Explainable Intelligence for Comprehensive Interpretation of Cybersecurity Data in Incident Management

by

Neda Afzaliseresht

StudentID: s4576843

Thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The Institute for Sustainable Industries & Liveable Cities
Engineering and Science
VICTORIA UNIVERSITY

July 2022

VICTORIA UNIVERSITY

Abstract

The Institute for Sustainable Industries & Liveable Cities
Engineering and Science

Doctor of Philosophy

by Neda Afzaliseresht
StudentID: s4576843

On a regular basis, a variety of events take place in computer systems: program launches, firewall updates, user logins, and so on. To secure information resources, modern organisations have established security management systems. In cyber incident management, reporting and awareness-raising are a critical to identify and respond to potential threats in organisations. Security equipment operation systems record 'all' events or actions, and major abnormalities are signaling via alerts based on rules or patterns. Investigation of these alerts is handled by specialists in the incident response team.

Security professionals rely on the information in alert messages to respond appropriately. Incident response teams do not audit or trace the log files until an incident happens. Insufficient information in alert messages, and *machine-friendly* rather than *human-friendly* format cause cognitive overload on already limited cybersecurity human resources. As a result, only a smaller number of threat alerts are investigated by specialist staff and security holes may be left open for potential attacks.

Furthermore, incident response teams have to derive the context of incidents by applying prior knowledge, communicate with the right people to understand what has happened, and initiate the appropriate actions. Insufficient information in alert messages and stakeholders' participation raise challenges for the incident management process, which may result in late responses. In other words, cybersecurity resources are overburdened due to a lack of information in alert messages that provide an incomplete picture of a subject (incident) to assist with necessary decision making. The need to identify and track local and global sources in order to process and understand the critical elements of threat information causes cognitive overload on the company's currently limited cybersecurity professionals.

This problem can be overcome with a fully integrated report that clarifies the subject (incident) in order to reduce overall cognitive burden. Instead of spending additional time to investigating each subject of incident, which is dependent on the person's expertise and the amount of time he has, a

detailed report of incident can be utilised as an input of human-analyst. If cyber experts' cognitive loads can be reduced, their response time efficiency may improves. The relationship between achieving incident management agility through contextual analytical with a comprehensive report and reducing human cognition overload is still being studied. There is currently a research gap in determining the key relationships between explainable Artificial Intelligence (AI) models and other technologies used in security management to gain insight into how explainable contextual analytics can provide distinct response capabilities. When using an explainable AI model for event modelling, research is necessary on how to improve self and shared insight about cyber data by gathering and interpreting security knowledge to reduce cognitive burden on analysts.

Due to the fact that the level of cyber security expertise depends on prior knowledge or the results of a thorough report as an input, explainable intelligent models for understanding the inputs have been proposed. By enriching and interpreting security data in a comprehensive human-readable report, analysts can get a better understanding of the situation and make better decisions. Explainable intelligent models are proposed in cyber incident management by interpreting security logs and cybersecurity alerts, and include a model which can be used in fraud detection where a large number of financial transactions necessitates the involvement of a human in the analysis process.

In cyber incident management application, a wide and diverse amount of data are digested, and a report in natural language is developed to assist cyber analysts' understanding of the situation. The proposed model produced easy-to-read reports/stories by presenting supplementary information in a novel narrative framework to communicate the context and root cause of the alert. It has been confirmed that, when compared to baseline reports, a more comprehensive report that answers core questions about the actor (who), riskiness (what), evidence (why), mechanism (how), time (when), and location (where) that support making real-time decisions by providing incident awareness. Furthermore, a common understanding of an incident and its consequences was established through a graph, resulting in Shared Situation Awareness (SSA) capability (the acquisition of cognition through collaboration with others).

A knowledge graph, also known as a graph to semantic knowledge, is a data structure that represents various properties and relationships between objects. It has been widely researched and utilised in information processing and organisation. The knowledge graph depicts the various connections between the alert and relevant information from local and global knowledge bases. It interpreted knowledge in a human-readable format to enable more engagement in the cyber incident management. The proposed models are also known as explainable intelligence because they can reduce the cognitive effort required to process a large amount of security data. As a result, self-awareness and shared awareness of what is happening in cybersecurity incidents have been accomplished. The analyses and survey evaluation empirically demonstrated the models' success in reducing significant overload on expert cognition, bringing more comprehensive

information about the incident, and interpreting knowledge in a human-readable format to enable greater participation in cyber incident management.

Finally, the intelligent model of knowledge graph is provided for transaction visualisation for fraud detection, an important challenge in security research. As with the same incident management challenges, fraud detection methods need to be more transparent by explaining their results in more detail. Despite the fact that fraudulent practices are always evolving, investigating money laundering based on an explainable AI that uses graph analysis, assist in the comprehension of schemes. A visual representation of the complex interactions that occur in transactions between money sender and money receiver, with explanations of human-readable aspects for easier digestion is provided. The proposed model, which was used in transaction visualisation and fraud detection, was highly regarded by domain experts. The Digital Defense Hackathon in December 2020 demonstrated that the model is adaptable and widely applicable (received first place in the Hackathon competition).

Declaration of Authorship

I, Neda Afzaliseresht, declare that the PhD thesis entitled *Explainable Intelligence for Comprehensive Interpretation of Cybersecurity Data in Incident management* is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

"All research procedures reported in the thesis were approved by the [Victoria University Human Research Ethics Committee - HRE20-001]."

Signature  Date 01/02/2021

Dedication

To those who raised me personally (my parents), professionally (my supervisors), and patiently (my wonderful husband) – Majid.

Acknowledgements

I would like to thank the participants of the study for their invaluable insights, making this research possible,

- my supervisors, Prof. Yuan Miao, Prof. Hua Wang and Dr. Qing Liu, for their unwavering support, guidance, and encouragement throughout the project,
- the Csiro- Data61 team for collaborating in the project, particularly Dr. Qing Liu, the member of CSIRO's Data61 who kindly helped and supported this project,
- and my family and friends for their warm support throughout this process. In particular my father, Masoud Afzaliseresht, and mother, Azam Sheibani for encouraging me to study at an outstanding university, and my awesome husband, Majid Afzalirad, for regularly encouraging me to act in my own best interest!
- other members of the professional community who have been most kind in offering their thoughts and advice,

Publications

This section includes the list of peer-reviewed academic articles that I have published during my PhD research. Elements of these articles are included in this thesis particularly in Chapters 3, 4, 5 and 6 for introducing the explainable intelligence models. The inclusion of the papers is highlighted in the relevant section within the thesis.

Published Papers:

1. N AfzaliSeresht, Y Miao, S Michalska, Q Liu, H Wang.*"From Logs to Stories: Human-Centred Data Mining for Cyber Threat Intelligence"* **IEEE Access 19089-19099** *Q1 Journal (IEEE Access 8, 2020)* [1].

2. N AfzaliSeresht, Q Liu, Y Miao. *"An Explainable Intelligence Model for Security Event Analysis"* **Australasian Joint Conference on Artificial Intelligence.** *Rank: B (Springer, Cham, 2019)* [2].

3. N AfzaliSeresht, Y Miao, Q Liu. *"Design a Storytelling Model from Security Events"* **NSS 2019 International Conference on Network and System Security.** *Rank: B ({Poster}, 2019).*

4. N AfzaliSeresht, Y Miao, Q Liu, A Teshome, W Ye. *"Investigating Cyber Alerts with Graph-Based Analytics and Narrative Visualization"* **24th International Conference Information Visualisation (IV) 20.** *Rank: B (IEEE, 2020)* [3].

Under Review Papers:

1. N AfzaliSeresht, Y Miao, Q Liu. *"The Empirical Analysis of Cognitive Load Reduction of the Security Incidents Interpretation using Storytelling Approach: Questionnaire Survey and Analysis"* **Computers & Security - Journal - Elsevier.** *Q1 Journal {Under Review}.*

2. N AfzaliSeresht, Y Miao, Q Liu, H Wang *"Investigating Money Laundering by Graph-Based Analysis and Scenario-Based Matching"* **Elsevier-Expert system and application Journal.** *Q1 Journal {Under Review}.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Never underestimate the likelihood of an attack on the organisation. Attackers and intruders are constantly devising new tactics. To respond effectively, businesses must be adaptable, which means exchanging pertinent facts with the proper people at the right time. Analysts are put in challenging situations where they must make quick decisions based on massive amounts of data that exceed their human capabilities. Intelligence models are needed as an alternative to overcome this limitation. In terms of security decisions, fraud detection and characteristics, current systems must adapt and behave more like humans, a concept known as the cognitive model, which integrates technology solutions that assist decision systems with the cognitive processes of analysts [7].

Cognitive security systems can mimic human thought processes to solve complex problems such as threat detection by use of AI technologies [8]. Cognitive security systems aim to improve cybersecurity operations by integrating various technologies [7]. As a result, it is the most appropriate term to use to characterise our proposed approach. The phrase "cognitive security" refers to the use of cognition that affect cybersecurity decisions and business outcomes. I looked at how to generate cognition in cyber and financial environments, including both self and shared situation awareness (SA) because awareness in a multi-actor operation, like cyber incident management or fraud detection needs both self and shared SA. An organisation's created levels of cybersecurity awareness enable the identification of the type of threats and attacks, maintenance of acceptable levels of security, and the design of proactive plans to confront any lurking threats [7]. The level of financial landscape awareness is also beneficial in detecting frauds and preventing criminal behaviour from continuing.

This chapter explains more about the aim and objectives of the investigation, and further details about the problem.

1.1 Research Background and Motivation

Cybersecurity incident response is a continual process of incident management in which human beings have to be involved in the analysis process through a dedicated manual effort to identify, investigate, respond and learn from potential cybersecurity incidents in a timely and cost-effective manner [9]. This process plays an essential role in enterprises as preventing all breaches is not feasible, leaving the security hole open for potential attacks. Even though many automated technologies are available, incident response still involves many manual processes. A quick incident response to a cybersecurity assault can help businesses prevent financial loss while safeguarding their reputation and competitive edge [9]. Current approaches that: (i) rely on the knowledge and skills of security professionals, or (ii) are restricted in the depth of the insight supplied, do not allow for the discovery of the root causes of the incidents, hence preventing the appropriate response to possible threats [1].

The rapid rise of information technology (IT) has resulted in global security concerns. The use of technologies introduces new challenges (e.g., higher risks of data exposure, sensitive corporate data, phishing, electronic fraud, impersonations and information security issues). New risks and security attacks have emerged as a result of technological improvements, with cybercrime serving as a sophisticated international threat on an industrial scale. As a result, new and complex cybersecurity scenarios involving massive amounts of data and several attack vectors have emerged, possibly exceeding security analysts' cognitive ability to detect unseen trends in the acquired information and data. There is a significant chance that every network already contains hidden threats. Because technologies advance at such a rapid speed that by the time a new threat signature is found, no cybersecurity system is impermeable or capable of detecting or stopping every potential threat.

In an effort to deal with cybersecurity attacks and data breaches, human beings have to be involved in the analysis process, where humans act as security sensors to detect attacks, which known as human-as-a-security-sensor by Vielberth and et al. [10]. Human cybersecurity professionals, such as analysts, act as a barrier between malicious actors and an organisation's data that must be safeguarded [11]. Analysts are responsible for determining whether a network is under attack, how to mitigate the attack and, in many circumstances, how the attack was executed. Cyber analysts work for major corporations. They face three main questions to efficiently and effectively detect, examine, and respond to cybersecurity incidents: (1) 'What happened?', (2) 'Why did it happen?' and (3) 'What should I do?' [12]. Analysts must get SA of the critical components of network defense. The capability of a Security Operations Centre (SOC) or a Security Incident Response Team (CSIRT) to quickly collect, integrate and analyse all data relevant to a cybersecurity issue in order to digest enormous amounts of data and identify hidden linkages and dependencies, provides SA.

According to a cognitive perspective of SA, in a complex world, a human operator will learn to interpret the appropriate, essential elements of information, attempt to comprehend their context and use this understanding to make near-future predictions about the state of the environment [11]. The term cyber situation awareness (CSA) applies SA to cyber defense. It refers to the comprehension of essential factors, which promotes efficient and timely predictions about the state of the environment [13]. In this context, cognitive sciences, particularly CSA [11], can assist security analysts in establishing actions inside incident management in less time and with greater efficiency. In terms of cognitive sciences, the perspective in the field of cybersecurity is to find ways to improve human-factor capacities, which are required for incident management, especially in security systems that are generated by technologies like cloud, mobile, IoT, and social networks that all generate large amounts of data. Cyber-cognitive situation awareness (CCSA) is a new term that focuses on the cognitive processes that support SA [13]. CCSA is influenced by various factors, such as protocol investigation, security practices, the information generated by computer systems, security blogs, vulnerability bulletins, and security professionals' expertise based on the tasks they perform on a regular basis.

The interdisciplinary scientific study of psychology, computer science, linguistics, philosophy, and neuroscience to better understand the human mind is known as cognitive sciences [14]. It is now critical to look at the role of cognitive sciences in enhancing human abilities for cybersecurity tasks. In SOC and CSIRT as key security defence teams, robust modern technologies like Security Information and Event Management Systems (SIEMs) are utilised to analyse logs and generate alerts and incident reports gaining insight into occurrences for obtaining data triage automatons. The SIEM tools integrated the logs from different sensors, correlated them, and compared them with rules or signatures. If the correlated data triggers a rule, an alert with a summary is generated. When they use SIEM, they must make decisions about data triage. As a result, they are dealing with CTA tasks. Because they must comprehend what the incident is, how serious it is, what should be done in response to it, and why it occurred based on the features and evidence described in the report. Some things are related to cognitive task analysis, part of CCSA.

Inadequate information in incident reports generated by security technologies written in a machine-friendly manner rather than a human-friendly format causes cognitive overload on currently scarce cybersecurity resources. Based on the network flows features, the connections (inputs and outputs) have been analysed from whom, when, where, how, and why. It shows the network is in normal or suspicious states. When analysts interpret the network's state and forecast the network's future state, cognitive task analysis (CTA) reflects the aims of CCSA. The cognitive activities that cyber analysts use to execute specific tasks such as decision-making, problem-solving, memory, awareness, and judgment are described and represented using CTA approaches. Cyber analysts, SOCs, and CSIRTs, should respond to possible incidents based on what SIEM technology notices and reports. They must indicate whether or not the elements mentioned in the incident report depict the actual incident. They must also investigate what

happened, when, where, and why, and the appropriate response. Identifying crucial decision points and grouping, connecting, and prioritising them are all part of CTA.

Cyber professionals begin gathering more facts about network users and services implicated in the original alert to satisfy their CTA duties in incident management [15]. They spent a lot of effort gathering data on network users to have a better knowledge of network conditions [15]. '(Who) was it about?', '(What) happened?', '(When) did it happen?', '(Where) did it happen?', '(Why) did it happen?', and '(How) did it happen?' are all critical details missing from current incident reports generated by devices. A 5W1H method is a helpful tool for analysing reports and texts, and it takes six pieces of information (Who, When. What. Where. Why, and how) [16]. According to the 5W1H method's theory, if a study delivers answers to the aforementioned questions, it can be regarded as complete with the main purpose of explaining a subject [17].

The biggest weakness in the security chain is still people. SOC or CSIRT must deal not just with technological issues but also with issues involving people and procedures. The study of cognitive science to understand and improve the processes and cognitive tasks of security analysts is a research issue that has attracted attention [14]. Cognitive science has the ability to improve human understanding, comprehension, and projection processes, which are both parts of self-learning and contribute to the situation of human cybersecurity specialists [14]. Collaboration between humans and machines, statistical methodologies, machine learning, and Big Data have all assisted security specialists at SOC or CSIRT in developing or expanding their cognitive skills, captivating the attention of many researchers interested in using this science in cybersecurity processes such as [14] or cognitive AI in [18]; Cognitive AI is a next-generation technology for security operations centers that was launched in order to improve military decision-making and strengthen cybersecurity defenses. The primary challenge that these new technologies must address is how humans can attain situational awareness in an environment where AI systems are deployed [18]. The combination of cognitive theories with methodologies and models used in the field of cybersecurity can help players in cyberspace make better decisions.

The aim of this study was to create explainable intelligence models in a human-readable format using novel storytelling strategies derived from security logs and alerts with sufficient context enrichment. As a key axis in the validation and decision-making phases, the various solutions for automating the performance of cognitive activities recognised in cyber operation procedures were addressed in the presented models by SOC or CSIRT. It will help them be aware of any threats that may be lurking within an IT system. Since incident management necessitates many cognitive tasks from a cyber analyst, such as decision-making, problem-solving, and judgment [11], the most obvious goal to emerge from this study is understanding and digesting the large and varied volumes of data, which expanding cyber analysts cognitive skills. As a result, the proposed models sought to improve cognition and promote expert comprehension as verification of filtered data is necessary to identify events with few false positives. Although this study focused on

cybersecurity, it was not restricted to it. This research can benefit any dynamic, complicated environment with a large volume of data that requires analyst engagement.

Explanatory intelligence, in particular, will be used to enrich inputs and provide additional context for items and topics. As a result, the explainable model in a human-readable format will aid cognition efforts, providing sufficient awareness for both experts and non-experts to confirm what happened to the data. Thus, the development of storytelling reports at various levels of detail, from local and global knowledge bases, offers a holistic view of the cyber situation, filling a gap in the analysis of cyber events through the incident management process.

Furthermore, a knowledge graph was utilised to include multi-source heterogeneous data, visualise the data, and allow several individuals to engage in incident management [19]. It is vital to engage in incident management that enables for information exchange in order to keep local and global information up to date, and to build an understandable common ground among humans. By involving more individuals in incident management to improve the monitoring and communication procedures and supporting CCSA, a narrative report with a knowledge graph enables security professionals to better comprehend aspects of an environment. Cybersecurity incident management teams can better understand the current situation and respond accordingly when they have a complete picture of occurrences impacting the organisation's environment.

The proposed intelligence model using the knowledge graph is also useful in other applications, such as fraud detection, where a large number of transactions must be analysed and digested. Fraud detection through transactions is of the same nature as incident detection among logs, both transactions and logs are vast in the volume of data that gathers in non- explainable formats [20] Fraud detection is the set of activities to identify fraudulent behaviour among transactions [21]. This field is an important challenge in security research that causes much attention and outlines from the government [22, 23]. Detection methods that employ Machine Learning required a domain expert who should analyse historic transactions to identify the fraudulent behaviour [20].

The interpretability of models in financial applications is missing and typically gives a score indicating whether a transaction is likely to be fraudulent or not — without explaining why [24]. Focus on the interpretation of financial data is rare [24]. CoDetect [25] is a fraud detection system that focused on interpretability and reveals financial activities from fraud patterns on a graph-based similarity matrix. Contextual information is used to improve the interpretability of the clustering model, which can be useful in financial detection as an unsupervised method. In this study, connections of transactions are revealed in the knowledge graph with labels that provide much more interpretive to aware an analyst. It is also capable of searching which can be shared with others easily.

In crime detection systems, visualisations are used for a variety of tasks, including detecting developing fraud patterns, conducting criminal investigations, analysing fraud alarms, and

coordinating with other fraud and financial crime experts [26]. Traditional visualisation and human-in-the-loop methods, while useful in other domains, have significant challenges in fraud detection. Graph knowledge has inherent benefits when it comes to expressing and displaying data. Scaling solutions to satisfy the needs of industrial applications while coping with the challenges of speed and complexity is essential [26].

In order to assess whether or not a reported transaction is truly fraudulent, a real-world fraud detection system requires human intervention. To establish whether the detected transaction is truly fraudulent or not, a human user should be able to look at prior transaction patterns or call the client. For the fundamental scenario awareness of financial data, simple to comprehend and analyse transactions is necessary [27].

1.2 Research Problems

The modern enterprise uses various cybersecurity incident management systems to protect information resources. The majority of organisations rely on such systems to properly respond incidents. Security devices report any suspicious or anomalous activities once they have been detected. Once the incident has been specifically identified, it is then passed to SOC or CSIRT for deeper investigation and response.

To response properly, situation awareness as collecting inputs from a system which informs surroundings to act upon required [28]. To be aware about incidents, security professionals rely on the information presented in reports. Their unique blend of sharp thinking and quantitative and qualitative capabilities in collecting, integrating and analysing the events that occurred in order to discover the best reaction would be extremely beneficial to their organisation. A high-quality SA, better understanding of the threats and associated impacts of cyber events is essential to the decision-making process [29]. As much as situation awareness, particularly CSA gains, analyst response times were reduced.

Most of the current research, for instances [28, 30–32] draws attention to the digital environment of a system with prospective on SA's systematic factors. Among them, the most focuses are on the security systems and tools such as SIEM to support CSA [28]. The capability of the SIEM tools in data collection and correlation, and also adjustment of them with technologies such as AI and ML is the main point of focus of the studies [28]. SIEM solutions have evolved to become secured systems by concentrating on the technical features which help to improve their detection, correlation, and reaction capabilities by integrating AI/ML technologies [33]. In paper [34], different SIEM tools are analysed and their capabilities were reviewed by Gartner. For example, Bryant and Saiedian [35] proposed LogRhythm with the ability to aggregate data from many sources and make sense of unorganised data; where cyber threat modelling with kill-chains was

proposed to facilitate logical data aggregation. Moukafih et. al. [36] used neural networks as machine learning techniques to create high detection capabilities.

Reports are the one case of the outputs of the SIEM tools. Generate reports are analysis results about detected security incidents from the SIEM that will reach the security analysts; Who needs to be involved in the analysis of the incident to response. The incident reports automatically generate from sensors forms the SA based on data fusion and correlation with Cyber Threat Intelligence [37]. Enhancing reports rarely focused on the current SIEM solutions to improve CSA. Required detailed information does not completely cover in the existing incident reporting of SIEM tools [33].

Due to the lack of well-established reporting techniques, analysts are not well-suited to be aware of threats or provide insight into network activities. Inadequate, non-semantic, and context-less information delivered in a machine-friendly format necessitates a long, manual reaction time to collect, integrate and analyse data in order coordinate strategic and operational security measures.

According to D'Amico and colleagues, perception, comprehension and projection are the main levels of cyber analysis, aligned with SA and requiring the completion of many cognitive tasks by the cyber analyst [38]. The reports generated by cybersecurity incident management systems lack detail, making it difficult to comprehend the incident or predict future incidents. It demonstrates that the current reports were rarely focused on two levels of SA: comprehension and projection [39]. While generated reports are not enriched enough to deliver instant insights, human beings have to be involved in the analysis process, which is predominantly a manual task. A significant limitation exists in the current cyber data analysis process, which relates to message verbosity without annotation to cope with the enormous volume of events [40].

Additionally, decision-makers (cyber experts) frequently face new challenges in cyber data analysis settings due to new events. They have little time to explore alternative courses of action before being confronted with a threat. Making decisions is difficult for a variety of reasons. It is difficult to locate and integrate decision-relevant data. In incident reports, context is often omitted, tacit, sparsely represented, or incorrectly represented, necessitating laborious and error-prone internal reconstruction by decision-makers. Experts are frequently engaged in activities in ways and combinations with which they are unfamiliar. They may be unprepared due to the modern requirement for speed of response. This forces experts to multitask between several overlapping incident warnings at the same time. As a result, it is critical to determine to what degree a human decision-maker is conscious of the situation, whether they have achieved a certain level of SA, and how well they continue to retain and grow that awareness over time.

It may be difficult to pinpoint the answers to the core questions about the actor (who), riskiness (what), evidence (how), time (when), and location (where) of the event in incident reports because their structure is not in such a way that the essential aspects required for comprehension are easily

conveyed. However, it is easier for analysts to rely on the reports generated by security devices, particularly SIEM tools. The reports from tools are not in an understandable format, and enough detailed information is required to answer these core questions.

The review of existing incident management, numbers of issues that less studies have investigated about them are as following.

1.2.1 High-volume of events are logged but not comprehended

Millions of activities and authorised or unauthorised attempts are recorded on computer systems on a daily basis. As an example, a university of 3,000 staff and 40,000 students registers approximately 200 MLN events every year[1]. At the same time, only about 20% (or 40 MLN) of the logs will be analysed by specialised security systems. The cybersecurity team is small in comparison to the volume of events recorded. The incident response team at Victoria University, for example, is made up of no more than ten trained professionals.

Numerous algorithms have been proposed to automatically analyse the events and signal alerts for potential malicious activities [41]. There is a multitude of various types of monitoring systems in use that generate potential threat alerts. In order to appropriately respond to the suspected threat, the *synthesis* of currently *disintegrated* systems is required. Building context around a potentially malicious alert is predominantly a manual task which involves rich experience and knowledge regarding log files analysis [41]. Thus, comprehensive alert analysis has become a critical task in harmful event and fraudulent activity detection, their timely resolution, and future prevention [42].

Although monitoring systems are helpful in filtering through millions of logged events and generating security alerts, final human assessment remains part of the process. As such, thousands of potential security breaches received from different monitoring systems pose significant burden on cybersecurity team resources. Because of the *machine-friendly* rather than *human-friendly* format of such alerts, and the extensive domain knowledge necessary, interpretation of raised alerts is strictly limited to cybersecurity professionals.

Comprehensive and accurate alert assessment is also prone to the subjectivity aspect that forms an inherent part of any human evaluation process. Correct response then depends on the extensive experience of analysts from the cyber threat management field. The dramatically increasing number of security alerts is currently outgrowing scarce and expensive cybersecurity resources.

[1] Reported by cyber director of VU CYBER

1.2.2 Up-to-date local and global knowledge is required for analysis

Despite the overwhelming volume of security alerts, only a fraction requires further investigation, though a percentage of alerts are false positives [43]. Still, time and effort must confirm that the alert is indeed a false positive or a real incident. Knowledge outside security logs is required to properly assess the scale of risks.

Security response teams need to gather local (The system integrated information about the network infrastructure) and global (indicators of compromise (IOC), tactics, techniques, and procedures (TTPs), IP addresses, DNS blacklists, etc) information from various sources to feed the correlation process and support analysis of security events and identify alerts [19]. In most cases, cross-data-source analysis has been a focal point in the design, maintenance and supervision of human-machine systems [44], [45].

Comprehensive and integrated up-to-date information to cyber professionals, in broad terms, contains any information that may be used to characterise the situation of an IT entity that is considered as linking to locally and globally available information [19]. Local and global information is continuously updating. Implementing approaches to integrate the information into the data model to make full use of cybersecurity-related details from various resources, and associating all this security-related knowledge is difficult and usually incurs expensive modification costs [46]. One of the major challenges is the rapid variation of the network environments which has a potential impact on security posture; i.e. machines added and removed, various patches applied, applications installed/uninstalled, or confidential data uploaded or deleted [47]. "The problem is not lack of information, but rather the ability to assemble disparate pieces of information into an overall analytic picture for SA" [48].

Local domain knowledge determines the risk of an internal assets, and the potential risk of outsider is specified by global domain knowledge. As an illustration, consider the examples below:

- **Local domain knowledge required:** A server of the organisation X is used for temporary storage and web testing, and is labelled as a *non-critical* host. Most of the alerts from that server can be omitted unless a serious breach occurs. However, the server is located in the finance department for financial reporting and budget planning. A finance department usually holds critical information. If an alert for a serious breach occurs for one of the servers in this department, other servers can also be at potential cyber risk, warranting further investigation despite no explicit alert being raised. Thus, the exceptional defense strategy should be adopted in advance following the complete knowledge obtained from inside the organisation

- **Global domain knowledge required:** The organisation Y with limited number of experienced cyber professionals has to prioritise crucial alerts over a large volume of the remaining security breaches to provide a prompt response. The selection is based on prior knowledge and experience accumulated through the repeated alerts from historical records. An appropriate response for the new attack requires an in-depth investigation of the attacker's characteristics. However, the attacker may change its behaviour over the time of repeated activities. The level of expert knowledge is usually not increasing at the same speed as the complexity of attacks in today's digital environment. As a result, a critical alert may not be given the required priority, leading to a delayed response and potential escalation. Thus, knowledge obtained automatically from external sources is required to stay up-to-date with increasingly sophisticated and dynamically changing cyber attacks.

Both examples show that comprehensive alert analysis requires domain knowledge from both local *and* global. Expertise is required to reliably handle alarms, and integration with knowledge can reduce false alarms [49]. False alarm rates are compromised with minimal expert intervention, so similar knowledge needs to be modelled and incorporated into alerts to reduce human interaction.

1.2.3 Knowledge beyond security team is required for analysis

To properly assess the scale of the risk, the knowledge outside a cybersecurity department is frequently required. A human interpretation of the knowledge and security analysis will be needed to engage more staff from different departments to managed the security risk. As an illustration, consider the examples below:

1. **Security *escalation* required:** A server of organisation X is used for temporary storage and web testing, and is labelled as a *non-critical* host. Most of the alerts from that server can be ignored unless a serious breach occurs. However, this financial year's end of season was particularly busy. The server was borrowed by the Finance Department as a temporary server for financial reporting and budget planning. The server now holds critical information, and the security level lifted accordingly, with *all* of the alerts closely monitored. The information about the server transfer was not passed on to security team though. However, the Finance Department staff have little or no expertise in cybersecurity. A big security hole is left open to the potential attackers

2. **Security *exception* required:** Organisation Y repeatedly receives a high volume of security breach alerts from an internal host. This is a typical symptom of attack, and the security system blocks the host along with the related ports. Further investigation involving staff from other departments reveals that the host is an experiment server used by the development team. The host is located in the department A's laboratory. The department is

testing game engines that have a cloud-end. Low-level communication between the local host and cloud server is required, as is appropriate configuration with relevant security exceptions.

In both examples, the alerts analysis requires knowledge from the security team and other departments which cannot be modeled and integrated with the alert analysis. Either false alarms could be triggered, or high-risk alerts could be neglected. Therefore, generally engaging more staff from different departments is needed to solve the issues discussed in the examples. The development of a shared understanding of cyber SA (currently restricted to cyber professionals) is required.

1.2.4 Fraud transactions resemble normal transactions

Fraud detection is sometimes formulated as a binary classification issue in which users are classified as frauds or non-frauds based on their transaction histories. Many fraud detection algorithms and learning approaches have been suggested based on the grouping pattern of frauds.

Designing and evaluating these algorithms, on the other hand, is difficult [50]:

1. Transaction data include a large number of features characterising user activity, making it difficult to choose the most relevant aspects of fraud patterns
2. The choice of feature sets and algorithms is highly influenced by the domains and scenarios
3. There are very few or no fraud labels for training or evaluation.

Usually, fraud can cause enormous financial losses [51]. For example, author in [52] discovered 32 instances with fraud losses ranging from $1,209$ to 1.9 million. Meanwhile, a successful fraud detection procedure necessitates the deployment of a suitable algorithm, selection of valuable feature sets, and elimination of false positives. All of these procedures however, need the presence of human specialists for visualisation which is an essential component of any effective fraud detection system, and as long as the specialists are aware of the situation, visualisation may aid in their decision-making.

Analysts' knowledge is crucial in the fight against financial fraud because it enables them to detect financial crime within an organisation by evaluating current data using their own experience and abilities or through a comprehensive analytical report [53]. By building a model that allows for the comprehensive identification of suspicious behaviour patterns while taking into consideration the human component, fraud analysts may better understand transactions' elements to discover potential financial fraud instances.

This study addresses the aforementioned research gap by investigating the research question: **How might explainable intelligence help SOC and CSIRT gain comprehension awareness, particularly CCSA?**

In prior research, automated CSA tools and models aiming to enhance the cognition of experts have been proposed [54]. As defined by Endsley: *"SA is the perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"*. As such, a SA system was designed to compile, process and fuse data from several different perspectives [55]. However, existing CSA systems have not been able to address the continually evolving cybersecurity challenges completely [54]. Although they are helpful, the cyber experts still have to digest vast volume of data to discover hidden links and dependencies. They must perceive and work to elucidate awareness of the network (CCSA), to develop knowledge, comprehension and execution of security response actions [11].

The reporting tool was developed to enable real-time analysis for improving situational awareness [37] .Unfortunately, there have been very few attempts to enhance the reporting techniques about security incidents as the output of security management systems. The enriched report can help human to understand better and easier with regards to human cognitive. Although integrate the reporting into other systems, particularly safety reporting, was the main concern of some studies, usage of the contextual data in a human-understandable format that improves the efficiency of security decisions and CSA, have not been studied [56] . In [57], author investigated to address how the SIEM use case can improve cyber SA. In this thesis, cyber security kill chain models were proposed to interpret information by the SOC analyst consuming in case [57].

This study used explainable intelligence to interpret cyber data, including incident alerts created by security tools and security logs from Windows, and then applied information from local and global knowledge sources to enrich data to improve experts' cognition. The findings demonstrated how enriched context in human-readable formats could provide a thorough insight into potential threats and incidents. Context is a dynamic grouping mechanism that encompasses instances associated with a specific situation.

The extracted information must be properly represented in order to depict background behaviour at run time. Story is the good candidate. Storytelling can be used as a knowledge representation method to highlight the explicit and implicit information from data, and convert it into a human-readable format [58]. According to Mackinaly et al. [59], *"Data tells you what is happening. Stories tell you why it matters"*. A malware information as an example presented in data and storytelling format in the following:

- Data: 5:05 PM 02/27/2019 -10.233.62.247- *MALWARE-CNCOsx.Keylogger*

- Story: Alert at 5:05 PM on 02/27/2019 because of the connection to a suspicious site 10.233.62.247. The malicious site is *MALWARE-CNCOsx.Keylogger*, which is a spyware program for Mac OS X that records keystrokes, may take screenshots.

Acquiring information and knowledge about why it matters shows a highly developed awareness of appropriate sources for an investigation that have affected experienced experts. The story-like technical reports are produced based on personal preference, which positively impacts comprehension.

In this research various explainable intelligence models get logs, alerts, and transactions as inputs and then make a story from them to aware SOC and CSIRT. The story-like technical reports are produced based on personal preference, which positively impacts comprehension. Explainable intelligence that analyses, integrates and translates cyber data into human-comprehensible stories provide a holistic view of potential security breaches that improves incident management process. The use of interpretability approaches (data into story) extract various forms of information because it interpret data means to extract information from them [60] . So, it increases confidence and protection in the decisions making process.

To support the exchange of security-related information and share lessons learned with affected departments, external regulators, industry associations, or other professional bodies [36, 56], shared awareness is required. To emphasise the importance of participatory reporting and gaining shared-awareness by others inside the company that affected from the incident and its risk, the second research question has been raised.

How can explainable intelligence help organisational members in the incident management process gain situation awareness while ensuring the involvement of critical stakeholders?

Incident management process requires comprehensive information to support a decision-making process. Organisations main challenges are 1/) security team do not have the capacity to handle the volume of incidents; 2/) other affected departments are not aware of risks or incidents. According to [30], involvement of other types of stakeholders such as managers, higher-level decision makers, and non-expert users in the incident management process is missing. The information collected through incident reporting to improve situation awareness for SOC and CSIRT can be shared with other affected departments and also with higher-level decision makers.

Information exchange is vital to communicate the right information to the right people, such as the asset owner. A consensus view of a number of individual views about a particular activity or collection of activities is needed in multi-actor networks. When a security incident occurs, the analyst must decide what happened and why it happened, however the response is contingent on the organisation's mutual SA capability. Shared SA settings involve several of the right people trying to form a common picture [61].

Various insights from a cyber-security incident report helps to focus and justify subsequent investigations [56]. Most of the studies focused on the share security insights between different stakeholders, who maybe be helpful to validate security threat analysis [56]. The main concern of this research is to integrate information about the incident from different person who may affected from it includes, risk owner or who has updated knowledge about it. For instance, an asset owner who knows the importance of the asset which was under threats.

Knowledge graph is used to depict the context by use background information with more ability to exchange the context. Therefore, people from different departments are able to update engaging in the analytical process. As a result, in this study, the relationships between the cyber data and relevant information from the local and global knowledge bases are visualised through a shared understanding of CSA contextualisation with knowledge graph. I also implemented the visual analytic process using the intelligence model with a knowledge graph to absorb a bigger volume of information in a financial case since open data is demanded due to humans cooperating in the analytical process to acquire insights, draw conclusions, and eventually make appropriate decisions. As a result, money laundering analysts and other relevant parties can engage in the process of detecting fraudulent transactions with full insight gained from the knowledge graph.

1.3 Clarifying the Terminology

1.3.1 Knowledge base

In this study, two main databases (local and global) to obtain contextual insight into an alert have been used. In terms of completeness, internal sources and external sources are provided to enable a sufficient level of comprehension.

The local knowledge base includes supplementary information that is internally processed, as well as the raw data collected from security devices. The local knowledge base contains explicit knowledge about the event. Implicit knowledge is added to the knowledge base by predefined rules and procedures. The local knowledge base includes available information with the domain knowledge of experts and the raw data collected from the security devices, such as: (1) a list of the internal servers and hosts with the associated information, including domain name, administrator, severity (low, medium, high), location and installed applications, (2) story templates, (3) rules for analysis, (4) regular expressions, and (5) a list of keywords.

The global knowledge base contains supplementary information that is collected by external companies and researchers, and is processed internally. The following information is included in the global knowledge base. This list it is not exhaustive:

(1) Whois command and additional information about the IP address and domain registrants [62], (2) online scan engines such as Virus Total[2] or Threat Miner[3] that generate the "malicious" labels for a given URL or file, (3) online open threat exchange repositories such as AlienVault [4], Windows Defender Security Intelligence (WDSI) [5] and Symantec [6], (4) the Snort community rule set and (5) intelligence threat feeds.

1.3.2 Event

In this document, the event is the status of the action that is recorded in the log by the monitoring system.

1.3.3 Alert

The alert is a generated message when an abnormal event occurs [63]. The security devices generate an alert when they observe that a part of an event specification matches their predefined patterns [64]. The generated message (called an alert message in this thesis), provides a short description for further analysis.

1.3.4 Cyber security incident

An unwanted or unforeseen cyber security event, or a sequence of such events, with a high likelihood of compromising business operations [65]. In this study, cybersecurity incidents are identified by aggregating and correlating various events and patterns defined on the SIEM device.

1.3.5 Report (incident report)

The report is a document that presents detailed information about the alert to help the analysts to understand more about the abnormal events registered [66].

1.4 Thesis Outline

The structure of the thesis is as follows. Chapter 2 provides a literature review on the main domains of this study where explainable intelligent models are used. The explainable model is

[2] https://www.virustotal.com/
[3] https://www.threatminer.org/
[4] https://otx.alienvault.com/pulse/
[5] https://www.microsoft.com/en-us/wdsi/threats
[6] https://www.symantec.com/security-center/a-z/

most commonly used in cybersecurity, and the role of cognitive sciences (self and shared SA) in cybersecurity is highlighted; however, the model is proposed and used in financial applications to achieve SA. This chapter contains a research review on cyber incident management and fraud detection. Finally, the chapter identifies research gaps in the literature.

Chapter 3 present the designed explainable intelligent model in cybersecurity management, with a focus on interpreting logs as cyber data (input). This chapter demonstrates how logs are analysed and presented in chains of events (storytelling) using the proposed intelligent model to gain self and shared SA. This chapter provides a detailed description of the approach, methodology and evaluation used in this study. The proposed explainable intelligence also was published in Australasian Joint Conference on Artificial Intelligence conference [2].

Chapter 4 introduces the designed explainable intelligence model in cybersecurity management which is used to interpret incidents resulting from cyber alerts (input). This chapter shows how alerts are analysed and presented in a storytelling framework using the proposed intelligent model to improve self-awareness. It also describes the approach, methodology and evaluation in detail. In terms of evaluation, a real-world scenario was conducted with an empirical analysis, as well as a survey analysis, with a comparison to reports generated by the Secureworks (external vendor's tool). The proposed explainable intelligence also was published in IEEE Access Journal [1]. The survey design and evaluation after a revision is under review of Computers & Security Elsevier Journal[7].

Chapter 5 addresses how to achieve shared cyber SA through the use of the explainable intelligence model proposed in the previous chapter. This chapter includes a visualisation model (knowledge graph) that is used in conjunction with the storytelling framework to improve understanding by involving more people - achieving shared SA. The connection of this research to broader debates in the incident management literature is also presented. The proposed intelligence with the shared SA ability through the knowledge graph also was published in 24th International Conference of Information Visualisation (IV) [3].

Chapter 6 demonstrates how the designed intelligence model can also be used for fraud detection, transaction visualisation and awareness. To visualise a financial transaction, the knowledge graph used in Chapter 5 is utilised. It aids in raising both self and shared awareness, as well as making sense of the criminal behaviour hidden in particular scenarios. This chapter was prompted by the designed explainable intelligence with shared SA capabilities, and its employment in not cyber management (in fraud detection) is emphasised. The proposed model received first place in the Hackathon competition[8].

[7] N AfzaliSeresht, Y Miao, Q Liu. "The Empirical Analysis of Cognitive Load Reduction of the Security Incidents Interpretation using Storytelling Approach: Questionnaire Survey and Analysis" Computers & Security - Journal - Elsevier.

[8] Hackmakers DigitalDefence December 2020

Chapter 7 concludes the dissertation by summarising the research background, research method, and key contribution of this research project. The chapter also outlines directions for future research and highlights the limitations of this study.

Chapter 2

Literature Review

This chapter presents a survey of the literature in the primary fields of this research, including incident management and fraud detection where the generated explainable intelligent models will be used. The goal of explainable intelligence is to minimise human cognition efforts during the usage process (incident management or fraud detection). In order to attain self and shared SA, the function of cognitive sciences in the analytical process is emphasised. This chapter offers a review of the literature on cyber incident management, fraud detection and cognitive science in order to identify research gaps.

Finally, contemporary technological options for assisting human analysis and reducing cognition efforts will be divided into four categories: first, supervised/unsupervised detection approaches, also known as black box methodology, which are simply temporally pattern-based solutions used to integrate and synthesise data and trends. The second type of technique is visualisation, which aids in the understanding of systems and allows them to change over time. To address the lack of comprehensive analysis in the use of gathering all significant components, a third analytical methodology is exchange formats. Finally, narrative analytics and explainable intelligece methodology was presented as an analytics capability that can help organisations in developing a comprehensive analytical process for incident management or fraud detection, as well as improving the detection or response process.

2.1 Cyber Incident Management

2.1.1 An overview of cyber security skills and deficiencies in cybersecurity incident management

There are many individuals whose hard work has contributed to incident management research. It is vital to review their roles and skills at the beginning of the literature review. Security Operation

Centres (SOCs) and Security Incident Response Teams (CSIRT) are two of the most common security operations. Although both provide cyber incident responses, they don't have the same goals or methods. A CSIRT is mainly responsible for operational and technology-centered tasks, with the primary objective of facilitating organisational recovery to routine business operations [9]. Typically, SOC sets up for the first tier of analysis to do 24*7 monitoring, exploring, detection, and response [45]. CSIRTs may work under SOCs or independently, depending on the needs and size of the organisation. Depending on the demands of the enterprise, the CSIRT inside an organisation may be a structured unit or an ad-hoc team. Threats are encountered on a regular basis in sectors like health care, banking, or government, necessitating the creation of a formal, full-time CSIRT.

As the brain of a security organisation, a SOC performs a variety of tasks to protect data, including prevention, detection, incident management, reporting, and anything else that involves managing and protecting data within the company. They'll need some promotion skills and knowledge to complete these tasks. They must, for example, have a basic understanding of network security and security fundamentals, as well as the ability to ingest data from a variety of sources. The CSIRT team, on the other hand, relies heavily on organisational, problem-solving, and communication skills to carry out their duties. Since there is a strong lead in cyber defense activities based on SOC analytic thinking, they are called cyber analysts or cyber professionals in this research.

According to the survey [67] conducted by the SANS Institute[1] involving the observation of various organisations over a two-year period, cybersecurity analysts mostly spend an average of 24 hours or less on detection after a compromising incident has occurred. Approximately 40% of analysts require more than 24 hours, in some cases, 4-6 months to detect the initial compromise.

In their routine activities, they need to process vast amounts of information and combine data from multiple sources to understand incident states. Security analysts require great concentration and cognitive skills to undertake these activities and they can be affected by various factors [14]:

- High stress
- Low SA
- Limited experience
- Unstructured tasks
- Insufficient information in alert message produced in machine-friendly rather than human-friendly format
- Non-standardised methodologies to identify and respond to attacks

[1]The SANS Institute is a private U.S. for-profit company founded in 1989 that specialises in information security, cybersecurity training, and selling certificates.

- Large amounts of data and information
- Uncertain and out of date sources of information
- Lack of performance metrics
- Lack of collaboration with different departments for shared SA knowledge
- Lack of visibility into insider behaviors.

For the prevention and identification of security threats in cyber infrastructures, many technical solutions are currently available however, humans remain the weakest link in the security chain. SOC and CSIRT must deal with issues relating to individuals and procedures in addition to those arising from technology [14].

As cybersecurity incidents become more prevalent and affect organisations, it is critical that organisations be able to investigate, track, monitor and respond to cybersecurity incidents in a timely and cost-effective manner [9]. According to a review of the literature on cybersecurity incident response, the goal of many organisations' cybersecurity incident response approaches is to invest in sophisticated prevention measures aimed at controlling identified risks rather than an adaptive response mechanism to investigate and combat unknown dynamic and emerging risks [68]. Recent commercial discussions also indicate that emerging methods have inherent flaws when applied to real-world security incident responses [69, 70]. This is because the majority of cybersecurity incident response methods are organised in a sequential plan-driven manner, beginning with the planning process and ending with the identification of a cybersecurity incident. This is followed by containment, which allows the malicious act to be eradicated and eventually, lessons learnt are integrated into the next step of planning [71].

Despite the fact that the majority of the literature on cybersecurity incident response is concentrated on technological practices for implementing cybersecurity incident response capabilities inside organisations, researchers have also discussed and found numerous flaws in existing organisational cybersecurity incident response approaches. Some of these flaws include being too linear, failing to provide adequate insight into the causes of the incident, failing to maximise the advantages of digital forensic capabilities, and failing to represent the concurrent lifecycle of real-world incident handling [68, 72–75].

The goal of many organisations' cybersecurity strategies, according to [68], is to invest in sophisticated preventive measures aimed at controlling known risks rather than an adaptive response mechanism to examine and combat unknown dynamic and emerging risks. Organisations that operate under the prevention model are better prepared to deal with static and predictable cybersecurity threats, but they are more vulnerable to complex and unpredictable threats like Advanced Persistent Threats (APTs).

Tan et al. [74] investigated the reasons that led cybersecurity managers to refuse to investigate security incidents. A highly controlled industry that penalises companies for security incidents, a lack of advance planning, and an industrial focus on device recovery rather than incident investigation were among these factors [74]. Tan et al. [74] also noted that the company in their case study was unable to determine how the cyber-attack occurred and was unaware of the advantages of prosecuting perpetrators related to security incidents. They also lacked a specific understanding of what constitutes a cybersecurity incident.

Werlinger and his colleagues [75] looked at the cybersecurity incident response practices of professionals from a variety of organisations and sectors, focusing on the socio-technical aspects[2] of incident response in particular. They looked at the techniques used in cybersecurity incident response and how they could be enhanced.

