QUANTITATIVE COMPUTATIONAL FRAMEWORK FOR ANALYZING EVIDENCE TO IDENTIFY ATTACK INTENTION AND STRATEGY IN NETWORK FORENSICS

MOHAMMAD RASMI HASSUN MOSA

UNIVERSITI SAINS MALAYSIA
2013

QUANTITATIVE COMPUTATIONAL FRAMEWORK FOR ANALYZING EVIDENCE TO IDENTIFY ATTACK INTENTION AND STRATEGY IN NETWORK FORENSICS

BY

MOHAMMAD RASMI HASSUN MOSA

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

June 2013

This doctoral dissertation is dedicated to my late father

"Rasmi"

God bless his soul

ACKNOWLEDGEMENTS

All praise and thanks be to Allah, the Lord of the world, for giving me the energy to finish my research. I would like to thank my lovely mother, Hajah UmMaher who always makes dua'a (supplications to Allah) for me and for supporting me spiritually throughout my life.

I'm deeply grateful to my supervisor, Assoc. Prof. Dr. Aman Jantan, for his invaluable motivation, enthusiasm, patience and guidance. His guidance helped me in all the time of research and writing of this thesis. Without his offering to accomplish this research, I would not be able to finish my study at USM.

My sincere thanks also goes to my family; my brothers Maher, Hani, Tala't, Belal, Hassun and Ahmed, my sister UmHazem, my wife UmHashem, and my children's Jude, Sara, and Hashem who have always shown their faithful support during my study. I appreciate their everlasting patience during the long period of my study.

I'm much obliged to my group members AbudAlghani, Azri Hafiz, Evi, Safar, Dr. Ghassan, Dr. Fadli, Izham for their valuable comments and suggestions. I sincerely thank my friends: Dr. Ahmed Adel, Hani Al Mimi, Husameddin Dawoud, Mahmoud Baklizi, Muhannd AbuHashem, Bassam Al Tamimi, Mohammad Al Madi, Ashraf Maghari, and Dr. M. Al Momni for the stimulating discussions, help and supports.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

CHAPTER TWO: LITERATURE REVIEW

CHAPTER THREE: EVIDENCE ANALYSIS FRAMEWORK IN NETWORK FORENSICS

CHAPTER FOUR: DESIGN AND IMPLEMENTATION OF THE NFEA FRAMEWORK

CHAPTER FIVE: EXPERIMENTAL RESULTS, ANALYSIS, AND DISCUSSION

CHAPTER SIX: CONCLUSION AND FUTURE WORK

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

ActDF	Active Digital Forensics
ADF	Abstract Digital Forensics
AIA	Attack Intention Analysis
Asymp. Sig.	Asymptotic Significance
BHOs	Browser Helper Objects
BPA	Basic Probability Assignment
CBR	Case-Based Reasoning
CERT	Computer Emergency Response Team
CFFTPM	Computer Forensic Field Triage Process Model
CMC	Cumulative Match Characteristics
CSV	A Comma Separated Values
CVE	Common Vulnerabilities and Exposures
DBMS	Data Base Management System
DCOM	Distributed Component Object Model
DDoS	Distributed Denial of Service
DFRWS	Digital Forensics Research Workshop
DgmLen	Datagram Length
DIPL	Digital Investigation Process Language
DoS	Denial of Service
D-S	Dempster–Shafer
EEDI	End-to-End Digital Investigation
EIDIP	Enhanced Integrated Digital Investigation Process
FBI	Federal Bureau of Investigation
FNDR	False Negative Detection Ratio
FNRR	False Negative Ranking Ratio
FORZA	Forensics Zachman
FP	Frequent Pattern
FPDR	False Positive Detection Ratio

FPRR	False Positive Ranking Ratio
FTP	File Transfer Protocol
HTML	Hyper Text Markup Language
HTTPS	Hypertext Transfer Protocol Secure
ICMP	Internet Control Message Protocol
IDIP	Integrated Digital Investigation Process
IDS	Intrusion Detection System
IDSDFM	Intrusion Detection System Data Fusion Model
IP	Internet Protocol
IPS	Intrusion Protection System
IT	Information Technology
LISP	List Processing
LSA	Local Security Authority
LSASS	Local Security Authority Subsystem Service
MASP	Mining attack Sequential Pattern
MLP	Multi-Layer perception
MOSTI	Ministry of Science, Technology and Innovation
MPSVM	Multi-class Probability Support Vector Machines
NFATs	Network Forensics Analysis Tools
NFEA	Network Forensics Evidence Analysis
NIDS	Network Intrusion Detection System
NSM	Network Security and Monitoring
OS	Operating System
PC	Personal Computer
PCAP	Packet Capture
PDU	Protocol Data Unit
POP3S	Post Office Protocol 3 Secure
ProDF	Proactive Digital Forensics
R2L	Remote to Local80

RAM	Random Access Memory
ReDF	Reactive Digital Forensics
RPC	Remote Procedure Call
SAI	Similarity of Attack Intentions
SAS	Similarity of Attack Strategies
seq(be/le)	Sequence (big ending / little ending)
SLR	Systematic Literature Review
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
SVM	Support Vector Machine
TCL	Tool Command Language
TCP	Transmission Control Protocol
TNDR	True Negative Detection Ratio
TNRR	True Negative Ranking Ratio
TOS	Type Of Service
TPDR	True Positive Detection Ratio
TPRR	True Positive Ranking Ratio
TTL	Time to live
U2R	User to Root
UDP	User Datagram Protocol
USM	Universiti Sains Malaysia

LIST OF APPENDICES

LIST OF SYMBOLS

I, i	The attack Intention
A, a	The attack
E, ev	Evidence
P(ev)	The probability of the evidence (ev)
P(ev|i)	The probability of the evidence (ev) when the intention (i) occurs
P(i)	The probability of the intention (i)
P(¬ev|i)	The probability of the evidence (ev) when the intention is not related to (i)
P(¬i)	The probability that the intention (i) is not occurs
P(i|ev)	The probability of the intention (i) for a given evidence (ev)
BPA ( )	The basic probability assignment function
$\{EV_s\}$	The set consisting of all evidence of the intention (i)
Ø	Empty set
{PS(EVs)}	The subset of EV_s, it is the class of general propositions concerning the actual state of the attack (a) evidence's domain, including empty set
Be( )	The belief or support function
Pl( )	The plausibility function
{PA}	The set contains all of past attacks
{AE}	The set contains all of the attack evidence
{AI}	The set contains all attack intentions for all predefined attacks
{H}	The set of hypothesis that related to the attack detection accuracy and collection of evidence
$SimA_nI(A_k)$	The similarity of the attack (A_k) intention, k is the number of the new attack
$SimAI(A_k)$	The maximum similarity of the attack (A_k) intention
$\{A_kE\}$	The set of the priority value of the attack (k) evidence
PV	The priority value
$\{A_kV\}$	The set of the evidence group value of the attack (k)

