Cyber Security Visualization Effectiveness

A thesis
submitted in fulfillment
of the requirements for the degree
of
Doctor of Philosophy

at
The University of Waikato

by
Jeffery Garae

Department of Computer Science
Hamilton, New Zealand
April 2018

DEDICATION

This research contribution is dedicated to my Parents and Son - Caleb Garae. Liu.

Abstract

Security visualization utilises predefined data attributes and translates them into visual nodes to form images for the purpose of communicating critical security information to targeted audiences. It is commonly used for two reasons: exploring and reporting purposes thus, sharing insights on suspected security events. However, the challenge of selecting the best visualization out of two or more visualization samples, regardless of existing limitations such as screen dimensions and visual complexities, required users to utilise certain measurement criteria. These criteria urge security visualization researchers, developers and users (viewers) to ask themselves the following two questions: What makes a security visualization effective? How do we measure visualization effectiveness in the context of investigating, analysing, understanding and reporting cyber security incidents?

This thesis explores a range of effectiveness measurement techniques for web and mobile platforms. We investigated existing effectiveness methods for the design, implementation and user observation phases in security visualizations. Consequently, we identified effectiveness criteria and metrics in applications include visual clarity, visibility, distortion rates and user cognitive response (viewing) times. With the goal of aiding decision making in cyber security operations, we provided a distinctive security visualization paradigm of a full-scale effectiveness measurement (SvEm framework) approach for both theoretical and user-centric visualization techniques. Our framework facilitates effectiveness through our SvEm algorithm thus, providing various interactive three-dimensional (3D) visualization applications to enhance both single and multi-user collaboration.

The SvEm framework involves several key components: (1) web/mobile display dimensions and resolution, (2) security incident entities, (3) user cognitive activators and alerts, (4) working memory load, (5) threat scoring system and (6) the colour usage management. To evaluate effectiveness in our framework, we developed several use cases: (1) VisualProgger - a real-time security visualization analytic application (web and mobile platforms), (2) a security visualization with augmented reality and (3) a security visualization for intelligence tracking and monitoring. In addition, we developed and documented a new security visualization guideline (a SCeeVis pre-standard) as part of the SvEm framework to aid with the

design, implementation and observation environments.

This pre-standard further allowed us to develop our SCeeVis colour chaining standard and a new cognition and working memory (SvEm-CWML) instruction set to enhance the user's cognition and perception process for security visualizations. As a result, our visualization application outputs effectiveness measurement by capturing and increasing the user's attention span through the process of reducing cognitive load, while increasing the viewer's memory efficiency. Thus, users have a high potential to gain security insights from a given visualization. Our evaluation shows that, viewers perform better with the existence of prior knowledge of security events and if they operate in a comfortable visual environment. It has also indicated that circular visualization designs attracted and maintained the viewer's attention. Finally, these discoveries have revealed new research directions for future work relating to effectiveness measurement in security visualization.

Acknowledgements

Without doubt this research study and journey was by far one of the biggest challenges I have ever undertaken in life. Its completion is made a reality with continuous hard work and would not have been possible without the following respected people.

First and foremost, I would like to thank my supervisors. Dr Ryan K. L. Ko, for all your guidance, insights, advice, knowledge and support since day one and throughout my journey with this study. Your knowledge and help has motivated me to push through this far within the security domain. I have learnt a lot and grown into understanding why it is critically important to address cyber security events, issues and provide potential solutions. Dr. Mark Apperley, thank you for offering your knowledge, wisdom and time. Your wisdom and guidance have challenged me to think beyond tasks set in front of me. I have learnt to think critically and address all related aspects of this study because of your advice.

Secondly, I would like to thank CROW's from the Cyber Security Researchers of Waikato (CROW Lab). CROW, you stepped in and offered that opportunity for me to try out what I have been dreaming of. To the CROW team, you are my education whanau away from home, especially Mark Will, Alan Tan, Craig Scoon, and Cameron Brown. A journey made possible, easier and manageable because of your presence guys. Thank you, Mark and Alan, for bearing with me with the idea-curve balls thrown at you. I hope those curve balls will one day be an impact in your lives. I would also like to thank Dr. Harris Lin and Dr. Sivadon Chaisiri for your help both academically and socially. Special thanks to Baden Delamore, Dr. Raja Naeem Akram and Kang Du. I am glad that I have met you guys during the journey of this study. The coffee and dinner discussion have always been interesting and certainly have allowed me to observe and view concepts in broader perspectives. Yet, another special thanks to all exchange students (Nanyang Polytechnic and Temasek Polytechnic) whom I have worked alongside with their security projects at the CROW lab. I cannot mention you all, but your help and collaborations with visualization, provenance and attribution related projects has broadened my knowledge and experience in the academic domain. Thank you, guys.

A special thanks to Interpol and the IGCI Innovation and Cybercrime directorate mem-

bers, especially Mr. Noboru Nakatani, Dr. Madan M. Oberoi, Mr. Steve Honiss, Mr. Christian Karem, Mr. Yasuhira Toshinobu, Mr. Costel Ion and Mrs. Margaret Samuel. Your knowledge into law enforcement operations, security and cybercrime has broadened my views on security as a whole.

Thank you Michel Lilliord, Edward William, Russell Mujee, Willie Frank and Adrian Bule for the ongoing support throughout the research journey which started back in 2008.

I would also like to thank my son Caleb Garae Liu. You were and have been the rock of my education journey. Your presence, questions raised and giving up your time to be with me in the lab when you are supposed to be in the park playing games or resting, has provided me the motivation to push on from day one till completion. Your questions have always taken me out of my research head-space and placed myself into your shoes to think like you and explain my work to you in the simplest way possible. To this point, I am not sure if nodding your head means a thumbs-up. Thanks Caleb for your contributions.

To all my families and friends who have helped in a way or another. Thank you all for your patience and understanding with this career path which started many years ago. I would not have gone this far without your help.

To my valuable sponsors - The New Zealand and Pacific Foundation Scholarship (NZAID), and STRATUS (Security Technologies Returning Accountability, Trust and User-Centric Services in the Cloud) (https://stratus.org.nz), a science investment project funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE). - I wish to thank my sponsors for this amazing opportunity. Without your help, I would not have been here pursuing this study, attending academic conferences and research attachments. I am forever thankful for this opportunity. I also wish to thank the Department of Computer Science at the University of Waikato.

Finally with respect, I humbly thank the Government of the Republic of Vanuatu for the strategical opportunity given to pursue this scientific research journey to enhance, build and strengthen Vanuatu's Cybersecurity and National Security.

My sincere thanks and gratitude to everyone for your support!

Contents

List of Figures

List of Tables

List of Publications

1. CORE A Conference: Trustcom 2018 Conference Paper Title - A Full-Scale Security Visualization Effectiveness Measurement and Presentation Approach. Published Date: August 2018. (See pg.11)

2. Cyber Forensic and Security International Conference Paper Title - Security Visualization Intelligence Model for Law Enforcement Investigations. Published Date: August 2018. (See pg.11)

3. CORE A Conference: Trustcom 2017 Conference Paper Title - Visualizing the New Zealand Cyber Security Challenge for Attack Behaviors. Published Date: August 2017. (See pg.12)

4. ICCCRI 2017 Conference Paper Title - Returning control of data to users with a personal information crunch - a position paper (AWARDED ICCCRI-2017 BEST PAPER). Published Date: April 2017. (See pg.12)

5. CORE A Conference: Trustcom 2016 Conference Paper Title: - UVisP: User-centric visualization of data provenance with gestalt principles. Published Date: August 2016. (See pg.12)

6. User-centric Intelligence Visualizations for Ransomware Propagation, Bitcoin Transactions and Early Cybercrime Detection (2017 – Private Publication: International Law Enforcement Digital Security Research Seminar) (See pg.13)

7. Springer Book Chapter Title: - Visualization and Data Provenance Trends in Decision Support for Cybersecurity. Published Date: August 2017. (See pg.13)

8. IET Book Chapter Title: - Security visualization for cloud computing: an overview. Published Date: September 2017. (See pg.13)

Nomenclature

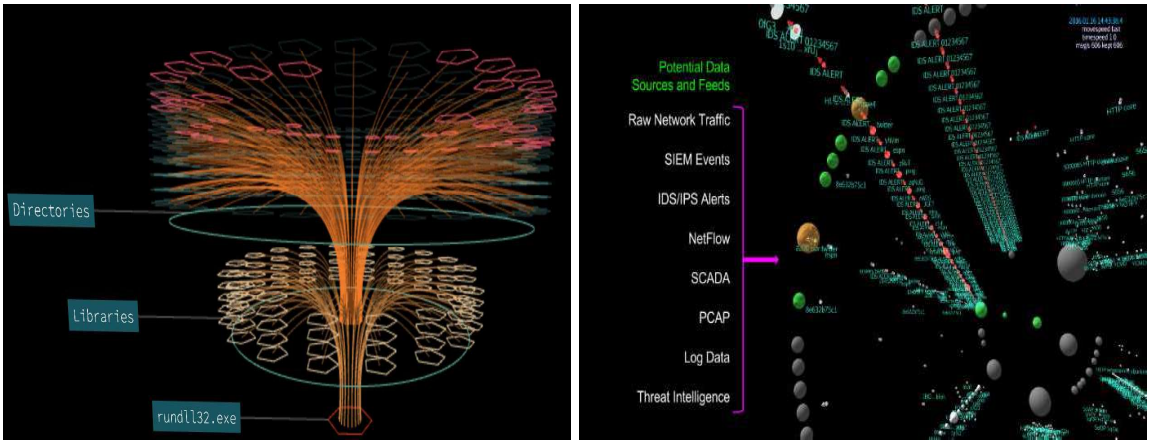
2D	2-Dimension
3D	3-Dimension
AR	Augmented Reality
BGP	Border Gateway Protocol
CCC	Command-Control-Centre
CTF	Capture the Flag
DC-R	Dataset Collection Requirement
DIO	Design, Implementation and Observation
DNS	Domain Name System
DoS	Denial-of-Services
DPaaSVS	Data Provenance as a Security Visualization Service
En	Entity
EnR	Entity Relationship
HD	High Definition
ID	Identity Document
IP	Internet Protocol
ISO/IEC	International Organization for Standardization
IEC	International Electrotechnical Communication
IT	Information Technology
ITU	International Telecommunication Union
IQA	Image Quality Assessment
NZCSC	New Zealand Cyber Security Challenge
PDCA	Plan-Do-Check-Act
QR	Quick Response Code
SvEm	Security Visualization Effectiveness Measurement
SCeeVis	SvEm–Security Visualization Standard
SVaaCS	Security Visualization as a Cloud Service
SvEm-CWML	SvEm–Cognitive and Working Memory Load

Chapter 1

Introduction

1.1 Background

The choice of selecting between two visualizations relies on specific forms of measurement techniques to help rate one over the other. Such ratings help users to choose between two visual images, such as Figure 1.1a and Figure 1.1b [6]. This thesis aims to measure and evaluate effectiveness of security visualization for cyber security operations. However, such objectives require us to ask the following questions: *How can we rate visual images*? *what are the contributing factors affecting the rating on visual images?* and *how do humans contribute to the evaluation of effectiveness in this process?*



(a) Ransomware Activity with VisualProgger **(b)** Deep Analytics Visualization

Figure 1.1: The Choice of Comparing Visualizations

On a day-to-day basis, humans process visual representations of information faster than they are able to read and understand a paragraph of text. They have tremendous capacity to naturally analyse visual information by combining colour gradients into one coherent picture. A picture is a term commonly associated with the arts domain, while science refers

to it as information visualization. '*Information visualization*' refers to visual representations of data [7]. It takes advantage of human perception by enhancing cognitive abilities [8] through the user's visual cortex to process and understand predefined visual information. Visual techniques are used to rapidly locate, discover, identify and compare various types of information. With this in mind, the use of visualization in industries, academia and other domains is becoming the norm for: (1) information exploration, (2) displaying of quantitative data, (3) reporting and (4) envisioning information by communicating potential ideas and insights across to targeted audiences. These are crucial aspects of cyber security operations. Thus, the ability to communicate or share security information and knowledge with the use of security visualization strengthens the entire cyber security operation process.

We begin this chapter by exploring the importance of security visualization and defining 'cyber security' as a key concept for this thesis. The International Telecommunication Union (ITU) [9] [10] defines '*cyber security*[1]' in a lengthy and detailed approach. The Merriam Webster dictionary [11] defines it as "measures taken to protect a computer or computer system against unauthorised access or attack." A clear understanding of what cyber security is and why it is important enables users of the Internet to identify cyber security issues and drawbacks. While the existence of global interconnection capabilities through the Internet offers undeniable advantages, it is a continuous cyber security drawback that brings security challenges. This reaffirms why cyber security is a unique integral component and a growing concern for industries, academia and users of the Internet [12].

Thus, new effective cyber security tools, techniques and standards are high in demand. An ideal solution to confront cyber security issues is by applying security visualization to facilitate information sharing, exploring and aiding decision making. Security visualization is a burgeoning component of cyber security used to identify cyber-threat/attack patterns and behaviours. For example, Figure 1.1b shows deep analytics with visualization, whereby mapping techniques are used to show relationships in network traffic[6]. It is a data representation-based visualization framework whereby all information is consolidated and presented in a single 3-dimensional view. Thus, with security visualization, complex data logs are transformed into simple visual representation forms that are easy for users to understand. We use this example to show how easy visualization is designed and implemented yet challenging if not used in the right manner with the proper audience.

Moreover, security visualization frameworks are useful in cyber security to gaining insights into security events. But how effective are the security visualizations? Do they help

[1] Based on ITU's ITU-T X.1205, cyber security is defined as "collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used to protect the cyber environment and organisation and user's assets." https://www.itu.int/en/Pages/default.aspx

improve cognition and decision making during critical security events in a given cyber security operation? This research presents a framework for measuring effectiveness in security visualization presentations. Our primary focus is improving our Security Visualization Effectiveness Measurement (SvEm) framework [13] by providing a full-scale security visualization effectiveness measurement approach across an entire visualization experience. We assume that users are motivated and competent when interacting with security visualizations. We address data processing performance, visual clarity and user interactive enhancement features for a better user-centric experience. Finally, to measure effectiveness of security visualizations, we require extensive assessments of both web, mobile platforms and their respective user response and reaction times.

1.2 Motivation for this Research

The effective presentation and reporting of security events (e.g., malware attack, data breach) within minimal time required has several challenges. These include: dataset complexity (dataset structure, multiple sources, 'noise' and data uncertainty), various range of user knowledge and preferences, hardware processing power limitations, web browser rendering, scaling, processing limitations and mobile platform screen limitations. Therefore, there is a need for effective security visualization for both web and mobile platform users. This thesis therefore has a core focus on security visualization 'effectiveness measurement' methodologies for both web and mobile platforms. We aim to reduce the time spent that it takes to analyse a given security visualization for precise insights into security events and overall to contribute effectively to the entire cyber security operations time frame.

In order to contribute with effective security visualizations during a security event, we need to understand the goals and strengths of security visualizations, moreover querying the importance of needing security visualization for cyber security operations. Thus, the goal of security visualization is to transform abstract data of security events into computer graphical representations that are easy to understand when reasoning and supporting decision making processes. For example, Minard's carte figurative of Napoleon's 1812 campaign map [14] utilises techniques to convey datasets into a simple and effective infographic representation with attribution [2] [15], provenance [3] [16], [17] and abstract aspects illustrating mostly information (patterns, behaviours, trends, etc.) depicted in the visualization (See

[2] Attribution in cyber security is defined as "determining the identity, source or location of an attacker or an attacker's intermediary." 'Traceback' or 'Source tracking' are common terms used as substitutes for the term attribution [15].

[3] Provenance in the context of this thesis is defined as "a series of chronicles and the derivation history of data on meta-data" [16], [17].

Appendix A1). After careful observation of Minard's visual representation, we can conclude that provenance is well depicted in that particular visual presentation approach. However, what is the key formula to attracting users to see and clearly understand Minard's carte figurative of Napoleon's 1812 campaign map? We argue that visual clarity, data representation simplicity and a user relevant (including motivation, interest, relevant knowledge, etc.) environment are required. Human cognitive '*preattentive processing* [7]' contributes to understanding Minard's carte figurative of Napoleon's 1812 campaign map. It has the ability to effectively and accurately analyse visual information with minimal help required. This example has enabled the user's choice and demand to fully utilise visualizations in daily exploring and reporting routines in cyber security operations.

Apart from exploring and reporting purposes, visualization is found across multiple research domains, industries and amongst end-users. It has contributed to improving decision making processes [18], especially when innovative ideas are conveyed across or discussed with the use of visualization mockups in boardrooms or project meetings. On the one hand, analysts and data scientists use complex visualization techniques to make sense of data collected and to '*connect the dots*' between key findings. However, complex visualization techniques require several analytical steps to acquire potential outputs. End-users, on the other hand, challenge themselves with trying to process and simplify such complex data into a simple-to-understand form with the use of visualization. Web browser capabilities, screen dimensions, hardware processing power and mobility are additional factors affecting how visualizations are portraying intended insights to users. These are pressing factors requiring the need to address effectiveness in standard visualizations and security visualizations.

Limitations in screen dimensions is one of the major factors affecting effectiveness in security visualization. For example, there is a visualization contrasting comparison in mobile platforms (small display size) versus a visualization in 52-inch-high definition (HD) plasma screens with better resolution. This is a challenge that mobile platforms bring when trying to visualize large data volumes. Due to rapid increase in technological services, such as application platforms moving into the cloud, a mobile device is now becoming person's 'digital self' [19]. This makes security a crucial component to a person's physical and online presence. Security visualization is a solution aiding users to keep up with their 'digital self,' practically addressing queries such as: *what is happening to my data?, Who can see my data?, or Can I get an eye-catching alert (notification) of when my mobile has been hacked?* The emphasis of this thesis highlights effectiveness in security visualization for mobile platforms. However, in order to clearly understand the key ideas and context of this thesis, we deliver a set of definitions.

1.3 Definitions

Security visualization in the context of this thesis means a graphical representation of a security event (e.g., malware attack, insider threat activity, data monitoring) associated with a mobile platform/device. Security visualization also refers to visual representations of security data collected in the following scenarios: security monitoring, forensic data extraction, malware analysis, anomaly detection and real-time data tracking for intelligence.

Mobile platforms refer to small screens basically belonging to the mobile devices (smart phones, tablets) and 13 inch laptops or less. Note that the terms mobile platform and mobile device are interchangeable in this thesis. Designs, sizes, shapes and purposes are mobile platform marketing criteria that dictate its usefulness and their contribution to how users use mobile devices.

Measurement in the context of this thesis is the assignment of a number to a characteristic of an object or event. The number assigned can be compared with other objects or events. The scope of a measurement in this thesis has direct association with our proposed security visualization framework.

Effectiveness in the context of this thesis refers to the degree in which something is successful in producing a desired result within the least time spent. The focus of our thesis is on an effectiveness measurement algorithm for security visualization in both web and mobile platforms. Therefore, our thesis deliverable is to provide a successful web and mobile security visualization whereby security insights are acquired by users within the least possible reaction time. In this thesis, we state effectiveness is best measured by results over time with prior knowledge of related security events. Our ideal effectiveness outcome is proportional to visual clarity, simplicity and a viewer's attention span, whereby a high effectiveness measurement rate means higher attention span period.

1.4 Thesis Statement

Our thesis makes two claims: firstly, due to existing visualization complexities resulting from data processing operations and complex visual representations, we provide an effective security visualization approach to improve user interaction with the visualization application used over their mobile platforms. Secondly, existing user preferences, level of security knowledge may hinder the users' ability to efficiently interact with existing security visualization. Hence, this thesis capitalises on understanding user needs, preferences and demands. As a result, we provide effective visual representations, a security visualization standard and a user cognitive instruction set to aid the user's observation environment thus increase their attention span.

The way in which web and mobile platforms are used for security visualization indicates how useful they are. Furthermore, both developers and marketing criteria have strong emphasis on web and mobile designs, display sizes, shapes and user-centric features. These are some of the components affecting security visualizations. On another level, the increasing rate of cyber-attacks in existing systems and networks pose a threat to both web and mobile platforms [20], [21] and applying security visualization to understand these cyber-attacks is an effective technique and approach. User-centric security techniques applied to security visualization frameworks have enhanced user cognition and perception therefore attracting the user's attention. Thus, there are critical demands to develop new security tools and frameworks to provide effective user-centred security detection, monitoring and mitigating of cyber-attacks. While these approach act as security situation awareness towards users, it is important to understand what different targeted audiences prefer when information is shared with them.

The application of security visualization into a user's go-to daily security tool provides the sense of user empowerment and a sense of control. In security visualization, user empowerment requires a well-developed vision that bridges the gap between past, present and predictable future data activities. Effective security visualization leads to discovery of new insights, novel approaches, fresh perspectives and a whole new span of understanding of security events[22], [23]. For example, security visualization for mobile platforms reduces the time taken to analyse complex logs. It provides users with simple, effective, easy-to-process visual outputs that can alert and educate users on what is happening to their system: for example, activities in their online banking application. This thesis addresses difficulties and security challenges users face when trying to comfortably visualize security events over their web and mobile platforms. Therefore, we aim to acquire the most security insights portrayed with effectiveness measurement techniques presented in this thesis.

1.4.1 Problem statement

In order to implement effectiveness measurement techniques and methodologies in security visualizations, we have to understand the problems and challenges within the existing security visualizations. To date, the frustrations and challenges for users to comfortably [24], [25], [26] replicate a complex visualization from a 52-inch, high display (HD), plasma screen on a mobile platform has contributed to the lack of transferring useful knowledge across to viewers. Mobile users [27] occasionally use their desktop machines to see what an intended visualization delivers from transformed preprocessed datasets. There is a lack of effectiveness measurement methodologies for security visualizations. Furthermore, rendering, scal-

ability and visual clarity complexities are additional security visualization challenges mobile platform users face. For example, Figure 1.2 depicts a demotivating factor for users when attempting to use security visualization in mobile platforms. This is the primary challenge for visualizing large amounts (volume) of data in both static or real-time visual approaches. Therefore, the challenge is critical for both web and mobile platforms and requires effectiveness measurement methodologies to aid user choices when confronting a security visualization.

Figure 1.2: A Visual Contrast between Large High Definition (HD) Display vs Mobile Screen [1]

The constraints identified in Figure 1.2 leads to establishing our problem statement:

> Can users effectively visualize security events over their web and mobile platforms in a split-second to help them decide what the next secure step to execute is?
> If '*yes*,' how can we measure effectiveness in security visualization for web and mobile platforms?

1.4.2 Hypothesis

Based on the motivation and problem statement of this thesis, the following are derived as our hypothesis:

Security events can effectively be visualised on web platforms and on small 2D screens of mobile platforms.

To effectively address the thesis hypothesis, the problem statement is further developed to help outline the research direction for this thesis. These are:

- How can we measure effectiveness in security visualization?
  - Can effective measurement methodologies and techniques justify which security visualization is better, given two types?
  - How can we effectively represent and present predefined security data using security visualization?
- How can we evaluate any given effectiveness in security visualization from a platform and user point-of-view?

1.5 Scope

In this research, we have set out to address security challenges encountered by users when using security visualizations. For example, a malware attack explained with the use of security visualization portrays different visual experiences for both web and mobile platform users. The focus of this thesis is on the use of security visualization as techniques/methodologies in aiding critical decision making for both web and mobile platforms. In the context of this thesis, the use of security visualization is not restricted to security events occurring in web platforms but also over mobile platforms. Mobile platform (devices) users are the targeted audience for this research; however, our emphasis is more towards end-users of three categories:

1. End-users with minimal technology knowledge
2. Law enforcement digital crime users
3. Security visualization application (web/mobile) developers

With the prime focus on these three sets of targeted audiences, an intended theoretical outcome of this study is to identify and establish a security visualization effectiveness measurement algorithm. Moreover, an intended user-centric outcome of this study is to identify

a security visualization standard to guide both security visualization developers and users. A security visualization standard provides key security visual attributes that trigger or activate a user's cognition while perceiving security events, therefore increasing the user's attention span. Establishing a user cognitive mindset with security prior knowledge enhances decision-making processes, therefore reducing the time spent on viewing and making sense of a given security visualization. Finally, a practical outcome for this thesis is to implement effectiveness attributes in security visualization to stimulate user interactivity and drive enthusiasm when confronting security visualizations.

1.6 Thesis Contributions

The achievements resulting from this entire security visualization research process are measured with deliverable research contributions within the field of cyber security and security visualization. This thesis makes the following contributions in the order of:

1. Security Visualization Effectiveness Measurement (SvEm) Algorithm:
    - It is possible for both web and mobile platform display dimensions (width, height) and resolutions to control how much information is given to be processed and visualized.
    - User cognition and perception enhanced by 'working memory capacity' with prior security knowledge are the visualization preliminary prerequisites and assumptions during the observation process. They help minimise the time spent analysing a given security visualization with the aim of understanding potential security insights.
    - Establish the SvEm-CWML instruction set to enhance the user's visual thought and observation process.
2. A Security Visualization Standard (SCeeVis):
    - It is possible to implement specific security visualization development guides to standardise visual security event representations and presentation.
    - Adding distinctive user-centric security features for specific security visualization scenarios and use-cases activates cognitive knowledge.
    - Establish the SCeeVis standard colour association rules. This colour association rules utilise the standard set of colours (red, yellow, green and blue) to assist users to visual recognise and follow distinctive visual nodes of interest.

3. Security Visualization Applications:

    - Web and mobile security visualization application proof of concept.
    - Law enforcement user-centric security visualization intelligence application.

1.7 Thesis Structure

Chapter 2 provides the literature review related to this thesis by categorising literature into the following five interested research areas: security visualization, user cognition activation enhancement techniques, user-centric applications methodologies, data processing and rendering techniques and security visualization use-cases. This related research basically helps identify gaps for this thesis area, therefore setting out requirements that are needed which are established in Chapter 3. These primary research requirements are datasets. Chapter 3 describes how datasets are selected and used for this thesis. Various dataset types, collection designs and processes are outlines in this chapter. The New Zealand Cyber Security Challenge (NZCSC) datasets, law enforcement bitcoin transaction dataset and mobile malware datasets are used for this thesis. These datasets are various types depicting multiple attack scenarios, attribution, provenance and tracking landscapes. Additional datasets are used for specific visualization samples within this thesis.

Chapter 4 describes our security visualization effectiveness measurement (SvEm) algorithm. It describes the main components and outlines how they all link together. The practical approach to SvEm consists of the mobile platform specifications and the users' cognitive knowledge prior to confronting any given security visualization. However, for such algorithm to be practical and implemented, rules and criteria with specific assumptions are outlined in Chapter 5. SvEm limitations and problems encountered are provided for use-cases if the required rules and criteria are not met. With specifications, assumptions and limitations outlined, Chapter 5 discusses the SvEm prototype design, implementation technique and the application scenarios (Chapter 5.9) with details outlining specific features. It also provides reason why this chapter is the core component to this thesis by outlining the importance of the SvEm applications. Based on past findings, guidelines and standards have the tendency to assist users with how such applications work, and this is why Chapter 6 is introduced. It presents our Security Visualization Standard (SCeeVis). Chapter 6.1 presents the roles of our security visualization standard and Chapter 6.2 provides an overview of existing standards that are inline with our proposed SCeeVis standard. Our SCeeVis Security Visualization standard is presented in Chapter 6.3. alongside all relevant details describing the functionalities, methods and how the standard enhances effectiveness measurement in

security visualization frameworks. However, to achieve our goals of this thesis, critical steps throughout the SvEm implementation required evaluation and validation. This is discussed in Chapter 7, with an evaluation of our thesis contributions whereby statistical data and several use-cases, including law enforcement evaluations, are delivered. Finally, Chapter 8 provides concluding remarks and future work for this thesis.

1.8 Publications Related to this Research

Throughout the duration of this thesis, key research contributions/ideas were written and submitted to academic security and visualization related conferences and book chapters. These publication titles, citation details and references are in related chapters of this thesis and are outlined in order of publication date. Additionally, book chapter contributions are outlined.

1. Paper Title - A Full-Scale Security Visualization Effectiveness Measurement and Presentation Approach (CORE A Conference Publication):

   A published paper on the core component of this thesis research i.e. Cyber Security Visualization Effectiveness.

   - Citation details are as shown:

     J. Garae, R. K. L. Ko and M. Apperley, "A Full-Scale Security Visualization Effectiveness Measurement and Presentation Approach," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 639-650. doi: 10.1109/TrustCom/BigDataSE.2018.00095

2. Paper Title - Security Visualization Intelligence Model for Law Enforcement Investigations (Cyber Forensic and Security International Conference (CFSRC 2018) Publication):

   A paper published and presented on security visualization for Intelligence model in law enforcement investigations.

   - Citation details are as shown:

     Garae, J., Ko, R. K. L., Apperley, M., & Schlickmann, S. J. (2018). Security visualization intelligence model for law enforcement investigations. In B. Cusack, & R. Lutui (Eds.), Proc 2018 Cyber Forensic & Security International Conference (2018 CFSIC) (pp. 165-177). Conference held in Tonga.

3. Paper Title - Visualizing the New Zealand Cyber Security Challenge for Attack Behaviors (CORE A Conference Publication):

   A published paper on how user-centric features are added to real-time security visualization to enhance user interaction and situational awareness around security events. Parts of this work are drawn from Chapter 5 - sections 5.6 and 5.7.

   - Citation details are as shown:

     Garae, J., Ko, R. K. L., Kho, J., Suwadi, S., Will, M. A., & Apperley, M. (2017). Visualizing the New Zealand Cyber Security Challenge for attack behaviors. In Proc 16th IEEE International Conference on Trust, Security and Privacy in Computer and Communications (pp. 1123-1130). Sydney, Australia: IEEE. doi:10.1109/Trustcom/BigDataSE/ICESS.2017.362

4. Paper Title - Returning control of data to users with a personal information crunch - a position paper (Conference Publication: AWARDED ICCCRI-2017 BEST PAPER):

   A published paper delivered to the International Conference on Cloud Computing Research and Innovation (ICCCRI). Key ideas presented are on the need to return control of data to users, particularly at a personal level. Parts of the work in this paper contribute to the identification of sensitive areas revolving around users, which contributes to Chapter 2 which is the thesis background and literature review section.

   - Citation details are as shown:

     Will, M., Garae, J., Tan, Y. S., Scoon, C., & Ko, R. (2017). Returning control of data to users with a personal information crunch - a position paper. In International Conference on Cloud Computing Research and Innovation (ICCCRI).

5. Paper Title: - UVisP: User-centric visualization of data provenance with gestalt principles (CORE A Conference Publication):

   A published paper delivered to the 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Trustcom 2016), detailing user-centric features, data provenance and provenance visualization with gestalt principles. This work largely contributes to Chapter 2 and to Chapter 6 which is a core chapter of this thesis.

   - Citation details are as shown:

     Garae, J., Ko, R. K. L., & Chaisiri, S. (2016). UVisP: User-centric visualization of data provenance with gestalt principles. In Proc 15th IEEE International Conference on

Trust, Security and Privacy in Computing and Communications (pp. 1923-1930). Tianjin, China: IEEE Computer Society. doi:10.1109/TrustCom.2016.0294

6. Paper Title: User-centric Intelligence Visualizations for Ransomware Propagation, Bitcoin Transactions and Early Cybercrime Detection:

   A Private Publication: International Law Enforcement Digital Security Research Seminar (2017). Held at the INTERPOL Global Complex for Innovation Centre in Singapore.

7. Book chapter Title: - Visualization and Data Provenance Trends in Decision Support for Cybersecurity (Book Chapter Publication - August 2017):

   A published book chapter in 2017 outlining visualization and data provenance trends aiding decision support for cybersecurity. This chapter contributes to Chapter 2, Chapter 6 and provides a benchmark for our Chapter 7 evaluations.

   - Citation details are as shown:

     Garae J., Ko R.K.L. (2017) Visualization and Data Provenance Trends in Decision Support for Cybersecurity. In: Palomares Carrascosa I., Kalutarage H., Huang Y. (eds) Data Analytics and Decision Support for Cybersecurity. Data Analytics. Springer, Cham. https://doi.org/10.1007/978-3-319-59439-2_9

8. Book chapter Title: - Security visualization for cloud computing: an overview (Book Chapter Publication - September 2017):

   A published book chapter outlining the contributions that security visualization is bringing into the cloud computing environment and outlines methodologies used in Chapter 4, Chapter 5, Chapter 6 and Chapter 7.

   - Citation details are as shown:

     Garae J., Ko R.K.L., Apperley M. (2017) Security visualization for cloud computing: an overview. In: Kumar V., Ko R. K. L., Chaisiri S. (eds) Data Security in Cloud Computing. The IET.

Chapter 2

Background and Literature Review

This background chapter sets two approaches: (1) identifying existing effective security visualization solutions and (2) identifying current user effectiveness methods in security visualization. From a technical approach, screen sizes, hardware specifications, distortion, visual clarity, and dataset complexities are the factors impacting effectiveness in security visualization [24], [25], [26]. On the other hand, the user's (viewer) approach has the following core factors contributing to effectiveness in security visualization [24]: (1) motives, (2) cognitive/perception capabilities, and (3) working memory (prior knowledge). Both technical and user approaches are discussed thoroughly in Section 2.2 through to Section 2.9. However, we establish the baseline for this research by providing a visualization overview outlining how visualization evolved. Moreover, we provide a visualization genealogy to show an overview of the visualization domain along with a thorough literature review to help understand existing literature and the concepts that make visualization effective and interactive. These concepts included assessing effectiveness in visualization tools, visualization implementation methods and correlation rankings in visualizations. However, it is important to understand the use of visualization throughout history to assess how effective it is in respect to different aspects regarding its application.

2.1 A Visualization Timeline

Visualization is vital to everyday life in a society, ranging from a child's visual awareness, education, research, public health and even for the law enforcement sector [28]. Visualization is both *art* and *science* [7]. History sees it as art, while research calls it the 'science of art, namely visualization.' In pre-17th century [29], as depicted in Figure 2.1, early maps and diagrams were commonly used to interpret abstracts of data and have shown great advantages of applying information visualization. At this stage, the terms '*information visualization*' and '*data visualization*' are often interchangeable. In a broader view, information visualiza-

tion is seen as predefined visual presentations of pieces of data. This takes us back in time to the earliest scratches of all forms on rocks to the development of *pictoria* as mnemonic devices in illuminated manuscripts. Another example involves the use of diagrams in the history of science and mathematics which dated back in the 1800s [30]. The use of diagrams led up to the rebirth of data visualization in the late 19th century. In 1975, data visualization evolved with the growth of data therefore requiring new presentation methods. This introduced high-dimension interactive and dynamic data visualization. Thus, we provide a visual history of how data visualization [29] evolved over the past years in Figure 2.1 to help understand the impacts affecting the use of visualization.

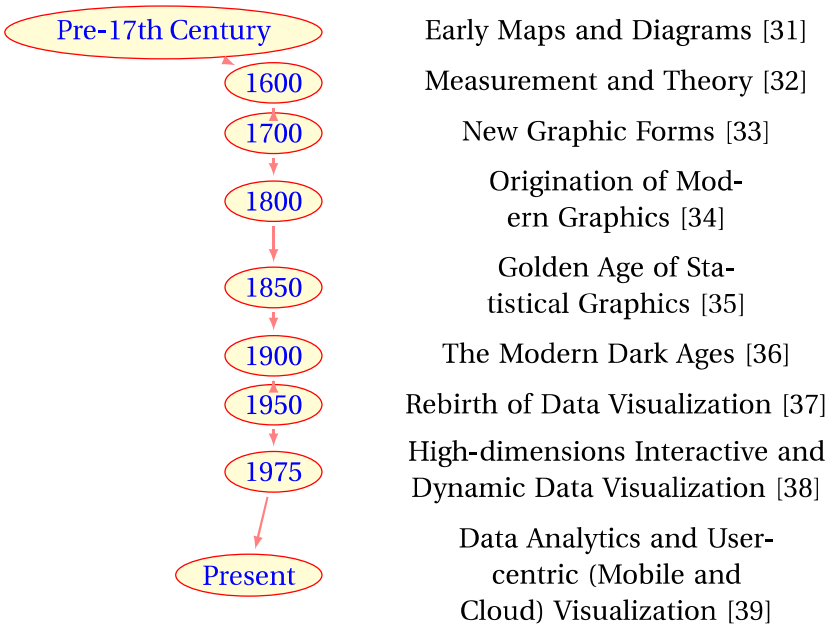


Figure 2.1: A Timeline of Visualization History

The modern visualization context introduced the high-dimensional view and interactive data visualization, whereby large datasets are transformed and presented. These presentation methods have been adopted and became the basis of security visualization. However, there are various reasons and purposes for visualization in the security research domain. Traditional security visualization research and solutions were founded upon information assurance, intrusion detection, general network traffic and host processes [40], [41], [42]. Information exploration and reporting of data abstracts were common in security visualization. Since 2011, holism, scaling and zooming techniques gained popularity in visualization by adding user experience features. However, user-centric challenges have revolved around the need to present data findings in a complete (*parts of the whole*) form, which has al-

lowed room for further visualization research [43], [44]. With an increase in global network connectivity, security visualization has gradually added new visual presentations based on enhancing hardware features, such as: emphasis on Border Gateway Protocol (BGP) anomalies, DNS traffic, malware network traces, routing anomalies, and IP flows [45], [46], [47], [48]. These network-centric visualizations with further designed insights along the years improved how visualization is used to show real-time web attacks and behaviour-based malware with the use of log files. These security visualizations are used to study and understand attack types, patterns and behaviours of malicious events [49], [50][51]. The notion of using security visualization to understand cyber-attacks and threats showed a change from exploratory to a more behavioural analytic visualization trend and environment whereby precise crucial information obtained from the data presented makes an impact on the next mitigation step and decision. This entire analytics and visualization requirement approach led to a whole new understanding of visualization in general and security visualization in particular. We present a visualization genealogy in Section 2.2 to help explain requirements, purposes, methods, techniques, prototypes and representation/presentation forms.

2.2 Security Visualization Genealogy

We begin with an introductory review of security visualization to establish background knowledge on the importance of visualization in security. In the context of digital and cyber security, visualization allows researchers to understand cyber-attacks and threats [52], [53]. Security visualization is used for '*exploring*' data to gain insights and '*reporting* on the findings [54] acquired. However, in a specific user-centric approach, visualization has two primary purposes: (1) communicating an idea using predefined pieces of information across to an audience and (2) seeking to understand pieces of data through the exploring of data [55], [56] and discovering new ideas. For example, tools and techniques such as SEEM [57] and DAVAST [58] have allowed security analysts to identify and address issues pertaining to specific cyber-attacks. These approaches are achieved with the application of data analytics, whereby users interactively query pieces of data through a visual representation platform. We developed a basic visualization genealogy to show how security visualization has been used over time with respect to purposes, classifications, representation and presentation methods and views. These features affect interactivity and effectiveness in security visualization platforms. The genealogy highlights important features and concepts as shown in Figure 2.2, and reflects existing user-centric methodologies in security visualization.

As part of the security visualization genealogy, we proposed the following features, approaches and concepts that contribute to security visualization in the following ways:

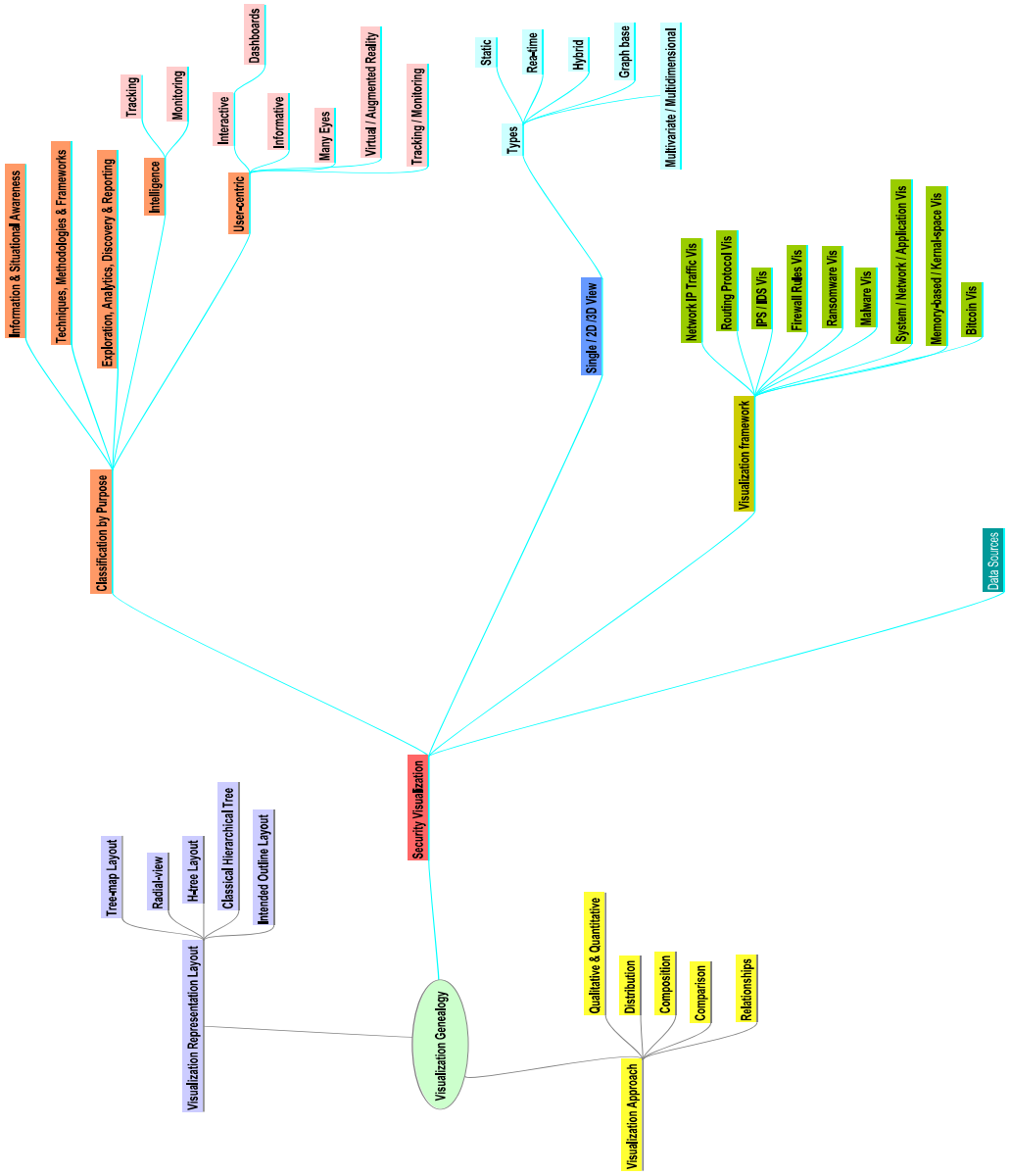


Figure 2.2: A Security Visualization Genealogy Approach

1. *Visualization representation layout*: the representational layouts provide the ability to tell a story, show patterns, behaviours and relationships in a simple and consistent manner in visualizations.

2. *Visualization approach*: these approaches provide an overall view of what is expected in visualization, i.e., presenting quality or large (volume) data, revealing data distribu-

tion, illustrating data content, comparing data attributes and viewing data relationships.

3. *Single/2D/3D view*: the views refer to the n-dimensional visual presentation plane, i.e., single, 2-dimension and 3-dimension view.

4. *Visualization framework*: these are some common examples of existing security visualization frameworks and tools.

5. *Classification by purpose*: the 'classification by purpose' shows various security visualization purposes and aims to address the question of *why use security visualization*?

With various data sources and dataset intentions, providing a classification of security visualization is a challenging task. Therefore, we began by associating security visualization with its core purpose in a way that it relates to this thesis. Across multiple research domains and industries, visualization is being used for distinctive purposes. We design and proposed a quick and easy way to view how visualization has been used in the past is shown in Figure 2.3. It basically classifies existing visualization according to these four purposes: *distribution, composition, comparison* and *relationships*. It also provides a basic classification according to the visualization trends over time. These main groups are: (1) the dashboards/graphs [59], (2) collaborative [60] visualization, (3) multivariate/multi-dimension [61] visualization and 3D/virtual and augmented reality [62], [63] visualization. Moreover, there are vast number of visualization representation designs presented in the visualization field.

The existence of this visualization genealogy (Figure 2.2) and classification (Figure 2.3) established the need to review past literature in security visualization, with an in-depth understanding of existing research gaps required. Therefore, the review's primary focus is in the following areas: (1) effectiveness measurement techniques, (2) security visualization types and purposes, (3) classification of existing security visualization research by purposes, (4) existing user-centric visualization examples, (5) security visualization for law enforcement applications, (6) information security and (7) user cognition and perception requirements in security visualization.

2.3 Identifying Existing Effectiveness Measurement Techniques

One aspect of providing user-centric visualization in the security domain is to identify 'user-attention-trigger' variables, such as: effectiveness features, techniques and attractive interactive features. These user-attention-trigger variables are accessed and assessed by users

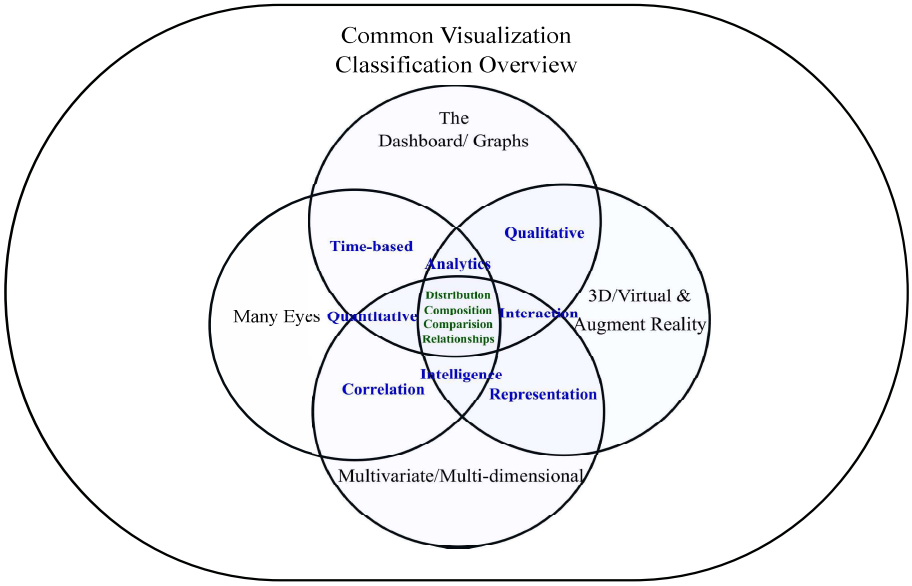


Figure 2.3: A Common User-centric Visualization Classification Overview with Relationships

to aid them in making better decisions when using security visualizations. Therefore, we surveyed existing effectiveness measurement techniques and provide a summary on our findings shown in Table 2.3 and in Figure 2.4.

Effectiveness Measurement Factor	**Measurement Range (Quantity)**
Cognitive Load [64]	High (Germane/Intrinsic /Extraneous Cognitive Load)
Working Memory (Prior Knowledge) [64]	High (Affects Cognitive Load)
NASA-TLX Test (Indirect-Work Load) [65]	Medium (Based on Work Load)
Subjective Workload Assessment Technique (SWAT) [66]	Medium (Scale Rating Factors - Mental Effort)
Image Quality Assessment [2]	Medium - High (Based on Distortion)
Eye Tracking (CODE Theory of Visual Attention) [67]	High (Eye Movement Based on Information Theory)
Brain Activity [68]	High
Response Time on Task (s) [64]	Low (Depends on Prior Knowledge and Effort)
Effort/Difficulty Rating [64]	Low (Based on Insights)
User Interactions/Performance [64]	Low (Based on Naive Physics; Body Awareness and Skills; Environmental Awareness Skills; and Social Awareness and Skills)
Visual Perception Metrics (Visualization Efficiency) [69]	Low (Based on Graphical Methods, i.e., similarities)

Table 2.3 highlights past work done across the computing science and psychology re-

search domains. These common research areas in visualization effectiveness are: (1) understanding cognitive load and working memory contributions, (2) image quality assessment, (3) eye tracking, (4) brain activity monitoring, (5) user interaction/performance and response time on tasks, (6) effort and difficulty ratings and (7) visual perception metrics for visualization efficiency. Although we have reviewed this research area thoroughly, certain areas of interests which directly linked and affect our research are discussed in Subsection 2.3.2, Subsection 2.3.3 and Subsection 2.3.4.

2.3.1 Security Visualization Effectiveness Issues and challenges

The challenge researchers encounter when designing and implementing security visualizations is the ability to maintain a security mindset to deliver effective security visualizations. Current researchers focus on the visualization presentation with information overload and visual features that result in cognitive biases. Thus, existing security visualizations are specific with a primary focus on achieving the goals of using visualization. At times, the lack of understanding the nature of the data by users, leads to misrepresentation and misinterpretations of the data in security visualization. This is a challenging issue for many researchers and developers. Figure 2.4 displays existing visualization effectiveness variables (highlighted in grey) and shows basic comparisons between information visualization, data visualization and security visualization. These comparisons highlight existing work in all visualization areas; however, we mapped out existing research gaps specifically for security visualization, as highlighted in yellow within Figure 2.4.

2.3.2 Correlation Effects on Ranking Visualization

Lane Harrison, Fumeng Yang, Steven Franconeri and Remco Chang have leveraged on 'perception laws' to quantitatively evaluate effectiveness of visualization designs [70]. This led on to the application of Weber's [70] law on the aspect of correlation of ranking visualizations, which has established methods around how visualization designs are affected when perception laws are applied. The ranking of visualizations by correlation techniques had enabled researchers to access statistical data that helped provide facts on certain problems and tests. However, the correlation of ranking concepts required that users (viewers) compare between two objects, ideas, and presentations to achieve an effectiveness rating.

2.3.3 Image Quality Assessment

In image quality assessment (IQA), error sensitivity measurements are applied to natural image complexity problems[2], for example the use of natural patterns such as spots, lines

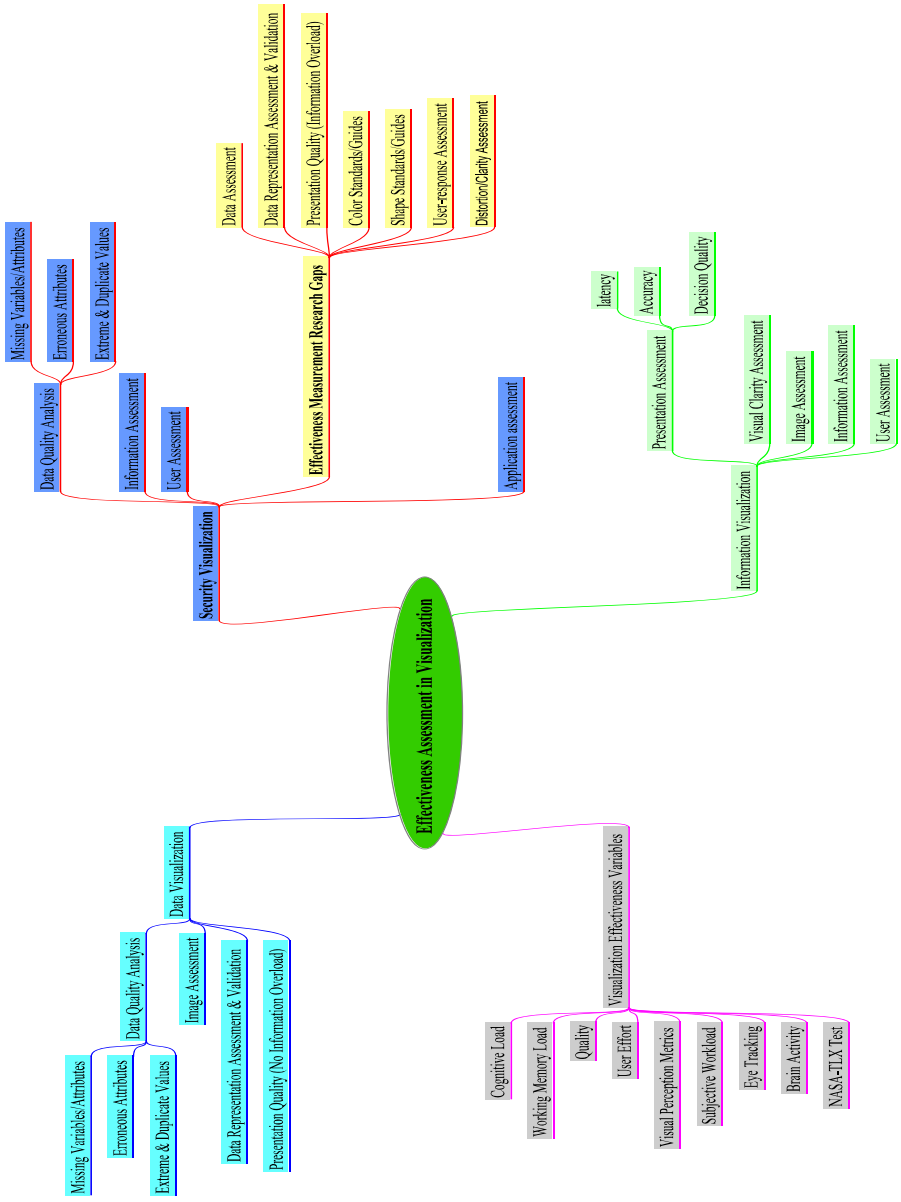


Figure 2.4: Effectiveness Measurement and Assessment Research Gaps

and bars to show errors rates in images. Visibility in images is assessed with respect to the distortion rating, based on human perception, which is another method used in error sensitivity assessment. This means that assumptions are established to provide better scope into how images can be assessed. Measurement of visual quality [71] in image and video processing applications with respect to perceptual space are facilitated with image quality

assessment (IQA) algorithms.

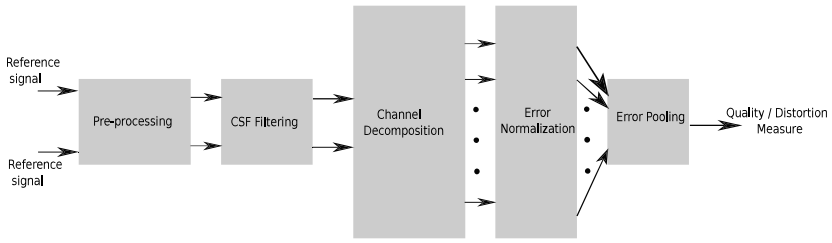


Figure 2.5: An Image Quality Assessment Prototype System Based on Error Sensitivity [2]

These techniques allow automatic assessment of image or video quality in a consistent manner. A major contributor to such techniques is human perception [71], which attempts to achieve consistency for physiological and psychovisual features with high quality predictions in human visual systems.