While various efforts, incident response tools and protocols may provide some support for the incident response process [76], Werlinger and his colleagues [75] believe that tool support for simulation should address not only the technical factors, but also include functionality that supports collaboration between different IT practitioners as they track simulations and evaluate their consequences. As a multifaceted activity, security incident response necessitates a combination of strong technical and communication abilities, as well as the ability to execute simulations in distributed systems supervised by a variety of practitioners [75]. This response process necessitated active participation from all of our participants as well as external stakeholders. Participants utilised a variety of technologies to help them complete their objectives, and when they didn't have the necessary security tools, they created their own [75]. Furthermore, according to [73], security forensic examiners, device managers, and cybersecurity incident handlers must collaborate more closely so that all important stakeholders understand the need for reporting even seemingly minor security issues.

2.1.2 Cybersecurity incident response process

Cybersecurity incident response is a continuous process of maintaining an adequate state of security in the organisation facing attacks [14]. This process is vital for organisations because there is always a security hole open for potential attacks, and organisations cannot prevent all breaches. Security specialists must identify, investigate, respond to, and learn from potential cybersecurity incidents in a timely and cost-effective manner in order to carry out the response activity [9]. A correct prompt response to the potential threat can avoid potential escalation like financial damage that impacts reputation and competitive advantage. Therefore, the primary goal of a SOC or CSIRT is to minimise the effects of an incident that may negatively affect the organisation along with managing a return to an acceptable security posture [9].

[2]Sociotechnical refers to the interrelatedness of social and technical aspects of an organisation or the society as a whole.

Existing solutions for generating data triage automatons, such as SIEMs, are designed to be reused at different investigation stages, which facilitates the analysis and detection actions. However, although numerous support tools come in a variety of implementations from machine learning algorithms that automatically analyse the events and signal alerts for potential malicious activities, the tools at their disposal are often specific to a particular data type [77]. Cyber analysts who must undertake cross-data-source analysis have, in most cases, been a focal point in the design, maintenance, and supervision of human-machine systems [44]. Although monitoring systems help filter through millions of logged events and generate security alerts, final human assessment is still part of the process. Cyber analysts must recognise and work to elucidate awareness of the network (CCSA) to develop knowledge, comprehension and execution of security response actions [11].

Cyber analysts use the available context for proactive decision support to enable them to make a judgment as to whether an incoming alert is a real threat or a false positive (Human-in-the-loop[3]). Aside from the mental challenges that come with making decisions based on human experiences and beliefs, cyber network settings are continually changing. [15]. Cybersecurity analysts, in coping with the fast-evolving threat landscape, follow complicated processes in their investigations of potential threats to their network. They have to trace each event and find the corresponding events to understand the symptoms and assumed the threat scenario that occurred.

A good understanding of threats is better compliant with security policies, extending this to raise awareness and improve incident management processes [78]. Raising awareness helps improve the incident management process [79]. While individual information security awareness is important, it is not a satisfactory level for organisations. Most analysts rely on operational reports from security devices. They are not comprehensive enough and contain meaningful content about the incident, so the threat cannot be fully understood. In addition, it is recognised that technical staff (such as SOC and CSIRT) are primarily involved in response and learning activities. Nonetheless, other affected departments, units, and management play an important role, especially in the case of serious incidents. de Souza et al. [80] have discovered that people are the most important source of information when dealing with complex cases. But according to a statement by the Information Security Manager in [81], sharing information, was one of the most challenging parts of incident management. The need for communication and cooperation particularly in the response phase is emphasised by Ahmad et al. in [81].

Several organisations, including the International Organisation for Standardisation [82] and the National Institute of Standards and Technology [83], have issued recommendations on cybersecurity incident investigation and recovery methods. Academic researchers have suggested cybersecurity incident processes in addition to best practices [84]. These cybersecurity incident response methods are based around a standard process, beginning with preparatory activities

[3]requires human interaction in the process.

before an incident happens, followed by incident detection and review, followed by incident containment which helps CSIRT to eliminate, recover and eventually provide input information into the planning stage. As a result, incident response is an orchestrated mechanism for dealing with and managing the consequences of a cybersecurity incident [9]. The primary objectives of the incident response process are to quickly mitigate the impact of the attack, the time required to recover from the attack, and develop countermeasures and instructions to aid in the prevention of similar attacks in the future.

To address and manage the consequences of a cybersecurity event, CSIRT or SOC follow the steps outlined below.

2.1.2.1 Preparation

This phase requires team preparation to be ready to handle a security incident in a timely and cost-effective manner [85]. The cybersecurity team establishes policies, a response planstrategy, a communication plan, and tools that can help mitigate any potential problems handling an incident [9]. Another essential key element to have implemented in the preparation phase is to train the organisation's employees. Without training, the cybersecurity team and the organisation's staff are unsure of their roles in the cybersecurity processes and policies, and the process of incident response handling could be a complete failure.

In LRZCSIRT [86], the roles, responsibilities, and tasks associated with incident response were suggested as a plan for incident response. From the experience of LRZCSIRT, it is recommended to clearly define the security incident so that it can be distinguished from other problems such as configuration errors. Ahmad et al. [81] and Hove and Tårnes [87] also emphasised that defining an impact assessment to determine how to handle an incident, is important in a plan.

Defining responsibility is very important, particularly in cases where IT has been outsourced [87]. For example, the response to high-impact incidents is coordinated by a High-impact SOC and CSIRT, while other incidents are handled by a team more independent [81].

Raising awareness related to information security was considered as part of the plan in some studies such as [87]. The study focused on raising end-user awareness of security issues [88, 89]. The focus of studies has been on raising end-user awareness of security issues [88, 89]. The role of security manager awareness in reducing incidents was performed by researchers such as Goodhue and Straub [90] and later by Straub and Welke [91]

2.1.2.2 Detection and reporting

In this phase, the CSIRT or SOC deals with detecting the cyber alert and determining whether it is indeed a false positive or an actual incident. To determine whether an event is an incident, particular steps are required to gather evidence (IOC) from various sources [85].

The analysts use open-source or public cyber threat intelligence (CTI) to collect IOC. Open-source threat intelligence management tools such as Collaborative Research Into Threats (CRITs) and the Malware Information Sharing Platform (MISP) for dynamic analysis require the execution of the software.

Although studies focused on manual or automatic approaches such as [92] and [87], the role of users in detecting and reporting suspicious activities (as the main focus of this research) has been done by others such as [93] and [79]. For example, in [93] not only the internal user role but also the external participation to obtain information about an incident that has occurred is emphasised to complete understanding of the true scope of the incident. According to Jaatun et al. [94] and Werlinger et al. [75] more communication between stakeholders is necessary.

A cybersecurity analyst from the response team considers both sides of a potential cyber incident and focuses on important links in their chain of reasoning. Making a judgment on a cyber incident from an alert is probabilistic in nature because the evidence is always incomplete, usually inconclusive, frequently ambiguous, commonly dissonant, and has various degrees of believability. The CSIRT should also prepare a document detailing the incident (incident report). Reports must document everything accomplished by concentrating on answering the Who, What, Where, Why and How questions regarding the events that have occurred [85]. The questions constitute a formula to determine whether a report gives a complete picture on a subject to facilitate any necessary decision making on the subject. A comprehensive incident report provides information related to these factors to prove judgments that require human analysis.

2.1.2.3 Assessment and decision

This phase, also called the containment phase, is intended to limit the damage and prevent any further damage. To accomplish this objective, cybersecurity policies are assessed and adjusted by the incident team, and the organisation's networks are reconfigured after making a system back-up and temporarily fixing the affected system to make sure that normal business operations continue without interruption [9].

Assessment may need to work with an external organization to verify that the incident actually occurred [79]. In addition, it requires specialised knowledge and knowledge of the normal state of the system [79]. Similarly, tracking the cause of anomalies often required specific technical

expertise and knowledge of attack patterns. so it was considered useful to work with other experts who may provide a new perspective. For example, to assess an incident on the computer X that may affect data Y, the owners of X and Y, have a view of whether the data has changed, the Incident Response Team can investigate any failure paths or symptoms, and managers must be involved in the recovery decision-making process. In a study by Werlinger et al. [75] some organisations stated that the potential cost of the incident was communicated to the manager and the manager decided whether to continue.

2.1.2.4 Eradication

In the eradication phase, the CSIRT cleans, removes or re-images systems. In other words, the incident team knows where the threat is so they are required to remove the threats or a piece of malware and clean the compromised assets. Their work depends on what that situation is and what is left in a compromised computer. Sometimes the task of cleaning requires a complete re-imaging to ensure that any malicious content was deleted and to prevent re-infection.

Metzger et al. (2011) Provide more insight into the types of responses that CSIRTs normally perform. Ahmadetal. (2012) Emphasise the need for communication and cooperation during the Eradication phase. The challenge reported in [86, 87, 95], is the lack of sufficient staff with the expertise to thoroughly investigate the chain of evidence. In some cases, the organization requests a third party [87, 92].

2.1.2.5 Recovery

To avoid further assaults, the compromised assets and services are returned to regular functioning during the recovery phase. Recovery also involves restarting the vulnerable production systems to prevent additional attacks. To guarantee that all of the affected systems are back in working order, the incident response team must test, check and track comprised systems. Finally, they must monitor the system and servers for a set period of time.

2.1.2.6 Lessons learnt

According to a study by [94], it was considered important to learn from the incident. However, the organisation actually found it difficult. The learning phase is important in incident management that can bring many benefits such as: providing security personnel with up-to-date information on current threats, who may obtain new ideas for resolving difficult incidents, and involves discussing the incident management process and it is potential for improvement of the incident management process and its activities [75, 86, 87]. This phase aims to conduct a retrospective of

the incident by completing any documentation after the conclusion of the incident. The incident response team should prepare full incident documentation that will help to prevent such incidents further. The documentation helps investigate the incident, understand **Who** was involved? **What** happened? **When** did it happen? **Where** did it happen? **Why** did it happen? **How** did it happen? And to understand whether anything could enhance the incident management processes.

The CSIRT develops measures that include adjusting security policies or any other future plans that will be used as training materials for new team members or as a benchmark.

2.1.3 Cybersecurity risk management

Organisations are like big biological units, and the challenge today is that every organisation is being bombarded by increasingly sophisticated threats. The attack surface that those threats are trying to exploit is increasing dramatically because of the rate of integrating new and disruptive technologies such as clouds and IoT[4]. The critical objective of protecting enterprises' assets and services from threats is how to build security inside these new disruptive technology platforms that they are adapting, while managing the legacy systems. Organisations need to have processes for addressing all of the new cyber risks that are coming out today. So, they are looking for protection processes by combining technologies, strategies and user education that can provide better diagnostics to protect their assets and services [9].

In particular, cybersecurity risk management is a continuous process that encompasses three processes (risk assessment, risk mitigation, and evaluation and assessment) to help organisations tackle the many security challenges they face on a daily basis [96]. These processes support risk-based decisions and improved cybersecurity, reducing costs related to managing security risk, and improving the overall cybersecurity posture. According to Humphreys [97], a risk cannot be properly managed unless it is thoroughly understood. Organisations use the cybersecurity risk assessment process to identify their assets and assess the impact and likelihood of a threat occurrence to provide a more rigorous management approach [9].

Despite the value of updating threat feeds to avoid zero-day attacks and strengthen the organisation's defense, it is important to identify local assets and their vulnerabilities as local information is important to ensuring a valid risk assessment and threat response. Risk assessment involves three factors: the importance of the assets at risk, how critical the threat is, and how vulnerable the system is to that threat. This risk is calculated based on the ***Likelihood = threat ranking × asset attractiveness × remaining vulnerabilities***, which is a countermeasure determination [98]. A review monitoring and reporting on the risk assessment are fundamental to choosing the most appropriate risk treatment options for effective cybersecurity. In other words, conducting a

[4]Internet of things

risk assessment enables enterprises to identify assets, threats and vulnerabilities in order to make informed decisions about which controls to use.

2.1.3.1 Analysis of cybersecurity risk management

Several studies have been conducted to show how the cybersecurity risk assessment process is done in practice. They are classified into three categories. The first group focuses on the organisations that whose assets are occasionally exposed [68, 99, 100]. The second group of literature focuses on organisations that consider the costs of risk as part of the business budget [101–104].

A review of these studies highlights the lack of security awareness and that security executives lack holistic security knowledge. The results imply that important security data are not gathered in the current decision-making processes, and the risk is not estimated. Cybersecurity risk management is not a distinct entity separate from other corporate processes; rather, it is a vital part of operating a modern business and assists an enterprise in gaining and maintaining a strategic edge over its business rivals. Typically, organisations do not have mature cybersecurity capabilities, and they are not incorporating risk assessment into other business processes. Attempts found in [105–107] helped businesses by proposing integrating risk assessment as part of management.

While standard cybersecurity solutions such as building stronger antivirus applications and firewalls to resolve cybersecurity risks and threats are still essential, they are no longer adequate. A comprehensive approach to cybersecurity risk management is expected across the entire organisation, including supply chains, networks and the broader environment. As a result, organisations must elevate cybersecurity risk management from a mid-level functional role to the boardroom and top management where strategic decisions are made.

Baskerville et al. in [68] state that the goal of several enterprises' cybersecurity risk management policy is to invest in advanced proactive controls aimed at mitigating proven threats, rather than in a sophisticated solution mechanism to counter uncertain complicated and emerging threats. As a result of implementing a prevention-focused approach, companies are well prepared to deal with static and predictable cybersecurity risks. They are, however, more vulnerable to dynamic and complex cybersecurity threats such as APTs [68]. Risk-driven and control-centered security management mechanisms have proved to be very useful in the static prevention of predictable attacks, but they are not well suited to complex response to unexpected threats such as APTs. In addition to building advanced response capabilities, organisations must make a radical structural change in which they leverage both preventive and response modes to their benefit as part of their cybersecurity risk management approach. Baskerville et al. [68] also advocated for the advancement of modern processes in cybersecurity environments that face comprehensive and sophisticated threats.

Even though organisations are aware of the importance of risk assessment, the cyber analyst still plays a vital role in assessing risks and identifying potential threats and vulnerabilities. Therefore, this thesis pays considerable attention to assisting cyber analyst capability that may help organisations achieve their business objectives.

2.2 Fraud Detection

Most financial transactions, such as utilising a credit card system, a telecommunication system, or a healthcare insurance system, may now be completed via electronic commerce systems as a result of the increased usage of computer technology and the continual expansion of businesses [108]. Unfortunately, both legitimate users and fraudsters use these platforms. Furthermore, fraudsters use a variety of methods to break into electronic commerce networks. Concept drift, support for real-time detection, skewed distribution, massive amounts of data and other challenges impede the effectiveness of fraud detection systems [108].

Many different types of studies have used various fraud detection techniques to conduct exploration, discovery and analysis. The majority of proposed detection solutions are supervised methods that use patterns to distinguish between known fraudulent and non-fraudulent transactions [109]; however, these methods rely on labels of fraudulent transactions in historical databases – information that is most often scarce or non-existent [110]. Fraud is very complex and constantly changes, so the supervised methods learned from samples that are not precise. Unsupervised methods that do not use prior information but are able to detect changes in behavior or unusual transactions are more interesting [110] however, transparency is required to increase the trustworthiness of these methods. In other words, the transactions and associated accounts must first be visualised before being investigated by analytical methods for fraudulent behaviour. Furthermore, the complexity of financial data necessitates the use of visual analytics which focuses on dealing with dynamic, heterogeneous and massive amounts of data, and is integrated with human judgement [111].

To analyse financial data, a great deal of research has gone into proposing advanced visualisation and interaction techniques. Visualisation refers to technologies that enable users to 'see' information to help them better understand and contextualise it [112]. FinanceVis [113] is a browser tool for searching papers related to financial data visualisation that contains more than 85 papers [114]. Furthermore, various surveys are presented to review the approaches to financial data exploration, such those conducted by Sugahn et al. or Roger in [114, 115].

The [115] is a survey exploring financial data in general. Despite reviewing many event detections in financial business, the fraud detection approach was not covered. The main highlight of this survey is its support to researchers who need to design and develop better systems to reach

dedicated goals. The [114] is a survey that studied 40 visual analytical approaches by focusing on fraud detection. The similarities and differences of fraud detection tasks and approaches in financial domains are highlighted. It looked at a variety of financial market tasks that have a lot in common with other domains. The most popular visualisation techniques are listed in this survey [114], and graphs (called node-link in the survey) are second most popular, after line plots. A graph is a node-links diagram used to analyse trading networks based on their behaviour.

When large entity-relationship datasets are visualised in a graph, some scalability problems such as visibility, usability and a high degree of nodes are likely to appear [114]. However, to be enabled to examine this massive, multi-dimensional, multi-source, time-varying information stream of datasets that change over the time, visualisation analytics is an ideal candidate. The importance of interactivity is emphasised in order to make proper progress in the visualised analysis process and to solve the challenges [116] - an interactive graphical analysis with queries to investigate the relationships [117]

In general, early approaches focused on providing interactive features rather than interpretation that generates meaningful and human-readable explanations from graphs [118]. The current approaches were proposed for automatic fraud detection. The approaches were designed from traditional anomaly detection to the latest deep learning models [119]. One of the difficulties in applying complex fraud detection models is that there is no easy way to explain how these methods work and how the model makes decisions [119]. There is no easy way to assess the predictive thinking of complex machine learning models and deep neural networks. For applications, in particular, providing effective interpretations for analysts is paramount and a regulatory requirement in many application domains. The generation of the communication graph and scenario-matching approach proposed in the explainable intelligence used in financial transactions, Chapter 6, are primarily concerned with this interpretation and transparency feature.

2.3 Cognitive Science

Philosophy, psychology, artificial intelligence, neurology, linguistics, and anthropology are all part of the multidisciplinary study of the mind and intellect known as cognitive science. Cognitive science's conceptual roots may be traced back to the mid-1950s, when academics from many disciplines began to construct theories of mind based on sophisticated representations and computational methods [120]. Paul Thagard's book [120], provides an introduction to this interdisciplinary field for readers who come to the subject from a variety of backgrounds and are looking for an integrated view of the achievements of cognitive science's various fields.

With the multiplicity of perspectives and methodologies that researchers from many areas bring to the study of mind and intelligence, cognitive science has unifying theoretical principles. The

nature of human knowledge has been explained via attempts to comprehend the mind and its activity [121]. Computational models of how individuals react in trials can be used to integrate psychology with artificial intelligence. Multiple techniques are the greatest way to comprehend the intricacy of human thought, especially when people are acting and detecting in complex settings like cyberspace or the financial landscape. Over the last decade, computer systems that detect, infer and act based on a deep understanding about human cognition's capabilities and limits have received a lot of attention.

In the cyber domain, cognitive science refers to the integration of human thinking into context to improve understanding, reasoning, and analysis that address cyber defence issues [122]. Mahony et al. [123] Presents the results of cognitive task analysis by subject area experts to clarify the design requirements of cyber situational awareness tools [123]. The cognitive aspects of situational awareness are related to the ability of humans to understand the technical implications and draw conclusions to make informed decisions. Therefore, cognitively, it is interesting to measure the degree to which human decision-makers are aware of the situation [32].

2.3.1 Cybersecurity situation awareness (CSA)

In prior research, automated CSA tools and models which aim to enhance experts' cognition have been proposed [124, 125]. As such, situation awareness systems have been designed to combine data from multiple sources to comprehend threat states [55]. The majority of existing research on CSA has not addressed the continuously evolving cybersecurity challenges well. Despite being helpful, security experts still have to digest vast volumes of data and discover the hidden links and dependencies, as a self-awareness [14].

D'Amico and colleagues [38] proposed perception, comprehension and projection as the main phases of cyber analysis aligned with SA. They demand substantial cognitive energy from a cyber analyst, such as decision-making, problem-solving and judgment [11]. In part, CTA reflects the goals of awareness when analysts comprehend the network's state and predict the future state of the network.

As stated in Chapter 1 CCSA [11] is defined as a new label-human awareness of the network for defense performance, which is different from the data fusion concepts, emphasising enhancing the SA of the human in cyber operations [11]. Despite the high levels of uncertainty and highly dynamic environments, an alternative to reducing this cognition limitation is a cognitive CSA system.

A CCSA system is defined as a system emphasising the use of cognition to make sense of the current situation and make decisions in real-time, and includes the term self-awareness and shared

awareness. Additionally, cyberdata investigation is improved through the better contextualisation of CSA and helps to establish an understandable common ground among human beings.

2.3.1.1 Self-Awareness

Some literature has focused on analysing the cognition of computer systems. This literature borrowed the term self-awareness from the field of psychology by adapting it to computer systems where systems generate knowledge about themselves and the environment in which they operate [126, 127]. In the field of cybersecurity, cognition solutions are expected to enhance human capacities in perception, comprehension and projection, which are part of human learning [14].

Self-awareness occurs in the re-construction of an incident, where an individual is in a complex reality but has enough understanding of the whole of the incident. Self-awareness is interpreted as an ability to obtain knowledge about the incident based on internal and external events [128]. The individual must then comprehend the situation (enabled by skills, training, competence and culture, automated tools, structural factors, situational factors etc.). Self-situational perception emerges during the construction of a conceptual model while a human is immersed in a dynamic world (rarely has a complete understanding of the system) and is limited by obstacles (language, lack of knowledge, etc.) [129].

The identifier of the incident will need to know the type of incident (what), the source (who), and the target (how, why) as well as time and location information. Hence, a story about an incident that occurs is crucial to empowering SA to make the correct decision as a response [61].

Self-SA is necessary because a clear understanding of the incident is required in order to make the correct decision. The expert will have to project the incident's consequences into the future. A person's internal heuristics are used to create a logical map of the event, which must then be communicated through a preliminary report [129]. This stage is critical since inadequate information can lead to bad decisions, but too much information can be overwhelming. End-users, for example, may be unable to determine the cause and target of an event and must rely on subjective judgement. However, with the help of additional enablers such as situational, structural and automated resources, the goal of self-SA is to allow the person to define, comprehend and project.

I highlighted instances for explaining of SA, in the context of this research, security logs, cyber alerts, and financial events.

- Self-awareness can be gained through logs when an individual recognises what the logs' relationships depict as a suspicious pattern. For example, numerous failed logins, followed by successful logins, followed by the creation of a user, can warn the analyst to the situation

- A cyber alert has already discovered malicious relationships: when a suspicious behaviour is carried out, a self-SA here obtains further information about the malicious activity, including what happened, who was involved, and so on, in order to understand why and how it occurred
- Many dollars are traded and documented in the context of financial events. Self-SA is obtained when an analyst can determine the relationship between the sender and the receiver, as well as whether the amount is reasonable or malicious, based on previous records and personal knowledge.

2.3.1.2 Shared-Awareness

Shared awareness involves all key role players (managers, asset-owner, end-users, etc.), including the response teams and analysts. Shared understanding, trust, coordination, common ground and commitment through the incident management process or fraud detection are positive consequences [61]. Interaction, visualisation, synchronisation and sense-making are key components of the process. Instead of being sequential, these core elements can be cyclic. Interaction involves the exchange of incident-related information in order to construct a common conceptual model of the situation through the use of communication methods. According to Franke and Brynielsson [32], the human-computer interaction may also be used to explain the cause of the event in order to gain a better explanation of the situation. The benefits of cyber-information sharing cannot be achieved unless a large number of parties engage.

Synchronisation is the process of bringing together people's mental models in order to achieve a shared interpretation of an event. In order to act on the details, sense-making entails developing a common interpretation of the collected data. Before experts can look into the future, they must first make sense of the situation. This will entail planning and scheduling tools as well as decision-making aids. Intelligent systems that automatically fuse vast data into concise meanings, process meanings, achieve observations and theories, access intuition, and present knowledge in meaningful ways can all be used to make sense [130]. Sense-making aids in the selection of the most appropriate frame from a variety of options. A frame is a mental model that detects gaps and predicts outcomes. The final and essential part is shared reporting, which informs the members of the full scope of the incident. Elements of sense-making by storytelling and interaction and visualisation through a knowledge graph create a shared model of the situation in more depth that will be used to understand the basis of the incident.

The use of graphic tools to map the incident is known as visualisation. Visualisation is an important part of SSA. For example, visualising big data graphs or cognitive task analyses may aid in the formation of a shared mental model. The co-operative exchange of knowledge between humans is important to an effective incident management process. Even a single contribution,

a new indicator or observation by an actor, can raise the whole community's knowledge and security.

I highlighted instances for explaining of shared SA, in the context of this research, security logs, cyber alerts, and financial events.

- Logs from various users, systems, components, and other sources are combined in a system for further analysis. The expert who analyses the logs in order to achieve self-awareness of what they are truly saying may or may not have sufficient knowledge of others to whom the logs belong. In an ideal state of SSA, all entities have access to the shared information and can act on it. For example, if a log shows a server was shut down for a period of time, the server owner is the only one who can prove whether the shutdown was intentional or unintentional. As a result, the owner should be included in the investigation to determine the right course of action
- Finding the relevant people from an alert investigation is easier than finding them from logs, because what is a compromised asset, where the incident occurred, and who owns the riskare usually provided in alert, or an analyst located them through a self-SA process. So, shared SA will be realised if the alarm can be shared with the owner of the threat risk, or the unit that owns the victim asset, so that they may be more engaged in the investigation
- • Shared SA can be valuable in financial events as there are a lot of hidden goals in transactions that an analyst isn't aware of. For example, one unit may validate any event's supply claim, while another unit may not. When one of the members of the unit to whom the transactions belong engages in fraud detection analysis, the relationship between transactions and where the money was actually paid becomes clear. As a result, shared SA will be obtained, making it easy to respond to detected factual behaviors.

2.3.1.3 Contextual Situation Awareness

Contextual awareness improves asynchronous awareness notions by explicitly including context information. Context-aware systems gather context data and adapt their behaviour accordingly [131]. Mechanisms for promoting awareness are based on observations made by supported collaborative working contexts. There are various contextual models that a content-aware system is designed based on it; The contextual models are classified into: *Key/value*, *markup schema*, *graphical*, *object oriented*, *logic base*, and *ontology base* [131].

The ontology-based model is popular and, according to [131] survey, has four main characteristics: simplicity, flexibility and accessibility, generosity, and expressiveness. Defining ontology is a common approach for finding the links to rich background knowledge. Examples of this approach

are Nimbalkar et al. [132], Chabot et al. [133], and Bonatti et al. [134]. They offer vocabularies for logs to leverage the linked data representation of vulnerable databases. Ontology engineering approaches, on the other hand, are more rooted in machine interpretation-based approaches to supplement existing detection techniques. The existing approaches are not evolved in the verbose textual description of individual events and linked relations between occurrences from the locality. Furthermore, the various types of information required to deal with cybersecurity issues are not typically proposed in a single ontology. Even while the proposed techniques are beneficial in terms of giving a more diverse range of data, a critical component is overlooked. People's capacity to understand actions within a given context is based on more than just the amount of information provided [135].

These approaches have been developed without providing a verbose textual description to interpret the individual events and linked relations between the events from a locality. Furthermore, various types of required information to cope with cybersecurity issues are not typically proposed in a single ontology. Providing the link between the evidence in the logs and the results in the report is not automatically and technically addressed. As a result, several studies have sought to pay more attention to the design of the reasoning module with the capability to trace the row information. Arasteh [136] proposed a model by labeling the term into a tree by expressing formulas of the logic. An association rule mining and automated planning approach was proposed by Khan and Parkinson [137]. The authors presented an activity plan with pre-defined status and actions to assist in the reasoning behind events' correlation.

This study's models for explainable intelligence is very similar to the ontology-based model, focusing on expressiveness instead of the fixed and non-verbal ontology.

2.4 Technologies to Support Analysis - Enhancing Cognitive Abilities

Technological solutions could integrate and synthesise information with the goal of reducing human interactions in the data analysis process and enhancing cognitive abilities. These solutions are divided into four: black box, visualisation, structured and narrative. In this section, the examples of the works falling within each group will be briefly introduced.

2.4.1 Black Box

The current analytical technologies that help analysts began with supervised/unsupervised detection methods is known as the black box methodology. To distinguish between normal and abnormal (malicious or fraud) activities, the results are usually presented in a boolean format.

They typically rely on temporally pattern-based solutions to integrate and synthesise data, and they frequently fail with time as there is no explanation of their internal structure to justify *how* and *why* the analytical process works. Furthermore, real-world data is frequently auto-correlated, and its classes are not evenly distributed. It is also difficult to discover data that has been labelled.

For instance, abnormal activity was recognised by the application of various machine learning techniques including Naive Bayes, K-Nearest Neighbours, and Support Vector Machines to high volumes of logs by Muggler et al. [42]. Bertero et al. [138] used NLP as a feature extraction tool in some classification algorithms to mine more precise information from log files and detect anomalies more effectively.

Some studies attempted to leverage more knowledge rather than focusing exclusively on boolean outcomes (normal or abnormal). As a result, they proposed analysing data based on predetermined criteria in order to comprehend. As a consequence, in order to better understand the mechanism, they have proposed analysing data using specific preset criteria. Tuor et al. [139] used a neural network to remove certain user characteristics such as roles and attributes. The results are used to predict the next function vector in order to identify insider attacks in an organisation early.

The black box method is often considered to be untrustworthy as there is insufficient reasoning about the situation and label assignment. As an example, a company simulates cyber attacks by a penetration test. Such activity should not be labelled as abnormal as an authorised person performed it. Given no explanation, how can vulnerability scanning operations be differentiated from real-world attacks? Without sufficient SA from data presentation, actual attacks could be mistakenly disregarded by an expert.

2.4.2 Visualisation

A significant body of literature has already sought to involve human supervision in the data analysis process with the use of visualisation techniques [140]. The use of visualisation as a presenting approach aids human cognition and aids in the detection of possible problems [141]. In the work by Xu et al. [142], for example, a decision tree (as a level of analytic presentation) was utilised to show how the system chooses whether to give a normal or abnormal label to a log record. As in the previous example, certain visualisation approaches are relatively simple and restricted by specified criteria which do not provide a full perspective.

A graph is another presentation that Aharon et al. [143] utilised to show the state of system behaviour. In their work, based on the clustering method, the graph depicts distinct groups of log messages together with their labels (normal process or failure process). It is easier to analyse important information and data about an event when messages are clustered together. For

example, all logs recorded from multiple devices associated with a specific user may be shown in a cluster.

The financial landscape has taken a keen interest in AI and machine learning solutions based on graph computing ideas. For the future of fraud and financial crime detection, graphing and developing adaptive solutions present appealing possibilities. However, incorporating graph-based solutions into financial transaction processing systems has revealed many challenges and issues.

Graphs have natural advantages when it comes to representing cyber events or financial transaction data. Companies, individuals, accounts, fund transfers, locations, devices, and other financial or non-financial data are frequently represented by nodes and edges. However, graphs require lots of cognition effort to be digested by analysts. Although clustering comparable messages on the graph is beneficial, it does not give more explanation as to *why* the messages belong to the same category.

Grouping events or transactions with the addition of search features(query-based form) provides more valuable information for analysis. Samii and Koh [4] considered more aspects of events by providing a search capability in an interactive query-based system. The information was displayed on an interactive visual interface from a high-level view to the original log files. In [144], Li and colleagues suggested a system to handle multiple forms of event logs by giving a simple approach to analyse them. The statistical data from the logs was extracted and displayed on a dashboard where the user could interact with the data through queries. Although an interactive dynamic query-based form has been provided to aid in the investigation of further information about an event, it is limited to particular graphical features, making it impossible for analysts to provide a comprehensive analytical understanding. For example, if the HTTP post method is not considered a design feature for searching in the interface or dashboard, an expert cannot search all connections with it. By considering more design features, a high level of knowledge and specialist training are required to understand what should be searched, and what should be expected from the results.

The proposed interactive visual interface by Samii and Koh [4] is demonstrated in Figure 2.1. As Figure 2.1 shows, the nodes in clusters in the visual interface are not easy to understand, especially for someone who is not trained well. Furthermore, the options for searching through logs are designed to a limited specification, making them incomplete.

As previously said, visualising events and offering search capabilities was not an understandable insight strategy for delivering a comprehensive view of occurrences. Certain visualisation techniques are simple and limited by predetermined criteria so they don't provide a complete picture. Azodi et al. [145] attempted to address the issue by identifying attack pathways and displaying them in a way that provides additional information. To gain a better picture of the

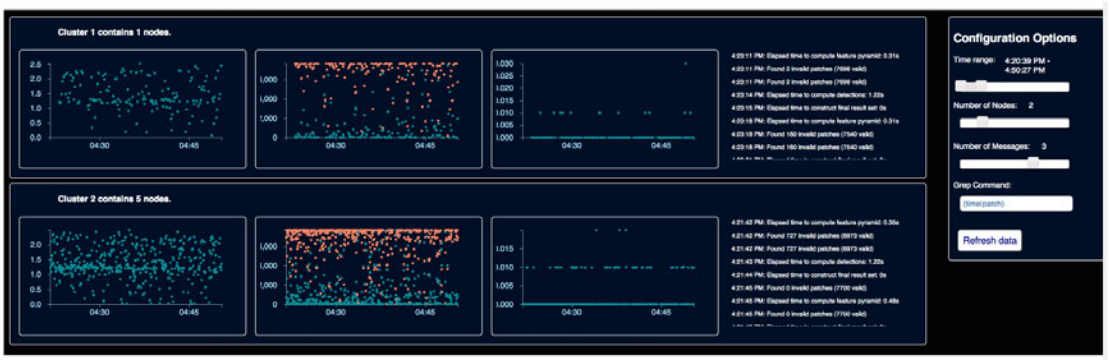


FIGURE 2.1: Interactive visual interface for analysing logs, proposed by Samii and and Koh [4].

attack's progress, events connected to an alert were found using regular expressions. A graph depicted attack pathways, and a connection demonstrated correlation between different attacks. Although the visual graph has more design features and information about the connections between sources and destinations, it lacks the necessary details and explanations for instant inference. For example, the graph may show a suspicious link between the organisation's inner server and an external web site however, it does not specify which functionalities were used over this connection, such as the HTTP method. Overall, the existing visualisation interfaces do not provide enough information to distinguish between normal and malicious connections, making it difficult for a cyber specialist to be fully aware of the situation.

2.4.2.1 Knowledge Graph

There has recently been a tendency to use background information, which may be represented as a knowledge graph [146], to better comprehend an entity's semantic connections. Knowledge graph applications have grown increasingly popular in recent years, with notable achievements in a variety of data-driven applications [147]. For several decades, knowledge graphs have been widely explored and used for information processing and organisation. Ontologies or semantical graphs are the names given to the earliest knowledge graphs.

With the emergence of big data and, in particular, the Google search engine, they have grown into sophisticated graph databases that represent many characteristics and the relationships between entities [148]. Visual question answering and connection extraction have been accomplished using knowledge graphs [148]. The model's ultimate output can be influenced by the structure of the knowledge graph (in a comprehensive model that humans better understand and require less cognitive effort). In the context of ambiguity, the use of AI to find the nearest entities in semantic knowledge graphs is useful, and can shape the output model to be more intelligent.

A knowledge graph is a data structure capable of modelling both real-world concepts and their relationships [149]. It is described as a graph to semantic knowledge. Each node in the knowledge

graph represents a real-world concept, whereas each edge represents the relationship between two concepts [147]. Crowdsourcing or automated extraction from online data or knowledge bases are generally required to create large-scale knowledge graphs. Once they detected an event or transaction, objects are mapped to knowledge graphs, and their connections to each other may be understood [148]. When two concepts have a high degree of semantic consistency, it is typically assumed that they are tightly connected in the same event or transaction [147].

Knowledge graphs act in a similar, though distinct, way to the human brain. They reflect the knowledge domain as well as the relationships that exist between entities [148]. Knowledge graphs have become popular for storing and organising massive volumes of data such as logs and transactions. Other approaches to store external knowledge in ways that allow for the calculation of semantical distances between items, in addition to knowledge graphs, are possible. These techniques can be used independently or in combination to increase awareness and comprehension of cyber incidents or financial activities.

Knowledge graphs may be stored in any type of back end, from files to relational databases to document stores. However, because they are graphs, it makes sense to store them in a graph database. This simplifies storage and retrieval considerably as graph databases have specialised structures, APIs, and query languages designed specifically for graphs. Furthermore, many graph databases provide much more than a data storage facility. They include graph analytics methods, visualisation capabilities, machine learning features and development environments. They have progressed from databases to platforms.

2.4.3 Structure and format of sourcing logs

Numerous studies have attempted to change the log structure into a rich format to improve understanding. Nimbalkar et al. [132] translated log files and added semantics keywords. The results are demonstrated in the semantic RDF linked data which is a machine interpretable representation. For cyber analysts, the lack of concept definitions and their relationships is one of the possible drawbacks of machine-readable formats. Furthermore, the representation format (machine-readable) is particularly challenging for non-experts who do not have time for training to interpret the data. To summarise, RDF as a structured data format is highly machine-readable, but it is not regarded a viable choice for human reporting and analysis. Machine-readable formats not only make it difficult to bring about self-cyber awareness, but they also make it difficult to achieve shared awareness when they are exchanged across incident management partners.

Despite the fact that much research has been conducted in the field of incident management and response in recent years, only a few researchers have concentrated on the data structures and processes for the sharing of security incident information. In incident management, information exchange formats were highlighted to enhance knowledge of every single participant to address

the lack of comprehensive analysis in the use of gathering all significant aspects [5]. Two famous formats for exchanging information are Structured Threat Information eXpression (STIX) [150] and Incident Object Description and Exchange Format (IODEF) [151]. STIX is focused mostly on cyber threat intelligence from a holistic perspective, and IODEF is concentrated on attackers' and defenders' information. They are created for various purposes [152], and their machine-readable format makes it extremely challenging to understand the components and the relations between them.

An in-depth analysis of the STIX contains representations for a wide range of elements. The root components of an incident are covered by incident and indication objects. The threat actor object, which can be traced to single events or attack campaigns, describes the attacker variable. It includes basic characteristics as well as details about the attacker's current personality and goals [5]. Objectives are represented as a mixture of the underlying intent and the desired result, backed up by a metric of estimating the degree of complexity of the attacker. While STIX does not demonstrate any direct representation of an attack object, it does provide representations for all specified attack components. The representations for 286 object properties and data types are possible with STIX [153]. STIX has a high machine-readability due to its extensive material coverage and sparse use of free-text properties. However, because of its intrinsic ambiguity, it has drawbacks in terms of human readability. As previously said, STIX has simple frameworks and distinct entity representations for events, markers and related object entities. As a result, depending on the usage case, it has a very limited propensity for unclear representations [5].

On review of the IODEF in depth, there is little need of the intended indicator items. IODEF includes various objects for representing threats which are specifically merged into the base object incident. The use of objects, which allows for the exact specification of network nodes, processes and utilities, allows for the description of operation behaviour as well as the determination of the impacted target. IODEF has an entity mechanism for expressing attack strategies and techniques, as well as basic information about the exploited vulnerability. [5]. IODEF offers implementations for 99 object properties and datatypes [153], resulting in slightly poorer content coverage than STIX formats. For communicating additional incident detail, IODEF heavily relies on free-text representations [5]. This dramatically reduces machine-readability capabilities while increasing human-readability. IODEF provides overlapping components of identical interpretations such as incident and event data [154]. As a result, it has uncertainty issues with format semantics.

X-ARF [155] is the only human-readable exchange format proposed. X-ARF is an approach to structured data representation that relies on a straightforward implementation and hence offers very limited functionality. It requires simple attack information as well as information about the attacker to be represented. Details regarding defense steps and indicator components are not provided [5]. Due to its slight sophistication, X-ARF only has very limited functionality for an automatic sharing of security information, as well as poor system readability. This, on

the other hand, results in excellent human readability. X-ARF is a basic format that can only exchange limited types of malicious alerts via an email. The email contains limited information such as alert description, alert category, and initial information about the attack and attacker [156]. The exchange formats transfer alert messages to a new structure and add descriptions to enrich it. Therefore, the main aim of them is sharing the alert message, not interpreting the alert message and providing more evidence for improved understanding. Furthermore, because of the low degree of ambiguity, there is a very low risk for unclear representations.

A comparative analysis of the most important incident reporting formats by providing an overview of the weaknesses and strengths results of them is done in [5]. Menges and Pernul [5] during their investigation discovered a table that shows the various degrees of criteria for each exchange format. Based on the table, there are not any exchange format with degree of fulfilment for the defining criteria, especially human-readability has the low level of attention among the exchange formats. The proposed table is shown in Figure 2.2.

	STIX	STIX2	IODEF	IODEF2	VERIS	X-ARF
Indicator	●	●	○	●	◢	○
Attacker	●	●	○	◖	●	◖
- Attributes	●	●	○	●	●	◔
- Objectives	●	●	○	○	●	○
Attack	◔	●	●	●	◖	◢
- Event	●	●	●	●	◖	◢
- Action	●	●	●	●	◖	◢
- Target	●	●	●	●	◖	◢
- Methods and tools	●	●	●	●	◖	◢
- Vulnerability	●	●	●	●	◖	◢
- Result	●	●	●	●	◔	◢
Defender	●	●	○	◖	●	○
- Reaction	●	●	○	◖	●	○
Contentual coverage	●	●	◖	●	◖	◢
Machine-readability	●	●	◖	◔	◖	◢
Human-readability	◢	◢	◖	◢	◖	●
Unambiguous semantics	●	●	◖	●	◔	●
Interoperability	◔	●	●	●	◖	○
Extensibility	◢	◔	◔	◔	◔	◖
Aggregability	●	●	●	●	○	○
Practical application	●	○	◔	○	◖	●
External dependencies	●	●	○	◔	○	○
Licensing terms	●	●	●	●	◖	●
Maintenance efforts	◢	●	◢	●	◔	◖
Documentation	●	●	◖	◖	●	◢

FIGURE 2.2: Analysis of incident reporting formats based on Menges and Pernul's comparison [5].

2.4.4 Narrative analytics

The narrative analytic methodology was introduced as a powerful capability that can assist organisations in developing a dynamic analytical process. A great deal of effort has gone into describing incidents and their elements in incident management [157, 158]. However, neither the potential partnerships across incident reporting formats nor the requisite changes for this method of use, have been discussed. Similarly, various representations of an event detection mechanism have been proposed. Incident prevention and response [159, 160], and computer forensics works [161, 162] are the key sources of certain recommendations.

While narrative activity is a sense-making process rather than a finished product [163], a narrative explanation can be a good candidate in analysis facilitation. Currently, no efforts have been made to assist cybersecurity analysts or financial auditors through the use of the narrative framework as a presentation format. Wu et al. [58] proposed a data-driven storytelling system for the improvement of social connections. The system transformed sensor data from IoT devices of elder's conditions for their loved ones in order to support a social connection between an alone elder and his/her family. Raw data was mapped to semantically meaningful variables through a GoalNet, and the dynamic storylines were generated based on a set of curiosity rules. Wu, et al. [58] only provided one level of explanation in their output results to attract the adult children's attention. Although the system could not explain the details of the elder's conditions and referred to a triggered sensor as evidence, they believed they reached their aims to captivate the adult children's attention. Their attempts to raise awareness through a common ground of understanding are both motivating and beneficial, but they are unrelated to cyber data.

A multi-level story derived from a cyber alert message can be a novel approach to assisting the analytical process in cybersecurity. It can also be applied to other domains easily. For example, explaining various financial transactions in criminal domains. Simple concepts in sequential sentences can be organised to discern where the events are heading. It is easier for human beings to identify correlations of events in the log files or financial transactions when they are modelled using a storytelling design [164]. Recently, technology has progressed in the direction of transparency, revealing more information about the situation and attempting to to analyse data automatically. As a result, using a storytelling framework is a good option for meeting current demand.

While the idea of generating story by machine is not new, the domain in which it is applied is. As an example, robot language generation or automated journalism is the recent accomplishment of current investigations which can be found in [165–168] and reviewed on the potential level of descriptions by Caswell and D*ö*rr [169]. Automated journalism refers to the process of using software or algorithms from natural language generation (NLG) technology to automatically generate certain routine writing tasks within news. The obvious use of automated journalism is

the writing of routine sports and financial news. An example of an earning report is shown in Figure 2.3, which is an Associated Press report that was published shortly after Apple released its quarterly figures in January 2015.

Apple tops Street 1Q forecasts

Apple posts 1Q profit, results beat Wall Street forecasts

AP. January 27, 2015 4:39 PM

CUPERTINO, Calif. (AP) _ Apple Inc. (AAPL) on Tuesday reported fiscal first-quarter net income of $18.02 billion. The Cupertino, California-based company said it had profit of $3.06 per share. The results surpassed Wall Street expectations. The average estimate of analysts surveyed by Zacks Investment Research was for earnings of $2.60 per share. The maker of iPhones, iPads and other products posted revenue of $74.6 billion in the period, also exceeding Street forecasts. Analysts expected $67.38 billion, according to Zacks. For the current quarter ending in March, Apple said it expects revenue in the range of $52 billion to $55 billion. Analysts surveyed by Zacks had expected revenue of $53.65 billion. Apple shares have declined 1 percent since the beginning of the year, while the Standard & Poor's 500 index has declined slightly more than 1 percent. In the final minutes of trading on Tuesday, shares hit $109.14, an increase of 39 percent in the last 12 months.

This story was generated by Automated Insights (http://automatedinsights.com/ap) using data from Zacks Investment Research. Access a Zacks stock report on AAPL at http://www.zacks.com/ap/AAPL.

FIGURE 2.3: Example of a descriptive quarterly earnings report generated automatically for the Associated Press by Zacks Investment Research data [5]

The lack of data required for application is the most significant obstacle to the development of automated journalism. In terms of the specific data inputs to an NLG platform, the representation handles new characteristics of an event [169]. For an instance, data has been extracted through a search by regular expression techniques from a particular website which contains enough information about technologies and customised by the tags based on the personal preference in [168].

Journalism is also applied based to event-driven stories. This approach, while still experimental, has become more practical through the availability of libraries of event, ontology and knowledge graphs [169]] in newsrooms. However, automated journalism is mostly confined to relatively short texts in limited domains [169].

This study looks into a number of aspects of the event detection process, particularly incident management and fraud detection. Basic features, on the other hand, that allow a detection system to better prepare materials for any information exchange are not presented in the current approaches.

Although the narrative technique has not been employed for cyber data or financial transactions analysis, it will be studied from several angles to establish its applicability to the cyber security and fraud detection arena.

2.4.5 Explainable Intelligence

In incident management, security tools like SIEM help detect, integrate anomalies, and correlate vulnerabilities and threats. Common approaches used in incident management are mainly based on expert knowledge that has made decisions based on the results of cyber tools [170]. Decisions were made as a result of observations collected from cyber tools and subsequent analyses were performed based on their outputs [29]. However, cyber analysts are potentially unable to understand and interact with the output required by tools' context [171]. Since explainability uncovers various aspects to enhance understanding of environments, explainable AI methods that amplify human intelligence, support this need [170, 171].

Early studies on explainable AI (XAI) focused on providing explanations of expert decisions in rule-based and logic-based AI systems, without addressing the current AI methods that are not quantitative in nature [172–175]. More recent studies of XAI are agents-based that used the Markov decision process, which can be helpful to make the decision based on different situations [176–178] Agent-based XAI was the first attempt to develop algorithms for automatically generating explanations with Markov [179]. Transparency has become a very interesting area for researchers, especially in decision-making [180–182]. Various researchers are trying to make decision-making more transparent. For example, Si and Zhu [183] proposed the ANDOR tree as an interpretable model. Shih et al. [184] described the ML process using a Bayesian network.

