

SECURITY VISUALIZATION INFRASTRUCTURES, TECHNIQUES, AND
METHODOLOGIES FOR IMPROVED ENTERPRISE SECURITY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

F.FERDA ÖZDEMİR SÖNMEZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JUNE 2019

THESIS TITLE

Submitted by F. FERDA ÖZDEMİR SÖNMEZ in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in The Department of Information Systems Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics**

Prof. Dr. Yasemin YARDIMCI ÇETİN
Head of Department, **Information Systems, METU**

Assoc. Prof. Dr. Banu GÜNEL KILIÇ
Supervisor, Information Systems

Examining Committee Members:

Assoc. Prof. Dr. Altan KOÇYİĞİT
Information Systems, Middle East Technical University

Assoc. Prof. Dr. Banu GÜNEL KILIÇ
Information Systems, Middle East Technical University

Prof. Dr. Şeref SAĞIROĞLU
Computer Engineering, Gazi University

Assoc. Prof. Dr. Pekin Erhan EREN
Information Systems, Middle East Technical University

Assoc. Prof. Dr. Sevil ŞEN
Computer Engineering, Hacettepe University

Date:

17.06.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : F. Ferda Özdemir Sönmez

Signature : _____

ABSTRACT

SECURITY VISUALIZATION INFRASTRUCTURES, TECHNIQUES, AND METHODOLOGIES FOR IMPROVED ENTERPRISE SECURITY

Özdemir Sönmez, F. Ferda

Ph.D, Department of Information Systems

Supervisor: Assoc. Prof. Dr. Banu Günel Kılıç

June 2019, 250 pages

This thesis focuses on providing designs to allow monitoring of the security status of enterprises at the organization level. The audience of this research is all enterprise level IT and security experts, and the other users who may be engaged in the security visualization designs, including the top level management. Numerous tools and programs are being used in organizations to analyze and overcome security vulnerabilities. However, the outputs of these programs are rarely understood clearly. During this study, existing security visualization requirements and designs along with the corresponding technologies used for security visualization were examined. For the sake of being user-centric, a visualization requirements survey was held. The results of the literature review and the survey were converted to a substantial requirement set for a generic enterprise security visualization infrastructure. This infrastructure was then implemented using industry's best standards and the contemporary big data solutions. The resulting design was validated through the use of expert reviews. Later, one of the favorite security visualization subjects for the enterprises, namely web application security was handled. A dashboard type holistic design to visualize black-box vulnerability test results was proposed along with forty plus metrics and measures. SIEM systems were also examined for their custom data visualization capabilities in parallel to this part of the study. Finally, security management related issues for the organizations was focused. In this part of the study, a decision support system for the optimization of security costs which relies on analytical methods and uses treemap type visualizations to visualize the threats, risks, corresponding precautions, and the costs was proposed. A real-world case study was used to demonstrate the benefits of this system.

Keywords: Security Visualization, Enterprise Security, Big Data, SIEM, Decision Support System

ÖZ

GELİŞTİRİLMİŞ KURULUŞ GÜVENLİĞİ İÇİN GÜVENLİK GÖRSELLEŞTİRME ALTYAPI, TEKNİK VE METODOLOJİLERİ

Sönmez Özdemir, F. Ferda

Doktora, Bilişim Sistemleri Bölümü

Tez Yöneticisi: Doç. Dr. Banu Günel Kılıç

Haziran 2019, 250 sayfa

Bu tez, kuruluşların işletme seviyesindeki güvenlik durumunun izlenmesini sağlayacak tasarımlar sağlamaya odaklanmıştır. Bu araştırmanın seyircileri tüm kurumsal düzeyde teknoloji ve güvenlik uzmanları ve güvenlik durumu görselleştirme tasarımları ile ilgili olabilecek üst düzey yöneticiler olabilir. Güvenlik açıklarını analiz etmek ve üstesinden gelmek için kuruluşlarda çok sayıda araç ve program kullanılmaktadır. Bununla birlikte, bu programların çıktıları nadiren açıkça anlaşılmalıdır. Bu çalışma sırasında, mevcut güvenlik görselleştirme gereksinimleri ve ilgili teknolojileri ile birlikte mevcut güvenlik görselleştirme tasarımları incelenmiştir. Kullanıcı merkezli olma adına, görselleştirme gereksinimleri anketi düzenlenmiştir. Literatür taramasının ve anketin sonuçları, genel bir kurumsal güvenlik görselleştirme altyapısı için belirlenmiş bir gereksinim setine dönüştürülmüştür. Bu altyapı, endüstrinin en iyi standartları ve çağdaş büyük veri çözümleri kullanılarak gerçekleştirilir. Sonuçta ortaya çıkan tasarım, uzman incelemeleri kullanılarak doğrulanmıştır. Daha sonra, işletmeler için favori güvenlik görselleştirme konularından biri olan web uygulaması güvenliği tasvir edilmiştir. Kara kutu kırılabilirlik testi sonuçlarını görselleştirmek için bir pano tipi bütünsel tasarım, kırk artı ölçümle birlikte önerildi. Çalışmanın bu bölümüne paralel olarak SIEM sistemleri de özel veri görselleştirme yetenekleri açısından incelenmiştir. Son olarak, kuruluşlar için güvenlik yönetimi ile ilgili konular ele alındı. Bu çalışmanın bu bölümünde, analitik yöntemlere dayanan ve tehditleri, riskleri, karşılık gelen önlemleri görselleştirmek için treemap tipi görselleştirmeler kullanan güvenlik maliyetlerinin optimizasyonu için bir karar destek sistemi önerilmiştir. Bu sistemin faydalarını göstermek için gerçek dünyadan bir vaka çalışması kullanılmıştır.

Anahtar Sözcükler: Güvenlik Görselleştirme, Kurumsal Güvenlik, Büyük Veri, SIEM, Karar Destek Sistemi

This dissertation is lovingly dedicated to
my father, M.Yalçın Özdemir for his ongoing love and support
and to
my mother, F. Semiha Özdemir.
Her support, encouragement, and constant love have sustained
me throughout my life, although she is not here to give me
strength and support now. I always feel her presence that used
to urge me to strive and to achieve my goals in life.

ACKNOWLEDGMENTS

I would like to express my sincere thanks and appreciation to

- my academic advisor, Doç.Dr. Banu Günel Kılıç for her belief in this work, support and encouragement throughout this study,
- my thesis follow up committee members Prof. Dr. Şeref Sağıroğlu and Assoc. Prof. Dr. Altan Koçyiğit for kindly accepting to take part in my jury and for their insightful comments and encouragement throughout this study, which enriched my thesis,
- Associate Prof. Dr. Aysu Betin Can, for her valuable comments and suggestions, and feedback for part of this study,
- Information Systems Department Head, Prof. Dr. Yasemin Yardımcı for her continuous care for the students which increases the love and fidelity to our department,
- I must thank to all my friends for listening, offering advice, and supporting me through this entire process.
- My special thanks are to the experts of information security who participated in the evaluation surveys of the study.
- Finally, I would like to thank my husband, parents, my sibling, and my entire family who are in the secret space of my heart, for their love, encouragement, motivation, support, and patience.

TABLE OF CONTENTS

ABSTRACT	vi
ÖZ	vii
ACKNOWLEDGMENTS.....	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xiv
LIST OF FIGURES.....	xv
LIST OF ABBREVIATIONS	xviii
1 INTRODUCTION	1
1.1 Significance and Motivation.....	4
1.2 Scope	5
1.3 Research Questions	6
1.4 Road Map	7
1.5 Thesis Contributions.....	14
1.6 Publications	20
1.7 Organization of the Thesis.....	21
2 LITERATURE REVIEW.....	23
2.1 Introduction to Literature Review	23
2.2 Background.....	25
2.2.1 Security Visualization	25
2.2.2 Design Issues.....	26
2.3 Methodology of the Review Study	28
2.4 Extended Review of the Selected Studies	30
2.4.1 Issues, Controversies, Problems.....	30
2.4.2 Examination Results.....	31
2.5 Validation of Security Visualization Studies	57
2.6 Future Research Directions	59
2.7 Concluding Remarks for the Literature Review.....	60

3	SECURITY VISUALIZATION REQUIREMENTS SURVEY.....	65
3.1	Introduction to the Survey	65
3.2	The Need for Security Visualization Requirements Analysis	66
3.3	Methodology of the Survey Study.....	67
3.3.1	Survey.....	68
3.3.2	Participants	68
3.4	Analysis and Results.....	69
3.4.1	Quantitative Results at a Glance	69
3.4.2	Further Quantitative Results.....	77
3.4.3	Qualitative Results	79
3.5	Concluding Remarks for the Survey	81
4	DESIGN AND DEVELOPMENT OF A GENERIC SECURITY VISUALIZATION INFRASTRUCTURE PROTOTYPE.....	83
4.1	Introduction to Design and Development of a Generic Security Visualization Infrastructure Prototype Study	83
4.2	Description of an Enterprise Network as a Ubiquitous Environment	86
4.3	Functional Requirements for an Enterprise Visualization System with Feedback from Users.....	88
4.4	Initial Design Features.....	93
4.4.1	Extensible Display Type Library and Dashboard Design	96
4.5	Evolving the Initial Design Using Big Data Technologies	97
4.5.1	Big Data Technologies	98
4.5.2	Big Data Technologies Related Design Decisions For Generic Enterprise Security Visualization Solution.....	100
4.5.3	Security Concerns.....	103
4.6	Results	105
4.7	Discussion (Including validation efforts)	110
4.8	Concluding Remarks for The Design and Development of a Generic Enterprise Security Visualization Prototype	113
5	APPLICATION SECURITY VISUALIZATION	115
5.1	Introduction to Application Security Visualization Study.....	115
5.2	Related Work.....	117
5.3	Holistic Web Application Security Vulnerability Visualization, HWAS-V .	121
5.3.1	Metric Definition Process and Data Measures	122

5.3.2	Web Application Visualization Measures/Metrics Based on Common Vulnerability Scan Outputs and Related Data	126
1)	Base Measures/Metrics - Measures Based on Vulnerability Scanner Tools 127	
2)	Metrics/ New Development-Bug Fix-Mainteanance Effect	128
3)	Effects of Previous Measurements and Time	129
4)	Metrics- Measures/Application Properties Effect.....	130
5)	Metrics/Classification Effect	131
6)	Metrics/Standards- Lists Effect	132
7)	Metrics/Protection Systems Effect.....	132
5.3.3	Visualization of Metrics.....	133
5.4	Case Study and HWAS-V	137
5.5	Discussion.....	145
5.6	Concluding Remarks for Application Security Visualization	148
6	EVALUATION OF SIEM SYSTEMS FOR CUSTOM DATA VISUALIZATION.....	149
6.1	Introduction to the Evaluation of SIEM Systems for Custom Data Visualization Study.....	149
6.2	Methodology.....	150
6.2.1	Selecting the SIEM Systems to be Evaluated	150
6.2.2	Evaluation Scenario	152
6.3	Results	153
6.3.1	Manage Engine Event Log Analyzer	153
6.3.2	Splunk	154
6.3.3	Rapid7 InsightIDR	155
6.3.4	Solar Winds Log and Event Manager	155
6.3.5	Micro Focus ArcSight	156
6.3.6	AlienVault	157
6.4	Discussion.....	157
6.5	Concluding Remarks for Evaluation of SIEM Sytems.....	161
7	A DECISION SUPPORT SYSTEM FOR OPTIMAL SELECTION OF ENTERPRISE INFORMATION SECURITY PREVENTATIVE ACTIONS ALONG VISUALIZATION	163

7.1	Introduction to Decision Support System For Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization Study	163
7.2	Literature Review for Analytical Methods for Security Domain	165
7.3	Methodology.....	170
7.4	Optimal Information Security Preventative Actions Along Visualization	173
7.4.1	Risk Assesment of Threats	174
7.4.2	Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization.....	176
7.4.3	Visualization of Outputs.....	178
7.5	OPISPA-V and Case Study	180
7.6	Discussion.....	185
7.6.1	Contributions	185
7.6.2	Limitations and Future Research.....	186
7.7	Concluding Remarks for a Decision Support System for Optimal Selection of Enterprise Information Security Preventative Actions along Visualization	187
8	CONCLUSION AND FUTURE WORK.....	189
8.1	Conclusions	189
8.2	Future Work.....	191
	REFERENCES	193
	APPENDIX	215
	A-Security Visualization Requirements Survey.....	215
	B-Survey Permission Form	235
	C-Larger Versions of Small Figures	236

LIST OF TABLES

Table 1- Research Questions.....	6
Table 2 - Thesis Facts	10
Table 3 - Thesis Contributions.....	14
Table 4- Thesis Publications	20
Table 5 - Types of Displays	25
Table 6 - Groupings of Use-cases	31
Table 7 - Apriori Rule Generation for Enterprise Software Systems	76
Table 8 - Apriori Rule Generation for Enterprise Security Systems	76
Table 9 - Apriori rule generation for other enterprise infrastructure elements	77
Table 10 - K-means clustering results of threat-data source, data attribute associations	78
Table 11 - Strategies and Suggestions	80
Table 12 - Examination of Data Structures for Suitability to Big Data Technologies	101
Table 13- Traceability of the Requirements & Detailed Design Features.....	105
Table 14 - Sample Data Sources and Their Representation in the Proposed Design ..	107
Table 15 - Expert Judgment Semi Structured Interview Results	112
Table 16 - Base Measures/Metrics Based on Vulnerability Scanner Tools.....	125
Table 17 - Metrics-Measures/New Developments-Bug Fix Maintenance Effect	128
Table 18 - Metrics/Measures Based on Effects of Previous Measurements and Time	129
Table 19 - Metrics-Measures/Application Properties Effect	129
Table 20 - Metrics/ Classification Effect	130
Table 21- Metrics/Standards -Lists Effect	132
Table 22- Metrics/Protection Systems Effect	132
Table 23 – Case Study Participant Characteristics	143
Table 24 - Evaluation Questions and Results	143
Table 25 - Summary of Feature Evaluation T-test Results	144
Table 26 - SIEM Configuration Table	158
Table 27 - Evaluation Summary	159
Table 28 -Types of Visualizations Available in OPISPA-V.....	173
Table 29 - Random Index(RI).....	176
Table 30 - Preventative Actions.....	177
Table 31 - Threatening Actions for the Case Study Based on (Hunter, 2012)	181
Table 32 - Acceptable Comparison Values for AHP Severity and Likelihood Comparison	183
Table 33 - Counter Measures Selected for the Case Study.....	184
Table 34 - The Results of Running OPISPA-V for Fixed Budget Optimization for Two Budget Options.....	184
Table 35- The Results of Running OPISPA-V for the Case Study for Fixed Acceptable Risk Level Optimization for Two Risk Level Options	185