2.3.4 E^3: Expressiveness, Efficiency and Effectiveness

'E^3: Expressiveness, Efficiency and Effectiveness [3]' by Ying K Leung and Mark D Apperley addresses graphics presentation techniques. The E^3 framework delivers the base for comparison of various presentation techniques. It emphasises three graphical data presentation aspects primarily as metrics to facilitate objective assessment of the technique used. These presentation aspects are: (1) Expressiveness, (2) Efficiency and (3) Effectiveness. This framework addressed problems of accessing large data sets for limited display surfaces. Fig. 2.6 shows the E^3 design framework on how data is presented with respect to the data space. The E^3 framework has the following components: information contents (I, I_1, I_2), data sets (original set (S), abstract set (S_1), data presented set (S_2)), set of presentation techniques (P), set of representation techniques (R), and the set of interactive tasks (T).

The E^3 framework focuses on large datasets with the notion of presenting data sets in abstract levels, with the use various techniques and interactive tasks. This concept of visual presentation in terms of expressiveness, efficiency and effectiveness enables an assessment of techniques against surface areas. Building on and utilising the E^3's concept, our thesis extends the focus into effectiveness measurement by reducing perception time spent (ef-

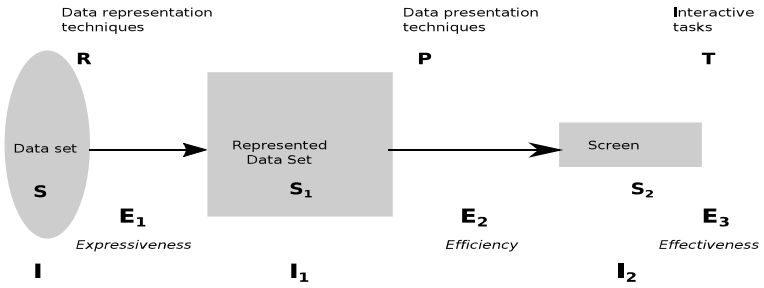


Figure 2.6: E^3 Design Framework: Stages in the Presentation System for Large Data Sets [3]

fectiveness) with simplified web and mobile security visualization methods. This includes the use of the SvEm algorithm, the use of colour application within visual interaction methods and the overview/cluster presentation approach for both web and mobile platforms. This leads to the need to understand different security visualization types and purposes to aid our research objectives.

2.4 Exploring Existing Security Visualization Types and Purposes

Security visualization expands across multiple domains for various purposes [7], [72]. For example, VISUAL [73] is designed to enhance network administrators' capabilities to monitor and understand network activities [54]. Adrian Perrig and Dawn Song developed 'Hash Visualization: Random Art [74],' a visualization technique used to improve real-world security. Their solution aimed at addressing human limitations, namely: (1) the ability to remember strong passwords and personal identification numbers (PINs) and (2) the difficulties of comparing meaningless strings. Other areas of approach included data management and the need to aid scientific visualization applications [75]. Capitalising on graphical user interfaces, boxes and arrows are used as indicators for database access with the aim of showing hierarchies of data abstracts.

The identification and attribution of cyber-attacks are made known through the use existing security tracking and monitoring techniques/tools. However, cyber-attacks evolve every time they are executed. This issue gives researchers the continuous need to design and im-

plement smarter reliable security tools, methods and techniques. Visualization is a promising method used to understand data activities that empower users to observe and interpret vulnerabilities and cyber-attacks [76] efficiently. In this thesis, we use the security visualization community (VizSec) case study to review and assess the work related around the types and security visualization purposes.

VizSec[77][1] began in 2004 by addressing research areas such as cryptography, encryption methods/techniques and network attacks (DoS attacks) with the intention of addressing security for privacy and trusted environments [79], [80], [81], [82]. These types of security visualizations were based on 2D/3D graphics [42]. They were intended for security researchers, with the sole purpose of providing educational information/situational awareness, visualization techniques and data analytics [83]. Due to the growing interest in visualization development, more security visualization researches were proven to be useful and have provided continuous insights to security related fields. We will now traverse through the VizSec research visualizations to identify the promising examples of novel innovative visualization researches and identify their strengths and weaknesses.

Fast-tracking into VizSec's history, the security visualizations were founded upon information assurance, intrusion detections, general network traffics and host processes. This boosted the invention of new network products with logging mechanisms implemented in them. Such features allowed the introduction of security visualization into hardware-based products. Visualization became an added feature/service to understanding the network logs collected, for example, visualizing BGP anomalies, DNS traffics, malware network traces, routing anomalies, and IP flows [45], [46], [47], [48]. With further insights along the years, security visualization provided visual views for web attacks, Log files and behaviour-based malware patterns [49], [50], [51]. This revealed a change from using visualization for exploring purposes to a behavioural understanding-based visualization purpose and trend.

The introduction of behavioural understanding-based visualization required that user-interfaces (UI) change to provide interactivity. User-interface (UI) based visualization developed to empower users by giving them control over tools with tasks such as filtering, mapping and multidimensional colour lookups. These are performed according to the users' needs and preferences [84]. Such tools were developed to address intrusion detection. Komlodi et al. [84] believed that allowing users to customise their display from simple or high-level (3D) over-views would support monitoring, analysis and diagnosis tasks. Their ap-

[1] "*The IEEE Symposium on Visualization for Cyber Security (VizSec) is a forum that brings together researchers and practitioners from academia, government, and industry to address the needs of the cyber security community through new and insightful visualization and analysis techniques. VizSec provides an excellent venue for fostering greater exchange and new collaborations on a broad range of security- and privacy-related topics [78]." Link: http://vizsec.org/*

proach has distinctive advantage due to the evoloving technology as time lapse. As a result, their use of customised displays, have allowed users to interact with visualizations.

Nowadays, mobile technologies are the go-to technology, whereby security visualization with user empowerment features are in critical demand. However, mobile security visualization is a current challenge for researchers, developers and users. These are due to hardware limitations, data volume challenges, rendering and scalability issues. Therefore, addressing these challenges by providing effective user-centric methods is the goal for this thesis. Although there are existing web and mobile visualizations, a potential effective security visualization approach is to develop three-dimensional approaches with 'simplicity' as the focus. These visualizations allow administrators and security analysts to analyse and understand systems and network events efficiently, often at any given time of the day [42]. Three-dimensional approaches provide multiple views thus allowing users to understand datasets and network logs better. This requires the need to design and develop security visualization frameworks from a more realistic hybrid architecture approach whereby both cloud and laptop/mobile platforms are used to provide a simple yet effective security visualization.

Finally, assessing various types of visualization over the past years indicated that most visualization in the security research domain revolves around specific purposes and types. The visualizations are specific with regards to the nature of the security event, i.e., malicious events.

2.5 Classification of Existing Security Visualization Research

In this section, we deliver a thorough literature review on past security visualization research released under the security visualization community (VizSec). Our emphasis for this review is to identify several aspects of security visualization and see the purpose of visualization in various applications. In addition, we observed how effectiveness in security visualization is applied. Therefore, we created a '*checklist*' and accessed a total of 166 papers of VizSec visualization research and categorise them. The assessment covered work done in security visualization since 2004 through to 2014. The idea behind having a check-list is to classify according to '*visualization requirement,*' such as purposes, attack type and data types. The checklist and its content are shown in Table 2.1:

The 'attack type' attribute on the checklist dictates the purpose of the visualizations, for which the right data types and data sources are identified to produce visualizations. However, because the attack types have been generic and standard throughout the past decade of the VizSec research, the 'purpose' attribute is used primarily to observe the trends of se-

Table 2.1: VizSec Survey Checklist Requirements

Checklist	Purpose	Attack Type	Data Type
✓	Info/Sit Aw	IDS	Net-Traffic
✓	Tech/Met/Fra	Malware	Log Files
✓	Exp/Ana/Dis/Rep	Net-Traffic	-
✓	Track/Mon	Host Pr	-
✓	User-centered	Ac-Control	-

curity visualization over the years.

The results in Figure 2.7 indicated all 166 VizSec papers were classified into various visualization purposes. Over the past years, visualizations implemented for the purpose of 'analysing' datasets rate the highest, at (39%), while visualizations providing information or situational awareness rate at (25%). These types of visualization range from 2D,3D and tree-mapping visualizations. However, there are very little security visualizations for end-users which were rated at (4%). Visualization for monitoring purposes is rated at (8%). The overall classification shows the need to research and develop security visualizations directed towards end-users and monitoring purposes. The results also indicated that since there is a low rate of visualization for end-users, a theoretical observation by default states that effectiveness in security visualization has a low rating as well.

Figure 2.8 provides a detailed trend of how security visualizations over the past years were carried out. It shows very interesting findings, such as throughout the past years, security visualization for analysis and reporting purposes has maintained its popularity. However, security visualization for information and situational awareness has reduced as we advance into the technology era.

Identifying different types of attacks requires 'User-centred' and 'Monitoring (Mon)' visualizations, which have been introduced into the VizSec research community[85], [86]. 'User-centred' in the context of this survey refers to products, frameworks, tools and systems which are tailored towards the 'end-user,' i.e., users who have little or no knowledge at all on a product, framework and tool. Forensic investigations [87], [88], [53] use such visualization techniques to analyse and report on their findings. Network attacks and host processes have also been visualized as part of VizSec. The purpose of applying security visualization to such an approach is to analyse host traffic and check for possible attacks. Apart from the

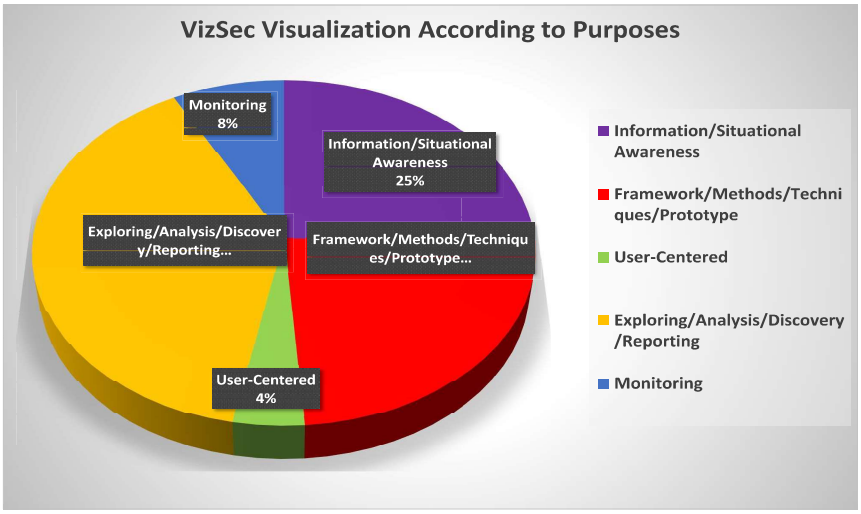


Figure 2.7: VizSec Visualization Classification

mentioned security visualization researches, visualizing distributed memory computations is also an area of important research in the VizSec community [89].

The findings of this survey have also identified some visualization dependencies, which are illustrated in Figure 2.9. These dependencies include: datasets collected and used; specific research purposes and theme; fundings allocated; and most importantly the security research priorities within the year of publication. Other technical dependencies affecting the outcome shown in Figure 2.9 is due to the fact publications discussing security techniques, methodologies, and framework approaches affect the purpose of providing a security visualization, i.e. visualizing for exploration of security data; analytical visualization; observatory visualizations; and utilizing visualizations for reporting of security findings. Moreover, if better multiple techniques/methods are deployed for such security visualizations, there are higher results of exploration, analysis and discovery from the security visualizations. This means reporting will be detailed and meaningful with possible new insights.

2.5.1 VizSec Visualization Classifications against Past Surveys

Past research surveyors of the VizSec community were evaluated against HCI research, system designs and visualization research in general. Others have evaluated security visualization against the application of visualization in general to explore and report on the data being analysed. For example, high-dimensional visualization types (2D/3D) have contributed

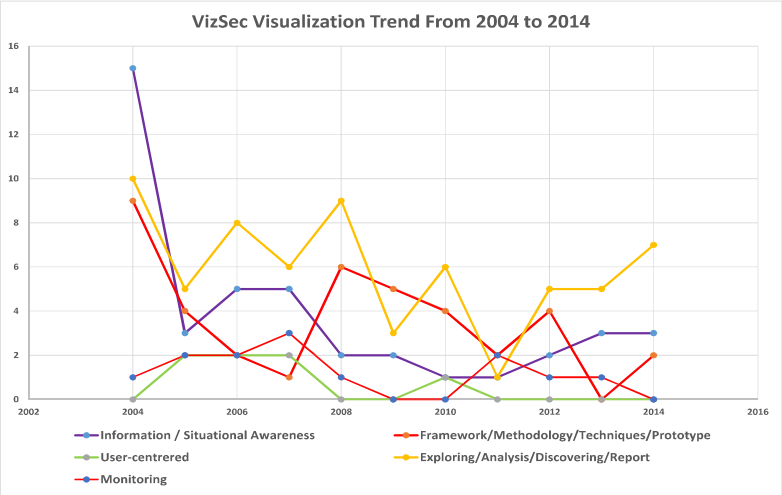


Figure 2.8: VizSec Visualization Trend by Purposes from 2004 to 2014.

massively in making security visualization successful [90], [91]. High-dimensional security visualization has addressed scalability issues in visualization for comparing multiple large datasets [34]. This evaluation method categorises the security visualization research into classification of dimensions, components and user type. We see this survey as information/situational awareness [92], [93], [94], [76]. Another past VizSec assessment had addressed seven challenges faced in previous security visualization research [92], which indicated the lack of effective security visualization and effectiveness measurement techniques.

Moreover, the use of security visualization began addressing research areas mostly in cryptography, encryption methods/techniques and network attacks (DoS attacks,) as a form of security approach for privacy and trusted environments [79], [95], [80], [81], [82]. The security visualizations were based on 2D/3D graphics and targeting information/situational awareness and in general were applying specific techniques and analytical methods, whereby the aim is to educate security researchers on potential security events [83].

The ability to provide information and situational awareness using visualization increased the need explore, analyse, and discover security vulnerabilities, malicious patterns and behaviours over the network layer. [96], [97], [98], [99], [100], [101], [102], [103]. This is due to more attacks targeting the network layer and the web. Such attacks over the network, has driven the security visualization community to focus on Intrusion detection systems. They were the common tool implemented data analytic capabilities and visualization fea-

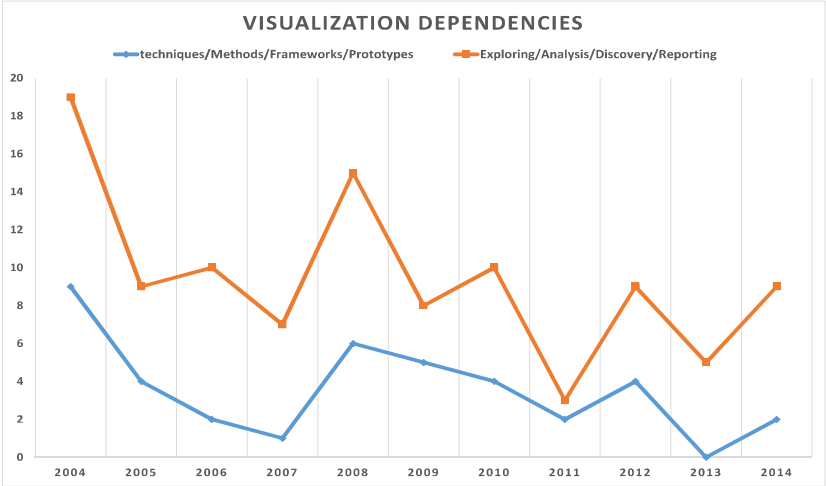


Figure 2.9: Security Visualization dependencies from 2004 to 2014.

tures to monitor network traffic for malicious and cyber-threats. These types of tools are implemented for specific reasons and for security experts to enhance their investigations on possible threats over their networks. The added visualization features in intrusion detection tools have allow security experts to process security information as a fast time frame. With such demands to produce better visualizations, the time or duration factor, was an aspect that led security researchers and developers to address it and offer specific yet simple visualizations that would aid security professionals in their decision-making process.

2.5.2 The Time-based Visualization Approach

Generally, security visualization serves as a powerful approach to monitor, detect/discover and analyse security related events [104]. The primary aim in security visualization surveyed over the past years was addressing the following areas of concern: the motivation behind the visualization implemented; the problem addressed with the use of visualization; who the targeted audiences are; the method and techniques used, and the type of data needed for visualization [105], [106], [107]. In Figure 2.10, we show how security visualization is used over a period of time with two assessment attributes: the attack types and attack landscape displayed with the use of circles. It is a time-based visualization utilising the use of colors and circle sizes to represent different attack attributes and attack statistics. In addition, most security reports require timestamps when reporting on certain cyber-attacks, particularly

when observing how an attack behaviour transitions over time. For example, time-bases security visualization is useful for visualizing cyber-attack histories, real-time security and provenance related events.

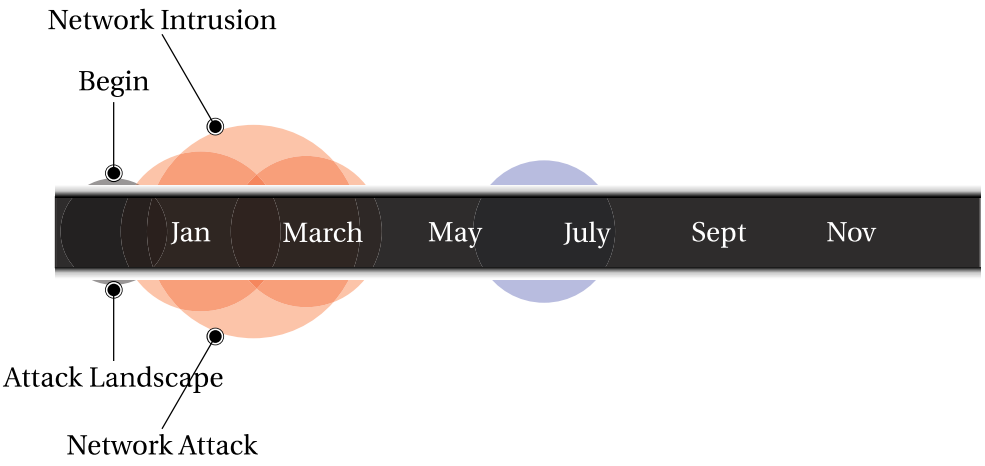


Figure 2.10: A timeline visualization sample

2.5.3 The Mobile Visualization Approach

The final section of our VizSec review, investigated further into existing mobile visualization. As seen across the visualization research and industry domain, modern visualization platforms are either web-based or mobile-application based. However, most web-based visualization platforms and frameworks are mobile platform compatible. Tableau [59], a web-based dashboard visualization, provides both web and mobile visualization experiences for data analytics. The Deepvis mobile application as shown in Figure 2.11, uses data analytics, particularly machine learning algorithms, to provide visualization clusters of data collected. It provides the ability to show clusters of data with the use of classification methods identify and distinguish certain attributes. Apple's mobile fitness application uses inbuilt sensors alongside data collected to provide monitoring statistics when doing exercises.

Small Screen Display Visualization

Small screen sizes, limited resolutions and hardware processing power limitations [108] are the challenges mobile platforms encounter when designing and implementing visualizations. Processing large datasets while attempting to visualize it is an additional challenge

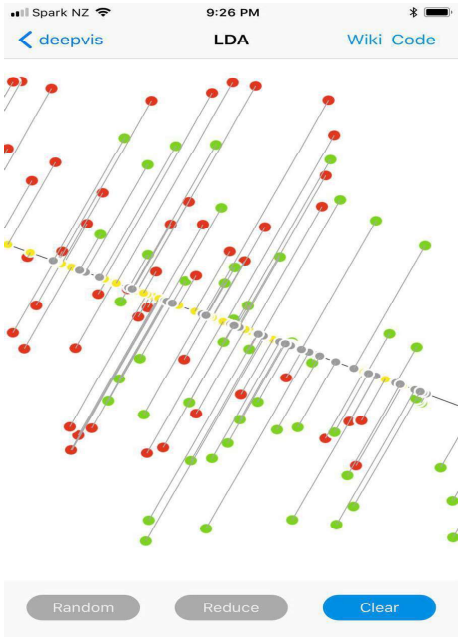


Figure 2.11: Deepvis Mobile Visualization showing Data Clusters

for small screen displays. There is the need for smarter, efficient and effective visualization for small screen platforms and a potential review and observation on existing visualizations for mobile platforms. Smart watches have shaped the way data is represented in small screen displays, allowing new interaction and visualization metaphors for users [109]. Integrated sensors and vibration feedback in new smart devices are other examples that enable new interaction with visual-feature applications to simplify information representation [109]. Generalising data collected and presenting them at an abstract level also enables better processing and rendering performance in mobile platforms [110]. In addition, user-interactive features such as one-hand thumb-based inputs have rapidly improved over the past years for mobile devices. The ability to zoom, drag, hold and filter [111] visual information on a mobile screen shapes the way for a better information processing experience for users [112]. Multi-touch [113] interactions, spatial input techniques and mobile collaborative features have yet added more user interaction capabilities in mobile platforms. While these features are not often directly related to how security visualization appears in mobile platforms, the ability to leverage on these interactive features enables users to comfortably interact and further assess what appears on their mobile screens. An example of this is a parental control application that tracks and monitors a child's location and their safety, utilises visual reporting capabilities is shown in Figure 2.12. It uses visual mapping techniques to plot and update Global Positioning System (GPS) coordinates on a near real-

time basis of people's movement. As seen here, the ability to leverage on existing mobile services and provide simple mobile visualizations for specific group of audiences is important. Despite going into details, such mobile parental mobile application uses mobile data stored in the cloud as part of the GPS service and transforms it into visualizations. Therefore, with both cloud and mobile technologies, using GPS signals and tracking features from the mobile being monitored enables a parental control application to locate the person of interest by utilising visual location markers.

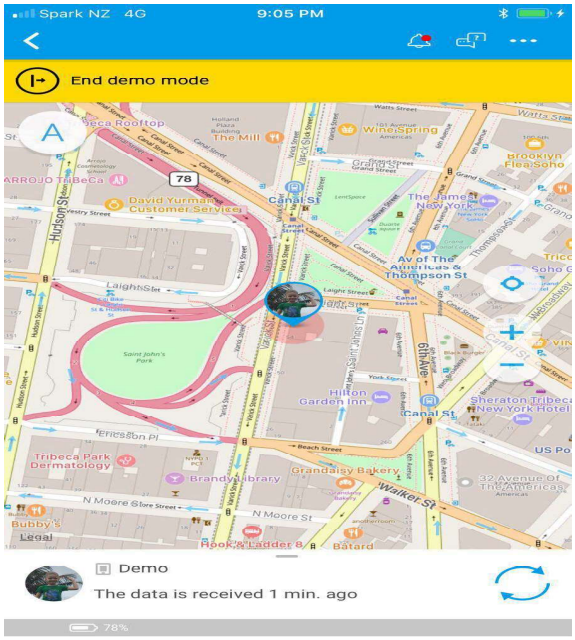


Figure 2.12: Parental Control Mobile Application with Security Tracking Features

Other popular visual analytic platforms like Tableau [59] and Microsoft Power BI [114] have addressed the need for data analytics, business intelligence and overall exploring the data at hand. Akamai's [115] internet monitoring mobile application provides a simulation of captured attacks around the world, as seen in Figure 2.13. Such analytics and exploring techniques have identified three distinctive activities to the entire analytic and iterative visualization construction process: (1) data attribute selection, (2) visual presentation template selection, and (3) visual mapping/representation specifications [116]. In security visualization for mobile platforms, the entire analytic and iterative visualization construction process adds a fourth (4) activity, i.e., visual performance and rendering process.

Figure 2.13: Akamai's Internet Monitoring Mobile Visualization Dashboard

2.5.4 Multivariate (n-dimension) Visualization

Assessing multiple sets of data within limited visual space during a security event scenario with visualization empower users to compare and critique why outputs, patterns and behaviours appear the way they are. This provides a user with learning moments to establish his/her ability to analyse what is presented. As a result, visual feedback in this case generates attention and interest, therefore motivating users to further explore a given visualization. Figure 2.14 illustrates a 'multivariate' visualization with n-dimensions showing multiple attack information represented within a small visual space.

The concept of multivariate representation and presentation techniques provides data visualization and visual analytics (DVVA) [117] techniques to utilising parallel coordinates (with n-dimensional visual view) visualization to extract crucial forensic science insights (information and knowledge) from large data sets. The time-tunnel (PCTT) [118] visualization tool is another multidimensional data visualization with parallel coordinates functionalities. It is used to represent IP packet data in intrusion detection due to the number of attributes present. The ability to view multiple security attributes (malicious payload details) in a multivariate visualization provides a way to control and enhance a user's visual view with minimal clicks and traversal ability around the display space, but still acquiring relevant information. Parallel coordinates enhanced in visualizations for pattern detection of changes enable observation of trends and correlations in datasets [119]. The use of existing parallel coordinate metrics such as entropy, edge crossing and class ordering are used

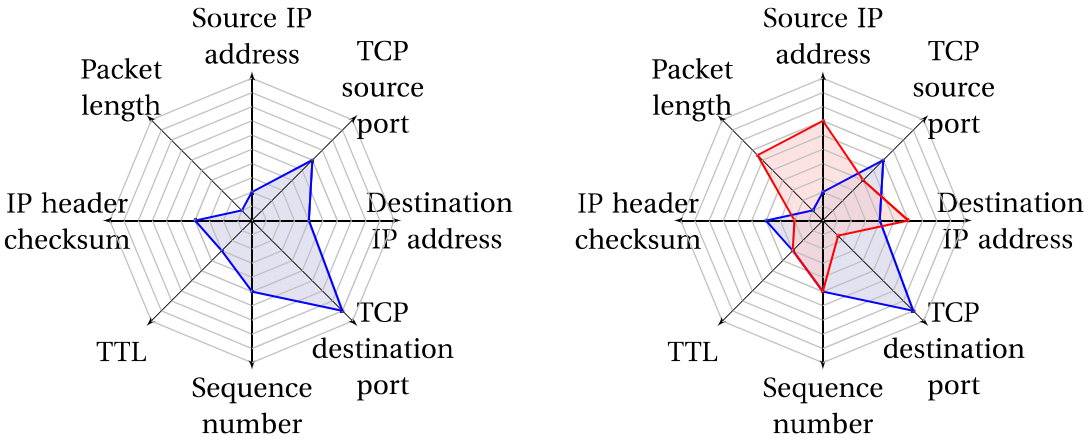


Figure 2.14: Multivariate visualization representation option

to facilitate data inspections. These techniques reduce the entire data analysis process for large datasets which is time-consuming when executed manually.

2.6 Security Visualization for Law Enforcement

Apart from VizSec, law enforcement organisations are another prominent security group worth addressing and observing for this research thesis. Observing security visualization usage by the law enforcement organisations has direct relationship with security and is a potential use-case for this research. The amount of data collected (e.g., notices and nominal, forensic data (fingerprints, DNA profiles, face recognition data, malware datasets and blockchain data)) from investigations on a daily basis for the purpose of law enforcement has far exceeded processing capabilities [120], [121], [122]. Therefore, the need to obtain security tools to aid investigations has been a demanding priority for law enforcement organisations.

2.6.1 Data Storage, Protection and Preservation

Apart from exploring and reporting on insights found during investigations, *information sharing*, *data protection*, and *data preservation* are critical to law enforcement organisations.

Law enforcement databases facilitate secure storage for data collected. The data stored are accessed with proper secure protocols under authorised security platforms. However, the nature of both cybercrime and physical crimes have become global and connected across zones, countries and jurisdictions [123], [124]. Thus, communication is the key to solving

the causes of these crimes. Secure, effective information sharing techniques are required for communication between countries and jurisdictions while preserving the underlying raw data i.e., preserving the underlying data ensures the protection against tampering with data.

2.6.2 Information Sharing and Attribution Process

However, law enforcement investigations often face a challenge of analysing data collected on a day-to-day basis. The complexity contributes to slowing down the entire investigation process. Investigators and cyber security specialists are continually exploring the best methods and tools [121], [125] to help them analyse and understand the nature of the data with the intention of obtaining useful insights [126], [127], [128], [121], [122]. However, international law enforcement organisations frequently have the challenges of: (1) attributing back to the source of the attack, and (2) ensuring and protecting data sensitivity and privacy issues, especially for the case of transnational cyber-attacks [129]. The trust complications for sharing and exchanging information with other countries becomes a challenge, which results in slowing down investigation processes.

There needs to be a method of sharing the data/information comfortably with other countries involved while maintaining the data integrity and authenticity, and without revealing the underlying raw data. Security visualization has proven to be a critical solution to this challenge in helping crime analysis [128].

2.7 Datasets and Information Security

The rise of cyber-attacks reported around the world requires better applications and tools to scan, detect and eventually mitigate them. This entire process is made possible with the use of datasets [130] at hand. The dataset gathering/collection process can be a difficult task for security research. However, in scientific research where a hypothesis is stated, testing with the use of reliable datasets is important [130].

Over the years, security visualization has proven to be a reliable technique and method for identifying attacks. Researchers have come to understand and extract useful information from visualizations provided where insights lead into investigations, and data analysis is executed for the purpose of identifying possible malware behaviours [101], [51], [102].

However, in order for investigations and reporting by the forensic researchers to be accurate or almost 100% accurate, researchers have explored new visualization techniques, such as the use of multidimensional visualization, to identify and expose threat landscapes and threat behaviours [131]. This makes reporting easier for security analysts and users of the visualization.

2.8 User Cognition and Perception Measurement Techniques

A user study on visualization effectiveness and cognitive load, (Anderson E.W. et (2011)) [64] evaluates visualization techniques by measuring brain activities, with a passive recording using electroencephalography (EEG). Addressing visualization techniques in relation to user (viewer) visual context (processes) requires a comparison between these techniques against an assessment of the user's (viewer) cognitive resources. This process brings about the concept of visual burden/workload. The outcome of this assessment is a result of statistical visual interpretation, cognitive load measurement indicators and user working memory estimates, while undergoing various user tasks.

Cognitive load plays a vital role to how successful visual information is processed through a user's visual cortex. It is made up of three parts: (1) intrinsic load, (2) extraneous (ineffective) load and (3) germane (effective) load [4]. Thus, in a security visualization observation environment, an ideal concept of cognitive load would be to ensure that the sum of intrinsic, extraneous and germane load should stay within the user's (viewer's) working memory limits. These would provide effective results with potential security visualization insights.

The psychological contributions to effectiveness measurement in visualization are identified and measured from assessing human cognition, perception, attention span and working memory load. The rational between a user's cognitive capability and working memory load can be understood and measured using their mental effort rating, as shown in Figure 2.15. For example, an ideal rating would fall in 'Region A' (Figure 2.15) [4], where a performance reading is high while mental effort is low. This also means the user's working memory load is high as well. User studies have provided means for cognitive load measurement [132], particularly mental effort and performance assessment techniques, which address visualization efficiency.

Insight-based evaluation [133], [134], [135] by InfoVis has elevated the use of insights as an evaluation measure for technologies. 'Insight' [133] is defined as gaining accurate and deep understanding of something, i.e., a unit of discovery. It is often not achieved by predefined tasks or procedures but there is a higher probability it is a by-product of exploring without an initial goal or destination. Moreover, 'sensemaking' [136] plays a major role in gaining insights. Although the model (Information -> Scheme -> Insight -> Product) of sensemaking includes insight as a component, the model enhances the entire experience of gaining and understanding insight.

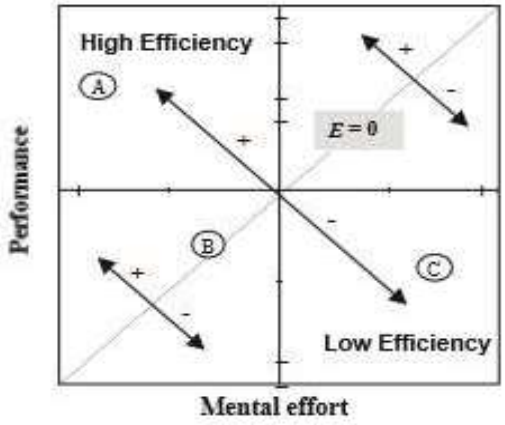


Figure 2.15: Mental Effort Efficiency Reading [4]

2.9 A Security Visualization Classification Review

Based on the literature review and assessment results, we have identified and classified existing visualization research areas into the following visualization purposes:

1. User-centred visualizations [81]
2. Monitoring visualization mechanisms [137]
3. Defensive visualization mechanisms [138]
4. Mobile device traffic visualizations
5. Cloud security visualizations
6. Visualization for ehaviour analysis [102], [131]

The review showed that there were less than 10% of the total ViZsec research were addressing user-centred security visualizations and security monitoring-based visualizations. However, the prominent security visualization research predicted for the next five to ten years is in areas of defensive (monitoring) mechanisms, mobile device traffic and cloud security. Security visualization for behavioural analytic purposes with both machine learning and deep learning methodologies is proving to dominate the data analytic research environment.

User-centred security visualization research is currently debatable as mentioned in the previous survey papers, due to the range of the user's requirements and preferences. However, referring back to our check-lists, we define 'user-centred' as 'end-users' of the products;

i.e., mobile device and cloud services customers. Usually these users are less IT-literate or no experiences at all with mobile devices, yet they are vulnerable to security attacks such as malware attacks.

Mobile device traffic and cloud security visualizations are the potential trend for security visualizations because most users of the 'Internet of Things (IoT)' which includes mobile devices, are becoming todays target of cyber-security threats. This is due to attacks on 'sensitive data.' In addition, the number of users using mobile devices compared to desktops or other hardware is far greater.

Even though "Defensive security visualization mechanisms" can be regarded as part of monitoring, surveying the past decade of VizSec research has indicated that there was only one paper published on defensive mechanisms [26]. Future defensive security visualization mechanisms can provide insights, prevent threat attacks and also tailor the users of such mechanisms to be cautious when browsing the net or even analysing real-time data logs/log files.

2.9.1 Existing Visualization Challenges

We have previously discussed the data types needed for visualizations and it is important the right types of data are processed and used. This, however, brings the challenge of having to visualize large volumes of data. Large data volume poses a challenge of preprocessed data being lost while processing, visualized and retaining all information within the visualization. Therefore, data lost effectiveness in security visualization.

2.10 Summary

In summary, this chapter provides a thorough literature review on existing visualization within most relevant research domains. We highlighted the general use and aspects of: why use visualization, what are the purposes and types of visualization, what is security visualization, and what are visualization applications? We also provide visualization limitations and challenges. In addition, we reviewed how visualization affects users (viewers). This is done by identifying existing key aspects where users and visualization merge to provide insights from respective datasets. We provided a prominent direction with regards to the VizSec research. This includes providing a 'checklist' to act as the visualization requirements. With the use of the checklist, we have assessed the past 166 VizSec papers and identified interesting findings that helped us identify research gaps. The first involves the 'dependencies' that contribute to providing better security visualizations. We have identified the trend areas in which the security visualization research is exploring and the possible future

direction of security visualization research. Our findings show that one of the core security visualization research interests for the VizSec community is Exploring/Analysing/Discovering/Reporting (EADR). Throughout the past decade, EADR has been consistent, while security visualization for 'Information and Situational Awareness' has been decreasing over the past years. Finally, with the current technological trend where most users are using mobile devices and are reluctant to use the cloud services, this indicates the rise in mobile-device threat attacks.

Finally, we assessed existing effectiveness measurement techniques and frameworks in visualization and provided research gaps within our research domain. The identified research gaps are as follows:

2.10.1 **Research Gap 1-** ***Effectiveness Measurement in Security Visualization***

While there are existing mobile device visualization applications, areas of approach are specifically targeting either certain groups as their audience or certain application purposes. For a user to make better decisions and select a type of visualization right from the begin the moment he/she confronts a visualization, there needs to be an effective approach which could aid the user's decision making. Effectiveness in visualization has been addressed; however, in the security domain and specifically for mobile platforms, this hypothesis is still in a grey area. Therefore, a full security visualization effectiveness measurement approach is the primary research gap which is addressed in this thesis.

2.10.2 **Research Gap 2 -** ***Solution Addressing Security Visualization Complexity***

Our review of this research area has indicated that security visualization complexities in mobile platforms are due to several causes: (1) hardware limitations, (2) processing power limitation, and (3) performance limitations. Although parts of this identified research gap are currently addressed, there is room for better techniques, technologies and hybrid solutions worth addressing to minimise this research gap. It involves finding new methodology around data management, performance management and optimisation techniques using machine learning and deep learning algorithms.

2.10.3 **Research Gap 3 -** ***Lack of Intermediary Mechanisms between Core Entities***

Current solutions around information and data visualization for users have grey areas that needed to be addressed. Based on our review, a derived conclusion indicated that there is lack of intermediary links, techniques and methodologies between the visualization presented, the visualization tool itself and the users (viewers). Therefore, our third research gap states the need for further efficient methods and ways to harmonise solutions between entities such as the links between users and the visualization, i.e., what can be done to improve user experiences? In addition, there are needs to address and harmonise research approaches from computing science, cyber security and the psychology fields. Theoretically, this would enable security researchers to understand users, user interactions and security visualization tools in a more effective way.

2.10.4 **Research Gap 4-** ***Lack of Security Visualization Standard and Guidelines***

Our fourth research gap identified is the requirement and need for a security visualization standard that could help facilitate information knowledge by establishing useful pieces of knowledge well in advance before confronting a visualization, to interact with it and gain insights.

Finally, with all key research areas mentioned and discussed in this chapter, we now have a clear understanding on existing work around effective measurement methodologies in visualization. However, our framework focuses specifically on security visualization with the aim of introducing effectiveness measurement in security visualization for mobile platforms with regards to the urgency of information presented. This will be discussed in the remaining chapters of this thesis.

Chapter 3

Constructing and Understanding the Required Datasets

Datasets provide security researchers and data scientists with the ability to process, analyse, test and evaluate scientific hypothesis. In security visualization, datasets [139] are the source and bloodline for providing meaningful insights to viewers. Datasets allow security experts to analyse and transform predefined data nodes into visualization whereby visual information is processed faster by humans compared to an attempt from reading logs.

We designed several data collection methodologies and techniques leveraging on both system and network logging tools. Multiple sources and security landscapes are used to simulate near-real world cyber-attacks. A primary thought and consideration when collecting logs is the intention of giving both web and mobile users a variety of possible security visualization. However, in order to collect datasets, ethics relating to data collection and participants are critically important. This includes getting consent from participants.

3.1 Importance of Datasets

The importance of using real-world data in cyber security is to enable and validate cyber security research. Unfortunately, obtaining real-world security data for research purposes is rare. Datasets or institution data are regarded as private or confidential and under most privacy laws, private data are not shared publicly unless authorised. This renders the entire real-world data collection process a difficult task. Therefore, as academic researchers we designed our own data collection framework whereby simulated security events (malicious attacks, SQL injection, etc.) are implemented, logged and collected to satisfy our research scope and requirements.

The absence of data creates a difficult task for cyber security researchers whereby affecting them in providing precise responses to cyber-attacks. Deep within the data collected are patterns and behaviours which when identified helps researchers to further understand

cyber-attacks. Cyber-attack patterns and behaviours are hidden insights shared when predefined data nodes are transformed into visual outputs. For example, provenance [17], a 'derivative history and series of chronicles of meta-data' derived from datasets have the ability to show the history and state of data at a required query time. Attribution is another example in security where identifying an attack and tracing it back to the source of attack can be illustrated through the use of security visualization. Both provenance and attribution [140], [141] are visible through the process of collecting/logging data in monitored networks. Therefore, from a research, threat intelligence and cyber response scenario, datasets are collected for the following security reasons:

- malicious attack (payload) identification.
- understanding the cyber-attack landscape.
- attribution and provenance knowledge or insights.
- daily intelligence (tracking and monitoring) on systems and networks.

However, privacy laws and regulations involving data and user 'personal data' require that proper ethics are applied, and approval is sought before collecting or gaining possession of data. The required dataset collection ethics for this research are discussed in Section. 3.2.

3.2 Ethics Around Datasets

Security datasets collected on a day-to-day basis are critical and confidential. This elevates the need to acquire proper ethical approval from the appropriate authorities involved. This means that outlining the required data collection scenarios and methodologies is important.

Our dataset collection requirements involve simulating threats and cyber-attacks which are prominent and associated with visualization on mobile platforms. This includes: (1) threats and cyber-attacks on both web and mobile platforms, (2) security events that can be viewed on mobile platforms and (3) applications and services associated with cloud technologies. When mobile platforms are the primary devices used daily, any data collected from web and mobile platforms associated with users require ethical approval. In this data collection process and research, the users involved are in these three categories: (1) desktop and mobile users at the University of Waikato, (2) law enforcement (digital crime officers) users, (3) industry experts and (4) end-users.

From the law enforcement data collection approach, an international law enforcement agency has been the point of collaboration. All ethics around dataset collected and used for

law enforcement security visualization purposes have been acquired through proper and approved processes. However, the ethics and data collected are kept confidential due to law enforcement policies and regulations. Datasets collected at the University of Waikato have gone through the computing science ethics approval processes and before carrying out data collection. Our approved ethics are for the following data collection items:

1. New Zealand Cyber Security Challenge 2015 (NZCSC2015) Dataset collection - See Appendix.B. 1 for a copy of the ethics approval letter.
2. New Zealand Cyber Security Challenge 2016 (NZCSC2016) Dataset collection - See Appendix.B. 2 for a copy of the ethics approval letter.
3. New Zealand Cyber Security Challenge 2017 (NZCSC2017) Dataset collection - See Appendix.B. 3 for a copy of the ethics approval letter.

The ethics cover data collection involving users, web application logs, kernel and system call logs, network (.pcap) logs, top logs and video logs capturing user inputs. The video logs contribute as our '*ground truth*' verification files against all other various logs collected during the New Zealand Cyber Security Challenges. Additionally, participants are briefed about the data collection process and are asked for their consent.

3.3 Dataset Collection Requirements

The prerequisite of obtaining data in any means comes down to expected quality and characteristics of data collected. These prerequisites are as follows:

1. Data collected is related to any security event (e.g., malicious attack).
2. Network, system, application and kernel logs.
3. Attribution and provenance related.

With these prerequisites, specific Data Collection Requirements (DC-R) are designed and scoped as stated below:

1. **DC-R1 - Security Event**: These events consist of various different security related scenarios. For our thesis, the events cover malicious activities recorded during the New Zealand Cyber Security Challenge (NZCSC) events. This includes the '*red vs blue team*' challenges. Such malicious events include URL-manipulation, Remote-code execution, SQL injection, and more. Other security events include normal data tracking and monitoring of systems and networks of interest.

2. **DC-R2 - Security Entities and Attributes**: Security Entities refers to all nodes affected during a malicious attack. This could be an 'IP address' or an 'apache web server.' Attributes refer to features and components of these entities. For example, an attribute belonging to an entity IP address is whether it is a private IP address or a global IP address.

3. **DC-R3 - Entity Relationships**: This refers to links between entities. Links connect entities, allowing them to either communicate with or be classified into a same group. These relationships also provide an attack landscape to an identified security event.

Once these requirements are identified, we set up a data generating mechanism that will allow data collection. These mechanisms are discussed in Section 3.4 and Section 3.5.

3.4 New Zealand Cyber Security Challenge (NZCSC) Datasets

The New Zealand Cyber Security Challenge (NSCS)[1] [142] event is an annual event established in 2014 by the Cyber Security Researchers of Waikato (CROW), the University of Waikato and its industry partners. The core purpose of establishing such a security event is primarily for several reasons. Firstly, it is an ethical hacking training ground where academics, industry security professionals and college students learn security concepts with the preventative approach of securing networks, systems and also cloud platforms. Secondly, it provides security awareness for all various audiences and, finally, the NZCSC event creates a safe and controlled cyber challenge environment for security research with data collection opportunities.

Data collection is a challenging task due to legislation around information sharing, privacy and related computer crimes. Unless there is a Memorandum of Understanding (MOU) between security searchers and firms, organisation (company) regulations and policies also restrict data collection and information sharing among security researchers. Therefore, creating cyber security challenges enable researchers to test security algorithms, applications and collect data.

In this thesis, we established the data collection requirement and methodologies to fit the NZCSC purpose. As part of a team effort to implementing the NZCSC and collect various datasets to meet multiple research needs, our thesis focuses on provenance and attribution related datasets. We acknowledged the design and developer of NZCSC-2015 Round-2 attack network [5] the contribution and ability to leverage and utilise the dataset collected. As part

[1]The New Zealand "National Cyber Security Challenge (NZCSC)" (https://cybersecuritychallenge.org.nz/) competition was established in 2014 by the University of Waikato (UoW) with all collaborating partners

of the data collection team, our focus is on extracting several Round-1 and Round-2 data log types which met our thesis requirement. These data log types and requirements are discussed in the remaining of this chapter.

3.4.1 NZCSC-2015 dataset

Due to difficulties in obtaining security datasets publicly, our NZCSC-2015 dataset was designed to capture all logs ranging from within the kernel, system and network layer. We have utilised Sysdig[2] [143], [144] to capture all kernel and system-level logs, and with Tcpdump[3] [145], [146] we collected network (.pcap) logs. In addition, video logs were captured of user-inputs in selected challenges to act as our verification 'ground-truth' dataset.

Figure 3.1 shows the NSCSC-2015 data collection infrastructure overview outlining the specific areas of the network that were used to collect our various datasets. As shown in Figure 3.1 data were collected from both Round-1(CTF challenge) and Round-2 (Red-Blue Team challenge) of the NZCSC-2015 challenge. However, designing a data collection approach for security visualization has challenges such as identifying realistic DC-R1 (security event) scenarios to visualize. Therefore, with all data collection requirements (DC-R), Round-1 has three challenges that were used to collect data, which are challenge-6, challenge-7 and challenge-10. In Round-2, all red and blue team machines have logging configurations for data collection. File, web, mail and tech machines in the blue team network environment are logged.

Figure 3.2 provides a full detailed NZCSC-2015 Round-2 data collection schematic [5] with the components required. The NZCSC-2015 Round-2 attack network involves four red teams (red1.csc2, red2.csc2, red3.csc2, baden.csc2 (Redundant backdoor machine)) machines and five blue team environments (file, web, mail and tech machines). A simulated public network and external network configuration to facilitate security, internet connection and separation of the networks are provided with various connection links (as indicated with coloured arrows) showing how all networks interact with each other.

In summary, various log collection types were implemented and managed by the NZCSC-2015 dataset team, which the author is a member of. These include:

- Audit access (access.log)logs.
- Audit error (error.log)logs.

[2]Sysdig is a universal system-level exploration and troubleshooting tool for Linux with native support for containers. (Link: https://sysdig.com/)

[3]Tcpdump is a free UNIX packet sniffing software used to gather network data (network packets).(Link: https://www.tcpdump.org/tcpdump$_man.html$)

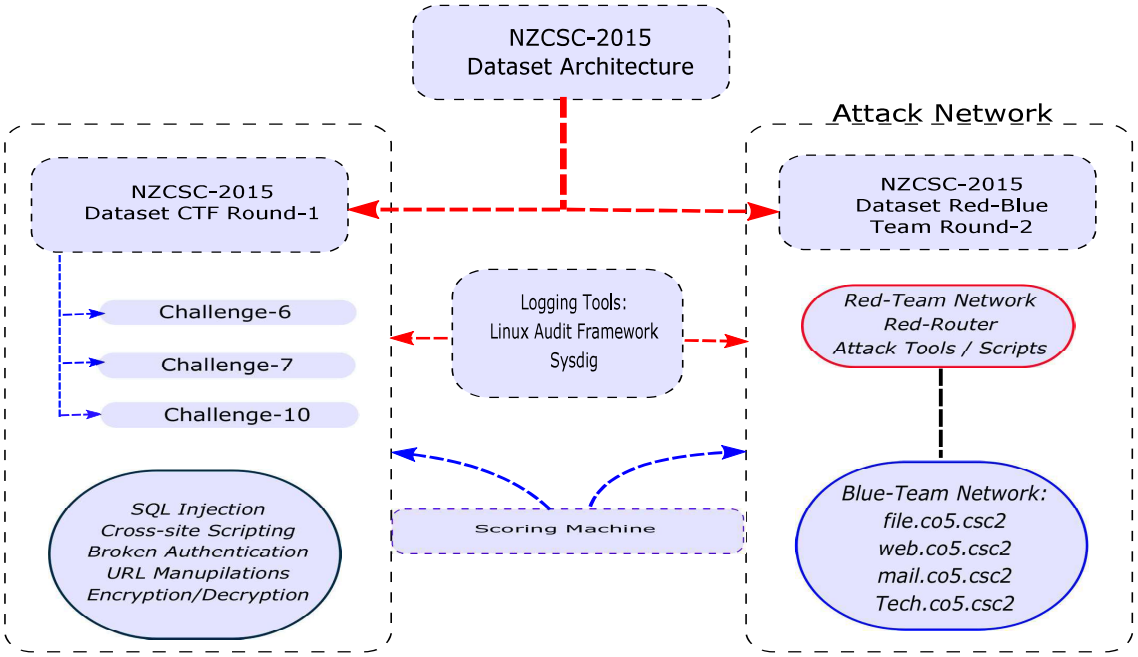


Figure 3.1: NZCSC-2015 Data Collection Architecture Overview

- Top info (top.log) logs.
- Mail server logs.
- Daemon logger (network traffic) logs.
- Ground-truth video (vlc) logs.

As part of the data collection process, all data sets are standardised and anonymised according to our security visualization needs. This step also complies with the standard ethics requirement around the usage of data and user privacy. The standardisation process is discussed in Subsection. 3.6.1.

3.4.2 NZCSC-2016 dataset

The NZCSC-2016 data collection infrastructure practically covers similar architecture to our NZCSC-2015 dataset collection environment; however, certain changes are applied to focus on specific security attack scenarios. This includes building NZCSC-2016 Round-1 challenges to a more realistic real-life attack landscape which makes the challenges more interesting for participants. Therefore, the dataset collection infrastructure was setup to collect logs in all Round-1 challenges and the Round-2 (red-blue team) challenge. Below, Table 3.1 outlines all different Round-1 security event (challenge) scenarios.

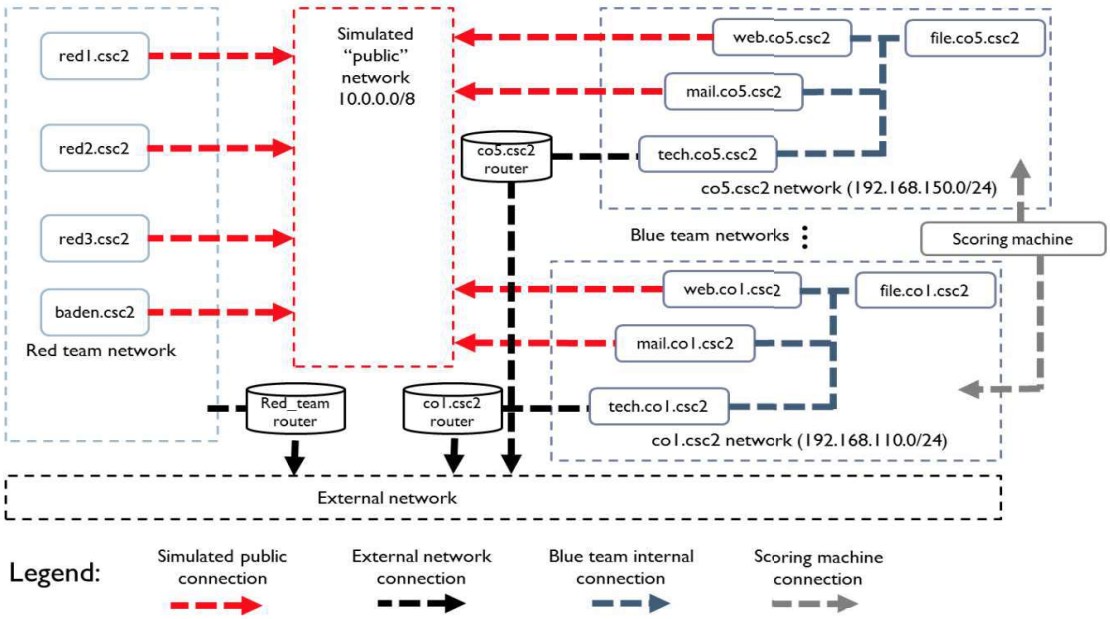


Figure 3.2: NZCSC-2015 Round-2 Data Collection Schematics [5]

Table 3.1: NZCSC-2016 Round-1 Security Challenge Types.

Round-1 Challenge	Security Challenge Type
Challenge-1	SQL Injection (easy) scenario.
Challenge-2	Phone encoding scenario.
Challenge-3	Social engineering Alice scenario.
Challenge-4	SQL Injection (hard) scenario.
Challenge-5	GPS tracking scenario.
Challenge-6	Cross site Scripting (XSS) scenario.
Challenge-7	Remote Bash Code scenario.
Challenge-8	Encryption/encoding scenario.
Challenge-9	Buffer overflow Information leakage scenario.
Challenge-10	Default login/PHP injection scenario.
Challenge-11	Steganography scenario.
Challenge-12	Forensics scenario.

Due to the nature of Capture-The-Flags (CTF), the NZCSC-2016 Round-1 logging approach was configured to capture audit access and error logs using Linux Audit Framework (LAF) for all volunteered participant machines. Sessions, user login details, request paths and server responses data logs are collected. While these logs can not directly show attribution of attacks, analysing the logs for traffic patterns and troubleshooting issues is the primary se-

curity visualization goal for the Round-1 challenge. Figure 3.3 shows the NZCSC-2016 data collection architecture overview outlining the different attack landscape, logging types and network infrastructure environment.