Most of the research focus is on making AI processes more transparent so that humans can understand what is happening in the ML system. This idea was mainly highlighted when DARPA created an explainable AI program in 2016 [185]. Examples are recommender systems for image recognition [181, 186, 187] and AI playing video games [188, 189]. The level of transparency of

ML algorithms, especially neural networks, addresses the need to explain the decisions behind the model [171]. This allows people to easily change the model based on their goals.

In the cyber arena, explainable intelligence research is still the first step of evaluation to support human cognition through transparent data. For example, in a recent study, Eric and Ning [171] proposed an XAI-driven virtual agent as junior cyber analysts analysed and presented available data on vulnerabilities and incidents related to the target system, with the ability to answer human questions. The agent is a recommender system that transparent data, which helps analysts to better understand incidents, vulnerabilities, and threats.

2.5 Summary

This chapter discussed incident management and fraud detection as applications under consideration for determining the role humans play in their processes, providing background on the key areas of this thesis. Briefly reviewing human capabilities in a SOC or CSIRT, cybersecurity incident response processes based on risk management factors, and analysis of cybersecurity incident management and fraud detection deficiencies are just a few examples. The findings of this chapter's literature review highlight the unavoidable role of humans in incident management and fraud detection processes.

In addition, I discussed the research into cognitive science, specifically self, shared, and contextual SA, and their implications for analysts. The findings showed that SA tools improve security analysts' knowledge, comprehension and execution of security response actions, allowing them to take action in less time and with greater efficiently. Because the thesis goal is to highlight the role of SA in humans in application domains, I brought these two reviews (applications and cognitive science) together to highlight the research gap.

The findings of a literature review of current technology solutions for supporting human analysis and assisting in the reduction of cognition efforts revealed that narrative is a good choice for use as the principal technology in the intended explainable intelligence. Narrative analytics technology was introduced as an analytics framework that may assist organisations in developing a dynamic analytical process in their complex and dynamic environments, such as incident management and fraud detection, in order to improve the detection or reaction process. The created explainable intelligence proposed in this work is described in depth in the following chapters.

Chapter 3

Explainable Intelligence to Interpret Logs - Self and Shared Situation Awareness

Huge volumes of events are logged by monitoring systems. Given the vast volume of records, analysts do not inspect or trace the log files until an incident occurs, which leaves a hole open for potential security breaches. Moreover, the large-scale and machine-friendly format of log files provides tremendous opportunity for task automation in order to reduce the burden on currently scarce cybersecurity professionals. Also, to adequately evaluate the situation and initiate the appropriate response, an extensive domain knowledge and experience is often required. Although numerous algorithms have been proposed to process potentially risky alerts, manual verification still needs to be performed to reduce the overwhelming number of false positives. What is more, event analysis out-of-context, i.e., without correlation with other events, proves not only limited, but oftentimes misleading.

This chapter discusses the proposed designed explainable intelligence to interpret logs that achieve both self and shared SA by providing a contextual background for comprehensive cybersecurity log analysis through innovative storytelling. The explainable intelligence model presents a sequence of events that are discovered automatically from the log file, along with details of subjects and objects to address based on the eventID, effectively reducing cognitive burden and manual analysis through context enrichment. As a result, analysts were learning and becoming more self-aware in the situation, particularly the incident. Because events are presented in a human-readable format, the chain of events presented in the storytelling framework makes sense for other participants in an investigation. As a result, shared SA is also achieved.

3.1 Introduction

Numerous activities take place in a computer system on a daily basis - applications start, firewall updates, user logs in, etc. As a result, millions of activities are recorded in computer system logs. Event logs, or simply logs, are machine-generated records to report sequences of events occurred during operations, and are in the form of text-entries [190]. The monitoring and reporting of events are classified to two main groups: *Network logs*for network traffics, i,e, firewall logs and *Host logs* for operating system or user activities, i,e. Windows security log events. Network logs are typically investigated to that show attempts have occurred, and host logs are used to learn more about an application, services and processes involved in an attempt to determine how an event occurred. In other words, the network logs record what happened and by whom, whereas the host logs detail how they happened. For example, a company that has captured network traffic may investigate a source port that is connected to a destination port, but the application or service that opened this port will be addressed in the host logs.

Logs are usually intended for security and diagnostic purposes. Their data can be extremely useful in system audits and forensic investigations. When monitoring systems generate alerts, the event logs are the first place analysts look. The log file contains rich information including when the problem occurred, what applications were running, and which application might have caused the problem. Until an incident happens, analysts do not audit or trace the log files which record the most critical occurrences. This leaves a security gap that can be exploited. Furthermore, the large-scale and machine-friendly structure of log data opens up a lot of potential for work automation, which may help to relieve the pressure on currently scarce cybersecurity specialists.

Human analysis is a tedious and inaccurate task given the vast volume of log files that are stored in a machine-friendly format. Furthermore, the analysts have to derive the context for an incident using their prior knowledge in order to fully understand what happened and initiate appropriate actions. The human verification of filtered data is required to justify events with minimal false positives. However, what seems to be missing from filtered data is the detailed and relevant knowledge to adequately evaluate the situation and initiate the appropriate response: the extensive domain knowledge and experience that is often required.

Given the complexity of modern systems and cyber attacks, algorithms have not been able to apply sufficient context to data, or contain enough intelligence to understand why certain classifications are important. Furthermore, there is no existing benchmark data, where a normal or malicious label is assigned to a log record. As an example, when a large number of files is deleted from the system, it can be considered either normal or malicious behaviour, based on whether the files were deleted by an owner or malware, respectively. Thus, human beings have to be involved in the analysis process to determine the relationship between the events. Furthermore,

event analysis performed in isolation, i.e., without regard for the context of other events, is not only limited, but frequently misleading.

The designed intelligence for self and shared SA in security event analysis for interpreting logs is proposed in this chapter. I propose a Log-Chain-Driven Storytelling Model (LDSM) as an explainable intelligence model for identifying periodic temporal associations with timestamps, as well as a conceptual model that provides context for comprehensive cybersecurity log analysis in an innovative storytelling fashion. The model is utilised to discover the relationships between events that persist for some duration of time. Since time plays an important role when representing the knowledge of events, the model is developed to recognise the events within the variation of the association rules over time.

In recent years, many studies have presented convincing arguments that time plays an important role in identifying knowledge of temporal data because data typically contains time stamping [191]. The timestamp is the most important part of a log because it conveys information about what happened in logs. In many scenarios, logging into a server after work hours is suspicious activity however, it is normal if it occurs during work hours. Only time can transmit knowledge. Retrieving knowledge from the log file by considering the time is also a very important factor for computer forensics investigations to reconstruct past events and find the relation between them [133]. Digital evidence is based on computer activities, however log files provide only a portion of the story [192]. Thus, analysts use software tools for demonstrating the activities through the timeline and compare them with other discoveries. Interesting events often occur within a specific period, therefore the time aspect is a very important factor in log file analysis [191].

On the technological side, mining is the most prevalent method for extracting information from logs and determining events' interconnections [137]. The sequence of events from log files is filtered out automatically and presented in a storytelling format. Storytelling is a novel analysis representation method that can highlight the semantic and implied information from log files and convert it into a human-readable format [58]. For automated sequential event discovery, the apriori-like algorithm for temporal pattern mining is used. The mining algorithm proposed in this study is similar to the one applied by Khan and Parkinson [137]. Our focus, however, is different. Although Khan and Parkinson [137] used timespan to determine the ordering of event sequences, it was not considered in the mining process for frequent item set. As a result, only the activities that appear multiple times in the log files are identified as frequent patterns, leaving out the activities with a short lifespan. Furthermore, the algorithm proposed in [137] deals with different unresolved conflicts by chaining events, which is addressed in this model. The approach proposed by Mahanta et al [193], on the other hand, takes into account time for retrieving partially periodic patterns, however the algorithm has only been tested in the market-basket problem [194]. Their approach is being used for security events. The proposed model mines the interesting events within the observed period and produces chains of sequential events by extending the apriori-like

algorithm. Through appropriate context enrichment, the interpretation of sequential event chains in a storytelling format, as a novel approach, is presented.

LDSM automatically interprets security logs by using appropriate context enrichment which improves cognition and makes it easier for experts to understand - self SA about logs has been achieved. The model's short narratives specify which subjects and objects must be addressed based on the eventID, effectively reducing the cognitive load associated with manual analysis. As a result, the story design model for security events contextualised interpretation is the chapter's main contribution, as it reduces human effort in identifying relevant relationships between events. As a result, potential risk incidents can be identified and exposed quickly, effectively preventing further escalation and reducing serious security risks, and giving security analysts a unified perspective that promotes comprehension and improves SA. Equipping security analysts with a cohesive viewpoint promotes understanding and improves situational cybersecurity awareness. It can also be beneficial to involve more individuals, such as analysts, IT support, asset owners and managers, in order to understand what happened among the massive volumes of logs and obtained SSA. The experimental results show the potential and benefits of the proposed storytelling model based on security logs, as demonstrated by three real-world case studies on the Windows platform.

3.2 Log-Chain-Driven Storytelling Model (LDSM)

The LDSM is a proposed explainable intelligence model for processing cyber logs that consists of four individual layers and main procedures, as shown in Figure 3.1. The first three layers process the logs and generate a short story which was explained in the paper [2]. Regular expressions are used to obtain both implicit and explicit properties from a log record. In the form of plain text, a short statement describing each event is used as the label for the generated vector from the properties. An apriority-like algorithm based on temporal pattern mining is used to find sequential events automatically. In other words, the interesting activities within the observed time are mined and modeled into chains of sequential events using an extension of the apriori-like algorithm in the LDSM. Furthermore, the text label is used to parse the sequential event chains in natural language by suitable context enrichment. Because the chain may contain more than one event, a chain of statements describing the most related occurrence is generated; like a short story. The final layer, Enrichment and Story, improves the output-short story by providing the object and subject of sentences, making it more human-readable. The following are the details of each layer, including primary purpose and associated steps:

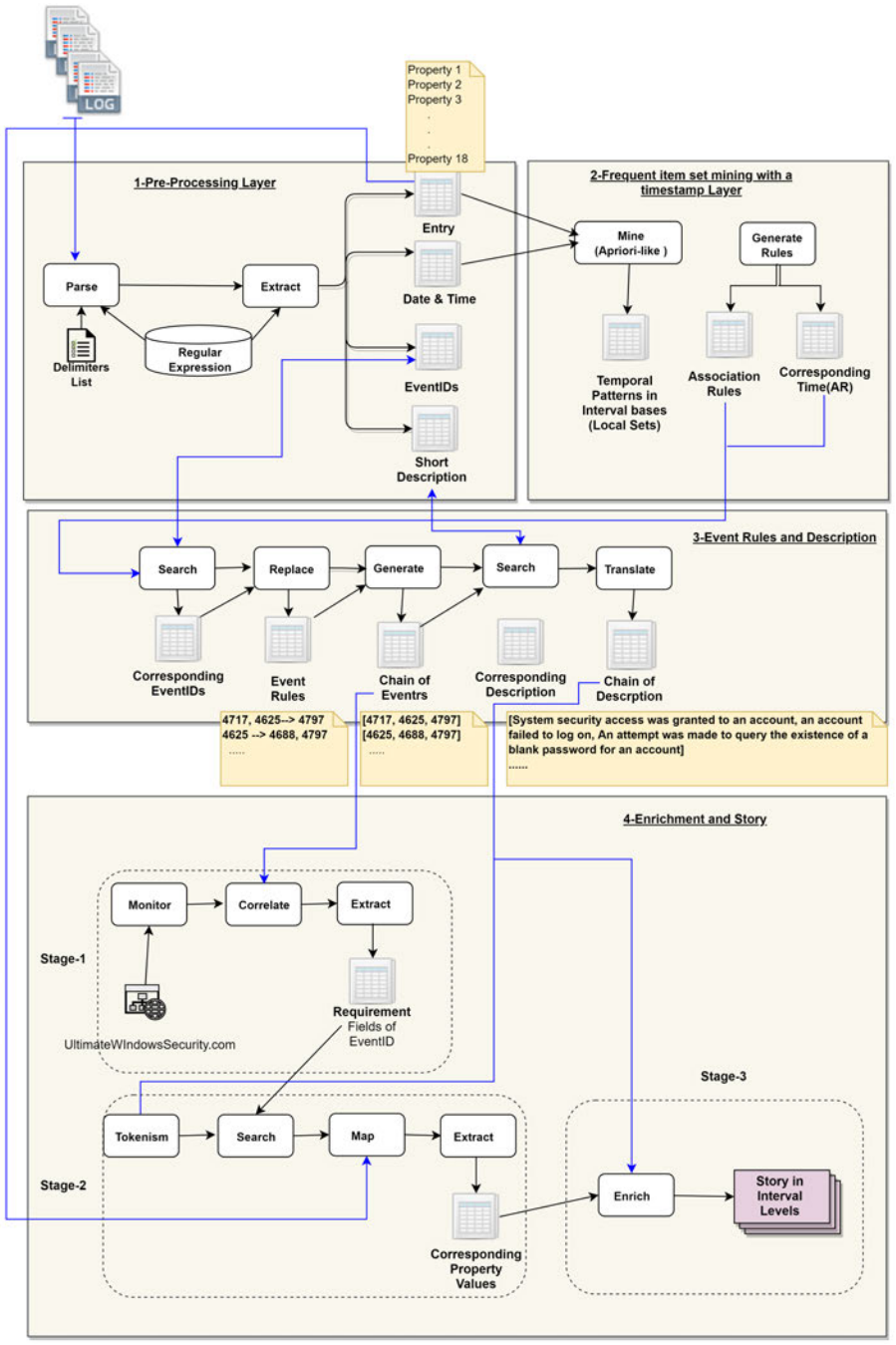


FIGURE 3.1: Overview of the Log-Chain-Driven Storytelling Model made of four layers (beige boxes) and operation procedures (white boxes). The Enrichment and Story layer represents final output multi-levels story (purple boxes) based on time intervals

3.2.1 Pre-processing layer

In this layer, the properties from the log file are extracted and stored in four tables: Entry, Date and Time, EventIDs, and Short Description. Let L={EventID, Date and Time, P_1,P_2,...,P_n} be a log record from a Windows log file, where EventID and Date and Time are the record's numeric event type IDs and the date and time of recording, respectively. P_i is the event property

symbol. Microsoft has defined the common properties of the Windows event log. The property P_i can be considered as an explicit field such as "User", or can be an implicit property which is embedded in an explicit property. For example, the "Error-code" from the "Message" property in the Windows security events is an implicit property. Both implicit and explicit properties are retrieved from the log record using regular expressions and deliminator lists. The message properties comprise the details of the logs that determine the subject and object of the events such as Logon ID, Object Name, Account Name, and Object type. Each event is described in a single sentence in plain text style. In the Windows platforms, the text is named "Short-Description" and is part of the message property (implicit property that is extracted from the message property and stored in the "Short Description" table). A list of the security event descriptions based on the eventIDs can be found in [6].

The Entry table comprises a set of properties which are intended to be a set of items for temporal association mining. A log record is separated into four sets with purposing to store in separated tables; L= {EventID, Date and Time, Entry, Short-Description}. Objects in the Entry table starting with p_{t_1} and ending with p_{t_m}, are log record properties. Each log record corresponds to a transaction with a unique identifier (EventID) and timestamp (Date and Time). Let T=t_1, t_2,..., t_m be a sequence of timestamps where $t_1 < t_2 < ... < t_m$. The tables are arranged in ascending order of timestamps. This means that the N_{th} record in the Entry table occurs at timestamp x in the N_{th} record of the "Date and Time" table, with a corresponding eventID in the N_{th} record of the EventIDs table, and that the definition of this event can be found in the N_{th} record of the Short Description table.

3.2.2 Frequent item set mining with a timestamp layer

It is simple to apply an association mining technique to discover co-occurred properties of log records by defining the log records based on the market-basket problem [194]. The properties that co-occur are associated with the relevant events and describe what is happening.

An apriori-like algorithm is a suitable candidate to discover temporal patterns in interval-based data. Apriori was the first algorithm to enable value-based pruning in order to avoid an exponential increase in the number of potential item sets. A support (support (X)) is defined as the proportion of event log properties that contain both X and Y. It is also known as the probability P (X $\cup$ Y). As the name of the algorithm shows, the main idea of the apriori algorithm is based on the inductive theory. Given an item set ABCD, they would first examine its subgroups ABC, AB, and so on. In other words, if X $\sqsubset$ Y, then it can be concluded that the support(X) $\leqslant$ support(Y). It means that if k-length item set could not be recognised as satisfying the pattern, there is no need to check any m-length item set, where m $>$ k [195].

The work of Mahanta et al. was the main source of inspiration for the periodic mining item set [193]. The local frequent set, according to Mahanta et al. [193], is a collection of item sets that are frequent in a certain period. The gap between the present time and the last-seen time of appearance of a given item set is validated using a threshold. If the gap is more than the threshold, it shows that the most recent time was the end of the previous local frequent set, and the present time is the beginning of the next (next set). The frequency of items in each local set is checked by minSupport. In other words, the local sets for each candidate are defined if the candidate is repeated more than the minSupport times from start to end of the corresponding time interval. If the transaction's timestamp fits inside the interval, it is placed in $[T_i, T_j]$. The $N[T_i, T_j]$ represents the number of transactions that took place within the time interval $[T_i, T_j]$, and the $N(x)_{[T_i,T_j]}$ represents the number of transactions that contained item set x. The following formula is used to calculate the support of a local item set: 3.1 3.1:

$$support(x)_{([T_i,T_j])} = \frac{\left|N(x)_{[T_i,T_j]}\right|}{\left|N_{[T_i,T_j]}\right|} \tag{3.1}$$

In each local frequent set, the amount of support is computed. The item sets could show up in many local frequent sets. As a result, item set x's support is determined by averaging the local support amounts, where each local support is bigger than the minSupport. Every k-length item set is created and recorded as an array. The set of candidates extracted is typically referred to as CK, where C stands for candidate and K stands for sequence length. If the average of local supports exceeds the minSupport, the item set is added to the selected candidates' sub-sequences (LK). CK for K $>$ 1 is pruned by dropping the candidates if their item set was not found in the previous LK. All of the time intervals of item set x, when x occurs frequently (more than minSupport), are saved in an array [193].

"An association rule is an expression of the form $x \Rightarrow y$, where x and y are item sets and $x \cap y = \varnothing$," according to [196]. To extract meaningful and intelligible patterns from a database, association rules are built from the observed frequent item sets. The rules of association differed between the research. The support of each association rule is defined as follows using the Equation 3.2:

$$support(x \Rightarrow y)_{([T_p,T_q])} = \frac{\left|N(x,y)_{[T_p,T_q]}\right|}{\left|N_{[T_p,T_q]}\right|} \tag{3.2}$$

Where N(x,y) is the number of transactions in the time interval that contain both x and y, and the time interval $[T_p, T_q]$ shows the intersection time of item sets x and y. A sub-sequence called "consequence" is taken from each LK in order to generate the association rules. If the item set from the LK is referred to as a "frequence", then the association is defined as **AR = freq-cons** $\Rightarrow$ **cons**, and the time interval for each is determined using the Equation 3.3, based on the TP array.

$$Time(AR) = TP[freq - cons] \cap TP[cons] \tag{3.3}$$

A confidence value is assigned to each association rule. The ratio between the number of transactions that contain x and y and the number of transactions that contain x determines the confidence of an association rule x $\Rightarrow$ y. Given that x is present in a transaction, the confidence provides the conditional probability of finding y in that transaction. The confidence is estimated using the Equation 3.4 Confidence AR based on the support values by defining an AR rule.

$$ConfidenceAR_{Time(AR)} = \frac{Support(freq)}{Support(freq - cons)} \tag{3.4}$$

If the confidence in the time period exceeds the user-defined minConf, an association rule is valid. The candidate has high confidence after utilising the timestamp and identifying in local frequent sets. It indicates that the timestamps aid in the selection and validation of candidates prior to the threshold being applied.

3.2.3 Event rules and description layer

The main concept of this layer is based on our efforts, which are detailed in [2]. The association rules define the event-based rules. Each log record (here L) has an eventID that corresponds to the Entry and is used to build the item sets. To create event-based rules, item sets are replaced with the corresponding eventID. The algorithm 1 shows how to find the log record that contains all of the properties from the item sets x and y separately.

Algorithm 1 Conversion of association rules to event-based rules.

```
1:  x ⇒ y is an association-rule
2:  for x and y in association-rule do
3:      Item1=x
4:      Item2=y
5:      for Each item do                                  ▷ Search Item in log record
6:          SearchEvents(item,evenTime,time(association-rule))
7:          if item is found then
8:              if evenTime in interval time(association-rule) then
9:                  return eventID
10:             end if
11:         end if
12:     end for
13: end for
14: LH=list of eventIDs for item1          ▷ Each item can belong to more than an eventID
15: RH=list of eventIDs for item2
```

The corresponding eventID to the property that occurred during the Time (AR) is searched from the eventID table because each association rule's property refers to a periodic time (AR). As a result, EventRules are replaced by the properties of association rules. "4717, 4425 ⇒ 4797,4426", for example. Because the properties of association rules can appear in multiple records with different eventIDs, a set of eventIDs is replaced in the rule's "Left Side" and "Right Side" (left and right side of the ⇒). While the searching and retrieving of eventIDs are based on time, the appearance sequence of the eventIDs in the time order table shows the series of occurrences. As a result, by considering the series from left to right, the chain of events from each event rule is generated.

The model of events that is used has a significant impact on awareness. Finding correlations of events in log files is easier for humans when a story (chains of events) is generated from discovered events. Because creating a story necessitates the generation of annotations, the short description property is used. Although the short description property interprets the main action (not subject or object), it is useful for users who want to follow the event sequence. A chain of subsequence events is translated into a story when each eventID is mapped to its corresponding short description. While the ordered sequence is kept in the chain and transplantation is done in that order, no loops or conflicts with the same source occur.

As Figure 3.2 shows, an example event chain includes 5 eventIDs (with loop), and the story is generated based on the sequence order without conflicts. Because the order of eventIDs is the same as their appearance in the chain, only the "⇒" symbol indicates the direction between two sequential eventIDs. Each event is translated to its own description by looking up the Windows guideline, which contains the eventID and its short description property. Table 3.1 displays a snapshot of the eventID, short description and explanation gathered from the online source [6]. Analysts and forensic investigators can obtain a more comprehensive perspective of security occurrences with this story.

After detecting an event in the logging data, it is critical to annotate it with metadata in order to assist administrators and experts with prompt incident analysis. Events can be linked to background knowledge as a brief description of the event by lifting raw log data and modelling their context. To interpret the context, Windows Logs stores the annotation metadata within the message property. The corresponding description to each eventID is discovered in the Short Description table and replaced in the event chain. As a result, event chains are translated in order to interpret the meaning of events from an unstructured log.

The story from the log files is chucked to M levels, according to Equation 3.5, where M is the number of levels that are determined based on the number of association rules (N), and the

Event chain {4717,4625,4688,4797, 4625}
Event by ordered symbol: 4717 ⇒ 4625 ⇒ 4688 ⇒ 4797 ⇒ 4625
Lookup the table for finding the short description for each eventID

Generate the story:
system security access was granted to an account ⇒ an account failed to log on ⇒ A new process has been created ⇒
An attempt was made to query the existence of a blank password for an account ⇒ an account failed to log on

Story after using the Template:
From Time x to y (if each event occurred exactly 1 second after the previous one) time is equal to [AR]:

- system security access was granted to an account
- an account failed to log on
- a new process has been created,
- an attempt was made to query the existence of a blank password for an accounts
- an account failed to log on

FIGURE 3.2: Translation of the chain of events into a story

TABLE 3.1: Snapshot of mapping of eventIDs to their descriptions according to [6]

Event_ID	**Short Description**	**Explanation**
4717	System security access was granted to an account	This event documents the grant of logon rights such as "Access this computer from the network" or "Logon as a service".
4625	An account failed to log on	It documents each and every failed attempt to logon to the local computer regardless of logon type, location of the user or type of account.
4688	A new process has been created	It documents each program that is executed, who the program ran as and the process that started this process.
4797	An attempt was made to query the existence of a blank password for an account	this event at least included the process that made the request

interval time threshold that is considered for finding the local set of each item set.

$$M = \frac{N_{[associationrules]}}{threshold_{[interval]}} \tag{3.5}$$

For Equation 3.5, it is assumed that each event in the log file occurred one second after the previous event. As a result, the story of each local set is explained on a single level. The first line of each level indicates the beginning and end of the story's time period. It enables analysts to make a more informed and timely decision about an incident reported by monitoring systems by

referring to the appropriate story level. Each level's story provides more information about what happened during that time period. The sequence of the most important events, chosen based on frequent item set mining, will be shown in a human-readable format.

3.2.4 Enrichment and story layer

The main foundation of this layer is how to make a more comprehensive semantically annotated extracted description. The short description is often a single general sentence that does not refer to the events' objects and/or subjects. Semantic metadata can be used to fill in gaps in event meaning. Through the conceptual model, I represent the essential semantic metadata containing the required subject or object for an event annotation. This layer is divided into three stages, each of which is used to filter relevant semantic data:

- **The first stage** looks for appropriate background context (relevant fields of eventID) to express the heterogeneous log data and enrich the descriptions. The information provided in [6] presents necessary appropriate properties to determine which subjects or objects must be addressed based on each eventID. Randy Franklin Smith is a highly trusted subject matter expert on the Windows security log that published UltimateWindowsSecurity.com (UWS) [6]. UWS spent years reverse engineering the events in the security log and isolating the arcane patterns to filter out the noise and mine the real gold that the Windows security log has to offer.

  The knowledge provided in [6] is a key principle in the design of our event ID-based conceptual model. The proposed conceptual model divides the requirement concept into specialised properties linked to the eventID and derived from [6]. As a result of mapping the conceptual model to log files, values of the mapped properties were taken from the logs and filled the absent meaning in the short description. Generic Classes (G) are formally defined, and the corresponding properties for each of them are defined in the Requirement Fields (RF). Generic classes are used to represent properties and event categories (Subject and Object) in general. Requirement Fields are G's detailed properties that are used to present the G's associated properties:

  - **Definition 1 (Class of 'G')**: A generic properties class refers to properties that can be included in a "Short Description". In other words, the Generic properties Class, denoted by main words, abstracts a meaning from an event
  - **Definition 2 ('RF')**: A Requirement Fields are G class properties (i.e. specific properties) that refer to event behaviour by providing event requirements fields on the logs.

Figure 3.3 shows how the proposed conceptual model's design is based on the necessary appropriate properties from UWS [6] for the eventID 6663. "Subject", "Process", "Access", "Object", "Account", "Network","Logon", "Group", "Cryptography", "Key File", "Member", "Service", "Transaction", "AuditPolicy", "Handle", and "Task" are the 16 generic classes (blue boxes) that our proposed conceptual model used to retrieve the requirement knowledge for various eventIDs. Then, to complete the list of requirement properties, RF properties (purple boxes) are created. Because I intend to use only valuable properties (those whose values are extracted from logs) to convey meaningful concepts, properties with constant values have a zero weight of meaning and are, thus, ignored. In other words, RF is only built with variable values, for example. As shown in Figure 3.3, although "Object Server" by referring to the UWS [6] is introduced as a requirement field under the G class, "Object", it contains the constant value "Security", which has no meaningful concept. As a result, the Rf for "Object" has no properties relevant to the object server.

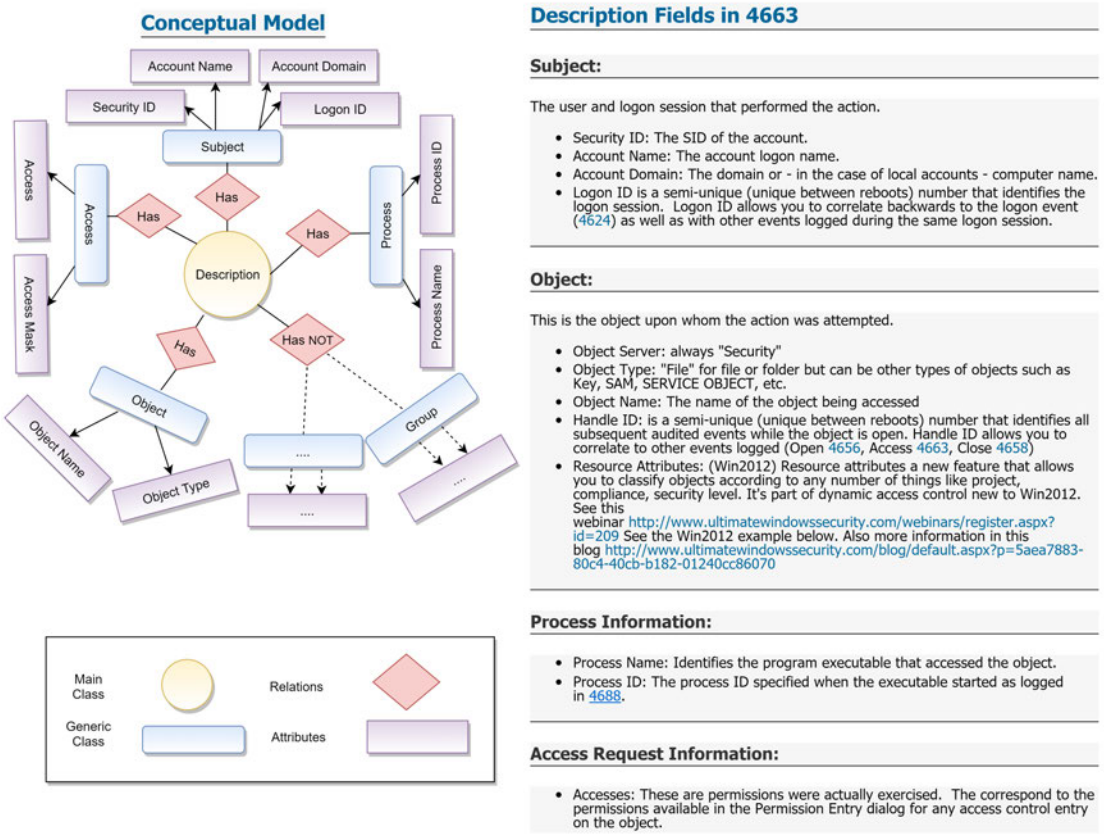


FIGURE 3.3: Conceptual model based on the Generic and RF classes by the extracted attributes from the external source [6]

- **The second stage** searches the conceptual model for matched properties that appear in the short description; then creates a map between the RFs and the corresponding properties from logs for extracting RF values. Based on the short description, corresponding properties from logs are recognised to achieve the goal of this stage. This task requires specifying a

section of a sentence as individual words. Each word mentioned in a given short description is linked to the entities in a given conceptual model (Generic classes). In the Generic classes, a short description may have one or more matching entities. For example, in the short description of eventID 4625, "An account failed to logon", two words, "account" and "logon", are linked to the Generic classes' "Account" and "logon" entities.

The second level of linking entails connecting RF properties to log properties (from the Entry Table). This task requires a mapping between the RF's entire properties under the matched Generic entity and the event's corresponding properties in the Entry table. The extracted values of RF properties can improve the description's quality

- **The third stage** replaces the values of the corresponding properties from the Entry table in the chain of descriptions with the matched entities from the Generic classes. As previously stated, each Generic class has its own sub-properties (RF), the values of which are searched through the Entry table (in the same event record) to enrich the description.

  For example, consider the short description of eventID 4663, "An attempt was made to access an object", where Object is replaced with the values of the RF properties "Object Type" and "Object Name" from the corresponding record of the Entry table.

  The detailed storytelling that is developed might be used to guide future actions in reaction to an incident or to raise awareness about what occurred.

3.3 Evaluation

In this section, I evaluate the Log-Chain-Driven Storytelling model using two empirical studies. The proposed model is compared to the model introduced by Khan and Parkinson [137], which I call "StoryPlan" in the first experiment. Although the authors did not refer to their proposal output as a human-readable story from the log file, they did refer to it as action plans, however, because the action plans are extracted from predefined statuses and actions, it is reasonable to assume that it is similar to a StoryPlan, providing a level of understanding and representing the log file knowledge. As a result, in order to make the comparison fair, I added a translation procedure (translation in Layer 3) to their model. The added procedure can translate a chain of events into descriptions, and the descriptions of the outputs can be compared based on a specific event being revealed from logs.

The LDSM's capacity to depict malware behaviours is proved in the second empirical trial. Through the proposed conceptual model (Layer 4) to describe the events among the logs, the chain of descriptions as output is supplemented to the needs of the objects and subjects.

In both steps of the experiment, I conducted an empirical analysis by launching Microsoft Windows Server 2012 R2 Base-64-bit instances on the Amazon Web Server (AWS). The security

TABLE 3.2: The details of process steps for three main types of activities which are used in scenarios in empirical analysis

	Process Steps		
Type	1	2	3
I	Install & Run Manager.setup.exe (Microsoft Toolkit)	Export 2 baselines (RemoteDesktopService MemberServiceSecurity)	Create & Run .bat for deployed baselines (with LPG*.exe)
II	Create Vb Script	Run Powershell-ISE	Run Script by Powershell
III	Unzip Malware Simulator	Run Command Prompt	Execute Simulator

I. Compliance Security Checking II. Security Script III. Malware Simulation

*LALR Parser Generator.

event logs are collected during the time when an administrator checks the Compliance Security or runs scripts. Because the output explains the events among log files, a malware simulation is run in one instance to assess the LDSM's ability to capture and describe the events. To accomplish this, a malware simulator tool [197] is run on the specific instance, resulting in Malware Simulator.txt being left in all accessible folders. The access attempt to objects from malware.exe is recorded in the logs as a result (folder auditing is enabled for all drivers and sub-folders in Windows). Both experiments demonstrate the LDSM's ability to explain malicious activities. As a result, the LDSM can assist analysts in dealing with a rapidly changing threat landscape, alleviate alert fatigue, improve SA, and speed up incident response.

Table 3.2 displays detailed information about process steps for security compliance checking, running security scripts (for checking blank passwords), and running malware simulator. Through the scenarios, each AWS instance uses a combination of these types of activities (Type I, II, and III based on Table 3.2). The following are the specifics of the experiments.

3.3.1 Empirical analysis 1

In this experiment, three AWS instances are launched with the following scenarios:

- **Scenario 1**: Log in as the admin remote user, clear the logs, create two local accounts, enable auditing policies, run script (Type II), check security compliance (Type I), and run script (Type II)
- **Scenario 2**: Log in as the admin remote user, clear the logs, verify security compliance (Type I), and run the script (Type II)
- **Scenario 3**: Log in as an admin remote user, clear the logs, run the malware simulator (Type III), enable auditing policies, and verify security compliance (Type I).

3.3.1.1 Pre-processing layer

The following are the extracted properties (both implicit and explicit) for the Entry tables: Entities={*User, Computer, Event_Source Name, Session_ID, SecurityID, AccountName, AccountDomain, LogonID, LogonType, LogonGUID, ProcessID, ProcesName, Caller_workstation, TargetAccountName, TargetAccountDomain, WorkstationName, Source-NetAddress, SourcePort*}

The corresponding eventIDs, short descriptions, as well as the date and time, are extracted and saved in the appropriate tables.

3.3.1.2 Frequent item set mining with a timestamp layer

To provide temporal patterns in local sets, apriori-like mining is performed on the properties of the Entry table. While each local set is explained in a single level of story, the minimum interval time threshold defines the period for tracing events. In this experiment, five minutes is set as the time limit. The minSupport and minConfidence thresholds are set to the same values used by Khan and Parkinson [137], 20% and 70%, respectively. The association rules are then generated and the corresponding interval time (AR) (period) is saved in the table.

3.3.1.3 Event rules and description layer

Because each association rule's property refers to a periodic time (AR), the eventID table is searched for the corresponding eventID to the property that occurred during the Time (AR). To generate event rules, eventIDs are replaced with the properties of association rules. For transforming the event rules into a chain of eventIDs based on Khan and Parkinson's proposed model [137] (each eventID from the right-side connects to all eventIDs from the left-side to produce a sequence of pairs), three major conflicts on the sequence order may occur:

- The same source conflict: two pairs of events with the same start eventID are discovered, i.e. for pair (a,b): "a" is the start eventID and "b" is the end eventID
- The same destination conflict: two pairs of events with the same end eventID are discovered
- The loop conflict: the start eventID of one pair is the end eventID of another.

Because our approach differs from [137] in ordering the sequence of chain of events, potential conflicts are ignored in the LDSM. Our eventIDs sequences are ordered sequentially by defining the local sets.

The corresponding description to each eventID is retrieved at this level, and the chain of description through the different levels is explained. A different item set with the same eventID may appear in a log record. As a result, the same chains can be generated from various item sets. Duplicated chains are removed from each layer to reduce the ambiguity of the translation.

The experimental results from the empirical analysis on the LDSM with comparison to the StoryPlan by Khan and Parkinson [137] are shown in Table 3.3. As Table 3.3 shows, for all scenarios, the number of association rules in the LDSM is greater than that generated by the StoryPlan. When the average confidence for the association rules is greater than the average confidence in StoryPlan, our proposed algorithm was able to discover more temporal association items from the log files with a higher reliable ratio. Since duplicated chains vary from level to level in the story, each level contains a different number of sentences. Table 3.3 shows the total number of unique chains as well as the number of generated sentences.

3.3.1.4 Enrichment and story layer

This level is not used in this experiment, as published in [2].

TABLE 3.3: Empirical results 1

		Storytelling Model			StoryPlan Model		
		Scenario-1	Scenario-2	Scenario-3	Scenario-1	Scenario-2	Scenario-3
Properties	1. *Num of Logs*	100	1276	4170	100	1276	4170
	2. *Num of association rules*	2509	24920	14982	54	1368	85
	3. *Average of confidence*	0.9661	0.9663	0.9684	0.9226	0.9573	0.9682
	4. *Num of unique chains*	48	1470	191	6	17	7
	5. *Num of conflicts*	0	0	0	5	5	5
	6. *Num of sentences*	48 in 8 levels	1470 in 143 levels	191 in 50 levels	1	12	2

3.3.2 Empirical analysis 2

In this experiment, three AWS instances awere launched with the following scenarios:

- **Scenario 1**: Log in as the admin remote user, clear the logs, create two local accounts, enable auditing policies, run script (Type II), check security compliance (Type I), and run script (Type II)
- **Scenario 2**: Log in as an admin remote user, clear the logs, verify security compliance (Type I), and run the script (Type II)

- **Scenario 3**: Log in as an admin remote user, clear the logs, run the malware simulator (Type III), enable auditing policies, and verify security compliance (Type I)

3.3.2.1 Pre-processing layer

The following are the extracted properties (both implicit and explicit) for the Entry tables: Entities={*User, Computer, Event Source _Name, SessionID, SecurityID, AccountName, DomainName, LogonID, LogonType, ObjectType, ObjectName, ProcessID, ProcesName, WorkstationName, SourceNetAddress, SourceAddress, GroupName, GroupDomain, ProviderName, AlgorithmName, KeyName, KeyType, FilePath, Operation, ReturnCode, Servername, AccessRight, TransactionID,NewState, ResourceManager, Category, SubCategory, SubCategoryGUID, changes, SoueceHandleID, SourceProcessID, TaskName*}

The corresponding eventIDs, short descriptions, as well as the date and time, are extracted and saved in the appropriate tables.

3.3.2.2 Frequent item set mining with a timestamp layer

To provide temporal patterns in local sets, apriori-like mining is performed on the properties of the Entry table. While each local set is explained in a single level of story, the minimum interval time threshold defines the period for tracing events. In this experiment, the threshold is set at five minutes. The minSupport and minConfidence thresholds are set at 20% and 70%, respectively. The association rules are then generated, and the corresponding interval time (AR) (period) is saved in the table.

3.3.2.3 Event rules and description layer

The eventID corresponding to the property that occurred during the Time (AR) is searched in the eventID table because each association rule's property refers to a periodic time (AR). To create event rules, eventIDs are replaced with the properties of association rules to convert event rules into a series of eventIDs. To put it another way, each eventID on the right-side connects to the next eventID on the right-side, and if that isn't feasible, connects to the eventIDs on the left-side to form a chain. Our eventID sequences are ordered based on time, with the appearance eventID on the right most side occurring sooner than the one on the left.

This layer retrieves the corresponding description for each eventID and explains the chain of descriptions through the various levels. In a log record with the same eventID, different item sets may appear. As a result, different item sets can produce the same chains. Duplicated chains are

removed to reduce the ambiguity of each layer's translation. An example of chain of description from the Scenario 3 is as the follows:

[(datetime.datetime(2019, 8, 1, 7, 36), datetime.datetime(2019, 8, 1, 7, 36)), (datetime.datetime(2019, 8, 1, 7, 41), datetime.datetime(2019, 8, 1, 7, 41)), (datetime.datetime(2019, 8, 1, 7, 46), datetime.datetime(2019, 8, 1, 7, 46))] [['A trusted logon process has been registered with the Local Security Authority.'], ['A handle to an object was requested.'], ['The handle to an object was closed.'], ['An object was deleted.'], ['An attempt was made to access an object.'], ['A privileged service was called.'], ['A new process has been created.'], ['A process has exited.'], ['An attempt was made to duplicate a handle to an object.'], ['A user account was created.'], ...

3.3.2.4 Enrichment and story layer

The study's main contribution is that it expands the context to be more comprehensive. In other words, "the handle to an object that was closed" in the example in Section 3.3.2.3 was enriched with corresponding properties to answer "what was the object type and object name that was closed?", as well as "what was the source handle ID and source process ID?" to clarify the ambiguous description.

As mentioned in Section 3.2.4, to accomplish this goal, a conceptual model with 14 Generic classes and corresponding RF properties is designed. Because formal evaluation of the narrative format of short descriptions is qualitative in nature, enrichment of sentences proves to be a difficult task. In this study, I concentrated on the conceptual model to demonstrate its utility in improving the description. Thus, the **Accuracy** as an evaluation criteria was defined. In this case, accuracy refers to the quality or state of being correct or precise in accordance with the conceptual model. The number of short descriptions that contain at least one matched word to the conceptual model (G classes) is compared to the total number of short descriptions to establish accuracy.

The repeated short descriptions were removed to be more precise. The proposed conceptual model was used to create and test a list of unique short descriptions extracted from the scenarios. The accuracy for three scenarios based on the proposed conceptual model are shown in Table 3.4. The accuracy of the three scenarios is over 80%, as shown in Table 3.4. The scenarios include a number of descriptions, but none of the words correspond to the G classes, which are 7, 2 and 6, respectively. The non-matched descriptions' corresponding eventIDs are *5033-1102-4902-4608-4739-4616-4647*, *1102-4616*, and *4957-4647-5158-5447-5156-4616*. For example, the short description for eventID *4616* is *"The system time was changed"*, there is no word for matching to the G classes in this. As a result, the accuracy demonstrates the conceptual model's potential for enriching the description. Although it fails to describe the descriptions for less than 20% of the time, the descriptions that fail are the general settings without process, object, and so on.

Each scenario is enriched. For instance, "The handle to an object was closed" is replaced with the sentence "A handle **Source Handle ID: 0x19fc**, **Source Process ID: 0xa00** to an object **Object Type: file**, **Object Name: C:/ Desktops/Malware.txt** was requested". The potential words that

Analysis Parameters	*Scenario 1*	*Scenario 2*	*Scenario 3*
Number of unique short descriptions	35	12	35
Number of total words	228	78	283
Number of matched words	33	12	48
Number of non-matched descriptions	7	2	6
Number of matched Generic classes	10	7	11
Number of matched RF properties	97	33	95
Accuracy	**0.8**	**0.833**	**0.828**

TABLE 3.4: Accuracy based on the statics analysis in three scenarios

helps the meaning is clarified by replacement that refers to the corresponding properties. This is helpful for tracing an event or a process. Table 3.5 contains detailed information from the empirical analysis.

3.4 Discussion

In this section, I will discuss how the model can be useful in interpreting logs into a chain of descriptions (story) that is comprehensive for humans by employing association rules and making a chain of events.

The generated story must be applied to a sequence of events that occurred multiple times in the same sequence because, in most cases, analysts check on repeated logs as malicious failures or errors. The generated story cannot be used in security breaches with a limited number or variety of attempts. However, mining associated items relevant to an occurrence enhance security analysts in better comprehending log records. The log sequences identified in a chain of events identify anomalous event sequences.

In the present logs, information from many sources is recorded, which may or may not be relevant to an event. By mining the many properties presented in a log record, the proposed model discovers the relevant logs and presents the chain of records which is statistically correct. The logs lacked detailed information on objects, services and processes that would have confirmed what was occurring. Although the log is often a resource for identifying more information about what occurred on the machines or servers after an event happens, the machine-friendly format with insufficient context makes it difficult for analysts to comprehend which log recorded the specific application or process. Knowing the time that the event or incident occurred can assist analysts in filtering logs and investigating them. However, numerous subjects and objects are still missing from the event logs.

A timespan is described as the window in the proposed model that is used in security log mining to discover abnormal event sequences that occur and repeat inside the stated timespan.

The model gets the necessary resources and creates a story with the chained events to assist security analysts and management in better understanding the abnormality that was reported based on the various time windows. In anomaly investigations, time is a critical factor; logs are filtered based on it.

Various events in the logs are detected by locating the relationship between the logs. The proposed model supplemented the sentences by referring to what objects and subjects were in the logs in order to generate much more readable event logs. Presenting context can assist analysts infer what happened and make it easier to trace each event in the chain of events. The chain of descriptions developed in a storytelling format was demonstrated to be more human-readable, facilitated comprehension, and provided a better awareness of the potential anomaly.

A comparison of one level of the LDSM-generated story and the logs from Scenario 3 is shown in Figure 3.4. As shown in Figure 3.4, the generated story describes malware activity in a human-readable and understandable format by finding correlations between events and enriching the context with requirement fields (instead of whole properties, which is provided in the Windows logs), which was difficult to achieve with the non-rich logs.