LIST OF FIGURES

Figure 1 - Visualization focus points of the thesis	5
Figure 2 - Thesis flow	8
Figure 3 - Short summary of thesis proposal phase	9
Figure 4 - Short summary of literature search phase.....	12
Figure 5 - Short summary of security visualization requirements survey phase.....	13
Figure 6 - Short summary of design and implementation of enterprise security visualization phase	15
Figure 7 - Short summary of identification of visualization topics phase.....	16
Figure 8 - Short summary of examination of SIEM systems for custom visualization generation capabilities phase.....	17
Figure 9 - Short summary of web application security visualization phase.....	18
Figure 10 - Short summary of visualization of enterprise security risks and costs phase	19
Figure 11 - Short summary thesis writing and closure phase.....	20
Figure 12 - Distribution of selection of data sources for visualization designs over years	37
Figure 13 - Graphical illustrations of simplified display properties- Part 1.....	41
Figure 14 - Graphical illustrations of simplified display properties – Part 2	42
Figure 15 - Graphical illustrations of simplified display properties – Part 3	47
Figure 16 - Graphical illustrations of simplified display properties –Part 4	48
Figure 17 - Distribution of display types of security visualization designs over years..	54
Figure 18 - Types of user interactions found in security visualization designs	56
Figure 19 - Validation data sources.....	58
Figure 20 - Primary sectors of the attendees	68
Figure 21 - Importance of data sources for the organizations	70
Figure 22 - Security visualization use-cases	70
Figure 23 - Evaluation of security visualization use-cases according to the enterprise size (number of employees).....	71
Figure 24 - Origin of existing security visualization solutions in the enterprises	71
Figure 25 - Most popular security visualization solutions in the enterprises.....	72
Figure 26 - Commonly used enterprise software solutions.....	72
Figure 27 - Hardware, networking and system components that are part of the infrastructures.....	73
Figure 28 - Associations of threats to data sources and data attributes.....	74
Figure 29 - Security visualization design issues	75
Figure 30 - Popular display types.....	75
Figure 31 - Popular security analyses.....	76
Figure 32 - Typical IT infrastructure of an organization.....	87
Figure 33 - Class diagram of the proposed structure.....	92

Figure 34 - States and activities of the proposed system	95
Figure 35 - a) XHTML content for a JavaScript-based display b) ContentAdapter structure for Flotr JavaScript library based Bar Chart visualization c) Sequence diagram for visualization display in dashboard form.....	96
Figure 36 - Methodology to find integration points with big data technology	97
Figure 37 - a) The first design architecture, b) The second design architecture, c) The third design architecture, d) Streaming details for third design e)Evolution summary	104
Figure 38 - Feedback loops	109
Figure 39 - The state of web application vulnerabilities between 2014 and 2017 (Imperva, 2018)	116
Figure 40 - Multi-project multi-phase vulnerability scan results.....	123
Figure 41 - Data structure and attributes of the proposed model.....	124
Figure 42 - Visualization mechanism	135
Figure 43 - General information dashboard.....	137
Figure 44 - Vulnerability scan results dashboard.....	138
Figure 45 - URL based scan details dashboard.....	139
Figure 46 - Alerts dashboard.....	139
Figure 47 - Alerts and data from security protection systems	140
Figure 48 - New developments, bug fixes, repeated alerts, fixed alerts	140
Figure 49 - Standards and Scan Rules.....	141
Figure 50 - Standards and Alerts.....	142
Figure 51 - Gartner magic quadrant for SIEM systems (Adapted from Gartner 2017) (Nicolett & Kavanagh, 2013) Systems marked in red color were included in the evaluation.	151
Figure 52 - Stepwise description of visualization of custom data using SIEM tools ..	152
Figure 53 - Flow of activities in the proposed model	169
Figure 54 - Multi-level threat model.....	175
Figure 55 - Sample visualization of costs (a) and risks (b) distributions to threats using OPISPA-V.....	179
Figure 56 - Sample visualization of cost (a) and risks (b) distributions to precautions	180
Figure 57 - AHP calculation and consistency check for threat groups.....	182
Figure 58 - AHP calculation for threats	183
Figure 59 - Threats and precautions model for the case study.....	188
Figure 28 - Associations of threats to data sources and data attributes (Large Scale).	236
Figure 35 - a) XHTML content for a JavaScript-based display b) ContentAdapter structure for Flotr JavaScript library based Bar Chart visualization c) Sequence diagram for visualization display in dashboard form (Large Scale).....	238
Figure 37 - a) The first design architecture, b) The second design architecture, c) The third design architecture, d) Streaming details for third design e)Evolution summary (Large Scale).....	240
Figure 43 - General information dashboard (Large Scale)	241
Figure 44 - Vulnerability scan results dashboard (Large Scale).....	242
Figure 45 - URL based scan details dashboard (Large Scale)	243

Figure 46 - Alerts dashboard (Large Scale)244
Figure 47 - Alerts and data from security protection systems (Large Scale)245
Figure 48 - New developments, bug fixes, repeated alerts, fixed alerts (Large Scale) 246
Figure 49 - Standards and Scan Rules (Large Scale)247
Figure 50 - Standards and Alerts (Large Scale)248

LIST OF ABBREVIATIONS

2-D	Two Dimensional
3-D	Three Dimensional
3DSVAT	3D Stereoscopic Vulnerability Assessment Tool for Network Security
ACID	Atomicity, Consistency, Isolation, and Durability
Avisa	IDS Alert Visualization and Monitoring through Heuristic Host Selection
Avisa2	IDS Alert Visualization and Monitoring through Heuristic Host Selection 2
AS	Autonomous System
BGPlay	BGP Routing Visualization,
BSA	Business Software Alliance
C&C	Command and Control
CAIDA	Center for Applied Internet Data Analysis
CCSVis	Cylindrical Coordinates Security Visualization
CIDEE	Cisco Intrusion Detection Event Exchange
Clique	Correlation Layers for Information Query and Exploration
CSV	Comma Separated Values
CVE	Common Vulnerabilities and Exposures
CWE	Common Weakness Enumeration
DARPA	Defense Advanced Research Projects Agency
DDOS	Distributed Denial of Service
DMZ	Demilitarized Zone
DoD	Department of Defense
Elisha	Experimental Visual Anomaly Detection
Enavis	Enterprise Network Activities Visualization
FTP	File Transfer Protocol
HnMaps	Hierarchical Network Maps
HViz	HTTP(S) Traffic Aggregation and Visualization
IDE	Integrated Development Environment
IDGraphs	Intrusion Detection and Analysis Using Histograms
IDMEF	Intrusion Detection Message Exchange Format
IDS	Intrusion Detection System
IDTk	Information Visualization Tool for Intrusion Detection
IETF	Internet Engineering Task Force
IoT	Internet of Things
IP	Internet Protocol

IT	Information Technology
ITGI	IT Governance Institute
KDD	Knowledge Discovery and Data Mining
LAN	Local Area Network
MAWI	Measurement and Analysis on the WIDE Internet
MOME	Meta-database for monitoring and measurement
Nagios	Nagios Core
NetsecRadar	Real-Time Visualization System for Network Security
NetVis	Network Security Management Visualization Tool based on Treemap
NFC	Near Field Communication
Niva	Network Intrusion Visualization Application
NV	Nessus Vulnerability Visualization for the Web
OWASP	Open Web Application Security Project
P3D	Parallel 3D Coordinate Visualization
PCI	Payment Card Industry
Portall	Visualizing Packet-Process Correlation
PortVis	Tool for Port-Based Detection of Security Events
RFID	Radio Frequency Identification
SABSA	Sherwood Applied Business Security Architecture
SDEE	Security Device Event Exchange
SIEM	Security Information and Event Management
SMTP	Simple Mail Transfer Protocol
SOM	Self Organizing Map
Svision	Network Host-centered Anomaly Visualization Technique
SYNEMA	Visual Monitoring of Network and System Security Sensors
Tamp	Threshold and Merge Prefixes
TNV	Time-based Network Visualizer
TrustVis	Trust Visualisation Service for Online Communities
Tudumi	Information Visualization System for Monitoring and Auditing Computer Logs
Vafle	Visual Analytics of Firewall Log Events
Visual	Visual Information Security Utility for Administration Live
VPN	Virtual Private Network
W3C	World Wide Web Consortium
WAN	Wide Area Network
WASC	Web Application Security Consortium
WIDE	Widely Integrated Distributed Environment
XML	Extensible Markup Language
ZAP	Zed Attack Proxy

CHAPTER 1

INTRODUCTION

Security visualization is a domain that links information security and cyber security fields to visualization technologies to improve the premier fields. This domain emerged due to the increased amount of information security or cyber security related data collected and wasted without proper examination, and late responses resulted by improper examination methods.

This study does not aim to develop a dedicated visualization software. The efforts will focus on proposing designs which are suitable for multiple data sources and which would serve for multiple use-cases. Therefore, existing visualization tools and libraries will be exploited whenever possible. This approach is more effective in improving enterprise security through visualization, compared to focusing on a single popular dataset or a single use-case, and assigning a complex time-consuming development process following the design phase for a visualization software.

Security visualization is the act of using information visualization techniques to help the decision-making process for security analysts. It provides situational awareness. It offers new representations of security data to increase understandability and help efficient processing of data. In general, there is a tendency to use the same type of display types for same use cases, or same type of display types for the data in similar formats. While this is the result of a consolidated learning in most cases, it may be beneficial to find alternative combinations of use cases, display types and data attributes for novel security visualization designs.

A solid knowledge and understanding on the set of actions treating the information systems of an enterprise are necessary in order to conduct the requirements of this study. Analysing data with the aim of security data classification and event identification is beyond the scope of this study. Existing classifications provided by protection systems, such as intrusion detection systems, vulnerability scanner systems, and alarm files are used as input data sources during the design of the visualization systems.

In order to gain solid knowledge and understanding of the domain, the existing work is classified according to display types, use cases, and data sources. The objective of this effort is to classify existing work according to their similarities, and by doing so to find out gaps such as data types which are seldomly used for security visualization purposes, new ideas of combining data coming from multiple sources, display types commonly used for some particular scenarios, which may also be suitable for some other scenarios.

Another effort, to understand and gain knowledge for the security visualization requirements of the enterprises was the security visualization requirements survey which was prepared to find answers related to visual representation of different use cases. The attendees of this survey were people with enterprise security expertise both from the academia and the industry. IT department staff including security experts and system admins from a diverse range of enterprises were called out for the survey.

The aim of the survey was to understand the existing situation in terms of utilization of security visualization solutions in the enterprises and to find out new requirements. For this purpose, the survey consisted of questions related to existing security analysis methods which encapsulate security visualization tools and techniques, data sources which are collected and/or, stored and/or, analysed as a part of security analyses methods, infrastructure of the enterprise including software, hardware and system components, security analyses methods which may be extended by including security visualization methods and the user practices and expertise.

One of the main difficulties tackled during this study was the deciding on the data sets to be visualized. There has been some attempts to standardize the log files however none of these attempts has been successful yet. Thus, the data formats offered by these standardization attempts are not internationally recognized. Using the survey outputs, both numerical importance values were assigned to the security data sources and the areas which happen to be neglected or given less importance were examined.

The survey results and the investigation of earlier literature work resulted in the definition of a set of enterprise security visualization requirements. Later on, these requirements were converted to a Enterprise Security Log Data Visualization System design. Although there are many visualization tools and data parsers, it is difficult to combine different tools. Most of the time, it is necessary to use multiple softwares/tools for multiple file types and this requires substantial amount of data preparation work.

The targeted visualization system of this study for the enterprises is not planned to be rigidly attached to the log file type. It may be improved by including new or additional log file types in time. Initially, writing parser programs for log files was not the aim of this work. Initial goal was to adopt suitable parser programs. However, it was seen that the existing parser programs do not suffice. Thus, data structures and corresponding parsers to enable parsing various kinds of security data were provided.

As part of this work, a generic log file reader and visualization system named EntSecVis which mainly depends on declarative data, which is the meta data of log files to be visualized was designed. This design has some parts which are totally generic, such as, data reading from various formats, storing various types of data with different number of attributes in a data store and querying and grouping the data in a declarative fashion. This design has also some other parts which are not fully generic but standardized in a way that data visualization scripts and applets from various sources and having various formats can be used in a standardized way using the generic data definitions.

After the generic security visualization architecture design and implementation, the focus was shifted to different kinds of web based applications: the Static Web (Shallow Web), the Dynamic Web (Deep Web), the Wisdom Web (Web 2.0), Mobile Web, and Semantic Web. Existing research on security visualizations mostly focuses on network security. Only a small number of studies considers the web application security vulnerabilities and possible visualization alternatives. Thus, this part of the research investigates new data and visualization alternatives to improve web application securities. As the number of web applications and corresponding number and sophistication of the threats increase each year, it is important to find new tools which are efficient and accessible for both expert and non-expert users. For this purpose, data sources which may be easily associated and which are commonly available for the majority of the web applications were examined. Later a set of web application security measures and metrics were defined. In the end, a set of dashboards which are presented in a related manner were prepared. In order to demonstrate the dashboard, sample data was generated using OWASP Zed Attack Proxy vulnerability scanner tool. This design allows investigation of about 40 metrics/measures for the multi-project multi-phase environment, which will enhance its benefits if the user aims to monitor a single security analysis result or consecutive analysis results and the changes between them for one or more projects.

Although some of the findings of the survey and literature search apply to the big data visualization systems, both the survey and the literature search did not have a specific focus on big data technologies. In parallel to web application security visualization project, the contemporary big data technologies were examined to improve the generic enterprise security visualization design. As a result of this investigation, big data technologies to integrate with the existing design were identified. Later, ways to make these integrations with the existing structure were proposed. The final improved design was presented using graphical ways together with brief descriptions of the used technologies. As a part of validation efforts, a detailed expert evaluation study has been conducted for the ultimate design.

Types and complexities of information security related vulnerabilities are growing rapidly and presenting numerous challenges to the enterprises. One of the key challenges is to identify an optimal set of precautions with a limited budget. Despite the fact that the majority of enterprises have a budget constraint for installing and maintaining the protection systems, majority of the previous work only focused on prioritization of

vulnerabilities and did not consider the budget constraints. Literature investigation resulted in the finding that the existing security visualization studies mainly neglected the security management issue. Due to this finding, a part of this research includes a method based on analytical hierarchical process and linear programming techniques to distribute a fixed amount of budget among precautions, while maximizing the amount of risk prevented. Visualization is part of this study. The resulting threats, precautions, costs, and risks relations were visualized in eight different visualization designs based on a treemap display.

1.1 Significance and Motivation

Security visualization field has been first encountered in the very beginning of 2000's. When security data sources are examined, it can be seen that the variety of data sources is very high. Investigation of existing studies and the data sources resulted in the conclusion that the number of metrics used in the security visualization domain is very low. Same measures (data attributes from security data sources) and metrics are visualized again and again, repeatedly.

This thesis is planned to touch many aspects of enterprise security visualization concepts. Thus, it does not cover a restrained issue. On the contrary, it is aimed to provide contributions distributed in various levels both technically and conceptually. Figure 1 shows the visualization focus points of this thesis. Initially, the log files coming from infrastructure elements and application security tools were focused. Later on, security management topics were included in the thesis scope.

The starting point was a known fact, which is the difficulty of manual examination of security log files. Security visualization domain has emerged due to this difficulty. It is very important to improve security level of the enterprises, and security visualization is one effective tool to help this improvement. Although there are numerous security visualization studies. There are only a very few studies focusing on enterprise security visualizations. Even, the requirements of enterprise security visualization have not been examined well yet.