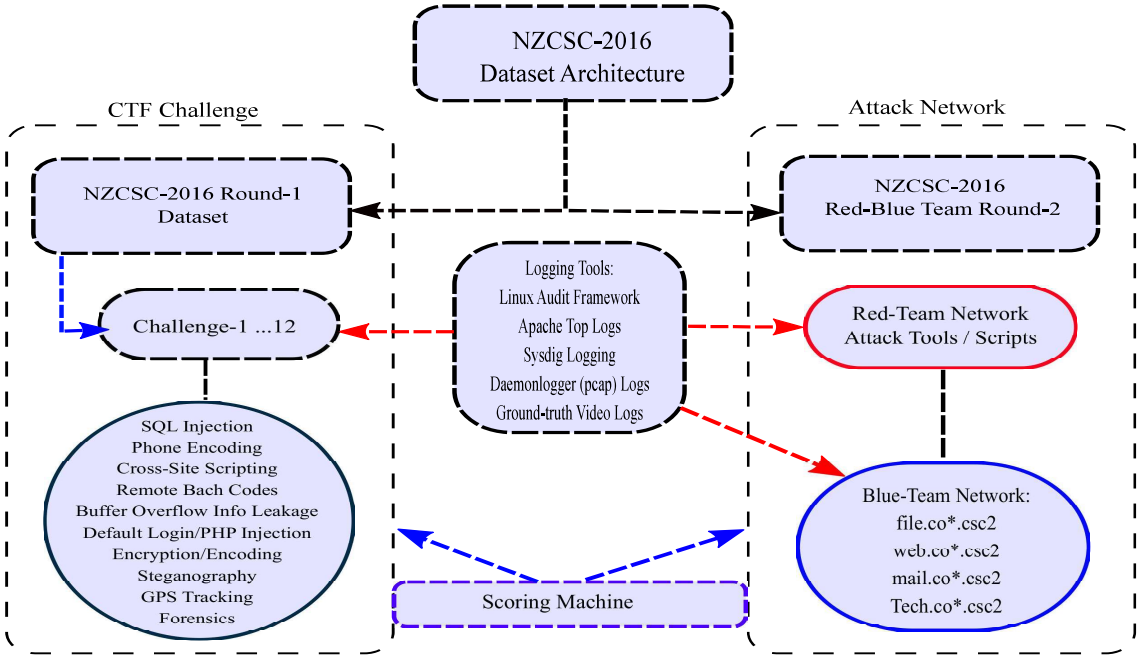


Figure 3.3: NZCSC-2016 Data Collection Architecture Overview

Our Round-2 data collection architecture and design is the important contribution component to this research. Inheriting the NZCSC-2015 Round-2 platform (a virtual machine environment: red-blue teams' networks and public network), we changed the data collection approach towards specific attribution, provenance and user-awareness purposes. Round-2 data collection schematics are shown in Figure 3.4 outlining all logging tools and setup details. Linux Audit Framework (LAF), Sysdig, Daemonlogger, Wireshark, Apache top logging, and VLC ground-truth video logging are configured in all red and blue team machines. All data logged are stored in an external data storage location during the collection process.

In security visualization [142] data logging and collection are important for monitoring systems and networks. It allows network and security experts to observe and maintain systems in a most known secure environment. This helps ensure regular implementation of security tools, protocols, rules and policies based on identified cyber-attacks and threats. A deeper understanding of cyber-attacks heavily relies on collected datasets from the captured attacks. Therefore, the selected NZCSC-2016 logging mechanisms aim to monitor and log all attack actions executed by the participating teams from all levels starting from network

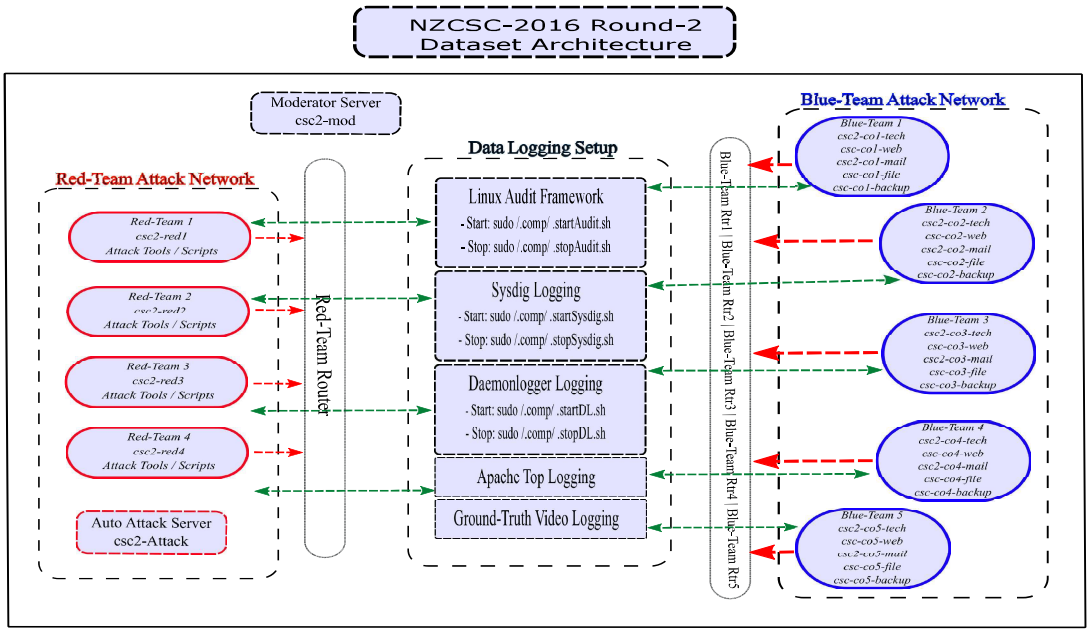


Figure 3.4: NZCSC-2016 Round-2 Data Collection Schematics

traffic, kernel level actions, user logs, system level actions, application access and error logs and user inputs. These logging mechanisms are configured for selected teams in all red-blue team challenge (Round-2) teams.

Raw security data consists of a large amount of information which requires cyber security experts to apply filtering techniques when processing them. In security visualization, the need to construct a desired filtering technique to collect security data with user-centric attributes that could empower users to see and understand security visualization is a challenging task. This is due to the wide range of targeted audience, often with different preferences. Thus, a NZCSC-2016 data collection filtering requirement approach for our security data is shown in Figure 3.5. Basic user-centred security attributes are at most familiar to views that contribute highly to setting up the '*data attribute checklist.*' This data attribute checklist consists of *suspected or malicious IP address (s), attack payloads, malicious data (events, inputs), host details, user-login details, attack landscapes* and more. Our checklist outlines the generic standard and common security attributes. It is flexible and dependable on the type and nature of attack logged and identified for security visualization. As seen in Figure 3.5, the *raw data* collection required the filtering technique, while *analytical data* relies on both the filtering technique, and standardisation process. Finally, Figure 3.6 provides a preview of the analytical data records stored in the database. These analytical data can also be regarded as raw data for other security needs.

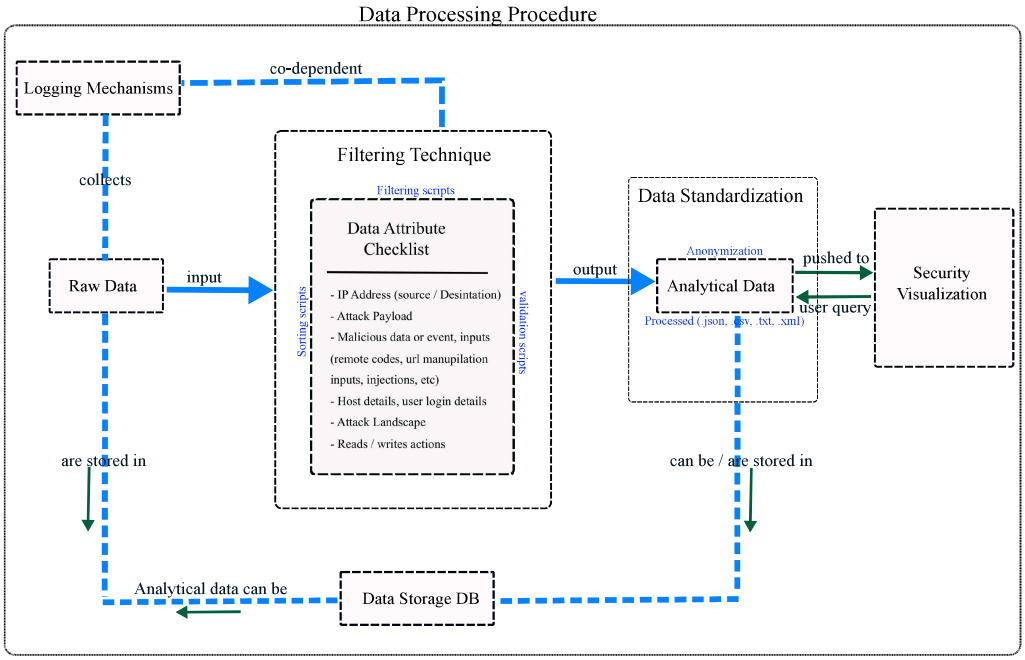


Figure 3.5: NZCSC-2016 Data Filtering Process Approach

ID	Time	Source	Destination	Protocol	Command	Attack Type
26	18:29:28	10.0.53.4	10.42.122.123	TCP	nmap 10.42.122.0/24	Reconnaissance
35	18:29:57	10.0.53.2	10.42.122.200	TCP	/usr/bin/python /usr/bin/sqlmap -u http://10.42.12...	SQL Injection
36	18:30:24	10.0.53.3	10.42.122.200	HTTP	GET /adminlogin action?username=&password=...	URL Manipulation
37	18:31:18	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=%27&password=...	URL Manipulation
38	18:31:29	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=Admin&password=...	URL Manipulation
39	18:31:59	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=Admin&password=...	URL Manipulation
40	18:32:49	10.0.53.2	10.42.122.200	TCP	/usr/bin/python /usr/bin/sqlmap -u http://10.42.12...	SQL Injection

Figure 3.6: A Data Record Snippet After Analytics

With ideal goals to visually show all attack information such as payloads, sources, destinations, and attack types, attribution and provenance related specifications were taken into consideration. All connections established between red and blue team machines such as source and destination IP-addresses, ports association and commands associated with these connections provide additional information that is vital for security visualization. For example, remote executing programs, file transfers, secure and transparent tunnel usage and managing public keys are logged and collected. As shown in Figure 3.7, interested data (secure shell (SSH) details, IP-address (s), read/writes, user-login details, uniform resource locator (URLs)) records were collected when a participant utilised the SSH library (libssh.org) during the security challenge.

In summary, our NZCSC-2016 log collection design involves the following logging tools:

- Audit (LAF) access (access.log) logs.

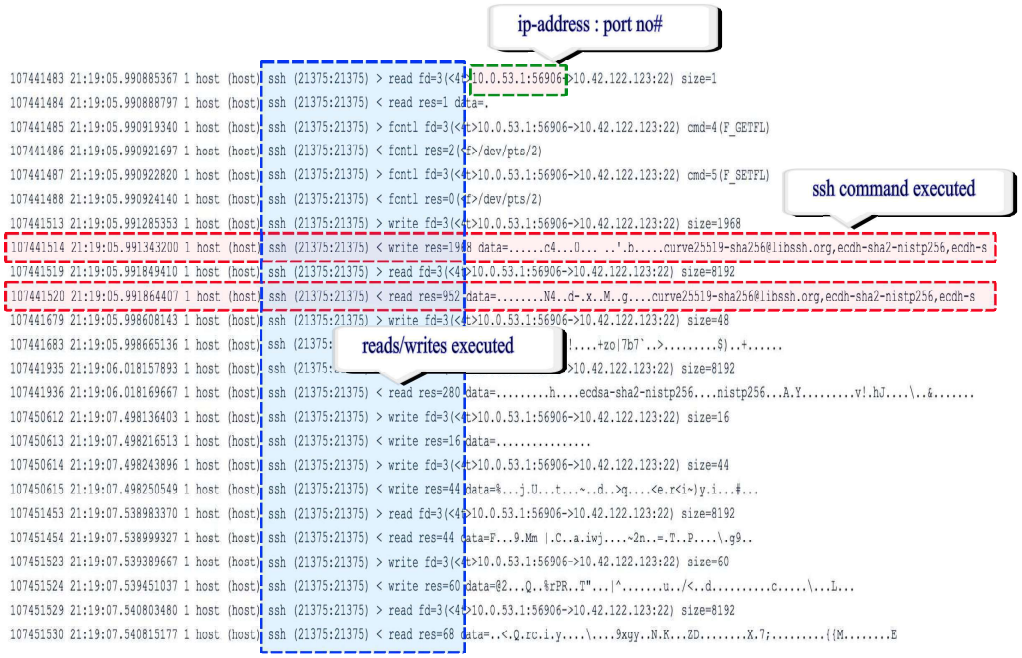


Figure 3.7: Data Collection Sample of Security Related Attributes

- Audit (LAF) error (error.log) logs.
- Apache top info (top.log) logs.
- Sysdig logs.
- Deamonlogger and Wireshark (.pcap) logs.
- Ground-truth video (vlc) logs.

The choice of using these opensource logging tools provides an affordable research budget and most importantly provided the security datasets which are needed for this research.

3.4.3 NZCSC-2017 dataset

As part of our main contribution to this research thesis, and leading the NZCSC-2017 data collection team, our the NZCSC-2017 Dataset collection infrastructure provides a twist to previous years of data collection during the New Zealand National Cyber Security Challenge (NZCSC). This is due to the demand of providing a user-centric security visualization. The need for user-centric security visualization enabled a real-time data collection and processing architecture design approach. Therefore, the overall goal for our NZCSC-2017 dataset

collection was to identify security events during the NZCSC-2017 Round-3 (red - blue (attack/defend)) team challenge. While data was collected during the NZCSC-2017 Round-2 (red - blue (attack/defend)) team challenge, a real-time security visualization providing red team attacks on the blue team networks was visualized. Progger (Linux and Windows versions) and Sysdig had been installed on all red and blue team virtual machines. Figure 3.8 shows the entire NZCSC-2017 attack network with specific logging mechanisms in all virtual machines.

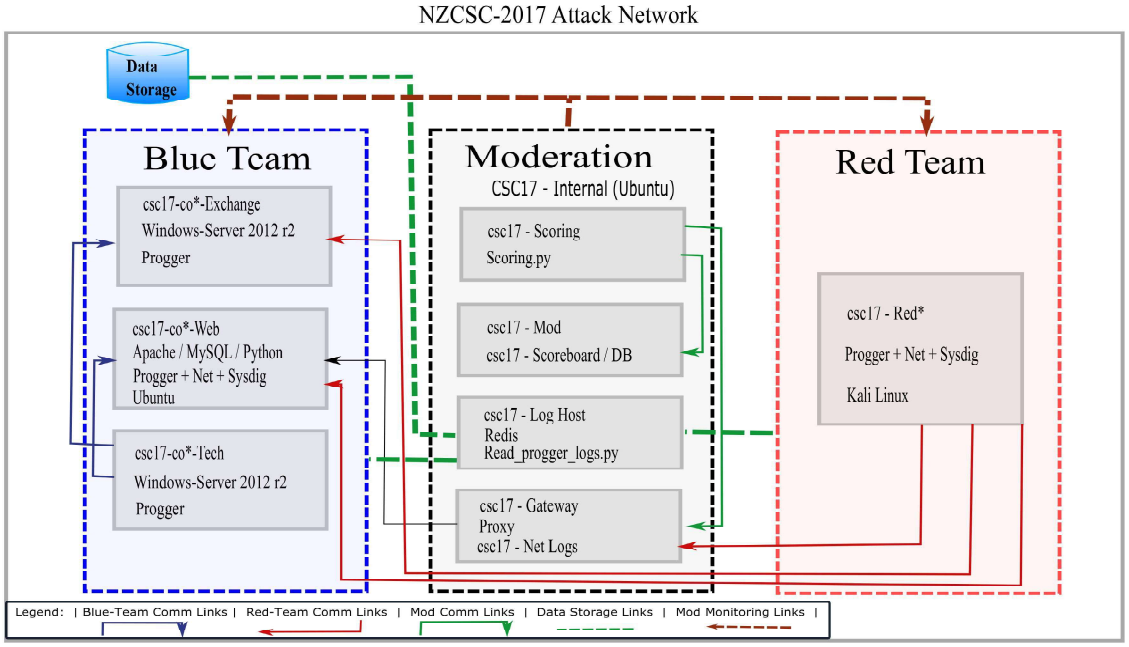


Figure 3.8: The NZCSC 2017 Attack Net Design

Deploying Progger and Sysdig into all red and blue team machines (network) with Redis as an intermediary between all red/blue team networks, log host and database storage, allowed the data collection process to be successful for a real-time processing scenario. Logs are collected and temporarily stored in Redis for real-time visualization use while the same copy gets written into specific database tables. Moreover, our NZCSC-2017 dataset is stored in MongoDB (a NoSQL database platform) and is queried as JavaScript Object Notation (.json) and comma-separated values (.csv) log entries for usage. Alternatively, these .json and/or .csv log entries were reused fo other security analytical purposes.

As seen in Figure 3.8, the log host is moderated by an internal moderation machine ensuring that the raw data collected from all red and blue team machines were used for the two following reasons: (1) processed by parser scripts and pushed to the security visualization frontend for visualization, or (2) data recorded were written into respected database tables

for storage and future use.

The real-time collection process and storage mechanism requires several main operations to ensure the overall livelihood of the backend security visualization infrastructure. A multi-threaded operation approach consists of the following:

- *Data collection collector (scripts)*: this operation primarily consists of the Progger and Sysdig logging scripts. The collector ensures all red and blue team machines are logged.
- *Listener process*: this operation listens for new information on the Redis server and extracts it for analysis as part of the real-time process.
- *Status process*: the status process checks and displays data updates of the collected data on the console at a regularly repetitious interval.
- *Database process*: the database process ensures that the processed data (analytical data) from raw data, are delivered and written into compatible MongoDB formats and stored.
- *Parser (scripts)*: the parser ensures that data is pushed from the back-end to the security visualization frontend for visualization.

While these operations are briefly described, a thorough description and analysis of these data collection operations are further discussed in Chapter 5 of this thesis.

3.5 Law Enforcement Datasets

Security visualization for mobile platforms should in reality be flexible and able to visualize what a mobile user prefers. This means data processed for visualization can be from multiple data sources for various security purposes. A core component of this thesis is providing security visualization for law enforcement. Some primary law enforcement demands for security visualization fall within these reasons: (1) intelligence, (2) tracking and monitoring, (3) reporting and finally (4) information sharing purposes. With these demands, multiple data sources ranging from both malware and bitcoin datasets are acquired through approved procedures, policies and regulations for our security visualization intelligence application.

3.5.1 The Bitcoin Dataset

This thesis leveraged on publicly available bitcoin datasets to show, track, monitor and report on bitcoin transactions that are of interest to law enforcement operations. However,

due to information sensitivity, privacy and law enforcement regulations, a generic bitcoin data collection schematic is discussed for this thesis. Figure 3.9 illustrates our entire bitcoin data collection process. Multiple bitcoin data sources were used for this data collection project. Further detailed explanation on bitcoin data is discussed in Chapter 5 of this thesis report.

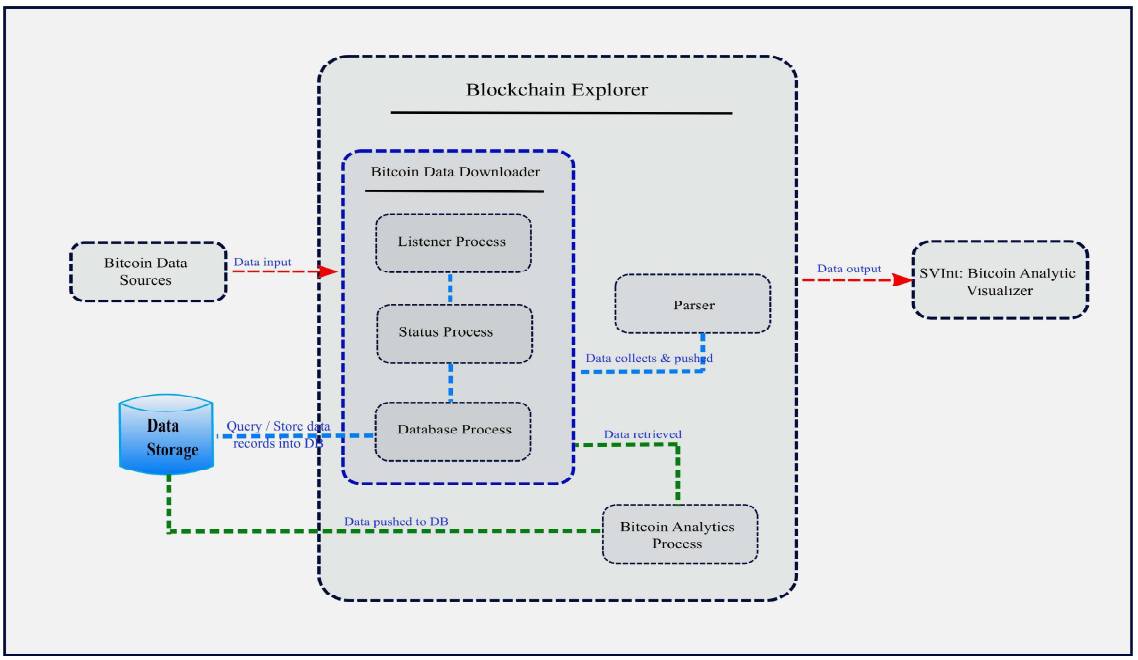


Figure 3.9: Bitcoin Data Collection Schematics

3.5.2 Mobile Malware Datasets

Apart from bitcoin datasets, malware datasets are required for intelligence, tracking and monitoring capabilities by law enforcement services. We provide a malware data collection schematic as a proof-of-concept for law enforcement security visualization. This data collection schematic includes the following components:

- Multiple data sources from trusted security firms.
- Database storage location.
- Data collector and parser.
- Update process.

Malware datasets are received on a weekly basis and are processed, stored and preserved in secure data storage locations. Data processed from these malware datasets are used for our law enforcement security visualization which is thoroughly covered in Chapter 5 of this thesis.

3.6 Dataset Requirements and Specifications

Security visualization tools and applications require datasets for analytic purposes. However, most tools and applications require standard data and not often raw datasets. In addition, sensitive datasets are not permitted for usage unless proper means of anonymisation and standardisation steps are executed to safeguard potential users. This is to preserve privacy and reduce the chances of attributing back to the dataset sources. Therefore, our thesis provides the anonymisation and standardisation process used.

3.6.1 Data Anonymisation and Standardisation Process

In order for such sensitive datasets to be used for cyber security research purposes, with the ultimate goal of publishing the available datasets publicly, 'Anonymising and Standardising' the dataset is crucial. *Why the need for a data anonymisation process?* Due to security, privacy and sensitivity reasons, this eliminates the chances of attributing back to distinctive network sources. The anonymisation method focuses on the following:

- Locate sensitive information (names, user-names, IP addresses, etc.) attributing to any known sources.
- Substitute all sensitive information (attributes) into new generic details, based on a created anonymised standard. This standard should allow researchers to attribute back any sensitive data but not the users of the visualization.
- Secure the anonymised standard for references.

Why the need for a data standardisation process? Standardization procedures are taken to allow datasets of various formats be used across numerous analytic tools. This allows interested security researchers to easily integrate these datasets with data analytics and threat intelligence tools. The entire standardisation process is executed after the required datasets are collected. The process of analysing the collected data is done using three methods: (1) manually analysing logs and identifying their existing format (knowing how many attributes and types of delimiters used), (2) identifying and categorising different attacks by

Table 3.2: Dynamically Storing Attacks into the Database.

ID	Time	Source	Destination	Protocol	Command	Attack Type
26	18:29:28	10.0.53.4	10.42.122.123	TCP	nmap 10.42.122.0/24	Reconnaissance
27	18:29:28	10.0.53.4	10.42.122.151	TCP	nmap 10.42.122.0/24	Reconnaissance
28	18:29:28	10.0.53.4	10.42.122.200	TCP	nmap 10.42.122.0/24	Reconnaissance
29	18:29:28	10.0.53.4	10.42.122.60	TCP	nmap 10.42.122.0/24	Reconnaissance
30	18:29:43	10.0.53.4	10.42.122.11	TCP	nmap -sT –top-ports=100 10.42.122.0/24	Reconnaissance
31	18:29:43	10.0.53.4	10.42.122.123	TCP	nmap -sT –top-ports=100 10.42.122.0/24	Reconnaissance
32	18:29:43	10.0.53.4	10.42.122.151	TCP	nmap -sT –top-ports=100 10.42.122.0/24	Reconnaissance
33	18:29:43	10.0.53.4	10.42.122.200	TCP	nmap -sT –top-ports=100 10.42.122.0/24	Reconnaissance
34	18:29:43	10.0.53.4	10.42.122.60	TCP	nmap -sT –top-ports=100 10.42.122.0/24	Reconnaissance
35	18:29:57	10.0.53.2	10.42.122.200	TCP	/usr/bin/python /usr/bin/sqlmap -u http://10.42.12...	SQL Injection
36	18:30:24	10.0.53.3	10.42.122.200	HTTP	GET /adminlogin action?username=&password=...	URL Manipulation
37	18:31:18	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=%27&password=...	URL Manipulation
38	18:31:29	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=Admin&password=...	URL Manipulation
39	18:31:59	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=Admin&password=...	URL Manipulation
40	18:32:49	10.0.53.2	10.42.122.200	TCP	/usr/bin/python /usr/bin/sqlmap -u http://10.42.12...	SQL Injection
41	18:33:26	10.0.53.2	10.42.122.200	TCP	/usr/bin/python /usr/bin/sqlmap -u http://10.42.12...	SQL Injection
42	18:35:01	10.0.53.3	10.42.122.200	HTTP	GET /adminlogin action?username=&password=...	URL Manipulation
43	18:35:48	10.0.53.1	10.42.122.200	HTTP	GET /post/create action?name=Admin&date=12%2F%...	Cross Site Scripting (XSS)
44	18:35:51	10.0.53.3	10.42.122.200	HTTP	GET /adminlogin action?username=&password=...	URL Manipulation
45	18:38:37	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=%3C%3ECoolGuy...	URL Manipulation
46	18:39:59	10.0.53.3	10.42.122.200	HTTP	GET /adminlogin action?username=	URL Manipulation
47	18:41:02	10.0.53.1	10.42.122.200	HTTP	GET /post/create action?name=NewAdmin&date=12%2F%...	Cross Site Scripting (XSS)
48	18:41:20	10.0.53.1	10.42.122.200	HTTP	GET /post/create action?name=%3C%3ECoolGuy...	Cross Site Scripting (XSS)
49	18:42:03	10.0.53.1	10.42.122.200	HTTP	GET /post/create action?name=%3C%3ECoolGuy...	Cross Site Scripting (XSS)
50	18:42:12	10.0.53.1	10.42.122.200	HTTP	GET /adminlogin action?username=Mark&password=...	URL Manipulation

analyzing all different types of logs, and (3) creating scripts to dynamically and automatically anonymise the dataset based on analysis and insert them into database tables. Table 3.2 shows anonymised data being categorised into different types of attacks and stored into databases. Such scripts include regular expressions that are being used to search and match rows or excluded rows in various logs. Examples of excluded rows are 'commented information' and 'duplicated information' which do not contribute to how security attacks are executed. This script acts as a '*Collector*' mechanism that checks for new data inputs, anonymise the inputs and inserts them into respective database tables.

3.7 Summary

In summary, the demand to provide security visualization for web and mobile platforms had pushed us to obtain multiple types of datasets for this thesis. However, the challenge of collecting and sharing datasets freely has proven difficult due to privacy and confidentiality reasons. This led to executing several steps to collect datasets: (1) obtain proper ethical procedures to acquire relevant ethical approvals for dataset collection, (2) establish research collaboration with interested organisations to acquire the relevant datasets for this research, and (3) simulate cyber-attacks using cyber security challenges to collect datasets. As a result, the datasets collected from multiple trusted sources are: The New Zealand National Cyber Security Challenge (NZCSC) events, malware datasets, and Bitcoin datasets. The raw datasets collected are of different types and sizes, therefore, they have been anonymised and

standardised to meet both ethical and visualization requirements before processing them for security visualizations.

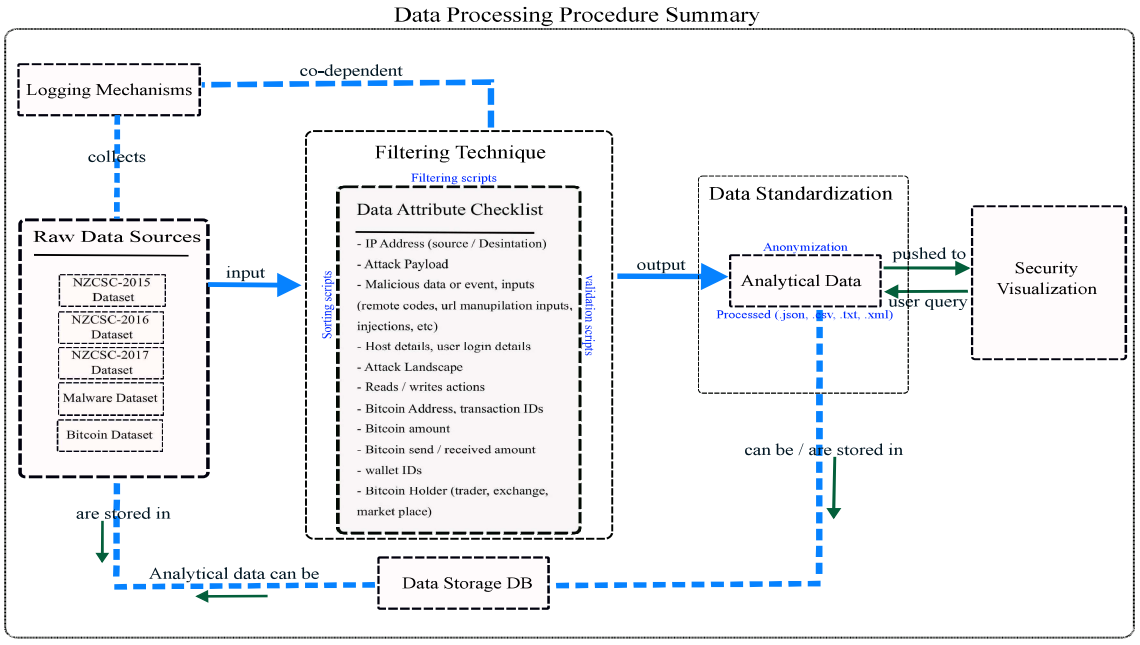


Figure 3.10: A Summary of our entire Data Collection Schematics

The multiple data collection scenarios and landscapes enabled us to implement and provide a wide range of security visualization for both web and mobile platforms. Hence, in this chapter, we delivered our NZCSC-2015, NZCSC-2016, NZCSC-2017 and the bitcoin data collection schematics to show further details on how data are collected for this research. In addition, we delivered the NZCSC-2017 round-2 attack network infrastructure that was designed as part of our data collection environment.

We provided a summary of the entire data collection schematics approaches shown in Figure 3.10. It outlines all data sources and relevant events associated with the preparation of data to meet our security visualization requirements. We conclude by stating the key components of our data collection framework and security visualization framework, which are: (1) logging mechanisms, (2) raw data sources, (3) filtering and processing techniques, (4) data standardisation method, (5) data storage (database) and (6) the security visualization frontend.

Chapter 4

The Effectiveness Measurement Algorithm for Security Visualization

The English Oxford dictionary [147] defines '*effectiveness*' as "the degree to which something works efficiently and produces the intended results." However, when using the term in the context of visualization, effectiveness is seen as a result with respect to representation and presentation of data. In different research fields of visualization, effectiveness appears basically in several areas: (1) perceptual context, (2) user context, (3) application context, (4) representation and (5) presentation context. This makes effectiveness a challenging task. Therefore, the demand of a measurement scale/indicator around effectiveness in visualization is needed. For example, effectiveness studies in information visualization have conducted research through the perceptual law context as seen in ranking of visualization correlations using Weber's law [148]. Scatterplot, parallel coordinates, bar and donut charts are used to show and compare correlation ranking presentation performance [70]. Moreover, assessing graphical data presentation techniques for large datasets with respect to expressiveness, efficiency and effectiveness is another form of measuring effectiveness [3].

However, adding the concept of security as the primary entity into the visualization scope and purpose, effectiveness becomes a 'full-scale' approach. This is due to multiple reasons: the landscape (nature), information sensitivity, urgency of efficient data representation and presenting security visualization. In addition, the need to present and communicate security related information through visualization has to be simple and effective. Furthermore, security information presented has to be accurate with a high degree of delivering precise correct and useful knowledge without allowing users to enter into the realm of cognitive biases. Therefore, this chapter looks at delivering a theoretical solution by providing an effective security visualization measurement algorithm for mobile and web platforms.

4.1 Motivation of This Chapter

Current mobile visualizations serve various specific purposes and often users respond with their views and preference around whether such visualizations are contributing to delivering the required information. However, one of our hypotheses for this thesis is to know if given two or more security visualization (SVis sample-1, SVis sample-2), can we say SVis sample-1 is better than SVis sample-2, or vice versa? This requires a proof to be made, tested and verified, for example, a theoretical proof to show the above hypothesis. Therefore, this chapter outlines our core contribution and presents our 'security visualization effectiveness measurement (SvEm)' algorithm, SvEm design, components, and application scenarios. As technologies and applications are customer (users) driven, mobile platforms are becoming the main and most popular personal electronic devices used on a day-to-day basis. However, on the flip side of technology, mobile threats have increased rapidly, being around 77% out of the total cyber-attacks recorded in 2014 by Kaspersky Labs [149], [150]. In 2017 Q2 and Q3, mobile cyber-attacks increased, ranging from data leakage, SMS attacks, mobile (banking, etc.) malware, phishing and ransomware attacks [151], [152], [153]. Despite existing security tools being implemented, users may not see how critical or severe a mobile cyber-attack is until it is too late to act in stopping the attack from escalating, creating further damage. Therefore, we propose and provide our SvEm framework to aid users with the use of security visualization. Our SvEm approach presents a full-scale effectiveness measurement methodology for security visualization for mobile platforms. Due to limitations around mobile platforms, the primary emphasis of our SvEm approach basically tackles these following areas:

1. **Security data representation and pre-processing technique**: this refers to the data representation design and pre-processing methodology
2. **Security visualization sample**: this refers to the visualization presented
3. **Users (targeted audience)**: this refers to specific users to whom our security visualization is presented
4. **User perception and attention span**: this refers to the SvEm methodology used to address effectiveness with respect to user perception and attention span
5. **Interaction process**: this covers both communication and response between users and the security visualization presented

Based on the five key areas of our approach, our SvEm algorithm and applications are built on a user-centred approach whereby predefined/pre-processed security information

is effectively communicated to the targeted audience. This involves understanding each component that makes up the algorithm and theory. Therefore, the remaining chapter provides the proposed algorithm, the SvEm architectural design, SvEm model requirements and finally expands around the SvEm applications.

4.2 Security Visualization Effectiveness Measurement (SvEm) Algorithm

The notion of *effectiveness* in any application and services is best measured by results over time. It relies on training data, guidelines, specific standards, performance results and feedback. Hence, the existance of Chapter 3 is to facilitate the required datasets for our SvEm algorithm. However, measuring effectiveness requires specific architectures and designs to enable users to set up measurement collection points and locations whereby effectiveness is measured. Both applications and users are used to collect effectiveness results. In our SvEm model, users are trained within the first few seconds of confronting a given security visualization to help capture their interest in the task provided. This is an example of an effectiveness measurement collection point whereby users' attention spans are being assessed with regards to time a user can interact and concentrate on a given visual task.

In the training and visual observation process, our SvEm framework equips a viewer with basic cyber-attack landscape knowledge and visual interaction methodologies by providing guidelines and visualization samples to preview. These training techniques leverage on both the users of our SvEm framework and the security visualization presented. Our training techniques enabled users to tap into the use of known application interactive features, user-trigger features and the application of famous colour usage. These famous colours are adopted from popular or familiar cases such as the use of traffic lights (red, yellow, green) and the Interpol's notice system [154]. This enabled the user's training progress to be efficient and effective. For example, with the traffic system, a 'red' colour symbolises Stop/Danger and 'red' alert in the Interpol Notice system represents 'wanted person'. Such representation is replicated across into security visualization to maintain familiarity and therefore enhances the user's working memory capacity. A 'red' coloured circle in our SvEm security visualization represents a known and identified malicious entity (e.g., malicious IP address or file). The establishment of this concept described in this paragraph provided us the means to expand our SvEm algorithm and to explain how the functionalities that make up the algorithm. The SvEm algorithm, seen in Equation: 4.2 and Equation: 4.3 is defined with the following components and variables.

(SvEm) Distortion (d_{svem}) Theory Assessment

$$\mathrm{SV}_{val} = \frac{(w * h)}{\mathrm{Sv}_f * d_n} \triangleright (\mathrm{Sv}_f * d_n) \neq 0 \tag{4.1}$$

$$\mathrm{SvEm} = \frac{\mathrm{SV}_{val}}{(\mathrm{Cl} * n_{clicks})/t_{me}} > 50\%(\mathrm{D}istortion) \triangleright ((\mathrm{Cl} * \mathrm{n}_{clicks})/t_{me}) \neq 0 \tag{4.2}$$

(SvEm) Time (t_{svem}) Theory Assessment

$$\mathrm{SvEm} = \frac{(\mathrm{Cl}/t_{me})}{n_{clicks} * \mathrm{Sv}_f/d_n} \geq 0.25sec(s)(\mathrm{T}ime) \triangleright (\mathrm{n}_{clicks} or (\mathrm{Sv}_f * d_n)) \neq 0 \tag{4.3}$$

Where:
*w * h : Web/Mobile display area (dimensions)*
Sv_f *: Security visual nodes (e.g., Infected-IP, timestamps, etc.)*
d_n *: n-dimensional view in security visualization*
Cl : Cognitive Load (Identifiable attributes (quantity) - Prior knowledge)
t_{me} *: Memory efficiency (Effort based on working memory - Time-base)*
n_{clicks} *: Number-of-clicks on visualization*

However, we begin with declaring and making certain assumptions to allow our SvEm algorithm to be valid. We state that the following variables: Sv_f; d_n; Cl; n_{clicks}; t_{me}, should have a default value of 1, thus, each SvEm variables should not have zero values. For example, in order to see a visualization, a click-event of value 1 is automatically executed by default. However, we establish several assumptions as part of the scope to achieve reasonable effectiveness measurement ratings. As a result, errors are minimised and appropriate SvEm results are achieved. These assumptions are as follows:

1. If a user is able to understand a given security visualization within *less than* 5 seconds, then visualization has effective SvEm output.

2. Input data has to be within the capable processing power of the mobile platform used.

3. User must always have some form of cognitive ability (Prior knowledge) before engaging the given security visualization.

4. Number-of-clicks (n_clicks) refers to number of clicks (navigating) on the mobile platform screen to the point where the first known security attribute has been identified.

4.3 The Security Visualization Effectiveness Measurement (SvEm) Components

In security visualization, measuring effectiveness requires certain techniques whereby required components and attributes act as the measurement entity. Thus, by analysing existing techniques and understanding how they function, we introduce our new effective measurement techniques which measures the effectiveness in both a given security visualization and the viewer's experiences in security events. It is based on both the visualization approach and user intuition. However, to achieve effectiveness, we need to minimise the time (duration) spent on viewing a visualization and making sense with the insights portrayed in a visualization. These SvEm theorem components are [13]:

1. *Web/Mobile platform screen surface area (w * h)*: This refers to the surface area used to display a security visualization. Screen sizes have a great impact on how visualizations appear.
   A "display screen dimension" parameter is a key requirement for this theorem. Various screen dimensions have a great impact on how users interact with the visualization presented. Bigger screen dimensions allow greater visualization information to be presented, on the other hand, small screen dimensions create visual challenges as well as allow room for further effective visualization development. Hence, having this parameter is critical to the effectiveness in user interactions.

2. *Security visual nodes (Sv_f)*: These are known security attributes identified in a visualization, e.g., an malicious or infected IP address. The security visual nodes variable range from 1 -> 1000 nodes.
   The "security visual node" parameter enables this theorem to tag malicious attributes identified in collected datasets. It provides visualization developers the ability to manage malicious or security attributes shown in a given visualization. The parameter also allows this theorem to help measure effectiveness when users are given a choice to visualize between two or more different visualization.

3. *N-dimensions*: The N-dimensions parameter refers to how many visual dimensions plane (view) are used to represent visualization. The higher number of dimensions used for the visualization presentation, indicates the greater depth of data able to be

represented over a visualization screen. The n-dimension variable ranges from 1 -> 6 dimensions. The performance or capability of how many dimensions utilized for a given visualization, is subjected to the processing power of the visualization system (e.g. mobile platform, plasma screen). The existence of this parameter contributes to effectiveness in visualization in respect to how many visual nodes and information can be presented instantly to a user.

4. *User Cognitive load (Cl)*: This is critical to how effectiveness is assessed and measured in this theorem. It is based on how much knowledge (prior knowledge) a user has around the expected visualization. It is the prerequisite security knowledge of the expected security events such as a malware cyber-attack. Users dictate how much information they are will to visually process within certain time and/or execute a self-assessment phase to measure effectiveness in the given security visualization. Hence, user cognitive loads provide the preliminary factor as to whether a user would be interested or not, and if they are interested, how fast can they interact wit the presented visualization and gain insights. The user cognitive load (Cl) variable has a default value of 1 and range increases accordingly based on the users.

5. *Memory efficiency (t_{me})*: This is a time-base attribute which measures how fast one can recall security related attributes. Memory efficiency is critical to this theorem because it is the primary time-assessment factor or attribute in this theorem. Time is a characteristics of effectiveness, i.e. obtaining a minimal time period to gain security insights when interacting with a given visualization relies on the users memory efficiency and cognitive load. The user's memory efficiency (t_{me}) variable has a default value of 1 and increases accordingly based on the time taken to interact with a given visualization.

6. *Number-of-clicks (n_{clicks})*: The "Number-of-clicks" parameter is regarded as the self-assessment factor, whereby users perform while navigating around a given visualization. Over a period of time visualizing and analyzing security visualizations, an average number-of-clicks recorded is used to understand effectiveness in security visualization. The number-of-click parameter refers to how many 'touch gestures' or 'clicks' one has to perform on the mobile platform screen or other display screens in order to view the presented visualization. With a default value of 1, the number-of-clicks variable require less clicks as possible to achieve the most ideal effectiveness value.

With a clear understanding of the SvEm algorithm and its components, we derived our SvEm theory (framework) to address two areas of effectiveness in security visualization for web and mobile platforms. These areas are: (1) '*distortion rate*' in security visualization and

(2) the observation'*time*' period. The ideal distortion rate baseline value has to be greater that 50% visual clarity. However, this minimum value provides an effectiveness measurement approach whereby the overall assessment is measured against a '*high*' or '*low*' rating. This allows our effectiveness measurement results to be more realistic. However, the factors affecting our SvEm-distortion rating are: (1) phone dimensions and resolution, (2) user knowledge and (3) the number of clicks users execute. In addition, SvEm-time component is measured against a minimum constant value: *0.25 seconds*, which is known in psychology research as the least minimal cognitive time required for a human to process and understand information. This process has been assessed through studying human perception in various psychology research studies. We adopt this concept to measure the time taken from the moment a visualization is presented to a user to the time users' highlight known visual nodes, patterns and knowledge. Thus, our overall SvEm-time assessment results are calculated and presented as an average rating against many other samples. Overall, the SvEm algorithm is summarised with the following points:

1. **Visualization Distortion**: Effectiveness in distortion is defined when a visualization has *greater than* 50% visual clarity
    - How clear and readable visualization is presented
    - Features and attributes presented are visible
    - Visual patterns emerge into reality for observers.
2. **User Response Time**: Duration in milliseconds (ms)
    - Analytical time of processing the visualization
    - Time of user- cognition to recall known memory

Based on this theory, all SvEm variables are defaulted to a numeric value of 1 (one). The factors highly contributing to a high SvEm value are: (1) *w * h: smartphone dimensions* and (2) d_n*: n-dimensional view of security visualization*, i.e., a 3-dimensional representation visualization view has proven less distorted than a single dimensional visualization view. The advantage of presenting more data information are visible in higher n-dimensional visualization views. This increases the user's (viewer) ability to provide a higher count for $\mathrm{S}v_f$. The less value of n_{clicks}, indicates the overall *time* spent on viewing the visualization is reduced therefore a higher effectiveness measurement outcome is achieved.

4.4 SvEm Design Architecture

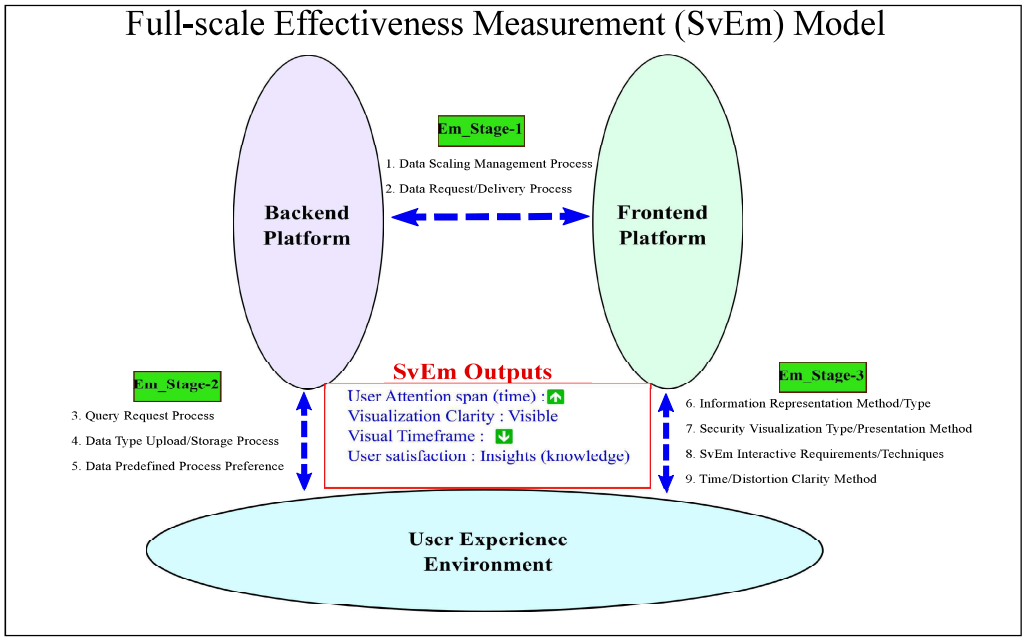


Figure 4.1: A Full-Scale Effectiveness Measurement Model

An overview of our proposed full-scale SvEm model with effectiveness measurement attributes, methods and processes is shown in Figure 4.1. This SvEm model has nine effectiveness measurement attributes and processes that are considered and attended to. They range from 'data scaling management process' to 'time/distortion clarity presentation method and output.' Our overall intention of producing such design was to address effectiveness throughout the entire framework. When less time is spent observing while there is an increase in the user's attention span, this provides and delivers an effective security visualization presentation to the viewers. This is a visual experience achieved by the audience, thus, we provide a full-scale effectiveness measurement model with a user-centric observation and interactive environment.

4.4.1 SvEm Data Processing Design

Due to large and unstructured datasets (raw data), scalability and rendering challenges have hindered effective presentation of security visualization in mobile platforms. The pitfall caused by various types/range of audiences with different visual preferences and levels of understanding on the information presented through security visualization adds further design and implementation challenges. Therefore, with the dataset at hand, several critical

implementation and monitoring stages, known as 'effectiveness insertion approach' stages (Em_Stage-1, Em_Stage-2 and Em_Stage-3 (See Figure 4.1)), are required and executed accordingly to provide a full-scale effectiveness experience in a mobile security visualization environment. It is critical to pay attention to these stages and execute them in order to provide an effective measurement technique in security visualization. These effectiveness insertion approach stages comprise of:

⋆ **Em_Stage-1**:

1. Data Scaling Management Process
2. Data Request/Delivery Process

⋆ **Em_Stage-2**:

3. Query Request Process
4. Data Type/Storage Process
5. Data Predefined Process Preference

⋆ **Em_Stage-3**:

6. Information Representation Method/Type
7. Security Visualization Type/Presentation Method
8. SvEm Interactive Requirements/Techniques
9. Time/Distortion Clarity Method

General Description of Em_Stage-1:

The question of concern here is: *How would implementing and monitoring this stage contribute to effectiveness measurement in security visualization?* While this stage is not new to any application implementation, our framework pays more attention to improving our implementation and monitoring environment of this effectiveness insertion stage (Em_Stage-1). This stage ensures an understanding of the data at hand and using it to its maximum potential for insights. It handles the entire effectiveness measurement implementation between the backend and frontend of our SvEm framework that executes two primary tasks. These are:

- Data Scaling Management Process: This task comprises several steps, with the primary role being preparing security data and making sure data is managed efficiently throughout the entire data journey of the security visualization process. Firstly, it involves standardising and normalising the data type, i.e., categorical, ordinal or interval type. Secondly, it involves understanding the required data process, i.e., hierarchical data, multivariate data, graph data, etc. Thirdly, it involves understanding the volume of data expected for the framework prior to processing and pushing it to the security visualization frontend platform. This requires the understanding of the data processed and the nature at which the data was collected. Hence, knowing if the data is in a real-time format, large data volume request, or static data request? Finally, applying a data scaling input process with respect to the mobile display screen size. The scaling mechanism is outlined below:

$$Let: = \left\{ \begin{array}{l} dm = \text{data size}, \\ n_s = \text{scaling}, \\ w = \text{mobile display screen width}, \\ h = \text{mobile display screen height}. \end{array} \right\} \tag{4.4}$$

whereby scaling factor: dm/n_s

therefore data scaling input:

$$(w * h) * dm/n_s = \left\{ \begin{array}{l} \text{if data input = 4 * for every 4000 input value}, \\ \text{if data input = 3 * for every 3000 input value}, \\ \text{if data input = 2 * for every 2000 input valu}, \\ \text{if data input = 1 * for every 1000 input value}. \end{array} \right\} \tag{4.5}$$

- Data Request/Delivery Process: This task involves listening and collaborating with the users (viewers) request from the security visualization frontend. Firstly, setting up data tables that allow fast query processing and polling is important. Secondly, due to the type of visualization implemented at the frontend, i.e., real-time circular visualization design, data processed can be efficiently handled to ensure proper delivery with minimum time required for the process.

General Description of Em_Stage-2:

The key factor to providing and improving effectiveness measurement in security visualization for Em_Stage-2, is to provide all means of help a user (viewer) needs for a seamless experience when handling data, data requests, data storage and ensuring multiple data preferences for both the security visualization platform and users. Em_Stage-2 handles our effectiveness measurement technique implementation between the back-end and the user experience environment. The Em_Stage-2 executes three tasks as outlined below:

- Query Request Process: This task performs similar roles to the 'Em_Stage-1: Data request/Delivery Process' but from a user's (viewer) receiving end. Providing visual instructions and locations such as a 'search-box' or a clickable 'upload' button within a visual and easy to reach security visualization environment is the approach implemented.

- Data Type/Storage Process: This task plays an important role in the overall effectiveness measurement process in security visualization. It refers to the malicious data containing cyber-attack details, patterns and behaviours. These data types distinguish the visualization environment (for example, network packets) and involves several states of processing and storing them in a database. Attack scenarios are then triggered by the frontend visualization platform dictate the types of data requested.

- Data Process Preference: This task relies on two components. First, from the backend perspective, raw data contains noise (unwanted visual data/information) which is not all required for the actual viewing or visualization. Secondly, with preprocessed sets of instruction, data is made available for the visualization frontend. This task is often processed according to the user's preference.

General Description of Em_Stage-3:

Em_Stage-3 covers the implementation process between the frontend and the user experience environment. It is the most crucial set of tasks involving presentation components that attract the users (viewers) to the security visualization. These tasks involve the following:

- Information Representation Method/Type: This task handles data transformation from raw data to predefined data attributes compatible to visual languages, i.e., transformed into visualization types. For example, entities are depicted using shapes (e.g., circles) and colours. However, the challenging task is to select the appropriate representation

method/type to use for security visualization. This should comfortably connect with the user's preference and relatedness to the cyber-attack visualized.

- Security Visualization Type/Presentation Method: This task handles the way security visualization is being presented. Based on the user preference, request and/or cyber-attack landscape, various security visualization presentation methods are implemented to accommodate user selected choices. Our SvEm framework provides several presentation approaches: (1) real-time, (2) static and (3) user-based interactive approach.

- SvEm Interactive Requirements/Techniques: The emphasis of this task is on triggering, capturing and maintaining the user's attention with interactive features. This is done with the purpose of increasing the user's attention span throughout the visualization confrontation period.

- Time/Distortion Clarity Method: This task contains crucial effectiveness measurement implementation and monitoring processes. It provides a set of measurement techniques by observing time and the visualization clarity presented in the security visualization frontend. Time is measured in two ways: (1) through an increase in user attention span therefore increasing the working memory load, (2) through a reduced or the least amount of time required, from the moment a user confronts a visualization and interacting with it till known security insights are gained. Distortion is, however, implemented and measured against the mobile display screen dimensions. Visibility with appropriate security entities shown via the visualization presented adds to how effective a security visualization appears.

Finally, as seen in all Em_Stages, the notion or idea of providing and measuring effectiveness in security visualization for mobile platforms is executed throughout the whole process from the moment data is received, processed right through to visualizing it on a mobile platform. This enables all potential 'SvEm Outputs' to be achieved.

4.4.2 SvEm Model Requirements

In addition to our SvEm design architecture, multiple theoretical requirements are needed to establish an understanding between the SvEm model and interaction features. These requirements allow users to understand how our SvEm framework works, particularly interactive relationships between mobile platforms, users (viewers) and the security visualization presented. Below are the requirements and their corresponding details:

1. **Correlation Indicators**: these are predefined indicators between interested entities. For example, a ransomware attack visualized on a mobile platform has payloads, infected IP and command-control-centre (CCC) link attributes. Correlation indicators are case-based depending on threat/attack landscape and method used.

2. **Correlation Links**: these links show the relationship between two entities of interest, e.g., both infected source IP address (attacker's IP address), a destination IP address (victim's IP address) and communication via network packets. These links also include the method of attack, whether it is via remote code execution, scripts and payload dumps.

3. **Security Entities**: these are devices, services, platforms and users. Entities allow security experts to measure how big the threat/attack landscape is.

4. **Perceptual Trigger**: this requirement exists within users (viewers) and has a direct relation to the existence of '*known working memory*'. It contributes to instant moments where a user (viewer) makes sense of the given security visualization.

5. **Colour Management**: This requirement is a practical requirement based on the choice of selecting a colour to aid all other SvEm model requirements. For example, using two colours with the same RGB could make the visualization confusing. Colour application and user-centric classification (e.g., clustering of data) features are seen as part of the colour management process.

6. **Precision and Accuracy Rating**: This rating involves the user's ability to select known or interesting security visualization feature to extract insights. Precision and accuracy depends on the ability of a user to either have security knowledge around the presented security visualization or not.

These SvEm requirements contribute to the entire framework with the aim of facilitating a perceptual concentration environment for users to feel comfortable when viewing security visualization on their mobile platforms. Equally important, these requirements maintain our effectiveness measurement technique with respect to distortion, interaction time and user response.

4.5 SvEm Technical Components and Aspects

In addition to our SvEm design architecture and model requirements, a breakdown of the proposed SvEm algorithm is explained in the remainder of this chapter. Each variables and

their functionalities are discussed with various assumptions which could affect the SvEm algorithm performance. We begin with addressing various variables in Equation 4.7 and Equation 4.8.