From 2019-08-01 07:31:00 to 2019-08-01 07:31:00 Story is: [['A handle,Source Handle ID:0$x19fc$, Source Process ID: 0$xa00$ to an object, object type: file, Object name: C:/Desktops/Malware.txt was requested.']] ['The handle, Source Handle ID:0$x19fc$, Source Process ID: 0$xa00$ to an object, object type: file, Object name: C:/Desktops/Malware.txt was closed.'], ['An object, object type: file, Object name: C:/Desktops/Malware.tmp was deleted.'], ['An attempt was made to access, Access Right: SeNetworkLogonRight an object, a folder type, C:/Desktops/Test.'], ['An operation was attempted on a privileged object, a folder type, C:/Desktops/test.'], ['A new process, process ID: 0xfa8 and process name: C:/Desktop/malwaresimulator-master/MalwareSimulator.exe has been created.'], ['A process has process, process ID 0xfa8 and process name C: /Desktop/malwaresimulator-master /MalwareSimulator.exe exited.'],

(A) *One level of the proposed story .*

Audit Success 01/08/2019 7:31:14 AM Microsoft-Windows-Security-Auditing 4656 Removable Storage "A handle to an object was requested.
Subject: Security ID: WIN-1RTKVJHP1CH/Administrator Account Name: Administrator Account Domain: WIN-1RTKVJHP1CH Logon ID: 0$x3CB85$
Object: Object Server: Security Object Type: File Object Name: C:/Desktops/Malware.txt Simulator Handle ID: 0$x19fc$ Resource Attributes: -
Process Information: Process ID: 0$xa00$ Process Name: C:/Desktop/malwaresimulator-master/MalwareSimulator.exe
Access Request Information: Transaction ID: { 00000000-0000-0000-0000-000000000000 } Accesses: READ_CONTROL ReadAttributes
Access Reasons: READ_CONTROL: Granted by Ownership ReadAttributes: Granted by $D : (A;OICI;FA;;;BA)$
Access Mask: 0x20080 Privileges Used for Access Check: - Restricted SID Count: 0"
Audit Success 01/08/2019 7:31:14 AM Microsoft-Windows-Security-Auditing 4658 Removable Storage "The handle to an object was closed.
Subject : Security ID: WIN-1RTKVJHP1CH/Administrator Account Name: Administrator Account Domain: WIN-1RTKVJHP1CH Logon ID: 0$x3CB85$
Object: Object Server: Security Handle ID: 0$x19fc$ Process Information: Process ID: 0$xa00$ Process Name: C:/Desktop/malwaresimulator-master/MalwareSimulator.exe
Audit Success Success01/08/2019 7:31:15 AM Microsoft-Windows-Security-Auditing 4690 Handle Manipulation "An attempt was made to duplicate a handle to an object.

(B) *The log records in security Windows log file*

FIGURE 3.4: The generated story by the Log-Chain-Driven Storytelling model VS the Windows logs to describe a malware activity

3.5 Summary

In this chapter, the LDSM has been proposed to automatically extract the cybersecurity events. The model's output reduces a large number of log records into more understandable sub- sets with sequence chains and time periods of occurrence. To supplement the message from the proposed conceptual model, the events are translated into their own descriptions, along with the corresponding subjects and objects. The ability to enrich the log description has been validated in terms of accuracy in finding the requirement properties and enrichment rate. The proposed

model has achieved three main principles: (1) automatically creating chains of subsequent events without pre-defined status, (2) generating a more accurate and easily traceable enriched story through time, and (3) identifying important short and long life span events. This chapter results was published in the paper 2 of publications.

TABLE 3.5: The detailed information of Empirical analysis 2

Classes		Scenario1					Scenario2					Scenario3				
Generic	RF	Reg G*	Orange G**	Total G	Total Matheed RF	EventIDs	Reg G	Orange G	Total G	Total Matheed RF	EventIDs	Reg G	Orange G	Total G	Total Matheed RF	EventIDs
Subject	Security ID Account Name Domain Name Logon Type Logon ID	0	0	0	0		0	0	0	0		0	0	0	0	
Account	Security ID Account Name Domain Name	8	5	13	39	4738-4724-4624-4634 4781-4726-4722-4720 4625-4723-4717-4797- 4776	3	1	4	12	4723-4624 4717 4718	5	2	7	21	4720-4624-4797 4722-4625-4776 4724-4634-4738
Process	Process ID Process Name	1	0	1	2	4688	1	0	1	2	4688	1	1	2	4	4616-4688
Network	Workstation Name Source Net Address Source Address	0	0	0	0		0	0	0	0		0	0	0	0	
Logon	Logon Type Security ID Account Name Domain Name	2	2	4	16	4672 4624 4648 4625	1	1	2	8	4624-4672	2	2	4	16	4611 4624 4625 4672
Object	Object type Object Name	1	0	1	1	4907	1	0	1	1	4674	7	1	8	16	4690-4674-4656-4907 4658-4660-4663-4670
Group	Security ID Group Name Group Domain	7	0	7	21	4737-4733-4731 4735-4732 4728-4729	0	0	0	0		4	1	5	15	4728-2729 4732-4733 4825
Crypt	Provider Name Algorithm Name Key Name Key Type	1	0	1	4	5061	0	0	0	0		0	0	0	0	
Key File	File Path Operation Return Code	1	0	1	3	5058	0	0	0	0		0	0	0	0	
Member	Security ID Account Name	4	0	4	8	4728-4729 4733-4732	0	0	0	0		4	1	5	10	4728-2729-4825 4732-4733
Access	Access Right	1	0	1	1	4717	1	1	2	2	4718-4717	0	2	2	2	4825-4663
Service	Server Name	0	2	2	2	1100-5024	1	0	1	1	4673	1	0	1	1	4673
Transaction	Transaction ID New State Resource Manager	0	0	0	0		1	0	1	3	4985	1	0	1	3	4985
AuditPolicy	Category Subcategory Subcategory GUID Changes	0	0	0	0		1	0	1	4	4719	0	0	0	0	
Handle	Source Handle ID Source Process ID	0	0	0	0		0	0	0	0		1	2	3	6	4690-4656 4685
Task	Task Name	0	0	0	0		0	0	0	0		1	0	1	1	4702

* The number of matched words from short descriptions to the G entities where G is introduced as the requirement by [6]
** The number of matched words from short descriptions to the G entities where G is **NOT** introduced as the requirement by [6]

Chapter 4

Explainable Intelligence to Interpret Cyber Alerts - Self Situation Awareness

An average medium-sized organisation logs approximately 10 to 500 MLN events on the system each day. A much smaller number of threat alerts are investigated by specialised personnel, leaving a security gap open for future attacks. Inadequate information in alert messages written in a machine-friendly ormat creates cognitive overload on currently limited cybersecurity experts. The Alert-Driven Storytelling Model (ADSM) as the designed explainable intelligence is proposed in this chapter. It produces a report in natural language using a novel storytelling framework derived from cyber alerts. The solution caters to various reader experiences and preference levels by delivering adjustable models filled from both the local and global knowledge base.

To provide empirical evidence for the effectiveness of the ADSM in establishing self-awareness, a case study from an educational institution's SOC and a survey study were used to validate the model. Two structured questionnaire surveys were conducted to investigate the completeness and comprehension of the two output incident reports, namely the Secureworks report (baseline) and the Storytelling report (ADSM). The reports referred to the same events recorded at the SOC. The goal of the surveys was to quantify the ability of the proposed model (which used local and global knowledge) in a narrative framework to reduce the cognitive effort of security analysts.

Although both types of report provide information regarding incident occurrence, the specifically designed questions were used to elicit particular aspects affecting the analysis process from a report completeness point of view. In other words, how comprehensive is the report in terms of the incident problem clarification and overall cognitive load reduction? The success of the ADSM in its effectiveness goal was empirically demonstrated by the comparison of ratings obtained from the cybersecurity experts and students on both types of the reports. The designed ADSM model outperformed the existing baseline by 32.6% (R1), 28.0% (R2) and 23.5% (R3) on the comprehension scale (on average).

The evaluation results demonstrate the importance of explainable intelligence in interpreting cyber alerts, as well as its ability to reduce the expert's cognitive efforts in order to achieve the highest level of self-awareness.

4.1 Introduction

Different perspectives of cyber security can be displayed by human agents with varying levels of knowledge and expertise in the field. Higher security skills usually suggest more capacity as experience can affect decision-making capabilities [198]. Advanced knowledge and prior experience may improve security threat sensitivity and incident response ability. According to the researchers in [199], domain knowledge in information and network security and situated environmental knowledge established in an analyst's particular environment, are necessary to improve an analyst's ability to identify threats. Continuous interactions with a specific operational environment are used to develop situated environmental knowledge [200].

Cyber security knowledge and practical security response capabilities are two primary aspects of workforce security capacity; both of which are critical in establishing a secure operating response environment. The level of cyber security knowledge can also vary dramatically depending on prior knowledge or the results of a comprehensive report. Successful input (i.e., comprehensive report) of security knowledge and response capabilities, on the other hand, should be based on a thorough understanding of current capacity measures, as well as the identification of capability gaps that expose user vulnerabilities (weaknesses) to cyber attackers.

Analysts begin their threat hunting investigation with an alert message. According to the survey [67] performed by the SANS Institute[1] involving the observation of various organisations over a two-year period, cybersecurity analysts mostly spend an average of 24 hours or less on detection after a compromising incident has occurred. Approximately 40% of analysts require more than 24 hours, and in some cases, 4-6 months to detect the initial compromise. Alert correlation is carried out with a mix of machine algorithms and human investigation [201]. Given the complexity of modern systems and cyber attacks, algorithms have not successfully applied sufficient context to the message or provided enough intelligence to understand why certain alerts are important. Furthermore, human beings have to be involved in the analysis process, human-as-a-security-sensor into security analysis [202].

The main reasons for an organisation's effectiveness in responding to incidents include shortage of staff and skills, a lack of integration with other security and monitoring tools, and a lack of visibility into insider behaviours. In summary, the lack of a comprehensive and complete incident

[1]The SANS Institute is a private U.S. for-profit company founded in 1989 that specializes in information security, cybersecurity training, and selling certificates.

report to combine inside and outside visibility is the main weakness in detecting and responding to security incidents. The existing alert analysis procedure has a major problem which is the message's verbosity without annotation, making it difficult to deal with the massive volume of events [40]. To properly assess the scale of the risk and gain a solid understanding of cyber situations, knowledge outside security logs must be added to the report generated by current solutions. Local domain knowledge determines the risk of internal assets, and the potential risk of the outsider is specified by global domain knowledge. As an illustration, consider the examples below:

- **Local domain knowledge required:** A server of the organisation X is used for temporary storage and web testing, and is labelled a *non-critical* host. Most of the alerts from that server can be omitted unless a serious breach occurs. However, the server is located in the Finance Department for financial reporting and budget planning. Finance departments usually holds critical information. If an alert for a serious breach occurs for one of the servers in this department, other servers also can be at potential cyber risk, warranting further investigation despite no explicit alert raised. Thus, the exceptional defense strategy should be adopted in advance following the complete knowledge obtained from inside the organisation
- **Global domain knowledge required:** Organisation Y with limited number of experienced cyber professionals has to prioritise the crucial alerts (over a large volume of the remaining security breaches) for prompt response. The selection is based on prior knowledge and experience which, in turn, is based on repeated alerts from historical records. An appropriate response for the new attack requires an in-depth investigation of attacker's characteristics. However, the attacker may change its behaviour over the time. The level of expert knowledge does not usually increase at the same speed as the complexity of attacks in today's digital environment. As a result, a critical alert may not be given the required priority, leading to delayed response and potential escalation. Thus, knowledge obtained automatically from external sources is required to stay up-to-date with increasingly sophisticated and dynamically changing cyber attacks.

Both examples show that comprehensive alert analysis requires domain knowledge from the local and the global. If the complete knowledge cannot be modelled and integrated in alert analysis, either false alarms are triggered or high-risk alerts are ignored. In broad terms, effective input (comprehensive report) to cyber professionals can be interpreted as comprehensive and integrated up-to-date information that is linked to locally and globally available information [19].

Cybersecurity analysts must have a thorough awareness of cyber situations in order to guarantee that cybersecurity is a top priority throughout the organisation. Despite high degrees of uncertainty and extremely dynamic situations, a human-readable comprehensive report with complete

contextual information can be an effective input assisting analysts in making sense of the current situation and making decisions. Because of necessary the short response times, most analysts rely on the information supplied in the report to undertake further enquiry if necessary.

In this chapter, the ADSM that generates the incident report to interpret cyber alerts in natural language by means of applying novel storytelling framework is proposed. The main motivation behind the ADSM's incident reports is to provide both local and global information about a cyber incident to help cybersecurity analysts make better decisions. In terms of comprehension (improved cognition) and completeness, the report created outperforms the current technique (enriched context). The evaluation demonstrates the power of storytelling in the interpretation of potential threats in a cybersecurity context, where supplementary information presented in a human-readable format increases the level of security SA, saving analysts time in the incident management process.

I developed and conducted the survey instrument to evaluate CTA by comparing incident reports and demonstrating how a thorough incident report reduces necessary cognitive effort and, as a result, supports better comprehension by analysts. The proposed surveys address the aforementioned research gap by looking into how the ADSM could help the SOC team gain comprehension awareness. I created two questionnaires based on the 5W1H (who, what, why, when, where and how) method to evaluate the incident reports generated by Secureworks and the ADSM. The 5W1H method was also used by De Melo e Silva, et al. [17] as a fundamental methodology for evaluating the level of completeness of standards and platforms in cyber threat intelligence, which appears to be applicable to incident report completeness evaluation.

Given that 5W1H is based on cognitive theories, it is possible to assess how the proposal model contributes to reduced cognitive effort when combined with the ADSM as an applied model [14]. As a result, cybersecurity experts and students with various levels of security knowledge were asked to rank the output reports based on their comprehension by answering questions. In this chapter, I report on our experiences by conducting a between-subject experiment. That is, a group received and evaluated the Storytelling reports and a group received and evaluated the Secureworks report.

4.2 Alert-Driven Storytelling Model (ADSM)

The proposed ADSM consists of four individual layers and main procedures, and is illustrated in Figure 4.1. The design of the ADSM includes a set of layers covering various aspects of the alert correlation process. The results are demonstrated in narrative incident reports, enriched by context to be used by cybersecurity analysts to weed out less relevant alerts and to better understand the progress of the attack. Security analysts explore the details of alerts and various

reports, filtering the data of interest for further in-depth analysis and correlating relevant data [45]. The ADSM uses two knowledge bases, local and global, to interpret the incident alerts. Context and vulnerability information about internal hosts, online scan engines, online public threat exchange repositories, story templates and the Snort [203] community rule set in the knowledge bases are mapped with alert records to report an incident. The details of each layer, i.e. primary purpose and associated steps, are as follows:

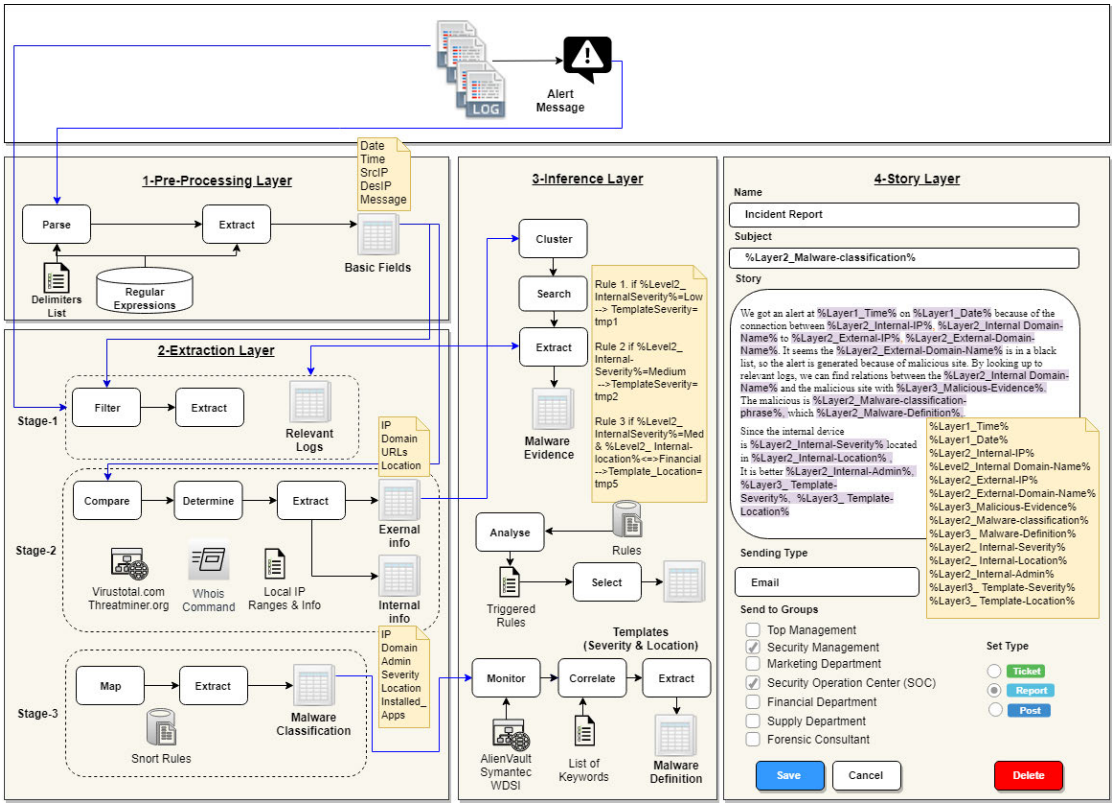


FIGURE 4.1: Overview of the Alert-Driven Storytelling Model made of four layers (beige boxes) and operation procedures (white boxes, except the story layer). The story layer represents the final output with modification capability

4.2.1 Pre-processing layer

In this layer, the alert record is parsed to extract the basic fields of an alert. The fields include Time, Date, Source Internet Protocol (SrcIP), Destination Internet Protocol (DesIP), as they are relevant to the alert. An alert record L is defined by a tuple with 5 attributes, {Date, Time, SrcIP, DesIP, Message}.

The alert generated by the SIEM system[2] was used in the case study.

[2]The approach is not limited to SIEM systems, and can be easily adopted to alerts from other monitoring systems

Since the selected fields are primary properties in each alert message, the proposed approach does not depend on the specific device. The following is information about the five attributes of alert record:

- **Date and Time** values represent when the events are registered. These values can be different from alert Date and Time (as received after an event)
- **Source Internet Protocol (SrcIP)** value represents the address of the initiator of an event. In other words, who is the source of the connection (Subject or Object of an event)
- **Destination Internet Protocol (DesIP)** represents the objects of the events. In other words, DesIP is an address to which the connection has been made (Subject or Object of the event).
- **Message** value represents behaviors, which the Subject conducts towards the Object This value usually includes the classification group name for threat. Since this study focused on the malware category, the value contains terms such as 'malware' or 'trojan'.

A collection of regular expressions is used to parse and tokenise the alert messages. The delimiters include '/', '?', '.', '=', '-', and '_'. The extraction parsers and tools before this layer are applied as pre-processing. The outputs produced will be further used in the Extraction layer.

4.2.2 Extraction layer

Although selection and retrieval of basic fields from an alert message is performed, the basic information about the alert, the relationships between basic fields and corresponding information allows the analyst to spot the potential logical links.

In this layer, the alert message is complemented with supplementary information to compensate for the lack of data which leads to insufficient understanding [54]. As a result, full awareness about the alert situation from various heterogeneous sources, such as different departments and owners, can be achieved [204]. The relevant information with potential logical links to the alert record is filtered and extracted from the knowledge bases and stored temporarily. The extraction layer consists of three main stages, which is relevant information and its relationship to the five attributes of alert record:

The first stage looks into the aggregated logs files that use Date and Time when the events were synchronised. Every single log record in log file has the Date and Time references. Events are sorted based on the time sequence. Date and Time of an event comes from the basic fields in L and log files, which are gathered log records from a variety of network devices. Binary search

in terms of time is applied to retrieve events in a particular time interval. Since some logs are recorded based on the UTC, and others are recorded based on the local time, to cover all the related logs $+/-1$ day time span is applied. The log file also provides information about source and destination IP for the connection. Therefore, the corresponding connection between SrcIP and DesIP are found by tracing the entire interval. The output of this stage is a list of events that represent the connection between the source and destination that happened in the particular time interval

The second stage searches the local and global knowledge bases to identify the IP address and Domain Information (which IP belongs to the organisation and which is from the outside, thus suspected to be a source of infection). In this stage, the Whois command identifies the names within a given registrar's registry. Therefore, the other registry outside the organisation is used as external. Furthermore, each organisation provides a list of IP address ranges based on their own network architecture. The IP address matched to this list is considered to be internal. After determining the connection type, from internal to external or from external to internal, the corresponding information from the local and global knowledge bases is extracted. The global knowledge base contains a set of information based on the online public repositories such as *"Virus Total"* [205] and *"Threat Miner"* [206]. Each set represents the IP address, which is recorded in a black list, domain names, geography location of the server and URLs [3] that were repeated in previous infections (cause to be reported in a block list). The local knowledge base contains set of relevant information about internal hosts/servers (IP address, domain name, administrator, location, severity and installed application). Although updating the local and global knowledge base is computationally expensive, it is a trade-off between automatic and complete information extraction, and the time and effort required for manual search

The third stage uses an alert message to map to the Snort rules [203] to extract the complete malware classification phrase. Snort is a lightweight network intrusion detection system that uses rules to perform content pattern matching and detect a variety of malware. Snort rules are opensource and used in variety of security devices. By mapping the message field from alert record (L) to the Snort malware rules, the complete phrase for the infection is extracted. While Snort and Snort Rules are usually thought of as a list of independent - opensource patterns to be tested in matching engines of security devices, the alert message usually contains a Snort classification label, which defines the malware category [207]. The ADSM approach is limited to security devices that lie at the core of Snort as a matching engine. Since Snort is a popular intrusion detection system, this is not a severe limitation and a variety of commercial and

[3] Uniform Resource Locator (URL) forms a part of the Uniform Resource Identifier (URI), and serves as a pointer to where the resources are located and the procedure to fetch them.

opensource devices work with the Snort rules. Snort uses pre-identified attack signatures to conduct real-time traffic analysis, content searching and content matching to detect attacks.

4.2.3 Inference layer

In this layer, information is analysed by using artefact metadata and machine learning techniques to reconstruct the past events and answer the core questions that highlight the (who, what, why, when, where and how) features of an incident. Some questions were answered in the Extraction layer (when, where, who (victim and thread), but there is still insufficient detail to explain different aspects of the incident, riskiness (what), evidence (how), and type and mechanism (what)). To understand who is the actor and what is the purpose of the action, the relevant information from the malicious website is extracted by the Extraction layer. Connections from an external computer to a single port on a malicious network machine are recorded in an online repository. They provide more information about the malicious site that can be useful in answering what, and how questions. In the Inferring layer, online scan engines and public threat exchange repositories are used to analyse temporarily stored information and their relationships in order to explain the incident type and mechanism, as well as the evidence that supports occurrence and its potential impact.

Type and Mechanism (What) To obtain better insight into the incident, I searched several web articles for the incident type based on the sentence which is extracted from Snort [203] (matched to the alert message). The malware definition extracted from web articles explains the mechanism and how the malware typically behaves. To do this, a web scraper monitors pre-selected websites and the results are shown in the Document Object Model (DOM) tree. I borrowed this idea from [208] and used a scraper to monitor each website in the list of top security technical blogs to extract the associated supplementary information. It should be noted that, although the list of websites is limited, the approach is not restricted to scraping and the list can be customised. Examples of websites used in the case study are:

- *AlienVault* [4]
- *Symantec* [5]
- *Windows Defender Security Intelligence (WDSI)* [6]

The scrapers perform breadth-first crawling on each website to search for the *malware classification phrase* found in the extraction layer. DOM trees are generated for pages that are characterised

[4] https://otx.alienvault.com/pulse/
[5] https://www.symantec.com/security-center/a-z/
[6] https://www.microsoft.com/en-us/wdsi/threats

by the same HTML template. These pages contain relevant definitions as opposed to the ones with example logins, subscriptions and advertisements considered to be non-relevant. All pages' DOM trees are compared to identify the node with a combination of the tokenised phrase from the *malware classification phrase* + 'is' + text under the node with the title 'summary', 'definition' or 'behaviour', starting with 'this malware', 'this virus', or 'this trojan'. By traversing the tree, a node with the incident explanation is identified, and using the Natural Language Toolkit (NLTK[7]), the incident definition phrase is extracted. This is away of providing further details about the malware and clarify the aim of an action.

Evidence (How) To provide evidence of the events which are relevant to an alert, the extracted information, namely the recorded malicious URLs, Downloaded files and Communication files belonging to the external host, is searched among the filtered logs found at a particular time. Using K-means clustering on the extracted URLs and the Ngrams function to iterate over N's values, the pattern of the URLs is searched. Input URLs are divided into disjoint subsets then, for each URL in each subset, the distance to all the other URLs in the same subset is computed and the URL that has the lowest sum of distances should be the centrist. To extract the max-length URL from each subset, the NLTK library, which offers an Ngrams function to iterate over values of N, is used. Then, the max-length URL from each subset, which presents the pattern of the URL, is searched among the relevant logs to extract the evidence. Repeated URLs are removed and the URL, as a symptom, is selected to enrich the report.

Riskiness (What) To obtain more information about the riskiness of an event, the information from the potentially compromised internal server is applied to the list of rules to derive proofs. The proof is a sequence of the conclusions that demonstrates the risk of an event based on the internal information. A set of rules is used to infer valid conclusion, which defines the risk. The risk is based on the internal assets' values in terms of location and severity. For example, a server in a Financial Department faces higher risk than other departments.

4.2.4 Story layer

Story generation from analytically enriched data is the main contribution of this study. It is much easier for human beings to find the correlations between events in the log files if they are modelled using a storytelling framework. A story can incorporate different aspects of an event and can convey the *meaning* of an alert. Therefore, both competence and comprehension are achieved by explaining the security alert in the storytelling design.

[7]NLTK is a free, open source, community-driven project. https://www.nltk.org/

The story can be personalised based on the needs and preferences of the individual reader [209]. As Figure 4.1 shows, the intended audience can be selected in the 'Send to Group' section of the interface and the appropriate template based on their preference is shown in the Story section. I divide the explainability space in the sense of the security domain into explanations of relevant information/data itself. This space addresses static versus. interactive variations in explanations seen by the user in response, as well as local versus. global explanations. Each template contains set of variables (the yellow border) that is initialised through the previous layers.

In this layer, the correlated information and inference results are used, based on the pre-defined rules, to enrich the narrative template report. Each template contains a set of variables that are initialised through the previous layers. The retrieved information and analytical results, which are automatically stored in the local knowledge base, are used to replace the variables in the story. Each variable contains its own original layer. For example, Date and Time are the variables extracted from the alert message in the Pre-processing Layer. The riskiness of the event is explained in the separate templates based on the triggered rules, and are used to enrich the message with more internal recommendation. The results are the knowledge sets and the relationships between them. In other words, the story is generated based on the template, and the relationships between retrieval information from previous layers. The template is modifiable and can be customised based on the preference and internal policy.

The generated story can be set as the 'Ticket' for future actions as a response to an incident, 'Report' for management, and 'Post' for broadcasting to increase awareness about what has happened. Although storytelling design is template-based, the templates and rules are easily modifiable without the need of extensive technical expertise. Customisation can be achieved based on organisational demands.

4.3 Evaluation

4.3.1 Empirical analysis

In order to validate the model proposed, the case study on a real-world scenario was conducted. More specifically, a report generated by ADSM was compared with a report generated by external vendor's tool, Secureworks[8].

Secureworks is the commercial cybersecurity analytical tool used by the SOC team at the education institution. More specifically, Secureworks provides Incident Response Services for potential cyber threats' detection among the monitored log files, and alerts their clients by appropriate report generation. The vendor claims to combine human-machine analytical

[8] https://www.secureworks.com/

capability to assist in information security services. According to Secureworks, *"to ensure that even if our machine learning models occasionally encounter an issue, Human and Machine are Working Together"* [210]. Thus, the report generation still relies on human assistance to derive actionable cyber threat intelligence.

As for technical details, the machine side of Secureworks manages the logs from approximately 800 servers at the education institution, $2000-6000$ MPS[9] (low - holiday period, high - semester period), and $600-700$ *high-risk* incidents per year. The human side involves manual assistance and human-readable report format generation about the incident registered (for the customer to understand their cybersecurity situation).

The example of the alert message produced by Secureworks is as follows:

```
MALWARE-CNC Osx.Keylogger-Elite - 10.233.62.247 -> 104.239.223.14
02/27/2019 5:05 PM
```

4.3.1.1 Pre-processing layer

The basic fields (i.e. $\{$*(Date, Time, SrcIP, DesIP, Message)*$\}$) were extracted from the alert using regular expressions presented in Table 4.1.

TABLE 4.1: Regular expressions used in the case study

Type	RegEx
For message	*([A-Z]—[a-z])\w**
For IPs	*(?:[0-9]1,3}3[0-9]1,3*
For Date and Time	*(?:0?[1-9]:[0-5]—1(?=[012])\d:[0-5])\d(?:[ap]m)?)*

4.3.1.2 Extraction layer

The information relevant to the basic fields were retrieved in the following stages:

The first stage: The relevant logs were identified based on Date and Time as well as source-destination connection. In order to ensure the coverage of the maximum number of potentially relevant events, the timespan was set to one day before and 1 day after an event. Since the Date and Time of the incident (based on the extracted basic fields) was 02/27/2019 5:05 PM, the timespan was set to the following: 02/26/2019 5:0 PM - 02/28/2019 5:0 PM (to allow all the devices to record their logs). In total, $644,434,681$ logs were recorded by monitoring devices at the university throughout the time interval specified. After filtering based on both SrcIP and

[9]Message Per Second

DesIP, the number of events was reduced to 12. This provides the final list of events that represent the connections that occurred between SrcIP and DesIP were within the timespan specified

The second stage: The SrcIP was marked as Internal (by comparing it with organisation IP addresses range), and the DesIP was marked as External (by applying the Whois command and comparing it with the registry).

The retrieved information (i.e. IP, Domain, Admin, Severity, Location, Installed Application) about the internal server in the alert message included IP 10.233.62.247, and stored in the local knowledge base was:

Internal Server = {*(10.233.62.247, Sev1.edu.au, Tommy Schart, IT-developer group, CoNsoleKit Microsoft Visual C++)*

The retrieved information (i.e. IP, Domain, URLs, Location) about the external server in the alert message including IP 104.239.223.14, and stored in the global knowledge base was:

External Server = {*(104.239.223.14, service.macinstallerinfo.com, URLs*[10]*, US)*

The third stage: Since this study focuses only on malware, only Snort rules related to malware with the following titles were searched to identify the matched classification phrases: *snort3-malware-backdoor.rules,snort3-malware-cnc.rules,snort3-malware-other.rules, snort3-malware-tools.rules*. The matched Snort rule, which was mapped to the message part from the basic field, was as follows:

```
alert tcp HOME_NET any -> EXTERNAL_NET HTTP_PORTS ( msg:‘‘MALWARE-CNC
Osx.Keylogger.Elite variant outbound connection"; flow:to_server,
established; http_uri; content:‘‘/read-mip.php",fast_pattern,nocase;
metadata:impact_flag red,policy balanced-ips drop,policy security-ips drop;
service:http; reference:  virustotal.com/en/file/e23cae7189d6
ca9c649afc22c638a45fd94f19ef6b585963164cca52c7b80f9b/analysis/;
classtype:trojan-activity; sid:41458; rev:1;)
```

4.3.1.3 Inference layer

The purpose of this layer is to answer the *what*, and *how* questions. The “MALWARE-CNC Osx.Keylogger.Elite variant outbound connection” was the malware classification phrase (according to: Extraction layer, 3rd stage). The definition of this malware was extracted from web

[10]Because of the large number of URLs, not all are defined here.

articles in cybersecurity field and stored in global knowledge base. The extracted definition for the case study was compiled as follows: *malware classification phrase + 'is' + behaviour*. The definition was found under the 'Behaviour' node from the Symantec website [11], and included:

> *"OSX.Keylogger is a spyware program for Mac OS X that records keystrokes, may take screenshots, and may also send the information to a predetermined email address."*

Then, the malicious URLs were classified into five classes, each represented by the max-length URL. These were searched among the 12 relevant logs to provide evidence for the incident. The URL that was matched in the relevant logs was randomly selected for use in the next layer. Since the infected server was not located in the financial department and the severity was Medium, two rules based on Severity and Location were triggered, and the corresponding template for each was selected.

4.3.1.4 Story layer

The story based on the automatic retrieval of the variables from the previous layers was generated in this layer. A complete template was contrasted against the report obtained from the commercial tool. The report produced by the proposed model (Figure 4.2A) was compiled fully automatically, while the Secureworks report (Figure 4.2B) required *both* machine processing and human assistance.

I are receiving alerts for "*MALWARE-CNCOsx.Keylogger.Elitevariantoutboundconnection*" reported by your device 10.30.207.199 for outbound traffic (Not Blocked) sourcing from 10.233.62.247 and destined to 104.239.223.14. This indicates 10.233.62.247 may be infected with the Osx.Keylogger.Elite malware. OSX.Keylogger is a spyware program for Mac OS X that records keystrokes, may take screenshots, and may also send the information to a predetermined email address. The external host at 104.239.223.14 is currently flagged as malicious by multiple vendors per VirusTotal. Reference: https://www.virustotal.com/ip-address/104.239.223.14 Please investigate this at your earliest convenience. If you have any questions, you can reach us either by updating this ticket and delegating it to the Security Operations Center or by phone at 877-838-7960.
Regards,
Secureworks SOC

(A) The *Secureworks* report

I got an alert at 5:05 PM on 02/27/2019 because of the connection between 10.233.62.247 , Sev1.edu.au to 104.239.223.14 , service.macinstallerinfo.com . It seems the service.macinstallerinfo.com is in a black list, so the alert is generated because of malicious site. By looking up the relevant logs, I can find relations between the Sev1.edu.au and the malicious site with http://service.macinstallerinfo.com/Mac/getInstallerSpecs/?&channel=3Db5002&info= 3D238749466&encinfo=3D1&newInstallerVM=3Dtrue&vm=3D&machineId= 3D05fdb9a9faf29d565c8dfd9b9dabf2de]
The malicious site is *MALWARE-CNCOsx.Keylogger.Elitevariantoutboundconnection* , which is a spyware program for Mac OS X that records keystrokes, may take screenshots, and may also send the information to a predetermined email address.
Since the internal device is Severity: Medium located in IT-developer group , it is better that admin: Tommy checks the system and images it to make sure it is clear from any infections. Tommy can also check all the computers in Location: IT-developer group , where the internal device is located, to check all other connections to this malicious site.

(B) The *Storytelling* report

FIGURE 4.2: The reports generated in response to the security alert by both (A) Secureworks and (B) proposed solution

[11] https://www.symantec.com/security-center/writeup/2010-041918-0517-99

4.3.1.5 Analysis

Since the formal evaluation of the narrative format of both reports is qualitative in nature, the improvement in cyber threat management is challenging. In this empirical case study, I focused on the core questions to be answered in the report (i.e. actor (who), riskiness (what), and evidence (how)) as a basic for the proposed model evaluation. Thus, the following two criteria were defined: completeness and comprehension. In our case, completeness refers to the amount of information required to obtain full comprehension about the situation. By assumption, the storytelling model due to its auto-fill function from various knowledge bases provides the complete information required to take action. On the other hand, the classic report (Secureworks in this case) entails a manual search for missing information. To increase results' reliability, an additional 10 alerts were investigated.

Since different types of alerts require different investigation times, a random sample of 11 alerts was selected. An expert from the SOC team was involved in the empirical alert analysis consisting of filling the missing information from internal and external sources (similar to the model proposed). The Secureworks reports for the alerts classified as malware (Potential Device Compromise) were obtained between 11/02/2019 and 28/02/2019. Table 4.2 shows the status of the knowledge required to complete the report ('Completeness' header). The expert retrieved the necessary information manually, and the extraction time was measured in seconds ('Completeness Time' header). The average extraction time across the 11 malware alerts was 1455.(36) s (approximately 25 mins). Thus, in total it took approximately 30 mins to answer the core questions about the actor, riskiness and evidence (completeness = 25 mins + comprehension = 5 mins). As a result, the proposed model reduced the time to respond based on the full understanding of the situation by approximately 83% (25/30). In the storytelling model, given sufficient information on *what*, *who*, and *why* aspects, the time taken to obtain complete comprehension about an alert was approximately 5 mins (= 300 s). The time required for understanding was directly related to the degree of completeness (missing information has to be searched and extracted manually).

I also investigated the scenario where the 11 alerts occurred in a consecutive manner (busy period). To avoid potential damage and further escalation, the alerts should be addressed immediately. Time to respond to all alerts was set as a cumulative sum of Completeness + Comprehension Time of each consecutive alert. Since the alerts are processed sequentially, the total response time builds up. Table 4.3 demonstrates the cumulative delay time to respond to an alert in the case of 11 consecutive alerts received in a day. Given the scenario, the proposed model has the potential to reduce the response time by approximately 17000 s (approximately 6 days) in comparison with the report derived in a semi-manual manner by the SOC team (existing approach). Please note that human limitation and environment limitation were not considered in the experiment.

TABLE 4.2: Empirical Results

Alert Messages	Completeness				Completeness Time				
	Definition	APPs in Definition	Malicious Site Info	Malware Evidence	Definition[s]	Relevant Log[s]	Malicious Site Info [s]	Malware Evidence[s]	**Total Time[s]**
1 *Win.Trojan.Donvibsvariant*	No	No	No	No	50	1200	33	201	**1484**
2 *Win.Trojan.Donvibsvariant*	No	No	Yes*	No	50	1140	21	60	**1271**
3 *Gamarue.Andromeda.Trojan.Phone.Home*	No	No	No	No	75	1380	24	360	**1839**
4 *Agent.known.malicious.Virut*	No	Yes	No	No	68	1260	19	130	**1477**
5 *Agent.known.malicious.RookIE*	No	No	Yes*	No	154	1020	22	80	**1276**
6 *Win32.Virut*	Yes	No	No	No	0	1200	25	194	**1419**
7 *DownloaderTrojan*	Yes	No	No	No	0	1140	22	380	**1542**
8 *MALWARECNCOsx.Keylogger.Elite*	Yes	Yes	No	No	0	1320	24	145	**1489**
9 *Agent.known.malicious.Virut*	No	Yes	No	No	68	1080	21	99	**1268**
10 *Win.Trojan.Donvibsvariant*	No	No	No	No	60	1260	24	128	**1472**
11 *MALWARECNCOsx.Keylogger.Elite*	Yes	Yes	No	No	0	1320	24	128	**1472**

* Only Location is provided

TABLE 4.3: Empirical evaluation of the consecutive alerts

Alert No	Total Time by SOC		Total Time by ADSM	
	Completeness	Comprehension	Completeness	Comprehension
1	1484	300	0	300
2	2755	600	0	600
3	4594	900	0	900
4	6071	1200	0	1200
5	7347	1500	0	1500
6	8766	1800	0	1800
7	10308	2100	0	2100
8	11797	2400	0	2400
9	13065	2700	0	2700
10	14537	3000	0	3000
11	16009	3300	0	3300
	19309 [s]		3300 [s]	

4.3.2 Survey evaluation

ADSM focuses on the comprehension phase of cyber analysis intending to reduce the cognitive load imposed on security analysts. The developed field survey instrument examined CTA by comparing the incident reports and revealing how a complete incident report helps in cognition effort reduction and as a result of easier comprehension by analysts. Therefore, in this section, I report on our experiences conducting a between-subject experiment, i.e., a treatment group received and evaluated the Storytelling reports and a control group received an evaluated the Secureworks report.

The aim of this survey is making an evaluation on human cognitive factors by means of applying novel storytelling techniques from security logs and alerts through appropriate context enrichment. The explanation stories are enriched to provide the supplementary information for the objects and subjects of an incident. Therefore, the explainable model in a human understandable format is expected to cover cybersecurity issues allowing an expert and non-expert to acquire appropriate awareness to confirm that an alert is indeed a false positive or a real incident.

The output of the proposed model is an informative human-readable format report from the incident alert. As the formal evaluation of the narrative format of reports is qualitative, I aimed to evaluate the format through a survey. At first, the participant faced three incident reports. For

each report, questions focused on the completeness criteria (the amount of information required to obtain full comprehension of the situation). Based on the presented report, participants were required to rate and answer the specific questions designed under six core questions based on the 5W1H method. The particular questions are asked about the 6 "WH" questions, which the answers are from the presented report. Based on how easy they found the answer from the report, participants were required to rate each WH question on a Likert scale.The participant should only answer/rate questions based on the given information in the report.

Due to evaluating the generated report in terms of comprehension (the level of understanding about the incident and the potential action to be acted upon), two report styles (Storytelling and Secureworks) will be shown together to be rated based on the specific questions.

4.3.3 The surveys and questionnaires

Using the Qualtrics platform[12], a structured questionnaire involving three randomly selected incidents was designed to compare the incident reports generated by the ADSM and Secureworks [13].

The narrative format of both reports is qualitative in nature and is generated from the same incident alerts. The comparison between Secureworks reports and Storytelling reports in terms of comprehensiveness helps increase the results' accuracy. The Secureworks reports and the Storytelling reports are shown in Figures 4.3, 4.4 and 4.5, respectively.

The information in both reports was supplied to assist SOC members who manually perform a data triage task, as indicated in Figures 4.3 to 4.5. Data triage analysis usually involves examining the details of alerts and various resources to find the relevant data about the actor (who), riskiness (what), evidence (why), time (when), location (where) and mechanism (how) of an incident. The information in the Storytelling reports was provided under the headers *"Type"*, *"The connection"*, *"The malicious"*, *"Local info"*, and *"Recommended actions"* and was organised in a framework that highlights the (who, what, why, when, where and how) features of an incident. The importance of a cyber incident is mostly determined by its severity and location. Furthermore, the action recommendation proved beneficial for a timely and coordinated response, and is part of the inference phase conducted from the modifiable organisation's rule. For example:

- Since the severity level is high, it is better that the administrator checks the system and images it to ensure that it is clear of any infections
- The administrator should also check devices connected to this server to monitor data breaching.

[12] http://www.qualtrics.com/
[13] a commercial cybersecurity analytical tool https://www.secureworks.com/

02/27/2019 5:05 PM

I are receiving alerts for "MALWARE CNCOsx:Keylogger:Elite variant outbound connection" reported by your device 10.30.207.199 for outbound traffic (Not Blocked) sourcing from 10.233.62.247 and destined to 104.239.223.14. This indicates 10.233.62.247 may be infected with the Osx.Keylogger.Elite malware.

OSX.Keylogger is a spyware program for Mac OS X that records keystrokes, may take screenshots, and may also send the information to a predetermined email address. The external host at 104.239.223.14 is currently flagged as malicious by multiple vendors per VirusTotal.

Reference:https://www.virustotal.com/ip-address/104.239.223.14.

Please investigate this at your earliest convenience.

(A) The *Secureworks* report

I got an alert at 5:05 PM on 02/27/2019 because of the connection between 10.233.62.247 (Sev1.edu.au) to 104.239.223.14 (service.macinstallerinfo.com). It seems the service.macinstallerinfo.com is on a blacklist, so the alert is generated because of a malicious site. By looking up the relevant logs, the relations between the Sev1.edu.au and the malicious site has been confirmed.

Type:
The Device (Sev1.edu.au) may have been compromised through this non-secure connection.

The connection:
http://service.macinstallerinfo.com/Mac/\getInstallerSpecs/?&channel=3Db5002&info=3D238749466&\encinfo=3D1&

The malicious: *MALWARE-CNCOsx:Keylogger:Elite variant outbound connection*
It is a spyware program for Mac OS X that records keystrokes, may take screenshots, and may also send the information to a predetermined email address.

Local info: *Sev1.edu.au is an internal server device located in the IT-developer group. It doesn't carry financial information, but it contains sensitive data.*
****It was supposed that the IT- group and financial group contains high-risk records. The importance of a record is determined by the label and location it is belonging.*

Recommended Actions: 1) Since the Severity level is High, it is better the administrative checks the system and images it to make sure it is clear from any infections. 2) The administrative should also check devices connected to this server due to monitoring data breaching.

(B) The *Storytelling* report

FIGURE 4.3: The reports generated in response to the security alert of the first incident by (A) Secureworks and (B) ADSM

Secureworks reports (Part (A) in Figures 4.3 to 4.5) did not follow a consistent format (such as headers) to describe an incidence and facilitate inference, despite the fact that there was data to address the incident characteristics.

The structured surveys were conducted from March to July 2020. A total of 998 questionnaires were distributed, and 104 respondents returned completed the questionnaires. Since the participants were given three different incident reports in each survey, I collected a total of 170 responses to the Secureworks reports and 170 responses to the Storytelling reports.

In an ideal world, all security experts should be involved in the survey to gain accurate information,

02/27/2019 6:33 PM

I are receiving "MALWARE-CNC Win.Trojan.Donvibs variant outbound connection" alerts from your 10.30.207.199 device sourcing from 10.233.62.247 and destined to host 112.87.43.213, indicating that host 10.233.51.94 may be infected with malware.

Please investigate this activity at your earliest convenience.

(A) The *Secureworks* report

I got an alert at 6:33 PM on 02/27/2019 because of the connection between 10.233.62.247 (Sev1.edu.au) to 112.87.43.213 (xmlconf.rcv.sandai.net). It seems the sxmlconf.rcv.sandai.net is on a blacklist, so the alert is generated because of a malicious site. By looking up the relevant logs, the relations between the Sev1.edu.au and the malicious site has been confirmed.

Type:
The Device (Sev1.edu.au) may have been compromised through this non-secure connection. .

The connection:
https://xmlconf.rcv.sandai.net/?appid=54

The malicious: *VBS/Donvibs Malware*
A threat downloads and installs other programs, including other malware, onto your PC without your consent.

Local info: *Sev1.edu.au is a public server device located in the financial group, and it is carrying financial information but not sensitive data.*

****It was supposed that the IT- group and financial group contains high-risk records. The importance of a record is determined by the label and location it is belonging.*

Since the Severity level is Medium, it is better the administrator checks the system and images it to make sure it is clear from any infections.

(B) The *Storytelling* report

FIGURE 4.4: The reports generated in response to the security alert of the second incident by (A) Secureworks and (B) ADSM

as the number of experts with cybersecurity backgrounds is not high; an average of 35 people in each company. I had 170 participants as sample size, which put the survey accurately based on the confidence level and margin of error. The error shows the level of precision with means the actual value of the population was estimated to be [211]. The confidence level is determined to ensure that the average value of the attribute is equal to the true population value [211].

As the aim of the survey is to establish the proportion of cognitive burden that can be reduced by the storytelling report, and cognitive burden will vary from person to person, an error rate of 5% is acceptable and a 80% level of confidence is reasonable. These means 80 out of 100 samples will represent the true population value within the range of precision specified [211]. The sample size is calculated based on a math standard formula, where
Sample Size = (Distribution of 50%) / $\sqrt{(\%Margin\ of\ Error/Confidence\ Level\ Score)}$
If a margin of error is considered to be 5% and the confidence level is 80%, the outcome of

02/27/2019 7:28 PM

I have detected suspicious outbound traffic from your host at 10.233.62.247 triggering on alert 'MALWARE-CNC User-Agent known malicious user-agent string - Virut'. This activity indicates that the host has browsed to an IP address hosting a malware domain. This does not by itself indicate that the host is compromised. However, this may lead to the host being compromised as it may download malware.

(A) The *Secureworks* report

I got an alert at 7:28 PM on 02/27/2019 because of the connection between 10.233.62.247 (Sev1.edu.au) to 50.63.161.112 (lempar.com). It seems the lempar.com is on a blacklist, so the alert is generated because of a malicious site. By looking up the relevant logs, the relations between the Sev1.edu.au and the malicious site has been confirmed.

Type:
The Device (Sev1.edu.au) may have been compromised through this non-secure connection.

The connection:
`http://www.lempar.com/update/extrasecure-smadav.txt`

The malicious: *W32.Virut Virus* A virus that infects executable files. Some variants also infect ASP, HTML and PHP files. The virus has worm-like behaviour and spreads by copying itself to fixed, removable and network drives. It also opens a back door on the compromised computer.