GV	The group value
$f_c(A_kE)$	The classifier function of the attack (k) evidence
$\mathbf{Euc}(A_kE_z, A_dE_z)$	The Euclidean distance function for the attack (A_k) and attack (A_d)
$\mathbf{Sim}(A_kE_z, A_dE_z)$	The cosine similarity between the attack (A_k) and attack (A_d) evidences from the same group
$\mathbf{SimA_nI}(A_k)$	The similarity of attack intentions metric for the new attack named (A_k)
$\mathbf{Weight}(A_d)$	The weight of the similarity of the attack (A_d) with the new attack (A_k)
$\mathbf{SimAS}(A_k)$	The similarity value of the attack strategy with the new attack (A_k)
{CrCase}	The set contains all cyber crimes cases
At	The cyber crime attributes
$\mathbf{Weight}(A_kAt_z)$	The weight of the attribute (z) for the cyber crime (k)
$\mathbf{Weight}\ (A_kI)$	The net weight of attack (k) intention
NC	The new cyber crime case
PreC	The predefined cyber crime case
NCAt	The attribute of the new cyber crime case
PreCAt	The attribute of the new predefined cyber crime case
SimCase( )	The similarity of the new cyber crime case response with others
CaseID, C	The cyber crime case identity
EW	The weight of the attack evidence
IW	The weight of the attack intention
SW	The weight of the attack strategy
SCID	The similar cyber crime case
SimC1	The weight of the similar cyber crime case for each new case based on the available evidence (baseline method)
SimC2	The weight of the similar cyber crime case included with the weight of the highest value of the similar attack strategy

SimC3 The weight of the similar cyber crime case included with the highest detection probability values of the intentions

SimC4 The weight of the similar cyber crime case based on the similar attack strategy and intentions

RANGKA KERJA PENGIRAAN KUANTITATIF BAGI MENGANALISA BUKTI-BUKTI UNTUK MENGENALPASTI TUJUAN SERANGAN DI DALAM FORENSIK RANGKAIAN

ABSTRAK

Peningkatan jumlah jenayah siber telah mendorong para pengkaji di dalam bidang forensik rangkaian membangunkan teknik-teknik yang baru untuk menganalisa dan menyiasat jenayah ini. Walaupun jenayah siber menghasilkan jumlah bukti yang banyak, analisis dan ukuran terhadap kesan daripada kerosakan yang disebabkan oleh jenayah ini adalah sukar kerana jumlah bukti yang terlalu besar di dalam setiap kes. Hal ini telah menjadikan kos penyiasatan kes jenayah siber masa kini begitu mahal dan memerlukan masa yang panjang. Tambahan pula, teknik-teknik ini menggunakan proses aktif dan reaktif untuk menganalisis jenayah siber, dan proses ini bermula selepas jenayah siber ini dikenalpasti, dan seterusnya menyebabkan pengenalpastian bukti-bukti penting menjadi sukar. Selain itu, maklumat yang diperlukan untuk memahami dan menganalisa faktor-faktor jenayah siber seperti tujuan dan strategi jenayah ini juga adalah terhad.

Tesis ini mencadangkan satu rangka kerja baru untuk menganalisis bukti-bukti jenayah siber. Rangka kerja ini bertujuan untuk menggunakan bukti-bukti jenayah untuk membina semula tujuan serangan dan menganggar strategi-strategi serangan yang serupa. Tujuan serangan dikenalpasti menerusi algoritma baru yang dikenali sebagai Analisis Tujuan Serangan, yang meramalkan tujuan jenayah siber dengan menggabungkan teori Dempster-Shafer dengan teknik rangkaian penyebab. Strategi serangan serupa telah dianggarkan dengan menggunakan salah satu daripada kaedah yang dicadangkan. Kaedah pertama ialah dengan mencipta satu model baru menggunakan bukti-bukti berkenaan apabila tujuan jenayah siber tidak dapat dikesan. Model ini bertujuan untuk mengukur bukti-bukti serupa antara kes-kes jenayah siber baru dengan yang lampau untuk menganggarkan strategi yang serupa. Kaedah kedua pula dijelaskan dengan mereka bentuk algoritma baru yang dikenali sebagai Persamaan Strategi Serangan yang menggabungkan kaedah pertama dengan pra-analisis faktor-faktor tujuan untuk menambah-baik keputusan analisa jenayah siber ini. Pra-analisis bagi tujuan serangan ini dinilai dengan menggunakan satu kaedah baru yang dikenali sebagai Persamaan Tujuan Serangan, yang menggunakan persamaan metrik untuk menganggarkan tujuan jenayah siber yang serupa. Tujuan dan strategi serangan digunakan untuk membandingkan kes baru dengan kes sedia ada yang didokumenkan untuk meningkatkan kebarangkalian bukti-bukti jenayah siber yang sepadan dengan potensi kes-kes yang serupa. Oleh itu, perbandingan ini telah memaksimakan kebarangkalian penemuan seperti padanan dan peratusan persamaan antara kes-kes dengan menggunakan teknik penaakulan berasaskan kes untuk menyediakan bukti berguna yang dapat membantu penyiasat kelak.

Rangka kerja yang dicadangkan ini telah dinilai dengan menggunakan data trafik rangkaian sebenar yang diperolehi daripada makmal kajian USM dan juga cabaran set data forensik oleh Projek Honeynet. Keputusan kajian menunjukkan rangka kerja ini mampu memaksimakan nilai kebarangkalian secara purata sebanyak (9.17%) untuk mendapat kes-kes serupa, yang dapat membantu penyiasat untuk menyelesaikan jenayah siber dengan mengkaji kes-kes yang serupa ini, sekaligus membolehkan mereka mengadaptasikan jalan penyelesaian untuk kes yang baru. Kajian ini menunjukkan suatu rangka kerja baru untuk menganalisa bukti-bukti yang dapat meningkatkan proses penyiasatan menerusi aktiviti membuat keputusan yang lebih baik yang dapat membantu dalam memberkas penjenayah sebenar.

QUANTITATIVE COMPUTATIONAL FRAMEWORK FOR ANALYZING EVIDENCE TO IDENTIFY ATTACK INTENTION AND STRATEGY IN NETWORK FORENSICS

ABSTRACT

The increasing number of cyber crimes has motivated network forensics researchers to develop new techniques to analyze and investigate these crimes. Although cyber crimes produce a large volume of evidence, analyzing and measuring the extent of the damages caused by these crimes are difficult because of the overwhelming amount of evidence involved in each case. Thus, current cyber crime investigation techniques are costly and time consuming. In addition, these techniques normally use active and reactive processes to analyze cyber crimes, and such processes start after the cyber crime has been identified, which makes identifying useful evidence difficult. Moreover, the information required to understand and analyze cyber crime factors such as the intention and strategy of the crime are limited.