The review papers written so far in the security visualization domain focus on a very limited number of works. The extended summary of security visualization designs, which was provided as part of this thesis, may help researchers who want to solve security visualization problems by applying novel designs, may guide those who investigate current trends in the security visualization domain, and may be used for educational purposes.

One important aspect of this thesis is its wholistic approach to enterprise security visualization topic. While proposing new metrics and new infrastructure designs, benefits of commonly used enterprise security monitoring systems, such as SIEM systems were also examined to some extent.

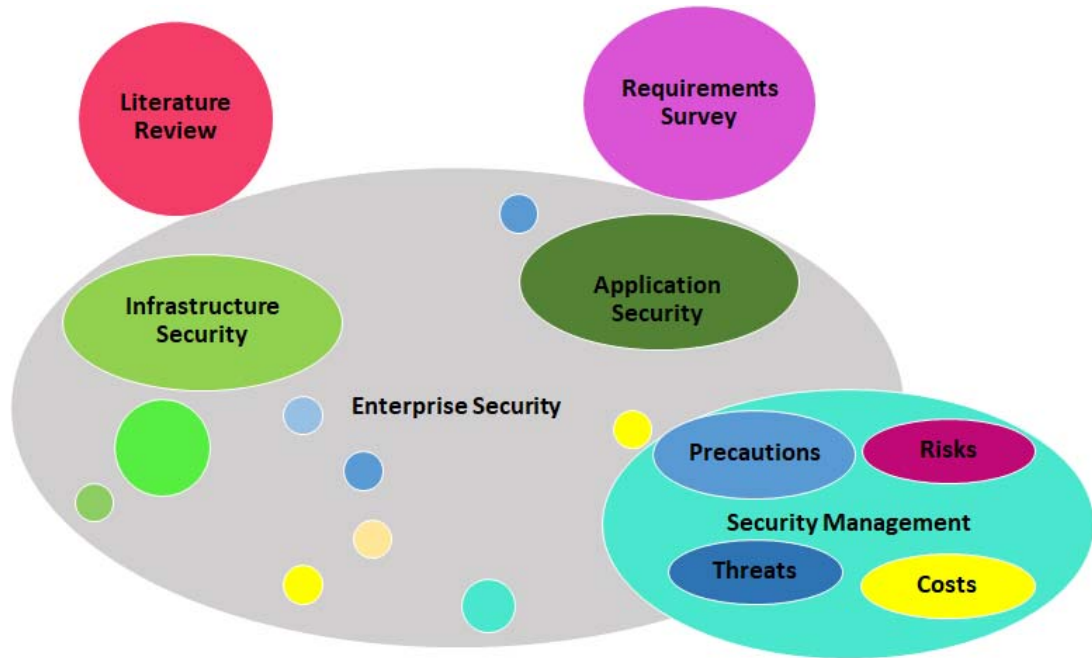


Figure 1 - Visualization focus points of the thesis

The earlier security visualization solutions for the enterprises did not have an explicit aim to form an enterprise security visualization knowledgebase. The enterprise security knowledgebase design based on user feedbacks, and threat, vulnerability, and display type classifications, which was presented as part of generic security visualization solution, is an innovative approach. This design may be improved by clever feedback processing mechanisms forming an enterprise security visualization cyber-physical structure.

The thesis provided ways to leverage security management related elements. Besides the data coming from infrastructure elements, and applications (commonly through log files), the data related to the examination of other security management issues, namely threats, precautions, risks, and costs, were also included in the thesis content.

1.2 Scope

Enterprise security is a broad area. In this study, it is aimed to provide visualization solutions to improve the security of the enterprises. However, some of the security related issues will be outside the scope of this work due to resource and time limitations. The use-cases which have highly changeable structures from domain to domain or from organization to organization is out of scope of this research. One example for this kind of use-case is identity management. The data sources generated as a result of some

unique requirement for an enterprise or due to a specialized background knowledge, such as smart card applications, cryptographic techniques, and biometrics are also out of scope of this study due to limited time and resources. The target is to provide design and methodologies which can be repeated or which can be used by most of the enterprises.

1.3 Research Questions

The initial aim of this thesis was to increase the visualization utilization for security tasks in the enterprises. The research questions to fulfill this initial aim arose as the thesis progressed gradually. Table 1 shows the list of the questions which gradually appeared during the research. Figure 2 presents the relations of research questions to the parts of the thesis while pointing out facts detected/achieved during the thesis studies through the assigned fact numbers for each part. The resulting contributions and the related publications for each part are also shown in this figure in a numbered manner. Details of research questions, facts, contributions, and publications are presented in Table 2, Table 3, and Table 4 respectively.

Table 1- Research Questions

Q1-How common is the visualization used in the enterprises for security analysis purposes?
Q2- What are the security data sources, common use-cases, display techniques, and the design issues for security visualization domain?
Q3- What are the requirements for an enterprise security visualization system?
Q4- How to design a data input structure which is suitable to be used and extended for most type of data sources?
Q5- How to design a display mechanism which benefits from the most type of existing display technologies?
Q6- How to design the enterprise security visualization system so that it would form an enterprise security visualization knowledgebase in the long term?
Q7- What are the gaps in the literature for enterprise security visualization?
Q8- What are the display types and difficulties in generating custom visualizations using custom data in the SIEM systems?
Q9- How to monitor and compare vulnerability scan results gathered in subsequent analyses while watching the effects of modifications made to the web application and extending the metrics in existing web application visualization tools?
Q10- How to provide a visualization system which deals with enterprise threats, precautions, risks, and costs?

1.4 Road Map

The research follows a mixed method approach. In each independent part of the study different ways including diverse techniques and data were used to achieve results. Figure 3 to Figure 11 include short summary of each thesis phase. After the proposal phase (Figure 3), the thesis continued with the literature review of the security visualization domain (Figure 4). During the literature review, an extended summary of existing work was given which may help novice researchers find out what has been done so far. The security visualization literature was classified according to use cases, display types and data sources to help researchers find out gaps and alternative combinations of data, display and usage scenarios. Notable features, interactivity and usability of the designs and their validation methods were depicted and a short trend analysis of the security visualization domain was made. The attributes related to data sources, and display types were defined as part of literature review work. The presented use-cases were also grouped into categories in this period. Since the focus point of the thesis is the security visualization for the enterprises, the use-cases were examined based on their suitability to the enterprises and their locational situation in terms of being protective for the enterprises. At the end, the use cases from the literature were categorized as use-cases inside the enterprise, use-cases among the enterprises, and use-cases beyond the enterprise. As part of literature review section, the future research topics of security visualization work was also identified.

A graphical library consisting of 51 images has been prepared by hand drawing after detailed examination of the literature. Later these images were converted to computerized images using Adobe Illustrator software. Although each paper mentioned in the literature work provides one or more images showing either the actual view of the design or graphical illustration of the presented design, it was difficult to capture the visualization attributes which were repeatedly used in the existing designs by examining the provided independent images. While some of the illustrations or images were too complicated and accommodating more than one display property, some others were simpler. Each of these 51 images corresponded to captured incomplex properties commonly used in different display types.

Following the literature review work, a security visualization requirements survey was prepared, to find answers related to visual representation of different use cases (Figure 5). The attendees of this survey were people with enterprise security expertise, from the academia, and the industry. IT department staff including security experts and system admins from a diverse range of enterprises were called out for the survey.

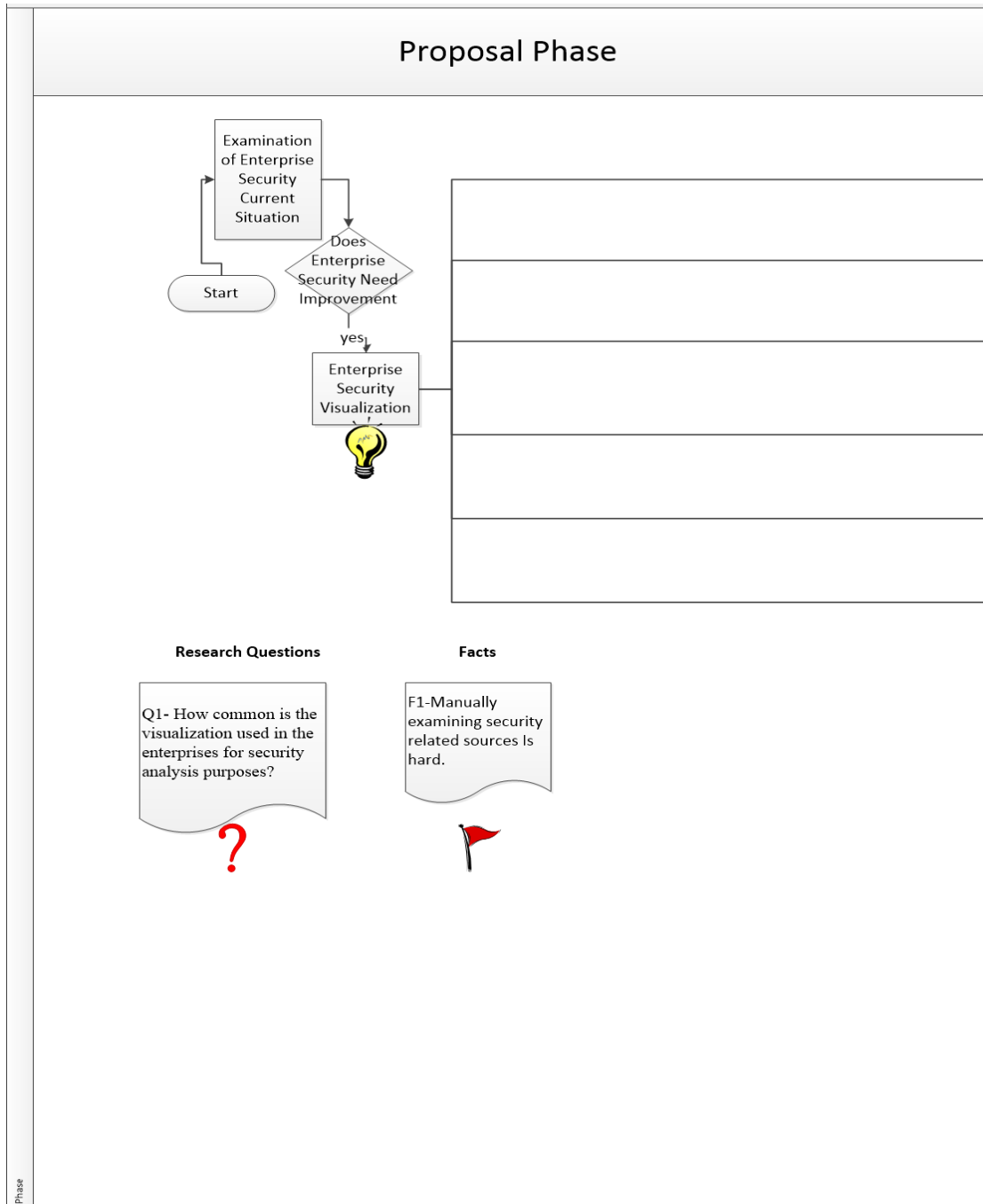


Figure 3 - Short summary of thesis proposal phase

The aim of the survey was to understand the existing situation in terms of use of security visualization solutions in the enterprises and to find out new requirements. For this purpose, the survey consisted of questions related to existing security analysis methods which encapsulate security visualization tools and techniques, data sources which were collected and/or, stored and/or, analysed as part of security analyses methods, infrastructure of the enterprise including software, hardware and system components, security analyses methods which may be extended by including security visualization methods and the user practices and the user expertises.

The survey results and investigation of earlier literature work resulted in the identification of some quantitative and qualitative facts related to enterprise security visualization domain. These facts were converted to a set of solid enterprise security visualization requirements. These requirements were converted to a Generic Enterprise Security Log Data Visualization System design later (Figure 6). This proposed design was developed in phases. Initial design was based on conversion of the requirements identified using industry standard Java technologies based web application design structure. This initial design was ameliorated by integrating various big data technologies, mainly to increase the scalability of the overall system and to improve data collection mechanism of the overall design. The final design also enables investigation of the security data in a near real-time fashion, due to use of third party tools. Some parts of the design were implemented using Java development language and Java development frameworks, Spring, and Hibernate. Some other parts of the design, which are mainly related to the integration with third party tools, remained at a conceptual state, although integration parts were planned as either code or configuration. The validation of the design was made through expert reviews.

Table 2 - Thesis Facts

F1-Manually examining security related sources is hard.
F2-The variety of data sources is high.
F3-Repeatedly same metrics and attributes are visualized in the literature.
F4-Although, in the literature it is repeatedly mentioned that having user-centric requirements is very important, there is no study which includes a comprehensive requirement analysis for the security visualization domain.
F5-Existing security visualization studies focus on network and infrastructure security but lack security management issues.
F6-Existing metrics are not well associated with threats by the users.
F7-Simple charts are most preferable among users.
F8-Dynamic and static web applications are the most widely used enterprise software applications.
F9-Enterprises do not benefit from security visualization systems well.

F10-A visualization infrastructure that can be easily extended is needed to overcome the difficulty of implementing visualizations for new data sources.

F11-The number of metrics in an enterprise security system may be high.

F12-Dashboard type of visualizations encapsulating the various type of display types are most suitable for a large number of metrics.

F13-Use of the third party big data technologies would increase the design quality and would help fulfilling the detected requirement set.

F14-Javascript based visualization mechanisms form the majority of the overall visualization techniques.

F15-Unlike other technologies, JavaScript-based visualization libraries have a structure which allows forming a uniform structure, so that developers can benefit from the various type of JavaScript libraries using the same standardized visualization structure.

F16-There is a gap for the web application vulnerability scan result visualization systems in the literature.

F17-Security management related issues are not given enough importance in the security visualization domain.

F18-SIEM systems should be investigated for their custom visualization generation capabilities.

F19-Gartner makes an evaluation of SIEM systems each year based on scenario and data independent criteria.

F20-SIEM systems are very complex in general and have high configuration costs.

F21-Using proper visualization tools will enable forming a convenient prototype to further evaluate the proposed measures and metrics.

F22-Tableau is a visualization software which enables quick generation of complicated dashboards.

F23-Case studying is a technique which is commonly used for the evaluation of security visualization designs.

F24-Existing information security visualization designs neglect security management related objects to be visualized.

F25-Existing analytical methods related to security management commonly make prioritization of the risks. However, they are not adequate to offer precautions for the associated threats and vulnerabilities.

F26-Although security cost issue is used as a decision parameter, there is no study which provides an end to end mechanism starting from threats to the precautions while spending a fixed amount of security budget in an optimum level.

F27-Every thesis has limitations due to time restrictions.

F28-The end of Ph.D. study is the start of a lifelong learning period.