Local info: *Sev1.edu.au is a public server device located in the education group, and it doesn't carry financial information and sensitive data.*
****It was supposed that the IT- group and financial group contains high-risk records. The importance of a record is determined by the label and location it is belonging.*

Since the Severity: Low, it is better the administrator checks the system to make sure at your convenient time.

(B) The *Storytelling* report

FIGURE 4.5: The reports generated in response to the security alert of the third incident by (A) Secureworks and (B) ADSM

the formula is 128 (Min); 170 is greater than this minimum, and is reasonable for the target population.

I sent the questionnaires to people working or studying in the cybersecurity area via LinkedIn[14] and social media. Only students enrolled in a cybersecurity course at postgraduate level were selected. The respondents who returned questionnaires were from different organisations/institutions in several countries and regions.

To ensure the randomisation of subjects, two separate surveys with the same questions but different incident report presentations were distributed to the cybersecurity experts and students. With the tracing distribution capability in the Qualtrics platform, I ensured that both groups of participants were involved in both questionnaires. According to the research topic, there is no

[14] `https://www.linkedin.com/`

clear risk. The participants were only required to read reports and score them based on their enriching level. The participants' involvement was beneficial for the body of knowledge.

The survey was comprised of four parts: consent, personal questions, completeness questions and comprehension questions. The details of each part are explained as follows.

4.3.3.1 Part 1 - Consent

Based on the research area, I were involved with human integrity and rights rather than animal and environment rights. According to the research topic, transparency was an issue in this research. Full transparency about what I are doing during the research was addressed on the primary agreement with the participants. In that way, our research meets the expectations of participants' transparency. I only asked participant to read the report and answer the questions, therefore no medical, health or human risk was raised by this research.

The participants were first asked to sign a consent form to indicate their voluntary participation in the survey, and ethical approval (Application ID: HRE20-001) was obtained from the Victoria University Human Research Ethics Committee (VUHREC). The first part of the survey also contained information about the study and details of what the respondents were required to do.

4.3.3.2 Part 2 - Personal Questions

The questionnaires were designed for particular participants: experts with cybersecurity experience in industrial companies, or students enrolled in a cybersecurity course at an institution. The questions regarding the participants' personal information included the respondent's background (cybersecurity expert or student), professional level and number of years of experience in the field of cybersecurity. The personal questions are shown in Figure 4.6.

4.3.3.3 Part 3 - Completeness Questions

The responders were provided with three incident reports of one type (Secureworks reports or the Storytelling reports). They were asked to answer and rate the information in the reports using the ***completeness*** criteria (the amount of information required to completely describe the situation). According to the 5W1H method, a report can be considered complete and have achieved its main objective of clarifying a topic if it answers the following core questions [17];

- **Who** was involved?
- **What** happened?

FIGURE 4.6: The questions for collecting personal information from the respondents

- **When** did it happen?
- **Where** did it happen?
- **Why** did it happen?
- **How** did it happen?

These questions constitute a formula to determine whether the report gave a complete picture of a subject (incident) to facilitate any necessary decision making by the subject. Thus, the 5W1H method was used on the cyber incident reports to evaluate and characterise the reported incidents. Each question required a factual answer to identify what experts must analyse and know about the incident [17]. One particular approach to support CCSA is from an analytic and model-based perspective that enriches content to reduce cognition effort of understanding:

- "**What**" was used to define the topic (incident) accompanied by the type of incident and its context
- "**Who**" referred to an organisation or an individual unit that is the owner of the cyber risk
- "**Why**" referred to the root causes of the incident

- "**Where**" referred to the location where the incident started or impacted
- "**When**" referred to the date and time of incident occurrence
- "**How**" was used to explain the mechanism applied in the incident.

The respondents were asked to put their answer in a box and then rate how easy it was to get the response from the reports using a 10-point Likert scale. Likert scales were used for cognitive state sensing and assessment for understanding human cognition augmentation. The level of difficulty in extracting information from the given reports was assessed generally, and was closely related to the report style and completeness. The scale intervals (0-9) were as follows: Not found **(Not at all easy)**, **(Extremely)** hard to find it, **(Very hard)** to find it, **(Hard)** to find it, **(Slightly)** hard to find it, **(OK)** to find it, **(Slightly easy)** to find it, **(Easy)** to find it, **(Very)** easy to find it, **(Extremely)** easy to find it.

Survey data were compromised when respondents did not find the factual answers in the way the researchers expected. Surveys needed to be reviewed and revised to produce more reliable and valid measurements. To prevent the submission of erroneous and inaccurate data, responses were checked. Knowing the correct responses and comparing them to the text that was typed into the blank boxes allowed the researchers to enhance their study by refining the rates. For instance, if a participant gave a wrong answer to the question, "Who is the victim?", and rated the question as "Very easy", the response should be revised to "Hard", as the answer was incorrect (he was not able to find it from the report easily). In Section 4.3.4, further details about the revision process are provided.

In summary, participants were asked to complete blank boxes with their responses in relation to the content presented in the reports so they could reflect on their opinions. Because the text was a quantitative variable, supplementary questions were created in advance using the Likert scale to gather respondents' responses in a measurable manner. These are referred to as "ratings" questions.

- **Rate 1**- How easy was it to identify the victim or the threat of the incident based on the given information in the report? (**Who?**)
- **Rate 2**- How easy was it to detect the type of the incident based on the information in the report? (**What?**)
- **Rate 3**- How easy was it to identify when the incident happened based on the information given in the report? (**When?**)
- **Rate 4**- How easy was it to identify which organisation or unit was involved in the incident based on the information given in the report? (**Where?**)

- **Rate 5**- How easy was it to identify the root cause of the occurred incident based on the information given in the report (**Why?**)
- **Rate 6**- How easy was it to identify the mechanism of the threat based on the information given in the report? (**How?**)
- **Rate 7**- How easy was it to identify the incident severity level based on the information given in the report? (**What?**)

4.3.3.4 Part 4 - Comprehension

In the fourth part of the survey, the incident reports were evaluated based on their effectiveness using the comprehension evaluation criteria (the level of understanding of the incident and the potential action to be taken). The degree of comprehension achieved by the report's narrative technique to minimise the cognitive burden imposed on cybersecurity analysts while processing a huge number of logs is referred to as comprehension. Likert-scale ratings can indicate how well analysts understand incident reports. Three five-point Likert-style questions were used to identify the participants' opinions of the effectiveness of the reports. The participants compared both types of incident reports (Secureworks and Storytelling) based on the comprehension criteria. To ensure that their responses were not impacted by the report type, respondents were unaware of which reports were Secureworks and which were Storytelling. The following three questions were asked to assess the level of comprehension of both report styles:

- **Q1**- How would you rate each report's effectiveness in relation to improving the analysts' cognition for a proper response? (understanding how to respond based on the information given in the Secureworks reports and the Storytelling reports)
- **Q2**- How would you rate each report's effectiveness in providing visibility into the incident? (understanding the incident based on the information given in the Secureworks reports and Storytelling reports)
- **Q3**- How would you rate each report based on ease of understanding? (by reading the report, you are able to comprehend it).

4.3.4 Analysis of responses

I began by analysing the responses to the completeness questions (Part 3) and then the responses to the comprehension questions (Part 4); both of which were categorised using the information obtained in Part 2. If necessary, I revised the responses before beginning the analysis because it was important that the questions be easily understood and result in low levels of response error.

The data entered into the blank boxes aided in the identification of errors when revising. The following rules were used to revise the responses:

1. The Likert-style rating was changed to zero levels because "no information was given in the report". Because no information was given about the organisational unit where the incident occurred in the Secureworks reports, the responses were changed to "0" if participants made any guesses or provided incorrect context
2. Errors were detected when respondents answered incorrectly. For example, the response was incorrect, but they assumed it would be "Very easy" to locate the information in the provided report. In such cases, the responses were revised so that no incorrect rating points were calculated. All incorrect answers were subjected to the same simple rule for this revision:
    - Rule 1- If a respondent did not correctly answer the question (in the black box) on a piece of information but rated it as "Slightly easy", "Easy", "Very easy" or "Extremely easy", their rate score was changed to "Hard"
    - Rule 2- If a respondent answered incorrectly and rated it "OK", "Slightly hard", "Hard", "Very hard"), their rating score was reduced by one level. For example, from "Hard" to "Very hard".

For both Parts 3 and 4, statistical analysis was carried out using both the descriptive and comparative models:

- **Descriptive Analysis**: The features of the data were described quantitatively, or summarised in a descriptive analysis. The average of the sample is referred to as the mean. The value that appeared most frequently in a set of data values is known as the mode. In relation to the responses collected from the questionnaires, the mean and mode were calculated. The results (Likert ratings) were summarised graphically and numerically for Part 3 (Completeness) and Part 4 (Comprehension)
- **Comparative Analysis**: In Part 3, a comparative analysis was used to compare the Storytelling and Secureworks reports' completeness levels. Comparative analysis was also used in Part 4 to assign a score to each report by comparing comprehension, effects, causes and consequences
- **T-Tests** were conducted to determine if there was a significant difference between the means of the responses for the Secureworks and the Storytelling reports. The t-test questions whether the difference between the report styles represents a true difference in the study or if it is possibly a meaningless random difference. In this case, the t-test allowed the model to be improved and re-estimated when some questions (parameters) were not significant.

4.3.5 Analysis - Completeness Level (Part 3)

4.3.5.1 T-Test

I performed t-tests for both the Secureworks and Storytelling reports by setting the threshold for significance to p = 0.05. A summary of the significance level tests adopted in this study is presented in Table 4.4. The t-value and p-value determined the significance level. As shown in Table 4.4, all seven questions are significant. Determining significance in the survey analysis meant "an assessment of accuracy" and shows that the survey results are accurate within a certain confidence level and not due to random chance [212]. In our case, this means a 95% confidence interval for the difference between the two groups, Storytelling reports and Secureworks reports.

TABLE 4.4: The t-test of the completeness rates for the seven questions (Part 3 of the survey) for the Secureworks and Storytelling reports

		T−VALUE	P−VALUE	Significant
Questions about completeness	R1. How easy was it to identify the victim or the threat of the incident (Who)?	-8.407	< .00001	YES
	R2. How easy was it to detect the type of the incident (What)?	-1.65	0.048	YES
	R3. How easy was it to identify when the incident happened (When)?	-3.6727	0.00016	YES
	R4. How easy was it to identify which organisation or unit was involved in the incident (Where)?	-12.875	<.00001	YES
	R5. How easy was it to determine the root cause of an incident (Why)?	-3.7521	0.00019	YES
	R6. How easy was it to identify the mechanism of the threat (How)?	-5.079	<.00001	YES
	R7. How easy was it to identify the incident severity (What)?	-9.5756	<.00001	YES

4.3.5.2 Descriptive and comparative analysis

Respondents assessed the seven Likert-style questions based on the 5W1H method for each incident. As a result, the average score (mean) across all questions may be determined; providing insight into how well respondents perceive the report to be complete. The mean has a high rate if the report is complete and provides rich information regarding the event subjects (What, Why, When, Who, Where and How). It implies that the respondents obtain all of the information that should have been provided.

The respondents were given seven questions about the ease of locating the information in the reports. The mean of the completeness level for the seven ratings for both styles of reports is shown in Figure 4.7. On average, the respondents' evaluations for the completeness of the Secureworks reports ranged from 0 to 6.1 (Not found to Slightly easy). For the Storytelling reports, the mean ranged from 4.8 to 8.8 (Slightly hard to Extremely easy).

The report is more complete and able to be understood by people if essential items can be found quickly. The ADSM was able to produce a more complete report. The Storytelling reports gave

participants more information to help them answer the 5W1H method's questions about the incident, and they were written in a better style, making it easier to discover the key items.

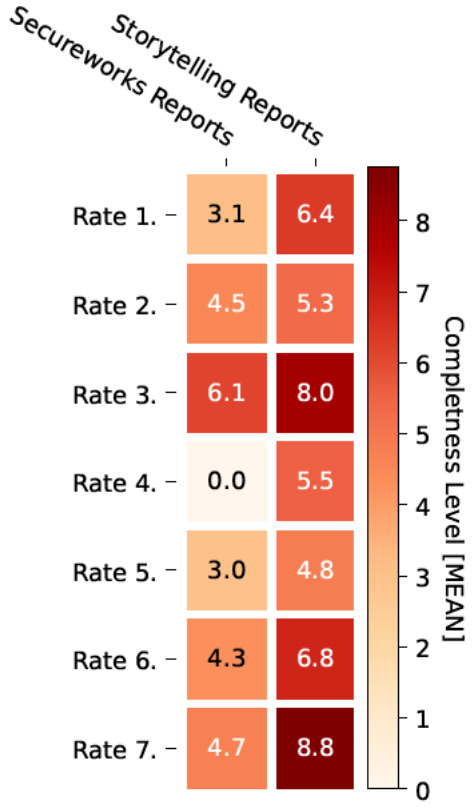
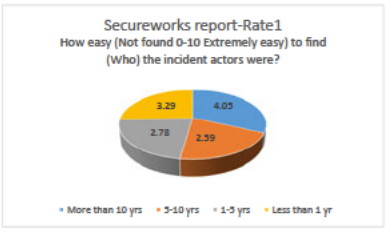


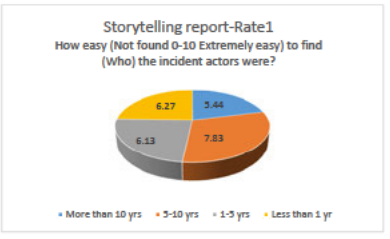
FIGURE 4.7: The mean of the completeness level of the seven ratings in Part 3 for the Secureworks and Storytelling reports

I will go over the responses question by question, grouping them by the personal information I acquired in Part 2. Figure 4.8 shows the mean of the completeness level in relation to answering "who" the incident actors were and how easily the report style assisted in finding the answer. In Rate 1, the participants were asked a "who" question to find the "victim" and "threat" as the incident's actors in the given reports. Then, they rated the report based on how easily they found the actors (Rate 1).

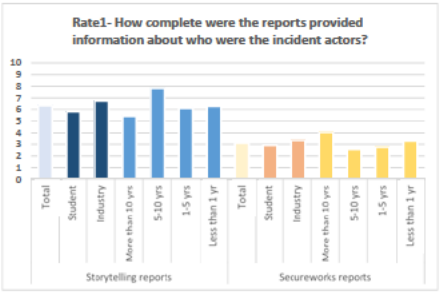
The pie charts 4.8a and 4.8b show the mean results for both the Secureworks reports and the Storytelling reports by the number of years of professional experience (either in industry or as a student) participants have had in the area of cybersecurity (in this study, this is referred to as professional years). The chart's blue section shows the average scores of the respondents. The orange, grey, and yellow sections illustrate the average scores of the respondents with 5-10 years', 1-5 years', or less than 1 year's experience in cybersecurity. By comparing chart 4.8a and chart 4.8b, it can be seen that all respondents, regardless of their number of years of professional



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**



(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



(C) **Rate 1 – How complete were the reports in providing information about the actors?**

FIGURE 4.8: Descriptive and comparative analysis of the Rate 1 distinguishing between types of respondents

experience, rated the Storytelling reports more highly than the Secureworks reports in terms of completeness (indicating that finding the answer to the items about actors was easier).

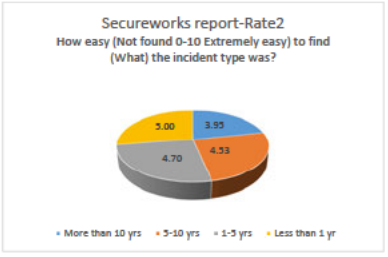
Chart 4.8c shows the results, based on the respondents' personal data. Each type of respondent is represented on the horizontal axis. Years of professional experience (less than 1 year, 1-5 years, 5-10 years, more than 10 years) and professional group are used to categorise the respondents (cybersecurity experts or students). The total average score (Total) for all responses relating to each type of report is also displayed on the horizontal axis. For each type of respondent, the horizontal axis of Figure 4.8c depicts the completeness level in respect to the first rated question.

The first light blue bar shows the total average score for all responses relating to the Storytelling reports and the first bright orange bar indicates the total average score relating to the Secureworks reports. As shown in chart 4.8c, in Rate 1, regardless of the type of respondents, everyone thought that, in the Storytelling report, finding the actors (victim and threat) to answer the Who enquiry was easy.

Figure 4.9 shows the mean of the completeness level in relation to answering "what" the incident type was, and how easy the report style was useful in finding the answer from the reports. These charts illustrate the results of Rate2.

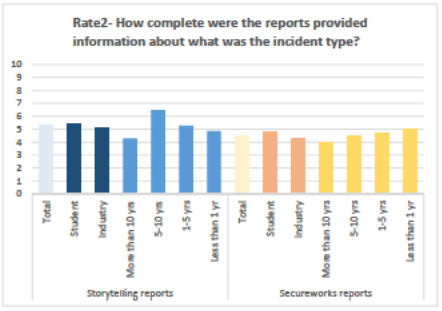
By comparing charts 4.9a and 4.9b, it can be seen that the participants with more than a year of experience gave the Storytelling reports a higher score than the Secureworks reports. The

type of incident was the same in both reports because they were derived from the same incident, but the reports' descriptions of the incident were different, which was noticed by those with a high-level of experience. Respondents were asked to choose the incident type from a list of possibilities, making it easy to respond to the question. The Secureworks reports were scored higher by participants with less than one year of expertise, but the difference in ratings between the Storytelling and Secureworks reports was small, less than 0.05. Since the total score for Storytelling reports is still higher, this can be ignored. Chart 4.9c summarises the total mean of rates, based on the personal information gathered in Part 2. As shown in chart 4.9c, the various groups rated the Storytelling reports slightly higher than the Secureworks reports in terms of completeness and it easier to them to locate the item from the given information to answer the "what" question, Rate 2.



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**

(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



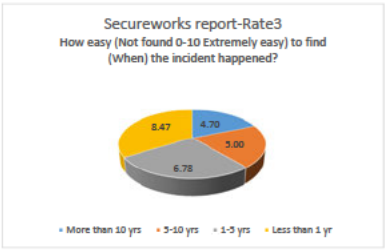
(C) **Rate 2- How complete were the reports in providing information about the incident type?**

FIGURE 4.9: Descriptive and comparative analysis of the Rate 2 distinguishing between types of respondents

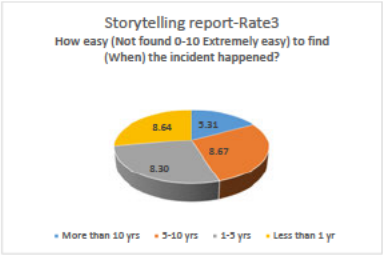
Figure 4.10 shows the mean of the completeness level in relation to answering "when" the incident occurred and how easily the report style assisted in finding the answer. The charts illustrate the results of Rate 3.

Pie charts 4.10a and 4.10b show the mean of the rating results for both the Secureworks reports and the Storytelling reports by the participants' number of professional years. A comparison of chart 4.10a and chart 4.10b shows that respondents with various professional years gave the Storytelling reports a higher score than the Secureworks reports. When compared to the

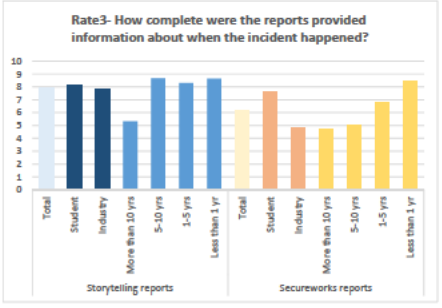
Secureworks reports, respondents with more than 10 years' experience found that the Storytelling reports were much clearer in explaining when the incident occurred, however both reports were rated low by experienced professionals. Others gave a higher score to both types of reports, but Storytelling received a higher rating. The report shows the ticket time when the alert was raised, not when the incident occurred, which is most likely why professionals with more than 10 years of experience gave such low ratings. Because activities are recorded in monitoring systems, it is difficult to determine the exact time the incident occurred. As a result, they may have expected the question to focus on the record time rather than incident time. Chart 4.10c, provides a better insight into how the participants with various years of expertise found it easier to identify the time of occurrence of the incident from the Storytelling reports.



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**



(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



(C) **Rate 3- How complete were the reports in providing information about when the incident occurred?**

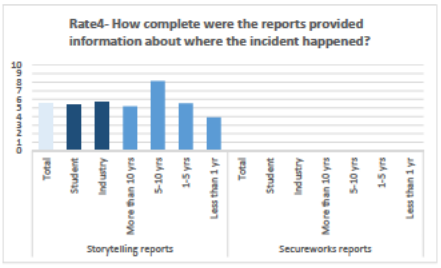
FIGURE 4.10: Descriptive and comparative analysis of the Rate 3 distinguishing between types of respondents

Figure 4.11 shows the mean of the completeness level in relation to answering 'where' the incident occurred and how easily the report style assisted in finding the answer. This figure illustrates the results of Rate 4.

The Secureworks reports did not contain any information as to where the incident happened or to which organisational unit the incident belonged, as illustrated in Figures 4.3, 4.4 and 4.5. Consequently, the reports contained no information about Rate 4. As a result, the responses were changed to 0 to indicate "No information was given," as per the first rule of revision. As a result, the charts and distributed analysis for the Secureworks reports are missing. In contrast, as shown

in Figures 4.3, 4.4 and 4.5, the Storytelling reports included information about the location of the incident.

As shown in Figure 4.11, the Storytelling reports made it easy for participants to identify the organisational unit where the incident occurred.



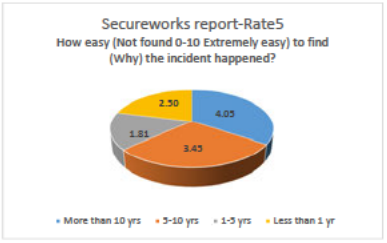
(A) **Rate 4 - How complete were the reports in providing information about where the incident occurred?**

FIGURE 4.11: Descriptive and comparative analysis of the Rate 4 distinguishing types of respondents. Since there is no information on this in the Secureworks reports, the means are zero

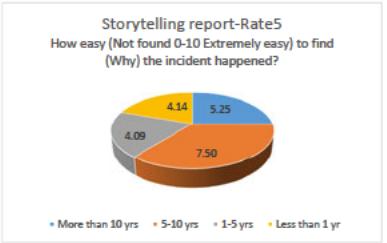
Figure 4.12 shows the mean of the completeness level in relation to answering 'why' the incident occurred and how easily the report style assisted in finding the answer. This figure illustrates the results of Rate 5. Most monitoring systems record events based on the triggered rules and patterns, whereas an event requires verification or enough evidence to be known as an incident.

A comparison of chart 4.12a and chart 4.12b shows that the participants with various professional years gave the Storytelling reportsa high rating for providing evidence as to why the incident happened. Chart 4.12c shows the total mean of rates, based on the personal information gathered in Part 2. TAccording to the findings, the Storytelling reports were more detailed in providing information about the possibility of an incident occurring than the Secureworks reports, as indicated by Rate 5.

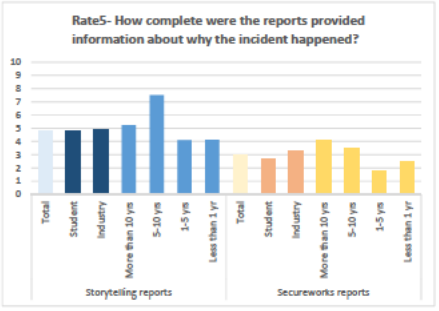
Figure 4.13 shows the mean of the completeness level in relation to answering "how" the incident occurred and how easily the report style assisted in finding the answer. The results in Charts 4.13a and 4.13b show that participants with more than 10 years' experience gave roughly similar ratings to both the Secureworks reports and the Storytelling reports when it came to describing the threat mechanism, whereas those with less experience or students thought the Storytelling reports provided a more complete explanation. It is possible that this is due to their extensive experience and knowledge of malware. As a result, they expected additional information and examples to round out the details of the malware detection explanation. The malware type and mechanism described in the reports did not convince participants with more than 10 years' experience that they understood how the incident occurred, hence none of the reports received a good ranking in terms of answering "how" the incident occurred.



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**

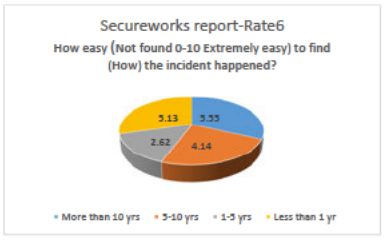


(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



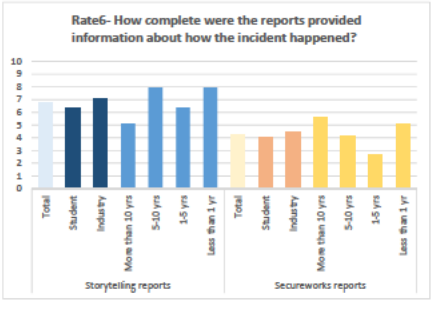
(C) **Rate 5-How complete were the reports in providing information about why the incident occurred?**

FIGURE 4.12: Descriptive and comparative analysis of the Rate 5 distinguishing between types of respondents



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**

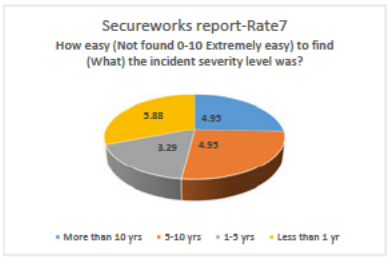
(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



(C) **Rate 6-How complete were the reports in providing information about how the incident occurred?**

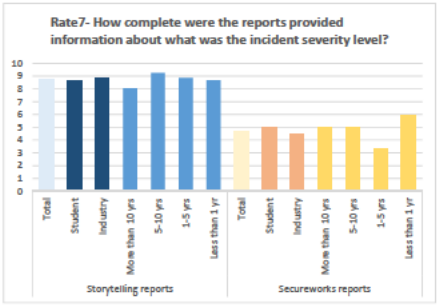
FIGURE 4.13: Descriptive and comparative analysis of the Rate 6 distinguishing between types of respondents

Figure 4.14 shows the mean of the completeness level in relation to answering the question about "what" in terms of the severity of the incident and how easily the report style assisted in finding the answer. A comparison of chart 4.14a and chart 4.14b shows that participants with several years' experience gave the Storytelling reports a higher score for offering a fuller explanation regarding the severity of the occurrence. Chart 4.14c shows that the level of completeness for the Storytelling reports is higher than the score given to the Secureworks reports.



(A) **Distribution of the responses (mean) based on professional years for Secureworks reports**

(B) **Distribution of the responses (mean) based on professional years for Storytelling reports**



(C) **Rate 7-How complete were the reports in providing information about the severity level of the incident?**

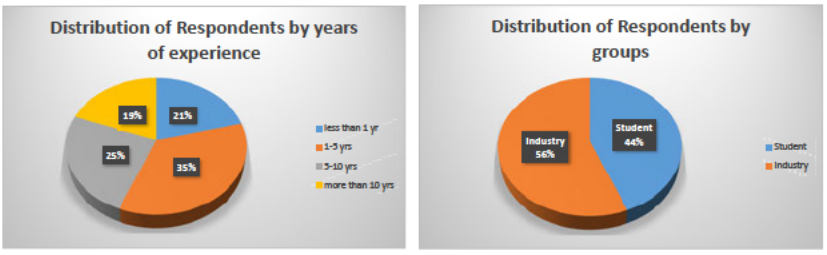
FIGURE 4.14: Descriptive and comparative analysis of the Rate 7 distinguishing between types of respondents

4.3.6 Analysis - Comprehension Level (Part 4)

Part 4 requires respondents to score the level of comprehension of both the Storytelling and Secureworks reports by answering three questions. Participants were able to compare and score the information supplied in the reports using the comprehension evaluation criteria because both reports were shown to them side by side. Participants in Part 3 were unable to compare the two reports since they did not know if the report they were given was the Secureworks or Storytelling report. As a result, the participants rated the information in the reports only using the completeness criteria.

4.3.6.1 Distribution of analysis

In Part 4, I received 171 responses that rated the level of comprehension of both types of reports. To identify who took part in the surveys, the results are classified based on the participants' personal information. Figures 4.15a and 4.15b depict participant distribution by years of experience and group. The bulk of participants had 1 to 5 years of experience, and only 19% had more than 10 years' experience. As shown in Figure 4.15b, there were 10% less students than professionals from cybersecurity organisations. I believe both groups had sufficient cybersecurity knowledge to assess the incident reports. Because the survey URLs were largely posted on LinkedIn (as a professional interactions platform), the majority of the respondents were from industry.



(A) (B)

FIGURE 4.15: Demographic analysis of survey respondents for Comprehension questions

4.3.6.2 T-Test

The independent samples t-test is used in hypothesis testing to compare sample means from two independent groups for an interval-scale variable since the distribution is approximately normal [213]. Statistical hypothesis testing offers a rigorous and objective approach to distinguishing truly significant differences in measurements from noise. I performed a t-test on the survey results by setting the threshold p = 0.05 to ensure the results were accurate within a certain confidence. By evaluating the t-test and calculating the p-value based on the t-value, the level of significance is determined. Table 4.5 shows that all the results for the three questions in Part 4 are significant and have a 95% confidence interval, which is not due to chance. Table 4.5 also shows the means of the rates from the Likert scales. On average, for all three questions, the Storytelling reports obtain better scores than the Secureworks reports in terms of effectiveness.

4.3.6.3 Descriptive and comparative analysis

The participants were asked three questions about both types of report in terms of comprehension. The comprehension level was calculated based on how effective the reports were in improving the analysts' cognition, providing visibility into the incident, and improving ease of understanding. Figure 8 shows that, on average, the effectiveness for the Secureworks reports was rated "Poorly

TABLE 4.5: T-test and descriptive analysis on Comprehension questions

	Secureworks	Storytelling			
	MEAN	MEAN	T-VALUE	P-VALUE	Significant?
Q1. Rate each report's effectiveness for improving the analysts' cognition for a response.	36.18	68.82	-12.702	< .00001	YES
Q2. Rate each report's effectiveness for providing visibility into the incident.	40.74	68.68	-10.4637	< .00001	YES
Q3. Rate each report based on ease of understanding.	54.41	77.94	-8.96918	< .00001	YES

to somewhat effective", (35% - 55%), while the Storytelling reports were rated "Somewhat to very effectiveness", (65%-75%).

In Question 3, the respondents were asked to rate the reports in terms of their comprehension level after reading the incident reports. In other words, they were asked to compare the reports and rate them based on the level of understanding, which was the main goal of the surveys. The results show that the Storytelling reports gained a higher score compared with the Secureworks reports (77.9 versus 54.4). Further insight into the comprehension level was obtained by calculating the mode as shown in Figure 4.17. The results show that most of the respondents felt that the Storytelling reports were 100% comprehensive incident reports.

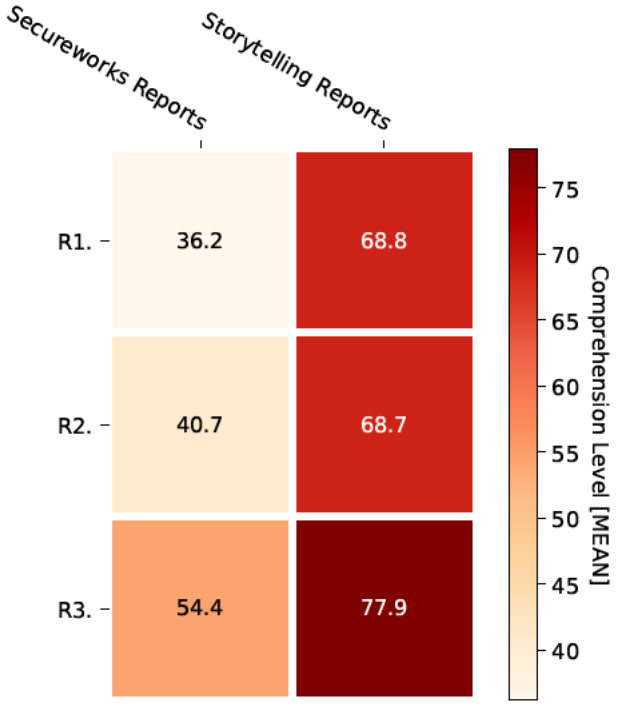


FIGURE 4.16: Comparing the Comprehension level (mean) of the Secureworks reports and Storytelling reports

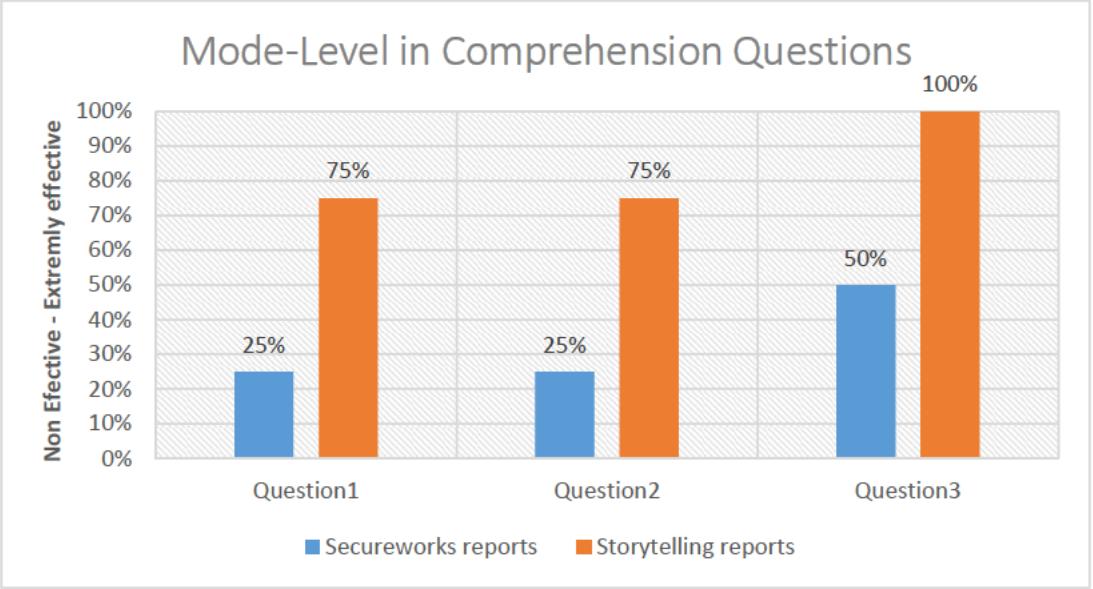


FIGURE 4.17: Comparison of the comprehension level (mode) of the Secureworks reports and the Storytelling reports

4.4 Discussion

The improvement from a human-computer interaction perspective in security alerts handling will be discussed using the main two criteria: (1) Completeness and (2) Comprehension. Based on the survey results from the cybersecurity experts and students, the Storytelling reports generated from the ADSM were evaluated and rated based on how helpful the insights into the incident were. To obtain a better understanding of the results, the Secureworks reports generated by the commercial vendor for the same incidents were rated and compared with the Storytelling reports.

Completeness: The information in the Secureworks report was insufficient for prompt inference, and the SOC member had to manually gather the complementary data from different sources. For instance, the information about the risk severity *(medium)* as well as the internal location of the device *(IT-developer group)* were missing. Also, the action recommendation *(check the system and images)* and person designation *(admin Tommy)* proved beneficial for timely and coordinated response. The utilisation of local and global knowledge bases aimed to provide rich and comprehensive context around the incident. The template was filled using both internal information as well as the external sources.

The Storytelling reports also obtained a better than average rating in comparison with the Secureworks reports. Some of the core items required to answer the questions did not exist in the Secureworks reports. For instance, information about the risk severity and the internal location of the victim device (risk-unit) was not mentioned in the Secureworks reports. This resulted in the Secureworks reports obtaining a lower rating in comparison to the Storytelling reports in terms of completeness.

The Storytelling reports were also recognised as having a better reporting style and the readability of the reports was sufficiently clear and detailed to allow analysts to digest the content of the reports comfortably. For instance, the respondents indicated that it was easier for them to find the answer to "Who was the victim?" or "What type of incident had occurred?" from the Storytelling reports, but it was difficult to find these answers from the Secureworks reports.

The Secureworks reports required human involvement in the generation process, however the utilisation of local and global knowledge bases in the ADSM meant that human involvement was not needed for the Storytelling reports, and the context about the incident was rich and comprehensive. Cybersecurity interpretation is also heavily reliant on analytical experience and knowledge (where and how to search for relevant information), which puts a strain on already scarce cybersecurity resources, but this was reduced using the ADSM.

Comprehension: The narrative technique was applied in both cyber incident reports with the aim of reducing the cognitive load imposed on cybersecurity analysts whilst processing the cyber alerts. The reports generated in a storytelling manner proved to be more human-readable, facilitated comprehension, and effectively allowed for a faster response to potential threats (the time factor is found to be crucial in the cybersecurity domain). Also, the human-readable format of the report contributed towards wider audience engagement with CSA (currently restricted to security professionals). As an example, the user of the infected device can receive the storytelling report and obtain an insight into the cyber situation instantly, thus preventing further problem escalation. The narrative format assists understanding despite a lack of expertise in cyber security domain.

As discussed in Section 4.3.4, the respondents obtained better insight into the cyber situation instantly by reading the Storytelling reports. The respondents found that generating and aggregating the necessary information with the Storytelling reports was more effective than with the Secureworks reports. As a result, the cognitive effort in information digestion and understanding was significantly reduced in the Storytelling reports and as analysts did not have to conduct a manual search, so a large amount of time was saved.

Finally, the ability to provide reports at different levels of detail automatically enabled the report to cater for various information needs and intended aim (i.e. low-level for the Security Operation Centre, high-level for top management).

Summary: By comparing the generated story and the Secureworks report, the following can be inferred:

- The storytelling report is fully generated automatically, reducing the burden on cybersecurity resources

- The implicit knowledge (what happened and why?) which analysts have to investigate manually, is included in the generated story
- The log files with private information that cannot be sent to the third party for further processing are protected
- The merging of information into a human-readable report to aid analyst cognition in gaining a better understanding of the incident and being more aware of the situation is provided by the Storytelling reports. As a result, analysts are able to save time by making quicker decisions and responding to incidents, resulting in fewer breaches
- Experts and students who were members of the sample gave Storytelling a high rating in terms of completeness and comprehension, indicating that it was effective in raising knowledge about current occurrences (knowledge digestion and comprehension required the least amount of cognitive effort).

In terms of current limitations, in this study I only focused on malware taxonomy for approach demonstration. Still, the model can easily be adapted to other types of incidents by providing the complementary sources in the local and global knowledge bases. Also, since an enriched report for a security alert in a story design is not available, I were not able to perform the direct comparison with the proposed storytelling model. Thus, the impact of the narrative format has been assumed to be beneficial for cognitive workload reduction based on empirical observation of the SOC team at the university.

4.5 Summary

In this chapter, local information is highlighted in the incident reports to help analysts better understand the incidents and their impact on the organisation. Global information is linked to the locally collected data aggregated in the incident reports which reduces the analysts' manual search and cognition loads. The incident generated in a storytelling manner is human-readable and facilitated improved comprehension. The report generated by the proposed model proved to be more complete and more comprehensible for the SOC team in comparison with the Secureworks report. As a result, the cognitive effort in information digestion and understanding was significantly reduced. Also, due to the human-readable format, a wide range of staff with different levels of expertise was able to be involved in the cyber risk management process.

Two surveys were designed and analysed to evaluate how cybersecurity experts/students found the reports generated from the LDSM and the commercial vendor in terms of completeness and comprehension. An analysis of the responses shows how the incident reports are useful for reducing expert cognition load. I assessed the human judgments of those who answered

the questions by rating how easily they found the items from the presented incident reports. To accomplish this goal, in Part 3, the participants were asked questions based on the 5W1H method to assess the completeness level. Part 4 measured the rate of comprehension gained by analysts while reading reports. The Storytelling reports, by focusing on both available local and global information to elucidate the environment's elements to describe the state of an incident, were rated as a beneficial report style in terms of completeness and comprehension. This chapter results was published in the paper 1 of publications.

Chapter 5

Explainable Intelligence to Interpret Cyber Alerts - Shared Situation Awareness

In real-world situations, several incident alerts are investigated by specialised staff. In order to provide prompt responses to serve incidents or ignore false alarms, alerts are prioritised and analysed. Security professionals rely on information provided in the alert message. Insufficient information in alert messages raise challenges for security analysts that require them to keep track of all local and global sources to identify the relevant information. The previous chapter discussed and evaluated self-awareness by proposing the ADSM.

This chapter emphasises shared-awareness in the process of cyber incident management in order to propose an explainable intelligence model. A Narrative Visualised Analytical Model (NVAM) is proposed, and a knowledge graph as a visualisation model is used in the proposed intelligence to present the relationships. The knowledge graph is proposed to capture the complex relationships between the alert and relevant information from the local and global knowledge bases to reduce the cognitive effort in information digestion and to understand a wealth of security data.

To enable cooperation in the cyber incident management process, it is necessary to generate a knowledge graph and interpret it in a human-readable format. The current machine-friendly formats for reporting incidents from alerts are extensive and complex. These characteristics hamper the readability and contribution, therefore preventing humans from understanding and being up-to-date with an incident.

NVAM contains four life cycles to help an analyst better understand the elements of the environment by involving more staff in the incident management: (1) analysing the alert, (2) designing the knowledge graph with the natural language sentences, (3) automatically implementing the

incident report in natural language by applying novel storytelling techniques from the knowledge graph and (4) maintaining the graph with the contribution of different levels of expertise. The performance of various NVAM's cycles is demonstrated in a case study with an example scenario from the SOC at an educational institution; highlighting its useability.

5.1 Introduction

5.1.1 Knowledge beyond security team is needed for the analysis

Most alerts are false alarms. To properly interpret the potential risk of an event, it often requires knowledge beyond the security department itself. Here are two examples.

As the first example, consider a scenario where a server of organisation XYZ is used for temporary storage and web testing, and is labelled as a non-critical host. Most of the alerts regarding the server can be omitted unless it is a serious breach. In the current financial reporting season, the server was borrowed by the financial department for financial reporting and budget planning. It now keeps critical information and the security level is raised to the highest level in the organisation to monitor all alerts closely. However, the role transition of the server is not passed on to the security team. Its users at the Finance Department have little expertise in security. A large security hole is left open to attackers.

In the second example, consider a scenario where an organisation repeatedly receives a high volume of a security breaches from a local host. This is a typical symptom of attack, and the security system blocks the host and related ports. A further analysis involving staff from different departments reveals that the host is an experimental server in Department A's laboratory which is used to test game engines' cloud end under development. The local host and cloud server require low-level communication and configuration with the corresponding security exceptions.

In both examples, the alerts' analysis requires knowledge from the security team and other departments which cannot be modeled and integrated with the alert analysis. Either false alarms could be triggered or high-risk alerts neglected. Therefore, engaging more staff from different departments is needed to solve the issues in the examples. They require developing a shared understanding of CSA (currently restricted to security professionals). Generating a report in a storytelling manner with an analytical graph to present the relationships is human-readable, facilitating comprehension, and allowing effective human involvement in the process.

5.1.2 Association analysis based on up-to-date local and global information

Security analysts need to gather local (the integrated information like the network infrastructure) and global (vulnerability, cyber threat and intrusion alert) information from various sources to feed the correlation process and support analysis of security events to explain the alerts [19]. The abundance of available cybersecurity knowledge raises challenges for security analysts and professionals who have to keep track of all the available sources and identify relevant information by provided by them [214]. Analysts' inability to identify the most relevant information that is not easily readable by humans can be considered a data quality issue, as can the overabundance of cybersecurity knowledge available in various formats [215].

In generic terms, comprehensive and integrated up-to-date information for security experts includes any details that can be used to characterise an IT entity's situation. The information is considered as linking to locally and globally available information [19]. Local and global information is continuously updating. Implementing approaches to integrate the information into the data model to make full use of cybersecurity-related details from various resources, and associating all these security-related knowledge is difficult and usually incurs expensive modification costs [46]. One of the major challenges is the rapid variation of network environments which has a potential impact on security posture, i.e., machines added and removed, various patches applied, applications installed/uninstalled and confidential data uploaded or deleted [216].

> "The problem is not lack of information, but rather the ability to assemble disparate pieces of information into an overall analytic picture for SA" [48].

To update the knowledge base, the data collected from humans can be converted into machine data either automatically or manually. To derive relationships to machine data, rule-based correlation and aggregation are the famous approaches. In order to facilitate the definition of rules, it can be helpful to visualise the generated data and separate the rules from the detection model [202]. Many state-of-the-art studies have been carried out, such as [217–220]. They propose visualisation of the knowledge graph to aid security analysts in their investigation.

At present, such approaches have been used in the application of knowledge graphs that consolidate data into a comprehensive picture [216]. Narrative reports accompany the analytical graph to compensate for the lack of data, leading to improved understanding. In this chapter, to enable cooperation in exchanging knowledge between humans and avoid peering into internal structure in machine-readable formats, the rules are defined in a human-readable format. Generating narrative reports with the interpretation of knowledge graphs for incident alerts facilitates comprehension because of human-readability, and effectively provides a better ground for faster response.

5.1.3 Threat intelligent sharing for mutual learning

Most organisations and participants now agree on the value of exchanging threat information for a variety of purposes. It has been shown that exchanging sensitive vulnerability data will help to prevent possible cyber attacks as well as counter current attacks and future risks. Leading cyber crime researchers, according to the Bipartisan Policy Center [221], agree that public-private cyber intelligence exchange speeds up the discovery and detection of attacks. As a result, if companies can detect an intruder during his active periods, they have a better chance of stopping the attacker before data is compromised [222]. Participation of various data security intelligence sharing exchanges is the most important concept. A higher SA of the threat environment, broader understanding of threat actors, and greater agility to protect against emerging threats are all advantages of sharing [223].

According to the findings of a recent survey [224], threat information sharing will help organisations strengthen their security posture and SA. Sharing risks, in general, improves collaboration for mutual learning and reaction to emerging threats, as well as reducing the risk of cascading consequences across a whole system, market, business, or across industries [223]. Despite the apparent advantages of sharing security data, there is a reluctance to disclose breaches. The authors of [225] reveal the numerous obstacles that hinder cooperation opportunities such as untrustworthy actors, consistency problems and so on. Regardless of the issues that exist in cooperation, sharing knowledge and having an on the ground understanding of the incident by those who are accountable adds intelligence to the incident management process, ensuring that no crucial event is missed unintentionally.

5.1.4 Graph-based analytical and storytelling representation

NVAM is proposed to enable security experts to analyse alerts and represent the incidents' intelligent analysis in enriched textual narrative reports. This approach combines an innovative manner of presenting the alert and corresponding data in a graph with an interactive human-friendly component for analysing and editing the threat information, as well as the interpretation of the graph's knowledge in a narrative report that is human-readable.

I chose our analytical strategy as a graph-based model to bring together isolated data and the varying update intervals of the data sources from the local and global knowledge bases of an organisation. The knowledge graph is built to carry out an alert correlation analysis. The knowledge graph gracefully supports human-friendly sentences that might be needed for revising/expanding the information due to exploring more security knowledge. It is well suited to the integration of many types of data that security analysts may wish to correlate. It also

brings together employees from various departments with varying levels of skill. The graph-based approach is used to expand network attack predictions by retrieving the target host's vulnerabilities and then retrieving the vulnerability associated with the alert.