4.5.1 SvEm-Distortion Theory Approach

SvEm-Distortion (d_{svem}) Theory Assessment

Calculate SV_{val} *ratio against mobile dimension:*

$$\mathrm{SV}_{val} = \frac{(w * h)}{\mathrm{S}v_f * d_n} \triangleright (\mathrm{Sv}_f * d_n) \neq 0 \tag{4.6}$$

Then:
Calculate SvEm-Distortion rate:

$$\mathrm{S}v\mathrm{E}m = \frac{\mathrm{SV}_{val}}{(\mathrm{C}l * n_{clicks})/t_{me}} > 50\%(\mathrm{D}istortion) \triangleright ((\mathrm{Cl} * \mathrm{n}_{clicks})/t_{me}) \neq 0 \tag{4.7}$$

Where:
*w * h : Web/mobile display area (dimensions)*
$\mathrm{S}v_f$ *: Security visual nodes (e.g., Infected-IP, timestamps, etc.)*
d_n *: n-dimensional view in security visualization*
Cl : Cognitive load (Identifiable attributes (quantity) - Prior knowledge)
t_{me} *: Memory efficiency (Effort based on working memory–Time-based)*
n_{clicks} *: Number-of-clicks on visualization*

Our *SvEm-Distortion* theory, seen in Equation 4.7, seeks to provide reasonable security visual nodes per appropriate mobile dimension. The number of visualization presentation dimension (d_n) used to represent security information enables us to manage how we present predefined security visual nodes to avoid distortion and complexity as much as possible. Firstly, a security visualization (SV_{val}) ratio is calculated against the mobile display area. This allows our framework to present reasonable security visual nodes per view. Secondly, user's cognitive load (Cl) with working memory (t_{me}) load are used in association with the number of clicks (n_{clicks}) done per visual view.

The SvEm-distortion (Equation 4.7) approach is implemented in such a way that the number of data visualized should be controlled while observed in a n-dimension visual space.

This triggered the need to deliver a theorem that would assess image distortion at a real-time based approach. Hence, the SvEm-distortion approach relies on the number of security visual nodes presented, n-dimension view, the users' cognitive and working efficiency. In addition, the web/mobile dimension provides the theorem with a visual area limit that can be used for effectiveness measurement. All variables required for this equation are described according to their contribution in the whole algorithm, thus, Algorithm 1 provides our SvEm algorithm pseudocode to help explain the SvEm theory.

Algorithm 1 Algorithm for Effectiveness Measurement with Respect to Distortion

1: **procedure** SvEm-TIME($\mathrm{S}v\mathrm{E}m = \frac{\mathrm{SV}_{val}}{(\mathrm{C}l * n_{clicks})/t_{me}} > 50\%$ (Distortion))
2: Input w
3: Input h
4: Input $\mathrm{C}l$
5: Input t_{me}
6: Input n_{clicks}
7: Input $\mathrm{S}v_f$
8: Input d_n
9: **while** $\mathrm{S}v_f \neq 0$,$\mathrm{d}_n \neq 0$ **do** ▷ Undefined if d_n and t_{me} is 0
10: $\mathrm{SV}_{val} = \frac{(w*h)}{\mathrm{S}v_f * d_n}$
11: Input SV_{val}
12: **while** $\mathrm{C}l \neq 0$,$\mathrm{n}_{clicks} \neq 0$,$\mathrm{t}_{me} \neq 0$ **do** ▷ Undefined if $\mathrm{C}l$, d_n and t_{me} is 0
13: $\mathrm{S}v\mathrm{E}m = \frac{\mathrm{SV}_{val}}{(\mathrm{C}l * n_{clicks})/t_{me}}$
14: **return** $\mathrm{S}v\mathrm{E}m$ ▷ SvEm is > 50% : 'high' or 'low'

Mobile Display Area (dimensions) Variable:

Mobile devices come with various screen display sizes and dimensions. A mobile device size refers to how much information it can store within the display space. For example, they may be small, medium and large in size. The term size in this thesis is interchangeable with mobile platform size in an overall perspective. A standard dimension refers to the level of details with respect to the mobile display length, width and height. However, in this thesis, a mobile display dimension refers to width (w) and height (h). In addition, the display resolution (pixel and density) is taken into consideration, however it is treated as a limitation factor in the distortion outcome. For example, the Apple iPhone 8 (Figure 4.2), with a physical size of 4.7 inches has dimensions - width: 67.3mm (750px), height: 138.4mm (1334px) and 326px per inch, compared to a Samsung Galaxy S8+ with physical size of 6.2 inch with dimensions

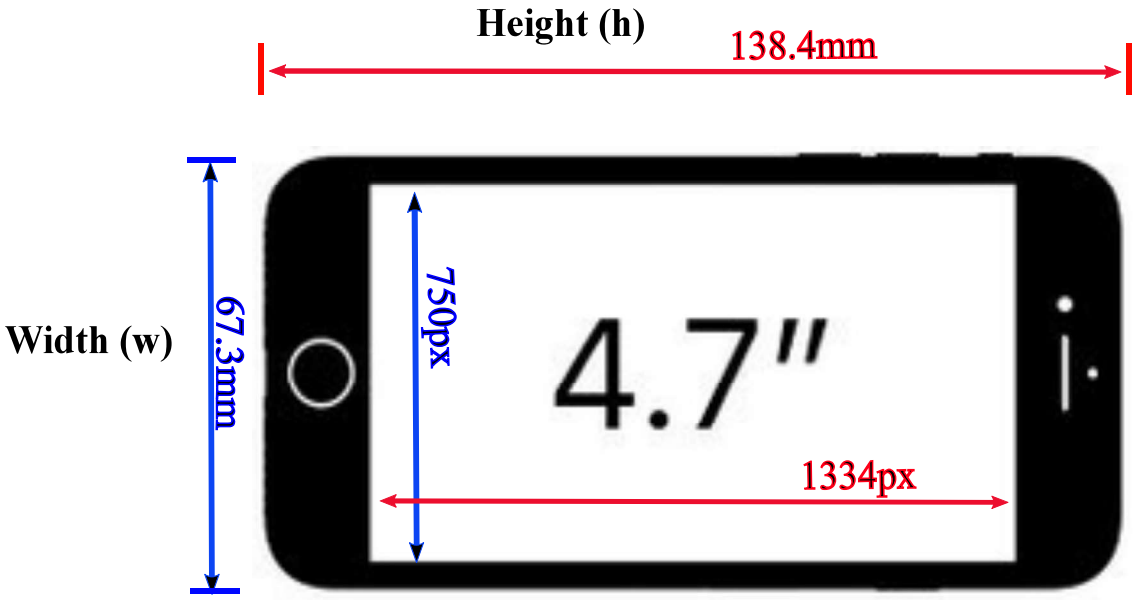


Figure 4.2: A 4.7inch iPhone 8 Mobile Dimension Details

- width: 73.4mm (2960px), height: 159.5mm (1440px) and 529px per inch. Both mobile devices would produce different outcomes due to different display area volume. However, effectiveness measurements in security visualization for mobile devices are measured based on the 'level of details' shown and understood by the user (viewer). These are discussed in Chapter 5 and Chapter 7 of this thesis.

Security Visual Nodes Variable:

Security visual nodes (Sv_f) are arbitrary variables such as malicious IP addresses and payloads. These arbitrary variables are subjected to the user's working memory capacity during the security visualization viewing process and experience. There are correlations between these security visual nodes (Sv_f), cognitive load (Cl) and memory efficiency (t_{me}), which contribute as a result of the level of details shown in the mobile display space. Figure 4.3 shows a sample of visual correlations distinguishing correlations between security nodes. The links between all correlation relationships and the level of detail in a given security visualization enabled effectiveness measurement in our framework. This is achieved with respect to the visibility of all security visual nodes during a user's observation period. The ability to identify security visual nodes in the shortest time contributes to the overall outcome of effectiveness measurement in security visualization.

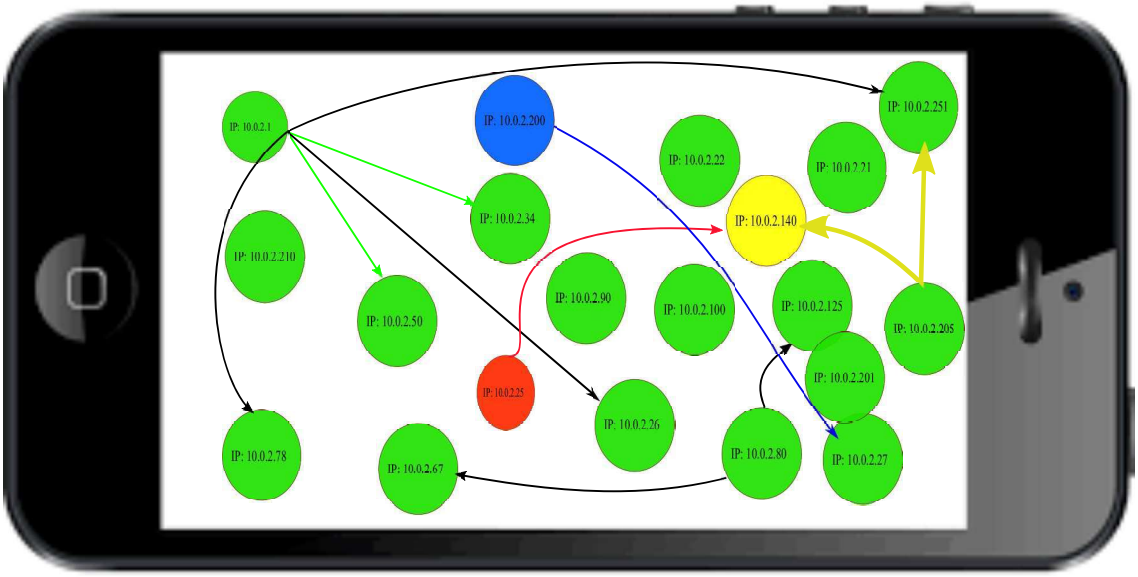


Figure 4.3: An Example of Correlation Relationships in Security Visual Nodes

N-dimensions View Variable:

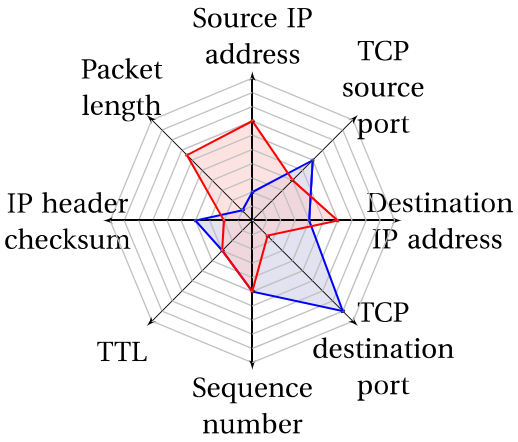


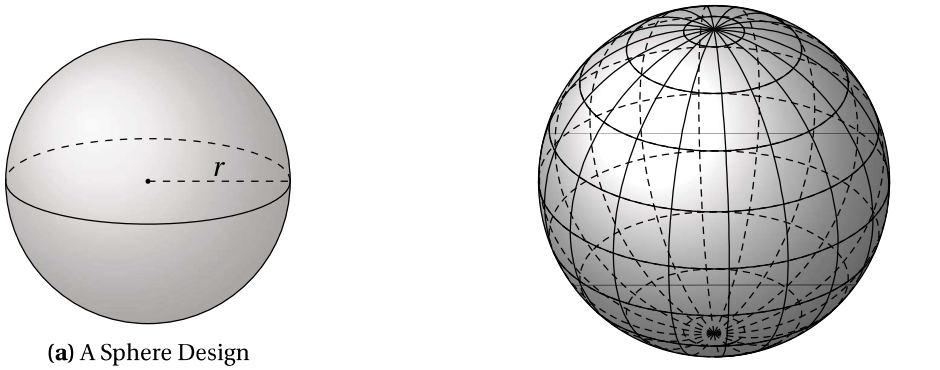
Figure 4.4: A Sample N-Dimension Visualization Representation Design

In theory, n-dimensional (d_n) visual models in visualization as shown in Figure 4.4 indicate the potential of showing multiple amounts and levels of detail in a visualization. We adopted the n-dimension view variable from the application of geometry with the use of parallel coordinates and spatial representation. With the notion of representing predefined security attributes and presenting abstracts of data, the core idea is to use projective n-dimensional (d_n) visual views to facilitate visualization of large dataset information in small

mobile platforms. With limited screen dimensions, obtaining a suitable and precise security visualization design such as a 3-dimensional (3D) 'circular' projective visualization model has advantages over a 2-dimensional (2D) model. A 3D visual model gathers for higher volumes of security attributes presented in a given visual space.

The Proposed 3-Dimensional Visual Representations

We proposed several 3-dimensional (3D) visual projective plane models: (1)sphere, (2) grid and (3) helix presentation designs for our security visualization framework. These projective models show data information using a 'coordinate' representation. Each coordinate in the visualization framework and model represent certain security data, especially when projecting volume data. Therefore, with the limited visualization space in mobile platforms, these projective plane models utilised the 3D environment with the n-dimensional presentation axis to show extensive load of security information.



(a) A Sphere Design

(b) A Sphere Grids and Presentation Points

Figure 4.5: N-Dimensional Sphere Visual model

1. *Sphere Visual Model*:

Our sequential sphere visual model had several variables of importance in the design phase and the purpose of offering such visualization. These variables, seen in Figure 4.5, include: (1) sphere size with respect to the radius (Figure 4.5a) of the projected sphere (usually signifies the volume of the data presented and visualized), (2) coordinates and grids (Figure 4.5b), which enables the visual mapping of security information and, finally, (3) the animated movement direction used to show visual information by allowing interactivity. These variables impact the way our visual model is presented on a mobile display platform. The ability to apply animated features to the sphere model, as depicted in Figure 4.6, provides

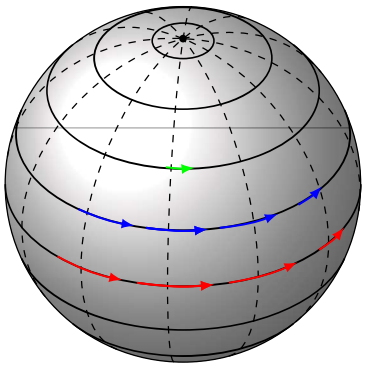


Figure 4.6: Sphere Visual Animation Movement Direction

the ability to scale and manage security visual representation attributes with respect to the mobile display dimensions and the load of data presented.

2. *Grid Visual Model*:

The projective visual grid model in Figure 4.7 is designed to accommodate large information volume and present it in a grid model with multiple layers. The layering model depicts a level of data granularity in an event of cyber-attack. By default, the 'top' and 'front' visual attributes are seen as recent information presented using our grid visual model. Moreover, security information is represented in an n-dimensional grid visual model to show different views of the attack landscape using various forms of classification, such as time-based details. In a grid 3-dimensional (*xyz*) visual model, large security datasets are presented in a small space (canvas) thus making it effective on mobile platforms.

3. *Helix Visual Model*:

In a similar concept to our sphere and grid visual models, our helix (spiral) visual model uses a 3-dimension (*xyz*) design approach to represent data in a sequential manner. The core purpose of this design is to enable data chaining or linkage with the concept of data provenance. The ability to add/push real-time data with rendering, scaling and compacting all in a helix form allows the user's view (eyes) to maintain visual concentration, therefore maintaining a positive visual attention span. As a result, the user does not exhaust his/her mental load and effort to acquire security knowledge. This contributes to providing an effectiveness measurement approach for our helix visual model. Figure 4.8 presents the model design whereby security visual nodes can be tagged and chained in a helix model to display security information with provenance related details.

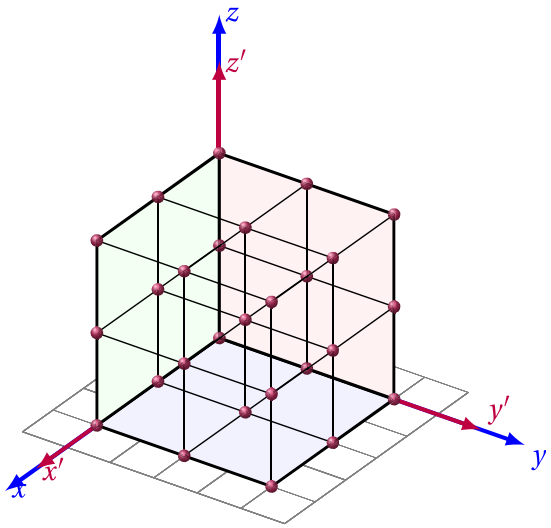


Figure 4.7: A Projective Grid Visual Model

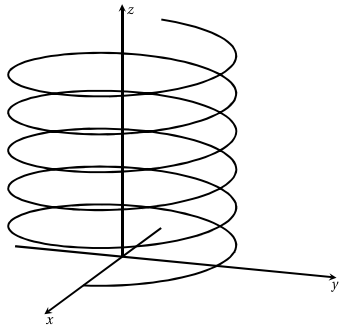


Figure 4.8: A Projective Helix Visual Model

To sum up, our n-dimensional view variable has three basic representation and presentation designs: a sphere model, grid model and Helix visual model. Using these n-dimensional visual models, we are able to represent and present large volumes of security data in a limited mobile platform display space while maintaining visualization clarity. The ability to comfortably render visual information using these 3 visual design models allows for a better visualization clarity approach with less distortion. Further details and evaluation will be discussed in Chapter 5 and Chapter 7 of this thesis.

Cognitive Load Value Variable:

The cognitive load (*Cl*) value variable is a crucial component to our whole SvEm algorithm and framework. In the context of this thesis, cognitive load is defined as a multidimensional construct load representing security events imposed on the user's (viewer) cognitive learn-

ing system while performing a particular security visualization task (s) [4], [155]. In this thesis, cognitive load is not simply considered as a by-product of the user's learning process but as the major factor of processing security data related information that determines the success of an instructional intervention. As such, theoretical effectiveness measurement in security visualization for mobile platforms has high rating and performance.

Cognitive load is observed as 'total cognitive load,' a sum of intrinsic, extraneous and germane cognitive loads. 'Mental effort' (resulting from working memory limits) and user 'performance' are key components contributing to the overall cognitive load measurement and visual experience. Thus, executing security visualization tasks such as identifying security visual nodes and applying visual classification of them affects the measurement of cognitive load in users (viewers). We utilise working memory load (mental load) and user performance to measure cognitive load. Both components are measured over a period of time with respect to 'time.' We inherited (see Appendix C: Appendix C1 and Appendix C2) methods from existing cognitive load measurement studies to design our simplified 'Secondary Task technique (ST)' [4], [156]. This allowed us to assess and measure the viewer's cognitive load with respect to the observed security visualization presented. Our secondary task technique measurement approach concentrates on measuring effectiveness through the user's interactivity with the visualization. This is done by assessing mental load and mental effort as key components of a user's working memory load and performance against several security visualization presentations.

Our derived secondary task technique (ST) involves assessing user interactions with the presented security visualization in two rounds. The first ST round involves presenting the security visualization without the use of instruction sets, standards/guidelines and the absence of 'user-trigger features' in our security visualization platform. The second ST involves presenting each security visualization with instructions, standards around the use of user-trigger features and restoring working memory knowledge of the security visualization presentation environment. We assess the changes in interaction ability and base the cognitive load rating on the number of security visual nodes identified against time. However, in terms of effectiveness measurement challenges, our current assessment uses a 'Best-case' scenario and a 'Worst-case' scenario to draw a scope of possible results.

Best-case scenario of measuring cognitive load: This involves the user (viewer) capitalising his/her working memory, knowing the sum of all; intrinsic load, extraneous load and germane load does not exceed the working memory limit.

Worst-case scenario of measuring cognitive load: This involves two cases: (1) the user chooses by default to bypass his/her working memory capacity during the mental processing period, thereby circumventing the limitations of working memory [4], and/or (2) the

user's cognitive load exceeds the working memory load limit and allows cognitive biases into the visual observation realm therefore affecting the user's final judgements.

Overall, activating cognitive load within the limits of working memory enhances the user's attention span by increasing his/her ability to concentrate while interacting with the security visualization presented.

Working Memory Variable:

Building on the concept of cognitive load, the human working memory (t_{me}) variable uses 'mental load' and 'mental effort' as aspects triggering the user's security knowledge cognitively. This originates from the interactions between the user and the visualization presented. Working memory values are based on prior (known) knowledge around the presented SvEm visualization, tasks and the security (subject) related characteristics of a cyber-attack landscape. Ideally in a SvEm visualization scenario, a viewer with working memory load has a higher chance of instantaneous understanding of the security visualization from insights gained. In our thesis, working memory is measured as part of the cognitive load measurement process as a 'time (t_{me})' value, whereby security visual nodes are identified by a user when observing a security visualization.

Number-of-Clicks Variable:

The number-of-clicks (n_{clicks}) variable refers to the number of times, a user would utilise a 'touch gesture' or 'click' event on the mobile platform surface to interact with a given security visualization. Ideally, the less number-of-clicks, the better the effectiveness measurement result. Interactive features such as 'zooming' and 'dragging' are not recorded as a click. Furthermore, a 'mouse-over' interaction feature function does not classify as a click but rather a user-trigger feature. In this thesis, the n_{clicks} variable is regarded as an added architectural feature with intentions of giving user eligibility to navigate their way around within a security visualization observation space. The notion of effectiveness measurement with respect to the n_{clicks} variable changes the moment users (viewers) confront a security visualization. This happens because different users (viewers) see and perceive visualizations differently and often for different intentions and reasons. Users naturally process visual models in various structural ways, therefore affecting how many times they apply a click function on a web or mobile security visualization.

4.5.2 SvEm-Time Theory Approach

Our *SvEm-time* theory approach has emphasis on the user's interaction period with the security visualization presented. The purpose of this approach is to assess 'time' issues during the observation period of security visualization. An ideal goal for our SvEm-time theory approach is to ensure the user's (viewer) attention span increases from the moment a user confronts a security visualization on his/her mobile platform. Thus, all variables in our SvEm-time theory play significant roles in achieving an effective outcome. This is done when the user utilises his/her attention span through the use of security visual nodes (Sv_f), n-dimension visual models (d_n), fewer numbers of clicks (n_{clicks}) and the ability to activate his/her cognitive load (Cl) within the limits of working memory (t_{me}) to gain security insights. As seen in Figure 4.8, our *SvEm-time* theory is calculated and measured against a time *constant: 0.25* seconds. This constant is a known standard in psychology studies as a rating of the least cognitive time required for a user to successfully process information with a high insight gain rating.

(SvEm) Time (t_{svem}) Theory Assessment

$$SvEm = \frac{(Cl/t_{me})}{n_{clicks} * Sv_f/d_n} \leq 0.25 sec(s)(time) \tag{4.8}$$

Where:
Sv_f *: Security visual nodes (e.g., Infected-IP, timestamps, etc.)*
d_n *: n-dimensional view in security visualization*
Cl *: Cognitive load (Identifiable attributes (quantity) - Prior knowledge)*
t_{me} *: Memory efficiency (Effort based on working memory - Time-base)*
n_{clicks} *: Number-of-clicks on visualization*

4.5.3 Mobile Platform Features

Mobile platform features are referred to as 'indirect effectiveness measurement-factors' in our SvEm theory process. While they do not affect direct ratings towards the entire SvEm calculation outcome, they enhance the user's performance, which affects the visual time rating results during viewing of various security visualizations. These mobile platform features in-

Algorithm 2 Algorithm for Effectiveness Measurement with Respect to Time

1: **procedure** SvEm-TIME($S\nu Em = \frac{(Cl/t_{me})}{n_{clicks} * S\nu_f/d_n} \leq 0.25 sec(s)$) ▷ Cl constant: 0.25sec.
2: Input Cl
3: Input t_{me}
4: Input n_{clicks}
5: Input $S\nu_f$
6: Input d_n
7: **while** $d_n \neq 0, t_{me} \neq 0$ **do** ▷ Undefined if d_n and t_{me} is 0
8: $n_{clicks} * (S\nu_f/d_n)$
9: (Cl/t_{me})
10: $\frac{(Cl/t_{me})}{n_{clicks} * (S\nu_f/d_n)}$
11: **return** $S\nu Em$ ▷ SvEm is in seconds

clude: (1) ability to click (2) zoom, dragging, swipe across and hold (select) and (3) toggle the mobile display from a portrait to landscape view. These features have different roles which enable the user to interact for various reasons. For example, clicking, zooming and holding (select) on a mobile visual display allow distinctive parts of our security visualization to be observed further in detailed.

4.5.4 Mobile Platform Types and Specifications

There are other concerns affecting our overall effectiveness measurement outcome and rating. These are with respect to the mobile platforms. They vary in shapes, sizes, hardware capabilities and, most importantly in terms of different display dimensions and resolutions. In particular, display resolutions do affect our effectiveness measurement rating in security visualization. However, in the context of this thesis, these factors are regarded as baseline variables which are seen as performance assumptions and range. In addition, size variation in mobile platforms has contributed to the challenge of measuring effectiveness consistently across all platforms. Therefore, other effectiveness measurement variables are used such as mobile interactive features, user-trigger features and user attention span. Moreover, mobile platform specifications refer to the hardware processing power and visual capabilities in resolutions and performance. Although the mobile specifications are seen as hardware limitations, this thesis acknowledges their existence to eliminate doubts from viewers and all their contributions to our effectiveness measurement techniques.

4.5.5 The SvEm Human Subject Component

With relevance to mobile platform types and specifications, we classify a 'SvEm human subject component' as a concern affecting our effectiveness measurement rating. In the context of this thesis, a SvEm human subject component is the security visualization audience (viewer). Different audiences have various visualization preferences, which affect our ability to consistently measure how they interact, what they would like to see and how well can they identify security insights. As a result, audience provides a challenge to consistency in measuring effectiveness for security visualization. However, with assumptions made and a careful scoping of the targeted audience used for this thesis experiment, we are able to produce significant progress with calculating effectiveness in security visualization for mobile platforms and understanding how effectiveness could be assessed and measured within various users (viewers). We observed users based on their interactions during their visual observation period.

4.6 A Practical (user-centric) SvEm User Model Explanation

The entire concept of our SvEm security visualization model is better explained with the use of an SvEM conceptual model to show how the core components link together and how they function to achieve our effectiveness measurement framework. From an overview perspective, the conceptual model illustrates how the user, user cognition and visualization function together with various requirements to achieve security visualization insights. This conceptual model adds to our 'effectiveness insertion approach stages (Figure 4.1)' to explain the effectiveness measurement process. As seen in Figure 4.9, a user's working memory, perception and the effectiveness-visualization techniques/features connect all main components. Users rely on 'working memory (known security knowledge)' to perceive visual images. From a technical aspect, users rely on effectiveness techniques and features to assist them when viewing visual images. However, user cognition relies on the perception processes to activate internal security insights within the underlying security visualization presented.

However, without going deeply into the psychological explanation of how human cognition and perception function, our SvEm conceptual model works well given the following assumptions are met:

1. Working Memory: activating the viewer's working memory is the primary step, with the aim of educating the viewer with visual previews of what is expected to be seen.

2. Emotions: human emotion affects cognition. Positively, a user has to be open-minded with a creative thinking mindset.

3. Attention strength: a user has to be focused only on the given visualization with the intention of learning and gaining security insights.

4. Mood: a user has to be in a positive mood with less mental distraction and be motivated to learn.

4.6.1 The SvEm Conceptual Model

Based on intensive research across the computing science and psychology domain, we have constructed our SvEm conceptual model to translate an overview of our full-scale effectiveness measurement approach across various vital entities: the user, human (user) cognition and security visualization. This led us to develop our conceptual model that incorporates the user's entire visualization experience with the security incident presented.

Figure 4.9 reveals the SvEm model with all the entities contributing to effectiveness measurement features: (1) *User*, (2) *Visualization* and the connecting variable of (3) *User Cognition*. These components incorporated together create the environment of linking relationships between the user and the visualization presented. In the SvEm conceptual model, the user relies on working memory through their mental effort to activate the user's cognition to perceive security information presented in the visualization. Predefined data are then processed through perception, whereby preattentive data processing is executed to form visual recognition and psychologically build (form) visual images of the information presented, i.e., the perceiving stage in users. As a result of the preattentive data processing operation, there is an establishment of the relationship between users, the users cognition and the visualization. Identifying this relationship aids uses to utilise their cognitive and preattentive processing capability to effectively interact with any given security visualization and observe comfortably in a least time possible.

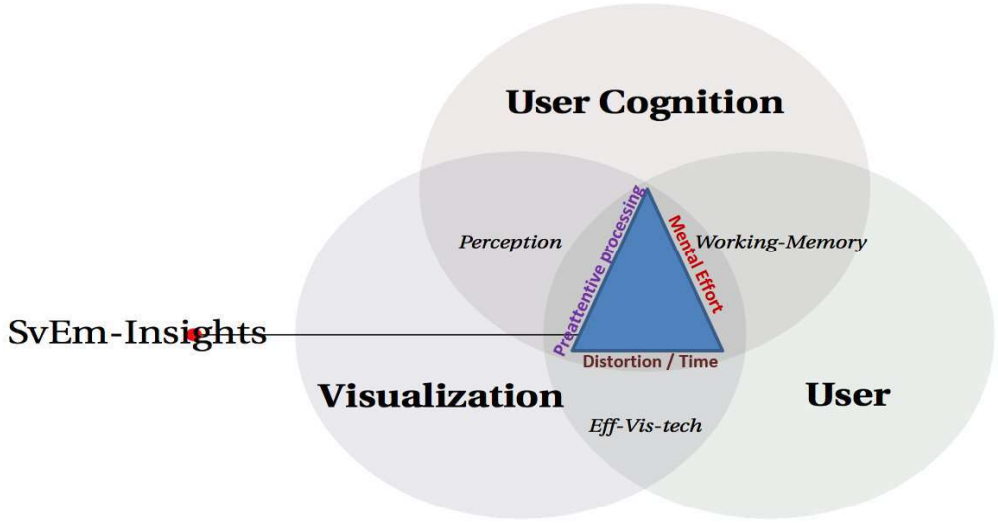


Figure 4.9: SvEm Model Illustrating all Components linked together

4.7 SvEm Usability Components and Aspects

In any theory and application, users have a high impact on theories and applications that are designed and implemented. Users do contribute to how theories and applications should function and operate for all different kinds of targeted audiences. Such functionalities and operations are referred to as user-centric or usability components. SvEm User-centric or usability features and components contribute to effectiveness in the entire security visualization presented.

In addition, users provide user-feedback, which affects and contributes to how these applications work. From the perspective of a technology user relationship link, user cognition and perception also play a large role in the impact and feedback from when they confront security visualization applications. Hence, the usability process and features in any form of technology products and applications are vital for evaluating and assessing products and applications. We categorise user impacts and identify them according to three groups: (1) user-trigger features, (2) user cognition with perception and (3) cognitive knowledge with working memory factor. These user impacts largely contribute to how effective our security visualization effectiveness measurement framework is, and how effective it delivers or translates security information and insights to the targeted audiences. These usability components and aspects categorised have different specific purposes on ensuring the security visualization presented in our SvEm framework allows effectiveness throughout the visualization experience.

4.7.1 User-trigger Features

Users have a choice and decision to make when confronted with a security visualization. User choices and decisions affect the outcome of the visualization rating, performance and judgement. This happens because users have preferences and existing knowledge on certain types of visualizations. Thus. user ratings are triggered by attractive and interactive attributes which are often presented within the given security visualization. In our SvEm security visualization framework, the attributes provide distinctive perceptual factors which this thesis refers to them as '*user-trigger*' features. User-trigger features help users to interact with the given security visualization in order to gain their interest, attention and help them explore what the security visualization has presented. Our SvEm framework outlines and implements the following user-trigger features:

1. Interactivity feature: These features are often generated mentally during the process of interacting with the security visualization. It involves the notion of 'directness' in visualization which utilises interactive features to interact with data directly, therefore enhancing the user's understanding of the data and what the data is presenting.

2. Cognitive activators:

    a) *Semi-permanent hold activator*: a visual animated feature allowing critical (suspicious) files of interest, being pushed out from the normal visual pattern and behaviour for at least 3 seconds to capture the viewer's attention.

    b) *Permanent hold* activator: a permanent coloured file indicator, marking out a malicious (suspicious) file. Red or yellow is used, depending on how critical the file is.

    c) *Critical-file detected* activator: this is an alert identifier to gain the viewer's attention.

    d) *Sound alert* activator: an additional alert identifier to gain the viewer's attention, particularly for colour-blind people.

    e) *Pre-attentive attribute/object*: a red dot in a pool of blue dots

    f) *Destructor attribute/objects*: 'blue dots'

The existence of both interactive and cognitive features activates the user's 'visual cortex' of his/her brain through the process of visualization with the use of colour indicators within

a group of entities used. This is known as 'colour perception.' In addition, applying 'Gestalt Principles of Perception' [157] as part of the user's known (prior) knowledge enables further motivation and interactivity with the security visualization. This elevates the user's concentration through the use of user-trigger features, therefore allowing the entire perceptual experience to increase the user's attention span, which positively affects our effectiveness measurement (SvEm) outcome (results). Further assessment of the use of user-trigger features will be discussed in Chapter 5 and Chapter 6.

4.7.2 User Cognition and Perception Attributes

As part of our SvEm usability approach, we take a close theoretical look at defining effectiveness in terms of user-centricity. This requires further attention to the 'SvEm conceptual model' presented in Figure 4.9 whereby the 'preattentive processing' component is the key feature into understanding how users activate their cognitive capabilities and perceive security visualization. Therefore, in this thesis applying a user-centric approach refers to four aspects: (1) user preference in relation to security visualization, (2) security events' representation convenience, (3) the presented security visual nodes and (4) enabling the user to validate (proof) their findings with the means of security visualization. For example, the use of user perception and cognitive activators in colour perception techniques is to provide visual security knowledge. This has an impact on our working memory load. Our memories store patterns we understand and, on the ability, to form and use patterns during the process of interacting with our SvEm framework. However, there is always that need to understand cognitive biases and avoid them during the visualization process. This could be avoided with the theory of 'simplicity' whereby 'visual features' such as colours and sizes activate the users' power for visual system in an incredibly fast way before they can even begin to think about it. This means users can overcome their cognitive biases. Another factor used to avoid cognitive biases in security visualization to boost effectiveness is to understand and execute the preattentive processing procedure adequately, as discussed in Subsection 4.7.3.

4.7.3 The SvEm Preattentive System

An important human factor that has an effect on users effectively perceiving a security visualization is their ability to utilise the their preattentive system during the visualization observation experience. The SvEm preattentive process workflow diagram explains how users perceive leading up to interacting, understanding and acquiring useful security knowledge. As seen in Figure 4.10, the preattentive processing procedure requires the users (viewers) to mentally interact and validate their visual output as they perceive the data given through

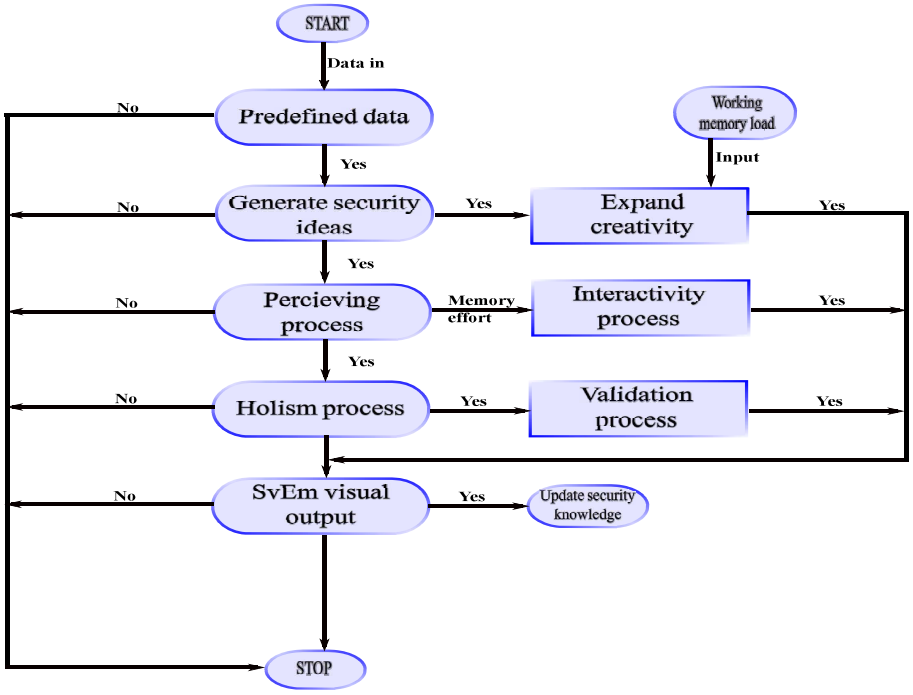


Figure 4.10: Workflow Diagram: The SvEm Preattentive Process Visualization System

security visualization. Therefore, with a careful execution of the users' preattentive system and acknowledging that users have prior knowledge (working memory load), effectiveness is measured within a reduced time frame.

Furthermore, from the SvEm conceptual model approach, both user cognition and visualization components are linked through the perception process by executing the preattentive processing system. However, the security visualization presented has to have all of the above features mentioned such as user-triggers and cognitive load activators.

4.8 Summary

In summary, this chapter provides our 'Security Visualization Effectiveness Measurement (SvEm)' theory with details of the SvEm algorithm. We provided the various features and aspects associated with our theory and explain in detail the SvEm variables with their functions and how they are calculated, assessed and achieved. We state that our theory and framework relies on two factors: (1) '*distortion rate*' and (2)'*time*.' To achieve this theory, we provided our SvEm-distortion and SvEm-time algorithms. These algorithms are explained with the use of pseudocodes. Hence, we presented a full-scale effectiveness measurement model (Figure.4.1) with our SvEm user conceptual model (Figure. 4.9) to highlight all con-

tributing components that add effectiveness measurement in our framework.

The concept around our Security Visualization Effectiveness Measurement (SvEm) theory is explained with the use of our SvEm conceptual model (Figure 4.9) and the SvEm preattentive process work flow diagram (Figure 4.10). This conceptual model provided an overview of what makes up the SvEm algorithm and how they are linked together to justify each contributing components of the algorithm.

Chapter 5

SvEm Framework: The Security Visualization Applications

This chapter delivers our practical implementation of our Security Visualization Effectiveness Measurement (SvEm) theory and framework. With the datasets obtained and described in Chapter 3 and the SvEm algorithm designed and introduced in Chapter 4, we a able to design and implement our practical SvEm framework to test and evaluate our effectiveness approach proposed in this thesis. We described all the user-interactive functionalities and designs that contributed to our security visualization effectiveness measurement and evaluation process for cyber security operations. All pre-requisites stated and explained in Chapter 3 and Chapter 4 provided the required knowledge, requirements and design methodologies used to construct our security visualization framework. We model all theoretical features discussed in Chapter 4 using both web and mobile security visualization applications. Moreover, our implementation process has encountered challenges which are discussed in Chapter 7. We begin by describing our web and mobile platform representations.

5.1 Security Visualization: Web and Mobile Representations

An important fact and approach to constructing a security visualization application with an intention to provide a full-scale effectiveness measurement approach required us to understand how security data/events are represented over the web and mobile platforms. We needed to ask these set of questions: *Are we representing real-time data?; Are we representing static data?; Who are the targeted audience?*; and *Is this visual representation suitable for large data sets/volumes?* These questions drive the need to effectively represent data in security visualization platforms. Hence, security visualization is an effective method, considering the urgency to communicate important information, for example, reporting a cyber-attack.

In this thesis, visual data representation is driven by the framework design and by the au-

dience preference and demand of the security visualization. Mobile platform users are our targeted audiences. However, existing mobile hardware limitations, required further investigations beyond the implementation provided by the visualization framework. A full-scale effectiveness approach requires both the mobile visualization framework and the users' interaction environment. Thus, we designed a centralised web-based security visualization framework to handle most resource operations. These includes performance, storage, scaling and rendering processes. Our centralised framework uses a web or cloud base approach to provide the mobile-based platform that queries all operations. Using this approach, we are able to bypass or manage existing mobile limitations and represent predefined data through the use of visualization presentation techniques. Hence, our effectiveness measurement technique aims to provide effectiveness measurement through the entire security visualization implementation process and the user's visual experience through effective data representation methods. However, representing data is a generic method in visualization. This thesis concentrates on the importance of security data and how the data are represented in a way, critical underlying raw data are preserved with less tamper evidence.

5.1.1 The Importance of Security Data Representation

The need to observe security data, information and security events affects how visualizations are developed and presented over web and mobile platforms. It is a current challenge whereby the volume of data visualized grows rapidly, and visualization researchers and developers find it difficult to represent data in a timely manner with the intention of translating useful knowledge to users. Therefore, our SvEm framework considers the data representation process, a vital component. We investigated existing data representation methods and designs, from which we concluded that representing data is a case by case task. Users are able to naturally process and comprehend why data is represented in a particular way whereby appropriate data representation design and visual implementation types are represented to suit particular data at the given state (i.e., data can be suspicious, or malicious). Thus, effectiveness in security visualization representation can be assessed through the rate at which users can comfortably process the data represented.

Equally important, representing data for provenance, attribution and intelligence purposes also affects the design and methods. These representation design methodologies are discussed in this chapter as part of our web and mobile platform features.

5.1.2 The Requirements for Data Representation in Web Platforms

A common aim for most web-based platforms is not only to gain global presence but to enable mobile services for users. Such purposes, intentions and goals contributed to the designs and implementation of our security visualization web-based infrastructure. Our SvEm framework utilises certain web technologies to build a central security visualization framework that would manage resources, services and functionalities. It is developed with the intention of providing security visualization for mobile platforms, whereby users can gain access to our security visualization framework from anywhere and use it to their advantage.

Therefore, the need to provide a mobile security visualization framework with user preference in mind required our SvEm framework to provide a multi-layered security visualization approach. This multi-layered security visualization framework provides several web and mobile visualization types for users to interact with based on their interests and needs. However, for a user to interact efficiently with our security visualization framework, our data representation and presentation approach has to be visually attractive to capture their interest. Therefore, providing several visual representation and presentation techniques was our key approach to enable users to have an option of which visualization to interact with in a given time with the aim of acquiring useful security insights.

The representation demands above have triggered the need to outline our representation and presentation techniques. Hence, our dataset requirements discussed in Chapter 3 and Chapter 3.6 have contributed to our design selections and outcomes. These data representation techniques include:

1. Volume data representation design.
    - Real-time security visual nodes representation.
    - Static security visual nodes representation.
    - Customised base data representation.
2. Abstract data representation design.
    - Real-time security visual nodes representation.
    - Static security visual nodes representation.
    - Customised base data representation.

These representation techniques require data analysis processes to understand the links and relationships between malicious data packets involved in an attack. For example, representing malicious relationship between data packets involves understanding what type

of entities to show, and why these entities are shown. Thus, applying data analytics at the backend of our framework and implementing an n-dimensional web-based security visualization platform provided the means to accommodate all possible data representation techniques and designs. Our web platform is built with the following tools and components: (1) the use of NodeJS[1], (2) Three.js[2], (3) WebGL[3] and (4) JavaScripting[4]. In addition, adopting the use of canvas frames helped create a visual space which enabled us to gather for data representation and scalability needs. We deployed WebGL representation visual models for our real-time security visualization, static visualization and other visualization types.

However, the continuous challenge in a web or mobile based security visualization platform is the struggle to manage how data are visually represented against the available display space allocated for the visualization. This created the need for continuous representation changes and often requires the need to invest time in creating the correct attractive visual representation (nodes and attribute) designs. These designs easily capture the user's attention when interacting with our security visualization framework. Our SvEm framework has been designed with the aim of providing several visualization types from multiple designs. For example, we utilised an n-dimensions (d_n) visual presentation approach to create 3D visual nodes with force-directed effects that would automatically scale when more data are pushed into the security visualization space. In addition, our use of shapes and colours in security visualizations has produced interactive representations, particularly when using common day-to-day events, that resemble popular and easy to understand visual images. An example is, the use of spheres to represent different forms or stages of data nodes/attributes such as a shaded red, yellow, green or blue to distinguish associated relationships amongst visual nodes. The use of shapes and colour choices are covered in detail in Chapter 6, which involves laying out a new security visualization standard and guidelines to aid security visualization developers and the users of our SvEm security visualization framework. However, a brief preview of what is covered in Chapter 6 is shown in Figure 5.1. It shows the common shapes and colours used for our security visualization representations.

In addition, representing files (data) as visual nodes in security visualization also involves providing a 3D model design that includes file labels and interactive features which aimed

[1]Node.js is an open-source, cross-platform JavaScript run-time environment for executing JavaScript code server-side. (Link: https://nodejs.org/en/)

[2]Three.js is a cross-browser JavaScript library/API used to create and display animated 3D computer graphics in a web browser. (Link: https://threejs.org/)

[3]WebGL is a JavaScript API for rendering interactive 2D and 3D graphics within any compatible web browser without the use of plug-ins. (Link: https://www.khronos.org/webgl/)

[4]JavaScript, often abbreviated as JS, is a high-level, dynamic, weakly typed, prototype-based, multi-paradigm, and interpreted programming language (Link: https://www.javascript.com/)

(a) Malicious and Suspicious Nodes **(b)** Normal and Intelligence Nodes

Figure 5.1: Data (Visual) Nodes Representation Design Samples

at capturing the user's attention. Some examples of file labels include: (1) file name, (2) file type, (3) file path and other details which are seen as important and related, and which would make the visualization meaningful. Interactive representation features include: (1) design appearance, (2) static/stationary file node, (3) animated file nodes, and more.

5.1.3 The Requirements Data Representation in Mobile Platforms

Our SvEm mobile platform capitalises on our web-based platform to present security visualization to the audience. With both web and mobile advantages, users are able to access our SvEm framework anywhere and interact with it. This allows end-users particularly mobile platform users to utilise their devices and maximise their visual experience whenever a security event (malicious event alert) has popped up on their phones.

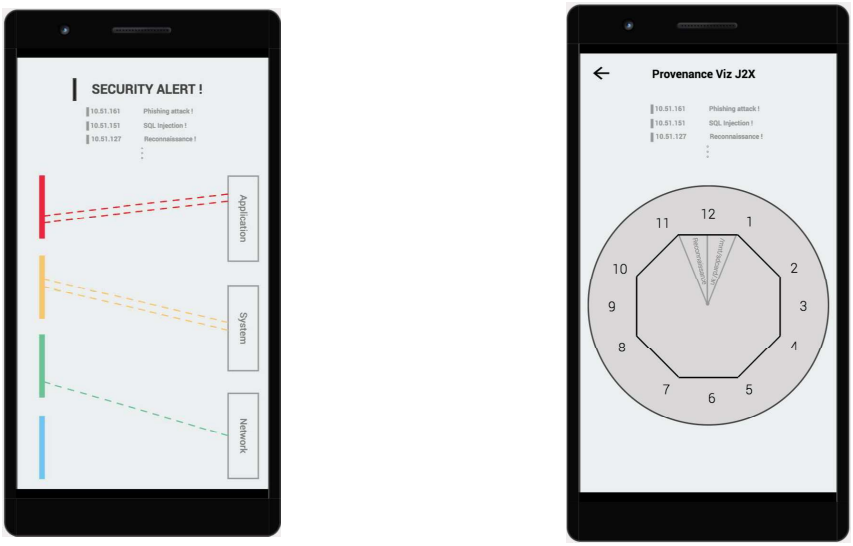
Our mobile platform is deployed as an additional web feature with mobile-friendly compatibility. We utilise the web capabilities to handle performance, data management, storage, load balancing and optimisation features to provide a light weighted mobile platform. Thus, the outcome of our SvEm visualization in mobile platforms was, it could handle existing hardware challenges such as rendering and data scaling. In order to achieve a light weighted security visualization framework for mobile platforms, we had to change how data is represented in mobile platforms. Volume data are represented in abstract forms and designs whereby we implemented a drill down approach to view more information related to a security event if needed. We applied a geometry and parallel coordinates visual approach to design and implement basically two data representation techniques for our mobile visualizations. These security visualization designs shown in Figure 5.2 and Figure 5.3 have represented data in the following two methods:

1. Abstract data representation design.
    - For real-time security visual node representation.
    - For static security visual node representation.
    - For customised base data representation.

2. Circular top-down (drill-down) data representation design.

    - For real-time security visual node representation.
    - For static security visual node representation.
    - For statistical nodes representation.

Based on abstract data representation designs and circular top-down (drill-down) data representation design, large volumes of data analysed at the backend are able to appear comfortably as an overview visual presentation. This allows users from all forms of preferences and levels of understanding to interact well with less complex visualizations. In addition, Figure 5.3 shows a statistical visual form of representing provenance related data in mobile platforms. The ability is to show provenance visualization with time as the provenance tracker enables mobile users to see simple security related data that could tell a history or journey and movement of malicious data.



(a) Parallel Coordinates Representation Designs **(b)** Provenance Representation Designs

Figure 5.2: Data Type Representation Designs

The ability to raise the level of detail from the more technical content and complex detail structures into simple abstract forms makes it easy to understand security events better. We developed parallel coordinate visualization types to show cyber-attack locations. For example, Figure 5.2 shows the use of parallel coordinates to show where the malicious data could be, i.e., within the application, software and/or the network layer. This is rated against a rating of malicious (red), suspicious (yellow), and/or normal (green) colour alert indicator. Likewise, intelligence monitoring is represented in blue in the security visualizations.

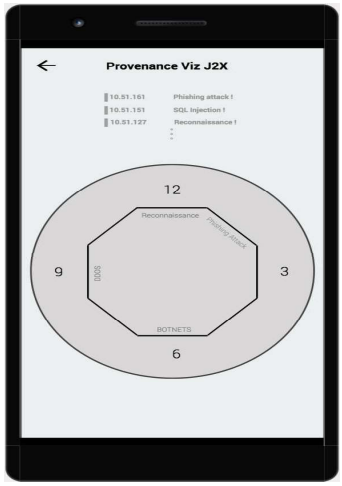


Figure 5.3: Circular Data Type Representation

Finally, with a clear understanding of how data or security visual nodes are designed and represented, the remainder of this chapter provides a detailed discussion into how our SvEm framework is designed and implemented. This involves providing deep insights into how we have designed and implemented it with the goal of providing: (1) effective security visualizations, (2) a security visualization effectiveness measurement technique and (3) the mechanisms involved in our SvEm framework.

5.2 The SvEm Security Visualization Application

Our security visualization model is a generic application that serves both web and mobile applications, i.e., visualizing various security events from both web, cloud and mobile appli-

cations. We reiterate that this security visualization application serves primarily for mobile platforms with effectiveness measurement techniques implemented and assessed throughout the entire security visualization visual experience for users. The core emphasis is on the ability to represent attack data clearly on mobile platforms. Often cyber attack or malicious activity datasets are complex and difficult to understand using visualization. Therefore, we are addressing 'representation effectiveness' in security visualization for mobile platforms.

Our SvEm mobile-based framework aims to address scalability, rendering and hardware processing power limitations. It is designed with mobile user features for mobility, performance, efficiency and effectiveness capabilities. We implemented a cloud storage infrastructure using Amazon Web Services' (AWS) application features to facilitate our backend infrastructure, while our security visualization frontend uses NodeJS, Three.js and WebGL. MongoDB, a NoSQL database framework is the choice for our data storage infrastructure. This allowed us to process and handle JavaScript Object Notation (.json)[5] files efficiently for the visualization frontend. Redis [6] is implemented as part of our network queue protocol which also acts as a data cache. It facilitates data movement between Progger, MongoDB and our security visualization frontend. However, to understand the entire data processing stages (downloaded, stored, processed) before parsing it to the frontend, we must address the backend design and architecture.

5.2.1 The Server-Side: Backend Design Architecture

The SvEm security visualization framework server-side (backend) infrastructure is designed to accommodate both static and real-time visualization scenarios. It handles all data analytic processes occurring within the database, collectors and parser environments. Our system architecture includes the following components: Windows Progger (Logging Tool), Redis, MongoDB, Nodejs and WebGL. Windows Progger (a windows version of Linux Progger [158]) is an internal system/kernel level provenance logging tool currently being developed with emphasis on providing security within computer and cloud systems. Redis [7] [159] facilitates our cache/database link between Windows Progger and MongoDB. All data are stored permanently in MongoDB [160], while Nodejs [161] and WebGL [162], [163] facilitate the client-side frontend security visualization framework.

[5] JSON (JavaScript Object Notation) is an easy to read and write text format, lightweight data-interchange format that allows machines to parse and generate effectively. JSON is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. (Link: https://www.json.org/)

[6] It is an open source (BSD licensed), in-memory data structure store, used as a database, has the ability to cache data requests and a message broker (Link: https://redis.io/)

[7] Redis is an open source (BSD Licensed) in-memory data structure store, used as a database, cache and message broker

We designed a hybrid server-side (backend) infrastructure to address hardware limitations and challenges. This enables efficient mobile use of the SvEM framework, i.e., can be accessed using both web and mobile platforms and anywhere possible. Most data storage and processing tasks occur in the cloud using Amazon Web Services (AWS) virtual machines. MongoDB and scripting tasks such as the 'collector' and 'parser,' keep track of new data required for processing. Progger, a provenance logger, is installed in the preferred logged machines. In between Progger and MongoDB, we installed Redis, a network queue protocol and has the ability to cache data requests. This enables efficient data processing and passing from the logger to either MongoDB or the client-side security visualization framework.

The server-side architecture is designed to handle data processes while managing data storage. Preprocessed data are engineered based on the visualization scenario. For example, a real-time logging of a computer's kernel system for provenance purposes is visualized to show and keep track of file creation, modification and deletion. Such architecture requires hardware processing power and large storage space. In addition, data are standardised to meet the effectiveness assessment of the security visualizations in web and mobile platforms. Both web and mobile requirements enabled the need to provide efficient data querying, processing, parsing, rendering and scaling tasks for a security visualization. Figure 5.4 shows the basic tools and libraries required to host the SvEm security visualization server-side backend.

The choice of this architecture resulted from leveraging on both open-source and commercial tools which were seen as viable for this research. We further discuss performance testing, results, advantages, limitations and challenges in Chapter 7 of this thesis.

5.2.2 The Client-Side: Web Frontend Architecture

Our web-based client-side architecture is built using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript, Bootstrap and webgl. The client-side frontend design involves three key effectiveness areas: (1) multiple and optional visualization views, (2) visualization representation scalability and (3) the ability to represent and show different granularities of security details with user-centric features.

The ideal visualization features presented in the SvEm security visualization framework includes user-centric, time-based, real-time based, dynamic, provenance and attribution-based visualization landscapes. These features are indicated to aid users (viewers) through their decision-making process. In order to achieve these visualization features, we created various visualization representation designs to illustrate security events.