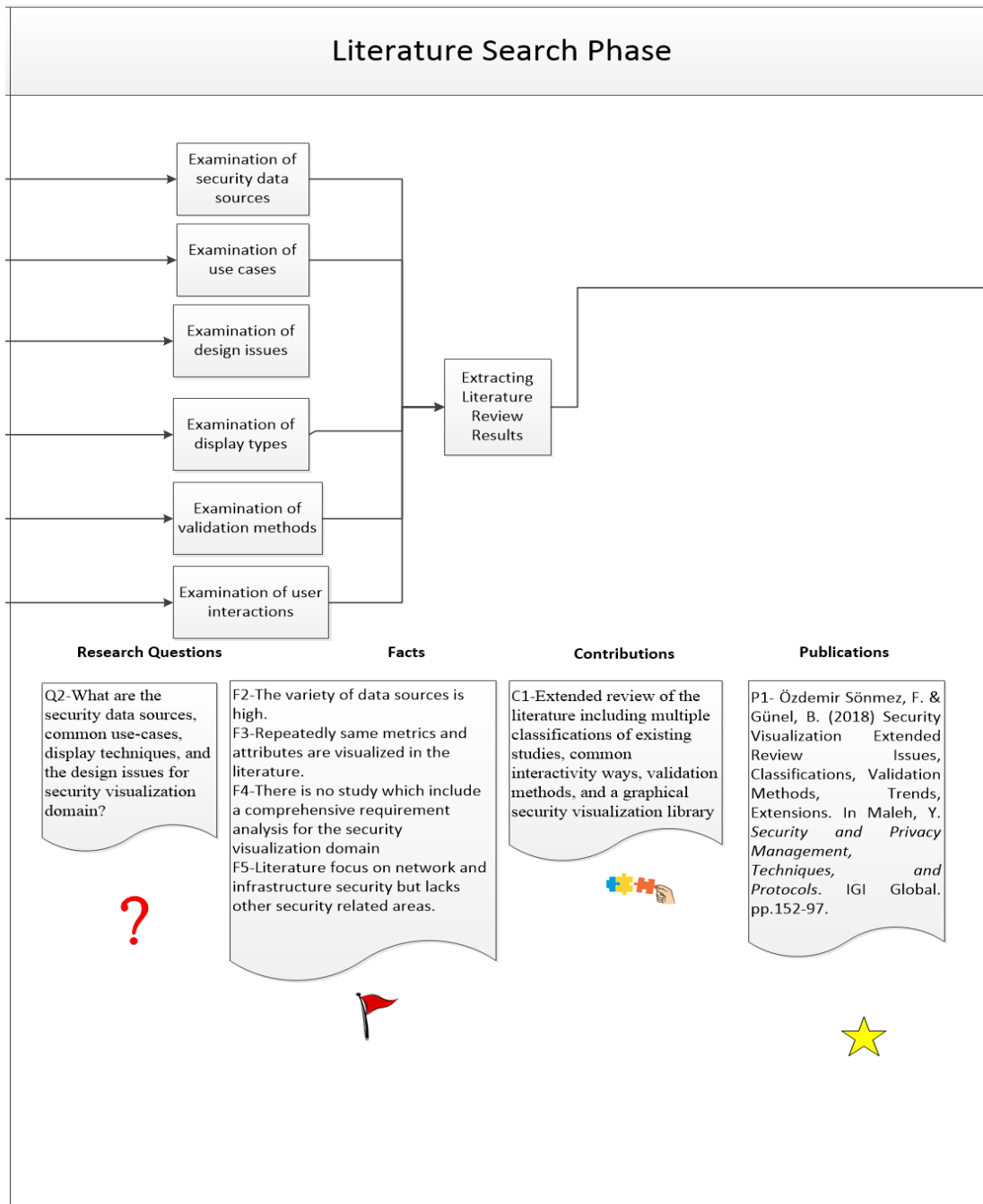


Figure 4 - Short summary of literature search phase

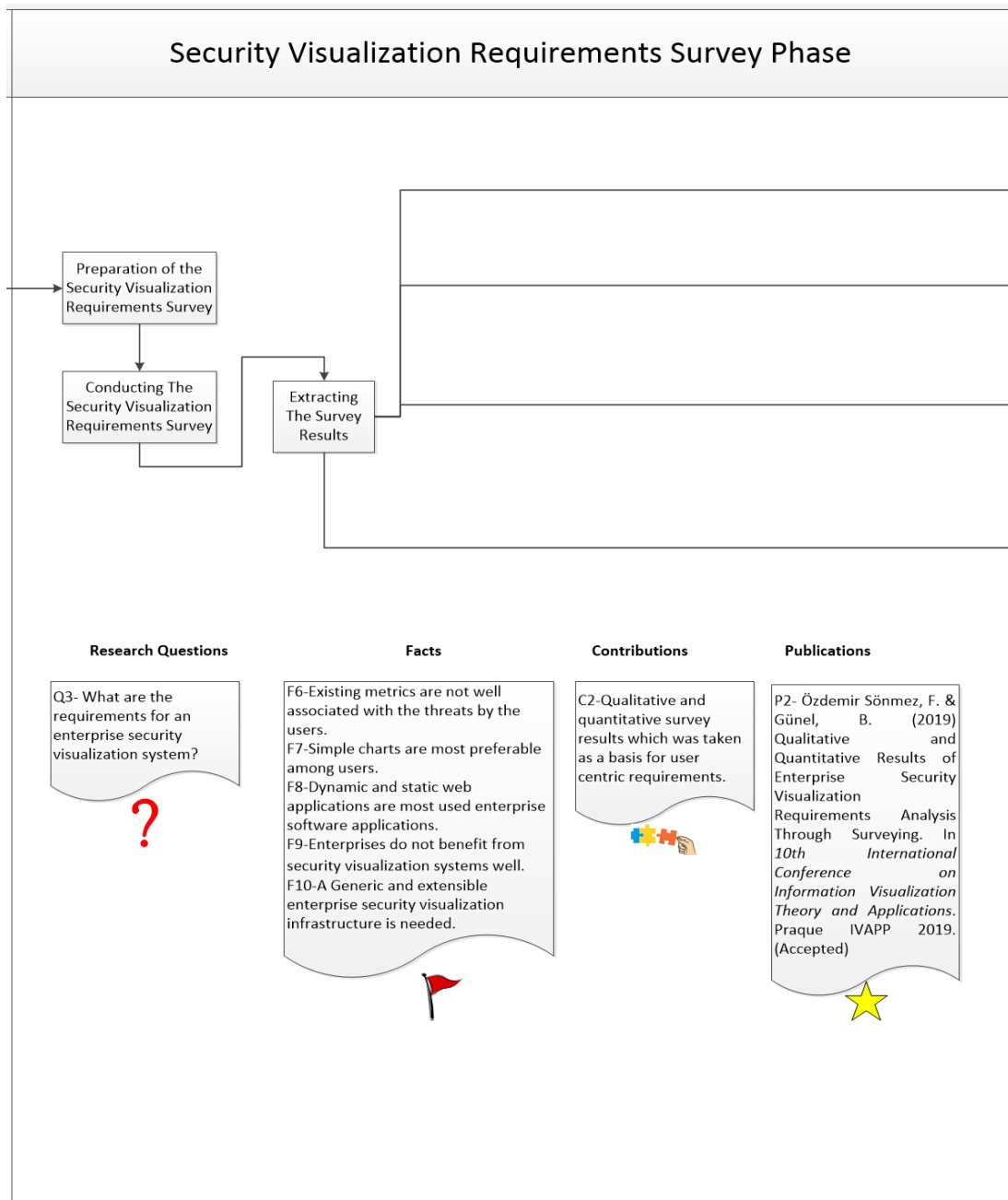


Figure 5 - Short summary of security visualization requirements survey phase

Besides the requirement set detected through the survey analysis and the literature search, which resulted in the design and development of generic enterprise security visualization framework, there were some other facts determined in the previous steps. These facts pointed out new substantial visualization topics which would be beneficial, and thus, should be handled as a part of this thesis. During the recognition of new visualization

topics (Figure 7), the following facts were identified: There is a gap for the web application vulnerability scan result visualization systems in the literature; security management related issues are not given enough importance in the security visualization domain, and SIEM systems should be investigated for their custom visualization generation capabilities.

During the evaluation of existing SIEM systems Gartner Magic Quadrant report has been used as a basis (Figure 8). Selected SIEM systems from this report were evaluated for their custom visualization capabilities based on an evaluation scenario created as part of the thesis study. Following this, web based application vulnerability scan results were examined, and a data structure with a set of measures and metrics were proposed (Figure 9). As part of validation efforts, a set of dashboards with the proposed metrics were prepared and these dashboards are evaluated for their practicality, efficiency, decision informing, and difference detection capabilities.

In the final part of the thesis, the focus was integrating security management related issues to the security visualization domain. An analytical model which aims to optimize security expenditure for the enterprises was proposed (Figure 10). This model encapsulates visualization as to help decision makers. With the thesis writing and consolidation of contributions, the thesis study ended (Figure 11).

1.5 Thesis Contributions

Table 3 shows the short descriptions and corresponding numbers for each contribution. The detailed content and description of each contributions is provided in the corresponding part of the thesis. Each contribution is associated to one part of the thesis as described in Figure 2.

Table 3 - Thesis Contributions

C1-Extended review of the literature including multiple classifications of existing studies, common interactivity ways, validation methods, and a graphical security visualization library.
C2-Qualitative and quantitative survey results which was taken as a basis for user centric requirements.
C3-Design and implementation for the enterprise security visualization infrastructure based on a user centric requirement set.
C4-Evaluation results of SIEM systems for their generation of custom visualization based on a unique evaluation scenario.
C5-A dashboard tool including forty plus measures and metrics for the web application vulnerability visualization.

C6-A Decision Support System Based on AHP, LP, and Visualization aiming to improve security management through visualization.

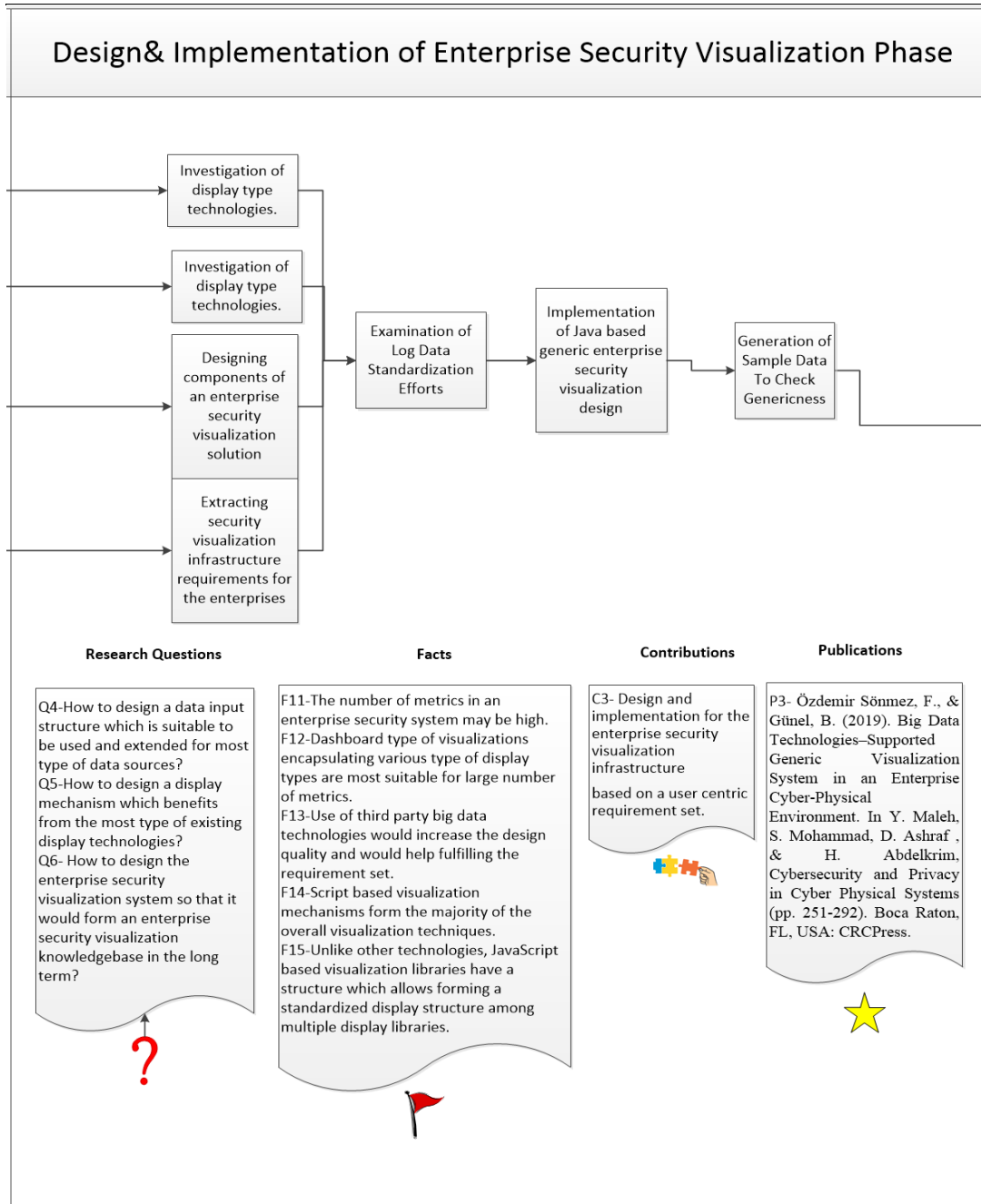


Figure 6 - Short summary of design and implementation of enterprise security visualization phase

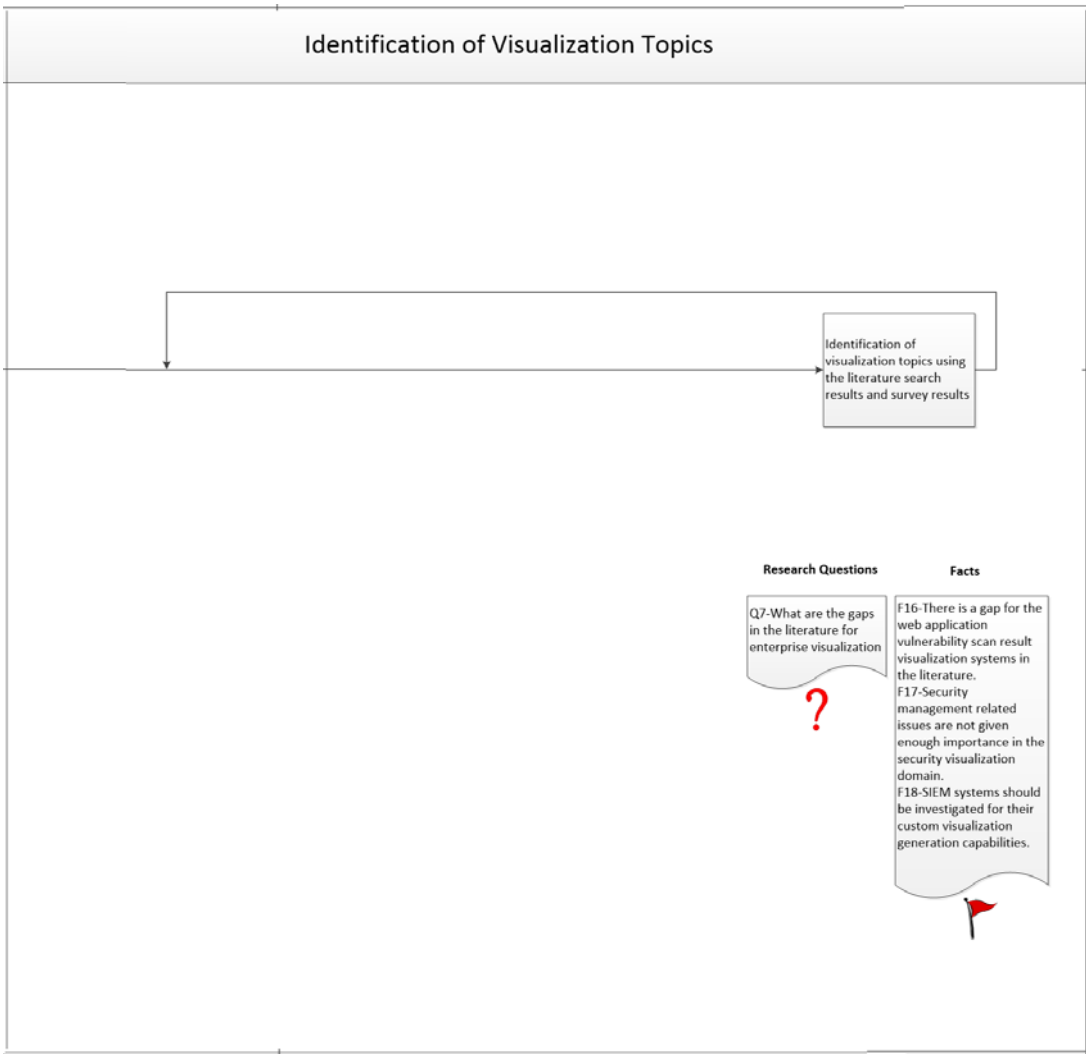


Figure 7 - Short summary of identification of visualization topics phase

Examination of SIEM Systems for Custom Visualization Generation Capabilities Phase