This thesis proposes a new framework to analyze cyber crime evidence. The proposed framework aims to use cyber crime evidence to reconstruct attack intentions and estimate similar attack strategies. The intentions are identified through a new algorithm called Attack Intention Analysis, which predicts cyber crime intentions by combining Dempster-Shafer theory and a causal network. Similar attack strategies have been estimated by using one of the two proposed methods. The first method creates a new model that uses evidence when the intentions for a cyber crime are undetected. This model aims to measure similar evidence between new and pre-existing cyber crime cases to estimate similar strategies. The second method is illustrated by designing a new algorithm called Similarity of Attack Strategies, which integrates the first method with the pre-analyzed intention factors to improve the results of the cyber crime analysis. The pre-analyzed intention is evaluated by using a new algorithm called Similarity of Attack Intentions, which uses a similarity metric to estimate similar cyber crime intentions. The attack intentions and strategies are used to compare a new case with existing documented cases to increase the possibility of matching cyber crime evidences with potential similar cases. Thus, this comparison maximizes the probability of finding such matches and maximizes the percentage similarities between cases by using the case-based reasoning technique to provide useful evidence that will assist investigators.

The proposed framework was evaluated by using real network data traffic obtained in the USM research labs and from the forensics challenge dataset by the Honeynet Project. The experimental results showed that the proposed framework maximize the probability value in average by (9.17%) of retrieving similar cases, which can help the investigators to resolve a cyber crime by studying similar cases, thereby enabling them to adapt a solution for the new case. This study presents a new framework for analyzing evidence which can enhance the investigation process through better decision making activities that will help in apprehending the real perpetrators.

CHAPTER ONE
INTRODUCTION

1.1 General Overview

Nowadays, cyber crimes are increasing and have affected large organizations with highly sensitive information. For example, the International Monetary Fund information system was compromised by a sophisticated attack for over a month in 2011 (Wolf and Maclean, 2011; BBC, 2011). The databases of major companies such as Sony Group and Google have also been penetrated by several anonymous computer hackers who stole personal data such as passwords from customer accounts (Jeremy, 2011, Runciman, 2012). Consequently, the affected organizations spent more resources analyzing the cyber crimes rather than detecting and preventing these crimes. Network forensics plays an important role in investigating cyber crimes; it helps organizations resolve cyber crimes as soon as possible without incurring a significant loss.

In general, the evidence is everything that is used to demonstrate and determine the truth of an assertion in order to support resolving of cyber crimes. Cyber crimes produce a large volume of evidence through network monitoring and capturing tools. Nevertheless, a significant amount of time is required to discover the real perpetrator. According to the 2011 CyberSecurity Watch survey, 21% of digital crimes are caused by “unknown” perpetrator (CERT et al., 2011). This fact encourages the perpetrators to repeat the cyber crimes. In the case of Sony, the 2011 attacks launched by a group of hackers who call themselves “LulzSec” penetrated a number of Sony sites and stole customers’ data. This incident indicates that the current network forensic investigation approach, which is reactive, is time consuming, costly, and error prone as it requires much effort to analyze the overwhelming amount of evidence presented in each case. Moreover, gathering useful evidence through the reactive approaches such as proposed by Rogers et al. (2006), Freiling and Schwittay (2007), and Almulhem (2009) is difficult because evidence is collected right after the detection of the cyber crime. Thus, a new approach is needed to analyze evidence and enhance the investigation process.

Most existing frameworks and models in network forensics such as proposed by Carrier and Spafford (2003), Baryamureeba and Tushabe (2004), Rogers et al. (2006), Freiling and Schwittay (2007), Almulhem (2009), Pilli et al. (2010), and Alharbi et al. (2011a) serve as a guideline in the investigation of cyber crimes without enough information or details on how to analyze the evidence. In addition, the vagueness of each phase processes is a gap exists in the network forensic phases of these frameworks and models. This gap exists because investigators have difficulty understanding how the phases work and how the outcomes for each phase are achieved. Considerable time is consumed to understand the phases as the researchers focus on the number and ordering of phases rather than the core operations inside these phases (Almulhem, 2009; Pilli et al., 2010; Alharbi et al., 2011a).

Based on various existing digital forensic approaches, Pilli et al. (2010) introduced a generic process model for network forensics. The proposed model has multiple processes embedded into nine phases: preparation, detection, incident response, collection, preservation, examination, analysis, investigation, and presentation. The investigation phase plays an important role in decision making to resolve cyber crimes. However, as mentioned by Casey (2005), the investigation phase is complex. The analysis phase supports the investigation phase in the latter's aim to improve the quality of decision making. The analysis phase analyzes the evidence of a cyber crime and generates important observations to establish the intention and strategy of the crime (Pilli et al., 2010). Attack intentions are plan instances selected for processing to achieve a goal and infers the motive of an attack based on the cyber crime actions. Attack strategy explains how the cyber crime is done and identifies the steps of the attack to generate a scenario. The analysis of large volumes of cyber crime evidence is a challenging issue (Wang et al., 2006a).

In conclusion, given the large amount of cyber crime evidences, considerable effort, time, and resources are required in collecting, analyzing, and summarizing useful evidence that help investigators establish a suitable decision. However, identifying intentions and strategies of cyber crimes is difficult for most investigators in network forensics. In general, the analysis phase attempts to establish the motive of a cyber crime and how the attack occurred by identifying the intentions and strategies of the crime. Unfortunately, with the increasing number of cyber crimes, these issues remain unaddressed and require more time and budget. This study demonstrates the need to improve the quality of the investigation phase through enhancing the process of evidence analysis. Such improvement includes producing useful evidence such as predicting the intention and establishing similar strategies of cyber crime cases to discover similar cases, thereby reducing the effort and processing cost during the investigation phase.

1.2 Problem Statement

The analysis phase clarifies the intentions and methodology of the attack and provides a feedback to improve the security tools (Pilli et al., 2010). The analysis phase support the investigation phase of network forensics, and has a knowledge gap, which reconstructing useful evidence of a cyber crime is difficult (Baryamureeba and Tushabe, 2004; Freiling and Schwittay, 2007; Almulhem, 2009; Pilli et al., 2010). This gap caused by the vagueness of the analysis phase processes. The analysis phase is challenging because it provides detailed information on the intention and strategy of the attack. Therefore, generating useful evidence in the investigation phase to measure the impact of a cyber crime is difficult and more costly in terms of capital and resources.

The main challenge faced by this research is to design a series of processes that analyzes the evidence of an attack to increase the overall speed of the investigation process. This challenge is phrased as the following research problem:

How can an efficient framework that analyzes attack evidence for network forensics be designed?

Addressing this problem requires a new framework to analyze evidence. The framework proposed in this study involves designing a set of processes and algorithms in the network forensic analysis phase that uses cyber crime evidence to reconstruct cyber crime intentions and establish similar strategies. This framework aims to directly analyze the cyber crime evidence and to maximize the probability value of retrieving similar cases.

1.3 Research Motivation

According to CERT et al. (2011), cyber crime attacks incurred an average monetary loss of $123,000 per organization in the USA in 2011. Ponemon (2011) reported that the annual cost of solving cyber crimes is $5.9 million. Ponemon's study is based on a representative sample of 50 organizations in various industrial sectors in the USA. The cost incurred by cyber crimes per company ranges from $1.5 million to $36.5 million each year, as shown in Figure 1.1. In reality, a strong relationship exists between the time required to resolve a cyber crime and the cost. Based on a previous study (Ponemon, 2011), cyber crimes could become costly if they are not resolved quickly. Current investigation techniques are very costly and time consuming because extensive effort is required to analyze the overwhelming amount of evidence presented in each cyber crime case. In addition, gathering useful evidence is difficult because most techniques utilize active and reactive processes to analyze cyber crimes; such processes start right after the detection of the cyber crime.