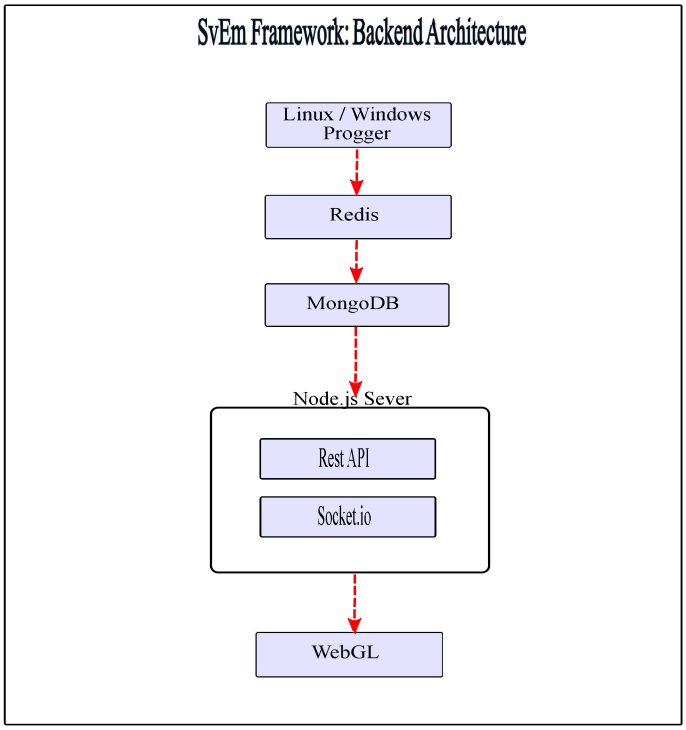


Figure 5.4: SvEm Backend Design and Architecture

5.3 The SvEm Security Visualization Designs

In conjunction with Chapter 4 – Section 4.4 and Subsection 4.5.1 our SvEm mockup designs include creating effective visualizations to resemble different security events (cyber-attacks) and their malicious landscapes. Thus, the visualization mockups address several views, including: abstract and overviews, circular (meet-the-eye) views, tracking and monitoring, granularity and layering, n-dimensional and parallel coordinate views and real-time interactive views. These mockup designs factor in the rate and volume of data collected, processed and used for visualization. For example, a ransomware attack uses the granularity and layering visualization view to show the malicious payload traversing through a computer system searching for files to encrypt. This shows the links between files, processes and libraries affected. Given that example, various visualization mockups are discussed and explained in detail in the next subsections of this thesis.

5.3.1 SvEm Design–1: Abstract and Overview Visualization View

The *abstract and overview* visualization mockup view is designed to show security events in an abstract level of detail. This design provides a visual summary of the malicious event

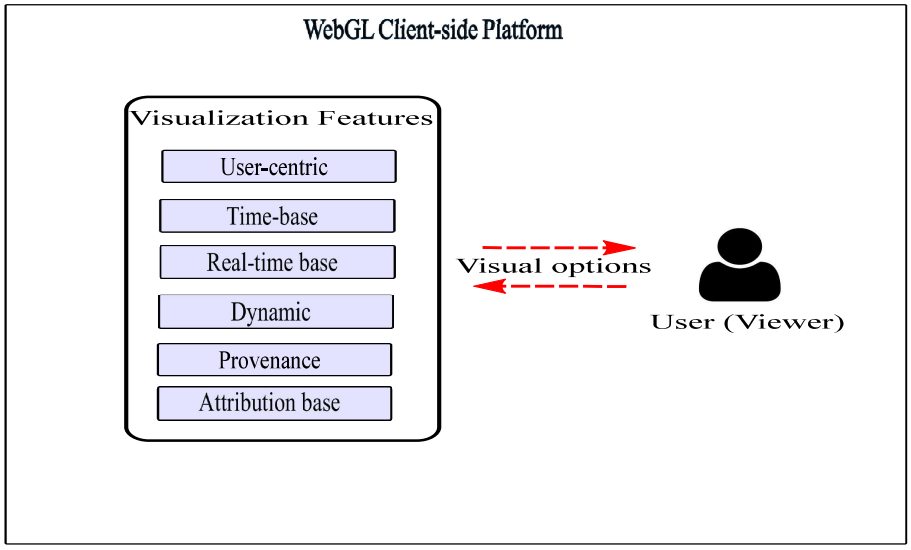


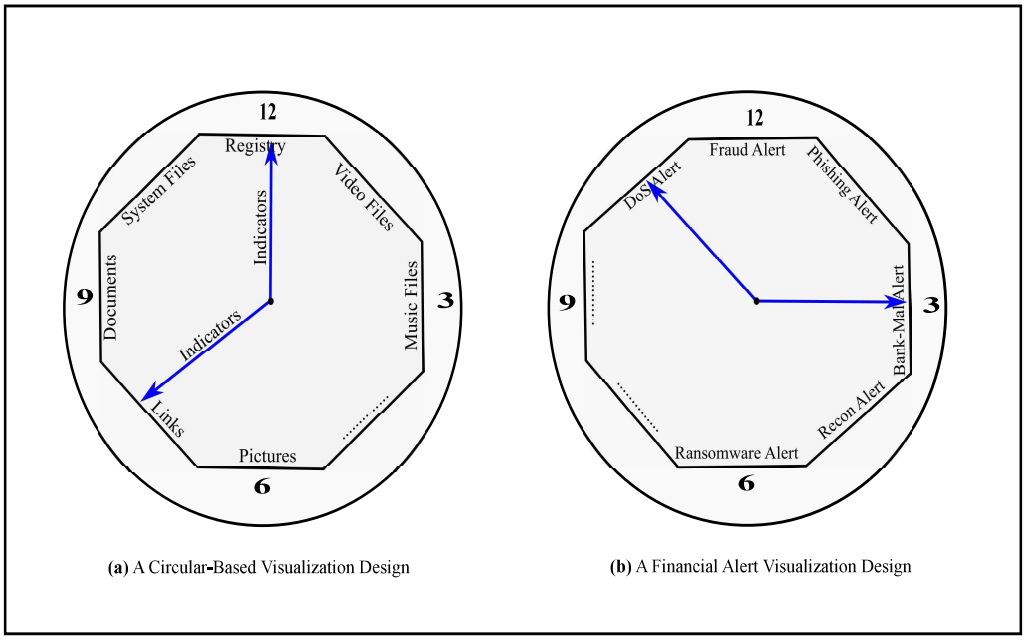
Figure 5.5: Frontend Data Type Representation

by giving generic and overview details of the malicious landscape. This type of visualization view is designed with simplicity and to be an introductory visualization view. Such designs target upper level management people and end-users who have minimal technical knowledge in security, by capturing the user's attention. The key goal of this design view is to provide only vital information relating to the malicious event with less visual noise (unwanted information). These designs are not restricted to a single visualization design, i.e., it can be a treemap visualization or an interactive design presenting a security alert. For example, Figure 5.6 shows an overview visualization design whereby a critical alert signal is implemented as a splash screen visual view.

Figure 5.6: An Abstract Visualization Alert Sample

5.3.2 SvEm Design–2: Circular (Meet-the-eye) Visualization Design View

Security datasets are complex when analysing them and usually contain multiple levels of important details. It requires higher level of user concentration with proper visual pattern perception and mapping. Such requirement seeks a simple, clear and easy-to-use visualization view. A *circular (meet-the-eye)* visualization view whereby all vital security information is presented in one page/screen is the solution. This keeps the user's eyes and concentration focused on a central visual view, therefore increasing the user's attention span while decreasing cognitive biases. A sample mockup design is shown in Figure 5.7, containing two samples illustrating data represented in a time-based visualization. Figure 5.7b simply shows a financial statistical based malicious attack lifetime (e.g., duration of attack) over time.



(a) A Circular-Based Visualization Design

(b) A Financial Alert Visualization Design

Figure 5.7: A Circular Statistical-time Based Visualization Design

5.3.3 SvEm Design–3: Intelligence Visualization Design View

The concept of the *SvEm intelligence* visualization mockup view facilitates the opportunity to use security visualization for attribution, provenance and intelligence purposes. We provide the ability to track and trace back cyber-attacks to sources and observe or monitor real-time attacks. The mockups offered for our intelligence design view offer the ability to track and follow specific visual nodes for more details relating to such a malicious event. For

example, Figure 5.8 displays an intelligence visualization [164] mockup displaying circular visual nodes with connecting lines/arrows. These are mockup features commonly representing tracking indicators in visualizations. Attribution and provenance-based visualizations are other examples of such visualization designs. Overall, our intelligence visualization mockup design seeks to represent and present insights through pattern and behavioural visual analytics.

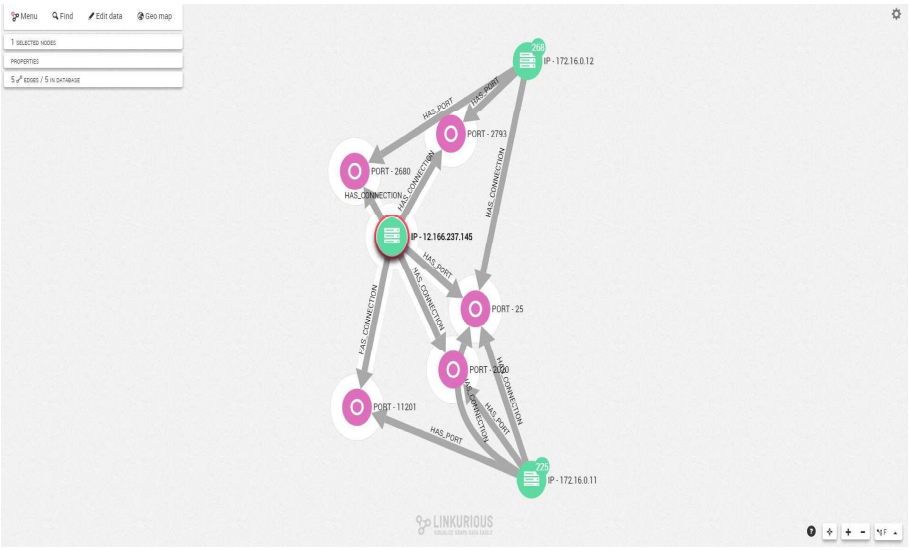


Figure 5.8: An Intelligence Visualization Design

5.3.4 SvEm Design–4: Granularity and Layering Visualization Design View

The *granularity and layering* visualization mockup design addresses the need to represent large security data volumes in a small visual space. In addition, the layering design provides the opportunity to visually observe malicious events such as data movements and payloads escalating into other layers in the systems. Ransomware visualization is an example used for our granularity and layering visualization design. The ransomware behaviour of traversing through a computer system-files, directories, processes and libraries requires a layering visualization view. It gives users the opportunity to observe, track and analyse the ransomware actions. Figure 5.9 illustrates our mockup view.

5.3.5 SvEm Design–5: N-Dimension Visualization Design View

The *n-dimension* visualization mockup design view shares similar visualization intentions and purposes with the granularity mockup designs. It facilitates data visualization views into

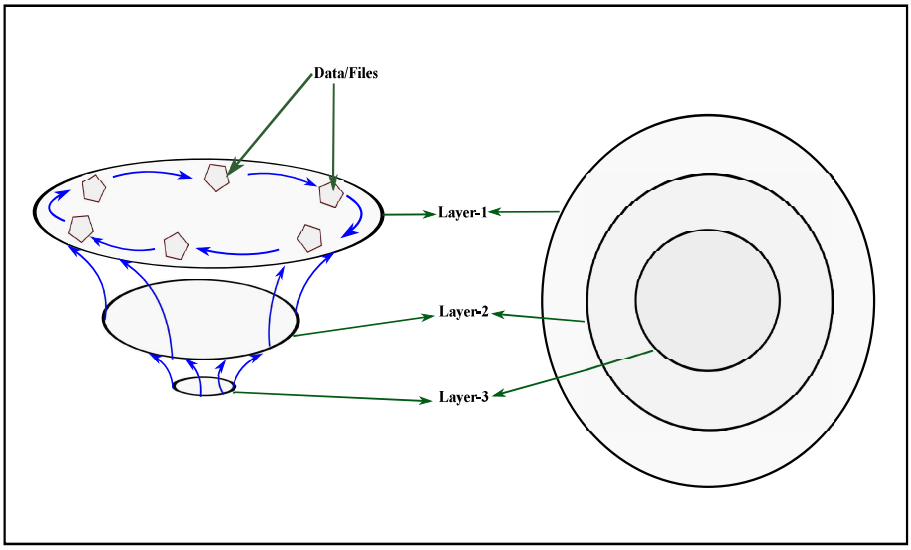


Figure 5.9: A Granularity and Layering Visualization Design

a small visual space with the intention of keeping the user motivated and concentrated on important security information. Such mockups are not only restricted to the level of information provided for visualization but with the primary focus of providing multiple insights through visualization. Our n-dimension visualization mockup utilises the concept of parallel coordinates and geometry, as shown in Figure 5.10, to represent security information in visualizations.

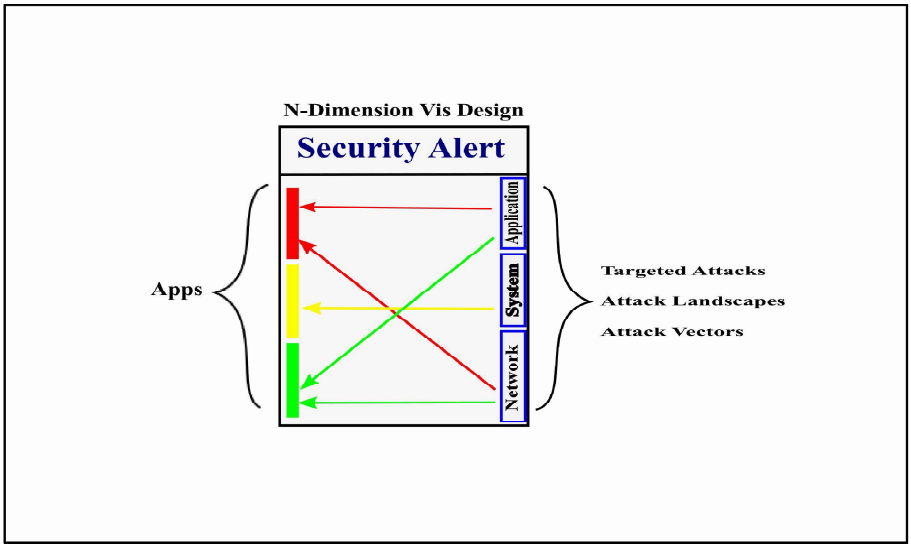


Figure 5.10: An N-Dimension Visualization Design

5.3.6 SvEm Design–6: Interactive Visualization View

Users play an important role to the outcome of every visualizations presented. Whether, it is a static-based or a real-time-based visualization, the end goal for every visualization is to interact with the users and communicate useful information. However, communicating useful information across to users requires an understanding of the data used and the potential interactive presentation techniques required for visualizations. The common interactive approach comes about with real-time visualization. It requires more planning and design preparation, which includes developing *interactive* visualization mockups. These mockups leverage on user interactive features and components to showcase attractive designs and implementations to show cyber-attacks events. A major interactive design part of our SvEm framework uses the augmented reality (AR) visual experience with a 3-dimensional visualization view. Figure 5.11 illustrate the mockup design view whereby users have the opportunity use their mobile platforms to interact with real-time security events.

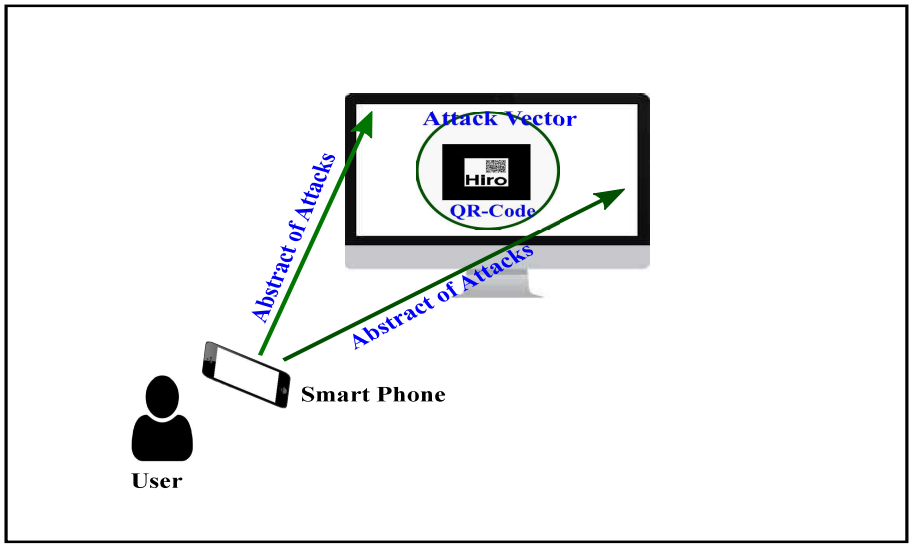


Figure 5.11: An Interactive Security Visualization Mockup with Augmented Reality

In addition, another mockup utilising the augmented reality experience to show in real-time, simulated cyber-attacks executed during a cyber security challenge event. The mockup in Figure 5.12 facilitates an interactive visualization experience, thus empowering users (viewers) to use their mobile platforms and observe a red versus blue team cyber security challenge. In a red-blue team cyber security challenge scenario, such real-time visualization shows attack packets (network data) executed from the red team on the blue teams. For example, visualizing a Dynamic Denial of Service (DDoS) attack by a red team machine onto a blue team machine. Hence, the ability to provide interactive visualizations in this scenario

naturally motivates the users to continue observe the challenge.

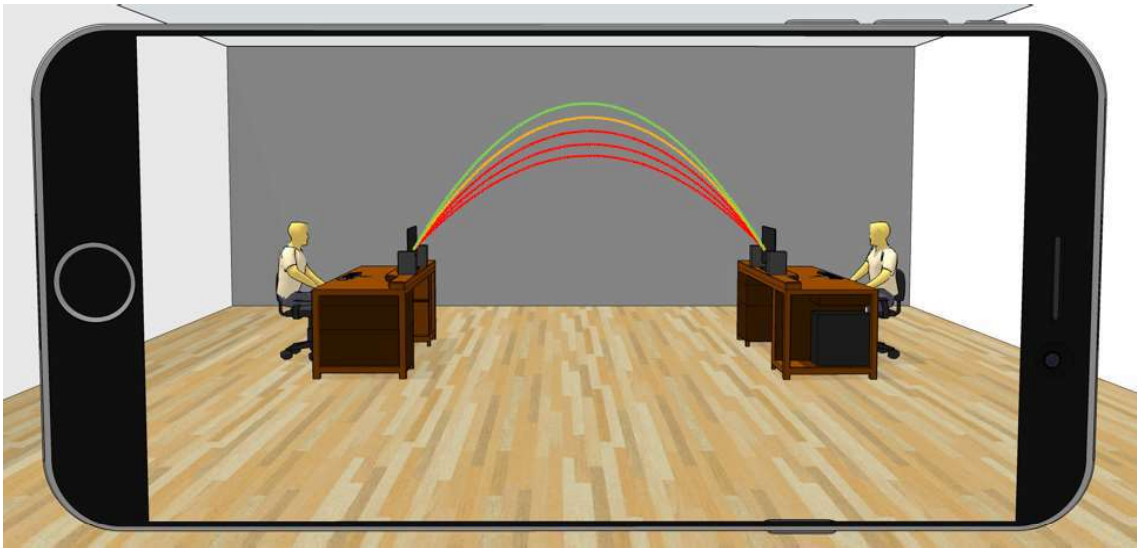


Figure 5.12: A Security Visualization Desired Output with Augmented Reality

Finally, the use of AR for security visualizations enables different audiences to participate and be aware of the simulated cyber-attacks and learn at their own pace with the use of mobile platforms.

5.4 The Mobile Security Visualization Platform and Features

Effectiveness measurements in security visualizations are assessed in many aspects during presentation. For mobile platforms, effectiveness the in security visualization requires specific features that gain the user's attention in an event of a malicious threat. Our SvEm web-based security visualization framework is compatible with mobile platforms. It provides several effectiveness techniques, namely user-trigger features and interactive visual presentation types. For example, Figure 5.13 shows our '*permanent hold*' interactive feature that aimed at capturing the user's attention while showing certain critical file of interest. These critical files can be malicious or, suspicious, and are recognised with the use of a '*semi-permanent hold* visual feature. The semi-permanent hold visual feature communicates the importance of viewing abnormal file behaviour in comparison with a normal file behaviour to help distinguish the malicious file.

Furthermore, our mobile security visualization provides an augmented reality visualization experience for the users to interact with and understand cyber-attacks in a deeper visual environment. The augmented reality approach empowers users to see and observe security events at their own pace, therefore capturing their attention and increasing their attention

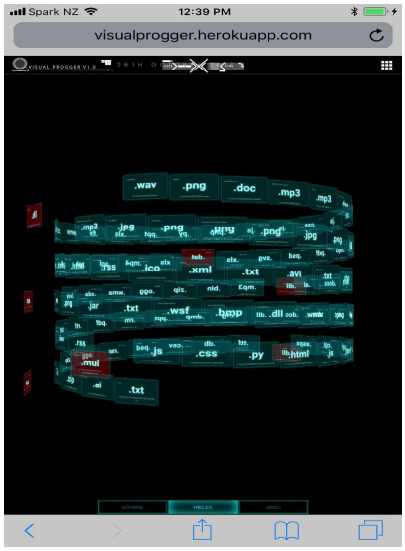


Figure 5.13: 'Permanent hold' user-trigger Feature

span. This visualization experience allows users to build their working memory load as they visualize cyber-attacks.

5.5 The Security Visualization Landscapes

Security landscapes are important for visualizations to be attractive and meaningful. They provide the means of understanding cyber-attacks in a visual environment. This thesis categorised visualization landscapes into two common location for cyber-attacks occurrences: (1) the system visualization landscape and (2) network visualization landscape. An overview of the landscape is shown in Figure 5.14, outlining all concerned components that could be affected in the event of a malicious attack. A system visualization landscape includes: directories, files, libraries and processes. However, a network visualization landscape has more components: IP addresses, routers and switches, file servers, web servers, email servers, desktops, virtual machines, hostnames, URLs and DNS names. In addition, the network infrastructure (i.e., public and/or private networks) is an important landscape factor for security visualizations. Therefore, security visualization researchers and developers need to understand the visualization landscape when designing their visualizations. This reduces data representation and presentation errors, therefore reducing cognitive biases in the user's visual observation process.

The overview of the landscape presented raised the need to go in depth into specific visualization landscapes. These landscapes are of various security events such as malicious threats and intelligence tracking. Thus, we derived several security landscapes in order for

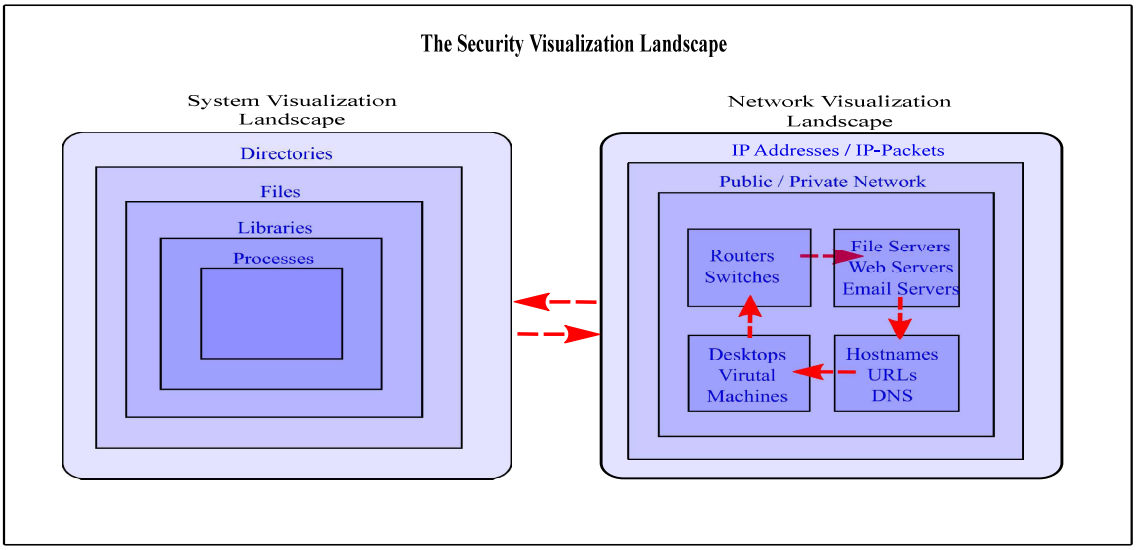


Figure 5.14: The Visualization Landscape Overview

our visualization designs, discussed in Section 5.3, to be implemented. These security visualization landscapes cover both systems and network components, with various attack scenarios. The security landscapes are as follows:

1. The New Zealand National Cyber Security Challenge Landscape
    - A system attack and tracking landscape.
    - A network attack landscape.
2. A ransomware attack landscape.
3. A bitcoin transaction intelligence landscape.
4. A malware landscape.

In order to show these landscapes in detail, we sample the New Zealand cyber security challenge landscape which involves both the system and network. We illustrated the landscape in Figure 5.15 with the NZCSC-2016 round two 'red and blue' team challenge and build on our interactive visualization design (Figure 5.3.6) to provide a security visualization of cyber-attacks. This design provided the visual environment for users to observe attacks executed on blue teams by the red teams.

Finally, accomplishing both the security visualization mockup designs and landscapes, we implemented them using several visualization application use cases to showcase our effectiveness theories. These are discussed in Section 5.6 and further analysis is discussed in Chapter 7.

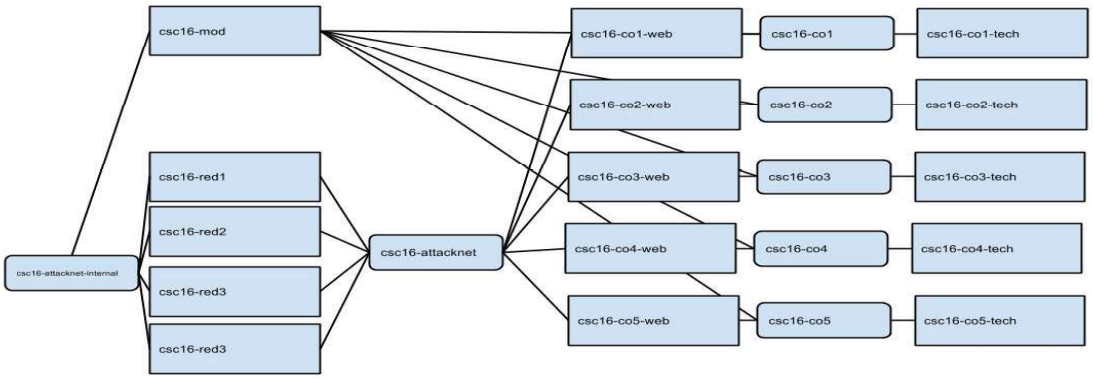


Figure 5.15: NZCSC 2016 Network Visualization Landscape

5.6 Use Case 1: VisualProgger – A user-centric Security Visualization Application

VisualProgger, a user-centric security visualization application, is implemented to showcase our visualization algorithm along with all user-trigger features discussed in this thesis. The core application features include: real-time security visualization, static visual analytics and an interactive augmented reality visualization. In the current state, this real-time application serves two purposes: (1) tracking and monitoring files of interest and (2) visualizing cyber challenges. The goal of VisualProgger is to empower users (viewers) to interact and concentrate on the given visualization provided. Therefore, in conjunction with all the effectiveness theories and approaches discussed in the early chapters, we present various security visualizations with the emphasis on showcasing effectiveness throughout the visualization framework.

5.6.1 VisualProgger Security Visualization Samples

Building on Chapter 4–Subsection 4.5.1, our VisualProgger security visualization application delivers three main visualization prototypes: (1) the sphere visualization prototype, (2) the helix (spiral) visualization prototype and (3) the grid visualization prototype. These prototypes serve different purposes and target a larger audience given we provide three visualization alternatives.

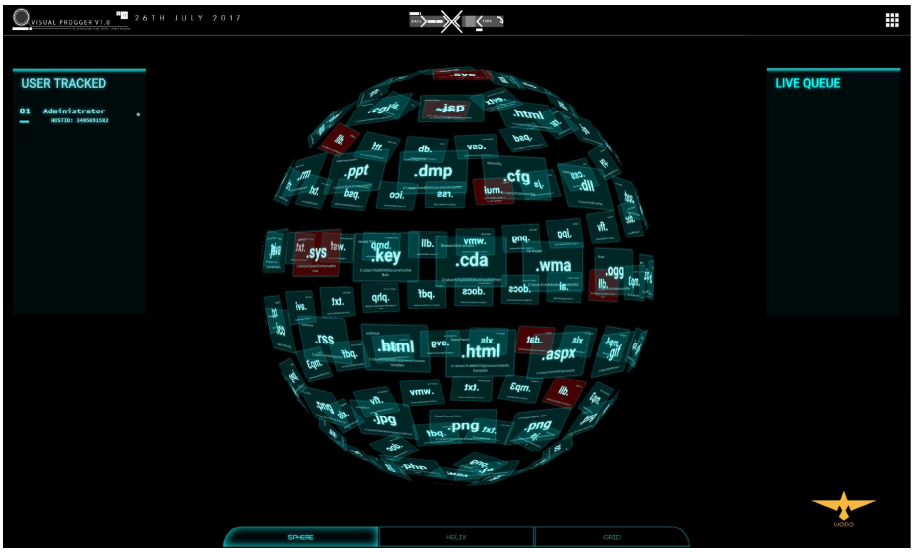


Figure 5.16: The VisualProgger Sphere Visualization View

1. The Sphere Visualization Prototype:

The sphere real-time security visualization prototype facilitates a user-centric visual environment and experience whereby users (viewers) have the opportunity to observe real-time events with interactive malicious alert features activating a 'semi-permanent' hold file position, a coloured (red or yellow) indicator and a sound notification. These alert features capture the user's focus and attention, therefore motivating them to interact with the visualization to further understand the malicious event. As seen in Figure 5.16, a sphere visualization sample shows file systems in a Windows operating system. The red coloured files are known as 'critical files' of interest.

2. The Helix (Spiral) Visualization Prototype:

With a similar objective to the sphere visualization prototype, our helix prototype seen in Figure 5.17 gives the users a perception of provenance with structured continuity of first-in, first-out visualization view. All data (system files, libraries, processes and directories) logged using Progger are preprocessed into visual nodes and are pushed into the top of the visualization stack for observation. These predefined visual nodes then move down the helix visualization queue for observation. A live queue of all files visualized is shown, giving users the option to observe the visual form of the data.

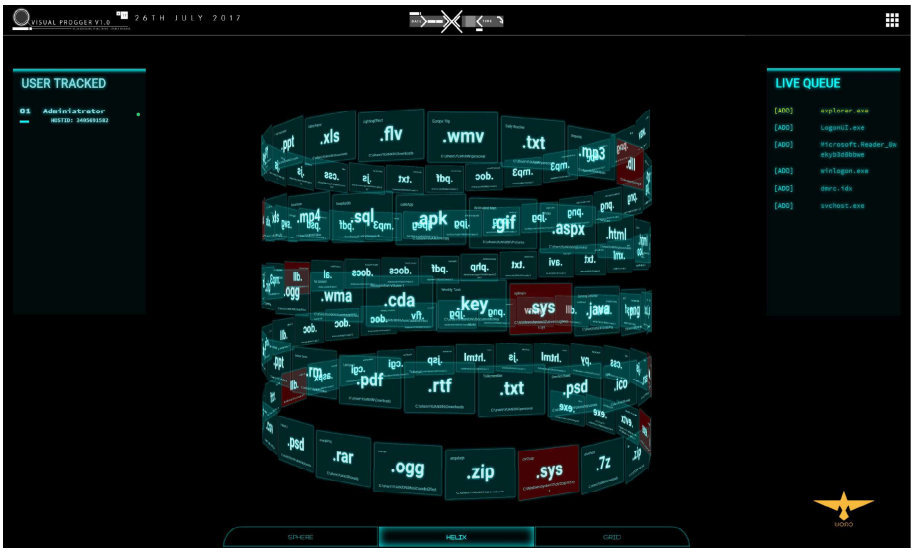


Figure 5.17: The VisualProgger Helix Visualization View

3. The Grid Presentation Visualization:

Our grid visualization prototype facilitates the concept of granularity and the layering of visualization views allows users to observe security events in various visual options, i.e., visualize in layers of time-series data occurrences. This visualization prototype seen in Figure 5.18 shows an example of the grid layout with new files prompting in front of the visualization view.

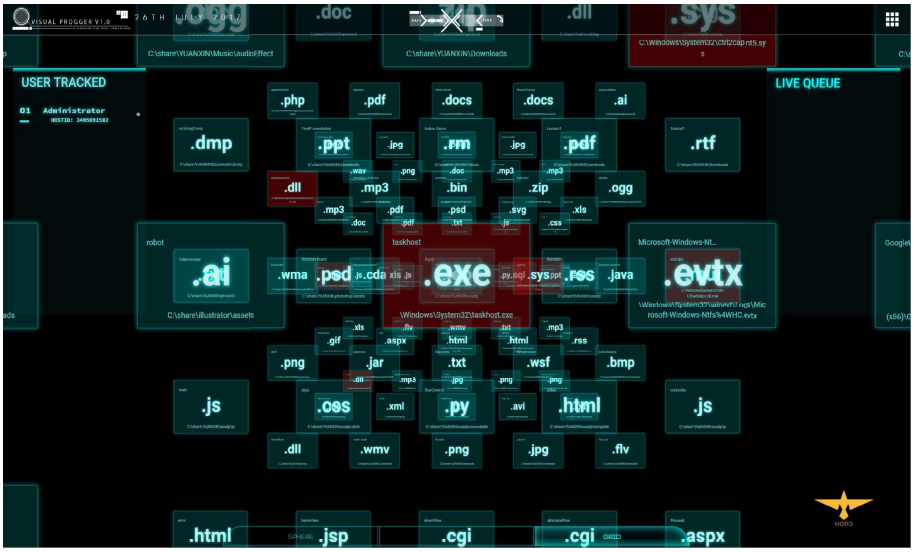


Figure 5.18: The VisualProgger Grid Visualization View

5.6.2 VisualProgger representation approach and features

VisualProgger shares several interactive visual representation features including zooming, dragging, selection, rotating and clicking. These features allow users to interact effectively with the security event visualized as users mentally process the event in a real-time scenario. Users have the ability to select and click on files-of-interest and investigate or execute further analysis.

There are additional user analytic features in VisualProgger implemented to enhance our entire effectiveness measurement approach. These features are: (1) visual nodes permanent-hold position, (2) viewing visual node dependencies, (3) file statistics capabilities, (4) critical file alert notifications and (5) user logged details. These analytic features are analysed thoroughly in Chapter 7.

5.6.3 Ransomware Visualization: Insights into Locky Visualization

In our VisualProgger locky ransomware analytics visualization, we utilise our granularity and layering design discussed in Chapter 5–Subsection 5.3.4. The ransomware security visualization is a static based reporting visualization. It uses data logged from a ransomware attack that are replayed using visualization to show users how the locky ransomware traverses through the system, scans, find and encrypts important documents found during the ransomware reconnaissance stage.

In the event of the ransomware scanning through files in the system, a visualization is shown whereby all files observed and touched, are classified into layers of directories, documents, libraries and processes. Dynamic Link Library (dll) files are also logged into a layer for visualization. In an event of identifying interested files, the locky ransomware locks all files causing the users not to be able to see and access their files and systems. These infected files are colour coded into red in the visualization, giving the users opportunity to select and click on the infected files to see additional details of interest.

Our VisualProgger ransomware visualization provides an interactive assessment on specific locky processes and visually show the files as they are encrypted. It indicated this while providing a security visualization from the time of reconnaissance up until the whole infected system is locked out. With VisualProgger, users are able to see information linking to the ransomware command control centre (CCC).

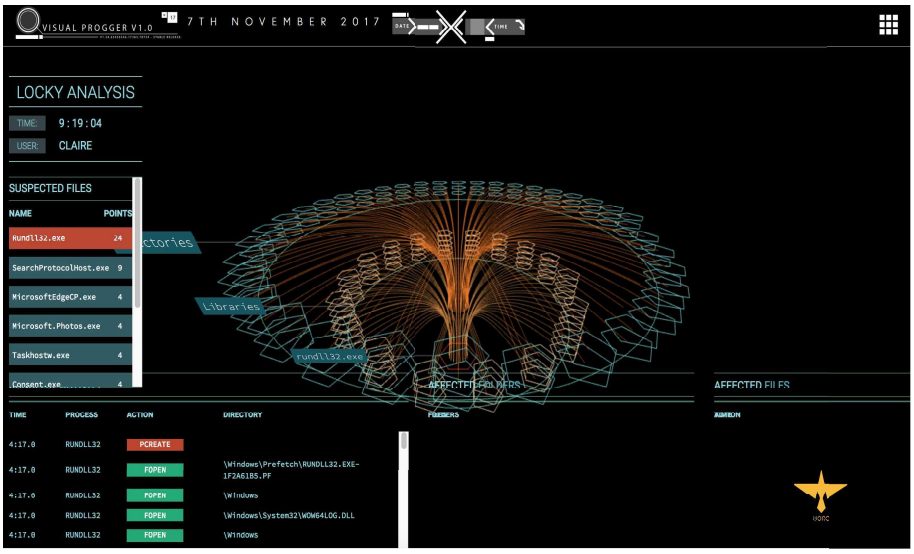


Figure 5.19: An Abstract Visualization Alert Sample

5.7 Use Case 2: Augmented Reality – A user-centric Security Visualization Application

Augmented reality environments are not new in aiding work processes, analysis and interactive visual experience. They are used in medical, engineering and computing science domains, whereby work processes are represented into various 3D visual models for analytics. Our SvEm Augmented Reality (AR) design emerged with the goal of empowering mobile platform users with opportunities to utilise their mobile platforms and observe while interacting with real-time cyber-attacks. AR delivers effectiveness with the mobile interactive capabilities with visualization. As a result, continuous user interactions with the security data are visualized. A visual experience with augmented reality for security visualization enables users to make sense of the security event scenario presented.

5.7.1 AR User Interface Design

Based on the designed augmented reality approach of visualizing cyber-attack events in real-time, we utilise the National Cyber Security Challenge (NZCSC) 2016 and 2017 event datasets to demonstrate our augmented reality experience with the red and blue team challenge. We maintained the use of colour and shape standards to allow a common and smooth user perception process when traversing through various security visualizations. A brief design of our AR security visualization shown in Figure 5.20, shows the red and blue team cyber security challenge visualization landscape. Each circle represents different types of attacks

and, a marker (Quick Response (QR) code), while the red and blue cubes represent red and blue team servers/machines. In addition, attack paths are also shaped in either red, yellow, green of blue respectively, indicating the type of traffic.

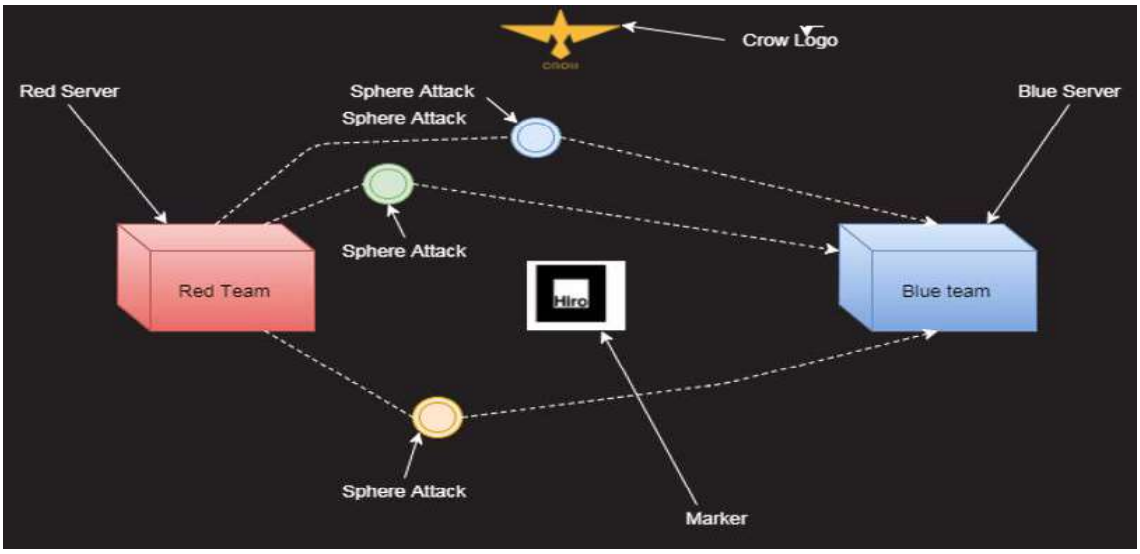


Figure 5.20: The SvEm Augmented Reality Security Visualization Design

Our Augmented Reality (AR) representation approach uses the interactive mockup design discussed in Chapter 5–Section 5.3.6. Data represented in our augmented reality environment are anonymised and analysed with the three filtering factors: (1) attack sources, attack destination (victim) and the type of attack executed. Finally, the augmented reality visualization has two vital roles, which are to provide visual real-time cyber security awareness and also a compatible visualization environment that is suitable for a larger range of audiences.

5.8 Use Case 3: SVInt Bitcoin Explorer – A Security Visualization Tracking and Intelligence Approach

In the law enforcement research domain, cyber security and cybercrime are growing concerns. The challenge of battling cybercrime activities across international jurisdictions brings the need for effective security tools. For example, the law enforcement operations require security tools and applications that are able to facilitate information sharing comfortably across transnational jurisdictions without revealing the underlying sensitive raw data due to privacy, integrity and confidentiality reasons. Security visualization methods and techniques are proven effective and crucial in aiding investigations and day-to-day cyber secu-

rity operations.

However, this use case uses the blockchain technology for a safe and reliable bitcoin transaction environment. We discuss the blockchain architecture as a core mechanism of our bitcoin explorer tool. An overview of the blockchain technology shown in Figure 5.21 explains how the blockchain technology works.

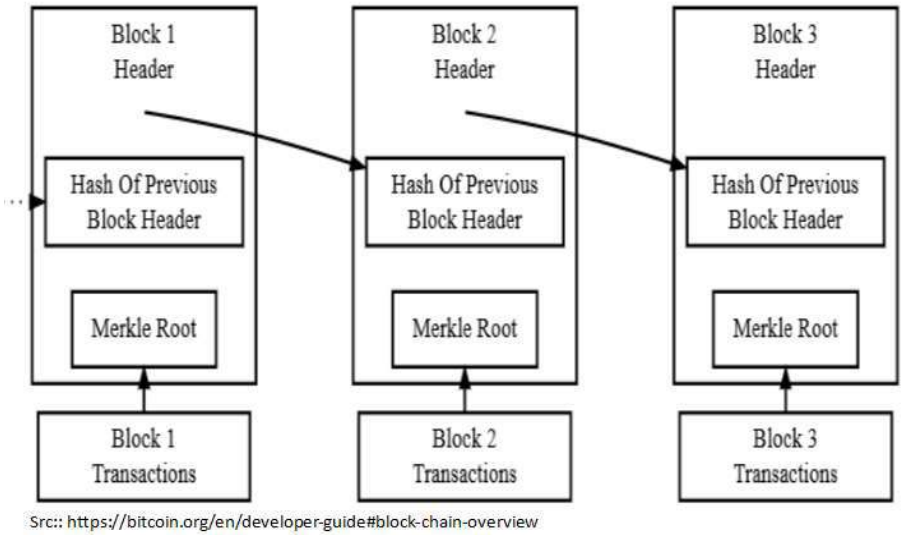


Figure 5.21: A Blockchain Technology Overview Design

Our SVInt Bitcoin security visualization use case is built as an added feature to an existing law enforcement in-house blockchain explorer tool [165]. With high relevance and importance to our security visualization framework, we showcase our bitcoin tracking and monitoring intelligence visualization prototype (application). This approach uses visualization to enhance law enforcement investigations by facilitating an efficient and effective method of tracking bitcoin transactions that are either suspected or associated with cybercrime activities. Law enforcement users (e.g., digital crime officers) are able to understand bitcoin transaction flows between bitcoin wallets, bitcoin exchanges and online market places (often dark markets) with this bitcoin visualization application. In addition, it provides statistics on malware attacks recorded by known security companies. Over a period of time, the visualization provides malware patterns and behaviours with geolocation mapping features, particularly of countries that are mostly affected. A design overview of our SVInt framework is shown in Figure 5.22, outlining all visualization functionalities.

5.8.1 Bitcoin Explorer user-centric Features

SVInt Bitcoin explorer user-centric features include the following features:

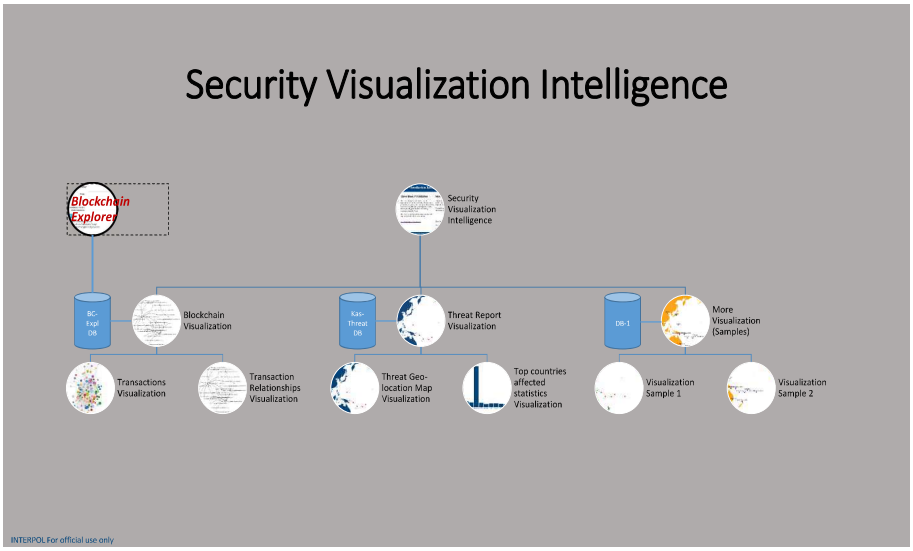


Figure 5.22: The SVInt Security Visualization Application Design

Standardised use of Colour Codes For Malicious Presentations

A key user-centric feature in our SVInt Bitcoin explorer visualization application is adopting Interpol's Notice system colour codes into our security visualization application. These colours are replicated across into our bitcoin visualization to represent similar offensive activities. For example, a red Interpol notice is an alert for a 'wanted' person, and with similar representation, a malicious file is shown in red and a suspicious file is shown in yellow. Our colour codes are further analysed and discussed in Chapter 6 and evaluated in Chapter 7.

One-to-One Bitcoin Transaction Mapping Representation

Another user-centric visualization feature implemented in our bitcoin explorer application is the ability for law enforcement users to track and monitor bitcoin transaction payments going in and out of bitcoin wallets. Transaction IDs (sender ID and receiver ID) are recorded along with the bitcoin amounts traded and are used for the visualization. However, due to the nature of how blockchain and bitcoin operates, our one-to-one bitcoin transaction mapping requires that we visually monitor the hops (jumps, transit points) of the transactions and trace/monitor the transactions through more than one hope. This allowed us to track and follow transactions through visualization and observe potential vital bitcoin transaction changes, which is important investigation information. In this visualization scenario, we applied the use of colour identifiers to coordinate transaction IDs based on the level of interest and suspicions, i.e., tracking and monitoring a suspicious transaction address involved in a cybercrime event. We utilise basic colour standards to track and monitor these

transactions as shown in Figure 5.23. A red coloured trace shows a targeted ID, which has been involved in a cybercrime, and a yellow link/trace indicates a transaction ID of interest for the investigation. Finally, a green trace shows normal legitimate bitcoin transactions.

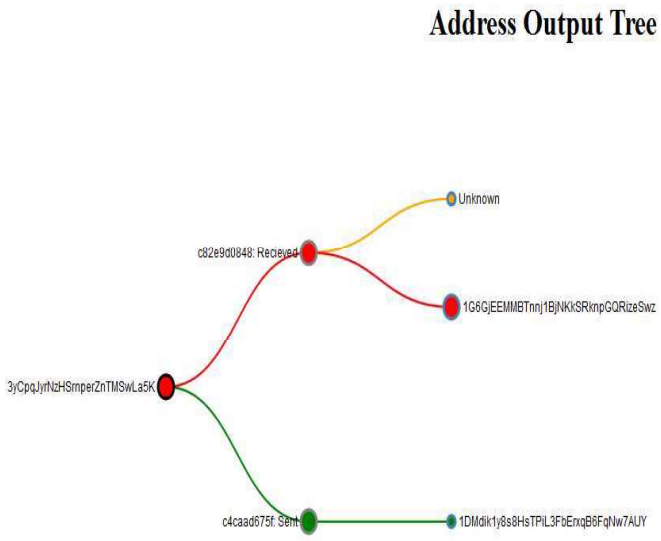


Figure 5.23: A Colour Coded Tree Visualization of Bitcoin Transaction Addresses

5.9 Security Visualization Application Services

One of the key advantages of using visualization in security is the ability to provide exploratory and intelligence services. In our SvEm framework, we have designed two application services to aid both web and mobile platform users. We designed and implemented 'Data Provenance as a Security Visualization Service (DPaaSVS)' and 'Security Visualization as a Cloud Service (SVaaCS)' to enhance the user's ability to use visualization as a security analytics tool, particularly in the cloud. Figure 5.24 delivers our two SvEm application services, with the aim of facilitating effectiveness in the user's experience.

5.9.1 Data Provenance as a Security Visualization Service (DPaaSVS)

The key idea of presenting DPaaSVS is to allow users (viewers) to push beyond their visualization experiences and use security visualization as a provenance service in cyber security. Daily operations such as command control centres are an example of where this service is used. Cloud applications that are given the ability to have logging mechanisms installed

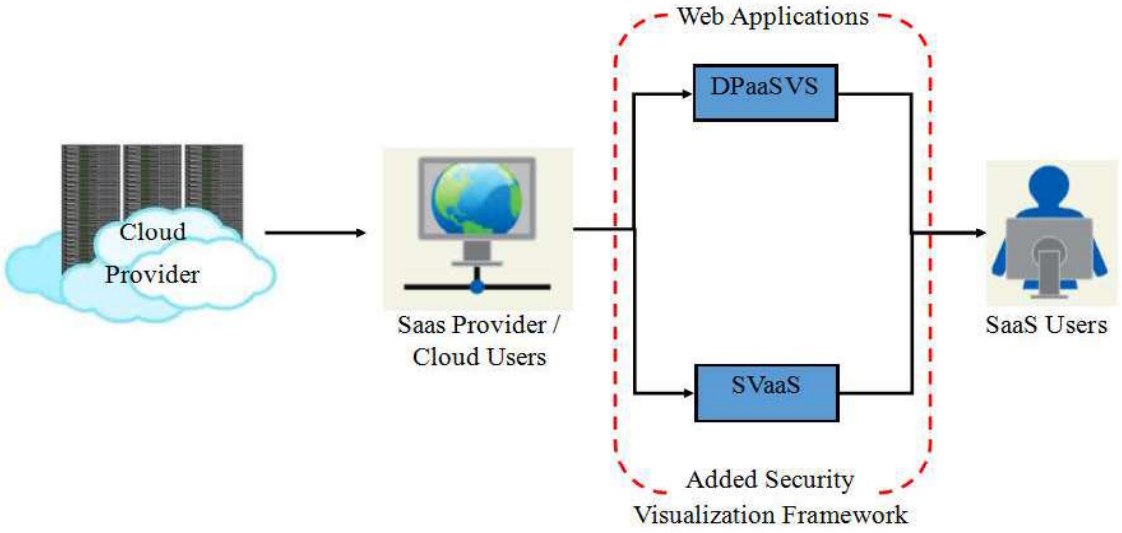


Figure 5.24: SvEm Security Visualization Services

are also seen as visible for DPaaSVS. It facilitates the ability to track and follow data (files) movements from the time of creation to the current state. It is an active visual provenance analytic tool with two roles, namely tracking and monitoring functionalities. It is designed for large datasets with exploring and reporting purposes in security visualization.

5.9.2 Security Visualization as a Cloud Service (SVaaCS)

Security Visualization as a Cloud Service (SVaaCS) is another potential security service that helps analyse complex security data for provenance purposes. The use of visualization to show the derivative history of meta-data associated with security events and malicious attacks is useful for security researchers and cloud users. SVaaCS offers the users an opportunity to see how security visualization can be effective. In addition, utilising this visualization with existing cloud applications for security tracking and monitoring, thus aiding security operations. We assess our proposed SVaaCS in detail in Chapter 7 and evaluate these services further.

5.10 Summary

In summary, this chapter delivers all the practical implementation aspects of our SvEm algorithm. We incorporated our SvEm designs discussed in Chapter 4 with our SvEm requirements to implement several use cases. In our applications, we focused on implementing user-centric features that capture the user's attention. These components are part of our

security visualization effectiveness measurement and evaluation indicators. An example is, implementing interactive security features and user-trigger components that maximise the viewer's working memory load when observing a given security visualization.

Moreover, we deliver several security visualization types including: (1) VisualProgger–a user-centric real-time interactive visualization application and a ransomware security visualization, (2) an augmented reality security visualization application implemented for a national cyber security challenge event, and (3) SVInt Bitcoin explorer–an intelligence security visualization application to monitor bitcoin transactions between suspicious wallet addresses (IDs).

Chapter 6

A Security Visualization Guideline - Towards developing a Standard (SCeeVis)

The increasing significance of information technologies and services create a crucial need for suitable measures in data/information security [166], [167]. As such, the existence of information security standards was implemented to enhance basically the following groups of people: (1) researchers and developers and (2) organisations wishing to implement security frameworks. These information security standards exist to aid auditing processes and measures. Standards and guidelines play vital roles into achieving full-scale effectiveness measurement and evaluation techniques in security visualization. This chapter delivers a proposed security visualization (SCeeVis) standard which addresses effectiveness performance and assessment throughout the design, implementation and observation (DIO) phases of our security visualization effectiveness measurement (SvEm) framework. It serves the purpose of setting the scope for visualization designers, implementers and users of the visualization. Our SCeeVis proposed standard aims to facilitate fast, effective information communication, and reduce and remove ambiguities (confusions) in both the security visualization presented and for the users. The application of this standard will ensure consistency in security visualizations throughout the DIO process, therefore increasing user motivation and interaction for maximum knowledge gained. However, due to the issue of visualization developers/users having different preferences, we established SCeeVis standard as a *referencing model for effectiveness measurement and evaluation* in security visualization. Thus, in this chapter, we: (1) state the importance of security standards, (2) provide a review of existing standards, (3) present our SCeeVis standard and its functionalities, (4) outline the presentation methodologies, (5) present the SCeeVis uses and applications and finally evaluate the standard presented. We begin this chapter by describing the role and importance of standards and review existing visualization standards that are related to this thesis.