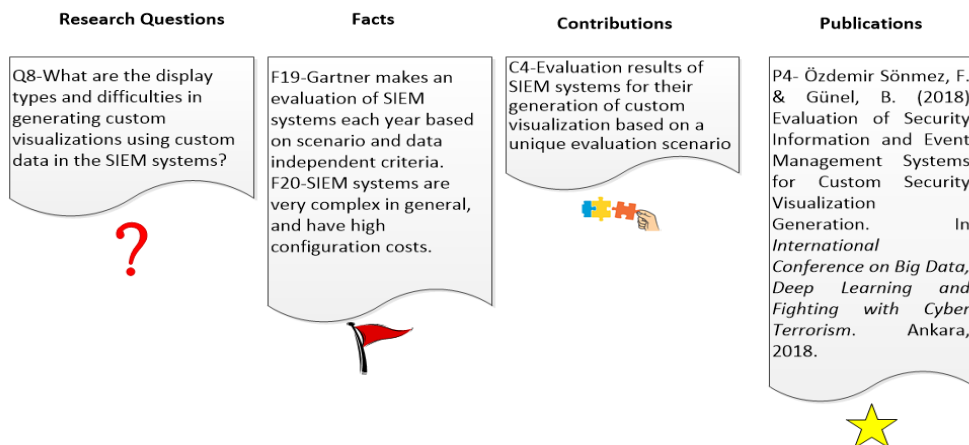
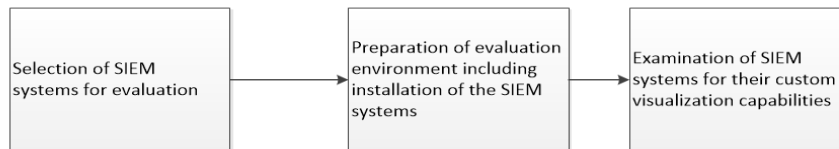


Figure 8 - Short summary of examination of SIEM systems for custom visualization generation capabilities phase

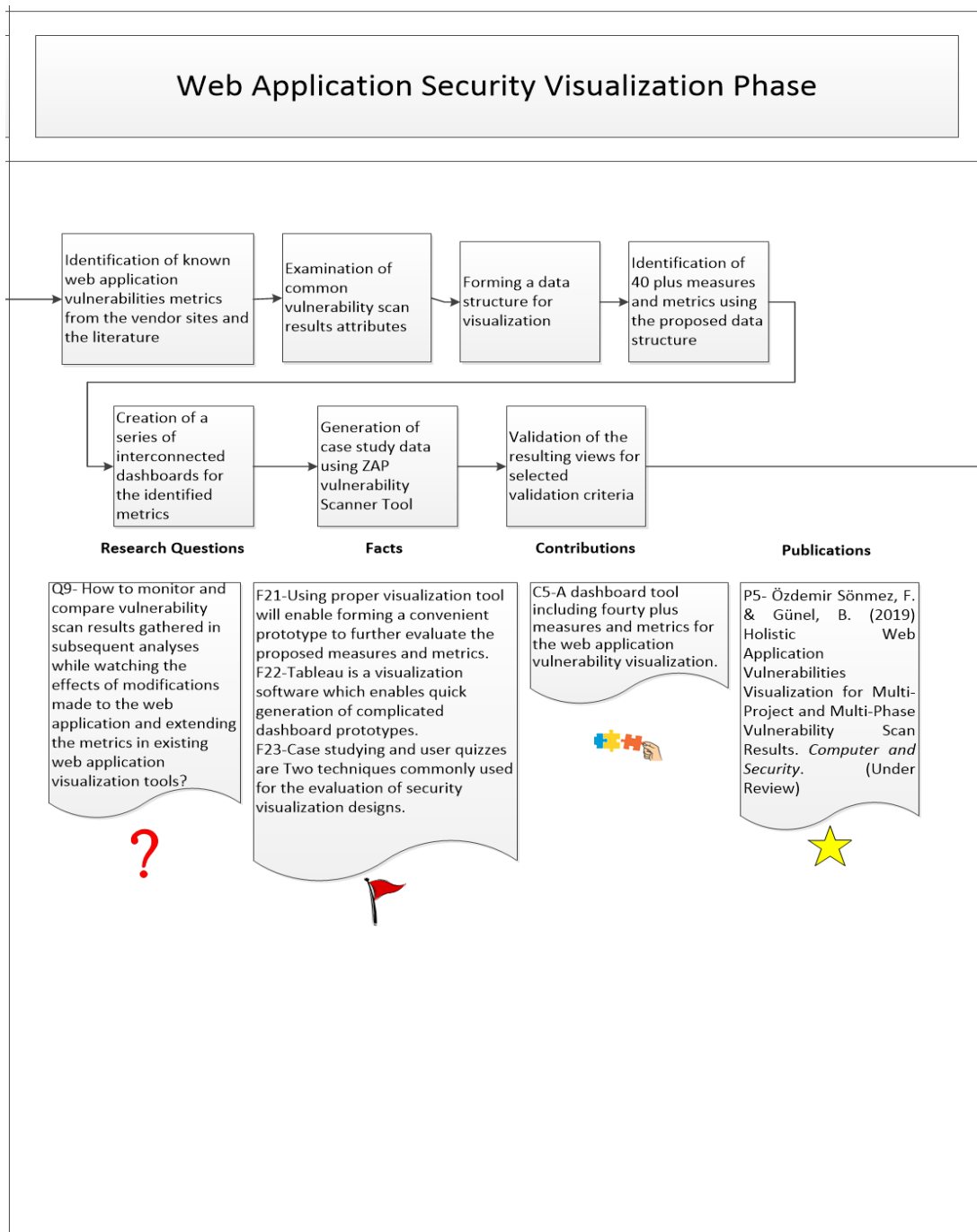


Figure 9 - Short summary of web application security visualization phase

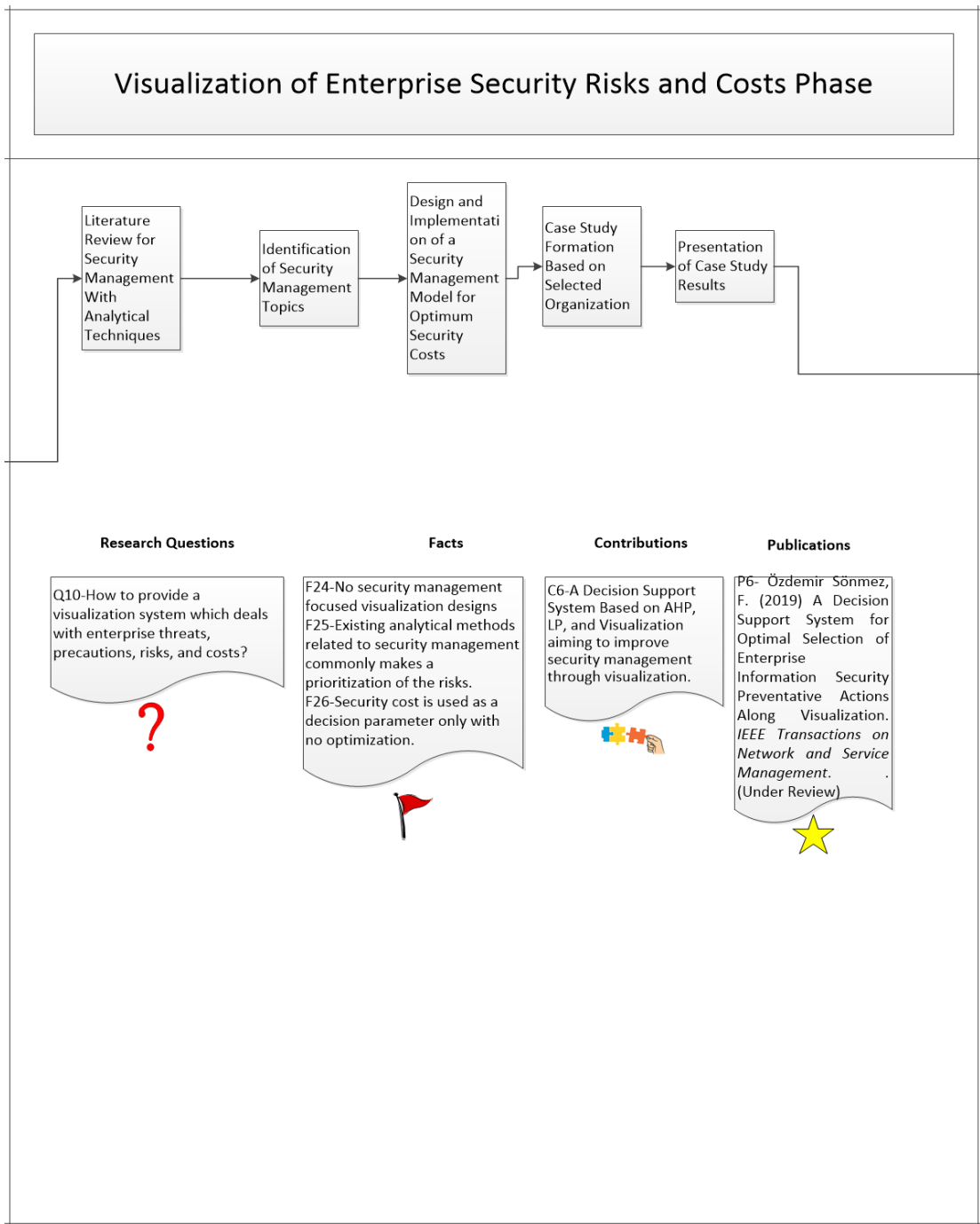


Figure 10 - Short summary of visualization of enterprise security risks and costs phase

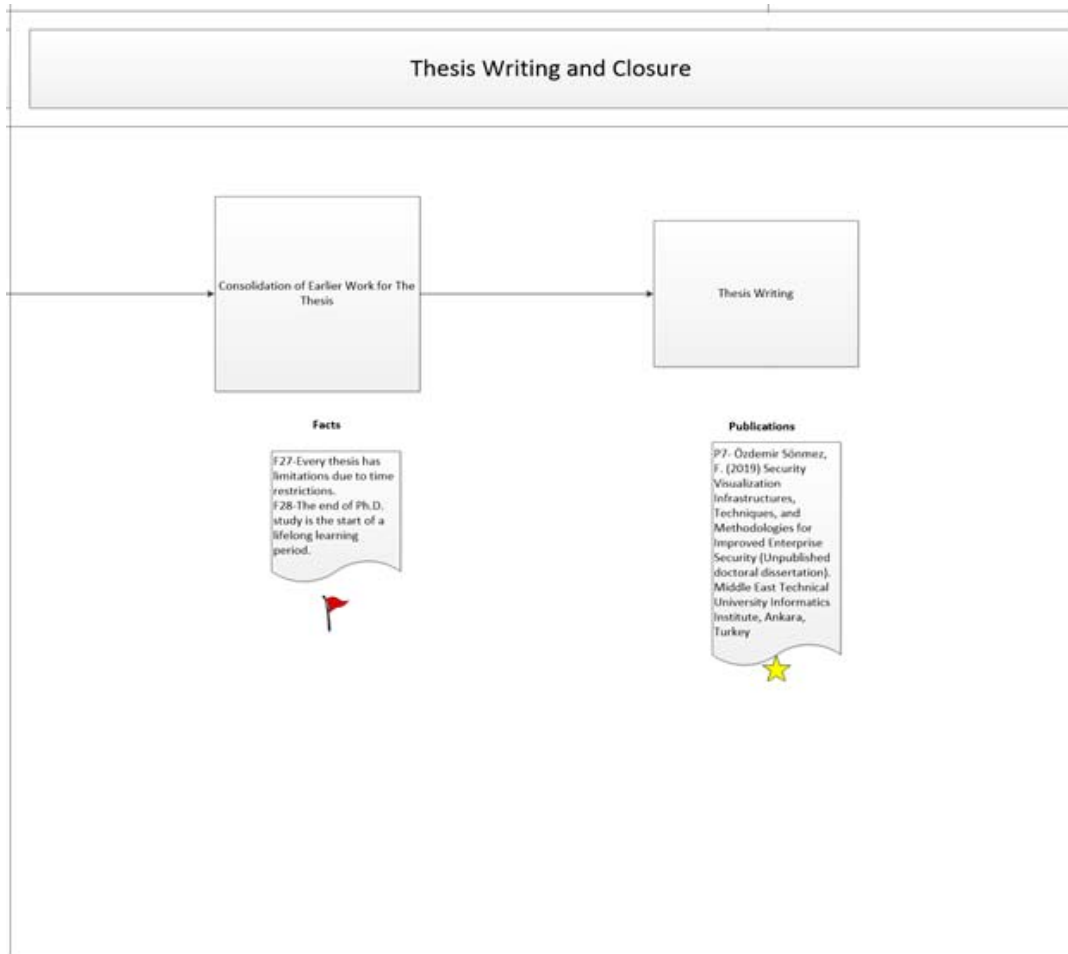


Figure 11 - Short summary thesis writing and closure phase

1.6 Publications

The publications including the thesis itself is listed in Table 4.

Table 4- Thesis Publications

P1- Özdemir Sönmez, F. & Günel, B. (2018) Security Visualization Extended Review Issues, Classifications, Validation Methods, Trends, Extensions. In Maleh, Y. Security and Privacy Management, Techniques, and Protocols. IGI Global. pp.152-97.