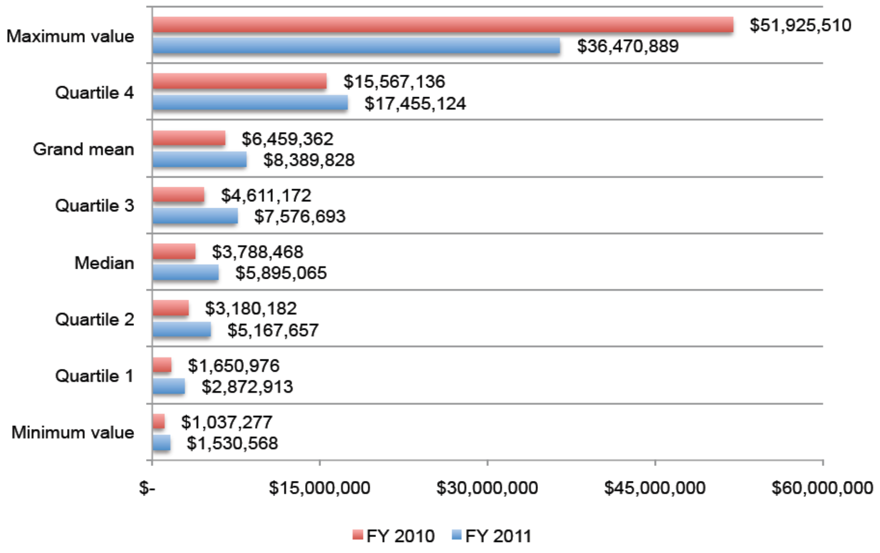

Figure 1.1: Key Benchmark Sample statistics on the Annualized Cyber Crime Cost, (Ponemon, 2011)

Figure 1.2 indicates that the average time to resolve a cyber attack is 18 days, with an average cost of $415,748 for the participating organizations over the 18-day period.

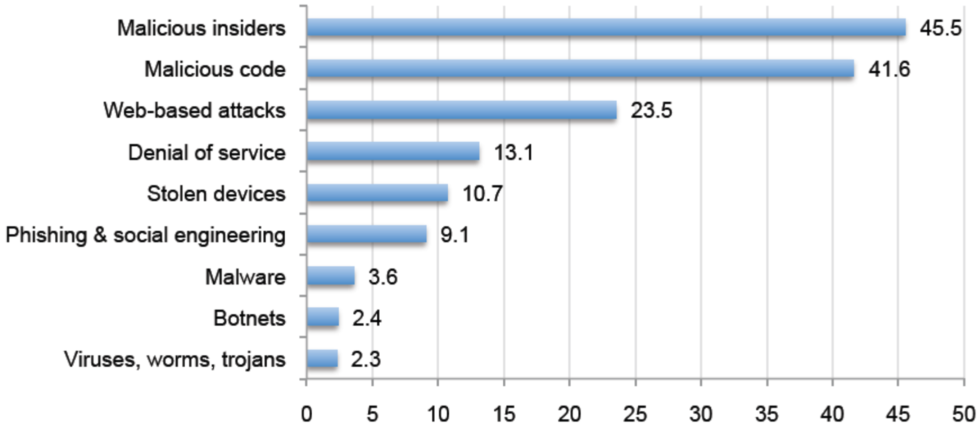

Figure 1.2: Average Days to Resolve an Attack for Seven Attack Types, (Ponemon, 2011)

In general, the amount of evidence collected by network forensic tools is huge. Most organizations do not pursue legal actions against perpetrators of cyber crimes because of the lack of useful evidence and sufficient information to prosecute the perpetrators (CERT et al., 2011). Defensive security approaches such as Intrusion Detection System (IDS) and Intrusion Protection System (IPS) were developed recently to detect, prevent, and establish a perspective of network attacks. Cyber crime evidence must be analyzed more intensively to generate clear and useful evidence and establish a more suitable decision in the investigation phase. Attack intentions should be predicted and similar attack strategies should be identified in the analysis phase. Maximizing the probability of retrieving similar cases is also helpful because it minimizes the amount of time and processing cost required to resolve cyber crime cases by analyzing the most similar cases.

Most attack analysis approaches are based on alert correlation techniques. These techniques are connected to network forensic tool for assistance such as IDS to understand and analyze the cyber crime occurrence. The drawback of most of these techniques is that they are developed to prevent future attacks and minimize damage and not to analyze the cyber crimes through network forensics (Wei and Thomas, 2008; Huang et al., 1999; Damiano et al., 2009; Wang and Peng, 2009). Thus, innovative methods and techniques are needed in the analysis of the attacks to increase the amount of evidence by establishing the attack intention and strategy in advance, and to help investigators in their decision making and in resolving cyber crimes (Almulhem, 2009).

The study of attack intentions provides more details about the features of the crime and the behavior of the attacker. The features are distinctive characteristics of the attack. It includes IP addresses, ports, type of services and protocol, etc. Attack intentions can be utilized as a useful piece of evidence to enhance the investigation process through decision making and to apprehend the real perpetrator. The most common techniques for attack intention analysis depend on determining the intention from the attack path as reported by Peng et al. (2009), Wang and Peng (2009), Wu et al. (2009), Feng et al. (2011), and Hao et al. (2011). The drawback of these techniques is that they are not suitable for large amounts of evidence limited to a specific type of evidence, and cannot present all the attack intentions. These techniques work only with a specific type of attack, such as Distributed Denial of Service (DDOS) attacks. Furthermore, these techniques were developed to enhance IDS, not to specifically analyze evidence in network forensics.

Determining the attack strategy allows network forensic investigators to easily draw a possible comprehensive frame of the cyber crime case. Thus, establishing similar attack strategies maximizes the probability of retrieving similar cases. Cyber crime attack strategies have become increasingly sophisticated, which makes the identification of an accurate attack strategy extremely difficult (Wei and Thomas, 2008). One example is a multi-stage attack in which the evidence is distributed among various sources. Most attack strategy techniques depend on alert correlation (Wei and Thomas, 2008; Damiano et al., 2009; Peng et al., 2009; Wang and Peng, 2009). However, these techniques have a number of limitations. For instance, they require a large number of predefined attributes, difficult to implement, and employed only to prevent future attacks and minimize damage.

The processes in network forensic evidence analysis require new techniques for different types of attacks. Attack evidence analysis requires computational techniques, such as a mathematical methods and graph theories for examining similar cases, thereby reducing the investigative efforts. This research is conducted to retrieve similar cyber crime cases by analyzing evidence through the identification of attack intention and establishment of similar attack strategies. The analysis of evidence helps investigators eliminate similar cases, which reduces the time and cost of investigation and allows investigators to solve new cases by analyzing the results of previous similar cases.