6.1 The Role and Importance of Standards and Guidelines

Information security is an important yet sensitive component to an individual or organisation on a day-to-day basis. It means protecting personal data and sensitive commercial information [168]. Thus, the existence of security tools and framework play the role of protecting this sensitive information. However, *how do we measure effectiveness of information (data) security*? This is a question security professionals and companies are asking. With that said, there has to be an indicator that should contribute to assessing efforts and judgements in a security process. This is why security standards are implemented. Their existence allows monitoring, analysis, measurement and evaluation of security frameworks [168]. Standards help explain methods of how to implement and manage measurement processes and assess and report on information (data) security metrics. With such processes, security metrics implemented for security standards produce insights into the effectiveness of a security framework.

The concept of implementing security standards is to provide users and organisations with a basic mutual understanding when using tools and frameworks. Standards facilitate better interactions, communication, measurement, design and implementations [169]. In comparison to a user-manual, standards help provide users with necessary information on how certain processes function. This information encompasses are well-defined and developed detailed characteristics, guidelines and rules with precise knowledge around certain products, materials, or services, and processes' execution.

However, different disciplines (research domains) classify standards into various names, references and are purpose-driven in terms of addressing specific security implementation areas in the technology work processes. The word 'standards' is the overarching term to describe a of set guidelines, rules, policies, regulations and reference models. For example, the ISO/IEC 27000 series of standards [167]–Information security management systems (ISMS) is an information technology (IT) initiative that ensures information assets are secure. Thus, organizations have sets of guides to meet in order to be certified under the ISO/IEC 27000 series standards. This is discussed in Subsection 6.1.1.

6.1.1 ISO/IEC 27000 Series of Security Standards

In information security, the International Organization for Stanardization/International Electrotechnical Commission 27000 (ISO/IEC 27000) family of standards is the standard that is internationally recognised. It covers the ISO/IEC 27001:2013 - Information security management and ISO 27032 - Guideline for cyber security established the guidelines and certification standard for Information and cyber security [170], [171], [172]. Table 6.1 deliv-

ers the summary of the information/cyber security standards which are related to our proposed standard. From a practical approach, the ISO 27001 formally provides a guideline for managing information risks through the use of information security controls within an organisation. For example, the ISO/IEC 27001:2005 provided the "PDCA (Plan-Do-Check-Act/Adjust) cycle or Deming cycle, which helps guide security experts throughout the implementation cycle of information security frameworks." ISO/IEC 27032, on the other hand, is a cyber security certification standard based on Internet Security [172]. Furthermore, the ISO/IEC 27000 series of standards allow efficient communication and support between users and companies when obtaining products and services [166]. Overall, security standards, guidelines and policies are implemented to contain and maintain the aspects of security events. Alternatively, the use of visualization in cyber security requires a security visualization standard [173], [174] to assist visualization developers and researchers.

Table 6.1: Standards Directly Related to Cyber Security

ISO/IEC 27000 Series	
ISO Codes	**Description**
ISO/IEC 27000	Information Security Management Systems (ISMS) - Overview
ISO/IEC 27001	Information Technology: Security Techniques in ISMS
ISO/IEC 27032	Guideline for Cybersecurity

In summary, the ISO/IEC 27000 family of standards provides an overarching framework used by organisations to assess how effective their information security frameworks are. Such standards allow consistency in information security processes with assurance of security implementation across the security landscape. With that said, implementing a similar set of standards and guidelines for visualizations prompts the need to review existing standards available in the visualization research domain.

6.2 Background: Visualization Standards

Over the past years, the discipline of visualization studies was commonly used in security for various purposes. International information/data visualization standards were implemented as visualization evolved. These standards facilitated the process of optimal decision making through innovative visual content. For example in the early history of visualization, Minard's Carte figurative des pertes map as seen in Appendix A1 (Figure A.1) illustrated both provenance and attribution through the use of visualization. This example is seen as a vi-

sual mapping reference model which ignited the popularity for visual mapping throughout the past years, increase of use of visualization for mapping purposes. Ralph Lengler and Martin J. Eppler delivered an interactive 'periodic table of visualization methods[175]' as seen in Figure 6.2. These two examples illustrated the need for innovation and guidelines to demonstrate the purpose behind visualizations. The periodic table serves the purpose of compiling and classifying different visualization methods, namely: (1) data visualization, (2) information visualization, (3) concept visualization, (4) strategy visualization, (5) metaphor visualization and (6) compound visualization methods. With the periodic table, researchers have the ability to understand various visualization methods, and their specific representation goals. For example, in the data visualization methods tables, pie charts, line graphs and scatterplots are used for statistical and comparison reasons.

The periodic table of visualization methods gives users the opportunity to interactively explore high quality visualization formats for specific needs and problems. Users are able to make better choices when given the ability to select a method of visualization from multiple options presented. However, in the cyber security research domain, visualizations are unstructured and often focused on specific problems. These are either visualization implementation issues and/or data representation problems, thus causing the entire visualization environment to be complicated. Therefore, we reviewed existing visualization standards and guidelines to identify research gaps which are implemented and addressed in our SCeeVis standard.

6.2.1 Industrial Visualization Standards

Although there is a wide range of proposed data visualization standards and best practices for industries, we identify several standards and discuss their importance as part of this chapter. This helps provide a comparison between academia and industry approaches. For example, the International Business Communication (IBCS) [176], [177] Standards provided the practical proposal and means for the design, presentation, dashboards, diagrams and tables used. This concept involves correlation of visual perception and semantic unification. Apart from the proposed IBCS standards, there are 'reference models' and 'best practices guidelines' that aid visualization developers. For example, the seven data visualization best practices include: (1) having a methodology, (2) knowing your audience, (3) defining resulting actions, (4) classifying your dashboard, (5) profiling your data, (6) using visual features, and (7) designing iteratively.

Microsoft's PowerBI [178], [179] has delivered best practices, which acts as a guide to visualization implementers. It addresses Power BI's best practices for designing visuals and re-

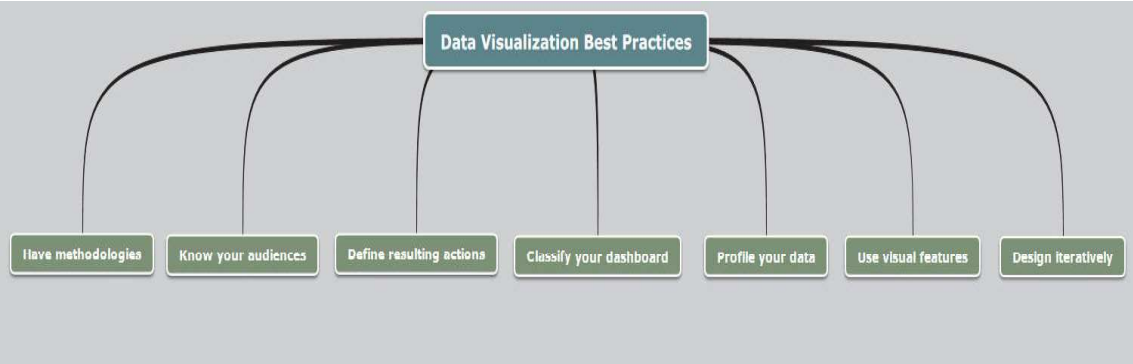


Figure 6.1: The Seven Data Visualization Best Practices

ports for dashboards beginning from the planning stage. Tableau [180], [181] also delivered best practices for building dashboard visualizations. It provided a set of basic instructions and tips on making dashboard visualizations effective. The features highlighted for tableau include trends over time, comparison and ranking, correlations, distribution, part to whole and geographical data. For example, commonly used methods when analysing data, is to track trends over time and observe changes, patterns and quantitative results. These features are the key to analysing data and producing effective visualizations.

Visualization best practices are very specific to visualization implementers and for certain types of visualization presentation. They are common for business intelligences due to the nature of the data analysed with respect to the targeted audiences. However, different visualization purposes require different forms of visualization implementation approaches. In addition, datasets with different landscapes and environments also affect decisions when designing and implementing visualizations. Therefore, the process of data analytics often dictates how a visualization is presented. Finally, the targeted audience is a crucial factor that affects the outcome of the visualization presented. Therefore, with respect to all the above mentioned, we reviewed taxonomies, reference models and best practices from a scientific stand-point to help enhance our SCeeVis Standard implementation and evaluation.

6.2.2 Additional Components: Visualization Taxonomies, Reference Models and Best Practices

While standards and guidelines are the core focus of this chapter, taxonomies, reference models and best practices are additional referencing materials that help provide additional features and understanding all forms of classifications throughout the visualization process. These are discussed in the subsections 6.22.1, 6.2.2.2 and 6.2.2.3 of this chapter.

6.2.2.1. Visualization Taxonomy–Data State Reference Model:

Ed H. Chi's taxonomy of visualization techniques [182] using the 'Data State Reference Model' provides an analysis of the information visualization design space with respect to data types and steps in process operation [183]. The data state reference model helps researchers to understand the space allocated for design and how information visualization techniques are applied in a broader approach. It categorised the data stages into four stages [182]: (1) value, (2) analytical abstraction, (3) visualization abstraction and (4) view. 'Value' refers to the raw data used and 'analytical abstraction' refers to data about data or information (meta-data). 'Visualization abstraction' refers to information that is viewable on screens using certain visualization techniques. 'View' refers to the visual end-product of the visualization mapping, where viewers see and interpret the images presented to them.

In addition, the data state reference model also provided transformation operators [182] containing processing steps. These are (1) data transformation, (2) visualization transformation and (3) visual mapping transformation. These processing steps basically outline analytical steps such as data extractions, specific detail extractions and processing information into viewable format then presenting it in a graphical interface.

Finally, in this data state reference model, the emphasis is on providing precise data analytics for better design space usage. This allows visualization developers to process data according to the available visualization space. While this is very useful for developers, our SCeeVis standard approach incorporates the idea presented by the data state reference model, and more features which will be discussed in Section 6.3 of this chapter.

6.2.2.2. Visualization Taxonomy–Information Visualizations:

Another important approach seen in existing visualization standards involves investigating interactions in visualization systems. In most interactive visualization frameworks such as dashboards, users can find it challenging when attempting to interpret multiple interaction events in more than two visual views. For example, executing basic visualization tasks such as data or graphical operations that are read from one dataset source in two or more views is challenging for users. However, Ed Huai-Chi and John T. Riedl's operator interactive visualization taxonomy [183] for visualization systems enables new exploring and evaluation methods for the design space of visualization operators. They incorporate Basic Visualization Interactions (BVI) to enhance data filters for more interactive detailed characteristics in information visualization, which thus helps users during their analytic tasks.

Moreover, analytical visualizations require standards and guidelines to assist with understanding the nature of the raw data collected, the predefined data presented and the avail-

able interactive features that are provided by the graphical user interface. Ben Shneiderman's [184] research contribution with 'The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization,' builds on the visual information-seeking mantra: overview first, zoom and filter, then details-on-demand to provide a taxonomy of task by data-type. It delivers seven data type tasks, namely: (1) 1-dimensional, (2) 2-dimensional, (3) 3-dimensional data, (4) temporal data, (5) multi-dimensional data, (6) tree data and (7) network data. Hence, the data-types offer seven sets of tasks, outlined in Table 6.2, namely: (1) overview, (2) zoom, (3) filter, (4) details-on-demand, (5) relate, (6) history and (7) extract. These tasks are attractive to users because of their functionalities, which enable data to be presented rapidly with the ability for users to explore further and achieve insights.

Table 6.2: A Set of Data-Type Tasks

The 7 Tasks by Data Type Taxonomy	
Tasks	**Description**
Overview	Gain an overview of the entire collection
Zoom	Zoom in on items of interest
Filter	Filter out uninteresting (noise) items
Details-on-demand	Select an item or group and get details when needed
Relate	View relationships among items
History	Keep a history of actions to support undo, replay, and progressive refinement
Extract	Allow extraction of sub-collections and of the query parameters

In summary, the taxonomies discussed address interactions, and data presentation features to enhance the user's ability to interact freely. This reduces perceptual confusions often due to multiple views and various interactive features used concurrently during the observation and interaction with visualizations.

6.2.2.3. Periodic Table of Visualization Methods:

As briefly mentioned in Section 6.2, Ralph Lengler and Martin J. Eppler's interactive 'periodic table of visualization methods[175]' is a reference model for visualization developers whereby various visualization methods are provided to aid decision making. This periodic table approach in Figure 6.2 allows researchers to recall knowledge and familiarity from known concepts of the periodic table of elements [185]. It aims to define and compile various visualization methods to provide an interactive systematic overview of graphical formats needed for visualization development. Hence, the periodic table of visualization methods supports researchers and developers in picking relevant methods with the appropriate

application parameters used for visualizations.

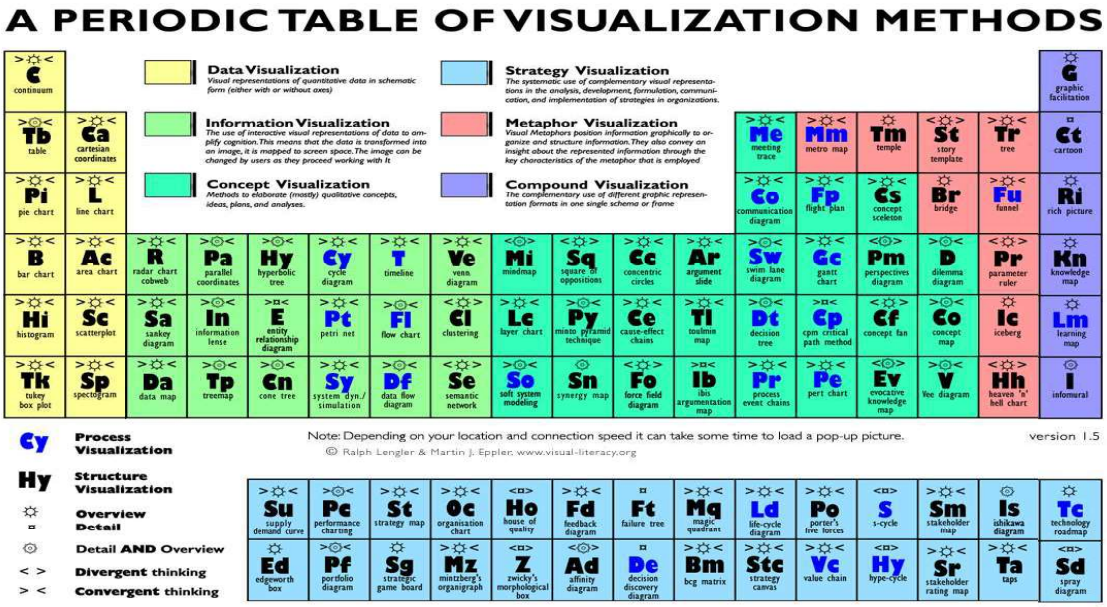


Figure 6.2: A Periodic Table of Visualization Methods

The similarities adopted into the periodic table of visualization methods eliminates the time spent on processing and learning what the functionalities are. This allows researchers to effectively build on known working memory and deliver the appropriate visualization method needed when required. However, the adoption of taxonomies, reference models and the periodic table approach still requires further standards and best practices targeting the visualization representation and users of the visualization. Thus, we address Gestalt principles (theory) of perception to connect all standard functionalities and identify possible research gaps.

6.2.3 The Gestalt Principles of Perception

The Gestalt principles of perception contribute to setting up parts of the security visualization standard by facilitating techniques of understanding how humans (users) perceive and process visual images. Gestalt theory covers a wide range of concepts and approaches to the entire psychology of visual perception. In this thesis, the inclusion of Gestalt principles of visual perception targets two areas: (1) the visualization designs and (2) the user's visual process, approach and experience in an event of visualizing security visualizations. Thus, building on scientific visual approaches, the ability to break down objects into single parts and define them allows the human brain to process visual perception and form

the 'sum of whole parts' [186], [187], [188]. Respectively, Gestalt principles of perception provide the basic user an observation approach around what to understand in visualization during the perceiving process [189], [190], [191]. It particularly refers to the 'principles of grouping,' whereby users perceive objects (security visual nodes) to understand visualization through an orderly, symmetrical and simple approach [192], [190], [188]. Thus, this process allows users to identify patterns faster and effectively. A brief overview of the Gestalt principles [190], [188] is outlined in Table 6.3 with descriptive summaries of each law.

Table 6.3: Gestalt's Laws (Principles) of Grouping Overview

Gestalt's Principles of Perception	
Laws	**Description Summary**
Proximity	States that things which are closer to each other are perceived as a group
Similarity	States that things which share visual characteristics are seen as belonging together
Continuity	Predicts the preference for continuous figures
Closure	Our eyes perceive and fill in the missing information to form a complete figure
Pragnanz (Good Figure / Simplicity)	Objects are perceived in the simplest way possible
Symmetry and Order	Provides the feeling of solidity and order

The outlined Gestalt principles of visual perception provided an approach around how to perceive objects through visualization. However, perception itself plays a very important part to our entire effectiveness measurement approach and standard. Hence, these principles demand the need to shift from the visualization provided and see two important components in users (viewers) that enhance the entire visual experience. Both components contribute directly to providing effectiveness and assessment measurement in visualizations. These are: (1) visual perceptual boundary and (2) a visual perception threshold. The visual perceptual boundary refers to the area established around a stimulus which improves visual processing of objects when observing visualizations. For example, given various groups of malicious patterns observed in a given visual space, users naturally establish a perceptual boundary to allow effective processing of visual information. In addition, the visual perception threshold establishes the user's perception limit (maximum and minimum) and capacity during the perceiving process. Finally, with these two factors identified, establishing a standard for the visualization framework and the user's visual environment bridges the gap between the visualization presented and the user (viewer) during the observation period. However, there is room for improvement; an example is by, incorporating Gestalt Principles

of Visual Perception concepts with existing security standards such as the Interpol notice system. This would add more user-trigger features, because by default, users relate effectively to colours for differentiation and comparison purposes.

6.2.4 The INTERPOL Notice System

Apart from the ISO/IEC 27000 series of standards - Information security management systems (ISMS)–there have been several visualization standards, best practices, information policies and reference models implemented over the years, e.g., the 'Gestalt Principles (law) of Perception and the periodic table of visualization methods. However, in the security visualization field there are lack of standards and guidelines to guide visualization researchers, developers and users in all aspect of designing and developing/implementing security visualizations. Thus, we explored the concept of 'INTERPOL's Notice System' to create and establish our security visualization Guideline (SCeeVis Guideline). The work of SCeeVis Guideline is the beginning of a research project towards developing security visualization Standards.

Figure 6.3: The INTERPOL Notice System Concept

The INTERPOL Notices system was designed to facilitate information sharing among member countries [154], [193]. In this case, critical crime-related information is shared by using a simple yet effective visual concept that distinguishes different crime types. The use of colours shown in Figure 6.3 shows the different types of notices implemented by INTERPOL to allow police intelligence to share information efficiently. Although, the notice system is

not seen as a standard in its existence due to the definition of standards and guidelines, the performance is effectively utilizing a visual information sharing method which behaves as a standard or guideline. For example, the red notice [193] simply alerts law enforcement to 'seek the location and arrest a wanted person.' A yellow notice alert refers to the need to locate a 'missing person or help identify people' who have difficulties identifying themselves. In addition, there is the blue, black, green, orange, purple and the INTERPOL-United Nations security council special notice system which is issued for groups of people or individuals who are the target of the United Nations security council sanctions committees[154].

The notion of using a colour identifier for information sharing is simple, effective and has a broader coverage of targeted audiences, given its simplicity to facilitate effective information sharing to help provide critical crime-related information shared among international law enforcement agencies. It does not require technical experts to analyse the visual notice system and explain the crime details. It is self-explanatory. Thus, our security visualization (SvEm) framework requires such concepts be harmonised into a security visualization framework to maintain the colour standard and generic meaning of what each colour meant. We adopted the INTERPOL Notice alert concept and built on both the security and colour concepts to show cyber-attacks. Hence, we address these concepts and describe our security visualization guideline (SCeeVis) and how it performs, i.e. the development and establishment of security visualization guideline towards developing a security visualization preliminary Standard.

Overall, these taxonomies, guidelines, reference models and best practices address various parts of the visualization design and implementation process. However, our development of SCeeVis guideline (preliminary standard) goes beyond these design and implementation phases of the visualization development and provides the means of effectiveness measurement for users through the implementation of the SvEm framework. It facilitates the additional user guideline and reference model targeting the users of security visualization frameworks and tools. These are discussed in Chapter 6.3.

6.3 The SCeeVis Security Visualization Guideline - towards A Cyber Security Standard

Due to the complexity resulting from user-preferences, data-complexities, visualization representation and presentation, we designed and implemented our SCeeVis standard to minimise these complexities by reducing doubts and cognitive biases in users. The SCeeVis outlines basic guides for researchers, visualization developers and users of security visualizations. As a result, the intended visualization purposes are observed in a comfortable

environment, thus critical security information is communicated across to the audience.

The SCeeVis guideline provides a set of rules, controls and scope for security visualization usage. It maintains and control all types of visual outputs required for the users to view. This guideline help viewers use security visualization and effectively make sense of cyber-attacks when visualizing complex transformed data. Hence, we establish our SCeeVis Guideline, a work towards development of a new standard, as a core component of our SvEm security framework. This guideline or pre-standard has a high impact in communicating critical security-related information across users. However, how this standard works is something that needs to be addressed.

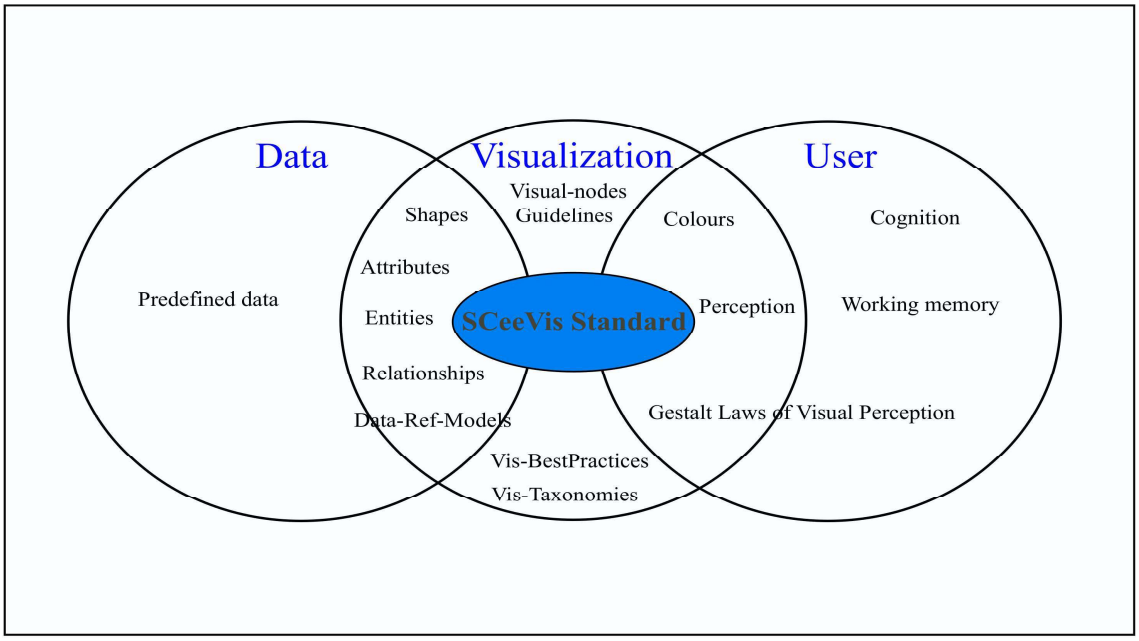


Figure 6.4: Our Proposed SCeeVis Pre-Standard Conceptual Model

The proposed SCeeVis Security Visualization Guideline overview is explained with the use of a conceptual model seen in Figure 6.4. The SCeeVis standard aims to provide developers and users with a guide to maintain a full-scale effectiveness approach throughout the SvEm visualization framework. Hence, there are several important components taken into account when developing this guideline:

1. *Data*: this refers to the security data collected and processed for visualization. The guideline aims to provide methods on anonymising and standardising the raw data. The emphasis is on managing rendering and scalability issues by ensuring the data are processed in a timely manner without overloading the backend and frontend of the security visualization framework.

2. *Visualization*: this refers to the security visualization presented. Our SCeeVis guideline basically addressed issues and complexities faced by both visualization developers and the viewers. The guideline adopts existing visualization best practices, reference models, guidelines and security standards, thus providing an additional user-centric visual standard component to the entire approach. The proposed contributing factors considered in this guideline are: colours, perception, Gestalt laws of visual perception, shapes, security entities/attributes and relationships.

3. *User*: this refers to the visualization viewers and developers of the visualization. The standard facilitates a user-trigger guide to activate cognitive abilities and increase user attention span. It is done in an enhancing environment through the use of colours, perception attributes, existence of working memory loads and visual interactive features.

Finally, the targeted audience is the visualization developers and users (viewers) of the visualization. Overall, the SCeeVis guideline aims to bring both developers and users together into one visual experience realm, by ensuring a common understanding throughout the design, implementation and perceiving stages of the SvEm visualization presented. Often, researchers and developers implement security visualizations for specific purposes and at their own liking, hence their choice of visual design. There is a need for the use of common colours and symbols in visualization representations to facilitate common information transformation, communication and sharing across various users. Thus, it is important to understand this critical component of the guideline, which is the use of colours in security visualizations and know what each colour meant (stands for) and the overall meaning behind their usage. These selection of colours are formalised into our SCeeVis guideline to help enhance the knowledge of researchers, developers and users around the use and interaction with visualizations. Thus contributing to achieving a full-scale effectiveness approach.

6.3.1 The Security Visualization Colour Identification Guideline

It is critical to standardise the use of colours in security visualization. A large amount of data with interested entities requires a simplified security visualization. This is crucial for enhancing rapid information processing. For example, using the red and orange colour in the same visual space automatically creates confusion to users, resulting in complications for the entire visualization experience. Our standardised colour selection shown in Figure 6.5 is:

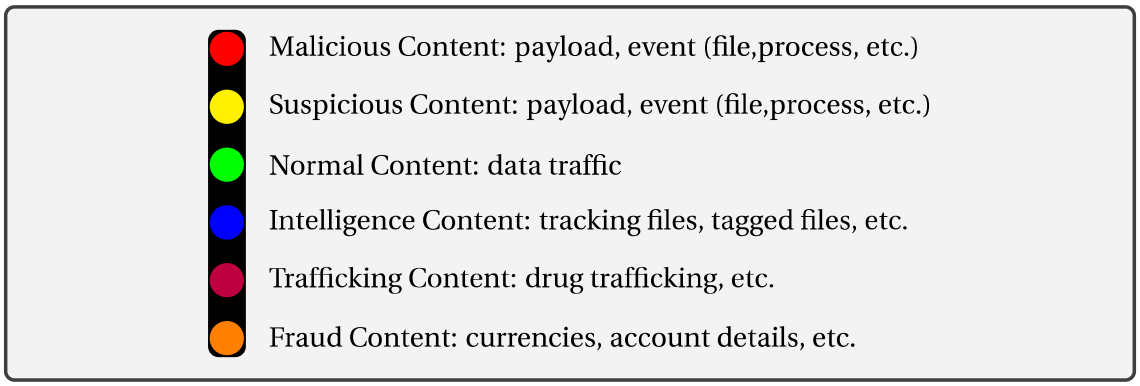


Figure 6.5: Our Security Visualization Colour Guideline

Red, Yellow, Green, Blue, Purple and Orange. These colours are grouped into two categories– primary and secondary groups. Our primary colour choices for security visualization are red, yellow, green, and blue. The secondary group are purple and orange. These additional colours are specifically for law enforcement security visualization with concepts matching the Interpol's colour-coded Notice system [194], [193]. Note, the use of red and orange in the same visualization is not encouraged. For example, orange is only used to show illegal trafficking content and it is regarded as an independent visualization type.

Overall, our colour standard addresses simplicity, familiarity and the establishment of a comfortable environment thus leverage on prior knowledge for an effective learning process. From a developer's approach, understanding the colours and matching them to security event attributes in visualization is important. It is part of the colour management process which avoids the issue of colour overlapping in representing security incidents. Colour overlapping in visual presentations contributes to scenarios where there is visualization misinterpretation or cognitive biases. Furthermore, the users have the opportunity to understand, compare and observe faster given the visualization presented uses known colours. Human vision is becoming naturally familiar with these colours proposed for this standard. The colours red, yellow, and green are commonly used in everyday occasions such as in traffic lights, road signs and more. Users are able to reassemble these colours used in security visualization and interpret them with similar meaning, i.e., red for malicious (stop/-danger), yellow for suspicious (warning) and green for normal. The choice of using such colours in a security visualization standard automatically reduces the user's working memory load during the visualization observation period and increases the attention span given the user is familiar with the use of colours. Theoretically, a user should be able to connect the dots within a given security incident much faster and effectively due to increased attention span and prior knowledge when using this colour guideline.

Finally, a good knowledge of what the colours mean and how they are used establishes the need to understand the SCeeVis guideline functionalities. It is also required to know what each entity and attribute means when used in security visualizations. These functionalities are outlined and require very basic knowledge to understand them.

6.3.2 The SCeeVis Functionalities

Our SCeeVis guideline or pre-standard covers and provides an introductory content overview of a new security visualization standard. It is important that visualization developers and users understand what is needed to visualize from the processed data. It is also important to know what visualization developers would want the users to see in security visualizations. Thus, with a clear understanding on both the user's demand and the visualization needs from the developers perspective, will establish a common ground with the overall visualization intention and output. Therefore, developing a visualization should contain all security features intended to be presented when observed by the users. There are two distinctive parts of the security visualization guideline and these are:

- *Part 1: Understanding What is Needed*: In this stage, researchers and developers have to understand the purpose of the intended visualization, nature of the data collected, and, most importantly, know the users who will be viewing the given visualization.

- *Part 2: Security Visualization Process*: In this stage, a clear understanding of who the targeted audience is very important. The design and implementation phase is critical; for example, choosing the right colours, objects and relationship links are critical for an effective visual output. These choices and decisions primarily focus on the data collected and the security event (e.g., cyber-attack) needed to be visualized.

The SCeeVis guideline Part 1 covers the preparation aspects of the entire backend design and implementation process of the visualization framework. This includes the visualization design thinking process which is fact seen as the information gathering process. It involves understanding that is needed from the security datasets collected for visualization. It is categorised into three sub-parts: (1) the problem, (2) the purpose and (3) the cyber-attack landscape. SCeeVis Part 1 basically guides visualization developers to analyse and understand the dataset from an end-user's mindset. This means breaking down the data collected into categories, and understanding the cyber-attack landscapes, relationship links and the representation techniques required for visual analytics visualization.

Security Visualization Standard (SCeeVis)

Part 1: Understanding what is Needed

1. **The Problem**: **-** Identification of security events
    - Identify security events (e.g., malware attack, SQL injection, etc.) and data type (raw data: log files, social media data, etc.)
    - Understand the nature of data (e.g., financial data, health data, etc.)
2. **The Purpose**: Know the visualization type and technique
    - Understand the intention of security visualization (e.g., show relationships, etc.)
    - Decision: exploratory or reporting visualization, security visualization technique (e.g., categorising: time-based, provenance-base, attack-based)
3. **Cyber-Attack Landscape**: Know the cyber-attack location (e.g., network, systems, application layer, etc.)
    - Know the point of attack (e.g., network attack, identify source and destination of attack, etc.)
    - Attribution of cyber-attack

The 'problem' sub-part requires the information gathering, data analysis and knowing what the security event is. The 'purpose' sub-part deals relies on the information and knowledge gathered in the 'problem' sub-part. With such information, a visualization type and technique is chosen and designed with interactive user-centric features. Once, the problem and purpose are identified and executed, it is important to know and understand the security event landscape, propagation methods and if possible know the source and destination of the security event. This means obtaining and understanding the provenance and attribution details associated with the event, if known.

Finally, in Part 1, visualization researchers and developers require indirect knowledge on the targeted audiences. This knowledge could the researchers and developers to predict what the users would want to see, how to view it, and the know the predictive information (knowledge) extracted and portrayed by the visualizations.

Part 2: Security Visualization Process

1. **Visual Presentation Methodologies**: How to present data visually
2. **Colour and Shape Standard for Security**: Decision on choice of colours
    - Standardising main colour choices
        - Color: Red = High attack nature or violation (e.g., malware process)
        - Color: Yellow = Suspicious process (e.g., IP address)
        - Color: Green = Good or normal process (e.g., network traffic)
        - Color: Blue = Informational (intelligence) process (e.g., IP address)
        - Color: Black = Deleted, traces: non-existed (e.g., deleted file)
    - Standardising main shapes choices
        - Shape: Circle = Nodes (e.g., network nodes)
        - Shape: Rectangle = File entities (e.g., .docs, .jpeg, etc.)
        - Shape: Square = Data clusters (e.g., IP address - network traffic)
        - Shape: Diamond = Web/social media entities, process (social media data)
    - Standardising Use of Line Types
        - Line: Single line (- - -) = Relationships, connections, links, provenance, time-base (e.g., between two network nodes)
        - Line: dotted line (– – –) = Relationship interactions, or possible relationships (e.g., .docs, .jpeg)
        - Line: Solid arrow (→) = Direction of relationship or interaction
        - Line: Dotted arrow (–– >) = Direction of predicted relationship or interaction
3. **Security Visualization Techniques**: Provenance and attribution-based, user-centred, real-time based
4. **Security Visualization Type**: Animated, 3D, static, etc.

Part 2 of the SCeeVis guideline involves the security visualization process. This process has four sub-processes, which include: (1) visual presentation methodologies, (2) colour

and shape standard for security, (3) security visualization techniques and (4) security visualization types. The visual presentation methodology covers the process of visual representation designs and the presentation method. It includes understanding what the security data contains and the nature of the security event. The colour and shape standard for security sub-process highlights the need to standardise the colours and shapes used for security visualization. This maintains the same level of understanding for both the developers and users; thus, browsing through multiple security visualizations does not require the user to re-learn what colours and shapes mean. Hence, the colour and shape standard are easy to observe, process, and an effective method to communicate critical security information through visualization. Although colour and shape representations are critical, the techniques used for security visualizations are important, as they help relate the story behind the data presented. For example, this involves presenting various visualizations to show provenance, attribution, real-time security events, patterns and behaviours of suspicious, or malicious events. This sub-process involves gaining access to the user-trigger features, which helps activate user cognition during the process of the visualization observation. Finally, the security visualization type sub-process is adding an additional user-centric feature to the entire visualization guideline. It is based on the type of data and the predicted knowledge required to guide the users through the security visualization.

6.3.3 SCeeVis Guideline Requirements and Presentation Methods

The key requirement for any standard to be functional is the notion of common understanding by security researchers, developers and users of the standard whereby a common language is communicated for visual representation and presentation in security visualization. This involves linking all SCeeVis guideline features together and knowing how each feature functions. Firstly, the SCeeVis guideline requires to building on the existing collection of standards, taxonomies, best practices and guidelines outlined in the SCeeVis guideline conceptual model (Figure 6.4), to establish the basis of this standard. Secondly, the adoption of our proposed colours into SCeeVis is crucial. Thirdly, the implementation requirements for the SCeeVis guideline need to be clearly outlined. For example, the choice of using selected shapes to represent various security visual nodes requires consistency across all security visualizations. Finally, it is important that researchers, developers and users understand how this security visualization guideline delivers to enhance neutral and common knowledge when applying the guideline for security visualizations.

In addition, our SCeeVis guideline delivers various presentation methodologies with emphasis on real-time and interactive security visualizations. Therefore, the use of colours,

shapes, and a manageable presentation space for visualization, is important to maintain effectiveness for security visualization. The ideal method involves using correct colours, shapes and relationship link attributes in a security visualization. Thus, it is important that researchers and developers understand the nature of the security incident while designing the visualization and adopting the guideline stated in this standard. Overall, the concept of providing an effective security visualization requires simplicity and clear visual presentation as appropriate and outlined in our SvEm algorithm. This leads to allowing security researchers and visualization developers to understanding the links between the security incident landscapes, entities and relationships.

6.3.4 Security Incident Landscapes, Entities, and Relationships

Security incidents are the derivation of security visualizations. Thus, a security visualization is effective and appealing when portraying useful interactive security information from the incident captured. Hence, researchers and developers need to know and understand the incident landscape, entities involved and the relationships between entities (e.g., malicious IP address, payloads, etc.), which describes the story of the security incident. Entities (En), relationships (EnR) and security landscapes (SL) are core components of our framework. Therefore, using this knowledge, a security visualization can be effectively designed and presented to the targeted audiences. We provide details for what each component means when applied in the SCeeVis guideline and in security visualizations.

Security Incident Landscapes:

Security landscapes (SL) provide the incident scope and environment for users to perceive clearly within the security incident visualized. A familiar SL enhances a user to intuitively establish a conceptual boundary that enables them to confront a visualization with confidence. Landscapes vary according to the type of security incident captured thus, contributing to the need for effective visual presentations.

Entities:

Entities (En) refers to the following: threat actors, malicious payloads, infected IP address and more. These entities contribute to the SvEm effectiveness measurement theory. For example, Figure 6.6 shows four different entity (Network nodes) presentations with the use of shapes and colours. Identifying these entities in visualization with the least amount of observation time affects the performance of our framework.

(a) e.g., Malware payload and suspicious IP address. **(b)** e.g., legitimate IP address and bitcoin node

Figure 6.6: A sample presentation of Entities for Security Visualization

Entity Relationships:

Entity relationships (EnR), also called links, are vital for our framework. EnR functions connect entities together, show relationships between entities, show directions of interactions and are also used to activate user-cognitive functionalities. This allows users to perceive hidden information, links and potential security insights. The SCeeVis guideline uses basic relationship representation, i.e., 'solid and dotted (dash)' lines to represent entity relationships. We added colours into various types of relationships to show the effect these entities are portraying in security visualizations. For example, a red line shows a malicious connection/link and a red dotted (dash) line shows interaction relationships and links. However, when changing the solid/dotted lines and making them solid/dotted lines, it shows the direction of the relationships and interactions. Finally, relationship representations in the SCeeVis are not restricted to straight solid/dotted lines, they can be curved solid/dotted lines as well. Such relationship representations reduce visual design and observation complexities.

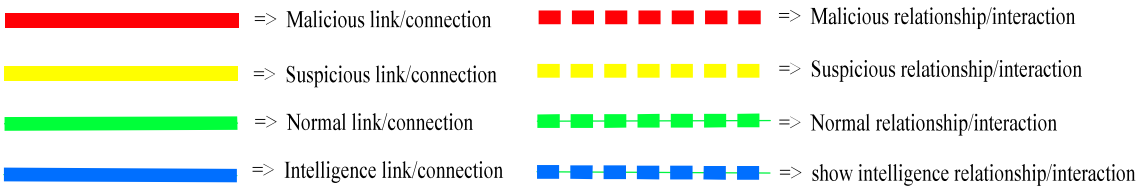


Figure 6.7: The sample Entity Relationships for Security Visualization

6.3.5 SCeeVis Application Use Cases

Security standards are technical documentations with no effects unless applied practically in information technology systems. Our SCeeVis guideline is implemented through use cases described in Chapter 5 of this thesis. The application of SCeeVis are implemented in our visualization use cases, namely (1) VisualProgger, a real-time security visualization application, (2) locky ransomware visualization and (3) the real-time cyber security chal-

lenge augmented reality visualization. In addition, we implemented a centralised security visualization framework which includes a bitcoin visualization for intelligence tracking and monitoring and malware reporting visualization. These were developed for the law enforcement security domain.

The application of this guideline has been applied during all stages: (1) design, (2) implementation and (3) observation stages. Thus, providing this standard to the visualization developers and the users (viewers) help establish a common knowledge on what will be provided (i.e., the developers task) and what is provided through security visualization for the users (viewers) to see and observe in the security incidents. In all use cases mentioned, the use of colours, shapes, relationships and interactive features presented through visual entities (nodes) shows a simple method of applying such a standard to communicate critical information analysed from the data and presented to the users (viewers).

6.4 Evaluation of the (SCeeVis) Guideline

Although our SCeeVis guideline shows strength and advantages in the security visualizations discussed in earlier chapters, there is the need for evaluation and seeing how effective this standard is. We evaluate the application of our SCeeVis guideline against other known security standards and state the strengths and weaknesses of our SCeeVis Guideline.

6.4.1 The SCeeVis Colour Association Rules

The SCeeVis guideline has strong emphasis on the need to standardise the use of colours and shapes for basic security visualization implementation and presentations. It has strengths, challenges and limitations. Firstly, the range of colours available for visualization usage can be limited thus creating difficulties when attempting to represent multiple security entities and relationships. Hence, we created a set of rules to introduce additional colour usage and colour relationships when applied in security visualizations. The rules are derived from and linked to each of the SCeeVis colours outlined, namely red, yellow, green and blue. These set of rules are effective only when associated with other colours, and when applied in security visualizations. Thus, the idea involves maintaining the standard colours as roots of visual colour representation so that users can still find their way around exploring security visualizations from a known point identified through the security visualization presentation. We call this set of rules as the '*SCeeVis colour chaining standard.*' These rules are:

1. Red ⇒ only used for malicious representations
2. Red ⇒ Yellow or vice versa is used for representing malicious and suspicious entities
3. Red ⇒ any other colour is used for malicious entity association with either unknown entities or classification of large data clusters
4. Yellow ⇒ Green is used for suspicious entity associated with normal entities
5. Yellow ⇒ Blue is used for suspicious entity associated with intelligence tracking entities
6. Yellow ⇒ any color is used for suspicious entity associated for classifications of either suspicious entities or unknown entities
7. Green ⇒ Blue or vice versa is used for normal entities associated with Intelligence tracking and monitoring
8. Green ⇒ and any other colour (except Yellow or Red) for security representation, comparison, classification and categorisation
9. Blue ⇒ and any other colour (except Yellow or Red) is used for intelligence tracking, monitoring, classification and categorisation
10. Green ⇒ Yellow or Red is used for representing normal entities associated with suspected or malicious entities
11. Blue ⇒ Yellow or Red is used for representing intelligence tracking entities associated with suspected or malicious entities
12. Purple ⇒ only for law enforcement use: for contents of trafficking (drugs, animal artifacts, etc.)
13. Orange ⇒ only for law enforcement use: for contents of fraud (currencies, account details, etc.)
14. Any other colour ⇒ other colours (**not**: red/yellow/green/blue/purple/orange), while it is not encouraged, we allow room for expansion. Thus, we use this option to represent either 'unknown' entities, clusters of entities or show purely classifications or grouping extracted from large datasets.

6.4.2 SCeeVis Challenges and Limitations

Our SCeeVis guideline provides a way forward for security researchers, visualization developers and users when designing and implementing security visualizations. As described in the earlier sections of this chapter, this standard addresses a full-scale effectiveness performance and assessment throughout the designs, implementation and observation (DIO) phases of our SvEm security visualization framework. It has leverage on existing security standards, taxonomies, best practices and guidelines to build a new security visualization standard. While there are strengths and advantages of this standard, there are challenges and limitations encountered when deriving this guideline and when applying it in security visualizations. We evaluate and discuss these challenges and limitations in the following way:

1. **Challenge-1:** ***Colour range limitations***: The choice of colours in visualizations has limitations due to the available range of colours. There are not many original sets of colours to use for visualizations. This makes visualization presentation a challenge when the datasets are large, and all security entities and attributes present in the visual space need to be provided.

2. **Challenge-2:** ***Colour overlapping issues***: A challenge encountered by developers is the ability to use colours comfortably when developing security visualizations. However, the similarities among various colours limits the range in terms of which colours can be used to differentiate entities. Examples of colour similarities are: (1) red vs orange, (2) yellow vs orange, (3) purple vs red/pink, (4) red vs pink and more. Thus, offering a security visualization that contains these colours in a visual space will automatically create colour overlapping issues. This also creates confusion for users (viewers), especially those that have vision issues and are not able to distinguish between such overlapping colours.

3. **Challenge-3:** ***Colour preference by users***: Users (viewers) of visualizations have preferences around colours used in visual images. It is often a case whereby users establish perception in an area they are comfortable with and also of known working memory load. However, that is a challenge for visualization developers when attempting visualization designs to please all targeted audiences. Users naturally decide and prefer certain colours over others, which can be a demotivating factor in increasing user attention span when observing security visualizations.

4. **Challenge-4:** ***Gestalt principles/laws and criticisms***: As part of our SCeeVis standard we adopted and incorporated Gestalt principles/laws concepts as a forms of practical

guidelines and situation awareness for developers and users of security visualizations. The concept of perceiving objects/entities as a whole and in groups is realistic; however, past literatures have criticised the concepts due to not enough user-testing and evaluation of these principles.

5. **Challenge-5:** ***Organisation Specific Standards***: Finally, a challenge stopping the full utilisation of our SCeeVis standard is related to organisation regulations and policies around which standard to adopt and use. This is due to demands from organisations to use certain specific standards that meet their requirements regardless of how effective that particular standard is. It restricts the potential of attempting to test and apply other standards such as our SCeeVis guideline.

6.5 Summary

In summary, this chapter delivers our Security Visualization Guideline (SCeeVis standard). Our Guideline is built on existing visualization designs and implementation concepts. These concepts are addressed towards visualization researchers and developers; however, the SCeeVis guideline adds another layer into the visualization designs and implementations to satisfy the users (viewers) of security visualizations. We provided a full-scale security visualization standard with emphasis around colour usage, user-trigger interactive features and delivered a set of our 'SvEm standard colour chaining' rules as part of the SCeeVis guideline colour association rules. These association rules serve the purpose of activating user cognition with the establishment of guideline colour chaining to enhance the users when traversing through a security visualization to make sense out of it. When applying the SvEm colour chaining rules, it establishes a pattern recognition method to trace known entities and relationships through the use of chaining (linking) with our standard colours, namely red, yellow, green, blue and black. This method helps establish a point of reference (comfort zone) for users to begin exploring security knowledge from the security visualization provided.

As mentioned, our SCeeVis guideline adopted designs and implementation concepts from existing security standards, taxonomies, best practices and guidelines outlined in past literature. We built on design and implementation research gaps which are identified as core user-centric features needed for a security visualization standard. These gaps aid the establishment of our security visualization to address simplicity and effectiveness in security visualizations. The introduction of a standardised colour system, adopted from Interpol's notice system and translated for security purposes, is the primary part of our standard. The introduction of standardised set of shapes and lines to represent entities and their relationships

is the second component of our standard. Finally, the introduction of our SCeeVis standard colour association rules is the third component of our Security Visualization (SCeeVis) guideline.

However, implementing security standards does have challenges and limitations that contribute to several evaluation factors. These are associated with the developer's designs and implementation process, plus the users of security visualizations, which are the end-users and organisations who operate on certain standard requirements. Challenges encountered in security visualization basically revolve around the use of this standard and the selection of components that make up the standard. Firstly, the standard colours mentioned can be a challenge if users confront the security visualizations with expectations, preferences and the ability to accumulate cognitive biases based on preferences. As a result, this standard will struggle to guide users to achieve potential security insights. Secondly, users and organizations have preferences and regulations that help them decide and interact with certain standards. They often have expectations that might demotivate them when attempting to apply our SCeeVis guideline to their needs. Finally, evaluating this security visualization guideline provides room for further improvement with added content around security and user guidelines.

Thus, we conclude that in the security domain, there are few sets of standards. The ISO/IEC 27000 family series provides guidelines for information security in business management systems. However, the security visualization field lacks standards to guide and help measure effectiveness in security visualizations offered. We provided our SCeeVis guideline with two primary emphases: (1) the introduction of standard colours, shapes, relationship links with an interactive user-trigger features guide, and (2) providing a set of standard colour association rules to enhance users' with pattern recognition and activate user's cognition point-of-reference (comfort zone) for users to establish their observation of the security visualization presented.

Chapter 7

Analysis and Evaluations

This chapter covers the assessment and evaluation section of our entire security visualization framework against the proposed objective of providing a full-scale effectiveness measurement approach in security visualizations for cyber security operations. Our thesis objectives are to provide an effectiveness approach for security visualization in the following stages: (1) design stage, (2) implementation stage and (3) user observation stage. With these objectives, we provided several interactive visualization designs which are built upon the need to visualize security incidents in limited spaces provided such as in mobile platforms. We also introduced our SvEm algorithm which incorporates the data used, visualization presented and user perception concepts. This algorithm provided two effectiveness outcomes: (1) measurement of distortion rates in security visualization and (2) the least time measurement required to acquire security knowledge. Furthermore, this thesis provided our security visualization standard (SCeeVis standard) as part of measuring effectiveness in security visualizations. Thus, we will assess and evaluate each effectiveness approaches stated to understand how effective this proposed framework can perform in security visualizations for web and mobile platforms.

7.1 SvEm Algorithm Evaluation

We begin with evaluating our SvEm algorithm. This algorithm provides a theoretical method of calculating effectiveness in security visualizations. It differentiates itself from other methods by not only addressing effectiveness in the design and implementation phases (stages) but also during the visualization observation phase. It is a unique full-scale security visualization effectiveness measurement approach. It begins with assessing and preparing the data processed which involves creating predefined data into visual nodes. Secondly, it associates the amount of data needed to be visualized against the available visual space. Thirdly, it delivers the implementation of security visual nodes in terms of representation and pre-

sentation methods. Finally, it delivers user-trigger features to activate user cognition and observes the viewers perception environment thus, ensures effectiveness in the security visualization presented.

7.1.1 SvEm Algorithm Variables Evaluation

We opt in to analyse and assess each algorithm variables as the SvEm algorithm evaluation strategy. This has allowed us to closely analyse how it is applied in security visualizations. Reiterating on our SvEm algorithm seen in Equations 7.1.1, 7.2 and 7.3, we preview the SvEm algorithm and all variables as follows:

(SvEm) Distortion (d_{svem}) Theory Assessment

$$\mathrm{SV}_{val} = \frac{(w * h)}{\mathrm{S}v_f * d_n} \triangleright (\mathrm{Sv}_f * d_n) \neq 0 \tag{7.1}$$

$$\mathrm{S}v\mathrm{E}m = \frac{\mathrm{SV}_{val}}{(\mathrm{C}l * n_{clicks})/t_{me}} > 50\%(\mathrm{D}istortion) \triangleright ((\mathrm{Cl} * \mathrm{n}_{clicks})/t_{me}) \neq 0 \tag{7.2}$$

(SvEm) Time (t_{svem}) Theory Assessment

$$\mathrm{S}v\mathrm{E}m = \frac{(\mathrm{C}l/t_{me})}{n_{clicks} * \mathrm{S}v_f/d_n} \leq 0.25sec(s)(\mathrm{T}ime) \tag{7.3}$$

Where:
*w * h : Web/Mobile display area (dimensions)*
$\mathrm{S}v_f$ *: Security visual nodes (e.g., Infected-IP, timestamps, etc.)*
d_n *: n-dimensional view in security visualization*
Cl : Cognitive load (Identifiable attributes (quantity) - Prior knowledge)
t_{me} *: Memory efficiency (Effort based on working memory - Time-base)*
n_{clicks} *: Number-of-clicks on visualization*

We analyse each variable and understand its existence as part of our SvEm algorithm and hence understand its roles in achieving effectiveness measurement in security visualizations. The variables are described accordingly, including:

1. *$w * h$: Web/Mobile display area (dimensions)*: A visualization display area (space) plays a vital role in our theory. It affects how information is presented. The ability to view security visualizations on different screen sizes creates a need for an effectiveness measurement consistency rating. For example, a 52-inch plasma screen with high resolution delivers totally different visualization clarity and user experience compared to a 4.7-inch smart phone. Such screen size range makes it difficult to obtain a consistent effectiveness assessment rating. Hence, for consistency in assessing and measuring effectiveness in this research, we narrowed down our visualization display space and area to focus on small display screens. This includes small screen mobile platforms and laptops with 13 inches or less display dimensions.

   As observed in our security visualization use cases, effectiveness in security visualization has additional factors such as the number of visual nodes or objects (entities, variables) displayed in a given time, concurrent visual views in the same display space and rendering capabilities. In addition, the use of colours in these visualizations also affects how a visualization displays itself within the display space. Hence, we conclude that our web/mobile display area (w * h) variable primarily contributes to our effectiveness measurement achievement and it dictates the outcome of the security visualization presented.

2. Sv_f *: Security visual nodes (e.g., Infected-IP, timestamps, etc.)*: Similarly, security visual nodes (Sv_f) are critical in security visualization, as they drive the core existence and purpose of offering security visualizations and communicating security information across to the targeted audiences. Patterns and behaviours emerge from observing and analysing multiple links between interested visual nodes. Therefore, a user confronted with a security visualization has the potential to either use their memory efficiency to perceive security information or explore and learn from the visualization presented. Thus, by users identifying security visual nodes and perceiving them via their intuitive cognitive ability and strength enables them to acquire security knowledge.

   However, there are challenges in designing and representing predefined data attributes into practical security visual nodes that in reality resemble the security events. Hence, the concept of providing a security visualization standard such as our SCeeVis standard helps connect the dots between each security visual node and their meaning/representation. In addition, designing security visual nodes required the concept of simplicity, fitting into the visualization space, and it has to be as realistic as possible to aid user perception. Furthermore, the challenge resulting from the data size needing to be visualized concurrently elevates the challenge of creating multiple security

visual nodes to represent different data attributes. For example, visualizing a cluster of malicious attributes in a static visualization environment is deemed a visual clarity challenge compared to visualizing the same cluster of malicious attributes in a real-time/dynamic security visualization space. Hence providing an effective security visualization required that we implemented an effective backend infrastructure that manages proper and timely data processing.

3. d_n *: n-dimensional view in security visualization*: The challenge of having a fixed visualization space (w * h) for visualization due to restricted hardware dimensions triggered the need to introduce our n-dimensional view (d_n) variable as an effectiveness component. Hence, n-dimensional views in security visualizations provide various visual views to distinguish between different security events extracted from the data collected. In addition, the n-dimensional view variable is utilised to show granularities in visual processes, clusters and the possibility of identifying different system and network layers. The n-dimensional view concept is also designed to accommodate the growing size of data needed for visualization. This enables multiple visual perspectives, for example viewing certain cyber-attack landscapes from multiple dimensions.

4. *Cl : Cognitive load (Identifiable attributes (quantity) - Prior knowledge)*: The cognitive load variable is a vital component in our SvEm algorithm as it addresses the users effectiveness contribution. It is both a theoretical and practical component that relies on user observation and assessment to address effectiveness in our SvEm framework. A cognitive constant has been established to act as the minimum possible cognitive measurement that can exist to activate user cognition in human beings. In this framework, cognitive load has a limit where users are required to utilise their working memory to perceive interactively with a high effectiveness rating.