P2- Özdemir Sönmez, F. & Günel, B. (2019) Qualitative and Quantitative Results of Enterprise Security Visualization Requirements Analysis Through Surveying. In 10th International Conference on Information Visualization Theory and Applications. Praque IVAPP 2019.

P3- Özdemir Sönmez, F., & Günel, B. (2019). Big Data Technologies–Supported Generic Visualization System in an Enterprise Cyber-Physical Environment. In Y. Maleh, S. Mohammad, D. Ashraf , & H. Abdelkrim, Cybersecurity and Privacy in Cyber Physical Systems (pp. 251-292). Boca Raton, FL, USA: CRCPress.

P4- Özdemir Sönmez, F. & Günel, B. (2018) Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation. In International Conference on Big Data, Deep Learning and Fighting with Cyber Terrorism. Ankara, 2018.

P5- Özdemir Sönmez, F. & Günel, B. (2019) Holistic Web Application Vulnerabilities Visualization for Multi-Project and Multi-Phase Vulnerability Scan Results. Computer and Security. (Under Review)

P6- Özdemir Sönmez, F. (2019) A Decision Support System for Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization. IEEE Transactions on Network and Service Management. (Under Review)

P7- Özdemir Sönmez, F. (2019) Security Visualization Infrastructures, Techniques, and Methodologies for Improved Enterprise Security (Unpublished doctoral dissertation). Middle East Technical University Informatics Institute, Ankara, Turkey

1.7 Organization of the Thesis

Chapter 1 introduces the general subject and discusses the need for undertaking this work, and lists the research questions, the road map, contributions, and the publications. Chapter 2 gives comprehensive information on the subject based on a literature review, which provides a proper framework for the entire study. Chapter 3 describes the security requirements survey and the data obtained through the survey. Chapter 4 describes the implementation and design of generic enterprise security solution and reveals the effect of using big data technologies. Chapter 5 characterises the web application security visualization solution. Chapter 6 presents the evaluation results of the SIEM systems for custom data visualization capabilities. Chapter 7 deals with the optimization of security expenditure. Chapter 8 summarizes the various conclusions drawn from the current research work and also provides scope for further research in this area.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Literature Review

The actions threatening information security have a variety of categories. For example, “web based attacks” is a name given to express a set of harmful activities targeting web-based information systems. The occurrence rates of these harmful events can be gathered from the numeric information provided by vendors of information security protection systems. Symantec programs blocked 190000, 464100, and 568700 “web-based attacks” in 2011, 2012, and 2013, respectively, showing a 23% increase between 2012 and 2013 (Symantec, 2014). This single example shows that there is a trend of increase in the occurrence of harmful events threatening information security. The number of actions is not increasing alone; indeed, the type of threats, their sophistication levels and impacts are also getting higher by time. This makes the field of information security very important. A single computing device without any network connections can still have security vulnerabilities. However, as the computing devices get connected to each other and to the Internet, the level of threats increases exponentially. These threats may be unintentional or intentional.

In order to detect and prevent these intentional or unintentional actions, systems such as intrusion detection, intrusion prevention, and firewalls are commonly used in the enterprises. The security analysts investigate the outputs of these systems either in real time or in a delayed manner. The main source of information provided by these systems is the log files. In order to warn against momentary or future events, some of the IDS systems or firewalls include some visual or audio alert systems.

Although the alternatives and capabilities of protection systems are getting better, there are problems with the usability of these systems. The main source of problems affecting the usability of these systems is the size of the data they process. The log files are often too large to be investigated manually. The frequency of alerts is often high which overwhelms the analysts. Each alert may not point out a correct situation. This results in omissions or ignorance in the long term. Numerous tools and programs are being used in

order to overcome security vulnerabilities of the organizations. However, the outputs of these programs are rarely understood clearly.

Security visualization is the act of using information visualization techniques to ease the decision-making process for security analysts. It provides situational awareness. It offers new representations of security data to increase the comprehension and provide an efficient processing of the data. In general, there is a tendency to use the same type of display types for the same use cases, or the same type of display types for the data in similar formats. While this is the result of a consolidated learning in most cases, it may be useful to find alternative combinations of these use cases, display types, and data attributes for novel security visualization designs.

To this end, while introducing the selected existing work in this chapter, these works are classified according to display types, use cases, and data sources. The objective of this chapter is to classify the existing work which are similar to each other, and by doing so to find out gaps, such as the data types which are seldomly used for security visualization purposes. In this way, it is expected to find new ways of combining data coming from multiple sources and display types commonly used for some particular scenarios which may also be suitable for some other scenarios. This extended summary of security visualization designs may help researchers who want to solve security visualization problems by applying novel designs and those who investigate current status and trends in the security visualization domain.

The reviews written so far in the security visualization domain focus on a limited number of works. Survey results that depend on few designs can provide only an incomplete perspective of the domain information. In this chapter, the number of designs that are examined in detail is 79. This examination results in a detailed perspective of the security visualization domain. The contribution of this work to the existing literature can be summarized as follows:

- An extended summary of the existing work is given which may help novice researchers find out what has been done so far.
- The security visualization literature is classified according to use cases, display types and data sources to help researchers find out gaps and alternative combinations of data, display and usage scenarios.
- Notable features, interactivity and usability properties of the designs and their validation methods are depicted, and a short trend analysis of the security visualization domain is made.

The next section is the Background Section including the design issues, and the common security visualization classification methods. The Methodology of the review study describing the overall procedure taken through the study including the scope definition is described next. This section is followed by the main section, Extended Review of the Selected Studies, including the classified findings, and validation methods of security

visualization studies. The next section is the Future Research Directions Section. Finally, there is the Concluding Remarks dedicated for this chapter.

2.2 Background

2.2.1 Security Visualization

Due to the increase of data in information technologies, visualization has become a popular technique for analyzing, and communicating the big data. Using visualization in the security domain is a relatively new research area. The first published work appeared in 2004. The major reason for the emergence of security visualization is the necessity of analyzing the huge size of security related data in a timely manner. Security visualizations enable human assessment of large size log files efficiently, which results in timely and improved decision making.

Marty (Marty, Applied security visualization, 2009) described the benefits of security visualization as being able to answer questions, posing new questions, allowing exploration and discovery, supporting decisions, communicating information, increasing efficiency, and inspiring the researchers. Security visualization designs may have different purposes such as summarizing the data, simulating past incidents, allowing pattern discovery, detection of malicious activities, anomalies, misconfigurations, and outliers. Security visualization may provide multiple views of the same data simultaneously or it may visualize different data in the same view.

Table 5 - Types of Displays

Category	Display Types
Simple 2-D Charts	Line Chart, Bar Chart, Pie Chart
Simple 3-D Charts	3-D Line Chart, 3-D Bar Chart, 3-D Pie Chart
Stacked Charts	Stacked Pie Chart, Stacked Bar Chart, Stacked Line Chart
Histograms	Histogram Chart
Box-Plots	Box-Plot
Matrixes	2-D Matrix, 3-D Matrix
Scatter Plots	2-D Scatter Plot, 3-D Scatter Plot
Parallel Coordinate Views	2-D Parallel Coordinates, 3-D Parallel Coordinates
Link Graphs	2-D Node Link Graph, 3-D Node Link Graph
Maps	Geographic Map, Globe View

Treemaps	2-D Treemap, 3-D Treemap
Advanced Views	Animation, Gamification, Simulation Views

Basic elements of visualization are data type, which can be categorical, ordinal, interval and ratio, the color, size, orientation, and shape of the graph, position, length and space allocated by the data on the graph and the use and purpose of chart axes (Marty, Applied security visualization, 2009). Types of displays vary from simple line charts to 3-D gamification and/or simulation displays. A list of alternative display types is shown in Table 5.

Choosing the right display type depends on the maximum number of data values, the number of data dimensions, data types and use-cases. Designing security visualizations needs expertise on security, data analyses techniques and visualization techniques. In order to make a contribution to the existing work, one should gain expertise in both security and visualization techniques. The majority of the journals concerning security visualization and conference papers visualize network traffic data. However, there are also other designs which visualizes other types of data.

2.2.2 *Design Issues*

There are numerous survey articles which depict the design issues of security visualization. Langton and Newey (2010) listed the design challenges and requirements of security visualizations as scaling for the size and dimensionality of cyber security datasets, displaying both historical and real-time data, addressing different data types, designing new human interaction techniques and improving them.

Harrison and Lu (Harrison & Lu, 2012) evaluated several state-of-the-art approaches such as Clique (Best, Hafen, Olsen, & Pike, 2011) and Clockview (Kintzel, Fuchs, & Mansmann, Monitoring large ip spaces with clockview, 2011) ending with a conclusion that existing tools solve many of the design problems; however, there is still a need for the improvement stating that while some visualization designs are scalable for a high volume of data, such as histograms, they are not scalable for a high number of dimensions. Other tools are better for high dimensions such as parallel coordinates. Security analysis techniques mentioned in network traffic analysis section require the analysis of information coming from a combination of data sources such as firewalls and IDSs to get meaningful results. Most of the existing security visualization tools, however, visualize data coming from a single source. Harrison and Lu also state that the existing security visualization studies lack risk awareness and management and analysis reporting. Another important issue identified by Harrison and Lu is that some of the attack types such as worms and persistent threats are still not detectable by the existing visualization systems. Another way of identifying the design issues is the bottom up

approach. Luse (2009) defined the necessary components of a network security visualization tool as overview, zoom, filter, details-on-demand, relate, history, extract, and primary notification. Luse examined the architectures of several visualization frameworks such as Tudumi (Takada & Koike, 2002), TNV (Goodall, Lutters, Rheingans, & Komlodi, 2005), NVisionIP (Lakkaraju, Yurcik, & Lee, 2004), Visual (Lee, Tros, Gibbs, Beyah, & Copeland, 2005) and IDSRainstorms (Abdullah, Lee, Conti, Copeland, & Stasko, Ids rainstorm: Visualizing ids alarms, 2005) by checking the existence of particular components in each framework. The necessary component sets can be enlarged by considering the new technological achievements.

2.2.2.1 Common Security Visualization Classification Methods

The majority of the taxonomies made so far for visualization tools or prototypes use three major types of categories. The input data driven type of categorization puts the visualization techniques using the same type of log files into the the same group. The second method, use-case driven, categorizes the security visualization designs according to usage scenarios, and the third method takes the categories of graphical approaches as the categorization criteria. There are also some sub-categorization systems which may be based on techniques such as being signature-driven or not and being real time or not.

All three major categorization systems are beneficial in their unique ways. The first method enables to find out and compare various types of visualizations for the same data types, such as traffic data, firewall log data, operating system data, and network structure data. For example, a traffic data may be displayed using a link graph or parallel coordinates graph, or firewall log data may be represented using a simple histogram chart or a complex graph showing more attributes.

The second method focuses on details of use cases better. Some survey articles in the security visualization research area depict possible use-cases of visualization for information security. Shiravi et al. (2012) classified the existing security visualization systems based on use case scenarios. The first group of use cases is the host-server monitoring which intends to monitor the current state of the hosts and the servers in a network. The second group of use cases is internal hosts with external IP numbers. The third group of use cases depends on port activity monitoring which aims to detect abnormal activity on ports to detect trojans, worms, and viruses which in general show abnormal patterns in port activities. The fourth group of use cases is the monitoring of attack patterns. The visualization systems in this group aims to visualize not only a snapshot of an attack but the behavior of the attack over a time period. The fifth group of use cases focuses on routing behaviors aiming to understand border gateway routing evolution in time.

The third method is useful to learn about and evaluate the technical diversity in security visualization techniques. Zhang et al. (2012) made a classification of security visualization techniques based on display types. The categories foreseen by the authors are text-based visualization which include works using geolocation such as the work by

StoneGate Management Center (Geolocation Map, 3), wireless network tools such as IntraVue (Conti & Abdullah, 2004), Wi-Viz (McPherson, Ma, Krystosk, Bartoletti, & Christensen, 2004) and WVis (Bogen, Dampier, & Carver, 2007), parallel visualization techniques such as PicViz (Tricaud, 2008), Rumint (McRee, 2008) and Visual Firewall (Lee, Tros, Gibbs, Beyah, & Copeland, 2005), hierarchical visualization techniques such as Treemaps (Johnson & Shneiderman, Tree-maps: A space-filling approach to the visualization of hierarchical information structures, 1991), three-dimensional visualization techniques such as INetVis (McRee, 2008), Flamingo (Oberheide, Goff, & Karir, 2006), Mineset (Brunk, Kelly, & Kohavi, 1997) and other visualization techniques which do not fall into any categories such as a border gateway protocol (BGP) based design (Teoh, Ma, Wu, & Zhao, 2002).

The classifications of general information visualization techniques are different from the security visualization techniques. The first type of classification is made based on the complexity of the model, which is calculated as the number of the dimensions of the data. The second type of categorization is made to make a differentiation between infographics studies and data visualization studies. The graphic generation methods are compared according to the amount of algorithmic work and the amount of manual drawings, the amount of aesthetic work, the capability of running for different data, and the quality of the data used. The third type of categorization is based on the purpose of the visualization, which may be exploration, explanation, and a combination of both, hybrid. Information visualization categorization attributes are also examined throughout the study.

2.3 Methodology of the Review Study

This study aims to synthesize the existing knowledge in the domain of security visualization through the review of existing literature. This research has two principle parts. The first part concentrates on reviewing the literature with keywords, which critically affects the quality of any review study. The second part includes investigating the selected literature. The methodology used during the preparation of this chapter is explained through the description of the tasks carried out, as in the following order:

The identification of research questions: The main concern at the start of this review study was to find out the parameters of the preparation of novel security visualization designs. Therefore, a preliminary literature research was done in the security visualization domain to find out the attributes affecting the overall structure and purpose of the design. After the preliminary research, the curiosity about whether a systematic method of alternative display types, alternative data sources, and alternative security visualization usage scenarios are related to each other, and whether these choices changed over time guided the study. As a result of this rigorous work, new associations of these alternatives can be made, such as making a new association of a display type with a use case or associating a data source with a use case, which were not done before.

After the initial curiosity about the parameters of security visualization designs, some secondary research questions arose which are: “What type of user interactions exist?”, “In what ways can these interactions be improved?”, “Which type of interactions are more beneficial for some display types?”, “Are the level and ways of evaluation and validation of existing designs adequate?” “In what ways can these evaluation and validation methods be improved?” during the evaluation of the selected work.

Conducting the search: Web of Science database was used as the main source of the review. The details of query results were as follows:

keyword = «Security Visualization», # of results = 96.

keyword= «Network Traffic Visualization» # of results = 21.

keyword = Security Visualization (without quotation marks), # of results = 936.

Selection of earlier work: The query results were stored in three Excel files containing title, authors, and abstract of the studies. The lists were examined according to their relevance to the study and their correspondence to the research questions. The works on implementation architectures were not included in this study, but reserved for future work. Similarly, although earlier review work were benefited from, these were not included in the final set of studies. The articles which introduce the design and/or implementation of a new security visualization work were selected. Numerical details related to the selection results are as follows:

- 21 of the results were eliminated from the first group.
- 1 of the results was eliminated from the second group.
- 263 of the results were eliminated from the third group directly due to their irrelevance to the “security visualization” topic. The rest of the works were examined manually in the order of relevance until the content of articles was comprehensive enough to be included in the study for a correct perspective of the domain.