1.4 Goal, Objectives and Scope of the Research

The main goal of this research is to propose a new framework to analyze digital evidence and increase the possibility of obtaining cyber crime evidence. The proposed framework can increase the probability value of retrieving similar cases. The work is divided into the following three main objectives:

- To identify the attack intentions of a cyber crime through a method that focuses on the reason for the attack for uncertain intentions. Attack intentions will be analyzed to better understand the motivation behind cyber crimes.
- To establish similar cyber crime strategies through evidence and pre-analyzed attack intentions. The established value will be used to increase the possibility of obtaining evidence to retrieve similar cyber crime cases.
- To evaluate the new evidence analysis framework. The new framework will show the significance of identifying intentions and establishing similar cyber crime strategies in increasing the probability value of retrieving similar cases during the investigation process.

The scope of this research focuses on the analysis phase of the generic process model for network forensics proposed by Pilli et al. (2010) because the model is comprehensive and is based on various existing digital forensic models. This research analyzes cyber crime evidence to efficiently and clearly establish the attack intentions and similar attack strategies, thereby supporting the investigation phase. Figure 1.3 presents the general research overview that indicates the scope of the study.

The main assumption of this research is that evidence collection and classification are predefined in the previous phase. Moreover, the research assumes that the depository of proactive network forensics is utilized to preserve and restore cyber crime evidence and analysis results. Thus, this research presents the components of each evidence classification and the proactive forensics depository to show the integrity and the dependency relationship between the analysis phase and other phases in the general network forensic model.

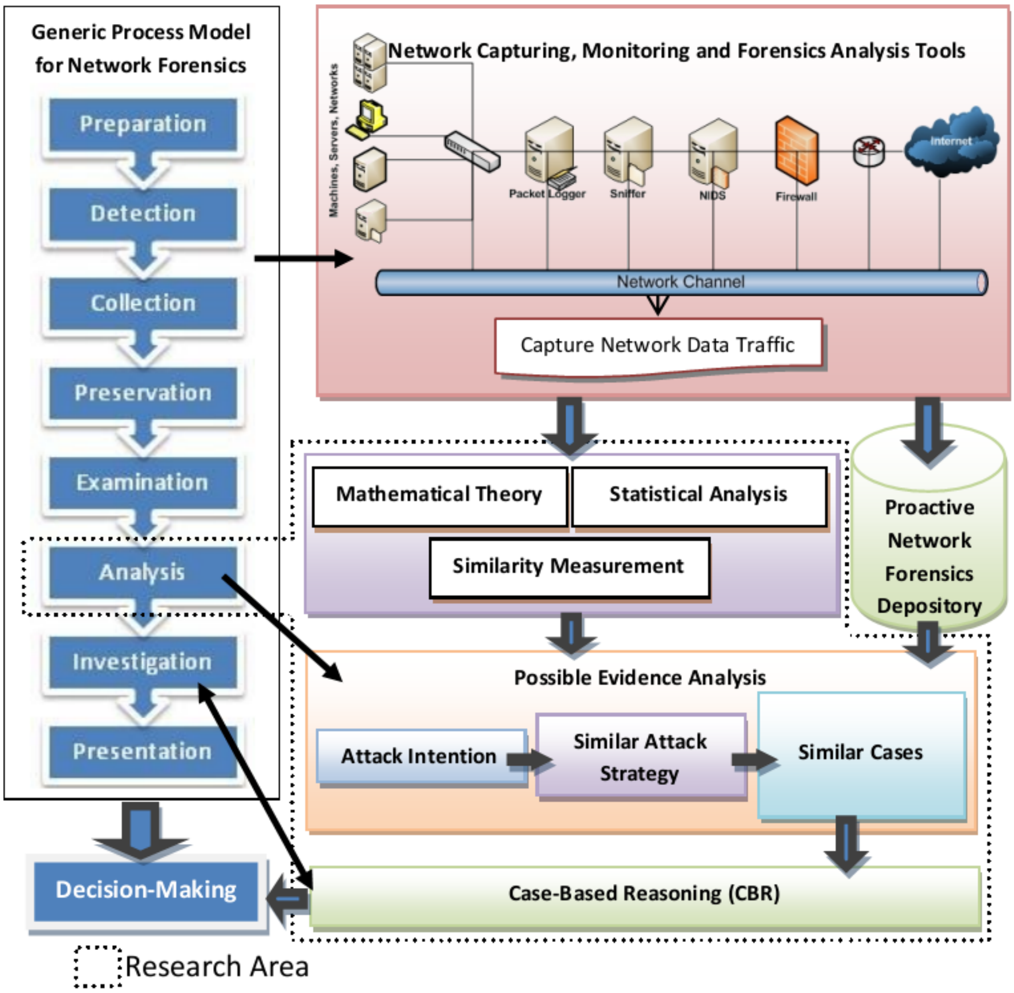

Figure 1.3: Scope of the Research

1.5 Research Methodology

This research applies a series of steps by combining methods from the statistical and similarity measurements to analyze evidence. This research proposes a new framework to analyze cyber crime evidence from different perspectives to generate useful evidence that can be utilized to improve the investigation phase. Evidence is analyzed to identify the intentions of the attack and establish similar attack strategies. The identified intentions and strategies are then compared with pre-existing documented cases to increase the possibility of matching cyber crime evidence with potential similar cases, thereby maximizing the probability of discovering a precise match and increasing the percentage of similarities among cases. Figure 1.4 shows the research process, which includes the input process, processing, and output process.

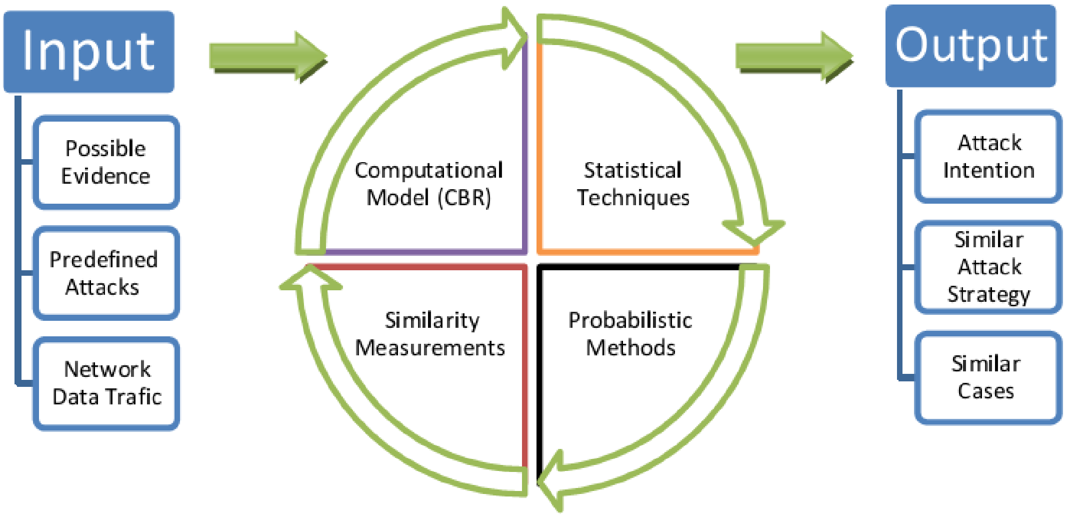

Figure 1.4: Process of the Research Overview

The research is conducted in four phases, as shown in Figure 1.5. The first phase is a preliminary study of the research problem, which is analyzing cyber crime evidence. The second phase determines the requirements that support the evidence analysis, such as tools, data sets, theories, and techniques. The theoretical framework is established, as shown in Figure 1.6. The requirements determined in this phase verify the integrity of the analysis phase with other phases in network forensics.