Because the success and efficacy of every security action relies greatly on the involvement of security specialists, NVAM turns alerts and association knowledge into human-comprehensible narrative [217]. Even persons with no specialist security expertise about the issue can benefit from the summarised accounts since they provide a holistic view for better tracing security alerts. In the processing of a large number of alerts, detecting and tracing malicious events with a combination of machine and human powers is a more reliable method [226]. Our strategy will aid in the establishment of an intelligible common ground between humans and machines by using summarised stories.

Storytelling is a method to assist or engage people to explore and interpret complex real-world problems. In other words, telling stories in the problem formulation stage merges synthesis and analysis, and makes abstract concepts more concrete [227]. Storytelling can be used as a knowledge representation method to highlight the semantic and implied information from events into a human-readable format [58].

Given the huge volume of events and corresponding alerts, the stories need to be generated automatically. In this chapter, interpreting security alerts from different aspects, from a holistic view to technical solutions, into a natural language is proposed. Therefore, expert and non-expert analysts could analyse data beyond the security rules and policies.

The remainder of the chapter is organised as follows. First, NVAM and its development cycles are introduced and then the aspects of the designing NVAM are explained in Section 5.2. Section 5.3 illustrates the usefulness of the NVAM by means of an example use case. The chapter ends with a conclusion and consideration of the current limitations.

5.2 Narrative Visualised Analytical Model (NVAM)

This research's initial overarching aim was to present the vulnerabilities and threat factors and local elements associated with the incident alert, which is effectively understandable by humans for alert validation. Based on an idea proposed by Afzaliseresht et al. in [1] that was explained in chapter 4, NVAM aims to reduce the cognitive load imposed on cybersecurity analysts while processing alerts. This chapter presents a revised analytical strategy of the knowledge graph model which is capable of generating the automatic story by exploring the subgraphs. A shared understanding of CSA is developed in order to engage more people in updating their knowledge.

The NVAM comprises four cycles to assist an expert in CSA. As a result, full awareness of the alert situation from various heterogeneous sources, such as different departments and owners, can be achieved. The main development life cycles are illustrated in Figure 5.1 and explained as follows.

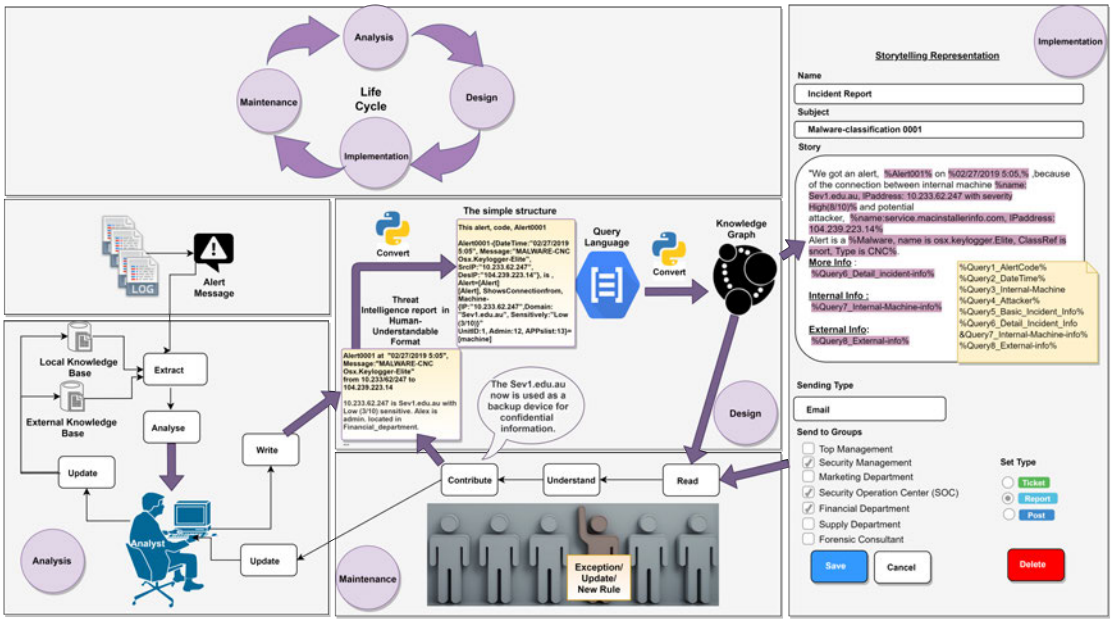


FIGURE 5.1: Overview of the NVAM's development life cycles consisting of four cycles (the story representation is the output of the implementation cycle which capable of revision based on updated knowledge from the maintenance stage)

5.2.1 Analysis cycle

Although monitoring systems help filter millions of logged events and generate security alerts, final human assessment is still part of the process. The analyst analyses the alert by using the extracted information from the local and global knowledge bases. The alert record is processed using regular expressions to extract the fundamental fields of any incoming alert recognised in the network flow. From the knowledge bases, relevant information with potential logical linkages to the alert record is filtered and extracted [228]. To explain the incident evidence and its possible impact, Internet scan engines and public threat exchange repositories are used to analyse information and their relationships. The correlated data and analytical findings are used in the report [66]. What is the fundamental component of the incident? This version has extremely simple human-readable sentences to explain, What is it? Who is the attacker? Who is the victim? When and where did the incident occur? What evidence that it happened appears in the logs? How much the asset was affected? All of the data is used to fill in the report's pre-defined template. A draft incident report is written for sharing the knowledge with others. The knowledge bases are regularly updated based on the available updated information (from other staff or public resources).

5.2.2 Design cycle

The analyst writes the analytical results in an incident report as a series of sentences which are converted into a simple structure, and then into query language which generates the knowledge graph automatically. Given that such alerts are machine-readable rather than human-readable, interpretation of raised alerts and cyber threat intelligence information is still required. Sharing is considered effective in a variety of cybersecurity situations, that provide a comprehensive situational awareness (shared SA) of potential threats and current incidents [229]. Cybersecurity analysts share the incident report informally as a text. Writing a lot of relevant information in a human-readable format reduces the cognitive load on cybersecurity resources, which are already limited.

5.2.2.1 Knowledge graph construction

The synthesised knowledge is visualised using a graph-like structure to support the analytical reasoning [230]. Usually, a historical collection of domain-specific knowledge is designed and developed prior to constructing a knowledge graph [231]. Here, a predefined set of vocabularies and relationships are not required. The graph is intended to be a flexible middle-ware analytical tool to avoid a solid set of rules. In other words, the initial alert messages and corresponding information from the local and global knowledge bases are combined in the graph to visualise the ongoing threats, to better comprehend an entity's semantic connections.

Graph databases are classified into two types based on the graph model they support: property graph and RDF. In general, RDF graph databases provide an emphasis on semantics and interoperability, whereas property graph databases place an emphasis on usability and performance. I chose the Neo4j to explain graph data modelling [232], though the simplified exchange syntax for ontology language is Resource Description Framework (RDF). Compared to Neo4j, RDF triples are machine-friendly syntax not suited to human understanding. RDF is very strongly index-based, which should be defined in the triple-oriented-language, while Neo4j is navigational (implements index free adjacency) and stores the connections between connected entities without scanning indexes [232].

In addition, Neo4j is opensource, significantly improving the processing efficiency of massive RDF data replication. Its Cypher query language is a very expressive query language built ground-up for humans to perform graph queries [233]. A Neo4j knowledge graph is a semantically enhanced insight layer of linked data that allows reasoning with the underlying data and reliably utilising it for sophisticated decision-making.

The following concepts are used to model a Neo4j graph [233]:

- **Nodes**: Concepts or entities in the domain
- **Labels**: Tags adding more meaning to nodes besides adding constraints and indices
- **Relationships**: A directed, semantically connection between nodes to depict the relations between them
- **Properties**: Key-value pairs that depict more information about nodes and relationships.

A graph consists of nodes, attributes and relationships. The knowledge entities as nodes and properties of entities as attributes must be recognised in the domain. How these entities and attributes are related to one another, and what entities are introduced at particular times should be captured [234]. The majority of graph queries aim to find an explicit pattern within the graph database. In a regular data processing system, graph queries have the expressive power to return the entity to the level of an analytic. Because the analysis is beyond the simple relations that can be maintained in tables, graph databases require various models. Before calculating specific data, certain searches need to follow numerous links. Initially, each graph database developed its own query language. Companies have recently begun cross-pollinating by introducing new implementations, aiming towards an opensource standard. The following are the most common graph query languages:

- **Gremlin**: A graph-searching language that was created for the Apache Tinkerpop project and enables declarative or procedural queries.
- **Cypher**: This declarative language, first developed by Neo4j and then adopted by others as OpenCypher, lets users search for nodes and edges that meet specific criteria
- **GQL**: This proposed standard aims to bring the Cypher, GSQL and PSQL styles together
- **SPARQL**: A query language for knowledge graphs represented in the RDF format
- **AQL**: ArangoDB's initial procedural language
- **PGQL**: Oracle's original language for searching and gathering data from nodes that fit certain criteria
- **GSQL**: TigerGraph's initial procedural language.

Cypher is a well-known querying and updating language for property graph databases and is used most widely [235]. Cypher has features for searching and altering data, as well as defining schema definitions such as linear queries, pattern matching, data manipulation and pragmatic queries [235].

5.2.2.2 Report into query converter

The knowledge graph, Neo4j's data model, is made up of nodes which represent entities (such as users, alerts, departments and so on), and relationships which indicate the links or relationships between the entities. Based on the domain, these entities should be recognised. Knowledge bases that integrate multi-source heterogeneous data can be helpful in identifying entities and their relationships. For obtaining contextual insight about an incident through the aggregation of isolated data, two main domain knowledge bases (local and global) are required.

The local knowledge base includes available information with domain knowledge of experts and the raw data collected from the security devices. The local knowledge base allows the exchange of explicit knowledge about the situation of an incident relevant to the company/institution. The global knowledge base contains supplementary information that is collected by external companies and researchers. Existing instance data in the global knowledge base is comprised of the following information: (1) Whois[1] and additional information about the IP address and domain registrants, (2) online scan engines such as Virus Total[2] or Threat Miner[3] that generate the "malicious" labels for a given URL or file, (3) the Snort community rule set[4] and (4) online open threat exchange repositories such as AlienVault[5], Windows Defender Security Intelligence (WDSI)[6] and Symantec[7].

The process of identifying entities and creating graphs can be done manually or automatically. If a manual approach is used, a person should specify the types of entities and connections that are permitted based on local and global knowledge bases. Not only is this a time-consuming task, it is also impossible in incident management. First, current global knowledge bases and resources are rarely publicly available, and those that are do not provide enough detail to represent the repositories I expect to model. Second, local network environments change frequently and must be manually updated.

The automatic approach is the best preference because I want a flexible graph that can be updated quickly and affordably by updating the knowledge bases that can be used in different incident types. A human-friendly convertor is intended to automatically convert simple vocabularies from an incident report to nodes, properties and relationships. As a result, Python scripts that accept human-readable sentences as inputs are used as automatic convertors, assisting in the collection and facilitation of knowledge from various organisational units, as well as from/to other existing conceptualisations or knowledge bases. A tool that uses two Python scripts to

[1] WhoIs. http://www.whois.com/s
[2] https://www.virustotal.com/
[3] https://www.threatminer.org/
[4] https://www.snort.org/downloads
[5] https://otx.alienvault.com/pulse/
[6] https://www.microsoft.com/en-us/wdsi/threats
[7] https://www.symantec.com/security-center/a-z/

convert an incident report written in human-readable format to a simple structure and then into query language without requiring any changes to the scripts.

To reason over statements expressed in these vocabularies, a human-readable format improves the cyber threat intelligence accessibility for security experts. A Python script is a good candidate which converts the natural language sentences into Cypher queries that are capable of quickly transforming into a knowledge graph. The script assists the integration of different schemas and formats for the presentation of the graph. The NLTK library, which offers a sentence tokenizer function to split the words based on the grammar, is used to extract the entities, attributes, and relationships between entities from each human-readable report. Then, the extracted terms are categorized into nouns, verbs and adjectives to describe the entities, relationships and properties. The groups are automatically converted into a simple structure that can be transformed into Cypher queries. The nodes and relationships can be associated with any number of attributes (hence referred to as properties) in the form of key-value pairs. This allows for data analysis modelling and querying.

A few simple rules are needed when simple structure sentences are created. For example:

1. {} is used to describe the property of the entity that could be used for searching
2. Relationship has to be one word, for example, in the expression: "Malware, is a, OSX-keylogger", "is a" is not correct, it shall be "is_a"
3. Shortcode [] is used to represent a whole entity, either the head or tail. For example, the investigation Alert=[Alert0]; later, I can use [Alert0] to refer to this node.

5.2.3 Implementation cycle

To fill the gaps in the current incident report, predefined scripts are used to query the knowledge graph. After a graph is constructed, subgraphs can be retrieved over Neo4j using a Python API. Py2neo is a Python Neo4j API with imperative and declarative features [236]. It is used to execute a more necessary and performant querying method. By traversing the graph, the queries look for terms that can be used to fill the text templates. I installed Neo4j APOC library extension to provide the analyst with more power and flexibility for crafting queries that can be successively constrained while maintaining a simple and readable syntax [237]. These extracted results from the queries are applied to generate an automatic story.

The query results are specifications for matching subgraph patterns to complete the incident report template of interest. Both templates and queries are modifiable and can be customised based on an organisation's preference and its local policy. To highlight the graph's implicit and explicit knowledge and convert it into a human-readable format, storytelling is used as a

knowledge representation method. The ability to automatically provide the story at different levels of detail enabling it to cater to various information needs and intended aims (i.e., low-level for the user at the Financial Department, high-level for top management).

The template and information to be filled out are updated when the audience group to receive the narrative report is chosen. In each level of the story, some pre-written templates are prepared to be completed with the enriched data. Readers' insights can be used to tailor each level of the story. For example, a top manager can utilise the top level of a story with no technical concerns, while an analyst can use a more thorough report to detect and trace a problem. The following are two examples of a story from an incident in the top and low levels:

At the Top level:
I got an alert on 11/02/2021 about an attack on server X, Sev1.edu.au,, which contained low severity financial data. There are no indicators that the server has been compromised. The server will be re-imaged to prevent it from being further compromised.

At the Lower level:
I got an alert, Alert0001, on 11/02/2021 5:05, because of the connection between local Machine, name: Sev1.edu.au, IPaddress: 10.233.62.247 with severity high(3/10) and Potential attacker, name:service.macinstallerinfo.com, IPaddress:104.239.223.14. Alert is a Malware, Name is Osx.Keylogger.Elite, Class Ref is snort, Type is CNC.More Info:Malware classifiedin snortrule, Classification is malware-CNC, Title is MALWARE-CNC Osx.Keylogger.Elite variant outbound connection. It found in ThreatExchange,Defintion is OSX.Keylogger is a spyware program for Mac OSX that records keystrokesmay take screenshots and may also send the information to a predetermined emailaddress, Reference is Symantec.local Info:Machine contains APPs, App1 is Schart, App2 is CoNsoleKit Microsoft Visual, App3 isC++ Machine administrated by Staff, Name is Alex, Mail is alex@vu.edu.au Machine located in Department, Name is FinancialUnit, Security rank is High(8/10), Unit is 1, Address is Block3 Level2.

A person who should be aware about the incident, can read the narrative report supplied by the knowledge graph and identify the fundamental source of errors and respond quickly by making key decisions. Similarly, a manager who isn't generally involved with technological concerns can quickly and accurately gather enough information about cyber occurrences.

5.2.4 Maintenance cycle

The generated narrative story, which is the graph accompaniment, is presented as a report, ticket, or post to the particular audiences (i.e., device administrator, risk owner, manager, analytical expert, and others.). The report's human-friendly nature contributes to a larger audience involvement in CSA (currently restricted to a security analyst). As an example scenario, the infected

device user receives the storytelling report with the graph and obtains insight into the cyber situation instantly, thus preventing further problem escalation. For instance, the sev1.edu.au was a primary server without confidential information. Although the server's security level was low because of the non-critical information stored on it, the owner used it as a backup device for confidential data without passing the role transition of the server to the security team. By reading the narrative report and the analytical graph, the owner understood the current situation of the server not being updated and the security professionals not being aware of its status. The user can contribute to complete the intelligent threat report and update the server's sensitivity.

5.3 Evaluation (Self-Evaluation)

In this section, I illustrate the applicability of the NVAM. For testing purposes, an incident alert example was randomly selected, which was raised from an external vendor's tool, Secureworks.

The example of the alert message produced by Secureworks is as follows:

`‘‘MALWARE-CNC Osx.Keylogger-Elite - 10.233.62.247 → 104.239.223.14 02/27/2019 5:05 PM’’`

The incident alert is correlated to the local and global knowledge bases to represent a possible threat scenario. To be comparable with the incident report, the extracted knowledge from knowledge bases is shown in a human-readable format in Table 5.1. As shown in Table 5.1, to represent the knowledge, the NameOfSource_Field (Value) is used as a structure, where: (1) *NameOfSource* is the name of the source from which the data is retrieved, (2) *Field* is used as a property which is searching from the source, and (3) *Value* presents the extracted value. i.e., Alert_SrcIP(10.233.62.247): Alert is a local source that source IP (SrcIP) is searched, and 10.233.62.247 is the extracted value for source IP.

The tool, Python scripts are used as convertors. The incident report is automatically converted into the simple structure format and then into Cypher queries. The alert_id is the key as an essential identification for each alert, and it used in corresponding nodes and relationships. A snapshot of the simple structure format transformed from the incident report is as follows.

This alert, code, Alert0001
Alert0001-{DateTime:“02/27/2019 5:05”, Message: “MALWARE-CNC Osx.Keylogger-Elite”, SrcIP:“10.233.62.247”, DesIP:“104.239.223.14”}, is , Alert=[*Alert*]
[*Alert*], Shows_Connection_from, Machine-{IP:“10.233.62.247”, Domain:“Sev1.edu.au”,Sensitively:“low(3/10)”, UnitID:1,Admin:12, APPslist:13}=[*machine*]
[*machine*], located_in, Department-{Unit:1,Name:“Financial_Unit”, Address:“Block3 Level2”, Security_rank:“High(8/10)”}

TABLE 5.1: Part of local and global knowledge associated with the incident alert

Local	Global
Alert_DateTime(02/27/2019 5:05)	Snort_Header(MALWARE-CNC Osx.Keylogger.Elite variant outbound connection)
Alert_Message(MALWARE-CNC Osx.Keylogger-Elite)	ThreatExchange_Definition (OSX.Keylogger is a spyware program for Mac OSX that records keystrokes, may take screenshots and may also send the information to a predetermined email address)
Alert_SrcIP(10.233.62.247)	ThreatExchange_Name(Syemantec site)
Host_SrcDomain(Sev1.edu.au)	WhoIs_DesDomain(service. macinstallerinfo.com)
Host_Seneitivity(Low(3/10)	ScanEngine_Name(ThreadMiner)
Host_Location(Fincanctial_Unit, Block3 Level2)	ScanEngine_URL (http://service.macinstallerinfo. com/ Mac/getInstaller-Specs/ ?&channel=3Db5002& info=3D238749466&encinfo= 3D1&)

By defining the sentence "This alert, code, Alert0001", a new alert_id as a new key is generated. Figure 5.2 shows a snapshot of the output generated from the Python script where human-readable sentences are converted into Cypher queries after transformation to a simple structure format. The output queries can be easily copy-pasted into Neo4j to generate the corresponding graph. Nodes, relationships, and properties are created based on the knowledge provided in human-readable sentences and corresponding queries. As Figure 5.2 shows, nodes and their properties are created, then the head and tail of a connection are defined, and then relationships are linked. As establishing nodes and links with Cypher is complicated for humans, the Python script translates the human-readable sentence to Cypher queries. Thus, a (human-readable format) bridge is provided for people to pass this stage.

The graph is generated directly from the Cypher queries. Figure 5.3 shows the generated graph for the incident alert with node labels and relationships type in the Neo4j. The corresponding knowledge related to the alert is shown by nodes and relationships in the graph. The local knowledge is presented through the "Machine" node and its dependencies, and the global knowledge is shown through the potential "Attacker" and "Malware" nodes and their associations. The nodes are shown in circles and they are classified into two groups: **sentence_entity** defined by the pink circles in the graph that bring a fact or isolated piece of knowledge. And the red circles are defined as **sentence_entity_shortform**, and deeper insights about them are provided in the graph (the nodes were converted by adding [] to represent a whole entity).

create(alert_2_shortform_0_:sentence_entity_shortform
{content:'Alert',content_lower:'alert',short_form:'Alert',alert_id: 'Alert0001' })
create(alert_2_shortform_1_:sentence_entity_shortform {content: 'Machine', content_lower:
'machine', IP:"10.233.62.247", Domain:"Sev1.edu.au",Sensitively:"low(3/10)", UnitID:1,
Admin:12, APPslist:13, short_form: 'machine', alert_id:'Alert0001'})
create(sentence_entity_alert_2head_2:sentenc_entity {content:'Alert0001',
content_lower:'alert0001',alert_id:'Alert0001', DateTime:"02/27/2019 5:05",
Message:"MALWARE-CNC Osx.Keylogger-Elite",SrcIP:"10.233.62.247",
DesIP:"104.239.223.14"})
create (sentence_entity_alert_2head_2)-[:is] → (alert_2_shortform_0_)
create (alert_2_shortform_0_)-[:shows_connection_from] → (alert_2_shortform_1_)
create(sentence_entity_alert_2tail_6:sentence_entity {
Content:'Department',content_lower:'department',
alert_id:'Alert0001',Unit:1,Name:"Financial_Unit", Address:"Block3 Level2",
Security_rank:"High(8/10)"})
create (alert_2_shortform_1_)-[:located_in] →(sentence_entity_alert_2tail_6)

FIGURE 5.2: Snapshot of the generated Cypher queries from the threat intelligence report (human-readable format)

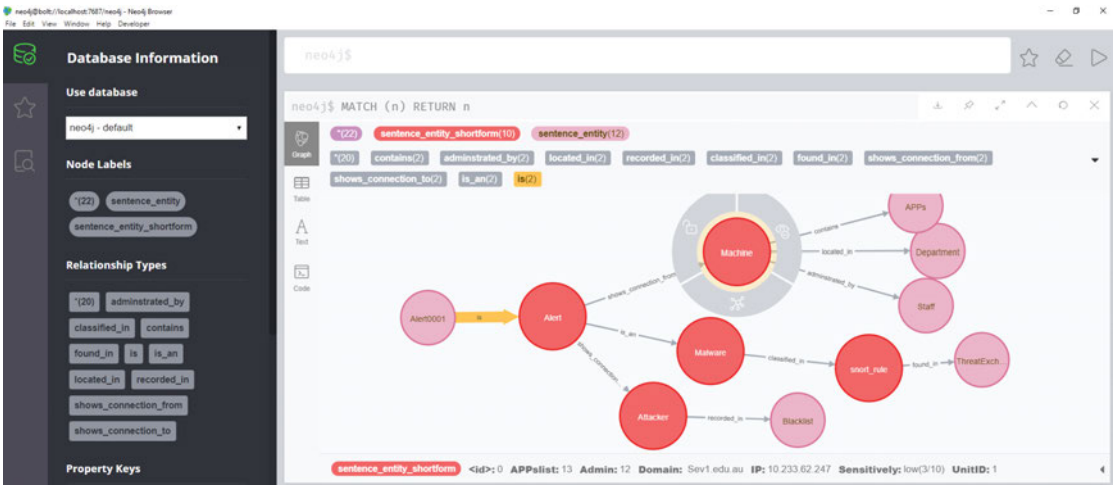


FIGURE 5.3: The generated graph in Neo4j from the Cypher queries (nodes are illustrated as circles and the relationships are shown as directed arrows)

Cypher is very similar to SQL, consisting of clauses, keywords and expressions like predicates and functions [232]. Each node represents an entity table, and its properties are the columns of the table. For instance, Machine as the highlighted node in Figure 5.3 containing properties like IP, Domain, Sensitively and others. It has relations with other nodes (Application, staff, and department).

For a Python graph database, Neo4j is installed on a system and then accessed via its binary and HTTP APIs, though the Neo4j Python driver, i.e., Database.driver [232]. Then, the graph for representing the knowledge of interest gives the analyst the power and flexibility for crafting queries. The query statements are easily and affordably defined and manipulated by users. A procedure on Cypher (APOC), an add-on library for Neo4j, is used for querying flexibly and

traversing the knowledge graph. The APOC library consists of many procedures to expand the subgraph nodes reachable from the start node following relationships to max-level adhering to the label filters. For example, in the query below, the collection of nodes in the subgraph and the collection of relationships between all subgraph nodes are returned. The given condition constrains the analytic results to focus on those that are linked from the node called Machine.

```
MATCH (p:sentence_entity_shortform {content:'Machine'+
",alert_id:"+str(AlertCode)+ "" })
CALL apoc.path.subgraphAll(p, {
relationshipFilter: '>',
minLevel: 0,
maxLevel: 5
})
YIELD nodes, relationships
RETURN nodes, relationships;
```

A node in the knowledge graph is labeled by its name, as a noun, and its properties are adjectives, and its acting by a verb shows a relationship. Retrieved subgraphs as results of the queries, bring the opportunities to generate sentences with nouns, adjectives and verbs automatically. For example, the paragraph *"Machine contains APPS, App1 is Schart, App2 is CoNsoleKit Microsoft Visual, App3 is C++. Machine administrated_by Staff, Name is Alex, Mail is alex@vu.edu.au. Machine located_in Department, Name is Financial_Unit, Security_rank is High(8/10), Unit is 1, Address is Block3 Level2."* is generated automatically after traversing the graph and returning the nodes, relationships and properties associated with the specific node, "Machine". The flexibility of creating a graph, and extracting information from it provides advantages to make an automatic report based on the preferences. A generated story from the incident alert is shown in Figure 5.4.

Figures 5.3 and 5.4 show how NVAM describes the incident raised by the alert in narrative and visualised model form. The knowledge graph and story together provide a comprehensive understanding of what has occurred both locally and globally. If the person receiving the alert is the owner or another participant with updated information about the alert, such as asset, location, severity, and so on, NVAM can provide a high level of ground knowledge to facilitate engagement in incident management. If the receiver, such as a manager, does not have up-to-date information, he or she can be updated immediately without overloading cognitive efforts due to the benefits of the human-friendly manner.

5.4 Summary

I have demonstrated that NVAM with knowledge graph and narrative report can assist security professionals to have a better perception of the elements of the environment by involving more

"I got an alert, Alert0001, Alert0001, on 02/27/2019 5:05, because of the connection between local Machine,name: Sev1.edu.au, IPaddress: 10.233.62.247 with severity low(3/10) and Potential attacker, name:service.macinstallerinfo.com, IPaddress: 104.239.223.14
Alert is a Malware, Name is Osx.Keylogger.Elite, ClassRef is snort, Type is CNC.
More Info:
Malware classified_in snort_rule, Classification is malware-CNC, Title is MALWARE-CNC Osx.Keylogger.Elite variant outbound connection. It found_in ThreatExchange, Defintion is OSX.Keylogger is a spyware program for Mac OSX that records keystrokes may take screenshots and may also send the information to a predetermined email address, Reference is Symantec.
local Info:
Machine contains APPs, App1 is Schart, App2 is CoNsoleKit Microsoft Visual, App3 is C++ Machine administrated_by Staff, Name is Alex, Mail is alex@vu.edu.au Machine located_in Department, Name is Financial_Unit, Security_rank is High(8/10), Unit is 1, Address is Block3 Level2
global Info:
Attacker recorded_in Blacklist, IP is 104.239.223.14, Reference is ThreadMiner, URL1 is `http://service.macinstallerinfo.com/Mac/getInstallerSpecs/?&channel=3Db5002&info=3D2387494666&encinfo=3D1&` *"*

FIGURE 5.4: A generated story from the alert

staff in the incident management that originated from the incident alerts. The conversion of human-readable sentences to query language and the interpretation of the graph's knowledge in a human-readable narrative report, aids comprehension and allows human involvement in the incident management process.

Through a shared understanding of CSA contextualisation, the knowledge graph is used in NVAM to visualise the relationships between the alert and relevant information from the local and global knowledge bases. Based on the knowledge graph, alerts are integrated into a more understandable form, such as summarised stories, to represent security events. As a result, alerts investigation is improved through better contextualisation of CSA, which aids in the establishment of an understandable common ground among humans. As a result, shared CSA is achieved by involving people with diverse knowledge and roles in the incident management process.

In this study, I only focused on malware taxonomy for approach demonstration due to current limitations. Nonetheless, the model is easily adapted to other types of incidents by supplying complementary sources from the local and global knowledge bases. Furthermore, because the enriched incident report for a security alert in a story design is not available, I was unable to conduct a direct comparison with the proposed storytelling model. Further still, the impact of narrative format comprehension and engagement is not directly measurable. Thus, the usability and suitability of NVAM were demonstrated in a real-world case study. This chapter results was published in the paper 4 of publications.

Chapter 6

Explainable Model to Interpret Money Transactions - Self and Shared Situation Awareness

The intelligent model of knowledge graph that was proposed in chapter 5 is provided for transaction visualisation for fraud detection, an important challenge in security research. As with the same incident management challenges, fraud detection methods need to be more transparent by explaining their results in more detail. Despite the fact that fraudulent practices are always evolving, investigating money laundering based on an explainable AI that uses graph analysis assists in comprehending schemes. By providing insights beyond current descriptive explanations of cognitive processes and requirements related to cybersecurity job performance, quantitative characteristics of the activities undertaken to solve the problem have been achieved [238]. Explainable model emphasised the role of humans in the loop by adding explanations that enhance cognitive insights. The main focus of this investigation is to support SOC and CSIRT teams in the incident management process, but the models can also use an accountable model to detect fraud. In fraud detection issues, transactions of accounts are investigated to identify fraudulent transactions [239].

Large transactions are a result of digital payments which have grown at an unprecedented rate in the last decade. Criminal schemes have quickly evolved to take advantage of the payments landscape. In real-world situations, fraudulent transactions resemble normal transactions, making it difficult to detect them immediately. While the number of transactions has increased, the task of auditing or tracing the money transferred based on records is becoming overwhelming. The use of visualisation techniques to make data (transactions) transparent is critical. Open data is demanded due to cooperation in the analysis process to gain insights, draw conclusions, and

ultimately make accurate decisions. This chapter aims to aid humans' natural ability to absorb a larger volume of information in the visual model during the visual analytic process.

A knowledge graph is a visual model that is used in the proposed Visualised Fraud Analytical Model (VFAM) to display the complex relationships between the money sender and receiver in a transaction. Based on the scenario-matching method, the developed network connecting accounts makes exploring account nuances simple and rapid. The scenario-matching method involves checking the model in real time and creating a matching trace for further investigation.

I used fraud scenarios to filter the network and identify suspect accounts linked to the money laundering phenomenon. Interactive querying was used to prune the graph based on the stated fraud scenarios and the visual model presenting the network made the data transparent and easily understood by humans. As a result, humans can successfully participate in the money laundering detection process given to comprehensive awareness provided through visual representations and interactive techniques.

6.1 Introduction

Money laundering is the process of making illegally earnt money appear to be clean, often through complex bank transfers and transactions. Concealing the origin of money earnt is often used in criminal enterprises. In other words, illegal earnings are integrated into a legal financial system by adding cover layers to hide the funds' true origins, and the funds are integrated into a legitimate financial system without raising suspicion from governments [112, 240]. The business community, the Australian Transaction Reports and Analysis Centre (AUSTRAC), law enforcement and regulatory agencies all play critical roles in preventing, detecting and disrupting money laundering, serious organised crime and terrorism financing. Money laundering is a large and complicated problem that is only getting worse, and current methods are failing to prevent it [241]. The main reason for this failure is an inability to distinguish between fraudulent behaviour (malicious transactions) and ordinary financial activity [242]. Despite significant government support, Australia is not doing enough to combat money laundering, placing it 48th out of 133 nations in the 2020 Financial Secrecy Index[1].

The complexity of fraud detection amongst massive amounts of transactions and rapidly changing fraudulent behaviour requires human involvement in the process. Analysts need more power to identify correlations of events in finical transactions. Furthermore, analyst reports from suspected transactions are critical pieces of information that aid activities in developing a clear intelligent picture and make society a safer and better place by preventing serious crime. More power is given to investigators who enable crowdsourcing of analysis by reducing complexity as a result

[1] https://fsi.taxjustice.net/PDF/Australia.pdf

of opening up data to analysis whichemphasises handling massive and dynamic sets of data by incorporating human judgment into the process. Human collaboration is easier via visual representations and interaction techniques in the analysis process [111]. Aside from the difficult nature of money laundering, visualisation techniques aid in the detection of financially significant fraud events as quickly as possible [114].

In this chapter, VFAM is proposed as an explainable intelligence to aid in the exploration, detection and analysis of fraudulent behaviour in bank transaction events. The visual model used in VFAM is a network of transactions with links between accounts. By integrating the interaction technique and scenario-matching approach, different types of fraud (money laundering) are queried in the graph then suspicious accounts are displayed.

The principal reason for building VFAM is to allow the exploration of information in a graph. The scenario-matching approach is used to prune the fraudulent network during the interplay. In money laundering detection, the target is unknown, meaning the suspicious accounts and fake funds are not explicitly revealed in the graph. The analyst must test various scenarios for the clustering outcome to be generated [116]. According to Ceneda et al., in [116], "visual analytics is typically applied in scenarios where complex data has to be analysed". Scenarios are hypothesised to assist analysts in driving to the destination, taking on each decision, changing paths if necessary, while still allowing the analyst the freedom to use other analytical tools to arrive earlier at the destination or exchange the result for a better understanding of situations [116]. The VFAM scenario-matching interactive graph-based analysis approach emphasises how simple it is to create various fraud scenarios and filter the visual information to detect and verify fraudulent activities.

Because the knowledge of transactions and relevant background are displayed in the knowledge graph, the query is a method for traversing the graph and checking the fraud scenarios. As mentioned in Chapter 5, Neo4j is a knowledge graph model that makes use of the strong query language Cypher. Some other works have used query language, particularly Cypher query [3, 243, 244], which I used to filter graphs to find expected patterns and results. Although they are not presenting a solution for fraud detection in financial data, their translated queries do not require humans to construct complicated queries by an interactive approach [243, 244]. These researchers employed the same query language provided in this chapter and took advantage of Python to make the queries more intelligible [3].

The remainder of this chapter starts with a brief introduction of the proposed intelligence, VFAM. Then, the dataset that is used for fraud detection is introduced. The chapter then elaborates on testing the VFAM based on the fraud scenarios to identify suspicious accounts through transactions. Finally, a discussion and conclusion are provided.

6.2 Visualised Fraud Analytical Model (VFAM)

To determine whether an account is in a fraudulent state, an analyst's verification is required to review transactions [245]. The initial aim of this chapter is to examine financial data using the explainable intelligence proposed primarily in incident management. I borrowed the idea and applied its model to disclose the fake transactions associated with accounts in a way that humans can understand and gain both self and shared SA throughout the money laundering detection process. VFAM was partially designed to take advantage of the human perception system, allowing analysts to derive insights from data more quickly, and assisting analysts in fraud detection.

In VFAM, a knowledge graph that can visualise transactions is presented. Graph-based analytics enable analysts to create a variety of queries and find the answers they are looking for by traversing the visual model, the graph. In other meaning, a shared understanding of financial transactions is developed and illustrated to engage analysts in fraud detection. The VFAS comprises four phases to assist an expert in support of analysis. As a result, full insights into the financial situation from the massive transaction can be achieved. The main development phases of VFAS are illustrated in Figure 6.1.

6.2.1 Data modelling phase

The synthesised knowledge is visualised using a graph-like structure to support the analytical reasoning [230]. Usually, a historical collection of domain-specific knowledge is designed and developed prior to constructing a knowledge graph [231]. Data modelling organises all of the data manipulated by the various functions involved in the system's design [246]. Domain-specific knowledge is investigated in order to recognise graph entities (nodes) and graph relationships (edges) in order to develop a data model. The domain that includes entities and relationships is where the money laundering occurs. The entities of domains can vary depending on the primary business process. In most domains, however, "Transactions" and "Accounts" play the most important roles because suspicious transactions are added or modified illegally. Other entities and relationships are added to the data model based on the information recorded in the transaction logs.

6.2.2 Visualisation modelling phase

Because explainability is so crucial in big data analysis [241], we've created a prototype for visualising such data as well as models to supplement human analysis. In the majority of the times, graphs (or networks) are used to represent relationships between entities. Transactions

are represented as nodes in a graph, with edges indicating the relationship between transactions. Through a well-designed graph, the analyst can quickly gain insight into the cyber situation. A graph consists of nodes, attributes and relationships. The knowledge entities as nodes and properties of entities as attributes must be recognised in the domain. How these entities and attributes are related to one another, and what entities are introduced at particular times should be captured [234].

Because massive amounts of data are processed on a daily basis, scalability and relationships are the most important factors to consider when designing a data model and selecting a graph database [247]. Scalability is the key to a successful analytical process; for instance, scaling up to one billion nodes, still maintaining top performance is a selection criterion in the financial area. I chose the Neo4j[2] to explain visual data modeling. Neo4J has a lot of benefits which makes it one of the most popular tools in this field. According to the results of a comparison, [248], Neo4J is one of the best options, with most features receiving a "Great-4" rating in comparison to other databases. As one of the best graph databases, it stand out for its features [248]. A Neo4j knowledge graph is a semantically enhanced awareness level of connected data that allows users to reason with it and rely on it for critical decisions. Its query language, Cypher, is an extremely expressive query language that was designed from the bottom up for humans to make graph queries. A flexible schema, Cypher (a powerful query language), a dashboard for planning, performing and analysing data, scalability and cloud readiness are the most important features [248]. According to the Neo4j website, they claimed:

> "The unmatched scalability of Neo4j lends itself to emerging AI and machine learning use cases, which require graphs to scale reliably across massive datasets to give learning applications context and to make AI more explainable."

6.2.3 Analysis and inference phases

Although fraud detection systems help filter through millions of transactions and generate fraud flags, final human assessment is still part of the process. The analyst analyses the transactions using the extracted information from the sender and receiver accounts by matching the suspected fraud scenarios as queries. Suspected transactions can be recognised by in-depth analysis of the sender and receiver accounts and their relationships.

Judgment and decision making is a psychological construct and considered to be CTA. CTA partially reflects the goals of awareness which occur when analysts comprehend the state of a transaction and predict the relationship between it and another. Numerous support tools come in a variety of implementations with machine learning algorithms that automatically

[2] https://neo4j.com/

analyse the massive transaction for potential fraudulent behaviors [77]—Still, their intelligence and presentation are not sufficiently transparent for full comprehension, necessitating human engagement in the judgement verification process. The visual scene provides a framework understanding based on memory and reasoning that reduces cognitive overload for humans [249]. It is much easier for human beings to find the correlations between transactions in the logs if they are modeled using a graph database, here it is Neo4j. The Neo4J interface running on a web browser facitates the process of analysis. The interface shows database information, node labels, relationship types and property keys, and on top I has text section where I can query data in Cypher language. Integration and visualisation in a light software, helps an analyst to have a better understanding of the elements of the environment.

An analyst investigating transactions should consider and be aware of the many situations that could arise in the transactions. Most of the time, there are certain fraud scenarios that are being searched for as a pattern. However, fraudulent activity evolves with time, necessitating greater investigation flexibility. It is for this reason that human involvement in the analytical process cannot be eliminated. Analysts must determine if each transaction is regular or atypical, and whether the claim is valid and reasonable. The money flow is either legal or suspicious loops are formed. They may put a part of their analysis to the test by using a Cypher query in the graph, and the results will show up as in the knowledge graph. The visual model relieves the analysts' cognitive loads and allows them to gain SA more quickly. A graph is a simple model that may be shared with others to incorporate their participation in the analysis process. The forensics team may put their hypotheses to the test based on the situation that has been replayed in their heads, and the findings are displayed in the graph.

I want a flexible graph that can be updated fast and cost-effectively by upgrading the knowledge that is used to evaluate various fraud scenarios. While the test may be presented in a human-friendly style, it also provides a fascinating method of analysis by evaluating various assumptions. I created fraud scenarios for recognising patterns of individual user fraudulent conduct. Fraud scenarios are a collection of user actions that point to the possibility of fraud. Computer intrusion situations are related to fraud scenarios, and the fraud detection system works similar to a signature-based intrusion detection system. In contrast to intrusion scenarios, fraud scenarios focus on high-level user transactions on financial data rather than computer system states and events. As a result, I used the Cypher query language to apply the fraud scenarios to the knowledge graph. A fraud scenario consists of a name, description and scenario rules. Scenario rules specify the order and timing of each entity's occurrence in relation to the fraud indicated in the cypher query.

Cypher meets this necessary expectation. Cypher organises searches in a linear fashion. This allows users to conceive of query processing as starting at the beginning of the query text and working their way to the finish in a linear fashion. The query as a whole is therefore made up of

these functions. This linear sequence of clauses is only recognised declaratively; implementations are allowed to reorder clause execution if it does not impact the query's semantics. The projections as RETURN are possible with the WITH clause, including aggregations. It also allows for filtering using projected fields. Pattern matching is the key idea of Cypher queries. In Cypher, the MATCH clause employs such a pattern and introduces new records in the query graph with bindings of the matched instances of the pattern.

The fraud scenarios are tested in the shape of human-readable queries. The graph for representing the knowledge of interest gives the analyst the power and flexibility for crafting queries [3]. The query statements are easily and affordably defined and manipulated by users [3]. As an add-on library for Neo4j, a Cypher procedure (APOC) is used for flexibly querying and traversing the knowledge graph. The APOC library consists of many procedures to expand the subgraph nodes reachable from the start node following relationships to max-level adhering to the label filters. A few adaptive fraud scenarios based on query passing mechanisms are proposed to investigate context-aware node representations and update edge relationships for improved answer inference. Our scenario-matching mechanism passed the queries and revealed the matched nodes and relationships, which is a dynamic graph inference process. The graph inference process is an iterative exploration, and the retrieved subgraphs are the results of the queries.

The analyst puts the analytical results (matched network) in a report within an embedded dashboard and charts are converted into a simple narrative report. The interpretation of query results by searching the graph is very easy and convenient. The report is shared informally as a filtered graph with text caption by the analyst. Much exploring of relevant information about a fraudulent s behavior assists the forensic investigator working with limited resources.

6.3 Evaluation

The VFAM detects fraudulent activity by matching fraud scenarios to transactions and presenting them in a visual model in order to acquire self-awareness and share SA. This section assesses the VFAM's capacity to detect collusive fraud based on the fraud scenario in the designed knowledge graph. The described fraud scenarios involve sender and receiver communications in transaction logs.

6.3.1 Dataset

The test data comprised collected transactions and account information that was utilised in the APAC hackathon 2020, Digitaldefence Hack[3]. The Digitaldefence Hack is a bi-annual global

[3]https://hackmakers.com

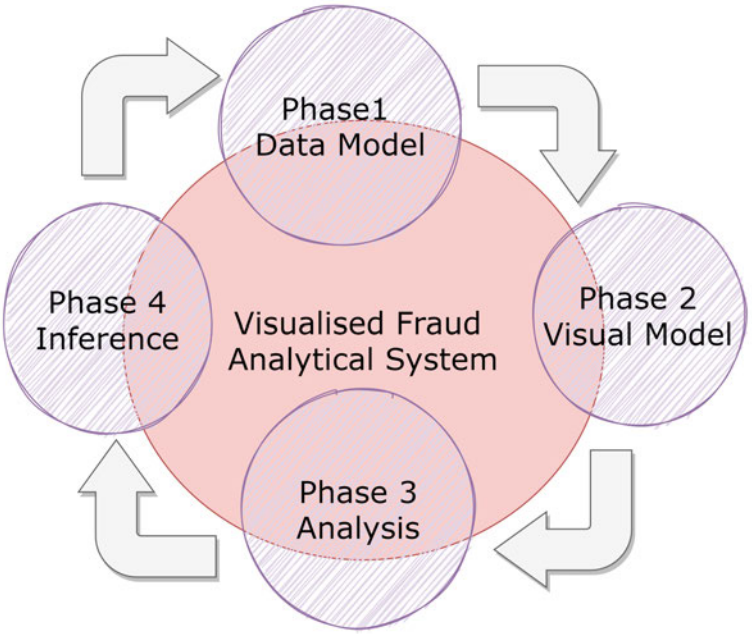


FIGURE 6.1: The main phases in the Visualised Fraud Analytical System for detecting and presenting fraudulent behaviours

hackathon addressing global challenges using best-practice cyber security and data science processes. With an estimated 2,000+ participants, 100+ mentors from 30+ countries and 10+ time zones working together to produce meaningful results. November 2020 had three areas: anomaly detection, deepfake analysis and cyber security hygiene.

Anomaly detection is the process of detecting unusual things or events in data sets that are out of the ordinary. Strange bank account activities, odd fraud activity, and company concerns can all be addressed using this method. Hackmaker's anomaly detection challenge requires competitors to create a system that can detect anomalies in a dataset. As a result, they make the dataset available to the competition. This dataset was distributed to only competitors to used for detecting frauds (money laundering) among massive transactions [4]. Through the request, the data set was made available for verification and review purposes.

The dataset was collected from a Charlies Crash Repairs (CCR) business. CCR is a well-established business located on the Sunshine Coast, north of Brisbane, Queensland, Australia. The business is strategically located close to all essential services, ensuring that the business is

[4]Datasets in various forms for the Anomaly Detection scenario was provided by oracle. However, after competition they are remove them from the servers.

- As a Zip file https://objectstorage.ap-sydney-1.oraclecloud.com/n/sdc90vkxb5rj/b/data/o/anomaly-detection%2Fanomaly-detection.zip
- as a tar.gz https://objectstorage.ap-sydney-1.oraclecloud.com/n/sdc90vkxb5rj/b/data/o/anomaly-detection%2Fanomaly-detection.tar.gz
- as individual files https://objectstorage.ap-sydney-1.oraclecloud.com/n/sdc90vkxb5rj/b/data/o/anomaly-detection%2Ffaccount.txt https://objectstorage.ap-sydney-1.oraclecloud.com/n/sdc90vkxb5rj/b/data/o/anomaly-detection%2Fftxn2.txt

highly accessible to passing traffic. The expected turnover for a business of this size is $2.1M. Instead, it is operating with a turnover of 14.3M. The operator's brother has a drug distribution network that needs to move approximately $1M a month. The transaction history and account information was able to ascertained for analysis, and some initial work on where the money is going has been done.

Because I used the core idea to propose the VFAM in the Digitaldefence Hack, I show the same result based on the database that hackmaker provided. Each challenge required a thorough understanding and examination of various data science and cyber security concepts. Each challenge could be completed by a group of two to eight people. The concept proposed in this chapter, which was based on the dataset, took first place in this big competition.