The overall result set included a total of 79 tool and prototype designs.

Investigation of the selected studies: Investigation was divided into steps which have various focus points as listed below. This part of the research included returning to previous steps multiple times as the new findings required re-examination, re-classification or clarification of the earlier findings. The focus points of these steps are explained below.

A. Investigate display types used in the selected work

1. Classify the display types and associate each selected study to a display type
 2. Examine display properties
 - i. Point out the designs providing properties which do not commonly exist in other designs
 - ii. Point out designs having higher usability level and/or looking more appealing
 - iii. Point out designs having lower usability levels
- B. Find out the data sets used in the selected work (While data sets are explicit in some earlier works, in others they were not.)
1. Point out data sets used independently from other information
 2. Point out data sets which are combined with other information
 3. Point out the attributes used most often in displays
 4. Point out dataset which are not used often in the existing designs
 5. Point out the attributes used most often in displays
 6. Point out data attributes which are not used often in existing designs
- C. Find out the objectives of the visualization designs
1. Point out and name the use cases of the design
 2. Categorize the resulting use-cases based on their usability for enterprises
- D. Mix the information gathered in the previous steps for comparison of designs with each other and to provide a solid domain knowledge
1. Give the example works using similar display types
 2. Give the example of works which have similar objectives
 3. Give the examples of works which use the same type of data

Discussion of the selected studies: In this part, whenever possible, recurring design properties were discussed along with missing points and subsequent design issues. Analysis results and the changes of design decisions over time were also included in this part of the study.

2.4 Extended Review of the Selected Studies

2.4.1 *Issues, Controversies, Problems*

Existing review studies focus on relatively small set of designs which result in a limited perspective of the area. Majority of the reviews also make the comparison of designs based on only one categorization item. However, examining the existing designs with multiple perspectives is necessary to understand current status of the domain. Examination of a larger set based on multiple criteria will provide better guidance to security visualization researchers.

The design issues introduced in the Background Section of this chapter point out common security visualization design problems. This provides an upper level

identification of the design issues without details. Hence, to fill this gap, an extended review of the domain is provided with details in this section to improve the upper level design issues mentioned so far. Additionally, further investigation of sets of existing designs prepared for some usage scenarios and having some particular display or data source attributes provided in this section, can be used as starting points during the creation of similar novel designs.

2.4.2 Examination Results

In this part of the chapter, the findings are classified under five sub-sections. Some particular designs are related only to one sub-section, such as having interesting display designs or noteworthy interaction features. However, the majority of the designs are related to more than one sub-section. There are also some designs which are related to all of the five categories. The findings which are not associated with the prior categories are included in the Other Notes Section.

Table 6 - Groupings of Use-cases

Among the enterprises: Use-cases which focus on network traffic visualization					
Visualization of end to end traffic: e.g. TNV, Visual, Visflowconnect	Port Monitoring: Spinning cube of potential Existence PortViz	Activity e.g. doom, plots,	Network detection: Linear Visualization	anomaly e.g. Tri	Monitoring of attack patterns of network traffic: e.g. P3D, Rumint, Krasser et al.
Visualizing web browser activities: e.g. HVIZ	Visualization of large-scale network data for monitoring and planning purposes: e.g. Histomap		Intrusion detection using traffic data: e.g. Abdullah et al., Netbytes viewer		Monitoring of attack patterns using IDS data: e.g. Snortview, IP-Matrix, Vizalert, IDTk
Inside the enterprises: Use-cases which are more applicable to enterprises/institutions					
Firewall configuration visualization: e.g. PolicyVis	Firewall visualization: Visual Firewall by Lee et al., Vafle	log e.g.	Network topology visualization for network planning and thrust relationships: e.g. TrustVis, Mansman et al., SecureScope		Visualization of application(s), service(s) and host(s) interaction: e.g. Tudumi, Enavis, Nagios
Visualization of hosts and network level vulnerability scanner results: e.g. NV	Application vulnerability level visualization: e.g. Goodal et al, Dang and Dang, Alsaleh et al.		Visualization of filesharings: e.g. Rode et al., Tri and Dang		Various type of network traffic visualization use-cases
Beyond the enterprises: Use-cases which focus on network routing and DNS protocols					
Visualization of BGP update messages: e.g. Teoh et al., BGP Eye	Visualization of BGP events (not updates): e.g. Teoh et al., Teoh et al.		Visualization of AS route changes and routing behaviours: e.g. Teoh et al., Tamp, LinkRank, Elisha, BGPlay		Visualization of DNS queries: e.g. SEO et al.

Other: Other use-cases focusing mostly on attack types			
Web attack scenarios in space and time coordinate systems: e.g. Dang and Dang	Visualization of Sybil attacks: e.g. Lu et al.	Visualization of Botnets: e.g. Dorothy project	Visualization of malware: e.g. Nataraj et al.

2.4.2.1 Findings Related to Use Cases

In this part, while considering the use cases of the earlier visualization designs, notable results, the designs more suitable for enterprises/institutions are pointed out.

Table 6 provides a summary classifying the use-cases identified during this study. The findings display that network traffic visualization is the most frequently studied use case. Visualization of the end-to-end traffic between internal hosts and external IP's, such as TNV (Goodall, Lutters, Rheingans, & Komlodi, 2005), Visual (Ball, Fink, & North, 2004), Visflowconnect (Yin, Yurcik, Treaster, Li, & Lakkaraju, 2004), port activity monitoring, such as Spinning Cube of Potential Doom (Lau, 2004), Existence plots (Janies, 2008), PortViz (McPherson, Ma, Krystosk, Bartoletti, & Christensen, 2004), network anomaly detection, such as Tri Linear visualization (Whitaker & Erbacher, 2011), monitoring of attack patterns such as P3D (Nunnally, et al., 2013), Rumint (Conti G. , et al., 2006), Krasser et al. (Krasser, Conti, Grizzard, Gribshaw, & Owen, 2005), Girardin (Girardin, 1999), visualizing web browsing activities of a host, such as Hviz (Gugelmann, Gasser, Ager, & Lenders, DFRWS 2015 Europe, 2015), visualization of large-scale network data for planning and monitoring, such as Histomap (Mansmann F. , Keim, North, Rexroad, & Sheleheda, 2007) are various types of network traffic visualization use cases.

The type of use cases which aim to detect alarm situations are, generally, grouped under the title “Monitoring of attack patterns” in the literature. Although network traffic data is also taken as data source for some designs, such as Abdullah et al. (Abdullah, Lee, Conti, & Copeland, IAW'05, 2005) and Netbytes Viewer (Taylor, Brooks, & McHugh, 2008) to detect network intrusions, intrusion detection and/or prevention system data is used more in this group to classify true and false alarms. Some designs are prepared for specific intrusion detection tools such as Snortview (Koike & Ohno, SnortView: visualization system of snort logs., 2004) which visualizes Snort IDS alarms. Other examples visualizing IDS data are IP Matrix (Koike, Ohno, & Koizumi, VizSEC 05, 2005), Vizalert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005) and IDtk (Komlodi, Rheingans, Ayachit, Goodall, & Joshi, 2005).

In addition to the network traffic and alarm situations monitoring related use cases listed above, there are also some use cases which seem to be applicable to improve enterprise security. Use-cases related to firewall utilization are in this group. These are use cases related either to firewall configuration or log monitoring. For example, firewall configuration visualization is focused on in PolicyVis (Tran, Al-Shaer, & Boutaba) aiming to help the investigation of complicated firewall rules. Besides investigating configuration rules, visualization finds other use cases for itself in using the firewall data. Visual Firewall by Lee et al. (Lee, Tros, Gibbs, Beyah, & Copeland, 2005) focuses on

visualizing firewall reactions to network traffic and Vafle (Ghoniem, Shurkhovetsky, Bahey, & Otjacques, 2014) focuses on firewall log visualization.

Another group of use-cases which are applicable to the enterprises include host/network topology visualization. Network topology visualization is used for network planning and trust relationship management in TrustVis (Peng, Chen, & Peng, 2012) aiming to visualize trust and to help identifying attacks in an organization. Visualizing the application - host interaction is a use case which combines the hosts' topology, the hosts' location and network traffic data in a way such as in Mansman et al. (Mansman, Meier, & Keim, 2008) and Securescope (Ferebee & Dasgupta, 2008).

Visualization of applications, services and hosts interaction is also beneficial for the enterprises. Visualization of network access and log-in information of a group of users to a server is studied in Tudumi (Takada & Koike, 2002) which is also similar in that sense. A visualization study which focuses on enterprise security visualization is Enavis (Liao, Blaich, Striegel, & Thain, 2008). In this study, the association between users and applications in an enterprise network is given in a link graph which consists of hosts, users, and applications. The aim of this work is to answer the question of "who does what in an enterprise network". The system is based on agent scripts deployed on hosts and servers which call Unix commands periodically. A subtype of monitoring hosts and services is the availability monitoring. Nagios Core (Josephsen, 2007) checks the availability of hosts and services and differentiates the unreachable or down machines and services. Nagios (Josephsen, 2007) uses multiple sets of command calls such as Ping, HTTP, SSH and MYSQL to collect data from different points. The designs based on periodical control of some information, such as this, include parameter settings. For example, as check and recheck interval, the maximum number of checks and a period for each check are among these parameters.

Vulnerability analysis scans visualization is also another category of visualization, which targets enterprise security. A design by Harrison et al. named NV (Harrison L., Spahn, Iannacone, Downing, & Goodall, NV: Nessus vulnerability visualization for the web, 2012, October), takes Nessus Vulnerability Scanner data, and visualizes the data using a combination of treemaps and bar charts. This model illustrates the level of vulnerability for each workstation in an enterprise. There are other ways of visualizing vulnerability levels. Specially designed vulnerability scanners search for application vulnerabilities using a number of code files. Visualization of outputs of such programs forms another visualization use case group. This type of visualization may be used to search for and make an analysis of enterprise application vulnerabilities. An example of this group of visualization designs is provided by Goodall et al. (Goodall, Radwan, & Halseth, VizSec '10, 2010).

Visualization becomes beneficial for evaluators as it facilitates collaboration during application security level examinations. A study for web application vulnerability visualization was made by Dang and Dang (Dang & Dang, 2014). This design enables

communication among vulnerability evaluators over visualization software. Another example of application level visualization focuses on PHP based web applications. This application visualizes security logs aiming to support security analysts for decision making during ongoing web server attacks (Alsaleh et al.) (Alsaleh, Alarifi, Alqahtani, & Al-Salman, 2015).

Monitoring the file sharings both with the insiders and/or with the external parties has high importance in terms of enterprise security. Visualization of file sharings among users is studied by Rode et al. (Rode, et al., 2006). This design provides additional features such as monitoring all the users' history who worked on the files before, providing list of the files which have not been shared at all yet. Another design which focuses on visualization of file sharings is by Tri and Dang (Tri & Dang, 2009). This design focuses on file events instead of user actions.

There are some use cases which are more related to the data beyond the interior of the enterprises to the Internet. Border gateway protocol is responsible to make the Internet routings between AS's on the Internet. It does not include any features related to the diagnosis of the routing decisions. Visualization is used to analyse and detect the anomalies in the Internet routing protocols. Visualization of AS route changes and routing behaviours are studied in Teoh et al. (Teoh, Ma, Wu, & Zhao, 2002), Tamp (Wong, Jacobson, & Alaettinoglu, DSN 2005, 2005), LinkRank (Lad, Massey, & Zhang, 2006), Elisha (Teoh, et al., 2003) and BGPlay (Colitti, Di Battista, Mariani, Patrignani, & Pizzonia, 2005). Visualization of BGP update messages is studied by Teoh et al. (Teoh, Zhang, Tseng, Ma, & Wu, 2004) (Teoh, Ma, Wu, & Jankun-Kelly, 2004). Visualization of BGP events (not BGP updates) is also studied by Teoh et al. (Teoh, Ranjan, Nucci, & Chuah, 2006) in BGP Eye. Use cases which visualize DNS queries are also in this group of use-cases. DNS queries data can be counted as more related to data beyond the enterprises to the Internet. An example of visualization of DNS queries is provided by Seo et al. (Seo, Lee, & Han, 2014).

While there are designs which aim to detect multiple types of attacks, there are other visualization designs which focus on a particular type of attack. An example use-case which focuses on a single attack type is the work by Lu et al. (Lu, Wang, Dnyate, & Hu, 2011) aiming to visualize the network topology in order to detect Sybill attacks. Another example is the Dorothy project (Cremonini & Riccardi, 2009) which visualizes the botnets using honeynet analysis results. The study by Seo et al. (Seo, Lee, & Han, 2014) also visualizes the botnet traffic. Another visualization study by Nataraj et al. (Nataraj, Karthikeyan, Jacob, & Manjunath, 2011) makes a the malware visualization. The study by. Seo et al. (Seo, Lee, & Han, 2014) uses DNS queries to detect botnets. Dorothy (Cremonini & Riccardi, 2009) project focuses on botnet detection based on a totally different approach. The researchers installed a honey-net and found out the hosts with some specific malwares installed through IRC channels. They used the resulting information to find out the zombie and C&C machines using some particular metrics. Visualization of malwares is a type of visualization which is commonly used for the

classification purposes. This use case has its own unique features different from other security visualization designs. In this type of designs, the binary of malwares are converted to 8-bit vectors, and these vectors are converted to grayscale images. Various part of images correspond to different sections of the binaries. Thus, malwares which belong to the same family have similar images.

The use-cases focusing on visualization of attack scenarios form another group of use-cases. Visualizing web attack scenarios in space and time coordinate systems is an interesting study by Dang and Dang (Dang & Dang, 2014), who offer that in order to understand intrusion detection attacks, it is important to understand the cause and effect relationships. Therefore, Dang and Dang (Dang & Dang, 2014), developed a prototype which visualizes attack scenarios. This visualization system is based on exploiting the links between pages of web applications and does not require the predefinition of cause and effect relationships (Dang & Dang, 2014).

2.4.2.2 Findings Related to Data Sources

One way of calculating the complexity of visualization designs is to identify the number of dimensions of the visualized data. While identifying the dimensions in this chapter, the following difficulties have been encountered.

- Although the majority of the designs describe the data sources, many of them do not explicitly identify all the data attributes.
- Some particular designs are able to visualize multiple types of data sources each having different number of attributes.
- The number of dimensions also changes due to the parameter selection of the users for some designs.

So, instead of defining attributes and dimensions, a categorization based on data sources is made. Examining the data sources of the visualization studies results in the following findings.