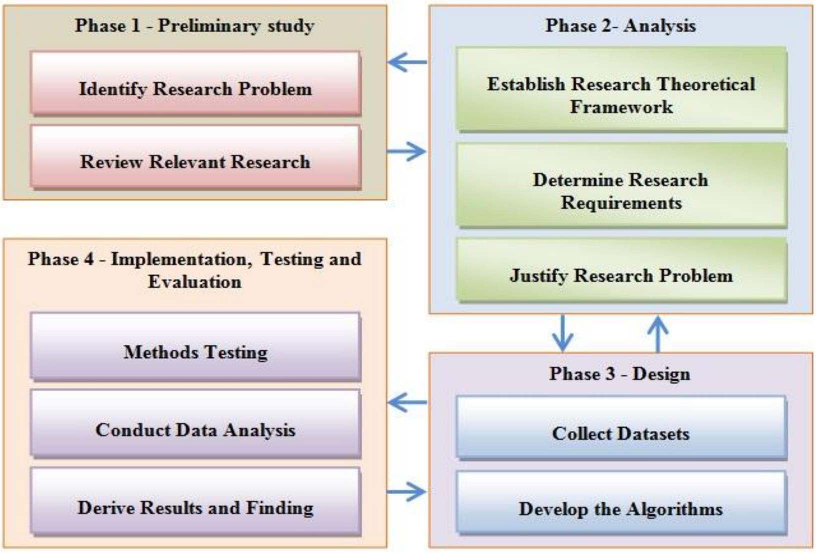
Phase 1 - Preliminary study
Identify Research Problem
Review Relevant Research
Phase 2- Analysis
Establish Research Theoretical Framework
Determine Research Requirements
Justify Research Problem
Phase 3 - Design
Collect Datasets
Develop the Algorithms
Phase 4 - Implementation, Testing and Evaluation
Methods Testing
Conduct Data Analysis
Derive Results and Finding

Figure 1.5: Research Methodology Process

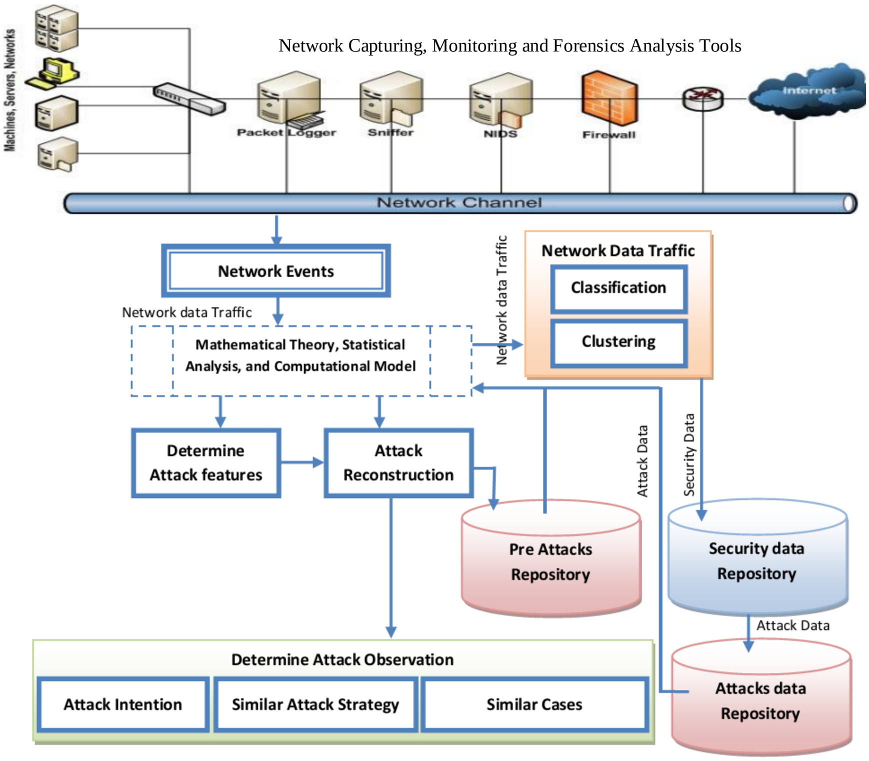

Figure 1.6: Theoretical framework

The third phase is the design of the components of the proposed framework. The data set is collected in this phase by capturing the network traffic from our university laboratories and monitoring for any suspicious attack. The proposed framework also utilizes the general network forensic datasets from the Honeynet Project (Werner, 2010). The datasets are manipulated and analyzed by using selected network forensic tools such as Wireshark (Wireshark, 2011) and Snort (Snort, 2010). The main purpose of this phase is to design suitable algorithms to analyze evidence. This phase also establishes a new algorithm that predicts attack intentions by combining the mathematical Dempster–Shafer (D–S) evidence theory with a probabilistic technique through a causal network. Furthermore, an extended algorithm is designed from the proposed attack intention algorithm to establish similar attack strategies through cosine similarity measurements.

Lastly, once the components of the attack intentions and strategies have been established, the proposed framework is applied to each component. The CBR technique is utilized to identify similar cases among the new cyber crime cases to help investigators solve cyber crimes efficiently. The proposed framework is evaluated to emphasize the significance and efficiency of the attack intention and strategy analysis in increasing the possibility of obtaining evidence as well as maximizing the probability of identifying similar cases.

1.6 Contributions of the Research

This research contributes a new framework for analyzing attack evidence to predict attack intentions and establish similar strategies, thereby increasing the possibility of obtaining cyber crime evidence. This contribution makes the investigation process even more effective by maximizing the probability of identifying similar cases as well as helping in apprehending the real crime perpetrator. The contributions are as follows:

- A new algorithm called Attack intention Analysis (AIA) to analyze attack intentions by combining the mathematical D–S evidence theory with a probabilistic technique through a causal network. The algorithm is utilized to predict attack intentions, thereby providing useful evidence.
- A similarity process model to estimate attack strategy when the intentions behind a cyber crime are unknown. The model utilizes cosine similarity measurements based on evidence classification to identify similar cyber crime strategies.
- A new algorithm called Similarity of Attack Strategy (SAS) to establish cyber crime strategies based on the intentions behind a cyber crime. The algorithm integrates the similarity process model and pre-analyzed attack intentions to expedite the investigation process by maximizing the ranking of similar cyber crime cases.
- A new framework to analyze evidence and evaluate the efficiency of identifying intentions and establishing similar cyber crime strategies through network forensic analysis tools and the CBR technique. The new framework retrieves similar cyber crime cases with a high probability value.