5. t_{me} *: Memory efficiency (Effort based on working memory - Time-base)*: Similarly, the memory efficiency variable co-exists with the cognitive load variable. It is a time-base measurement based on the user's effort when utilising their working memory load while processing the visual information presented in a security visualization. Thus, an ideal user experience with security visualization requires that memory efficiency would increase while cognitive load maintains a consistent reading.

6. n_{clicks} *: Number-of-clicks on visualization*: The number-of-clicks (n_{clicks}) variable is dependent on the visualization space (w * h) variable and the n-dimensional (d_n) view variable. Given a reasonable visualization space and that a 4-dimensional view performs and displays more visual nodes than visualizing on a 2-dimensional view secu-

rity visualization. Hence it affects the user's working memory load (t_{me}) time frame. Users tend to perceive better when interactive visual nodes are presented in high-dimensional views. Therefore, in such cases, the ideal output for the number-of-clicks in security visualizations should be the least recorded given the users has an active working memory load rate and high working memory load. In theory, a higher working memory load is a result of less number-of-clicks executed. However, we assessed user-clicks against zooming features and other user-trigger interactive features in both web and mobile user actions. Results have indicated that the number of clicks executed for the purpose of exploring the visualization purposes does not go in line with our intention of the concept of achieving minimal user-clicks. Therefore, the number-of-clicks (n_{clicks}) variable performs on a case-by-case visualization experience. For example, observing a security visualization for exploring purposes and observing a visualization for reporting purposes provide two different user-clicks outcomes. Secondly, the security visualization designs and types would affect the number-of-clicks variable in a visualization. Different security visualization presentation designs require certain number of clicks when observing security visualizations. Hence, comparing number of clicks in such situation require a set of approved visualization range to achieve better mean statistics in the number of clicks performed in a security visualization.

Overall, the success of our SvEm theorem and framework aims to achieve effectiveness in security visualization and enable an efficient measurement approach. This is achieved through all SvEm variables being interconnected and depending on each other. The existence of all SvEm variables co-exist in a way to provide a full-scale effectiveness measurement in security visualization.

7.1.2 The Cognitive Load and Memory Efficiency Calculation Evaluation

Assessments in software applications are performed through backend/frontend performance testing. However, from the application side linking to the users, it is a challenging task to evaluate performance for both the application and user linked together. This is often due to the assessment technique used to measure performance. However, our SvEm framework goes beyond the application and user perspectives. The primary focus is basically in an intermediate stage between the application and user (viewer), whereby user-trigger features in security visualizations and the cognitive/perception process in users are assessed to measure effectiveness. A major approach in this stage is to understand what users want and what they are looking for in security visualization. This is a challenging task; however, providing effective means to aid users and the SvEm framework is our approach, whereby users

are empowered to perceive and perform effectively in decision making. This is opposite to providing a security visualization to users with a direct intention of what a user would expect to see and extract from the visualization presented. Hence, the drive to empower and enhance the user's cognitive and working memory abilities contributed to effectiveness in the whole security visualization experience. As a result, users are comfortable in teaching themselves to understand security incidents in visualizations, thus improving their interactions and observations with the security visualization presented every time they are in such a visualization process.

However, the question of how cognitive and memory efficiency are measured is a current challenge. Past research in psychology has heavily invested in user studies and theoretical proofs [195], [196]. Operating less form a psychological approach, and more from a computing science perspective, we approached cognitive and working memory load with the concept of linking the user's perception with their cognition process and understanding the relationship between perception, cognition and the SvEm framework's user-centric features. These are executed when the user's mind has the ability to perceive and apprehend objects (e.g., security visual nodes) through sight with the use of visualization interactive features. Thus, this process allows users to think of key words relevant to the security visual nodes presented. The user's cognition process through the thinking process allows them to think and enhance their perception process and relate back to past/previous visualization experiences, hence evaluating the past objects perceived. This process is executed as a result of having a high memory working load/capacity in the security incident presented.

Based on the theoretical approach, we used past techniques to understand and calculate cognitive and working memory load in users. Hence, we assessed various simulated user experiments and got the following results, shown in Table C.1 (Appendix C: C3) and in Figure 7.1. This experiment has shown that there is consistency in the viewers' working memory and cognitive load performance: as working memory load increases, cognitive load also increased as well. However, in Figure 7.1, the lines of best-fit show that both cognitive and memory efficiency performance are linear and the cognitive load has shown a consistent load (capacity) limit. As a result, the cognitive load performance does not overlap working memory load performance in a user. In reality, this is the ideal situation a user (viewer) should be in when observing and analysing security visualizations.

Therefore, we set up an experiment to assess cognitive and working memory load and record user performance, with two sets of visualization experiences. This experiment is based on a dual-task methodology [197] with the assumption that the user's working memory will function effectively if users are supplied with our SCeeVis standard, SCeeVis colour chaining standard and SvEm-CWML instruction set. We provide a summary of the cognitive

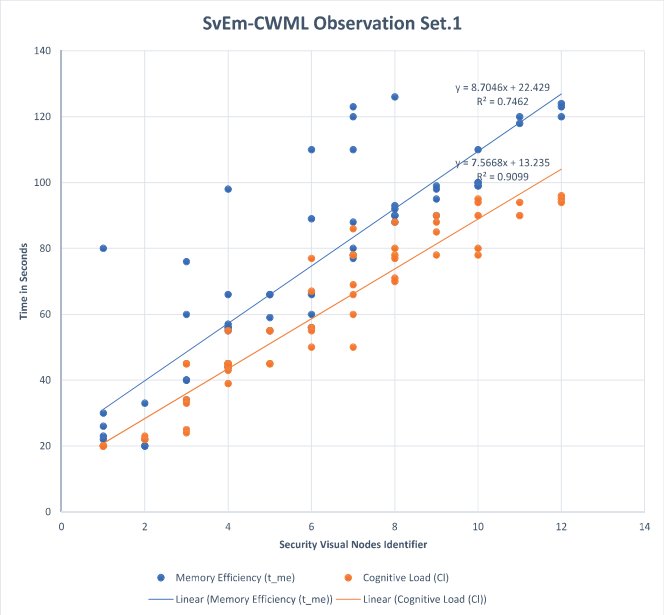


Figure 7.1: A Comparison Assessment of Cognitive and Working Memory Load in Viewers

and working memory load assessment by tasks executions. This is shown in Figure 7.2, and identifying how each entity (user, the security event, and SvEm framework) contributed towards the dual-task assessment process. The figure and results collected shows that users can effectively learn within the least time required before confronting a security visualization. We conducted the experiment in the following manner, as indicated in Table 7.1:

Table 7.1: The Cognitive and Memory Efficiency Observation Experiment

SvEm-CWML Observation Calculation	
Experiment 1	**Experiment 2**
No standard	SCeeVis standard
No SvEm-CWML Instructions	SCeeVis colour chaining standard
—	SvEm-CWML instruction set

The first experiment set is seen as the training set, whereby users were not handed instructions on how to fully utilise their working memory load and cognitive load capacity. Although this experimental process is proven difficult by past researchers [197], we designed a simplified version of the experiment to assess cognitive and working memory load. In

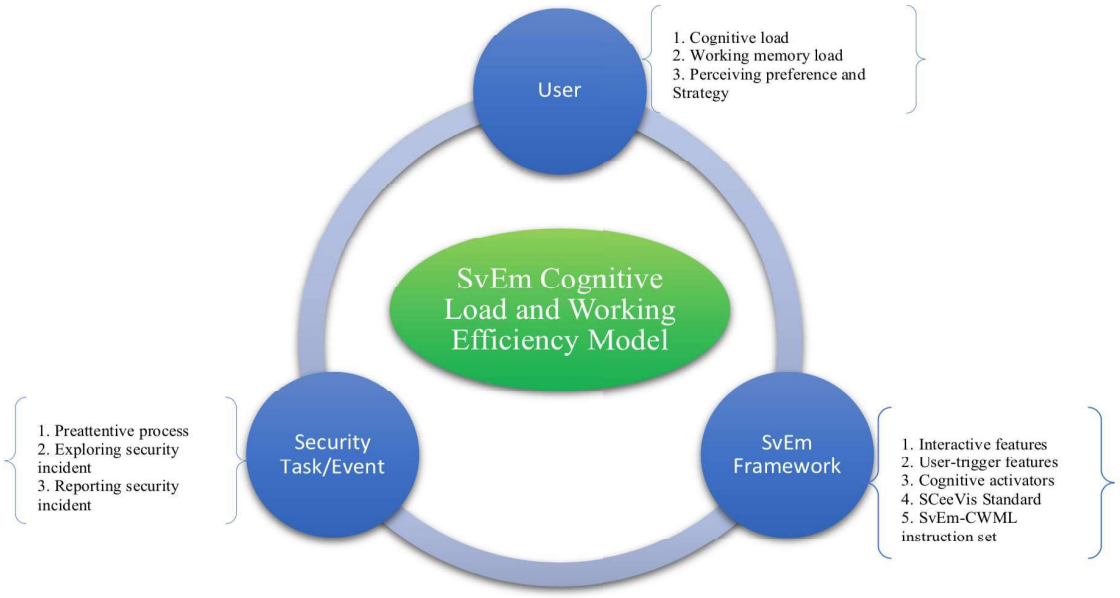


Figure 7.2: A Set of Tasks Executed for Cognitive and Memory Efficiency Load Measurement

Figure 7.3 samples of security visual nodes are identified from the two sets of security visualization presented. Both visualization samples are of different designs but maintain use of the SCeeVis colour standard. This helps users to identify interesting visual nodes faster and effectively, given they have a fair idea of what they are expected to visualize. In this example, predefined visual nodes are shaded in red, yellow, blue, green and other colours as well. Overall, the use of security visual nodes in our SvEm framework links known data attributes with the visualization to activate the user's cognition to an event while observing a security incident using the visualization.

In the first set of experiments, the number of visual recognition identifiers vary from 1 to 12 within the same set of visualizations offered. As observed, users (viewers) are asked to identify each visual node and explain what the visual nodes are by identifying the colour used for each visual node. The use of coloured visual nodes in the visualization also brings the ability to distinguish between distorted nodes from brighter nodes. This activates the user's cognitive load to further assess the visualization and respond with a record time duration.

This experiment is repeated in the second set but with changes made to it. The first step includes leveraging on using our SCeeVis standard guide with predefined security visual nodes, as visual recognition identifiers in security visualization trigger the user's cognitive load and increase the user's working memory load. The second step includes giving users

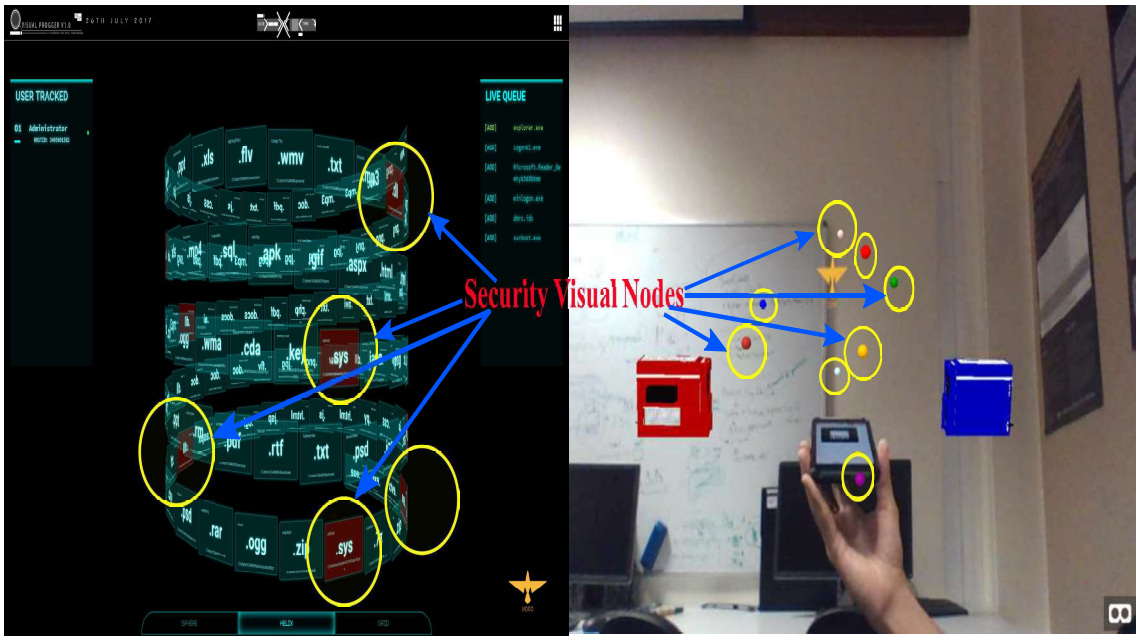


Figure 7.3: Visual Recognition Identifiers - Security Visual Nodes

an SvEm-CWML instruction set to enhance the user's ability to use their cognitive load and working memory load effectively to process visual nodes when observing security visualizations. Hence the number of security visual nodes identified is recorded against a time duration. This is used as a measure of cognitive load and working memory load. Thus, the number of visual nodes identified is seen as a memory efficiency (working memory load) identifier and measurement.

Hence, the user's cognitive and working memory load requires a practical set of assessment and measurement techniques to aid viewers and, at the same time, measure effectiveness in the entire visualization observation process. While this is not yet fully proven practically for security visualization, our theoretical approach enhances and measures the user's cognitive and working memory load by facilitating a '*security cognitive and working memory load instruction set.*' We called it the '*SvEm visual cognitive and working memory load (SvEm-CWML) instruction set.* This instruction set's co-function is to provide a mental set of procedures/guides for the users to use when confronting a security visualization. It should store and contain the following instructions:

START

1. Instruction Set 1 - Cognitive and working memory preprocessing stage:
    a) Step-1: Establishment of the SCeeVis standard in the user's mind
    b) Step-2: Recall, read, understand and store the colours, shapes and relationships codes (SCeeVis standard)
    c) Step-3: Recall the SCeeVis colour chaining standard
    d) Step-4: Identify and establish a 'visual recognition (point-of-presence)' point (this is done by identifying known security attributes and knowledge (e.g., security visual node/nodes) from the visualization presented).
2. Instruction Set 2 - Perceiving stage:
    a) Step-1: Explore with the use of the SCeeVis colour chaining standard in all direction.
    b) Step-2: Identify, group and classify security visual nodes.
    c) Step-3: Connect the dots between visual nodes and patterns.
    d) Step-4: Identify visual behaviours using patterns identified in the presented security incident
    e) Step-5: Mentally reassure the participant of the perception process.
3. Decision making process on final security knowledge is extracted
4. Conclude instruction set or revert back to Instruction Set 1 and start again.

END

The instruction set provides security experts the means to assess and measure the user's cognitive load and memory efficiency (working memory load) effectively. It is also a measure of how users visually observe security visualizations when no instructions are provided against the second set of experiment when instructions are provided to assist the visual process. However, there are challenges in these experimental sets around fully assessing various users with different visualization preferences and levels of understanding in security.

7.2 Cognitive Load and Working Memory Constraints and Limitations

The idea of investigating the user's knowledge and intuition ability to understand effectiveness in visualization processes had driven our research to focus on human cognition and memory efficiency. These two components affect how effectiveness is measured in security

visualization. However, evaluating and understanding what makes security visualization effective for users is a challenging task for security researchers. While there are potential methods used to link up security visualization and users, there are challenges as well.

Hence, human cognition (cognitive load) and memory efficiency (working memory load) execute constraints and limitations when assessing and measuring them. These limitations are often associated with the challenge of obtaining consistency in the experiments. Experimental entities (users) possess either similar or different preferences and concepts. This makes the cognitive and memory efficiency assessment difficult. Other constraints involve the state the user (viewer) is in, i.e., do they have prior knowledge of security related events, are they aware of how to use security visualization to explore security incidents, and are they motivated to use security visualization, (e.g., is the incident presented considered important to them)? Psychologically these questions affect the user's cognitive ability during a visualization observation process, thus affecting how cognition and memory efficiency are assessed. Hence, cognitive load theories have found that cognitive efficiency is constrained to limitations in working memory capacity to process and store information simultaneously [4].

Furthermore, security visualization is a new situational awareness approach for users in cyber security operations. Hence the concept of utilising the user's cognitive capacity and working memory load for specific visualization purposes is a challenging experience. It requires a training stage for the users of the SvEm security visualization framework to understand how human cognition and memory efficiency impacts effectiveness in security visualizations. This training process is challenging because of viewers' own distinctive visualization preferences and divergent methods of confronting, exploring and perceiving visual information. As a result, human cognitive load and working memory load affects effectiveness measurement in security visualizations.

7.2.1 Evaluating User Observation and Assessment

The ability of users to fully utilise human cognition and memory efficiency through the concept of working memory load or prior knowledge led us to evaluating the higher overview of the user's observation process and environment. As described earlier in this chapter, users rely on their cognition and perception capabilities to effectively confront, interact/observe and learn from the security visualization presented. Both cognition and memory efficiency are measured in time (seconds). Hence, effectiveness is not directly measured in users or in the security visualization presented but through the user's observation process and through the interactive user-trigger features presented in security visualizations.

A deeper understanding of how effectiveness is observed and measured required us to understand the entire visualization environment. The user's cognitive preprocessing and preattentive data processing realm allows users to naturally generate ideas to enhance their perceptual and visual experience. In this processing realm, users mentally perceive and sketch out ideas, links and, relationships, and generate patterns and behaviours from the security incident visualized. Thus, connecting the dots through this process helps share insights and overall brings motivation and comfort for users to interact further with the security visualization. This entire visual experience attracts users into interacting and exploring visual information thus removing and reducing cognitive biases in users.

The entire visual experience impacts both the users and the visualization presented. Hence, reiterating from the preattentive process seen in Figure 4.10, users generally execute several stages consistently and sequentially during their observation and exploring process. These stages include: generating ideas, expanding creativity with working memory load and mentally forming or building visual information. The stage of generating ideas links to user interactions with the security visualization presented, and it requires the user to perform interaction events such as mouse-clicks and, zooming, to enhance their perception to be creative and generate ideas. Hence, user-trigger features such as 'semi-permanent hold' allows the visualization to provide critical security information to users. This process activates the user's cognition and working memory load to expand on creativity, thus further exploring and gaining security insights.

Our user assessment is as vital as the SvEm security visualization framework. The assessment indicated that for a visualization to be effective, memory efficiency is crucial for a higher performance reading. User responses were observed and used to measure effectiveness in our framework. The SCeeVis colour standard contributed to the user performance by enabling users to process patterns and behaviours faster through classification and tracking/tracing of data relationships and links. In return, users understand the colour standards (guides), and approaching the visualization enabled them to process and connect the dots effectively through known security variables presented in the given visualization.

7.2.2 Evaluating SvEm's User-centric Features

The SvEm security visualizations are designed with user-centric features and capabilities to enhance the user's visualization environment. Thus, these features include user-trigger options, colour standards and guidelines. Although these features aid users, there are challenges and limitations. For example, user trigger events are executed based on the data filtered and processed, i.e., a malicious variable, attribute or file. The implementation of

SCeeVis colour guideline and pre-standard does provide and enhances the users with their visualization observation. However, from a design perspective, the range of colours available does pose a limitation when multiple data nodes and, attributes are needed to be represented concurrently in the same visual space (visualization). This creates confusion, which often leads to cognitive biases. As a result of cognitive biases, the effectiveness rating will be low. Hence, the ideal approach is to avoid pushing beyond the colour limitations.

In addition, the introduction of SvEm:cognitive-activators into our visualization framework creates a user-centric perspective of allowing users to be alert and watchful for security events. This automatically activates the viewer's cognition, thus motivating them to interact with the visualization presented.

In a limited visual space such as on mobile platforms, user-centric features are useful for interaction purposes. Our SvEm security visualization framework provides mouse-clicks, zooming and traversing visual views, which are useful for exploring and observing visual information. Users utilise these features effectively to interact, learn and communicate security information from the data presented, hence aiding them to make effective decisions. However, visual clarity is an issue when the data size required for visualization grows at a fast rate. Hence, representing all visual nodes in a limited space does cause visualization distortion. This is a challenge whereby effective data representation and presentation designs are implemented to manage high data volume, thus providing an ideal easy-to-view visualization.

7.3 SvEm Implementation and Application (uses)

Implementing our SvEm theory requires us to develop applications with specific features discussed earlier in this chapter to demonstrate our framework and approach. We incorporated all algorithms and designs into our SvEm framework and developed several use cases to show case our theorem. However, before performing our design, we surveyed popular existing visualization designs and picked the top four visual designs to use for our design benchmark. The designs chosen, after executing a literature survey and a user survey to find user preferences, are: (1) circular design, (2) treemap designs, (3) bubble charts, and (4) graphs. Although these four designs were ranked the most preferable and popular, effectiveness in security visualization does not only occur through the design phases in applications. There are more factors contributing to effectiveness in visualizations. These include how the security visualization application (SvEm framework) is used and what it is used for. Hence, the uses/purposes of each security visualization presented distinguish the security incident landscape and the urgency to present it through visualization. For example, a

real-time security visualization use case with VisualProgger serves the purpose of real-time tracking, monitoring and situation awareness. A ransomware visualization analysis serves the purpose of exploring and understanding the ransomware. Hence, the ransomware visualization allows the users to see and understand how it propagates through systems. In addition, augmented reality security visualizations are specific to capturing and providing critical security information in an interactive high-dimensional visualization environment. This is useful when data volumes grow rapidly thus removing the visual challenges from a 2-dimension visual plane and providing an augmented reality visual experience removes the stress-related cognitive biases in users.

7.4 SvEm Use Cases Evaluation

We evaluated our security visualization application use cases against several issues, misinterpretations, data presented, misunderstanding around the use of visualization and most importantly, the targeted audience. We begin with outlining two approaches that affect security visualization outputs:

1. When visualizations implemented are the work of designers with no background in security;

2. When visualizations implemented are the work of security professionals who do not understand data visualization

In these two security approaches, one visualization output is potentially attractive and colourful but not effective in transforming and communicating security information across to the targeted audiences. The other visualization output is potentially effective but often uncoordinated with visual complexities, thus leaving users (viewers) with confusion, which either demotivates the users or increases cognitive biases. Hence both approaches bring about misinterpretation and misunderstanding of the data presented. Such approaches also change the intention of providing visualization thus allowing users to interpret the visualization according to their own knowledge.

However, security visualization researchers can ask this question: '*What is it about visualizations that makes them an ideal method to solve big and critical security problems?* We evaluate our SvEm security visualization use cases to understand and evaluate this question.

7.5 VisualProgger Application Evaluation

The first use case developed to test our hypothesis is VisualProgger. It is a security application use case designed and developed with two security components: (1) a real-time security

visualization platform and (2) a visual analytic platform. The real-time security visualization platform was designed with the aim of simplifying data visualization by managing how predefined security data are represented in security visualizations. It provides a timely flow of real-time visual appearance with interactive features, which triggers the user's attention and activates their cognition to interact and explore security related events. As seen in Chapter 5, and particularly in Chapter 5.6, three visualization options are presented for the users (viewers) to select and interact with. These are the sphere visualization view (Figure 5.16), helix visualization view (Figure 5.17) and the grid visualization View (Figure 5.18).

7.5.1 Evaluation of Application User-Trigger Components

There are interactive user-centric features associated with these visualizations views. They serve a core purpose of providing security event alerts identified from the data visualized and also allowing users to interact with the visualization. These features shown and discussed in Chapter 4.7 and Chapter 4.7.1 are designed to link up the SvEm security visualization application with the users (viewers) by tapping into the user's perception environment from the moment the users confront the visualization. The three distinctive user-trigger features are: (1) 'semi-permanent hold' and (2) 'permanent-hold' and the (3) sound alert which brings up a critical alert splash screen to gain the user's attention. The 'critical' alert and 'semi-permanent hold' alert notifications shown in Figure 7.4 occur when malicious data is identified, processed and visualized. When triggered, a critical alert splash screen notification pops up with a sound and the SvEm application executes the malicious file (file is in red) identifier, namely the semi-permanent hold notification. This semi-permanent hold event is executed with a push-out event to distinguish itself from the normal visual nodes that are presented. These user-trigger features allow the users to interact by executing click, zoom and drag events on the selected files. With such capabilities in security visualization applications, users have the ability to explore, perceive and understand security events effectively. Hence, the user experience and interaction with such features enables effectiveness measurement from user observation studies.

However, there are extra user-centric visualization features mentioned earlier such as mouse-clicks, mouse-over labelling, zooming and toggling through several visualization details views. In Figure 7.5, file dependencies associated with the file of interest (e.g., malicious file) are visualized with expanded details and the contents of the files selected.

The concept of providing a 'file dependencies' visual view as seen in Figure 7.5a is to allow users to interact, generate ideas, increase creativity through their perception process, and understand the relationship between interested visual nodes (e.g., malicious attributes,

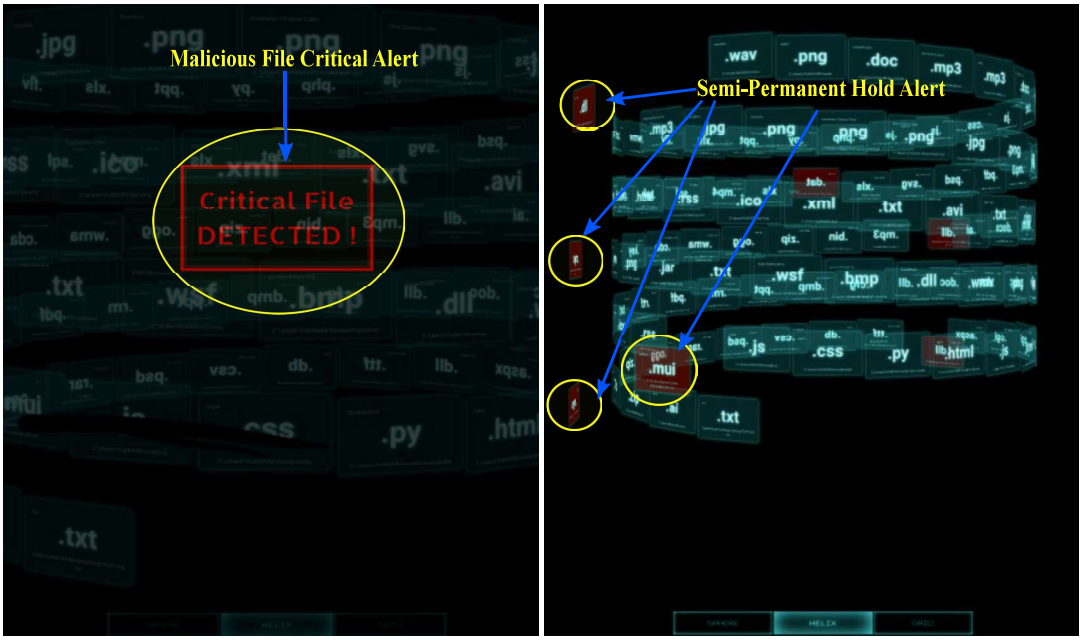
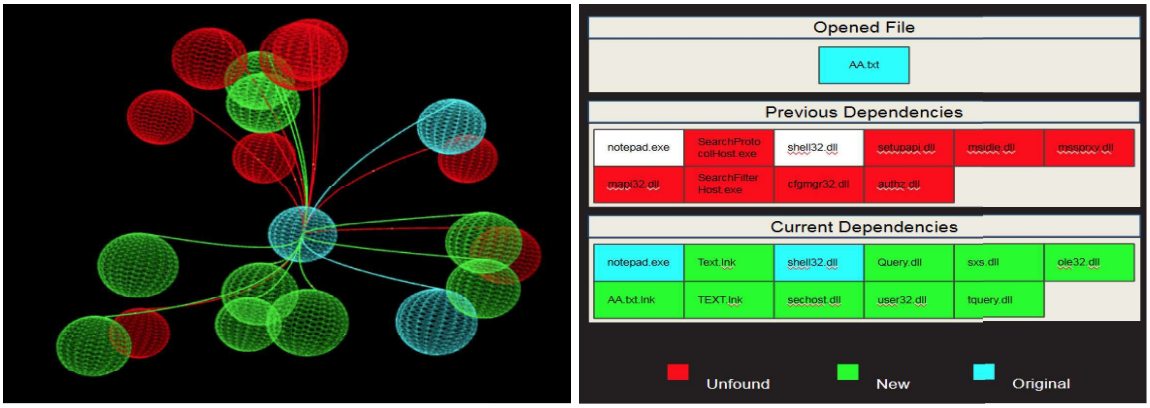


Figure 7.4: User-Trigger: (a) Critical Alert and (b) Semi-Permanent Hold Alert Notification



(a) A Visual View of File Dependencies **(b)** The File Dependencies Information

Figure 7.5: A File-Dependencies Visualization

files, etc.). In addition, the ability to provide 'file dependencies information,' as seen in Figure 7.5b, is an indicator of perception assurance for users. It enhances the user's perceptual preattentive processing operation while observing the visualization presented. Apart from these features, the ability to traverse through various visualization views and options with detailed information acts as a guide to the user while interacting with the visualization. Hence, with such simple features and multiple visualization options, users are motivated and driven to explore the SvEm visualization framework with VisualProgger freely and comfortably.

(a) The File Information **(b)** The File Dependencies Information

Figure 7.6: A Visual View of File with Detailed Information

Furthermore, a mouse-click event on a file of interest in VisualProgger will bring up the file details. As seen in Figure 7.6, file details/information (Figure 7.6a) are shown, such as file name, file type and file path. Additional file dependencies (Figure 7.6b) are also shown, such as system calls (write, read, open, delete), the date, and time of the file action. These features and detail options are useful for various reasons. For example, utilising VisualProgger for intelligence, provenance tracking, and attribution purposes provides the user the ability to trace, monitor and store visual information related to a security incident. These visual nodes contain details/information that are vital to gaining evidence and providing precise decisions when required.

The second VisualProgger component involves the visual analytic platform, which basically utilises both real-time and static data for visual analytic processes. This component targets the users who are concerned about the identified security incident and have the time to analyse the security incident with VisualProgger. It is designed to show patterns, behaviours, provenance and attribution. It uses data stored, based on user queries then processes and transforms predefined security data into useful and meaningful visualization. For example, the ransomware visualization presented in Chapter 5 and Chapter 5.6.3 is utilised by the visual analytic platform to help users to understand how locky ransomware affects systems. With VisualProgger, the interactive and user-centric features provide an effective analytical environment for users to comfortably observe the visualization presented. This visualization provides a deeper view of a computer system compromised by locky ransomware in three stages: (1) it executes a reconnaissance process then (2) encrypts the targeted files in the system, and finally (3) presents the encrypted files and payloads in red to show the level of infection. Hence, with VisualProgger, we are able to understand how ransomware attacks are executed from the point when systems are compromised and, infected,

through to the file encryption stage.

Overall, these VisualProgger capabilities contribute to effectiveness in security visualizations by enhancing the user's ability to interact, investigate and communicate security information extracted from the visualization presented. This enables the user to make sense out of and therefore aid the users in their decision-making process. Thus, evaluating effectiveness measurement in security visualizations relies on the way VisualProgger delivers and communicates predefined data representations through to the visualization presentation and observation stage. This makes the entire process a full-scale security visualization effectiveness approach.

7.6 Security Visualization with Augmented Reality Experience Evaluation

Apart from VisualProgger, the concept of exploring security visualization potentials in the augmented space emerged for four reasons: (1) it is mobile-centric and user-centric, (2) it is an SvEm effectiveness visualization insight-driven component, (3) it provides an interactive security visualization experience for the targeted audiences, and (4) there is a need to effectively enhance security events in a real-time environment which taps into the viewers cognitive and perceptual realm. We provide a proof-of-concept for our augmented reality security visualization shown in Figure 7.7.

Therefore, adding the concept and experience of augmented reality to both security analytics and visualization gives an added advantage and changes how users perceive security information. It gives an interactive capability to our SvEm framework, thus allowing the concept of n-dimensional view experiences to move to another level while utilising security visualization. Overall, it contributes with an effectiveness measurement approach in security visualization by actively activating the user's cognitive realm. In theory, this approach creates a comfortable environment for users (viewers) to observe security incidents, thus allowing them to explore further patterns and behaviours portrayed through the use of security visual nodes. This increases the user's attention span to concentrate on the visualization presented.

The use of our SCeeVis colour standard and instruction sets provided a procedural method of how to interact with security visualizations in a more meaningful and serious manner. Thus, establishing this environment through the use of augmented reality gains the user's attention, concentration and interest to interact with security data through the use of visualizations. However, there were challenges and limitations found when we explored augmented reality for security visualization purposes. Firstly, the concept of offering a mobile-

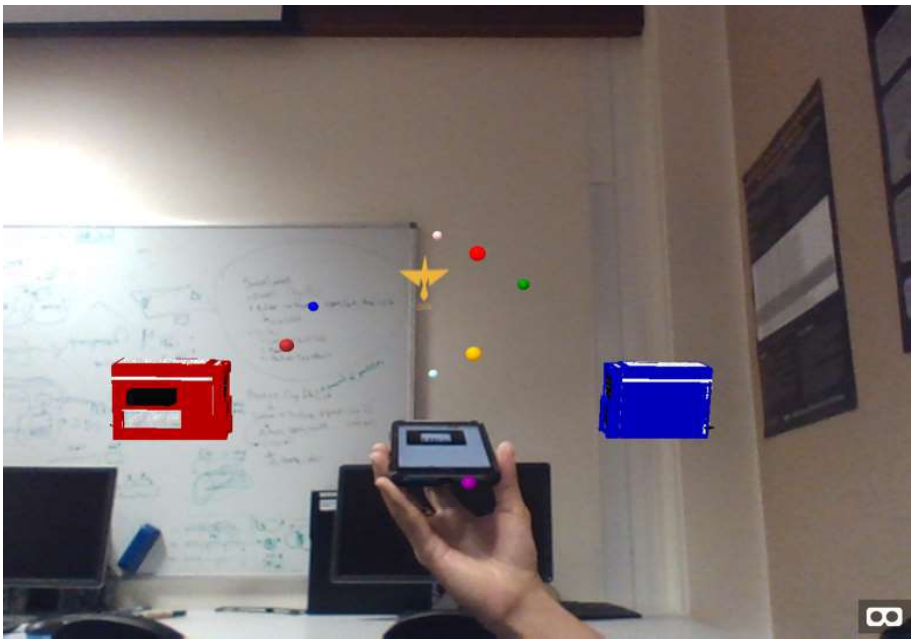


Figure 7.7: Security Visualization with Augmented Reality Experience

centric security visualization is controlled by the 'marker' and 'camera' capabilities. Hence, it can be limited to the distance (range) compatible with the camera's reach and the visibility of the marker used as contact point for the augmented reality visualization. Secondly, utilising the notion of using multiple markers concurrently for augmented reality visualization has proven a challenging task, especially around managing the data profiles from one marker to the other in real-time. This is yet to be further implemented and tested with new methods. Although our augmented reality visualization application is in its early stages of development, we explored the potentials of finding effectiveness approaches for security visualization. Hence, we concluded that there are useful effectiveness features provided by the augmented reality visualization domain.

7.7 SVInt: Bitcoin Explorer Security Visualization Evaluation

Another SvEm use case we implemented, is a security visualization (SVInt)for intelligence tracking and monitoring of bitcoin transactions. We refer to it as 'SVInt: bitcoin explorer security visualization. The purpose of this bitcoin visualization leverages on known data extracted from the bitcoin explorer. It serves with two purposes: (1) to track and monitor suspicious bitcoin transaction IDs and wallet IDs, as seen in Figure 7.8; and (2) to facilitate a centralised security visualization framework for information sharing in the law enforcement services.

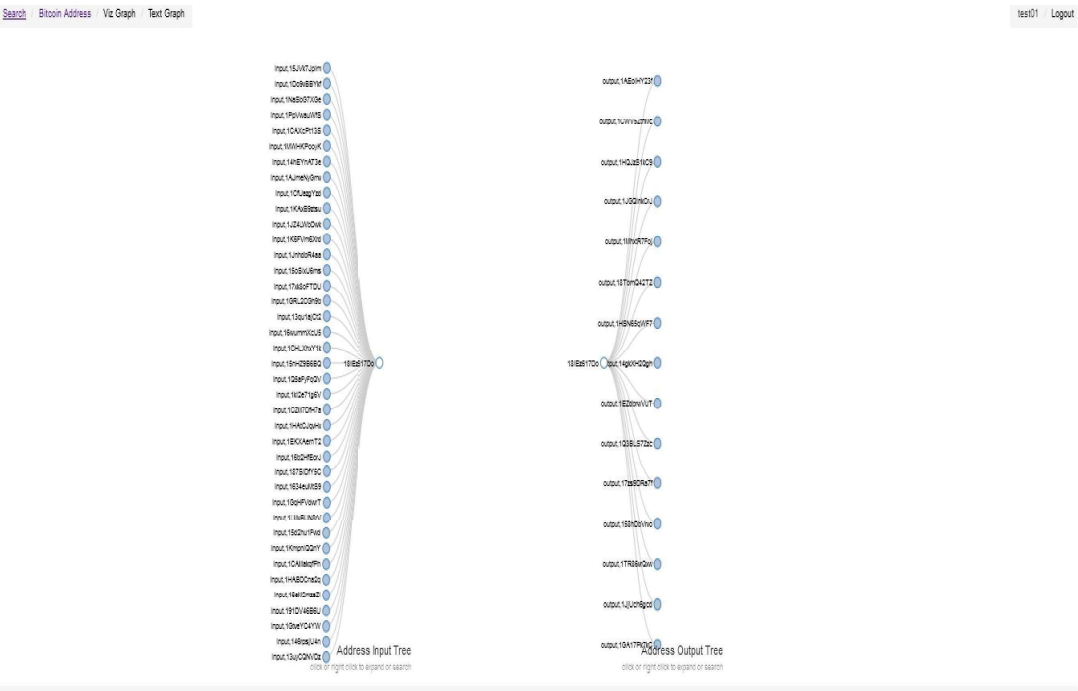


Figure 7.8: A Treemap of Bitcoin Transaction Addresses

7.8 SvEm Conceptual Model

The SvEm conceptual model provides researchers with an overview understanding of how our framework operates between the core entities. It is a full-scale effectiveness measurement approach with the following entities: the user, human cognition and security visualization. As discussed in Chapter 4, and particularly in Chapter 4.6.1, the three effectiveness contributing components are: (1) *user*, (2) *visualization* and (3) *user cognition*. With a clear understanding of these components and their requirements, security researchers and visualization developers can design, implement and communicate security ideas effectively from the data represented.

The inner SvEm model mechanism is where perception emerges as a result of incorporating user cognition and the visualization together. As a result of the user's observation, '*preattentive processing*' queries, occur during the time a user undergoes the perceiving process. In the event of confronting a security visualization, '*mental effort*' executed by the user depends on their cognitive capabilities, which involves their thinking process. The entire process builds the user's memory efficiency towards the security visualization presented. Taking into consideration the known links and relationships between users, user-cognition and the visualization, our effective-visualization techniques create the final relationship between the user and visualization presented. As a result, '*Distortion and/or Time*' are identi-

fied and measured. SvEm insights are triggered when all SvEm components are harmonised, thus mental effort reading is low; therefore, the right security information is transformed and translated for the user to perceive and process.

7.9 SvEm Performance Testing

Oue SvEm framework has executed performance testing in several SvEm areas of our framework. Hence, our security visualization applications were assessed for the following criteria: (1) visualization clarity and representation, (2) backend to frontend data transfer performance, and (3) SvEm:cognitive-activator presentation assessment. Visualization clarity and representation testing are executed during the application development stages through the use of visual observations. Our observations have shown that for a security visualization to be effective, there has to be some form of data and visual management mechanism established for a better user experience environment.

Therefore, the choice of using a WebGL visualization approach with a 3-dimensional visualization presentation view enables a new visual view for the users to interact with and experience the ability to process critical security information. Our application designs have allowed large volumes of data to be processed and presented in our security visualization frontend.

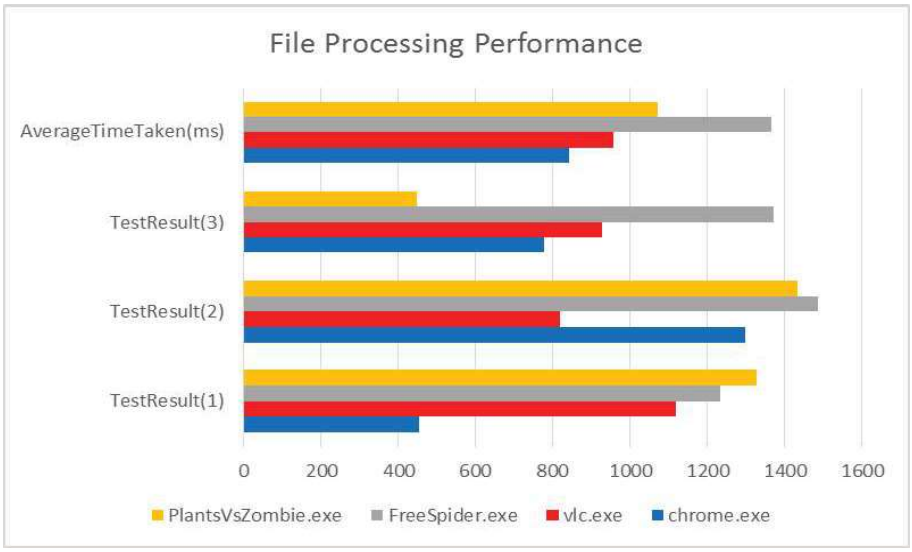


Figure 7.9: VisualProgger Application Performance Assessment

Hence, Figure 7.9 delivers the data transfer performance assessment and evaluation. In this example, we tested data transfer from our SvEm backend to the frontend with different executables (.exe) of various size and captured the data transfer time (ms) performance and

took an average time. This time is ideal for evaluating the time it takes to transform data attributes into predefined security visual nodes and push them respectively to the SvEm security visualization frontend.

7.10 Threat Scoring Detection System

As part of improving security effectiveness in security visualizations, our threat scoring detection system has to be accurate, with a high detection rate. This means proper malicious detection algorithms and systems are implemented for scanning and flagging malware and threats from the datasets obtained for visualization.

Therefore, a selection of known signature-base anomaly and malicious detection algorithms have been incorporated to filter collected datasets. Based on ground truth datasets, we initially selected *Local Outlier Factor (LOF)* [198], *DBscan* [199] and *K nearest neighbour (KNN)* [200] to see which satisfied our expected results. Thus, with several tests executed on the algorithms, we were able to verify our threat detection scoring system performance. For example, a normal action will be scored between 10-80, where as an anomaly behaviour will be scored with a negative value. Likewise, suspicious files in the systems are scanned against stored signature-based rules and databases. As seen in Figure 7.10, we illustrated normal versus abnormal behaviours and malicious records. In addition, scanning files within the system and having a preconfigured log history helps identify known file paths. Therefore, if a known or suspected file appears in another location, automatically this file is flagged through visualization with a yellow or red colour depending on the situation and stage of the file.

When evaluating our threat scoring system performance, ground truth datasets are used to test against the outputs. The known anomaly algorithms and malicious signatures are also used as filters to help classify or categorise files, attributes and entities visualized. Our scoring system is incorporated with Progger to flag out anomalies and malicious files within the system. As a result, 'files of interest' are visually represented in the following colour codes: malicious (red), suspicious (yellow), intelligence tracking (blue) and a normal legitimate data (green).

7.11 Summary

In summary, this chapter provides the analysis and evaluation of our key research contributions namely: (1) SvEm algorithm, (2) SvEm framework and our proposed (3) SCeeVis security visualization standard. We evaluated these key contributions with respect to achieving a full-scale effectiveness security visualization approach across the design, implementation

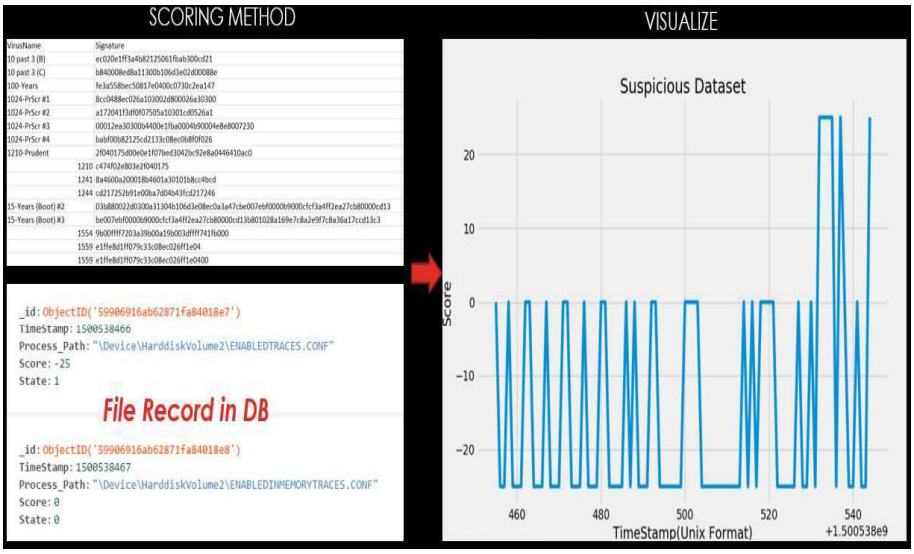


Figure 7.10: Anomaly Detection System Results

and user observation stages.

Hence, we identified that the effectiveness components between users, user cognition and the visualization presented are the connecting and linking components which include: (1) memory efficiency (working memory load/mental effort capacity), (2) cognitive activators, and (3) user-trigger and interactive features. Effectiveness is the result of these components mechanically harmonising together to achieve security insights, thus leaving the user to increase their attention span during the visualization observation period. To achieve effectiveness, we evaluated the relationships between user cognition and memory efficiency to prove a theoretical effectiveness approach. We evaluated that as working memory increases, cognitive load maintains consistency and is limited accordingly. This is an ideal stage whereby effectiveness in users is achieved. However, to improve this ideal state, we introduced a special security cognitive and working memory load instruction set to enhance standard users to trigger their cognitive load and perception. We call this instruction set 'SvEm visual cognitive and working memory load' (SvEm-CWML) instruction set. The result of the comparison between cognitive and memory efficiency in viewers is graphed and shown in Figure 7.1, whereby a linear result is achieved.

Furthermore, we evaluated our SvEm algorithm and SvEm framework to access all features provided, to aid and provide effectiveness in security visualization. The introduction of user-centric user-trigger features, SvEm cognitive interactive features and information presentation options through added visualization features has contributed to triggering effectiveness and motivation in users to interact and observe security visualizations for a longer

period of time, thus gaining insights. Finally, we executed several performance testing experiments to assess how data and visual nodes are represented, presented and managed during the entire visualization process.

Chapter 8

Conclusions and Future Work

8.1 Conclusion

The use of visualization in cyber security operations is proven effective for the purpose of exploring and reporting security events. It is widely used for security analytics and enhancing mitigation processes in cyber security investigations. Hence, the goals of this thesis is to investigate a range of security visualization effectiveness techniques and find effective user-centric methods to empower users with the appropriate means to understand cyber-attacks. However, empowering users to select an appropriate visualization sample (image) from a set of visualizations, let alone, encountering existing challenges to comfortably visualize and observing security information in small platform screens, require new 'effectiveness measurement' approaches. Reiterating the thesis problem statement of;

> Can users effectively visualize security events over their web and mobile platforms in a split-second to help them decide what the next secure step to execute is?
> If '*yes,*' how can we measure effectiveness in security visualization for web and mobile platforms?

with the hypothesis of;

> Security events can effectively be visualised on web platforms and on small 2D screens of mobile platforms.

Based on these challenges, thesis problem statement and hypothesis, this thesis explored and investigated a range of effectiveness measurement techniques for web and mobile plat-

forms. Effectiveness measurement techniques were designed, implemented and evaluated to provide effective user-centric security visualizations to empower users and aid decision making processes in cyber security operations.

8.1.1 Thesis Contributions Outline

This thesis presented a full-scale security visualization effectiveness measurement (SvEm) framework with techniques ranging from the technical designs and implementation, through to user-centric aspects. In order to achieve a full-scale effectiveness measurement in security visualization, we derived and delivered the following contributions:

1. **Dataset collection**: To achieve a full-scale security visualization measurement (SvEm) framework with techniques, datasets are the core component. A set of various data sources are utilised and approved for the purpose of collecting data for this thesis. Apart from testing our framework with public datasets, we focus on two data sources, namely:

    - *New Zealand cyber security challenge (NZCSC-2015, 2016, 2017) datasets*: Our dataset collection infrastructure is designed and incorporated as part of the University of Waikato's New Zealand national cyber security challenge (NZCSC). It is designed to collect a range of web logs, kernel logs and network logs. Hence, being part of a data collection team, the author of this thesis tailored our contribution to towards collecting attribution and provenance related logs which met the requirements for this thesis. Hence, our contribution is mainly the NZCSC Round-2 attack network data collection scenarios and the NZCSC-2017 data collection scenarios.
    - *Bitcoin dataset*: In addition, we utilise existing public based bitcoin data sources to collect our datasets. This dataset serves the purpose of intelligence tracking and monitoring of bitcoin transactions from and to bitcoin wallets. It is an approach specifically targeting law enforcement security visualization purposes, whereby the ability to track and monitor suspicious bitcoin transactions are the priority. Hence, to facilitate effectiveness, the entire data processing operation targets information sharing, data preservation and reporting on law enforcement operations.

2. **Security Visualization Effectiveness Measurement (SvEm) algorithm**: Our first contribution for this thesis is the design and development of our security visualization effectiveness measurement (SvEm) algorithm, which is delivered and discussed in Chapter 4 using Equation 4.2. With the challenge of understanding and selecting and mak-

ing a choice from two or more visualizations, our SvEm algorithm provides statistical measurement to aid the users decision making process. Effectiveness in this algorithm is measured using several key components namely: (1) visualization display dimensions, (2) the user's cognitive load and working memory load capacity, (3) n-dimensions of the visualization presentation view and (4) the number-of-clicks variable executed during the visualization observation process. Finally, to minimise error ratings and results, our SvEm variables are set to a default minimal numeric value of 1 (one), hence avoid the error of obtaining a denominator of 0 (zero) in our algorithm.

3. **SvEm conceptual model**: As part of achieving effectiveness in cyber security visualization, we have designed and constructed our SvEm conceptual model (Figure 4.9). This model is designed to incorporate the user's entire visualization experience with the security incident presented using visualization to achieve effectiveness in the user's environment. The conceptual model has three entities namely: (1) the *User*, (2) *Visualization* and the connecting variable (3) *User Cognition*.

4. **SvEm security visualization framework with use cases**: To test and evaluate our SvEm algorithm, we designed and delivered our second contribution which is the SvEm framework with 3 distinctive security visualization use cases. These use cases have been tested during the New Zealand cyber security challenge (NZCSC-2017) event. We achieved effectiveness through constructing three security visualization effectiveness use cases namely:

    - *Real-time VisualProgger*: A real-time effective security visualization approach whereby visualization distortion and clarity are the key focus in order to achieve effectiveness. Thus, backend data management process and frontend visual nodes representation management are the key areas of our real-time visualization presentation.
    - *Security visualization with augmented reality*: In addition, we provided an user-centric security visualization approach with an augmented reality visualization environment experience to empower and attract users to observe and interact effectively. Hence, providing an augment reality visualization environment increases the users cognitive and working memory capacity by allowing them to process visual information effectively.
    - *Security visualization for analytic and intelligence (tracking and monitoring)*: In this contribution, the targeted visualization audience are th law enforcement investigators. With primary goals of tracking and monitoring bitcoin transactions

and malware movement, we utilise security visualization by incorporating known and existing colour standards to resemble various security events.

5. **Security visualization (SCeeVis) guideline**: In order to obtain a full-scale security visualization effectiveness measurement approach with both the SvEm algorithm and the SvEm framework, we delivered our third contribution which is a security visualization (SCeeVis) standard. It is designed and serves as a guide to allow security visualization researchers, developers and users during the entire security visualization experience. This guideline is built on existing visualization techniques, security standards and best practices, however, the unique approach is making sure, the use of colours, shapes and relationships representation and presentation are standard and not complicated.

6. **SCeeVis colour chaining guideline**: To achieve a practical SCeeVis guideline approach, this thesis designed and delivered our additional SCeeVis colour chaining guideline contribution, which uses the standard colour association rules to help users connect with the SvEm framework effectively and navigate their way during the observation process. This design aids the user's entire visual analytic process to see, interact and understand security events effectively.

7. **SvEm cognitive and working memory load (SvEm-CWML) instruction set**: Similarly, our SvEm-CWML instruction set serves the user-side practical SCeeVis guideline approach. It ensures a timely and sequential security visual information process is executed by the users during their observation duration. Hence, our SvEm-CWML instruction set delivers simple basic instructions that user's can relate to easily and process effectively.

These contributions serve the purpose of achieving effectiveness measurement in the design, implementation and user observation/interaction stages of security visualization. Finally, we summarise our thesis contributions and state our future work in the remaining of this chapter.

8.1.2 Summary of Research Contributions

We began with designing and gathering the appropriate datasets required for our research study. Hence as part of the data gathering process we utilise logging mechanisms from the New Zealand cyber security challenge (NZCSC) event in 2015, 2016 and 2017. All datasets collected were anonymised and standardised to meet the ethical requirements and security visualization process requirements.

To achieve effectiveness in security visualization, we derived and introduced our SvEm security visualization algorithm which is the theoretical component of our research. This algorithm utilised the following components to accomplish effectiveness measurement: (1) web and mobile display dimensions, (2) security visual nodes, (3) an n-dimensional visualization view, (4) user cognitive load, (5) memory efficiency, and (6) the number of clicks executed when observing and interacting with the visualization presented. In this algorithm, specific key components are identified such as cognitive load, memory efficiency and the n-dimension visualization views that dictates effectiveness in security visualizations. We concentrated on these components to thoroughly understand how they are used to achieve the desired security visualization outcomes. Hence, our SvEm algorithm was tested and evaluated against certain user observation environment.

Furthermore, we provided a practical proof-of-concept (SvEm framework) for our algorithm, which allowed us to design, implement and evaluate all SvEm algorithm components. The design process includes implementing several design types that could fit easily in small display screens. These are circular visual designs, spheres, helixes, grid designs, treemaps and the augmented reality designs. We utilised these visualization concepts to allow a manageable data representation and presentation experience. Furthermore, the introduction of our distinctive SCeeVis colour and shape standard, alongside the execution of our SCeeVis colour chaining standard, allowed for visual nodes to be traced, tracked and monitored effectively. As a result, the user's perception environment is being enhanced, a comfortable viewing and observation process has been established allowing users to acquire relevant and critical security insights.