Other datasets can be beneficial in VFAM because it is not dependent on any particular properties. The data is displayed using a visual model in VFAM, and the scenario can be customised to detect fraud depending on the symptoms. For example, in one dataset, an individual or organisation with several accounts under multiple identities could be a sign of suspect activity. I employed the main scenario in this paper, which is known as fraudulent behaviour and may be used in most datasets. For instance, high volumes of in and out transactions being made in a short period of time. As a result, I placed a greater emphasis on the visual model and proposed VFAM to aid analysts, rather than the dataset, because any transactions or accounts, including fraudulent conduct, are sufficient to prove the VFAM. Because the database was large enough and contained accurate transaction data, it could imitate real-world transactions and meet the requirements for testing the proposed model. However, I will test our model with more datasets in future studies.

Table 6.1 Type of information (fields) provided in the dataset

6.3.2 Data model

To support explainability and retrieve full knowledge (semantic information) from massive transactions, 1052598 nodes from the dataset were recognised in the domain with the categories of entity labels and their properties as follows:

1. Customer [name(unique),DOB]
2. ACCOUNT[ACCID(unique)]
3. ACCtype[name (unique)]
4. ACCrisk[(name(unique)]
5. ACCcategory[(name(unique)]
6. Transaction[(TXN_ID(unique), TXTYPE, AMOUNT, TXDATE, REFERENCE)]

TABLE 6.1: The information provided in dataset for both accounts and transactions

	Field_Name	**Field_Type**	**Field_Description**
Transaction Information	TxId	Int	Unique Transaction Id (12 Digits)
	TxType	String	PAYMENT, TRANSFER, CHEQUE
	Amount	Float	Eg ., $50.23, $2398.43
	FromAcctId	Int	Unique Account ID (6 Digits)
	ToAcctId	Int	Unique Account ID (6 Digits)
	TxDate	DateTime	Transaction Timestamp
	Reference	Text	Reference Description I.e., “Dinner with Michele”, “Water Costs”, “Instalment”
	IsFraud	boolean	AutoGenerated isFraud=‘Y’, if Suspected based on the basic tool, However, no cases with the label “Y” (fraud) were discovered.
	IsFlagged	boolean	Flag for setting isFLAGGED=‘Y’, if flag is on for fraud detection tool
Account Information	AcctId	Int	Unique Acct Id (6 Digits)
	AcctType	String	CREDIT, BUSINESS , MAXI-SAVER, SAVINGS (based on bank’s category)
	AcctName	String	Account Name (Firstname and Surname)
	FirstName	String	First Name
	Surname	String	Surname
	AcctCreated	DateTime	Account Creation date and time
	PersonAge	DateTime	Date of Birth (Reveal age)
	AcctRisk	String	Predefined risk level (HIGH, MEDIUM, or LOW) based on account type and purpose)

* Reference Id for both ‘FromAcctId’ and ‘ToAcctId’

Furthermore, the dependencies between entities are defined as the following items (entity labels are shown in ‘’ and the relationship is shown by the “Rel” keyword).

- ‘Customer’ - Rel: has - ‘ACCOUNT’
- ‘ACCOUNT’ - Rel: type- ‘ACCtype’
- ‘ACCOUNT’ - Rel: category- ‘ACCcategory’
- ‘ACCOUNT’ - Rel: risk- ‘ACCrisk’
- ‘Transaction’ - Rel: from- ‘ACCOUNT’ (account is sender)
- ’Transaction’ - Rel: to- ‘ACCOUNT’ (account is receiver)

6.3.3 Visual model

The entities and relationships are depicted in the visual model. A graph made up of nodes, attributes and relationships is generated to help analysts visualise the dependencies between domain entities in a human-readable format.

Entities are nodes, and the properties of entities are known as attributes, which were identified during the data model phase. As a result, how these entities and attributes are related to one another, as well as which entities are introduced at specific times, are captured and illustrated [234]. I chose Neo4j to demonstrate graph data modelling. Neo4j is a navigational database that stores the connections between connected entities without the use of scanning indexes. Furthermore, Neo4j is open source, which improves the processing efficiency of massive data replication. Further still, its Cypher query language is a highly expressive query language designed from the ground up for humans to perform graph queries [233]. As a result, searching for and detecting anomalies is done using a graph-based analysis. A node in the knowledge graph is labeled with its name, as a noun, and its properties as adjectives, and its acting by a verb shows a relationship.

To model a Neo4j graph, the following concepts are used [233]:

- **Nodes**: Concepts or entities in the domain
- **Labels**: Tags for adding more meaning to nodes or adding constraints and indices
- **Relationships**: A directed, semantical connection between nodes to depict the relations between them
- **Properties**: Key-value pairs that depict more information about nodes and relationships.

In the graph, 1052598 nodes from the dataset were created, with 2105150 relations. Figure 6.2 depicts a snapshot of the nodes and their relationships. Figure 6.2 shows only 25 items including nodes and relations from the big graph. To support the analytical reasoning, the synthesised knowledge is visualised using a graph-like structure [230]. Before building the graph, a historical collection of domain-specific knowledge is typically designed and developed [231]. However, a predefined set of vocabularies and relationships is not required in Neo4j. The Neo4j graph is designed to be an adaptable middle-ware analytical tool that avoids a rigid set of rules. In other words, the graph depicts the initial dataset and the corresponding information from it in order to visualise normal and fraudulent transactions.

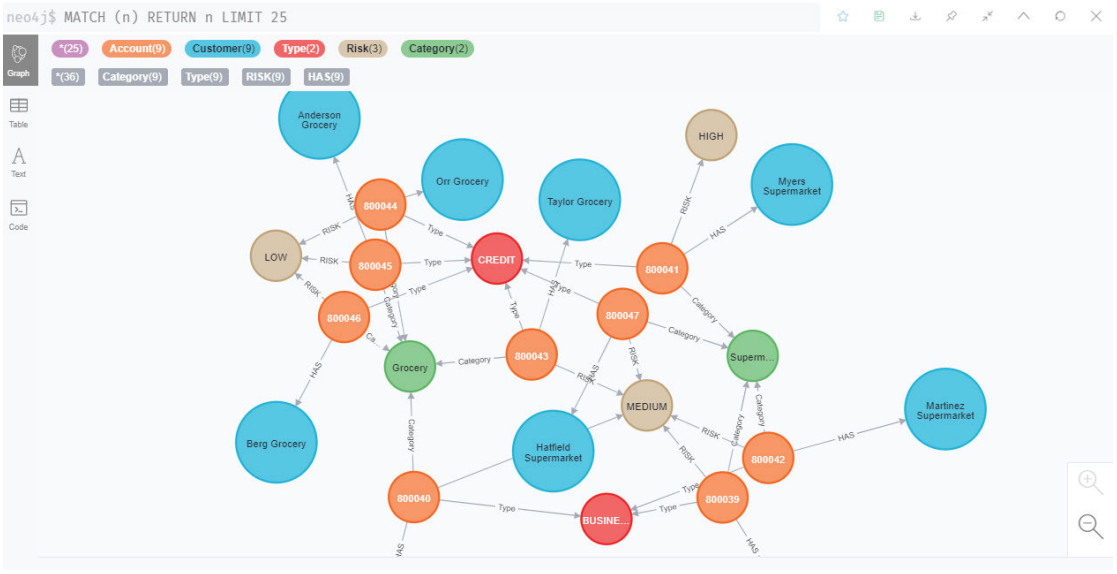


FIGURE 6.2: Snapshot of a generated graph in Neo4j that depicted nodes in a circle shape with their labels highlighted in different colours (i.e., blue for Customer entities, orange for Accounts entities) and relationships as directed edges with their tags

6.3.4 Analysis phase

During this phase, transactions are analysed using fraud scenarios and queries to investigate suspicious transactions by traversing the graph to detect fake transactions.

The Cypher query language is a well-known language for querying and updating property graph databases that was designed and implemented as part of the Neo4j graph [235]. The language includes linearly structured queries which enable analysts to think about query processing by imagining a scenario that starts at the beginning of the query text and progresses linearly to its end [235]. Retrieved subgraphs resulting from, brings opportunities to generate sentences with nouns, adjectives and verbs automatically.

The Cypher language is very like SQL[5] that returns the match results from the query. Some main clauses used to linear up the Cypher queries consist of the following:

- MATCH: a pattern that introduces by a record, such as (a)-[r]- >(b), where a is a node, r is a relation, and b is the end-node (directed arrow)
- RETURN: at the end of the query to return the results. It can be a record, node(s), relationship(s), paths or combinations
- WHERE: a clause to add a condition about the pattern, filtering based on the attributes.

Fraud scenarios are essential when writing queries. The scenarios are used to assess how the Cypher query language is described in order to develop and display some matched fraud

[5]Structured Query Language.

transactions in the graph. The fraud scenarios can reflect the analyst's thoughts as he or she investigates money laundering.

By using the scenario-matching approach and querying to display suspicious transactions and accounts by traversing the graph, the degree to which suspicious accounts in transactions are evaluated can be determined. When the results reveal misuse behaviour, all accounts in the results are flagged for further investigation, which includes contacting their owners for additional information. Three fraud scenarios are considered in VFAM to test and analyse transactions using the scenario-matching approach in Cypher query language. Each case of fraud discovered by the VFAM was investigated to ensure its accuracy. Records were added and removed as a supplementary check for each instance, and the scenario identification procedure was re-run. This allowed us to confirm that adding relevant transactions had the anticipated impact of generating a match where none previously existed, and that removing records had the expected effect of producing no match where one previously existed.

6.3.4.1 Fraud Scenarios

The following are the fraud scenarios consisting of a name, description and scenario rule which are defined in the Cypher query that were utilised as a pattern to see whether any fraudulent conduct was detected among transactions.

Scenario 1. Large transactions with $>$ threshold

- **Description** This scenario is looking for fraudulent behaviour that aims to conceal money in transactions. The scenario is created to find large sums of money with a high level of suspicion. Filtering large amounts of transferring is more important than filtering small amounts of transferring when analysing a transaction and recognising it has a fake amount (a significant amount) in a payment. The Transaction references are a description in the dataset that can be used to determine whether or not a transaction is reasonable. When looking for suspicious payments, comparing the transaction's reference details can help. Unusual spending of a specific account is a sign of unusual behaviour and possibly fraud [110]. For example, some transactions in the dataset show that more than $10,000 was transferred between two suspected accounts to claim a dinner payment. Suspicious payments are revealed by filtering different amounts and checking references. A hypothesis for transferring large amounts of money from/to suspected accounts is considered in this scenario. As a result, a threshold should be established which is gradually reduced (from the maximum amount) until the results reveal that an excessive amount was transferred from/to an account.

- **Cypher query**
  Match p=(a)-[R:To | From]- > (b) WHERE a.Amount- >TD Return p

  By defining a record called 'p' that has the connection between 'a' and 'b' as entities with a "To" or "From" relationship, this Cypher query looks for a matching pattern. A "WHERE" clause based on a transaction's property filters the results (Amount).

Scenario 2. Self to self transactions

- **Description** This scenario aims to detect fraudulent behaviour where a person attempts to conduct transactions without transferring funds. It happened when that exact amount of money was transferred from one account to another (self to self) in order to create a fake transaction history. Transferring money between a sender and a receiver (different accounts) is a real transaction, but in this case, both are the same. A hypothesis for transferring money from/to a suspected account (a fake circle) is examined in this scenario.

- **Cypher**
  Match n=(a:ACCOUNT)-[r:TO—From]-(b:Transaction) WHERE b.ToACCID=b.FromACCID Return n

  By defining a record called 'n' that has a connection between 'a' as Account entity "To" or "From" 'b' as Transaction entity, the query looks for a matching pattern. A "WHERE" clause is used to filter the results to ensure that the sender and receiver have the same ID.

Scenario 3. Circular transactions within the same day

- **Description** This scenario aims to detect fraudulent behaviour in which a person attempts to make fake transactions by transferring funds from one account to another using different references. The money is usually returned to the original account or another linked account. It is a transactional semi-circle. It occurred when the exact amount of money was transferred between several accounts and was then returned to the known account.

- **Cypher**
  Match (c:Transaction)-[r:To]-(a:Account) Match (d:Transaction)-[p:From]-(b:Account) where c.FromACCID <> d.ToACCID return a,b,c,d

  This query searches for a semi-circular path between accounts by using a two-match pattern and a condition to filter transactions that contain a link between two different accounts, not the same account. The symbol " > " represents the not equal clause, which reveals only the difference between the transaction entities' connection's sender and receiver. The unequal condition in the "WHEN" clause searches for a path (a transaction) from the "a"

to "c" entities and from the "d" entity, where the "c" and "d" accounts are not the same. As a result, the transactions' links between accounts may appear.

6.3.5 Inference phase

Finding correspondences between transaction items is critical in a variety of application contexts. In the fraud scenarios, a wide range of data is sought. In this case, matches are taken into account because they provide analysts with additional information. The VFAM enables the creation and execution of scenarios, allowing for specialised evaluations that yield comprehensive results. The results of the scenario-matching in the following subsections reveal several facets of deception with the intent to defraud (steal money in the transactions).

6.3.5.1 Scenario 1. Large transactions with > threshold

Figure 6.3 shows an example of fake transactions between two suspected accounts in which large sums of money were transferred (more than 90K). The nodes in Neo4j can be expressed by choosing an attribute. The "amount" is shown as a selected attribute in Figure 6.3 , but it can be changed to "reference" to help with the analysis. As Figure 6.3 illustrates, many large transactions are from/to accounts with IDs "800373" and "800377" on the same day. Therefore, based on the first scenario, they are listed as suspicious accounts.

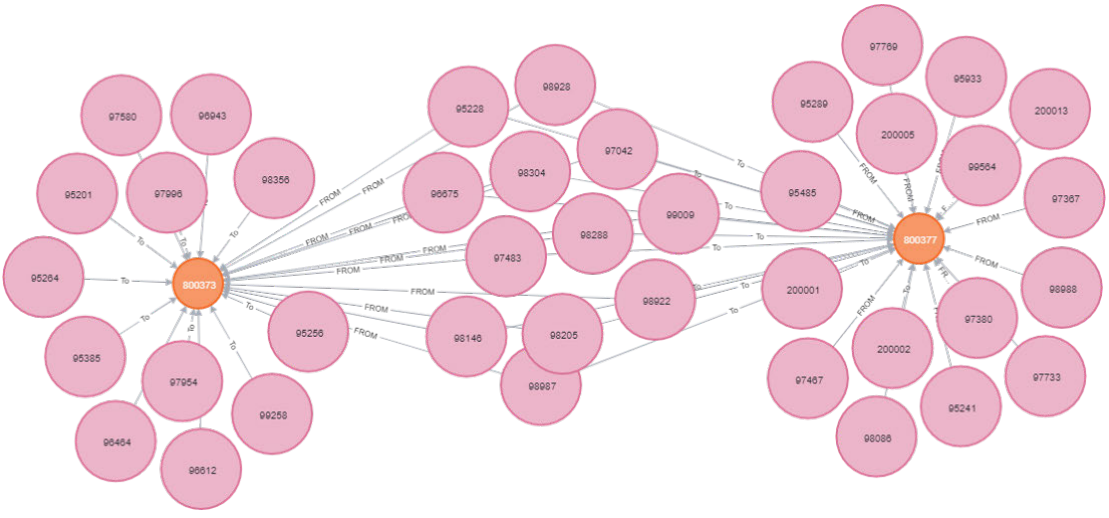


FIGURE 6.3: An illustration of fraud detection using the first scenario in which large sums of money are transferred over a short period of time. The orange circles represent the account ID, while the pink circles represent the transaction amounts to and from the account

6.3.5.2 Scenario 2. Self to self transactions

In the large graph, there were 300 nodes and 316 transactions, with 142 suspicious accounts in 158 transactions. Some accounts are flagged as suspicious if they have multiple self-to-self transactions. Figure 6.4 shows a snapshot of the detected fraudulent behaviours based on the scenario. As Figure 6.4 shows, many fake transactions were made without real transferring to two separate accounts only to build the transaction history.

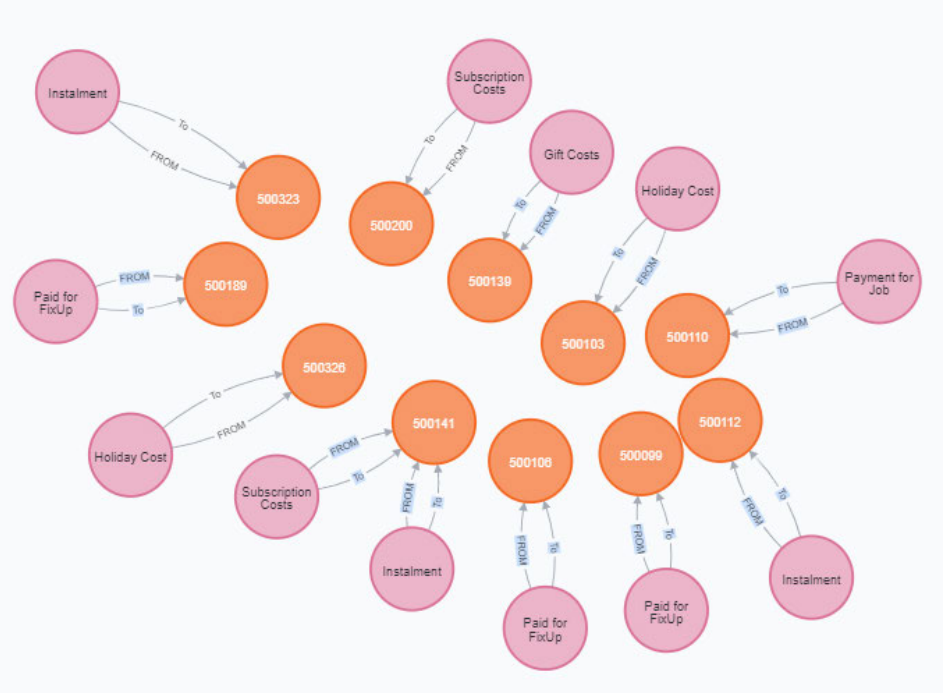


FIGURE 6.4: A snapshot of fraud detection using the second scenario in which money is returned to the same sender account after not being transferred between two separate accounts. The account ID is shown in orange circles, and the transaction reference is shown in pink circles as a selected attribute

6.3.5.3 Scenario 3. Circular transactions within the same day

The results found 300 nodes and 323 relationships involving 17 accounts and 283 transactions. Figure 6.5 shows a snapshot of the results, demonstrating how the circularity of the transaction can be seen by performing a large number of transactions. Many fake transactions were made to hide the money that had been sent to a suspicious account, as shown in Figure 6.5. The suspicious list includes all of the middle accounts.

6.3.6 Comparison

There are a variety of visualisation technologies that can assist with big data analysis, both with and without human involvement elements. However, the analyst must head up his or her

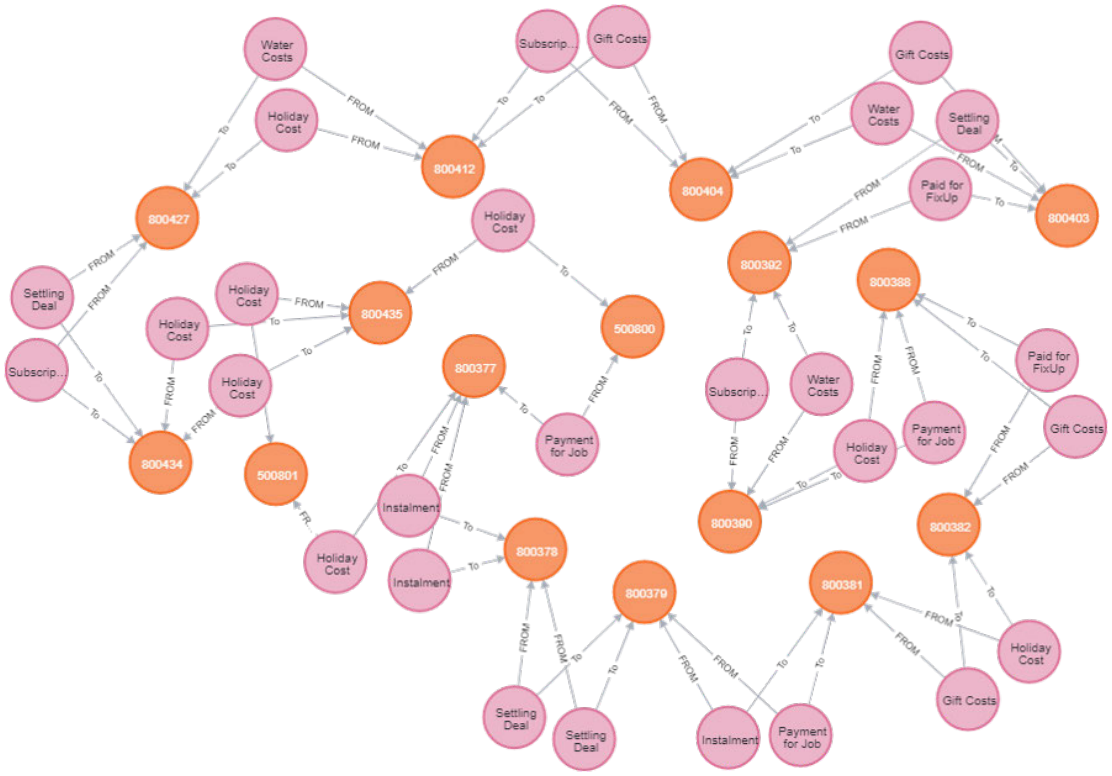


FIGURE 6.5: The third scenario is used to demonstrate how fraud can be detected. On the same day, money is transferred between multiple accounts using different references in a clockwise direction. The account ID is shown in orange circles, while the transaction reference is shown in pink circles as a chosen attribute

hypothesis in order to detect evidence and dig deeply using technologies, which are rarely focused in the existing technologies. The graph, as a visual model produced by VFAM and annotated in a human-friendly manner is intended to assist humans, unlike other tools that just display data. As a result, the analyst may lead his or her hypothesis without requiring technological assistance, and the network of data has less complexity to comprehend due to its human-friendly annotation. This implies that VFAM is considerably easier for analysts to utilise because the nodes and linkages are explained in a human-readable way. However, the transparent tools with comparable visual style that VFAM used to enable comparison are no longer available. However, I constructed and compared another visual tools, Graphistry[6] and Tableau software[7], to emphasise the benefits and transparency that VFAM provides for humans.

Graphistry delivers a human interface in the age of massive and complex data. It turns data into interactive, visual enquiry maps that are tailored to analysts' needs based on product claims. It quickly reveals links between events and entities without the need for searches or data manipulation. It can also capture all data without causing scalability issues, and then pivot on the fly to follow further investigations. NEO4J has the same capabilities, with the exception that analysts can use Cypher to filter their data by query in a human-readable format. All

[6]https://www.graphistry.com/
[7]https://www.tableau.com

data from large tables is converted to nodes and relationships using Graphistry. If the level of explainability could be measured, it would be clear that VFAM's graph is far more understandable than Graphistry's. The terms "comprehensible" and "explainable" are mutually exclusive. The analyst's awareness of how to investigate more fraudulent behaviour increased as a result of his or her perceptions of the graph on its own.

The outcome of Graphistry's first scenario is depicted in Figure 6.6. As shown in Figure 6.6, a large number of nodes and relationships were quickly created. The Graphistry data visualises the same dataset, but it is extremely difficult and unclear how an analyst should begin the investigation when compared to the graph created by VFAM, Figure 6.3. With Graphistry, there is no human-readable annotation to guide an analyst to what the graph revealed. For example, which nodes in Graphistry show suspect accounts? Answering these types of questions, which are necessary for generating hypotheses for further analysis, is not straightforward at all, especially when there are a large number of nodes.

VFAM, on the other hand, does not have such issues. Account nodes in Neo4J are coloured differently than transaction nodes, and they have additional data labelled on them; but, in Graphistry, there is no distinction between account and transaction, and the annotations are displayed in a separate table that is not human-readable and may cause more confusion. All data is visualised in a graph, or a mesh to use a more technical term, as illustrated in Figure 6.7, with the option to provide the dataset or more details in tables. These properties, on the other hand, are not human-readable and are useless for analysts trying to comprehend the meaning of the graph quickly.

As a result, if analysts don't have a thorough understanding of the accounts and transactions in this case, conducting an investigation will be more difficult. Because Graphistry's visual model does not alert the analyst, he or she may prefer to use other techniques to investigate more complicated cases. The graph and visual model in VFAM, on the other hand, are matched to the analyst's thinking style, giving him or her the best chance to dig deeper.

FIGURE 6.6: Graphistry depicts data set transactions and their links where the amount exceeds the threshold (the first scenario)

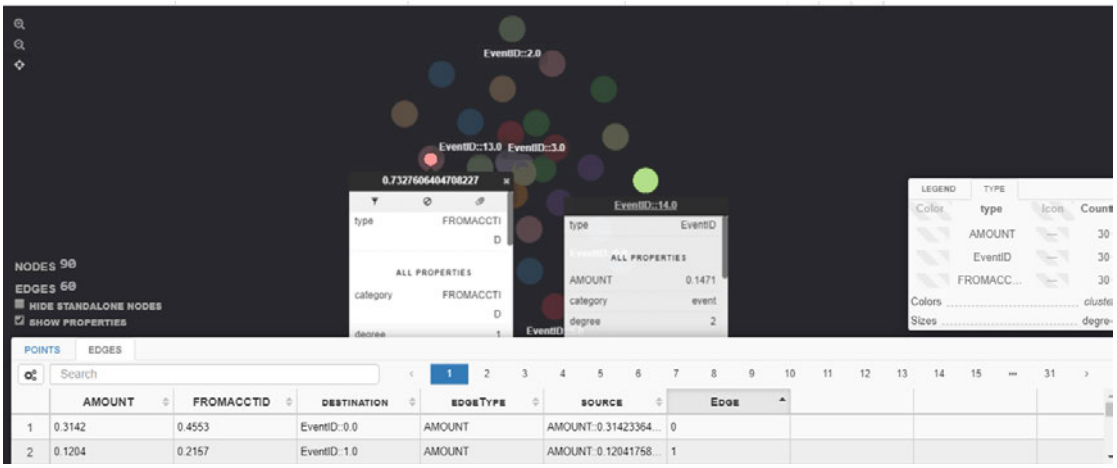


FIGURE 6.7: Graphistry features include bringing the dataset and detailed information into separate tables, as well as displaying the graph

The VFAM results are also compared to Tableau, a powerful non-graph interactive data visualisation tool. Tableau includes a dashboard with features found in datasets, allowing analysts to experiment with them and learn more about them. As can be seen in Figure 6.8 the transactions are displayed on bar charts with selected attributes in Tableau. Tableau's dashboard can be helpful when an analyst is looking for a case. To put it another way, the chart is less effective than the graph at bringing data and relationships together in one place and assisting with data digestion, particularly when fraudulent transactions are hidden amongst regular transactions. This means that when an analyst receives an alert about suspicious accounts from a graph, Tableau can help him or her go to the destination and double-check his or her theory with further evidence. However, Tableau is unable to provide first-insight to analysts who need to save as much time as possible when detecting fraudulent behaviour.

Finally, compared to Neo4J, VFAM, the existing visualisation model, Graphistry and Tableau in this example, is a poor choice, because present models do not seek to assist users in the analytical process and deduct their cognition efforts. In reality, while both Graphistry and Tableau are useful for quickly transforming and visualising data, they are less successful at lowering human cognition efforts.

6.4 Summary

I describe three fraud scenarios, including the technique of identifying and detecting their occurrence in the knowledge graph using the Cypher query, which I created in a human-friendly format. The intelligence model for fraud detection was created by combining the visualisation model derived from a Neo4J graph with its scenario matching supplied in a Cypher query. The

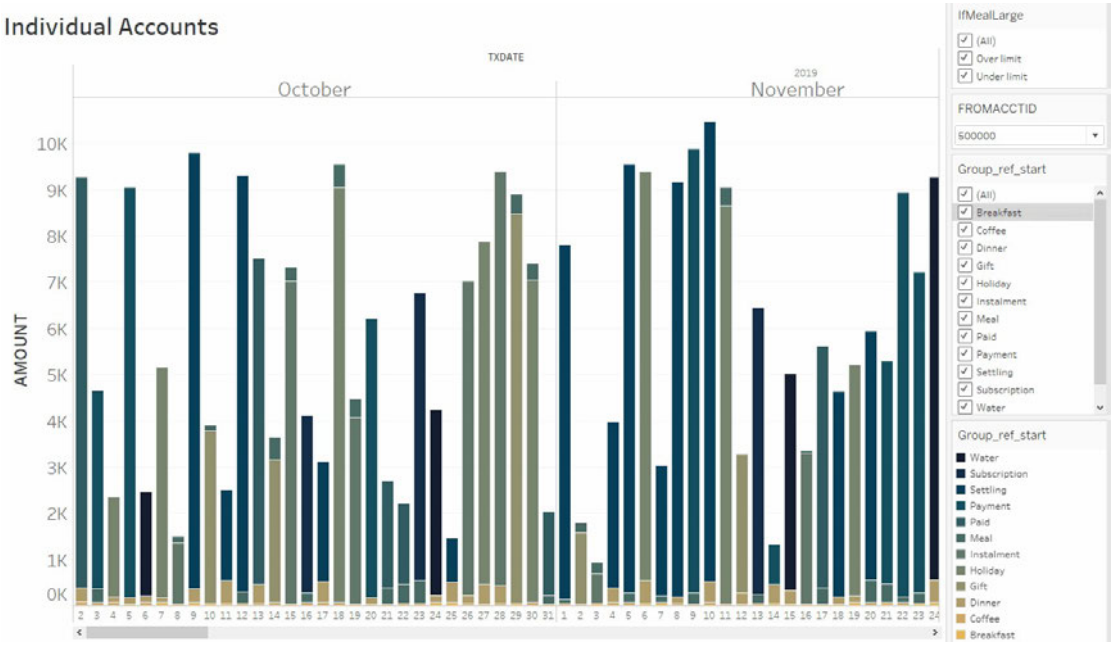


FIGURE 6.8: Snapshot of the Tableau Software with features from the dataset

findings of matching scenarios demonstrate the suspicious phenomenon of money laundering (suspicious accounts).

Suspicious accounts in transactions are an initiate phenomenon that occurs from time to time. The way two accounts are linked, on the other hand, is based on their suspicious behaviour, implying that all of the accounts involved in a transaction are fake. For example, money can be transferred between two accounts with different reference descriptions, but the sender and receiver accounts both belong to the same person. Two transactions have been identified as suspicious as a result of this enquiry. Other accounts could be selected at any time during the transfer process to conceal the fraudulent activity. The time and links between the money that was transferred and the fraudulent accounts that were queried in the scenario can be matched. Discovering the accounts' involvement in these transactions using VFAM has more benefits for the analyst, who can more easily comprehend the fraudulent activities.

Since fraudulent behaviors are always changing, and are automated, so supervised/unsupervised detection methods are only temporally pattern-based solutions and tend to fail with time [114]. Due to changing current conditions, VFAM for correlation analysis and pattern detection based on the scenario-matching approach, is more robust, as it integrates human perception into the detection process, which is precise and flexible and suited to spotting many different suspicious events. The visualisation technique, a graph analysis used in VFAM, is a dynamic drawing making links represented by different layouts, facilitating the understanding of systems and changing them from time to time. VFAM achieves the objective by providing a powerful visual analytical system to analyse massive numbers of transactions in the financial market.

I evaluated the data from a dataset with fraud scenarios to demonstrate the approach. The model be easily customised to different datasets and circumstances by providing human-readable Cypher queries. I was also unable to perform a direct comparison with the proposed VFAM because the result in this visual design is not available from other works. However, I did demonstrate other visualisations, including Tableau and Graphistry, for the same data to highlight the VFAM's distinct ability to provide a quick insight by interpreting financial entities in the visual model. The impact of comprehending the graph and gaining insight into data is difficult to quantify. As a result, fraud scenarios were created to demonstrate the usage and appropriateness of VFAM.

Fraud scenarios are patterns that may be found in data to determine whether or not a hypothesis regarding fraudulent conduct based on the pattern is valid. When researching a fraudulent behaviour, analysts normally have a lot of possibilities. I allowed individuals to simply verify their assumptions and obtain an immediate understanding of the outcomes by discussing the data and illustrating it with VFAM. Because the nodes, their labels, relations and attributes are shown in a human-readable style, the queries are simple to follow in a linear order. In comparison to other visualisation models used for fraud detection, VFAM focuses on a human-readable style for analysing data, which is done through both self and shared SA.

Chapter 7

Discussion and Conclusion

This research project sought to answer the following research questions: How can cybersecurity incident response teams or SOC have comprehension awareness in incident management using explainable intelligence? How can organisation members have a shared cybersecurity awareness through the process of incident management to make sure all relevant stakeholders are involved? Answering these research questions is important for both academia and industry. The incident management process includes many phases: plan and prepare, detect and report, assessment and analysis, response and learn. Nowadays, organisations face challenges to handle both internal and external cybersecurity threats that are more complicated and frequent than before.

Sabotage, embezzlement, theft, fraud and industrial espionage are counted as examples of threats' exploration. Organisation have to change their traditional cybersecurity incident management approach in the current complex and dynamic cybersecurity environment. They need to improve the agility of their cybersecurity incident response process by increasing attention to human factors. Security tools, however, have failed to defend organisations from the threats. In the design, maintenance and control of human-machine systems, human factors have become a focal point. Machine capabilities have improved so much since the beginning of industrialisation that human regulation of processes has changed from basic (with mechanisation) to cognitive (with computerisation), and even emotional (with semi/full automation).

Cybersecurity analysts require tools that provide decision supports in the current scattered and distributed systems. They need analytical support tools that can help them understand the environment's elements to comprehend the state of an incident, assist in prediction, detection, and respond to evolving cybersecurity threats, and justify their decisions. Making current security tools' data and alerts more understandable to humans by pulling data from multiple sources and building textual explanations out of them is the novel approach in this study that is used to reduce the experts' overall cognitive loads. The importance of augmenting human-cognition through human-automation interaction with technologies, such as security devices in complex

manufacturing and operational environments, reveals a variety of opportunities for developing novel methods for enhancing affective cognition and perception learning.

Collecting information about various events and detected vulnerabilities and reporting on them plays vital role in the detection phase. Reporting allows for detection through automatic tools, internal organisation collaboration and manual reporting. If all incidents with sufficient details were reported in documents, it will improve trend analysis, identify direct case of incident, identify severity measure, and allow good communication with stakeholders and suppliers.

There is a lack of formal standards for what information sources should be used and how this information should be put together to make a complete incident report. The main reasons impacting the organisation's effectiveness in handling to incidents include shortage of staff and skills, a lack of integration with other security and monitoring tools, and a lack of visibility into insider behaviors. In summary, the lack of a comprehensive and complete incident reporting to combine inside and outside visibility is the main weakness in detecting and responding to security incidents.

Complete incident reports analyse the voluminous monitored logs and data about the incident, and infer important features of the situation related to networks, threats and vulnerabilities that provide a solid understanding of cyber situations to help and support decision-makers and analysts. In order to have a complete report and perform this inference, which makes its output useful to human users, an incident report needs to have its outputs represented in a storytelling format with the use of vocabularies of well-specified terms and their relations from local and global knowledge bases.

I has developed and tested explainable intelligence models in the research of multiple, diverse case studies and also with a survey method. In explainable intelligence models, logs and cyber alerts with associated information related to an attack path or an event worth further investigation, are considered inputs. Outputs form a conceptual map of the incident that has applied awareness mechanisms of both self and shared situational awareness. Their reporting procedures, particularly the storytelling approach, and their visualisation models, a knowledge graph, are well-generated to communicate to stakeholders and assist in knowledge-based cognitive analysis.

Cyber analysts must not only maintain awareness of the current incident, they may also be affecting or affected by the local and global knowledge bases and cyber communities, and learn from incidents. An analyst's operation in performing a data triage task can be optimized by reducing its cognition overload via a complete incident report, which causes an agile incident response.

The reports highlighted important incident factors, in particular the actor (who), riskiness (what), time (when), location (where) of an incident and evidence (how), which had not previously been presented completely in incident reports or prominent in the literature. The interpretation of data

conversion for knowledge is accomplished by narrative incident reports that determined a level of CSA and the security analyst's tasks' automation development. Furthermore, open data by visualisation techniques is demanded due to cooperation in the analysis process by humans to gain insight, draw conclusions and, ultimately, make accurate decisions.

The chapter started by discussing the key findings of the study in relation to the existing cyber analytic capabilities, context-aware systems, human-centred cognition and cybersecurity incident response literature. The connection of this study to narrative representation, particularly story-telling, and its use to increase human cognition was explained. The limitation and future research directions are outlined at the end of this chapter.

7.1 Summary of Contributions

Modern companies operate in a diverse and complex cyber threat world. Organisations must respond to changes in their cyber threat environment on a regular basis in order to remain competitive. As cyber-attacks become more complicated, it is critical that organisations' incident response teams be able to track, investigate, report, react and, eventually, strengthen their overall organisational protection by implementing strong preventative and proactive response strategies.

Being agile in emergency management is one of the important characteristics of a responsive incident response approach, and a crucial aspect of it is getting the right knowledge at the right time to react appropriately. However, applying such a strategy is difficult and daunting for incident response teams. As a result, the primary aim of this study was to learn more about how cybersecurity incident response teams use explainable intelligence to improve effectiveness in their incident response method.

This study explores the use of explainable intelligence in cybersecurity incident management and cybercriminal-money laundering. Explainable intelligence models were used to develop human-readable, real-time, analytics-enabled dynamic capabilities and dynamic incident response strategies from the cyber data. The primary data in the cybersecurity incident management process from the cybersecurity incident response unit at educational institution were Windows logs and IDS alerts. The explainable intelligence models focused on the features of cybersecurity incident reports and the cognitive effort that security analysts must develop to interpret them. With a focus on a very broad problem which aims to identify a high level of intelligence, from the black box to transparent work, proactively detect abnormalities and address security challenges effectively.

This work contributes to the body of knowledge, including natural language processing, to enrich reports of cybersecurity incidents from both local and global knowledge bases. With enriching incident reports, cybersecurity analysts try to improve their understanding (self-awareness) during the incident handling processes. By creating a knowledge graph alongside narrative reports,

security professionals can gain a better understanding of the environment's elements by combining narrative and visualisation techniques. Because this combination is human-friendly, it is simple to understand, and involving more staff and gaining knowledge from the security team and other departments in incident management results in the creation of an shared SA.. A similar graph as a visual model was used in the explainable intelligence model to present the complex relationships between the money sender and receiver in a transaction to assist in exploring, detecting and analysing fraudulent behaviors. The network generated between accounts explores the nuances of accounts based on the scenario-matching approach quickly and easily.

The proposed explainable intelligence models include Log-Chain-Driven Storytelling model, Log-Driven Storytelling Model (LDSM), Narrative Analytics Assisted System (NAAS), and Visualised Fraud Analytical System (VFAS). In the current shape of study, there are strong elements to perform a fair evaluation of the intelligent models. Experimental results from empirical evaluations in various case studies compared to Secureworks' reports for the same incidents validate the main objective related to reducing the cognitive effort of cybersecurity analysts. A survey study was used to enhance the external validity/generalizability of findings in the form of security reports. Both professionals and students were recruited to participate in the surveys to validate the LDSM in terms of the completeness and comprehension of the generated storytelling incident reports compared to Secureworks' reports for the same incidents.

7.2 Key Insights

The state of cybersecurity is quickly changing. The ways in which focal organisations build, collect, store and exchange data accounts for the majority of changes in the cybersecurity landscape. Because this information is exposed across numerous targets including networks, software, data and physical components, hackers and criminals find it appealing. As a consequence, ongoing tracking of cybersecurity incidents around these targets is essential for business continuity, which is why cybersecurity incident response units integrate real-time analytics into their cybersecurity incident response capabilities.

While monitoring systems aid in the filtering of millions of logged events and the generation of security alerts, the final human review is still needed. Thousands of possible security breaches received from various surveillance services place a major strain on cybersecurity team personnel. Providing a human-friendly style of reports produced from such alerts, as well as the extensive domain expertise needed to understand the occurrence of raised alerts, the cyber security and response team is able to evaluate incidents as they occur and decide whether the event is actually an incident.

The right approach is then strongly dependent on the long-term expertise of researchers in the area of cyber threat management. Via an innovative approach, this research seeks to take the first step towards automating human-centric data triage. The concern is whether analysts' information requirements information can be elicited from their activities while conducting data triage functions and whether that intelligence can then be used to support analysts and reduce their workload.

To aid cybersecurity researchers, extensive analysis has been undertaken. Alert correlation is a priority sector in automatic data triage. To associate notifications, alert correlation techniques employ heuristic rules (a basic type of automation). However, this is constrained by the fact that it only analyses one data source (i.e., IDS alerts), while researchers in most situations would do cross-data-source analysis. Alert association analysis has since been combined with other types of analysis, but the approach remains based on heuristic rules. SIEM applications have been focused on security event correlations across various data sources, motivated by the advantages of cross-data-source research.

While SIEM systems make considerable progress in terms of generating more efficient data triage automatons, in order to produce a large number of complicated rules (complicated data triage automatons), specialist analysts must devote significant manual effort to creating the data triage automatons. As a result, it is necessary to make SIEM systems more commercially viable whilst still reducing the pressure on experts. This current human-centric viewpoint separates our practice from past attempts to create data triage automatons to assist experts. SIEM system outputs of fully explainable event reports boost data triage automation's robust analytics capabilities. The following are the main results of the proposed explainable robust analytics capability in cybersecurity incident response:

- Despite the difficulty, activity traces can be examined in a largely automated manner
- Despite the difficulties, the results of analysing the traces can be automatically converted into a state machine which can help generate automated threat alerts and prompt behaviour based on business rules
- Knowledge digestion and comprehension requires the least amount of cognitive effort
- It is possible to accomplish both on-demand and continuous real-time analytics that have their own time and position within the cybersecurity incident response process
- It assists businesses in integrating, expanding and reconfiguring their cybersecurity capabilities, expertise and practical competencies
- Organisations are increasingly merging proactive and reactive approaches to respond to cybersecurity incidents in a complex manner in order to gain greater insight into their cybersecurity setting

- Local and global cyber threat intelligence feeds are continuously created and modified to help organisations better understand what threats are on the forefront and how to react to them. Whereas, cyber threat intelligence feeds are used to constantly increase the reliability and efficacy of their cybersecurity incident response method
- The dynamic capabilities enabled by explainable human-centric analytics, such as cyber threat intelligence generation, dynamic risk assessment, and self and shared situational awareness, help organisations execute dynamic cybersecurity incident response strategies, enabling them to respond to a dynamic cyber threat environment in a fast, agile, creative, effective and proactive manner
- Self-awareness is exposed as a helpful way to make data-driven decisions in real-time, and is crucial in incident reporting and recognising new risks
- The storytelling report is generated fully automatically, reducing the burden on cybersecurity resources in exploring, detecting and analysing fraudulent behaviors
- The implicit knowledge (what happened and why?), which analysts have to investigate manually, is included in the generated story
- The log files with private information that cannot be sent to the third party for further processing are protected
- A common knowledge of CSA is established which facilitates comprehension and ultimately allows human participation in the incident handling process, which is primarily limited to security professionals
- Conduct of cyber data analysis across a diverse ecosystem of technology to explain root cause of incident
- The advantages of mutual sharing and learning from extended and accumulated threat intelligence by involving more people.

7.3 Study Limitations

Despite being promising, a number of limitations of explainable intelligence systems, particulaly contextual cyber awareness in its current state, have been identified and are as follows:

Data base and benchmarks: Lack of a database with labeled descriptive output similar to explain what is happened in the traffics and flows. I used our main databases, our gathered data and our own knowledge bases to obtain contextual insight into cyber data (logs or alerts). The local knowledge base includes supplementary information that is internally processed and the

raw data collected from security devices. The local knowledge base contains explicit knowledge about the situation of the event. Implicit knowledge is added to the knowledge base by predefined rules and procedures. So, it should be made and updated by the organisation under investigation. The main reason is the labor-intensive manual data collection and defining suitable benchmarks due to the lack of the publicly available dataset and benchmark corpus with fine-grained labels. The benchmark's limitations should be openly addressed. For example, I used parameters and criteria for the evaluation methods in Chapters 3, 4 and 5. A benchmark that does not provide a detailed discussion of limitations runs the risk of misleading readers. In extreme cases, this could even damage the wider research field by directing research efforts in the wrong direction. To avoid misleading researchers, I borrowed the criteria from the psychology areas to evaluate human cognition.

Access: This study's investigations depend on having access to people, organisations and their data. Because cyber data is typically classified as private, organisations are hesitant to share current cyber incidents with researchers. As a result, the data used to evaluate the proposed models was limited to what was collected from the education institute where the research was carried out.

Technique novelty: The study's aim was to empirically verify explainable intelligence methods on experts' cognition, followed by information discovery regarding cyber incidents in order to manage incidents more quickly. To compare their output and inform future study, state-of-the-art models were used. As a consequence, the novelty is evaluated in terms of its functional implementation in a real-world situation, or in terms of the void that has been found. As a result, the novelty is empirically evaluated in terms of its realistic application to a real-world situation where a gap was identified. Thus, the major part of the thesis focuses on gaining information to help decision-makers, cybersecurity providers, and incident response teams.

7.4 Future Research Directions

This section is intended to promote additional research on the theme of how organisations enhance resilience in their cybersecurity incident response by reducing cognitive overload on their experts, a subject with various academic possibilities. While this research has taken a step towards filling a crucial research gap, there are many possibilities for strengthening or building on the results. The following paragraphs address potential future research directions based on the study's results and limitations.

The study's background poses concerns about the proposed model's generalisability and identifies areas for future studies. It is difficult to generalise from interpretive research in the same way as quantitative research focused on statistical sampling methods can be generalised.

As a result, generalising the findings of this study to other situations involving the implementation of real-time analytics should be done with caution, as the findings might only be applicable to the incident response method and the sites analysed in this study. For example, this study acknowledges that the use of malware-related vulnerabilities, similar incident categories in the organisations studied, and relevant information collected in knowledge bases may limit the generalisability of the study's findings.

The results of this study may not be applicable to all types of vulnerabilities. As a result, further research is required to expand the knowledge base and support or refute the results of this study in other categories. Despite these limitations, the use of contextual analytics in the cybersecurity incident response process indicates a new direction for incident response research that considers the implications of contextual analytics capabilities in implementing dynamic incident response for both self and shared cyber awareness, and the results from this study will serve as a starting point for future research. The results of this study will serve as a foundation for future studies that will question, validate and expand the conclusions of this study.

Further study is needed to examine the conditions that promote or impede the implementation of contextual cybersecurity incident response. And boost the explanation models to gain higher accuracy from logs, alerts and financial transactions. Additional research may look at how various skills and activities of incident response units affect the creation of contextual incident response capability. Furthermore, before starting this research, the relevant information to enrich the cyber data in this study was collected and updated. The collection of data and the gathering of contextual knowledge were not intended for this study and can be avoided. The basic features of the data-driven incident would need further investigation in the future.

Finally, this study lays the groundwork for large-scale quantitative research into particular factors that aid organisations in improving incident management through qualitative analysis with narrative reports. The relationship between achieving agility through contextual analytical, self and shared CSA, and reducing human cognition overloads is a subject of ongoing research. Additional research is also needed to identify key differences between contextual analytics and other disruptive technologies in order to gain insights into how contextual analytics can offer distinct capabilities. The adoption of contextual analytics encourages a paradigm shift in the decision making process, reducing huge overload on expert cognition, however, more comprehensive studies are needed to examine its potential as well as the challenges it presents to organisations of all sizes.