1.7 Thesis Outline

This chapter presents the basic concepts and states the problem. In addition, this chapter presents the scope, goal, and objectives as well as the motivations, methodology, and contributions of this research. The remainder of this thesis provides the background and details of attack evidence analysis. Chapter two presents a literature review on the four domains of this research: current network forensic approaches, analysis phase in network forensics, attack intention analysis methods, and attack strategy analysis methods. The chapter focuses on the main challenges faced by network forensic analysis approaches and how previous studies addressed the disadvantages of the network forensic analysis process models and techniques.

Chapter three presents the proposed framework to analyze evidence in network forensics. This chapter describes the components of the proposed framework and clarifies the theoretical framework that includes all the components of the proposed framework. It discusses all the proposed algorithms for attack intentions and similar attack strategies. Chapter four presents the design and implementation of the proposed framework components. It discusses all the components for attack intentions and similar attack strategies

Chapter five presents the experimental results of the proposed framework. The chapter evaluates the efficiency of the proposed framework and reveals the significance of its components. Chapter six concludes this thesis and summarizes the main contributions of this research. The chapter provides suggestions for future studies.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

This chapter reviews literature on the analysis of cyber crime evidence in network forensics and cites previous and current studies related to the aforementioned research field. It focuses on previous network forensic approaches and how the analysis phase was presented in these approaches. It shows how this research fills the gap in network forensic knowledge, which depends on an efficient analysis of evidence. The chapter also reveals the necessity of this research and discusses the main tools and techniques utilized in the network forensic approach. Attack intentions and strategy methods are considered the main factors that improve the decision making process during the investigation phase in network forensics.

This research addresses the related studies in four main parts. The first part, which will be explained in the next section, discusses cyber crime and the fundamentals of network forensics, such as the definitions, main challenges, and network forensic analysis and monitoring tools. The second part, Section 2.3, introduces the current network forensic process models. These models are also compared to justify the proposed framework. The third part, Section 2.4, discusses the analysis phase in network forensics and the intentions and strategies of the attack analysis methods. The fourth part, Section 2.5, describes the implementation techniques utilized by network forensic approaches. These techniques are also compared to justify the Case-Based Reasoning (CBR) technique as a potential solution to retrieve similar cyber crime cases as well as to evaluate the efficiency of the proposed framework.

2.2 Fundamentals of Network Forensics

This section defines cyber crime as well as network forensics with its main challenges. It presents the main network forensic analysis and monitoring tools.

2.2.1 Cyber Crimes

"Cyber crime" refers to any crime that involves computer or network communication which may have been used to establish the crime or which may be the target. The United States Department of Justice defines computer crime as "any violation of criminal law that involves knowledge of computer technology for their perpetration, investigation, or prosecution" (Parker et al., 1989). Most cyber crimes occur because of the proliferation of different types of attacks, such as Trojan, phishing, and spoofing attacks, in computer networks.

According to Gandhi et al. (2011), the nature of the attack and the motive behind it should be identified to better understand cyber crime and to resolve it within a shorter time and at a lower cost. Most organizations utilize traditional investigation techniques, which are typically reactive, to solve cyber crimes, which could damage evidence that was gathered and analyzed after the occurrence of the cyber crime. Traditional techniques result in costly and time-consuming investigation processes. The methods in the investigation process need to be improved to better understand cyber crimes because it is a complex (Casey, 2005). For example, a proactive approach should be implemented in gathering and analyzing evidence to expedite the investigation process.

2.2.2 Network Forensics

According to Almulhem (2009), network forensics extends from network security and computer forensics; it works with the laws and guiding principles indicated in the judicial system, as shown in Figure 2.1. Traditionally, forensic specialists work hand in hand with law enforcement officers. The former utilizes scientific techniques to collect, examine, analyze, and document digital evidence from digital sources and network security programs. These techniques are incorporated into firewalls, intrusion detection systems, or network devices such as routers and switches to uncover facts related to cyber crime (Patel et al., 2011; Pilli et al., 2010).

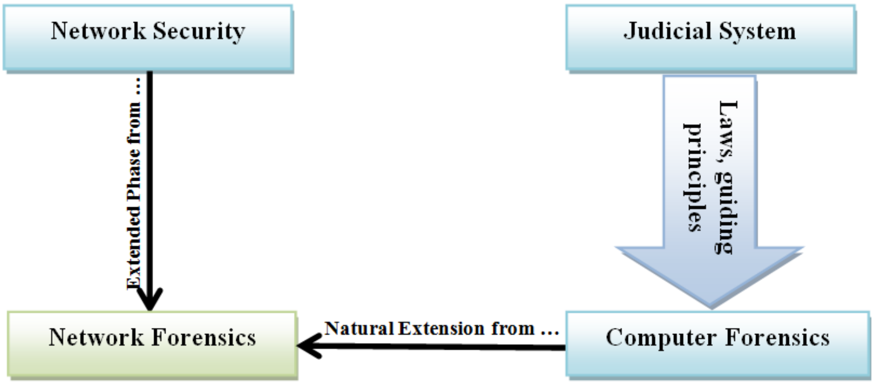

Figure 2.1: Network Forensics Locations

In early 2001, the first Digital Forensics Research Workshop (DFRWS) (Palmer, 2001) defined network forensics as “the use of scientifically proven techniques to collect, fuse, identify, examine, correlate, analyze, and document digital evidence from multiple, actively processing and transmitting digital sources for the purpose of uncovering facts related to the planned intent, or measured success of unauthorized activities meant to disrupt, corrupt, or compromise system components as well as providing information to assist in response to or recovery from these activities.” This definition indicates that the main phases of network forensics are collection, fusion, identification, examination, correlation, analysis, and documentation of digital evidence. These phases guide other researchers in proposing new approaches for network forensics. The identification of the deliberate intent behind cyber crimes is the main goal of network forensics.

Network forensic systems as reported by Pilli et al. (2010) can be classified depending on the three characteristics indicated in Figure 2.2. In general, there are two approaches in network forensics: proactive and reactive. Proactive network forensics is a new approach in live investigation that deals with the phases of network forensics during an attack. In contrast, reactive network forensics is a traditional approach that deals with cyber crime cases after a period of time, which consumes a considerable amount of time during the investigation phase. As reported by Alharbi et al. (2011a), Grobler et al. (2010), and Simson L. (2010), proactive forensic approaches reduce the time and cost of investigation by identifying potential evidence and reducing the resources needed in the investigation phase. These approaches are utilized in the preliminary analysis of a cyber crime and help improve and accelerate the decision making process.