In conclusion, we achieved a full-scale security visualization effectiveness measurement approach and presentation by improving the way security information is presented and how users perceive it. We designed and created our SvEm algorithm to help theoretically understand, obtain and measure effectiveness in security visualizations. Utilising our SvEm framework and our introductory proposed SCeeVis guideline and pre-standard, we tested and evaluated our effectiveness algorithm. Our framework effectively contributed to increasing the users' attention while reducing the time spent on observing security visualization to gain insights and communicate security ideas across various targeted audiences. Finally, the ability to increase the user's attention results in a longer user attention span thus allowing them to interact, observe, generate ideas, and perceive security information through the use of security visualizations. Combining all effectiveness methods and approached throughout our security visualization framework provided a full-scale security visualization capability with primary focus captures the user's attention instantly thus reduce cognitive overload and keeping users from loosing concentration on the security visualiza-

tion presented. Finally, providing a full-scale effectiveness measurement approach in Cyber Security visualization enhances real-time information sharing, aids decision making and most importantly translates security insights to users of all different level of understanding in security events.

8.2 Future Work

8.2.1 Reducing the Security knowledge Gap between Security Visualizations and Specific Users

In future, the need to further understand the user's security visualization needs in relation to the rate at which security information are acquired and processed through visualization, would contribute to providing effective security visualizations in cyber security operations. This will help reduce the time spent during the entire decision making process.

8.2.2 Extending SvEm Framework Cross Domain Evaluation

In future, we would like to further evaluate our framework by conducting effectiveness measurements for users across different domains (health, finance education, etc.), carrying out further analysis on how users interact and respond. Extending this research work across other research domains would allow accessibility, information sharing and collaboration in security events such as malicious attacks.

8.2.3 Exploring User Response to Security Visualization Evaluation

Moreover, our SvEm-CWML instruction set and SCeeVis colour chaining guideline theory requires further tests against multiple types of complex visualizations to observe its performance. Our current implementation and test covers a small portion of simple visualizations for small screen display dimensions. Hence, extending this standard and evaluating user performance would justify the capabilities and usefulness for security visualizations.

8.2.4 Extending Mobile-centric Security Visualization with Augmented Reality

Finally, extending this research into the augmented reality space alongside potential machine learning approaches for security visualizations is an ideal scope to address. In addition, executing further research into innovative mobile-centric approaches to explore the concept of offering multiple markers and an augmented projection approach for mobile

platforms can enhance effectiveness measurement in security visualizations. This would allow users to utilise their mobile platforms to explore security information in an interactive environment with options of training themselves, thus allowing them to make effective decisions in the event of a malicious incident.

Appendix A

Visualization History Related Materials

A1: "Carte figurative des pertes successives en hommes de l'armée française dans la Campagne de Russie 1812-13 (comparées à celle d'Hannibal durant la 2èmee Guerre Punique) / Minard. Régnier & Dourdet. 1869."

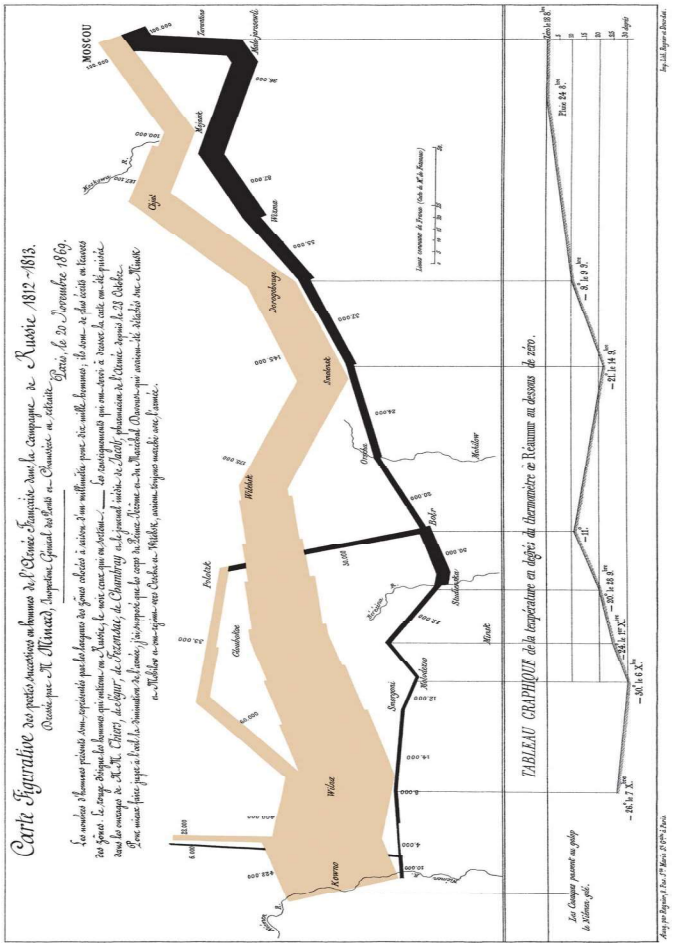


Figure A.1: Minard's carte figurative of Napoleon's 1812 campaign map

Appendix B

Data Collection and Ethics Materials

B.1 Data Collection Summary

The concept and idea behind any data collection process is to facilitate and enhance research testing environment. In the case of this thesis, our data collection process involves the Cyber Security Researchers of Waikato (CROW) team effort, comprising of both researchers, post graduate students and cyber security research programmers. With different research intentions, various logging mechanisms were designed and implemented during the New Zealand Cyber Security Challenge (NZCSC) events across the web, systems and networks to collect our datasets.

In the case for this thesis, our focus and contribution was on the web, kernel and network logs collected in the NZCSC events. The ideal approach taken for this thesis was based on attribution and provenance scenarios when collecting data. Hence, the more the types of logs collected during the NZCSC-2015, NZCSC-2016 and NZCSC-2017, the better it is for our security visualization use cases.

We acknowledged and credited the CROW team for their work and effort in the continuous annual data collection process which began in 2015.

The following attached ethical approval letters and participants agreement documents (B1, B2, B3, B4 and B5) are in order from 2015 to 2017:

1. B1: "New Zealand Cybersecurity Challenge 2015 (NZCSC2015) Data Collection Ethical Application Approval Letter"

2. B2: "New Zealand Cybersecurity Challenge 2016 (NZCSC2016) Data Collection Ethical Application Approval Letter"

3. B3: "New Zealand Cybersecurity Challenge 2017 (NZCSC2017) Data Collection Ethical Application Approval Letter"
4. B4: "New Zealand Cybersecurity Challenge 2017 (NZCSC2017) Participant's Agreement Form"

Faculty of Computing and Mathematical Sciences
Rorohiko me ngā Pūtaiao Pāngarau
The University of Waikato
Private Bag 3105
Hamilton
New Zealand

Phone +64 7 838 4322
www.fcms.waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

18 September 2015

Jeffery Garae
C/- Department of Computer Science
THE UNIVERSITY OF WAIKATO

Dear Jeffery

Application for approval under the Ethical Conduct in Human Research and Related Activities Regulations

I have considered your application to conduct interviews in New Zealand for your PhD research project entitled "Security Visualization for Mobile Development".

The procedure described in your request is acceptable. Participants involved in the study will not be identified in any resulting publications. At the conclusion of the research study the data will be stored in the FCMS Data Archive repository for five years.

The Participant Information Sheet and Research Consent Form comply with the requirements of the University's human research ethics policies and procedures.

I therefore approve your application to perform the research project.

Yours sincerely

Bernhard Pfahringer
Human Research Ethics Committee
Faculty of Computing and Mathematical Sciences

Faculty of Computing and Mathematical Sciences
Rorohiko me ngā Pūtaiao Pāngarau
The University of Waikato
Private Bag 3105
Hamilton
New Zealand

Phone +64 7 838 4322
www.cms.waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

22 May 2017

Jeffery Garae
C/- Department of Computer Science
THE UNIVERSITY OF WAIKATO

Dear Jeffery

Request for approval to conduct a user study with human participants

On the basis of the information you have provided on the FCMS Preliminary Ethics Application Form relating to your COMP 900-17C research "Cyber Security Challenge Data Collection", the Committee has given you approval to proceed with your proposed study.

We wish you well with your research.

Yours sincerely

Mike Mayo
Human Research Ethics Committee
Faculty of Computing and Mathematical Sciences

Faculty of Computing and Mathematical Sciences
Rorohiko me ngā Pūtaiao Pāngarau
The University of Waikato
Private Bag 3105
Hamilton
New Zealand

Phone +64 7 838 4322
www.cms.waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

22 May 2017

Jeffery Garae
C/- Department of Computer Science
THE UNIVERSITY OF WAIKATO

Dear Jeffery

Request for approval to conduct a user study with human participants

On the basis of the information you have provided on the FCMS Preliminary Ethics Application Form relating to your COMP 900-17C research "Cyber Security Challenge Data Collection", the Committee has given you approval to proceed with your proposed study.

We wish you well with your research.

Yours sincerely

Mike Mayo
Human Research Ethics Committee
Faculty of Computing and Mathematical Sciences

New Zealand Cyber Security Challenge

Ethical Agreement

Participant Declaration

In parts of the New Zealand Cyber Security Challenge (NZCSC) 2017 and training, you will be exploring cyber security techniques including techniques/methods that could be used to illegally access computer systems. This is a required part of learning to defend against cyber attacks. However, this knowledge could also be misused. To take part in this challenge, students must subscribe to the ethos that our individual and group purpose is to defend against cyber-attacks.

It is a condition of registration in this challenge that you agree to act ethically with respect to computer systems. This condition is enduring. Acting ethically includes not attempting to gain unauthorized access to a network or computer system and instructing or assisting others to do so. You will also adhere to the rules outlined on the back of this document. Should one of these be breached, the participant(s) in question will be disqualified from the competition immediately.

You will not deviate from the instructions or environments set out by that challenge organisers. The university and the challenge organisers will not be liable to your actions and you are responsible for your own choices, actions and consequences of non-compliance (which may include university expulsion or prosecution by law enforcement agencies).

I, (Name:) , accept these conditions.

Activity: Cyber Security Challenge

Date:

Witness 1 (Name and Signature):

Witness 2 (Name and Signature):

Rules of engagement

General Rules:

- No tampering with the challenge infrastructure
- No networking sniffing or scanning (Note: This rule excludes Round 3 challenge)
- No gaining of unauthorized access to the network and computer system
- Use only the tools provided to you. These are available on your machine (Note: This rule excludes Round 3 challenge)
- No spamming the challenges or scoreboard
- No screenwatching!
- No Social Engineering on other teams during NZCSC event

Round 3 Rules:

- Notify challenge organizers before restarting a Lab machines and VMs
- Do not close logging mechanisms (e.g. screen capture applications)

"Reminder, act ethically throughout the NZCSC event. Failure to comply with the above rules will result in an expulsion from the NZCSC event"

Appendix C

Cognitive Load, Working Memory, and Performance Assessment

C1: "A Cognitive Load Conceptual Framework Through Construct and Estimation Methods"

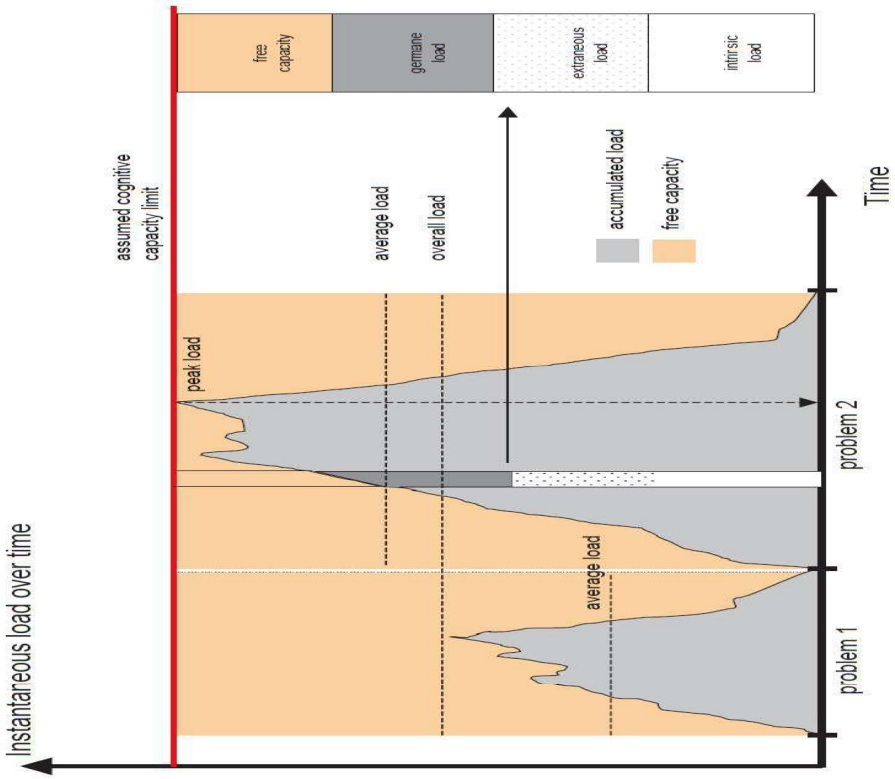


Figure C.1: Cognitive Load and Overall Working Load Explained

Citation details: Paas, Fred, Juhani E. Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. "Cognitive load measurement as a means to advance cognitive load theory." Educational psychologist 38, no. 1 (2003): 63-71.

C2: "A List of Studies Measuring Cognitive Load, Mental Efficiency and Techniques Used"

Studies That Measured Cognitive Load and Calculated Mental Efficiency and the Measurement Technique They Used

Studies	*Cognitive Load Measurement Technique*	*Mental Efficiency*
Sweller (1988)	PS, ST	
Paas (1992)	RS9	
Paas & van Merriënboer (1993)	RS9	ME
Paas & van Merriënboer (1994b)	RS9, HRV	ME
Cerpa, Chandler, & Sweller (1996)	RS9	ME
Chandler & Sweller (1996)	ST	
Marcus, Cooper, & Sweller (1996)	RS7, ST	ME
Tindall-Ford, Chandler, & Sweller (1997)	RS7	ME
Yeung, Jin, & Sweller (1997)	RS9	ME
de Croock, van Merriënboer, & Paas (1998)	RS9	
Kalyuga, Chandler, & Sweller (1998)	RS7	ME
Kalyuga, Chandler, & Sweller (1999)	RS7	ME
Tuovinen & Sweller (1999)	RS9	ME
Yeung (1999)	RS9	ME
Kalyuga, Chandler, & Sweller (2000)	RS7	ME
Kalyuga, Chandler, & Sweller (2001)	RS7	ME
Kalyuga, Chandler, Tuovinen, & Sweller (2001)	RS9	ME
Mayer & Chandler (2001)	RS7	
Pollock, Chandler, & Sweller (2002)	RS7	ME
Stark, Mandl, Gruber, & Renkl (2002)	RS9	
Tabbers, Martens, & van Merriënboer (2002)	RS9	
Tabbers, Martens, & van Merriënboer (in press)	RS9	
Van Gerven, Paas, van Merriënboer, Hendriks, & Schmidt (2002)	RS9	ME
Van Gerven, Paas, van Merriënboer, & Schmidt (2002a)	RS9	ME
Van Gerven, Paas, van Merriënboer, & Schmidt (2002b)	PR	
Van Gerven, Paas, van Merriënboer, & Schmidt (2002c)	RS9, ST	ME
van Merriënboer, Schuurman, de Croock, & Paas (2002)	RS9	ME

Note. Studies are listed in chronological order. PS = production system; ST = secondary task technique; RS = rating scale (9-point or 7-point scale); ME = mental efficiency; HRV = heart rate variability; PR = pupillary responses.

Figure C.2: Existing Cognitive Load Studies

Citation details: Paas, Fred, Juhani E. Tuovinen, Huib Tabbers, and Pascal WM Van Gerven. "Cognitive load measurement as a means to advance cognitive load theory." Educational psychologist 38, no. 1 (2003): 63-71.

C3: "SvEm - Cognitive and Working Memory Load Observation Experiment Results"

Table C.1: C3: "SvEm–Cognitive and Working Memory Load Observation Experiment Results"

SvEm-CWML Observation set: without SvEm-CWML Instruction Set		
Security Visual Node Identified	Memory Efficiency (t_me)	Cognitive Load (Cl)
1	80	20
3	40	25
2	20	22
6	89	56
5	66	55
4	56	45
4	55	45
7	123	78
8	126	80
4	98	55
3	60	34
2	33	23
6	110	77
7	120	78
4	45	44
1	23	20
1	26	20
1	30	20
3	76	24
4	56	45
5	55	55
6	56	67
7	80	86

Security Visual Node Identified	Memory Efficiency (t_me)	Cognitive Load (Cl)
8	90	88
3	40	45
3	40	45
9	95	88
10	99	90
4	66	44
2	20	22
10	100	94
8	93	70
9	98	78
8	92	71
7	110	69
4	66	44
2	20	22
10	100	94
8	93	70
9	98	78
8	92	71
7	110	69
11	120	94
12	123	96
12	124	95

Security Visual Node Identified	Memory Efficiency (t_me)	Cognitive Load (Cl)
1	22	20
12	120	94
11	118	90
10	110	95
9	99	90
8	90	77
7	88	60
4	57	45
5	59	45
6	60	50
4	44	43
7	77	50
10	100	78
10	99	80
5	66	45
6	66	55
7	78	66
8	88	78
9	90	85
2	20	22
3	34	33
4	44	39

References

[1] C. Dunne and C. Dunne, "App turns NYC subway maps into interactive data visualizations," [Online]. Available: https://www.fastcodesign.com/3030949/turn-nyc-subway-maps-into-interactive-data-visualizations-with-this-app, [Accessed 2017-09-03], May 2014.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[3] Y. K. Leung and M. D. Apperley, "E3: Towards the metrication of graphical presentation techniques for large data sets," in *International Conference on Human-Computer Interaction*. Springer, 1993, pp. 125–140.

[4] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educational psychologist*, vol. 38, no. 1, pp. 63–71, 2003.

[5] Y. S. Tan, "Reconstructing Data Provenance from Log Files," Doctoral Thesis, The University of Waikato, Hamilton, New Zealand, 2017. [Online]. Available: https://hdl.handle.net/10289/11388

[6] "Deepnode.com, 'Why Deep Node?'," [Online]. Available: https://www.deepnode.com/why-deep-node/, [Accessed 2018-02-23].

[7] G. Conti, *Security data visualization: graphical techniques for network analysis*. No Starch Press, 2007.

[8] J.-M. Hoc, *Psychology of programming*. Academic Press, 2014.

[9] "Itu.int, 'cybersecurity'," [Online]. Available: http://www.itu.int:80/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx, [Accessed 2017-10-07].

[10] R. Von Solms and J. Van Niekerk, "From information security to cyber security," *computers & security*, vol. 38, pp. 97–102, 2013.

[11] "Merriam-webster.com, 'Cybersecurity | Definition of Cybersecurity by Merriam-Webster'," [Online]. Available: https://www.merriam-webster.com/dictionary/cybersecurity, [Accessed 2017-10-07].

[12] J. Jang-Jaccard and S. Nepal, "A survey of emerging threats in cybersecurity," *Journal of Computer and System Sciences*, vol. 80, no. 5, pp. 973 – 993, 2014, special Issue on Dependable and Secure Computing. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022000014000178

[13] J. Garae and R. K. L. Ko, *Visualization and Data Provenance Trends in Decision Support for Cybersecurity*. Cham: Springer International Publishing, 2017, pp. 243–270. [Online]. Available: https://doi.org/10.1007/978-3-319-59439-2_9

[14] E. Tufte and P. Graves-Morris, "The visual display of quantitative information.; 1983," 2014.

[15] D. A. Wheeler and G. N. Larsen, "Techniques for cyber attack attribution," Institute for defence analyses Alexandria Va, Tech. Rep., 2003.

[16] J. Garae, R. K. Ko, and S. Chaisiri, "Uvisp: User-centric visualization of data provenance with gestalt principles," in *Trustcom/BigDataSE/I SPA, 2016 IEEE*. IEEE, 2016, pp. 1923–1930.

[17] Y. L. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," *ACM Sigmod Record*, vol. 34, no. 3, pp. 31–36, 2005.

[18] M. O. Ward, G. Grinstein, and D. Keim, *Interactive data visualization: foundations, techniques, and applications*. CRC Press, 2010.

[19] N. Yee and J. N. Bailenson, "The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation," *Media Psychology*, vol. 12, no. 2, pp. 195–209, 2009.

[20] I. Seo, H. Lee, and S. C. Han, "Cylindrical coordinates security visualization for multiple domain command and control botnet detection," *computers & security*, vol. 46, pp. 141–153, 2014.

[21] H. Choi and H. Lee, "Pcav: Internet attack visualization on parallel coordinates," in *International Conference on Information and Communications Security*. Springer, 2005, pp. 454–466.

[22] P. A. Facione, "Critical thinking: What it is and why it counts," *Retrieved June*, vol. 9, p. 2004, 1998.

[23] R. Marty, *Applied security visualization*. Addison-Wesley Upper Saddle River, 2009.

[24] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, "Challenges in visual data analysis," in *Information Visualization, 2006. IV 2006. Tenth International Conference on*. IEEE, 2006, pp. 9–16.

[25] L. Harrison and A. Lu, "The future of security visualization: Lessons from network visualization," *IEEE Network*, vol. 26, no. 6, 2012.

[26] S. Liu, W. Cui, Y. Wu, and M. Liu, "A survey on information visualization: recent advances and challenges," *The Visual Computer*, vol. 30, no. 12, pp. 1373–1393, 2014.

[27] J. Gong and P. Tarasewich, "Guidelines for handheld mobile device interface design," in *Proceedings of DSI 2004 Annual Meeting*, 2004, pp. 3751–3756.

[28] C. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans, and T. S. Yoo, "Nih-nsf visualization research challenges report." Institute of Electrical and Electronics Engineers (2005), 2005.

[29] M. Friendly, "A brief history of data visualization," *Handbook of data visualization*, pp. 15–56, 2008.

[30] M. Friendly and D. J. Denis, "Milestones in the history of thematic cartography, statistical graphics, and data visualization," *URL http://www. datavis. ca/milestones*, vol. 32, 2001.

[31] N. J. Thrower, *Maps and civilization: cartography in culture and society*. University of Chicago Press, 2008.

[32] A. Hald, *A history of probability and statistics and their applications before 1750*. John Wiley & Sons, 2003, vol. 501.

[33] E. Halley, *The Description and Uses of a New and Correct SEA-CHART of the whole World, shewing the Variations of the COMPASS*, 1702.

[34] B. G. Survey and W. Smith, *A Delineation of the Strata of England and Wales with Part of Scotland; Exhibiting the Collieries and Mines, the Marshes and Fen Lands Originally Overflowed by the Sea, and the Varieties of Soil According to the Variations in the Substrata, Illustrated by the Most Descriptive Names 1: 625 000.[Reproduction at Half Scale, 1: 625 000, of William Smith's Original 1815 Map]*. British Geological Survey, 2003.

[35] H. G. Funkhouser, "Historical development of the graphical representation of statistical data," *Osiris*, vol. 3, pp. 269–404, 1937.

[36] M. Friendly and D. Denis, "Discussion and comments. approche graphique en analyse des données. the roots and branches of modern statistical graphics," *Journal de la société française de statistique*, vol. 141, no. 4, pp. 51–60, 2000.

[37] E. Tufte and P. Graves-Morris, "The visual display of quantitative information.; 1983," 2014.

[38] A. Inselberg, "The plane with parallel coordinates," *The visual computer*, vol. 1, no. 2, pp. 69–91, 1985.

[39] F. W. Young and C. M. Bann, "Vista: The visual statistics system," Technical Report 94–1 (c), UNC LL Thurstone Psychometric Laboratory Research Memorandum, Tech. Rep., 1996.

[40] A. D'Amico and M. Kocka, "Information assurance visualizations for specific stages of situational awareness and intended uses: lessons learned," in *IEEE Workshop on Visualization for Computer Security, 2005. (VizSEC 05)*, Oct. 2005, pp. 107–112.

[41] K. Levchenko, R. Paturi, and G. Varghese, "On the Difficulty of Scalably Detecting Network Attacks," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, ser. CCS '04. New York, NY, USA: ACM, 2004, pp. 12–20. [Online]. Available: http://doi.acm.org/10.1145/1030083.1030087

[42] A. Oline and D. Reiners, "Exploring three-dimensional visualization for intrusion detection," in *IEEE Workshop on Visualization for Computer Security, 2005. (VizSEC 05)*, Oct. 2005, pp. 113–120.

[43] E. Lindquist, "Surveying the world of visualization," *Australian National University*, 2011.

[44] T. Webb, "Exploration of biogeographic databases: zoom lenses, space travel, and scientific imagination," *Journal of biogeography*, pp. 7–9, 2000.

[45] W. Zhuo and Y. Nadjin, "Malwarevis: entity-based visualization of malware network traces," in *Proceedings of the ninth international symposium on visualization for cyber security*. ACM, 2012, pp. 41–47.

[46] J. J. Fowler, T. Johnson, P. Simonetto, M. Schneider, C. Acedo, S. Kobourov, and L. Lazos, "IMap: Visualizing Network Activity over Internet Maps," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 80–87. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671501

[47] F. Stoffel, F. Fischer, and D. A. Keim, "Finding Anomalies in Time-series Using Visual Correlation for Interactive Root Cause Analysis," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec '13. New York, NY, USA: ACM, 2013, pp. 65–72. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517966

[48] S. Walton, E. Maguire, and M. Chen, "Multiple Queries with Conditional Attributes (QCATs) for Anomaly Detection and Visualization," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 17–24. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671502

[49] S. Chen, C. Guo, X. Yuan, F. Merkle, H. Schaefer, and T. Ertl, "OCEANS: Online Collaborative Explorative Analysis on Network Security," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671493

[50] L. Harrison, R. Spahn, M. Iannacone, E. Downing, and J. R. Goodall, "NV: Nessus Vulnerability Visualization for the Web," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 25–32. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379694

[51] C. Humphries, N. Prigent, C. Bidan, and F. Majorczyk, "ELVIS: Extensible Log VISualization," in *Proceedings of the Tenth Workshop on Visualization for Cyber*

Security, ser. VizSec '13. New York, NY, USA: ACM, 2013, pp. 9–16. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517959

[52] D. M. Best, S. Bohn, D. Love, A. Wynne, and W. A. Pike, "Real-time Visualization of Network Behaviors for Situational Awareness," in *Proceedings of the Seventh International Symposium on Visualization for Cyber Security*, ser. VizSec '10. New York, NY, USA: ACM, 2010, pp. 79–90. [Online]. Available: http://doi.acm.org/10.1145/1850795.1850805

[53] T. J. Jankun-Kelly, J. Franck, D. Wilson, J. Carver, D. Dampier, and J. E. S. Ii, "Show Me How You See: Lessons from Studying Computer Forensics Experts for Visualization," in *Visualization for Computer Security*, ser. Lecture Notes in Computer Science, J. R. Goodall, G. Conti, and K.-L. Ma, Eds. Springer Berlin Heidelberg, 2008, no. 5210, pp. 80–86. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-85933-8_8

[54] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Transactions on visualization and computer graphics*, vol. 18, no. 8, pp. 1313–1329, 2012.

[55] M. Bastian, S. Heymann, M. Jacomy *et al.*, "Gephi: an open source software for exploring and manipulating networks." *Icwsm*, vol. 8, pp. 361–362, 2009.

[56] T. L. Naps, G. Rößling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, M. McNally, S. Rodger *et al.*, "Exploring the role of visualization and engagement in computer science education," in *ACM Sigcse Bulletin*, vol. 35, no. 2. ACM, 2002, pp. 131–152.

[57] R. Gove, J. Saxe, S. Gold, A. Long, and G. Bergamo, "SEEM: A Scalable Visualization for Comparing Multiple Large Sets of Attributes for Malware Analysis," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 72–79. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671496

[58] T. WÃŒchner, A. Pretschner, and M. Ochoa, "DAVAST: Data-centric System Level Activity Visualization," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 25–32. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671499

[59] A. Nandeshwar, *Tableau data visualization cookbook*. Packt Publishing Ltd, 2013.

[60] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon, "Manyeyes: a site for visualization at internet scale," *IEEE transactions on visualization and computer graphics*, vol. 13, no. 6, 2007.

[61] A. Inselberg and B. Dimsdale, "Parallel coordinates for visualizing multi-dimensional geometry," in *Computer Graphics 1987*. Springer, 1987, pp. 25–44.

[62] T. Langlotz, T. Nguyen, D. Schmalstieg, and R. Grasset, "Next-generation augmented reality browsers: rich, seamless, and adaptive," *Proceedings of the IEEE*, vol. 102, no. 2, pp. 155–169, 2014.

[63] S. Ortiz Jr, "Is 3d finally ready for the web?" *Computer*, vol. 43, no. 1, 2010.

[64] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva, "A user study of visualization effectiveness using eeg and cognitive load," in *Computer Graphics Forum*, vol. 30, no. 3. Wiley Online Library, 2011, pp. 791–800.

[65] S. M. Casner and B. F. Gore, "Measuring and evaluating workload: A primer," *NASA Technical Memorandum*, vol. 216395, p. 2010, 2010.

[66] G. B. Reid, S. S. Potter, and J. Bressler, "Subjective workload assessment technique (swat): A user's guide," *Wright Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory*, 1989.

[67] A. T. Duchowski, "Eye tracking methodology," *Theory and practice*, vol. 328, 2007.

[68] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[69] A. Çöltekin, S. I. Fabrikant, and M. Lacayo, "Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1559–1575, 2010.

[70] L. Harrison, F. Yang, S. Franconeri, and R. Chang, "Ranking visualizations of correlation using weber's law," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 12, pp. 1943–1952, 2014.

[71] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[72] R. Tamassia, *Handbook of graph drawing and visualization*. CRC press, 2013.

[73] R. Ball, G. A. Fink, and C. North, "Home-centric visualization of network traffic for security administration," in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*. ACM, 2004, pp. 55–64.

[74] A. Perrig and D. Song, "Hash visualization: A new technique to improve real-world security," in *International Workshop on Cryptographic Techniques and E-Commerce*, 1999, pp. 131–138.

[75] M. Stonebraker, J. Chen, N. Nathan, C. Paxson, and J. Wu, "Tioga: Providing data management support for scientific visualization applications," in *VLDB*, vol. 93, 1993, pp. 25–38.

[76] D. Staheli, T. Yu, R. J. Crouser, S. Damodaran, K. Nam, D. O'Gwynn, S. McKenna, and L. Harrison, "Visualization Evaluation for Cyber Security: Trends and Future Directions," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 49–56. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671492

[77] *VizSEC/DMSEC '04: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*. New York, NY, USA: ACM, 2004.

[78] "Vizsec.org, 'VizSec'," [Online]. Available: http://vizsec.org/, [Accessed 2017-09-10].

[79] T. Nunnally, K. Abdullah, A. S. Uluagac, J. A. Copeland, and R. Beyah, "Navsec: A recommender system for 3d network security visualizations," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*. ACM, 2013, pp. 41–48.

[80] T. Kohno, "Attacking and Repairing the winZip Encryption Scheme," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, ser. CCS '04. New York, NY, USA: ACM, 2004, pp. 72–81. [Online]. Available: http://doi.acm.org/10.1145/1030083.1030095

[81] B. Waters, A. Juels, J. A. Halderman, and E. W. Felten, "New Client Puzzle Outsourcing Techniques for DoS Resistance," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, ser. CCS '04. New York, NY, USA: ACM, 2004, pp. 246–256. [Online]. Available: http://doi.acm.org/10.1145/1030083.1030117

[82] D. Yao, N. Fazio, Y. Dodis, and A. Lysyanskaya, "ID-based Encryption for Complex Hierarchies with Applications to Forward Security and Broadcast Encryption," in *Proceedings of the 11th ACM Conference on Computer and Communications Security*,

ser. CCS '04. New York, NY, USA: ACM, 2004, pp. 354–363. [Online]. Available: http://doi.acm.org/10.1145/1030083.1030130

[83] F. Fischer and D. A. Keim, "NStreamAware: Real-time Visual Analytics for Data Streams to Enhance Situational Awareness," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 65–72. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671495

[84] A. Komlodi, P. Rheingans, U. Ayachit, J. R. Goodall, and A. Joshi, "A User-centered Look at Glyph-based Security Visualization," in *Proceedings of the IEEE Workshops on Visualization for Computer Security*, ser. VIZSEC '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 3–. [Online]. Available: http://dx.doi.org/10.1109/VIZSEC.2005.1

[85] W. Fang, B. P. Miller, and J. A. Kupsch, "Automated Tracing and Visualization of Software Security Structure and Properties," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 9–16. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379692

[86] D. Inoue, M. Eto, K. Suzuki, M. Suzuki, and K. Nakao, "DAEDALUS-VIZ: Novel Real-time 3d Visualization for Darknet Monitoring-based Alert System," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 72–79. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379700

[87] T. Jankun-Kelly, D. Wilson, A. Stamps, J. Franck, J. Carver, and J. Swan, "A visual analytic framework for exploring relationships in textual contents of digital forensics evidence," in *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, Oct. 2009, pp. 39–44.

[88] T. R. Leschke and C. Nicholas, "Change-link 2.0: A Digital Forensic Tool for Visualizing Changes to Shadow Volume Data," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec '13. New York, NY, USA: ACM, 2013, pp. 17–24. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517960

[89] S. Engle and S. Whalen, "Visualizing Distributed Memory Computations with Hive Plots," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 56–63. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379698

[90] F. Mansmann, T. GÃ¶bel, and W. Cheswick, “Visual Analysis of Complex Firewall Configurations,” in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec ’12. New York, NY, USA: ACM, 2012, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379691

[91] O. Tsigkas, O. Thonnard, and D. Tzovaras, “Visual Spam Campaigns Analysis Using Abstract Graphs Representation,” in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec ’12. New York, NY, USA: ACM, 2012, pp. 64–71. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379699

[92] D. M. Best, A. Endert, and D. Kidwell, “7 Key Challenges for Visualization in Cyber Network Defense,” in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec ’14. New York, NY, USA: ACM, 2014, pp. 33–40. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671497

[93] P. Ren, “Ensuring the Continuing Success of Vizsec,” in *Proceedings of the 3rd International Workshop on Visualization for Computer Security*, ser. VizSEC ’06. New York, NY, USA: ACM, 2006, pp. 67–70. [Online]. Available: http://doi.acm.org/10.1145/1179576.1179591

[94] R. F. Erbacher, “Visualization Design for Immediate High-level Situational Assessment,” in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec ’12. New York, NY, USA: ACM, 2012, pp. 17–24. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379693

[95] “Ieee symposium on visualization for cyber security , ‘About VizSec’,” [Online]. Available: http://vizsec.org/about.html, [Accessed 2015-07-25].

[96] M. Alsaleh, A. Alqahtani, A. Alarifi, and A. Al-Salman, “Visualizing PHPIDS Log Files for Better Understanding of Web Server Attacks,” in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec ’13. New York, NY, USA: ACM, 2013, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517958

[97] D. Barrera and P. Van Oorschot, “Security visualization tools and IPv6 addresses,” in *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, Oct. 2009, pp. 21–26.

[98] F. Fischer, J. Fuchs, P.-A. Vervier, F. Mansmann, and O. Thonnard, "VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 80–87. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379701

[99] J. Glanfield, S. Brooks, T. Taylor, D. Paterson, C. Smith, C. Gates, and J. McHugh, "Over flow: An overview visualization for network analysis," in *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, Oct. 2009, pp. 11–19.

[100] S. Papadopoulos, G. Theodoridis, and D. Tzovaras, "BGPfuse: Using Visual Feature Fusion for the Detection and Attribution of BGP Anomalies," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec '13. New York, NY, USA: ACM, 2013, pp. 57–64. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517965

[101] J. Saxe, D. Mentis, and C. Greamo, "Visualization of Shared System Call Sequence Relationships in Large Malware Corpora," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12. New York, NY, USA: ACM, 2012, pp. 33–40. [Online]. Available: http://doi.acm.org/10.1145/2379690.2379695

[102] J.-E. Stange, M. Dörk, J. Landstorfer, and R. Wettach, "Visual filter: graphical exploration of network security log files," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*. ACM, 2014, pp. 41–48.

[103] D. Quist and L. Liebrock, "Visualizing compiled executables for malware analysis," in *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, Oct. 2009, pp. 27–32.

[104] B. Cheswick, "Visual tools for security: Is there a there there?" in *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, Oct. 2009.

[105] Y. Ghanam and S. Carpendale, "A Survey Paper on Software Architecture Visualization," Technical Report, 2008. [Online]. Available: http://dspace.ucalgary.ca/bitstream/1880/46648/1/2008-906-19.pdf

[106] W. J. Matuszak, L. DiPippo, and Y. L. Sun, "CyberSAVe: Situational Awareness Visualization for Cyber Security of Smart Grid Systems," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec '13. New York, NY, USA:

ACM, 2013, pp. 25–32. [Online]. Available:
http://doi.acm.org/10.1145/2517957.2517961

[107] A. Long, J. Saxe, and R. Gove, "Detecting Malware Samples with Similar Image Sets," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 88–95. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671500

[108] V. Paelke, C. Reimann, and W. Rosenbach, "A visualization design repository for mobile devices," in *Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa.* ACM, 2003, pp. 57–62.

[109] G. Bieber, T. Kirste, and B. Urban, "Ambient interaction by smart watches," in *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments.* ACM, 2012, p. 39.

[110] L. Chittaro, "Visualizing information on mobile devices," *Computer*, vol. 39, no. 3, pp. 40–45, 2006.

[111] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on.* IEEE, 1996, pp. 336–343.

[112] J. Hao and K. Zhang, "A mobile interface for hierarchical information visualization and navigation," in *Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on.* IEEE, 2007, pp. 1–7.

[113] A. Butler, S. Izadi, and S. Hodges, "Sidesight: multi-touch interaction around small devices," in *Proceedings of the 21st annual ACM symposium on User interface software and technology.* ACM, 2008, pp. 201–204.

[114] T. Lachev and E. Price, *Applied Microsoft Power BI: Bring your data to life!* Prologika Press, 2015.

[115] E. Nygren, R. K. Sitaraman, and J. Sun, "The akamai network: a platform for high-performance internet applications," *ACM SIGOPS Operating Systems Review*, vol. 44, no. 3, pp. 2–19, 2010.

[116] L. Grammel, M. Tory, and M.-A. Storey, "How information visualization novices construct visualizations," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 943–952, 2010.

[117] W. Wang and M. L. Huang, "Parallel coordinates visualization of large data investigation on hdds," in *Computer Graphics, Imaging and Visualization (CGIV), 2013 10th International Conference*. IEEE, 2013, pp. 93–99.

[118] Y. Okada, "Network data visualization using parallel coordinates version of time-tunnel with 2dto2d visualization for intrusion detection," in *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*. IEEE, 2013, pp. 1088–1093.

[119] K. Zhao, B. Liu, T. M. Tirpak, and A. Schaller, "Detecting patterns of change using enhanced parallel coordinates visualization," in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*. IEEE, 2003, pp. 747–750.

[120] H. Chen, H. Atabakhsh, C. Tseng, B. Marshall, S. Kaza, S. Eggers, H. Gowda, A. Shah, T. Petersen, and C. Violette, "Visualization in Law Enforcement," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '05. New York, NY, USA: ACM, 2005, pp. 1268–1271. [Online]. Available: http://doi.acm.org/10.1145/1056808.1056893

[121] H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga, and J. Schroeder, "COPLINK: Managing Law Enforcement Data and Knowledge," *Commun. ACM*, vol. 46, no. 1, pp. 28–34, Jan. 2003. [Online]. Available: http://doi.acm.org/10.1145/602421.602441

[122] J. Xu and H. Chen, "Criminal Network Analysis and Visualization," *Commun. ACM*, vol. 48, no. 6, pp. 100–107, Jun. 2005. [Online]. Available: http://doi.acm.org/10.1145/1064830.1064834

[123] B. Marshall, S. Kaza, J. Xu, H. Atabakhsh, T. Petersen, C. Violette, and H. Chen, "Cross-jurisdictional criminal activity networks to support border and transportation security," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. IEEE, 2004, pp. 100–105.

[124] J. Xu, B. Marshall, S. Kaza, and H. Chen, "Analyzing and visualizing criminal network dynamics: A case study," in *ISI*. Springer, 2004, pp. 359–377.

[125] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Information visualization*, vol. 7, no. 2, pp. 118–132, 2008.

[126] S. Barnum, "Standardizing cyber threat intelligence information with the structured threat information expression (stix™)," *MITRE Corporation*, vol. 11, pp. 1–22, 2012.

[127] E. Casey, G. Back, and S. Barnum, "Leveraging cybox™ to standardize representation and exchange of digital forensic information," *Digital Investigation*, vol. 12, pp. S102–S110, 2015.

[128] H. Chen, H. Atabakhsh, C. Tseng, B. Marshall, S. Kaza, S. Eggers, H. Gowda, A. Shah, T. Petersen, and C. Violette, "Visualization in law enforcement," in *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 2005, pp. 1268–1271.

[129] "Interpol.int, 'research / Cybercrime / Crime areas / Internet / Home - INTERPOL'," [Online]. Available: http://www.interpol.int/Crime-areas/Cybercrime/Research, [Online: accessed 2016-08-02].

[130] T. I. S. on Visualization for Cyber Security, "Vizsec.org,'data Sets'," [Online]. Available: http://vizsec.org/data/, [Acessed 2017-09-04].

[131] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato, "Problem Characterization and Abstraction for Visual Analytics in Behavior-based Malware Pattern Analysis," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, ser. VizSec '14. New York, NY, USA: ACM, 2014, pp. 9–16. [Online]. Available: http://doi.acm.org/10.1145/2671491.2671498

[132] W. Huang, P. Eades, and S.-H. Hong, "Measuring effectiveness of graph visualizations: A cognitive load perspective," *Information Visualization*, vol. 8, no. 3, pp. 139–152, 2009.

[133] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko, "Understanding and characterizing insights: how do people gain insights using information visualization?" in *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*. ACM, 2008, p. 4.

[134] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 4, pp. 443–456, 2005.

[135] P. Saraiya, C. North, V. Lam, and K. A. Duca, "An insight-based longitudinal study of visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1511–1522, 2006.

[136] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *Proceedings of international conference on intelligence analysis*, vol. 5, 2005, pp. 2–4.

[137] L. Hao, C. G. Healey, and S. E. Hutchinson, "Flexible Web Visualization for Alert-based Network Security Analytics," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, ser. VizSec '13. New York, NY, USA: ACM, 2013, pp. 33–40. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517962

[138] J. Rasmussen, K. Ehrlich, S. Ross, S. Kirk, D. Gruen, and J. Patterson, "Nimble Cybersecurity Incident Management Through Visualization and Defensible Recommendations," in *Proceedings of the Seventh International Symposium on Visualization for Cyber Security*, ser. VizSec '10. New York, NY, USA: ACM, 2010, pp. 102–113. [Online]. Available: http://doi.acm.org/10.1145/1850795.1850807

[139] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *journal of the Association for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[140] C. Guitton and E. Korzak, "The sophistication criterion for attribution: Identifying the perpetrators of cyber-attacks," *The RUSI Journal*, vol. 158, no. 4, pp. 62–68, 2013.

[141] T. Rid and B. Buchanan, "Attributing cyber attacks," *Journal of Strategic Studies*, vol. 38, no. 1-2, pp. 4–37, 2015.

[142] J. Garae, R. K. Ko, J. Kho, S. Suwadi, M. A. Will, and M. Apperley, "Visualizing the new zealand cyber security challenge for attack behaviors," in *Trustcom/BigDataSE/ICESS, 2017 IEEE*. IEEE, 2017, pp. 1123–1130.

[143] M. Azab, B. Mokhtar, A. S. Abed, and M. Eltoweissy, "Toward smart moving target defense for linux container resiliency," in *Local Computer Networks (LCN), 2016 IEEE 41st Conference on*. IEEE, 2016, pp. 619–622.

[144] J.-W. Selij and E. van den Haak, "'A Visitation of Sysdig', (2014)," 2014.

[145] F. Fuentes and D. C. Kar, "Ethereal vs. tcpdump: a comparative study on packet sniffing tools for educational purpose," *Journal of Computing Sciences in Colleges*, vol. 20, no. 4, pp. 169–176, 2005.

[146] V. Jacobson, C. Leres, and S. McCanne, "The tcpdump manual page," *Lawrence Berkeley Laboratory, Berkeley, CA*, vol. 143, 1989.

[147] A. Stevenson, Ed., *Oxford Dictionary of English*. Oxford University Press, Jan. 2010. [Online]. Available: http://www.oxfordreference.com/view/10.1093/acref/9780199571123.001.0001/acref-9780199571123

[148] M. Kay and J. Heer, "Beyond weber's law: A second look at ranking visualizations of correlation," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 469–478, 2016.

[149] K. Labs, "Kaspersky Labs,'kaspersky Lab Reports Mobile Malware in 2013 More Than Doubles from Previous Year | Kaspersky Lab US'," [Online]. Available: https://usa.kaspersky.com/about/press-releases/2014_kaspersky-lab-reports-mobile-malware-in-2013-more-than-doubles\-from-previous-year, [Accessed 2018-01-31].

[150] "Securelist.com, 'KSB 2014. Overall statistics for 2014'," [Online]. Available: https://securelist.com/kaspersky-security-bulletin-2014-overall-statistics-for-2014/68010/, [Accessed 2018-01-31].

[151] "Securelist.com, 'IT Threat Evolution Q3 2017. Statistics - Securelist'," [Online]. Available: https://securelist.com/it-threat-evolution-q3-2017-statistics/83131/, [Accessed 2018-01-31].

[152] "Securelist.com, 'Mobile Malware Evolution 2016'," [Online]. Available: https://securelist.com/mobile-malware-evolution-2016/77681/, [Accessed 2018-01-31].

[153] "Kaspersky Labs, 'Mobile Malware Threats | Android Security Issues | Kaspersky Lab'," [Online]. Available: https://www.kaspersky.com/resource-center/threats/mobile, [Accessed 2018-01-31].

[154] Interpol.int, "'notices / INTERPOL expertise / Internet / Home - INTERPOL'," [Online]. Available: https://www.interpol.int/INTERPOL-expertise/Notices, [Accessed 2018-01-21].

[155] F. G. Paas and J. J. Van Merriënboer, "Instructional control of cognitive load in the training of complex cognitive tasks," *Educational psychology review*, vol. 6, no. 4, pp. 351–371, 1994.

[156] B. Xie and G. Salvendy, "Prediction of mental workload in single and multiple tasks environments," *International journal of cognitive ergonomics*, vol. 4, no. 3, pp. 213–242, 2000.

[157] M. G. Ash, *Gestalt psychology in German culture, 1890-1967: Holism and the quest for objectivity.* Cambridge University Press, 1998.

[158] R. K. Ko and M. A. Will, "Progger: an efficient, tamper-evident kernel-space logger for cloud data provenance tracking," in *Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on.* IEEE, 2014, pp. 881–889.

[159] J. L. Carlson, *Redis in Action.* Greenwich, CT, USA: Manning Publications Co., 2013.

[160] K. Chodorow, *MongoDB: The Definitive Guide: Powerful and Scalable Data Storage.* O'Reilly Media, Inc., 2013.

[161] J. R. Wilson and J. Carter, *Node. js the right way: Practical, server-side javascript that scales.* Pragmatic Bookshelf, 2013.

[162] J. Congote, A. Segura, L. Kabongo, A. Moreno, J. Posada, and O. Ruiz, "Interactive visualization of volumetric data with webgl in real-time," in *Proceedings of the 16th International Conference on 3D Web Technology.* ACM, 2011, pp. 137–146.

[163] T. Parisi, *WebGL: up and running.* " O'Reilly Media, Inc.", 2012.

[164] A. Sahagun, "Linkurious - Understand the connections in your data," [Online]. Available: https://linkurio.us/product/#benefits/, [Accessed 2018-04-04], 2015. [Online]. Available: https://linkurio.us/product/#benefits

[165] H. Kuzuno and C. Karam, "Blockchain explorer: An analytical process and investigation environment for bitcoin," in *Electronic Crime Research (eCrime), 2017 APWG Symposium on.* IEEE, 2017, pp. 9–16.

[166] G. Disterer, "Iso/iec 27000, 27001 and 27002 for Information Security Management," *Journal of Information Security*, vol. 4, no. 02, p. 92, 2013.

[167] E. Humphreys, *Implementing the ISO/IEC 27001 Information Security Management System Standard.* Artech House, Inc., 2007.

[168] K. B. Head, "How to measure the effectiveness of information security," [Online]. Available: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/news/2016/12/Ref2151.html, [Accessed 2018-03-05].

[169] S. Brown, D. Pyke, and P. Steenhof, "Electric vehicles: The role and importance of standards in an emerging market," *Energy Policy*, vol. 38, no. 7, pp. 3797–3806, 2010.

[170] "ISO/IEC, 'ISO/IEC 27001 certification standard'," [Online]. Available: http://www.iso27001security.com/html/27001.html, [Accessed 2018-03-05].

[171] "ISO/IEC, 'ISO/IEC 27001 certification standard'," [Online]. Available: http://www.iso27001security.com/html/27001.html, [Accessed 2018-03-05].

[172] I. Ltd, "ISO/IEC, 'ISO/IEC 27032 cybersecurity guideline'," [Online]. Available: http://iso27001security.com/html/27032.html, [Accessed 2018-03-05].

[173] S. Ramanauskaitė, D. Olifer, N. Goranin, A. Čenys, and L. Radvilavičius, "Visualization of mapped security standards for analysis and use optimisation," *Int. J. Comput. Theor. Eng*, vol. 6, no. 5, pp. 372–376, 2014.

[174] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.

[175] R. Lengler and M. J. Eppler, "A periodic table of visualization methods," *available at www. visual-literacy. org/periodic_table/periodic_table. html*, 2007.

[176] I. hichert.com, "IBCS Standards • HICHERT+FAISST IBCS Institute," [Online]. Available: https://www.hichert.com/standards/, 2018, [Accessed 2018-03-04].

[177] "hichert.com, IBCS – Using Data Visualization best practices to improve business communication," 2018, [Online]. Available: http://www.datacommunitydc.org/blog/2015/10/ibcs-using-data-visualization-best-practices-to-improve-business-communication, [Online: accessed 2018-03-04].

[178] M. Hart, M. Sparkman, JiayueHu, C. Caserio, and A. Saxton, "Best design practices for reports and visuals (whitepaper) - Power BI | Microsoft Docs," May 2018. [Online]. Available: https://docs.microsoft.com/en-us/power-bi/power-bi-visualization-best-practices

[179] "'Getting around in Power BI service - Power BI'," [Online]. Available: https://docs.microsoft.com/en-us/power-bi/service-the-new-power-bi-experience, [Accessed 2018-03-04].

[180] Tableau, "Tableau.com, 'best Practices for Effective Dashboards'," [Online]. Available: https://onlinehelp.tableau.com/current/pro/desktop/en-us/dashboards_best_practices.html, [Accessed 2018-03-04], May 2018.

[181] T. Team, "Tableau.com, 'visual Analysis Best Practices: A Guidebook'," [Online]. Available: https://www.tableau.com/learn/whitepapers/tableau-visual-guidebook, [Accessed 2018-03-04].

[182] E. H.-h. Chi, "A taxonomy of visualization techniques using the data state reference model," in *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on.* IEEE, 2000, pp. 69–75.

[183] E. H.-h. Chi and J. T. Riedl, "An operator interaction framework for visualization systems," in *Information Visualization, 1998. Proceedings. IEEE Symposium on.* IEEE, 1998, pp. 63–70.

[184] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The Craft of Information Visualization.* Elsevier, 2003, pp. 364–371.

[185] A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard Iii, and W. Skiff, "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American chemical society*, vol. 114, no. 25, pp. 10 024–10 035, 1992.

[186] "Dewey, R. 'The Whole is Other than the Sum of the Parts' | in Chapter 04: Senses | from Psychology: An Introduction by Russ Dewey," [Online]. Available: http://www.intropsych.com/ch04_senses/whole_is_other_than_the_sum_of_the-_parts.html, [Accessed 2018-03-07].

[187] G. M. Heider, "More about Hull and Koffka." *American Psychologist*, vol. 32, no. 5, pp. 383–383, 1977.

[188] M. Soegaard, "Gestalt principles of form perception," [Online]. Available: https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/gestalt-principles-of-form-perception, [Accessed 2018-03-07].

[189] S. Bradley, "Design Principles: Visual Perception And The Principles Of Gestalt," [Online]. Available: https://www.smashingmagazine.com/2014/03/design-principles-visual-perception-and-the-principles-of-gestalt/, [Accessed 2018-03-07], 2014.

[190] S. Lehar, *The world in your head: a gestalt view of the mechanism of conscious experience.* Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers, 2003, oCLC:

52051454. [Online]. Available:
http://public.eblib.com/choice/publicfullrecord.aspx?p=452317

[191] "S. lehar, gestalt Isomorphism and the Primacy of Subjective Conscious Experience: A Gestalt Bubble Model," Feb. 2012. [Online]. Available:
https://web.archive.org/web/20120217033046/http://sharp.bu.edu/~slehar/webstuff/bubw3/bubw3.html

[192] "Webpagefx.com, 'Gestalt Principles | Design Strategy | Tips & Tricks'," [Online]. Available: https://www.webpagefx.com/blog/web-design/gestalt-principles-applied-in-design/, [Accessed 2018-03-07].

[193] J. J. Imhoff and S. P. Cutler, "Interpol: Extending law enforcement's reach around the world," *FBI L. Enforcement Bull.*, vol. 67, p. 10, 1998.

[194] M. Anderson, *Policing the world: Interpol and the politics of international police co-operation*. Clarendon Press Oxford, 1989.

[195] C. Firestone and B. J. Scholl, "Enhanced visual awareness for morality and pajamas? perception vs. memory in 'top-down'effects," *Cognition*, vol. 136, pp. 409–416, 2015.

[196] "C. Firestone and b. j. Scholl, cognition does not affect perception: Evaluating the evidence for top-down effects," *Behavioral and brain sciences*, vol. 39, 2016.

[197] R. Brünken, S. Steinbacher, J. L. Plass, and D. Leutner, "Assessment of cognitive load in multimedia learning using dual-task methodology." *Experimental psychology*, vol. 49, no. 2, p. 109, 2002.

[198] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," in *Proceedings of the 2003 SIAM International Conference on Data Mining*. SIAM, 2003, pp. 25–36.

[199] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*. Australian Computer Society, Inc., 2005, pp. 333–342.

[200] Y. Liao and V. R. Vemuri, "Use of k-nearest neighbor classifier for intrusion detection," *Computers & security*, vol. 21, no. 5, pp. 439–448, 2002.