJCSNS International Journal of Computer Science and Network Security, 8, No.10, October 2008.

[13] Seamus O Ciardhuain. "An Extended Model of Cybercrime Investigation", Journal of Digital Evidence Summer 2004; 3, Issue 1.

[14] T. Bass, "Multi-sensor Data Fusion for Next Generation Distributed Intrusion Detection System", In Proceedings of the IRIS National Symposium on Sensor and Data Fusion, 1999.

[15] Varshney, "Distributed Detection and Data Fusion", Springer-Verlag, New York, NY., 1995.