To summarise, the author hopes that the results of this study will be beneficial to both theory and practise in the pursuit of a deeper understanding of how contextual analytics increase resilience in the process of cybersecurity incident response and improve incident management processes. The cybersecurity threat landscape is dynamic and nuanced, making it difficult to understand or pin down in rules. Organisations have no way of knowing what types of cyber threats and

assaults they will encounter in the future. They can, however, use analytics to build a proactive approach to cybersecurity incident response. This will ensure that their cybersecurity incident response is fast, agile, informed and creative.

This research sheds new light on how contextual analytics, a specialised analytics capability, can help organisations identify and react to cyber threats more rapidly by enhancing cognition capabilities in cybersecurity incident response such as situation understanding, complex risk assessment, and cyber threat intelligence. These capabilities assist organisations in improving incident management processes and dealing with both predictable and unexpected cybersecurity threats by incorporating complex incident response techniques such as active protection, continuous monitoring and active reconnaissance. Contextual analytics, the narrative approach, human cognition skills and cybersecurity incident response management are all topics that these concepts add to current research and invite future research.

Bibliography

[1] Neda Afzaliseresht, Yuan Miao, Sandra Michalska, Qing Liu, and Hua Wang. From logs to stories: Human-centred data mining for cyber threat intelligence. *IEEE Access*, 8: 19089–19099, 2020.

[2] Neda Afzaliseresht, Qing Liu, and Yuan Miao. An explainable intelligence model for security event analysis. In *Australasian Joint Conference on Artificial Intelligence*, pages 315–327. Springer, 2019.

[3] Neda AfzaliSeresht, Yuan Miao, Qing Liu, Assefa Teshome, and Wenjie Ye. Investigating cyber alerts with graph-based analytics and narrative visualization. In *Proceedings of the 24th International Conference Information Visualisation (IV)*, volume 20, pages 518–526, 2020.

[4] Armin Samii and Woojong Koh. Interactive visualization for system log analysis. 2019.

[5] Florian Menges and Günther Pernul. A comparative analysis of incident reporting formats. *Computers & Security*, 73:87–101, 2018.

[6] Smith Randy Franklin. Windows security log event id. https://www.ultimatewindowssecurity.com. Accessed: 2019-10-21.

[7] Roberto O Andrade and Sang Guun Yoo. Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications*, 48: 102352, 2019.

[8] Witold Kinsner. Towards cognitive security systems. In *11th International Conference on Cognitive Informatics and Cognitive Computing*, pages 539–539. IEEE, 2012.

[9] Humza Naseer. *A framework of dynamic cybersecurity incident response to improve incident response agility*. PhD thesis, 2018.

[10] Manfred Vielberth, Florian Menges, and Günther Pernul. Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity*, 2(1):1–15, 2019.

[11] Robert S Gutzwiller, Sarah M Hunt, and Douglas S Lange. A task analysis toward characterizing cyber-cognitive situation awareness (ccsa) in cyber defense analysts. In *International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 14–20. IEEE, 2016.

[12] Sushil Jajodia, Peng Liu, Vipin Swarup, and Cliff Wang. *Cyber situational awareness*. Springer, ISBN (9781441901408), 2009.

[13] Chanel Macabante, Sherry Wei, and David Schuster. Elements of cyber-cognitive situation awareness in organizations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 1624–1628. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[14] Roberto O Andrade and Sang Guun Yoo. Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications*, 48: 102352, 2019. ISSN 2214-2126. doi: https://doi.org/10.1016/j.jisa.2019.06.008. URL https://www.sciencedirect.com/science/article/pii/S2214212618307804.

[15] Lyndsey Franklin, Meg Pirrung, Leslie Blaha, Michelle Dowling, and Mi Feng. Toward a visualization-supported workflow for cyber alert management using threat models and human-centered design. In *Symposium on Visualization for Cyber Security (VizSec)*, pages 1–8. IEEE, 2017.

[16] Tony Jinks. The 5w1h method. In *Psychological Perspectives on Reality, Consciousness and Paranormal Experience*, pages 41–44. Springer, 2019.

[17] Alessandra de Melo e Silva, João José Costa Gondim, Robson de Oliveira Albuquerque, and Luis Javier García Villalba. A methodology to evaluate standards and platforms within cyber threat intelligence. *Future Internet*, 12(6):108, 2020.

[18] R Bahrevar and K Khorasani. Note for national defence: Cognitive security analyst and why we need it.

[19] Alireza Sadighian, José M. Fernandez, Antoine Lemay, and Saman T. Zargar. Ontids: A highly flexible context-aware and ontology-based alert correlation framework. In Jean Luc Danger, Mourad Debbabi, Jean-Yves Marion, Joaquin Garcia-Alfaro, and Nur Zincir Heywood, editors, *Foundations and Practice of Security*, pages 161–177. Springer International Publishing, 2014. ISBN 978-3-319-05302-8.

[20] 4 major challenges facing fraud detection; ways to resolve them using machine learning. https://medium.com/razorthink-ai/4-major-challenges-facing-fraud-detection-ways-to-resolve-them-using-machine-le

[21] Michael Edward Edge and Pedro R Falcone Sampaio. A survey of signature based methods for financial fraud detection. *computers & security*, 28(6):381–394, 2009.

[22] Enisa. Financial fraud in the digital space. https://www.enisa.europa.eu/publications/enisa-position-papers-and-opinions/financial-fraud-in-the-digital-space, 2018.

[23] Australian Government. Create an incident response plan. https://www.counterfraud.gov.au/fraud-countermeasures/create-incident-response-plan.

[24] Tie Li, Gang Kou, Yi Peng, and S Yu Philip. An integrated cluster detection, optimization, and interpretation approach for financial data. *IEEE Transactions on Cybernetics*, 2021.

[25] Dongxu Huang, Dejun Mu, Libin Yang, and Xiaoyan Cai. Codetect: Financial fraud detection with anomaly feature detection. *IEEE Access*, 6:19161–19174, 2018.

[26] Eren Kurshan, Hongda Shen, and Haojie Yu. Financial crime & fraud detection using graph computing: Application considerations & outlook. In *2nd International Conference on Transdisciplinary AI (TransAI)*, pages 125–130. IEEE, 2020.

[27] Thushara Amarasinghe, Achala Aponso, and Naomi Krishnarajah. Critical analysis of machine learning based approaches for fraud detection in financial transactions. In *Proceedings of the International Conference on Machine Learning Technologies*, pages 12–17, 2018.

[28] Uğur Ünal, Ceyda Nur Kahya, Yaprak Kurtlutepe, and Hasan Dağ. Investigation of cyber situation awareness via siem tools: a constructive review. In *2021 6th International Conference on Computer Science and Engineering (UBMK)*, pages 676–681. IEEE, 2021.

[29] Veikko Siukonen. Human factors of cyber operations: decision making behind advanced persistence threat operations. In *European Conference on Cyber Warfare and Security*, pages 790–XIX. Academic Conferences International Limited, 2019.

[30] Liuyue Jiang, Asangi Jayatilaka, Mehwish Nasim, Marthie Grobler, Mansooreh Zahedi, and M Ali Babar. Systematic literature review on cyber situational awareness visualizations. *arXiv preprint arXiv:2112.10354*, 2021.

[31] Salvador Llopis, Javier Hingant, Israel Pérez, Manuel Esteve, Federico Carvajal, Wim Mees, and Thibault Debatty. A comparative analysis of visualisation techniques to achieve cyber situational awareness in the military. In *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–7. IEEE, 2018.

[32] Ulrik Franke and Joel Brynielsson. Cyber situational awareness–a systematic review of the literature. *Computers & security*, 46:18–31, 2014.

[33] Gustavo González-Granadillo, Susana González-Zarzosa, and Rodrigo Diaz. Security information and event management (siem): Analysis, trends, and usage in critical infrastructures. *Sensors*, 21(14):4759, 2021.

[34] O Rochford, KM Kavanagh, and T Bussa. Critical capabilities for security information and event management. *Resource Document. Gartner Inc., Stamford*, 2016.

[35] Blake D Bryant and Hossein Saiedian. A novel kill-chain framework for remote security log analysis with siem software. *computers & security*, 67:198–210, 2017.

[36] Nabil Moukafih, Ghizlane Orhanou, and Said El Hajji. Neural network-based voting system with high capacity and low computation for intrusion detection in siem/ids systems. *Security and Communication Networks*, 2020, 2020.

[37] Tero Kokkonen and Samir Puuska. Blue team communication and reporting for enhancing situational awareness from white team perspective in cyber security exercises. In *Internet of things, smart spaces, and next generation networks and systems*, pages 277–288. Springer, 2018.

[38] Anita D'Amico, Kirsten Whitley, Daniel Tesone, Brianne O'Brien, and Emilie Roth. Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 49, pages 229–233. SAGE Publications Sage CA: Los Angeles, CA, 2005.

[39] Mica R Endsley, Daniel J Garland, et al. Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, 1(1):3–21, 2000.

[40] Flora Amato, Giovanni Cozzolino, Antonino Mazzeo, and Francesco Moscato. Detect and correlate information system events through verbose logging messages analysis. *Computing*, 101(7):819–830, 2019.

[41] Iain Dickson. Text classification of network intrusion alerts to enhance cyber situation awareness and automate alert triage. https://www.dst.defence.gov.au/publication/, 2017.

[42] Michael Muggler, Eshwarappa Rekha, and Celikel Ebru. Cybersecurity management through logging analytics. *International Conference on Applied Human Factors and Ergonomics*, pages 3–15, 2017.

[43] Mathew Vermeer, Michel van Eeten, and Carlos Gañán. Ruling the rules: Quantifying the evolution of rulesets, alerts and incidents in network intrusion detection. 2022.

[44] Jianxin Jiao, Feng Zhou, Nagi Z Gebraeel, and Vincent Duffy. Towards augmenting cyber-physical-human collaborative cognition for human-automation interaction in complex manufacturing and operational environments. *International Journal of Production Research*, 58(16):5089–5111, 2020.

[45] Chen Zhong, John Yen, Peng Liu, and Robert F Erbacher. Automate cybersecurity data triage by leveraging human analysts' cognitive process. In *2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, pages 357–363. IEEE, 2016.

[46] Kai Zhang and Jingju Liu. Review on the application of knowledge graph in cyber security assessment. *IOP Conference Series: Materials Science and Engineering*, 768:052103, 2020. doi: 10.1088/1757-899x/768/5/052103.

[47] Steven Noel, Eric Harley, Kam Him Tam, Michael Limiero, and Matthew Share. Cygraph: graph-based analytics and visualization for cybersecurity. In *Handbook of Statistics*, volume 35, pages 117–167. Elsevier, 2016.

[48] MITRE Case Study. Graph technology powers cybersecurity situational awareness that's more scalable, flexible and comprehensive. https://neo4j.com/case-studies/mitre/, 2019.

[49] Karim Tabia and Philippe Leray. Alert correlation: Severe attack prediction and controlling false alarm rate tradeoffs. *Intelligent Data Analysis*, 15(6):955–978, 2011.

[50] Jiao Sun, Yin Li, Charley Chen, Jihae Lee, Xin Liu, Zhongping Zhang, Ling Huang, Lei Shi, and Wei Xu. Fdhelper: Assist unsupervised fraud detection experts with interactive feature selection and evaluation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[51] Zuraidah Mohd Sanusi, Mohd Nor Firdaus Rameli, and Yusarina Mat Isa. Fraud schemes in the banking institutions: prevention measures to avoid severe financial loss. *Procedia economics and finance*, 28:107–113, 2015.

[52] Marion R Fremont-Smith. Pillaging of charitable assets: Embezzlement and fraud. *Exempt Organization Tax Review*, 46(33):333–346, 2004.

[53] Marco Sánchez, Jenny Torres, Patricio Zambrano, and Pamela Flores. Fraudfind: Financial fraud detection by analyzing human behavior. In *8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 281–286. IEEE, 2018.

[54] Massimiliano Albanese, Hasan Cam, and Sushil Jajodia. Automated cyber situation awareness tools and models for improving analyst performance. *Cybersecurity systems for human cognition augmentation*, pages 47–60, 2014.

[55] PWDC Jayathilake, NR Weeraddana, and HKEP Hettiarachchi. Automatic detection of multi-line templates in software log files. In *17th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 1–8. IEEE, 2017.

[56] Chris W Johnson. Contrasting approaches to incident reporting in the development of safety and security–critical software. Safecomp, 2015.

[57] Petri Toropainen. Utilizing cyber security kill chain model to improve siem capabilities. 2020.

[58] Qiong Wu, Zhiqi Shen, Cyril Leungy, Huiguo Zhang, Yundong Cai, Chunyan Miao, et al. Internet of things based data driven storytelling for supporting social connections. In *International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pages 383–390. IEEE, 2013.

[59] Jock Mackinaly, Robert Kosara, and Michelle Wallace. Data storytelling using visualization to share the human impact of numbers, 2014.

[60] W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

[61] Keshnee Padayachee and Elias Worku. Shared situational awareness in information security incident management. In *12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 479–483. IEEE, 2017.

[62] Joan Gargano and Ken Weiss. Whois and network information lookup service. Technical report, Whois+, 1834.

[63] Frédéric Cuppens. Managing alerts in a multi-intrusion detection environment. In *Seventeenth Annual Computer Security Applications Conference*, pages 22–31. IEEE, 2001.

[64] John P Rouillard. Real-time log file analysis using the simple event correlator (sec). In *LISA*, volume 4, pages 133–150, 2004.

[65] Erik Hollnagel. *Barriers and accident prevention*. Routledge, 2016.

[66] Prashanth Rajivan and Nancy J Cooke. Information-pooling bias in collaborative security incident correlation analysis. *Human factors*, 60(5):626–639, 2018.

[67] Matt Bromiley. Sans 2019 incident response (ir) survey: It's time for a change analyst paper (requires membership in sans.org community). https://www.sans.org/reading-room/whitepapers/incident/paper/39070, 2019.

[68] Richard Baskerville, Paolo Spagnoletti, and Jongwoo Kim. Incident-centered information security: Managing a strategic balance between prevention and response. *Information & management*, 51(1):138–151, 2014.

[69] Jason Creasy and Ian Glover. Cyber security incident response guide. Technical report, Council for Registered Ethical Security Testers (CREST), 2013.

[70] Robert Eastman, Michael Versace, and Alan Webber. Big data and predictive analytics: on the cybersecurity front line. *IDC Whitepaper, February*, 2015.

[71] George Grispos, William Bradley Glisson, and Tim Storer. Rethinking security incident response: The integration of agile principles. *arXiv preprint arXiv:1408.2431*, 2014.

[72] Piya Shedden, Atif Ahmad, Wally Smith, Heidi Tscherning, and Rens Scheepers. Asset identification in information security risk assessment: A business practice approach. *Communications of the Association for Information Systems*, 39(1):15, 2016.

[73] Eoghan Casey. Investigating sophisticated security breaches. *Communications of the ACM*, 49(2):48–55, 2006.

[74] Terence Tan, AB Ruighaver, and Atif Ahmad. Incident handling: Where the need for planning is often not recognised. In *1st Australian computer, network & information forensics conference*, pages 1–10, 2003.

[75] Rodrigo Werlinger, Kasia Muldner, Kirstie Hawkey, and Konstantin Beznosov. Preparation, detection, and analysis: the diagnostic work of it security incident response. *Information Management & Computer Security*, 2010.

[76] John Bailey, Eser Kandogan, Eben Haber, and Paul P Maglio. Activity-based management of it service delivery. In *Proceedings of the symposium on Computer human interaction for the management of information technology*, pages 5–es, 2007.

[77] Iain Dickson. Text classification of network intrusion alerts to enhance cyber situation awareness and automate alert triage. https://www.dst.defence.gov.au/publication/, 2017.

[78] Tejaswini Herath and H Raghav Rao. Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*, 18(2):106–125, 2009.

[79] Inger Anne Tøndel, Maria B Line, and Martin Gilje Jaatun. Information security incident management: Current practice as reported in the literature. *Computers & Security*, 45: 42–57, 2014.

[80] Cleidson RB de Souza, Claudio S Pinhanez, and Victor F Cavalcante. Information needs of system administrators in information technology service factories. In *Proceedings of the 5th ACM Symposium on Computer Human Interaction for Management of Information Technology*, pages 1–10, 2011.

[81] Atif Ahmad, Justin Hadgkiss, and Anthonie B Ruighaver. Incident response teams–challenges in supporting the organisational security function. *Computers & Security*, 31 (5):643–652, 2012.

[82] International Organization for Standardization. *ISO/IEC 27001: 2013: Information Technology–Security Techniques–Information Security Management Systems–Requirements*. International Organization for Standardization, 2013.

[83] Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. Computer security incident handling guide. *NIST Special Publication*, 800(61):1–147, 2012.

[84] Robin Ruefle, Audrey Dorofee, David Mundie, Allen D Householder, Michael Murray, and Samuel J Perl. Computer security incident response team development and evolution. *IEEE Security & Privacy*, 12(5):16–26, 2014.

[85] Patrick Kral. Incident handler's handbook. https://www.sans.org/reading-room/whitepapers/incident/incident-handlers-handbook-33901, 2011.

[86] Stefan Metzger, Wolfgang Hommel, and Helmut Reiser. Integrated security incident management–concepts and real-world experiences. In *2011 Sixth International Conference on IT Security Incident Management and IT Forensics*, pages 107–121. IEEE, 2011.

[87] Cathrine Hove and Marte Tårnes. Information security incident management: an empirical study of current practice. Master's thesis, Institutt for telematikk, 2013.

[88] Hennie A Kruger and Wayne D Kearney. A prototype for assessing information security awareness. *Computers & security*, 25(4):289–296, 2006.

[89] Charlie C Chen, RS Shaw, and Samuel C Yang. Mitigating information security risks by increasing user security awareness: A case study of an information security awareness system. *Information Technology, Learning & Performance Journal*, 24(1), 2006.

[90] Dale L Goodhue and Detmar W Straub. Security concerns of system users: A study of perceptions of the adequacy of security. *Information & Management*, 20(1):13–27, 1991.

[91] Detmar W Straub and Richard J Welke. Coping with systems risk: Security planning models for management decision making. *MIS quarterly*, pages 441–469, 1998.

[92] Andy Johnston and Jessica Reust. Network intrusion investigation–preparation and challenges. *digital investigation*, 3(3):118–126, 2006.

[93] Erka Koivunen. "why wasn't i notified?": Information security incident reporting demystified. In *Nordic Conference on Secure IT Systems*, pages 55–70. Springer, 2010.

[94] Martin Gilje Jaatun, Eirik Albrechtsen, Maria B Line, Inger Anne Tøndel, and Odd Helge Longva. A framework for incident response management in the petroleum industry. *International Journal of Critical Infrastructure Protection*, 2(1-2):26–37, 2009.

[95] Suhaila Ismail, Arniyati Ahmad, and Mohd Afizi Mohd Shukran. New method of forensic computing in a small organization. *Aust J Basic Appl Sci*, 5(9):2019e25, 2011.

[96] Peter Katsumata, Judy Hemenway, and Wes Gavins. Cybersecurity risk management. In *Military Communications Conference (MILCOM)*, pages 890–895. IEEE, 2010.

[97] Edward Humphreys. Information security management standards: Compliance, governance and risk management. *information security technical report*, 13(4):247–255, 2008.

[98] Thomas L Norman. *Integrated Security Systems Design: A Complete Reference for Building Enterprise-wide Digital Security Systems*. Butterworth-Heinemann, 2014.

[99] Donn B Parker. Risks of risk-based security. *Communications of the ACM*, 50(3):120, 2007.

[100] Jeb Webb, Atif Ahmad, Sean B Maynard, and Graeme Shanks. A situation awareness model for information security risk management. *Computers & security*, 44:1–15, 2014.

[101] Lawrence A Gordon and Martin P Loeb. Budgeting process for information security expenditures. *Communications of the ACM*, 49(1):121–125, 2006.

[102] Lara Khansa and Divakaran Liginlal. Valuing the flexibility of investing in security process innovations. *European Journal of Operational Research*, 192(1):216–235, 2009.

[103] Thomas R Peltier. *Information security risk analysis*. CRC press, 2005.

[104] Jackie Rees and Jonathan Allen. The state of risk assessment practices in information security: An exploratory investigation. *Journal of Organizational Computing and Electronic Commerce*, 18(4):255–277, 2008.

[105] Atif Ahmad, Rachelle Bosua, and Rens Scheepers. Protecting organizational competitive advantage: A knowledge leakage perspective. *Computers & Security*, 42:27–39, 2014.

[106] Piya Shedden, Wally Smith, Rens Scheepers, and Atif Ahmad. Towards a knowledge perspective in information security risk assessments–an illustrative case study. 2009.

[107] Margareth Stoll. From information security management to enterprise risk management. In *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*, pages 9–16. Springer, 2015.

[108] Aisha Abdallah, Mohd Aizaini Maarof, and Anazida Zainal. Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68:90–113, 2016.

[109] Geoffrey J McLachlan. *Discriminant analysis and statistical pattern recognition*, volume 544. John Wiley & Sons, 2004.

[110] Richard J Bolton, David J Hand, et al. Unsupervised profiling methods for fraud detection. *Credit scoring and credit control VII*, pages 235–255, 2001.

[111] Daniel A Keim, Florian Mansmann, Jörn Schneidewind, Jim Thomas, and Hartmut Ziegler. Visual analytics: Scope and challenges. In *Visual data mining*, pages 76–90. Springer, 2008.

[112] Kishore Singh and Peter Best. Anti-money laundering: Using data visualization to identify suspicious activity. *International Journal of Accounting Information Systems*, 34:100418, 2019.

[113] Maxime Dumas, Michael J McGuffin, and Victoria L Lemieux. Financevis. net-a visual survey of financial data visualizations. In *Poster Abstracts of IEEE Conference on Visualization*, volume 2, page 8, 2014.

[114] Roger A Leite, Theresia Gschwandtner, Silvia Miksch, Erich Gstrein, and Johannes Kuntner. Visual analytics for event detection: Focusing on fraud. *Visual Informatics*, 2(4): 198–212, 2018.

[115] Sungahn Ko, Isaac Cho, Shehzad Afzal, Calvin Yau, Junghoon Chae, Abish Malik, Kaethe Beck, Yun Jang, William Ribarsky, and David S Ebert. A survey on visual analysis approaches for financial data. In *Computer Graphics Forum*, volume 35, pages 599–617. Wiley Online Library, 2016.

[116] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2016.

[117] Yusuf Sait Canbaz, Uğur Doğrusöz, Mehmet Çeliksoy, Fatma Güngör, and Koray Kurban. Hydra: detecting fraud in financial transactions via graph based representation and visual analysis. In *4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pages 1–6. IEEE, 2020.

[118] Susie Xi Rao, Shuai Zhang, Zhichao Han, Zitao Zhang, Wei Min, Zhiyao Chen, Yinan Shan, Yang Zhao, and Ce Zhang. xfraud: Explainable fraud transaction detection on heterogeneous graphs. *arXiv preprint arXiv:2011.12193*, 2020.

[119] Ismini Psychoula, Andreas Gutmann, Pradip Mainali, Sharon H Lee, Paul Dunphy, and Fabien Petitcolas. Explainable machine learning for fraud detection. *Computer*, 54(10): 49–59, 2021.

[120] Paul Thagard. *Mind: Introduction to cognitive science*. MIT press, 2005.

[121] Michael I Posner. *Foundations of cognitive science*. MIT press Cambridge, MA, 1989.

[122] Nancy Cooke, Michael McNeese, et al. Preface to special issue on the cognitive science of cyber defence analysis. *EAI Endorsed Transactions on Security and Safety*, 1(2):e1, 2013.

[123] Samuel Mahoney, Emilie Roth, Kristin Steinke, Jonathan Pfautz, Curt Wu, and Mike Farry. A cognitive task analysis for cyber situational awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(4):279–283, 2010. doi: 10.1177/154193121005400403.

[124] Massimiliano Albanese, Hasan Cam, and Sushil Jajodia. Automated cyber situation awareness tools and models for improving analyst performance. In *Cybersecurity systems for human cognition augmentation*, pages 47–60. Springer, 2014.

[125] Robert S Gutzwiller, Sarah M Hunt, and Douglas S Lange. A task analysis toward characterizing cyber-cognitive situation awareness (ccsa) in cyber defense analysts. In *International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 14–20. IEEE, 2016.

[126] Funmilade Faniyi, Peter R Lewis, Rami Bahsoon, and Xin Yao. Architecting self-aware software systems. In *Conference on Software Architecture*, pages 91–94. IEEE, 2014.

[127] Michael McNeese, Nancy J Cooke, Anita D'Amico, Mica R Endsley, Cleotilde Gonzalez, Emilie Roth, and Eduardo Salas. Perspectives on the role of cognition in cyber security. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 56, pages 268–271. SAGE Publications Sage CA: Los Angeles, CA, 2012.

[128] Peter R Lewis, Marco Platzner, Bernhard Rinner, Jim Tørresen, and Xin Yao. Self-aware computing systems. *Natural Computing Series*, 2016.

[129] Albert A Nofi. Defining and measuring shared situational awareness. Technical report, Center For Naval Analyses AlexandriA VA, 2000.

[130] Gary Klein, Brian Moon, and Robert R Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE intelligent systems*, 21(4):70–73, 2006.

[131] Matthias Baldauf, Schahram Dustdar, and Florian Rosenberg. A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277, 2007.

[132] Piyush Nimbalkar, Varish Mulwad, Nikhil Puranik, Anupam Joshi, and Tim Finin. Semantic interpretation of structured log files. In *17th International Conference on Information Reuse and Integration (IRI)*, pages 549–555. IEEE, 2016.

[133] Yoan Chabot, Aurélie Bertaux, Christophe Nicolle, and Tahar Kechadi. An ontology-based approach for the reconstruction and analysis of digital incidents timelines. *Digital Investigation*, 15:83–100, 2015.

[134] P. Bonatti, W. Dullaert, J.D. Fernndez, S. Kirrane, and A. Milosevic, U.and Polleres. The special policy log vocabulary. https://aic.ai.wu.ac.at/qadlod/policyLog/. Accessed: 2019-10-31.

[135] Markus Rittenbruch. Atmosphere: a framework for contextual awareness. *International Journal of Human-Computer Interaction*, 14(2):159–180, 2002.

[136] Ali Reza Arasteh, Mourad Debbabi, Assaad Sakha, and Mohamed Saleh. Analyzing multiple logs for forensic evidence. *digital investigation*, 4:82–91, 2007.

[137] Saad Khan and Simon Parkinson. Eliciting and utilising knowledge for security event log analysis: An association rule mining and automated planning approach. *Expert Systems with Applications*, 113:116–127, 2018.

[138] Christophe Bertero, Matthieu Roy, Carla Sauvanaud, and Gilles Trédan. Experience report: Log mining using natural language processing and application to anomaly detection. In *28th International Symposium on Software Reliability Engineering (ISSRE)*, pages 351–360. IEEE, 2017.

[139] Aaron Tuor, Samuel Kaplan, Brian Hutchinson, Nicole Nichols, and Sean Robinson. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. *arXiv preprint arXiv:1710.00811*, 2017.

[140] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. Fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE transactions on visualization and computer graphics*, 20(12):1773–1782, 2014.

[141] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics*, 22(1):240–249, 2015.

[142] Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael I Jordan. Detecting large-scale system problems by mining console logs. In *Proceedings of the 22nd symposium on Operating systems principles*, pages 117–132. ACM, 2009.

[143] Michal Aharon, Gilad Barash, Ira Cohen, and Eli Mordechai. One graph is worth a thousand logs: Uncovering hidden structures in massive system event logs. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 227–243. Springer, 2009.

[144] Tao Li, Yexi Jiang, Chunqiu Zeng, Bin Xia, Zheng Liu, Wubai Zhou, Xiaolong Zhu, Wentao Wang, Liang Zhang, Jun Wu, et al. Flap: An end-to-end event log analysis platform for system management. In *Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining*, pages 1547–1556. ACM, 2017.

[145] Amir Azodi, Feng Cheng, and Christoph Meinel. Towards better attack path visualizations based on deep normalization of host/network ids alerts. In *30th International Conference on Advanced Information Networking and Applications (AINA)*, pages 1064–1071. IEEE, 2016.

[146] Yuan Fang, Kingsley Kuan, Jie Lin, Cheston Tan, and Vijay Chandrasekhar. Object detection meets knowledge graphs. International Joint Conferences on Artificial Intelligence, 2017.

[147] Zhiliang Qin, Chen Cen, Wang Jie, Teo Sin Gee, Vijay Ramaseshan Chandrasekhar, Zhongbo Peng, and Zeng Zeng. Knowledge-graph based multi-target deep-learning models for train anomaly detection. In *International Conference on Intelligent Rail Transportation (ICIRT)*, pages 1–5. IEEE, 2018.

[148] Alina Nesen and Bharat Bhargava. Knowledge graphs for semantic-aware anomaly detection in video. In *3rd International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 65–70. IEEE, 2020.

[149] Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.

[150] STIXTM. Structured threat information expression (stixtm). https://docs.google.com/document/d/1IcA5KhglNdyX3tO17bBluC5nqSf70M5qgK9nuAoYJgw/, 2017.

[151] Robin Cover. Incident object description and exchange format (iodef). http://xml.coverpages.org/iodef.html, 2008. Accessed: 2019-09-06.

[152] Jana Komárková, Martin Husák, Martin Laštovička, and Daniel Tovarňák. Crusoe: Data model for cyber situational awareness. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, page 36. ACM, 2018.

[153] Elchin Asgarli and Eric Burger. Semantic ontologies for cyber threat sharing standards. In *Symposium on Technologies for Homeland Security (HST)*, pages 1–6. IEEE, 2016.

[154] Stefan Fenz, Andreas Ekelhart, and Edgar Weippl. Semantic potential of existing security advisory standards. In *Proceedings of the 1st Conference-Forum of Incident Response and Security Teams*, 2008.

[155] X-arf. x-arf network abuse reporting 2.0. http://xarf.org/, 2008. Accessed: 2019-09-06.

[156] Paweł Pawlinski, Przemysław Jaroszewski, Janusz Urbanowicz, Paweł Jacewicz, Przemysław Zielony, Piotr Kijewski, and Katarzyna Gorzelak. Standards and tools for exchange and processing of actionable information. *European Union Agency for Network and Information Security, Heraklion, Greece*, 2014.

[157] John D Howard and Thomas A Longstaff. A common language for computer security incidents. Technical report, Sandia National Labs., Albuquerque, NM (US); Sandia National Labs, 1998.

[158] Clive Blackwell. A security ontology for incident analysis. In *Proceedings of the 6th Annual Workshop on Cyber Security and Information Intelligence Research*, pages 1–4, 2010.

[159] Onur Deniz and Hilmi Berk Celikoglu. Overview to some existing incident detection algorithms: a comparative evaluation. *Procedia Social and Behavioral Sciences*, 2011: 1–13, 2011.

[160] Victor Ion Munteanu, Andrew Edmonds, Thomas M Bohnert, and Teodor-Florin Fortis. Cloud incident management, challenges, research directions, and architectural approach. In *7th International Conference on Utility and Cloud Computing*, pages 786–791. IEEE, 2014.

[161] Séamus Ó Ciardhuáin. An extended model of cybercrime investigations. *International Journal of Digital Evidence*, 3(1):1–22, 2004.

[162] Yunus Yusoff, Roslan Ismail, and Zainuddin Hassan. Common phases of computer forensics investigation models. *International Journal of Computer Science & Information Technology*, 3(3):17–31, 2011.

[163] Matti Hyvärinen. Analyzing narratives and story-telling. *The SAGE handbook of social research methods*, pages 447–460, 2008.

[164] Andreas Graefe. Guide to automated journalism. https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ, 2016.

[165] Daniel Paul Barrett, Scott Alan Bronikowski, Haonan Yu, and Jeffrey Mark Siskind. Robot language learning, generation, and comprehension. *arXiv preprint arXiv:1508.06161*, 2015.

[166] Kim Dongwhan and Joonhwan Lee. Designing an algorithm-driven text generation system for personalized and interactive news reading. *International Journal of Human–Computer Interaction*, pages 109–122, 2019.

[167] Carlson Matt. The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital journalism*, pages 416–433, 2015.

[168] Naeun Lee, Kirak Kim, and Taeseon Yoon. Implementation of robot journalism by programming custombot using tokenization and custom tagging. pages 566–570, 2017.

[169] David Caswell and Konstantin Dörr. Automated journalism 2.0: Event-driven narratives. *Journalism Practice*, 12(4):477–496, 2018. URL https://doi.org/10.1080/17512786.2017.1320773.

[170] Arnak Poghosyan, Ashot Harutyunyan, Naira Grigoryan, and Nicholas Kushmerick. Incident management for explainable and automated root cause analysis in cloud data centers. *Journal of Universal Computer Science*, 27(11):1152–1173, 2021.

[171] Eric Holder and Ning Wang. Explainable artificial intelligence (xai) interactively working with humans as a junior cyber analyst. *Human-Intelligent Systems Integration*, 3(2): 139–153, 2021.

[172] William R. Swartout and Johanna D. Moore. Explanation in second generation expert systems. In Jean-Marc David, Jean-Paul Krivine, and Reid Simmons, editors, *Second Generation Expert Systems*, pages 543–585, Berlin, Heidelberg, 1993. Springer Berlin Heidelberg. ISBN 978-3-642-77927-5.

[173] W. Lewis Johnson. Agents that learn to explain themselves. In *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, AAAI'94, page 1257–1263. AAAI Press, 1994.

[174] Michael Van Lent, William Fisher, and Michael Mancuso. An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the national conference on artificial intelligence*, pages 900–907. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[175] Mark Core, David Traum, H Chad Lane, William Swartout, Jonathan Gratch, Michael Van Lent, and Stacy Marsella. Teaching negotiation skills through practice and reflection with virtual humans. *Simulation*, 82(11):685–701, 2006.

[176] Francisco Elizalde, L Enrique Sucar, Manuel Luque, J Diez, and Alberto Reyes. Policy explanation in factored markov decision processes. In *Proceedings of the 4th European workshop on probabilistic graphical models (PGM 2008)*, pages 97–104, 2008.

[177] Thomas Dodson, Nicholas Mattei, and Judy Goldsmith. A natural language argumentation interface for explanation generation in markov decision processes. In *International Conference on Algorithmic Decision Theory*, pages 42–55. Springer, 2011.

[178] Omar Zia Khan, Pascal Poupart, and James P Black. Automatically generated explanations for markov decision processes. In *Decision theory models for applications in artificial intelligence: Concepts and solutions*, pages 144–163. IGI Global, 2012.

[179] David V Pynadath, Heather Rosoff, and Richard S John. Semi-automated construction of decision-theoretic models of human behavior. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 891–899, 2016.

[180] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[181] Lisa Anne Hendricks, Zeynep Akata, Marcus Rohrbach, Jeff Donahue, Bernt Schiele, and Trevor Darrell. Generating visual explanations. In *European conference on computer vision*, pages 3–19. Springer, 2016.

[182] Wenbo Guo, Dongliang Mu, Jun Xu, Purui Su, Gang Wang, and Xinyu Xing. Lemna: Explaining deep learning based security applications. In *proceedings of the 2018 ACM SIGSAC conference on computer and communications security*, pages 364–379, 2018.

[183] Zhangzhang Si and Song-Chun Zhu. Learning and-or templates for object recognition and detection. *IEEE transactions on pattern analysis and machine intelligence*, 35(9): 2189–2205, 2013.

[184] Andy Shih, Arthur Choi, and Adnan Darwiche. A symbolic approach to explaining bayesian network classifiers. *arXiv preprint arXiv:1805.03364*, 2018.

[185] David Gunning. Broad agency announcement explainable artificial intelligence (xai). Technical report, Technical report, 2016.

[186] Shixia Liu, Xiting Wang, Mengchen Liu, and Jun Zhu. Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, 1(1):48–56, 2017.

[187] Chun-Hao Chang, Elliot Creager, Anna Goldenberg, and David Duvenaud. Explaining image classifiers by adaptive dropout and generative in-filling. *arXiv preprint arXiv:1807.08024*, 2, 2018.

[188] Anurag Koul, Sam Greydanus, and Alan Fern. Learning finite state representations of recurrent policy networks. *arXiv preprint arXiv:1811.12530*, 2018.

[189] Guilherme Fião, Teresa Romão, Nuno Correia, Pedro Centieiro, and A Eduardo Dias. Automatic generation of sport video highlights based on fan's emotions and content. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology*, pages 1–6, 2016.

[190] Antonio Pecchia. *On the use of event logs for the analysis of system failures.* PhD thesis, University of Naples Federico II, Italy, 2011.

[191] Mazaher Ghorbani and Masoud Abessi. A new methodology for mining frequent itemsets on temporal data. *IEEE Transactions on Engineering Management*, 64(4):566–573, 2017.

[192] Andrew Marrington, Ibrahim Baggili, George Mohay, and Andrew Clark. Cat detect (computer activity timeline detection): A tool for detecting inconsistency in computer activity timelines. *digital investigation*, 8:S52–S61, 2011.

[193] Anjana Kakoti Mahanta, Fokrul Alom Mazarbhuiya, and Hemanta K Baruah. Finding calendar-based periodic patterns. *Pattern Recognition Letters*, 29(9):1274–1284, 2008.

[194] Duc Trong Le, Hady W Lauw, and Yuan Fang. Basket-sensitive personalized item recommendation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, volume 25, pages 2060–2066. International Joint Conference on Artificial Intelligence (IJCAI), 2017.

[195] Hoda Memarzadeh, Nasser Ghadiri, and Sara Parikhah Zarmehr. A graph database approach for temporal modeling of disease progression. In *8th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 293–297. IEEE, 2018.

[196] Iyad Aqra, Tutut Herawan, Norjihan Abdul Ghani, Adnan Akhunzada, Akhtar Ali, Ramdan Bin Razali, Manzoor Ilahi, and Kim-Kwang Raymond Choo. A novel association rule mining approach using tid intermediate itemset. *PLOS ONE*, 13(1):1–32, 2018. URL https://doi.org/10.1371/journal.pone.0179703.

[197] Bruno Caseiro. Malware simulator. https://www.linkedin.com/pulse/malware-simulator-bruno-caseiro. Accessed: 2019-10-22.

[198] Farzaneh Asgharpour, Debin Liu, and L Jean Camp. Mental models of security risks. In *International conference on financial cryptography and data security*, pages 367–377. Springer, 2007.

[199] John R Goodall, Wayne G Lutters, and Anita Komlodi. Developing expertise for network intrusion detection. *Information Technology & People*, 2009.

[200] John R Goodall, Wayne G Lutters, and Anita Komlodi. I know my network: collaboration and expertise in intrusion detection. In *Proceedings of the conference on Computer supported cooperative work*, pages 342–345. ACM, 2004.

[201] CBEST Intelligence-Led Testing. Cbest implementation guide, bank of england. www.bankofengland.co.uk/-/media/boe/files/financial-stability/financial-sector-continuity/cbest-implementation-guide.pdf. Accessed: 2019-10-30.

[202] Manfred Vielberth, Florian Menges, and Günther Pernul. Human-as-a-security-sensor for harvesting threat intelligence. *Cybersecurity*, 2:1–15, 2019.

[203] Martin Roesch et al. Snort: Lightweight intrusion detection for networks. In *Lisa*, volume 99, pages 229–238, 1999.

[204] Komarkova Jana, Martin Husak, Martin Lastovicka, and Daniel Tovarnak. Crusoe: Data model for cyber situational awareness. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*. ACM, 2018.

[205] Virus total. Virustotal-free online virus, malware and url scanner. https://www.virustotal.com/en.

[206] ThreatMiner. Data mining for threat intelligence. https://www.threatminer.org/.

[207] Sergei Egorov and Gene Savchuk. Snortran: An optimizing compiler for snort rules. *Fidelis Security Systems*, 2002.

[208] Xiaojing Liao, Kan Yuan, XiaoFeng Wang, Zhou Li, Luyi Xing, and Raheem Beyah. Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In *Proceedings of the Conference on Computer and Communications Security*, pages 755–766. ACM, 2016.

[209] Andreas Graefe. Guide to automated journalism. https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php, 2016.

[210] Davod Stevenson. Foresee: Human and machine learning working together. https://www.secureworks.com/blog/foresee-human-and-machine-learning-working-together, 2018.

[211] Glenn D Israel. Determining sample size. 1992.

[212] How to analyze survey data. https://www.surveymonkey.com/mp/how-to-analyze-survey-data/.

[213] Evie McCrum-Gardner. Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, 46(1):38–41, 2008. URL https://www.sciencedirect.com/science/article/pii/S0266435607004378.

[214] Elmar Kiesling, Andreas Ekelhart, Kabul Kurniawan, and Fajar Ekaputra. The sepses knowledge graph: An integrated resource for cybersecurity. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web (ISWC)*, pages 198–214. Springer International Publishing, 2019.

[215] Daniel Schlette, Fabian Böhm, Marco Caselli, and Günther Pernul. Measuring and visualizing cyber threat intelligence quality. *International Journal of Information Security*, pages 1–18, 2020.

[216] Steven Noel, Eric T. Harley, Kam Him Tam, Michael Limiero, and Matthew Share. Chapter 4 – cygraph: Graph-based analytics and visualization for cybersecurity. *Handbook of Statistics*, 35:117–167, 2016.

[217] Fabian Böhm, Florian Menges, and Günther Pernul. Graph-based visual analytics for cyber threat intelligence. *Cybersecurity*, 1:1–19, 2018.

[218] Steven Noel and Sushil Jajodia. A suite of metrics for network attack graph analytics. In *Network Security Metrics*, pages 141–176. Springer, 2017.

[219] Wei Wang, Rong Jiang, Yan Jia, Aiping Li, and Yi Chen. Kgbiac: Knowledge graph based intelligent alert correlation framework. In *Cyberspace Safety and Security*, pages 523–530. Springer International Publishing, 2017.

[220] Aviad Elitzur, Rami Puzis, and Polina Zilberman. Attack hypothesis generation. In *European Intelligence and Security Informatics Conference (EISIC)*, pages 40–47. IEEE, 2019.

[221] Bipartisan Policy Center. Cyber security task force: Public-private information sharing. *Bipartisan Policy Center Homeland Security Project. Washington DC*, 2012.

[222] Kacy Zurkus. Threat intelligence needs to grow up. Technical report, CSO online, 2015.

[223] Denise E Zheng and James A Lewis. Cyber threat information sharing. *Center for Strategic and International Studies*, 2015.

[224] April Ponemon. Exchanging cyber threat intelligence: there has to be a better way. Technical report, Ponemon Institute Research Report, Ponemon Institute LLC, 2014.

[225] Wiem Tounsi and Helmi Rais. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & security*, 72:212–233, 2018.

[226] Shixia Liu, Wang Xiting, Liu Mengchen, and Zhu Jun. Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, pages 48–56, 2017.

[227] Josina Vink. Storytelling | design research techniques. http://designresearchtechniques.com/casestudies/storytelling/, 2010.

[228] Moshe Ben-Bassat and Amos Freedy. Knowledge requirements and management in expert decision support systems for (military) situation assessment. *IEEE Transactions on Systems, Man, and Cybernetics*, 12(4):479–490, 1982.

[229] Florian Skopik, Giuseppe Settanni, and Roman Fiedler. A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, 60:154–176, 2016.

[230] Yedendra Babu Shrinivasan and Jarke J van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1237–1246, 2008.

[231] Richard Haberlin, Paulo Cesar G da Costa, and Kathryn B Laskey. A reference architecture for probabilistic ontology development. *STIDS*, 2013:10–17, 2013.

[232] Neo4j graph platform. https://neo4j.com/.

[233] Mahesh Lal. *Neo4j graph data modeling*. Packt Publishing Ltd, 2015.

[234] Brian E Ulicny, Jakub J Moskal, Mieczyslaw M Kokar, Keith Abe, and John Kei Smith. Inference and ontologies. In *Cyber Defense and Situational Awareness*, pages 167–199. Springer, 2014.

[235] Nadime Francis, Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Mats Rydberg, Petra Selmer, and Andrés Taylor. Cypher: An evolving query language for property graphs. In *International Conference on Management of Data*, pages 1433–1445, 2018.

[236] Georgios Drakopoulos and Andreas Kanavos. Tensor-based document retrieval over neo4j with an application to pubmed mining. In *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–6. IEEE, 2016.

[237] Apoc user guide 3.5. https://neo4j-contrib.github.io/neo4j-apoc-procedures/#introduction., 2019.

[238] Robert G Abbott, Jonathan McClain, Benjamin Anderson, Kevin Nauer, Austin Silva, and Chris Forsythe. Log analysis of cyber security training exercises. *Procedia Manufacturing*, 3:5088–5094, 2015.

[239] Piotr Juszczak, Niall M Adams, David J Hand, Christopher Whitrow, and David J Weston. Off-the-peg and bespoke classifiers for fraud detection. *Computational Statistics & Data Analysis*, 52(9):4521–4532, 2008.

[240] Bonnie Buchanan. Money laundering—a global obstacle. *Research in International Business and Finance*, 18(1):115–127, 2004.

[241] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.

[242] Toyotaro Suzumura, Yi Zhou, Natahalie Barcardo, Guangnan Ye, Keith Houck, Ryo Kawahara, Ali Anwar, Lucia Larise Stavarache, Daniel Klyashtorny, Heiko Ludwig, et al. Towards federated graph learning for collaborative financial crimes detection. *arXiv preprint arXiv:1909.12946*, 2019.

[243] Robert Pienta, Fred Hohman, Acar Tamersoy, Alex Endert, Shamkant Navathe, Hanghang Tong, and Duen Horng Chau. Visual graph query construction and refinement. In *International Conference on Management of Data*, pages 1587–1590. ACM, 2017.

[244] Robert Pienta, Acar Tamersoy, Alex Endert, Shamkant Navathe, Hanghang Tong, and Duen Horng Chau. Visage: Interactive visual graph querying. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 272–279, 2016.

[245] Yonghui Chen, Gregory Gancarz, and Scott M Zoldi. Visualization for payment card transaction fraud analysis, 2019.

[246] Lynda Ait Oubelli, Yamine Aït Ameur, Judicaël Bedouet, Romain Kervarc, Benoît Chausserie-Laprée, and Béatrice Larzul. A scalable model based approach for data model evolution: Application to space missions data models. *Computer Languages, Systems & Structures*, 54:358–385, 2018. URL https://www.sciencedirect.com/science/article/pii/S1477842418300447.

[247] Carole Nagengast. *Reluctant socialists, rural entrepreneurs: class, culture, and the Polish state*. Routledge, 2019.

[248] Fernandes Diogo and Bernardino Jorge. Graph databases comparison: Allegrograph, arangodb, infinitegraph, neo4j, and orientdb. In *DATA*, pages 373–380, 2018.

[249] Yusarina Mat Isa, Zuraidah Mohd Sanusi, Mohd Nizal Haniff, and Paul A. Barnes. Money laundering risk: From the bankers' and regulators perspectives. *Procedia Economics and Finance*, 28:7–13, 2015. URL https://www.sciencedirect.com/science/article/pii/S2212567115010758.