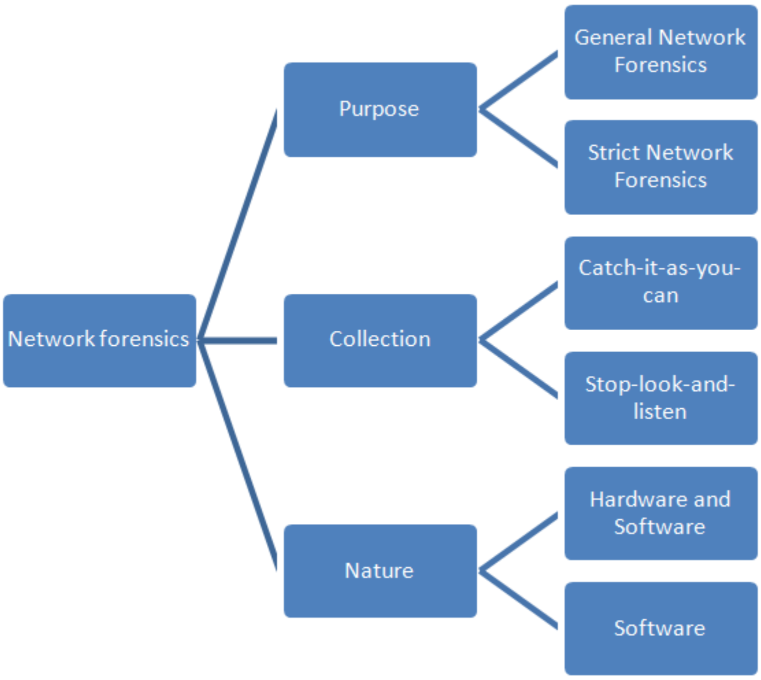

Figure 2.2: Network Forensics System Classifications

This research proposes a new framework to analyze evidence. The framework is proactive if it analyzes evidence and provides sufficient information on the intention behind the cyber crime and similar strategies and cases in the investigation phase during the occurrence of the cyber crime through the proactive depository.

2.2.3 Main Challenges in Network Forensics

Network forensics involves several challenges such as various data sources, data granularity, data integrity, data as legal evidence, privacy issues, and data analysis (Almulhem, 2009). Pilli et al. (2010) also reported the following challenges in network forensics, such as

- Identifying useful network events and recording the minimum representative attribute for each event

- The need for a full capture of the malicious behavior to reconstruct the attack behavior
- Integrated and aggregated logs and traffic data from various tools
- Distinguishing legitimate traffic from attack traffic by extracting the features of the patterns of anomalous network events
- Classification and clustering of network events
- Parsing and analysis of complex protocols
- Reconstruction methods that were utilized to understand the intention and strategy of the attacker
- Accurate determination of the geographical location of the attacker by building a topological database and IP location mapping

The above mentioned challenges indicate knowledge gaps in network forensics. This study proposes a solution to fill the gap in the analysis of evidence to identify the intentions behind the crime and establish similar cyber crime strategies. The solution aims to better understand the motive and methodology of cyber crimes, which will help improve the quality of the investigation process.

Network forensic processes are distributed among the general phases of evidence collection, preservation, analysis, and investigation. The investigation phase depends on the analysis phase in providing useful evidence of the cyber crime. Network forensic investigation is generally complex and very costly (Casey, 2005), and analyzing network data traffic is time consuming, error prone, and difficult (Simson L., 2010; Lin et al., 2009; Casey, 2007; Mathew et al., 2006; Yasinsac A. and Manzano Y, 2002).

Even the best attack detection and prevention techniques, such as IDS and IPS, also have limitations which are exploited by the attackers and allow the attackers to learn new strategies to circumvent these techniques (Benjamin et al., 2005). For example, a buffer overflow attack depends on a part of the execution code at a period of time during the operation of the program, which produces a change in the attack strategy. Caloyannides (2009) claims that a smart attacker has sufficient knowledge and skills to remove evidence of a crime, which then makes the identification of the real perpetrator difficult. The main reason for the difficulty, as reported by Brian (2006), is the complexity of the attacker's techniques, such as using Trojan files to modify the nature of the network forensic tools. The network security field continues to develop techniques for analyzing attack behavior based on the intentions behind the crime (Peng et al., 2009). Most studies in the fields of IDS and IPS depend on alert correlation and intrusion scenario techniques to understand and analyze attack behavior, which still have the above mentioned limitations.

Generally, observing and analyzing sophisticated attacks are difficult (Zhijie et al., 2008). Most multi-stage attacks generate huge volumes of alerts through IDS, which make the attack strategy difficult to recognize during the analysis process. Several researchers such as Alserhani et al. (2010) believe that at present, no technique can efficiently detect a multi-stage attack.

Anti-forensic methods are another challenge in network forensics. Data concealment and overwriting techniques hinder network forensics tools from accomplishing their purpose, which lengthens the investigation time. Anti-forensic methods also affect the quality of evidence collection and the accuracy of crime detection (Garfinkel, 2007). According to Alharbi et al. (2011b), the main reason for shifting to a proactive approach is to minimize the effects of anti-forensic methods.

Attack analysis is a critical and challenging task in security management (Qin and Lee, 2004). The limited capability of security sensors and network monitoring tools makes attack observation inaccurate and incomprehensible. This research believe that no complete library for all the possible attack strategies in network security exists, which increases the difficulty of the analysis of attack evidence and the recognition of the attack intention and strategy.

There is a large number of attack methods, make pattern recognition more difficult. According to Huang et al. (1999), changing attack patterns is a challenge in attack analysis, which also affects the network forensic process, especially for a large-scale distributed infrastructure. The growing amount of cyber crime evidence makes collecting significant evidence for the analysis process difficult because the performance of the network forensic tools changes frequently, as mentioned by Carrier (2009). However, Merkle (2008) stated that analyzing raw traffic in network forensics with the increasing amount of evidence is a complex task. Investigators need to identify and classify evidence to conduct an efficient analysis.

Recent studies, such as those conducted by Mouhtaropoulos et al. (2011), Alharbi et al. (2011b), Grobler et al. (2010), and Rebecca (2005), reported that the analysis and implementation of network forensic techniques in either the private or public sector encounter numerous difficulties. For example, these techniques require expertise and a certain level of network forensic standardization. Law enforcement officers and academic researchers need to collaborate in advance to improve and enhance the body of network forensic knowledge. Rogers and Seigfried (2004) believe that network forensics needs to focus more on the education, training, and certification sectors to improve inadequate network forensic processes.

2.2.4 Network Forensics Analysis and Monitoring Tools

Network forensic processes aim to resolve cyber crime cases and select a suitable response for such cases to discover the real perpetrators. The key to achieve this goal is network traffic, which is captured, recorded, and analyzed through network forensics to collect evidence for the analysis of cyber crimes. These processes require a particular tools to help investigators establish a suitable decision in response to a cyber crime (Pilli et al., 2010a).

According to Pilli et al. (2010a, 2010b), Network Forensic Analysis Tools (NFATs) are classified into two categories based on the source code, i.e., proprietary and open source tools, as shown in Figure 2.3. The same authors classified the Network Security and Monitoring (NSM) tools based on the purpose of these tools, namely, for packet capturing, statistics, pattern matching, manipulation, fingerprinting, and IDS, as shown in Figure 2.3.