In Figure 12 is demonstrated. While the majority of the security visualization studies focus on visualization of network traffic data such as TNV (Goodall, Lutters, Rheingans, & Komlodi, 2005), Visual (Ball, Fink, & North, 2004), Visflowconnect (Yin, Yurcik, Treaster, Li, & Lakkaraju, 2004), Spinning Cube of Potential Doom (Lau, 2004), Existence plots (Janies, 2008), PortViz (McPherson, Ma, Krystosk, Bartoletti, & Christensen, 2004), Tri Linear visualization (Whitaker & Erbacher, 2011), P3D (Nunnally, et al., 2013), Rumint (Conti G. , et al., 2006), Krasser et al. (Krasser, Conti, Grizzard, Gribschaw, & Owen, 2005), Girardin (Girardin, 1999), Hviz (Gugelmann, Gasser, Ager, & Lenders, DFRWS 2015 Europe, 2015) and Histomap (Mansmann F. , Keim, North, Rexroad, & Sheleheda, 2007) and visualization of IDS data such as Snortview (Koike & Ohno, 2004), IP Matrix (Koike, Ohno, & Koizumi, 2005), Vizalert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005), IDtk (Komlodi, Rheingans, Ayachit, Goodall, & Joshi, 2005), IDS Rainstorm (Abdullah, Lee, Conti, Copeland, & Stasko,

2005), Avisa (Shiravi, Shiravi, & Ghorbani, 2010) and Avisa2 (Shiravi, Shiravi, & Ghorbani, 2012), there are others, which work on disparate data types or combinations of them. There are many alternative data sources such as firewall data (Lee, Tros, Gibbs, Beyah, & Copeland, 2005), network topology data (Peng, Chen, & Peng, 2012), application code data (Goodall, Radwan, & Halseth, VizSec '10, 2010), event classification data (Zhao, Zhou, & Shi, 2012), web site topology data (Dang & Dang, 2014), file sharing data (Tri & Dang, 2009), and vulnerability scanner data, (Nunnally, Uluagac, Copeland, & Beyah, 2012, October), and NV (Harrison L., Spahn, Iannacone, Downing, & Goodall, NV: Nessus vulnerability visualization for the web, 2012, October).

The reason of selecting a specific data source or a group of data sources for a visualization study case can sometimes be easily predicted but not always. Vulnerability scanner data is used to visualize vulnerability levels of a group of hosts. This data is combined with IDS data, firewall log data, key logger data and network traffic analyser data in 3DSVat (Nunnally, Uluagac, Copeland, & Beyah, 2012) aiming to allow quick response in case of vulnerability level increase for a host. In order to incorporate application vulnerability levels, software codes are used as the visualization data source. An example is from Goodall et al. (Goodall, Radwan, & Halseth, VizSec '10, 2010). Social interaction data of members of a network is used to visualize the trust in an environment in Trustvis (Peng, Chen, & Peng, 2012).

When going one step further from internal network activities and server calls, alternative sources of data visualization become available. Such a data source is DNS log files. Lai et al. (Lai, Zhou, Ma, Wu, & Chen, 2015) used DNS log files in a large campus network to find out tendencies of the web users. Internet routing protocol data is one source of data used to detect BGP routing anomalies and for planning large-scale networking decisions. Ren et al. (Ren, Kristoff, & Gooch, Visualizing DNS traffic, 2006) used DNS query data gathered from a diverse set of caching servers in order to provide situational awareness for system administrators.

The majority of the designs use only one type of data such as netflow data or IDS data, but there are also some designs which use multiple data sources. NetSecRadar (Zhao, Zhou, & Shi, 2012) uses netflows, firewall data and host health status data. Netvis (Kan, Hu, Wang, Wang, & Huang, 2010) uses IDS data together with a huge department and user management data. Visual Firewall (Lee, Tros, Gibbs, Beyah, & Copeland, 2005) uses Firewall data along with IDS data. Dang and Dang (Dang & Dang, 2014) uses web site hierarchical structure data along with multiple web site vulnerability scan results. Tamp (Wong, Jacobson, & Alaettinoglu, DSN 2005, 2005) combines BGP routing data with IGP data, network traffic data and internet routing policies.

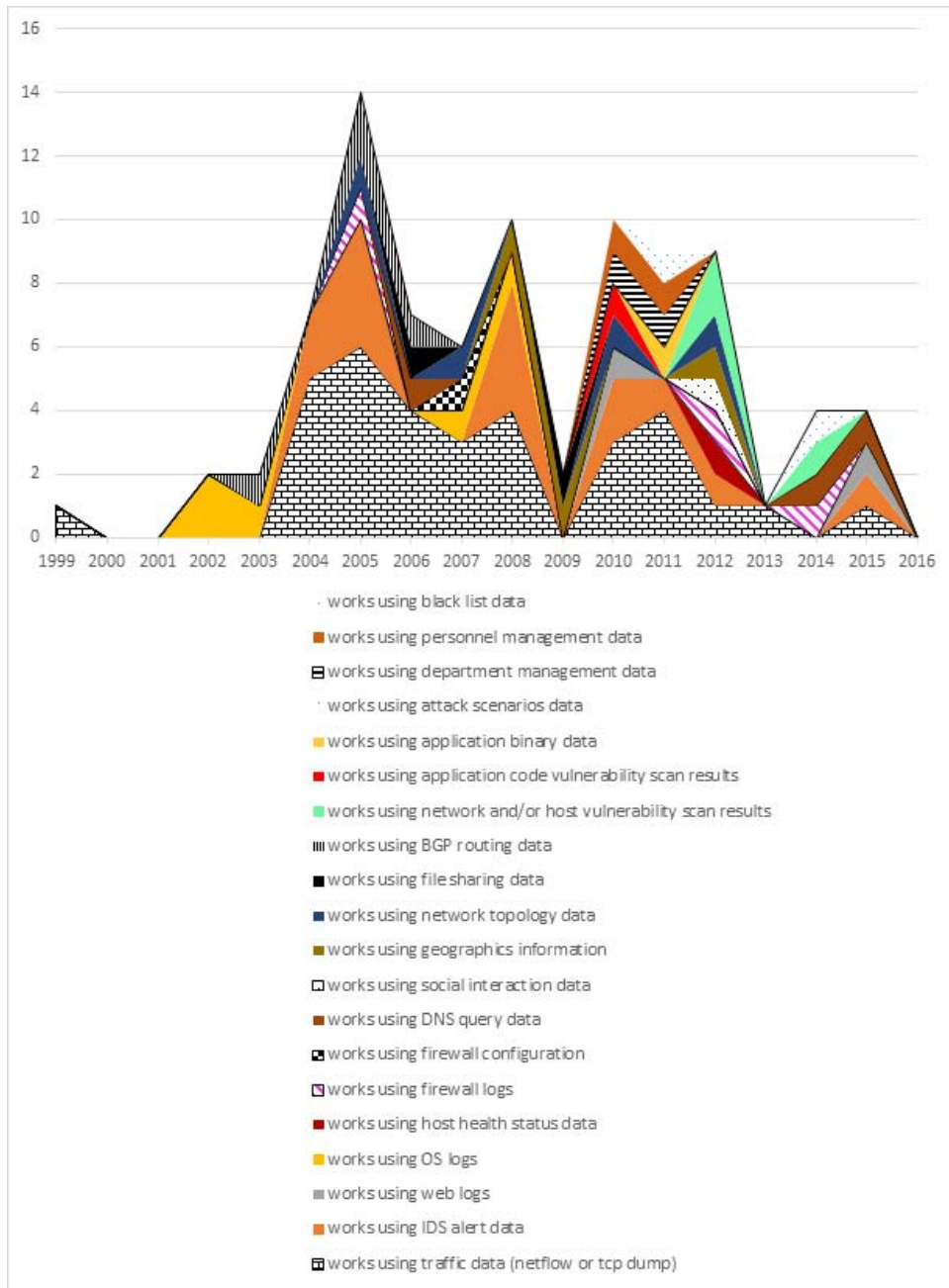


Figure 12 - Distribution of selection of data sources for visualization designs over years

When protocol distribution of studies using network traffic data is examined, it is seen that TCP traffic data is the main source for the majority of network traffic visualization systems. TCP protocol data is used as the main and/or only source for these systems, such as Security Quad and Cube (Chang & Jeong, 2011), Abdullah et al. (Abdullah, Lee,

Conti, & Copeland, IAW'05, 2005). InetVis (Riel & Irwin, 2006) extends this by including ICMP and UDP protocols.

While the majority of the network traffic visualizers do not attempt to isolate the data in terms of traffic types, HTTP(S) traffic aggregation and visualization, Hviz (Gugelmann, Gasser, Ager, & Lenders, DFRWS 2015 Europe, 2015), designed by Gugelmann et al. is an interesting example since it distills HTTP traffic from overall traffic data to visualize the web browsing activities of the users. It may be a good idea to make designs for other traffic types, such as streaming data traffic and/or bit torrent data traffic.

Since the log files are occasionally big, it is always necessary to find a way to reduce the data size. Another way of reducing the data is using only some part of the data, such as Abdullah et al. (Abdullah, Lee, Conti, & Copeland, IAW'05, 2005). Abdullah et al. (Abdullah, Lee, Conti, & Copeland, IAW'05, 2005) uses only the packet header part of the data to show port activity and eliminate the rest of the data.

Many of the visualization studies focus on generic data formats. However, visualization of data sources should not have to be generic. There are visualization systems which visualize the data belonging to in-house applications. For example, Impromptu (Rode, et al., 2006) visualize the data of a file sharing application developed as a test bed for the visualization system.

Visualization of log files of middleware servers is valuable in terms of security. However, most of the server applications/types are not given enough importance in security visualization domain. Web server log data is used by Alsaleh et al. (Alsaleh, Alarifi, Alqahtani, & Al-Salman, 2015) along with attack logs as a part of dashboard visualization. Ballora and Hall (Ballora & Hall, 2010) visualize the web log data along with sonification technique. This design aims to help intrusion detection activities and may handle both delayed and real time data. Other alternatives of server log files, such as application server logs, proxy server logs, mail server logs may be the source of novel security visualization designs.

Operating system logs are also valuable in terms of security. However, using system logs belonging to single or multiple hosts and/or servers is rare. Tudumi (Takada & Koike, 2002) uses Sys log file to gather network access to a server, Wtmp log-file to gather user log ins and log outs to the server and Sulog log file to gather user substitution messages.

Independent of the data source, most frequently visualized attributes are time, source IP, source port, destination IP, destination port and classification of the event, such as alert type. Including other TCP fields such as RST, FIN, ACK, SYN as in P3D (Nunnally, et al., 2013) may allow better understanding and increase the detection of attack patterns (scenarios) either manually or automatically.

2.4.2.3 Findings Related to Display Types

Examining the display types of the visualization works results in the following findings. Although each of the selected visualization design provides one or more graphics illustrating either the actual picture of the designed software display or graphical illustration of the presented design, it is difficult to capture the repeatedly used visualization attributes looking at those graphics. The complexity of the provided images from earlier work is variant. Some of the graphics accommodated more than one display property which increased the difficulty of capturing the useful display properties. There are also other difficulties about using the graphics from earlier work, such as copyrighting issues. Due to the listed difficulties, as a contribution to the area, a graphical library consisting of 51 images was prepared by using the hand drawing method. The hand drawn figures were converted to computerized images using Adobe Illustrator software. This graphical library is demonstrated in parts in Figure 13, Figure 14, Figure 15, and Figure 16.

Having a set of security visualization illustrations may serve many purposes. The motivation for creating these graphics set include using them during requirement analysis for capturing security visualization requirements, during the design of novel work, during all phases of security visualization studies to improve the communication of display properties, and for educational purposes.

The findings and contributions concerning the various types of display elements are as follows: Each of the illustrations in the graphics set corresponds to captured simplified property(ies) commonly used in security visualization designs. Some of the previous visualization designs depend on simple graphical charts, like pie chart or histogram, Figure 13 (a,b,c,d). For example, Abdullah et al. (2005) uses histograms to visualize the network traffic. Specifically, they visualize the aggregated port activities and demonstrate that time-dependent aggregated histogram charts can capture worm traffic and botnet activity. Line charts are used for web usage trend analysis in a campus network (Lai, Zhou, Ma, Wu, & Chen, 2015). In Net IQ Manager tool (Ferebee & Dasgupta, 2008) histograms are used to visualize the number of events for each host as a part of security trend analysis. Although, these charts look simple, the visualization designs in this group are highly comprehensible. These type of designs mostly focus on only some part of the data which results in clear understandability.

Parallel axis views allow visualization of multi-dimensional data where hosts are shown as nodes and flow of them are shown between vertical parallel axes, Figure 13 (e), Rumint (Conti & Abdullah, 2004), Visflowconnect (Yin, Yurcik, Treaster, Li, & Lakkaraju, 2004), IDSRainStorm (Abdullah, Lee, Conti, Copeland, & Stasko, 2005), Krasser et al. (Krasser, Conti, Grizzard, Gribshaw, & Owen, 2005). It is possible to visualize the end to end flow between external world and internal hosts by using three parallel axes together for external hosts, internal hosts, and external hosts respectively. This enables flow direction visualization, Figure 13 (f). In general, the x-axis is reserved for time dimension in parallel axis views. Some additional display features are included

in some designs for different purposes. For example, colored lines, Figure 13 (j) are used pointing out interactions coming to or going out from specific ports to help users “recognize and diagnose the problems” (Nielsen, 1995). Animation of network flow data over time is used to find out trends and detect anomalies Visflowconnect (Yin, Yurcik, Treaster, Li, & Lakkaraju, 2004). Rumint (McRee, 2008) uses parallel axis view to visualize a massive amount of network traffic data. This design allows selection of visualized attributes as text rainstorms for each axis which would “help recognize and diagnose” (Nielsen, 1995) some specific type of attacks. Krasser et al. (Krasser, Conti, Grizzard, Gribshaw, & Owen, 2005) uses both 2-D and 3-D parallel coordinate plots in combination with time varying scatter plots to monitor large-scale network traffic. The effective use of labelling, fading, scaling, and animation are also investigated in these designs in order to improve the visualization quality of large-scale network traffic, Figure 13 (g). Unlike other parallel coordinate systems, Krasser et al. (Krasser, Conti, Grizzard, Gribshaw, & Owen, 2005) show the protocol type by using coloring and the packet length by using vertical lines at the end of parallel axis connecting lines, Figure 13 (h), in the same view. Displaying these additional attributes minimizes the “requirement of remembering” during navigation among multiple views and increases the overall “recognition” (Nielsen, 1995).

IDS Rainstorm (Abdullah, Lee, Conti, Copeland, & Stasko, 2005) is an advanced parallel axis view based design which includes several parallel axes representing IP address groups. Horizontal dividers exist in this design to isolate departments, Figure 13 (i). IDS Rainstorm (Abdullah, Lee, Conti, Copeland, & Stasko, 2005) takes its name from its rainstorm like display. The area between axes is reserved for time varying number of IDS alarms generated for each IP for a time frame. Incorporating multiple axes in a display in this way allows visualization of IDS alarms in large networks.

Some of the designs use glyphs as a less important attribute of the visualization. For some other designs the overall design is based on the use of glyphs, for example, Clockview (Kintzel, Fuchs, & Mansmann, 2011, July), Erbacher et al. (2002), and, Erbacher (2003). The latter group may also have their own glyph designs instead of using standard shapes. For example, in Clockview (Kintzel, Fuchs, & Mansmann, Monitoring large ip spaces with clockview, 2011, July), clockview shaped circular glyph design divided into 24 parts is used to indicate the hourly traffic rate for each host in a matrix shaped display. This resulted in 24 times fewer number of cells for the total number of hosts. Erbacher et al. (2002) use a set of arrow designs which represent various network behaviors for intrusion and misuse detection purposes.

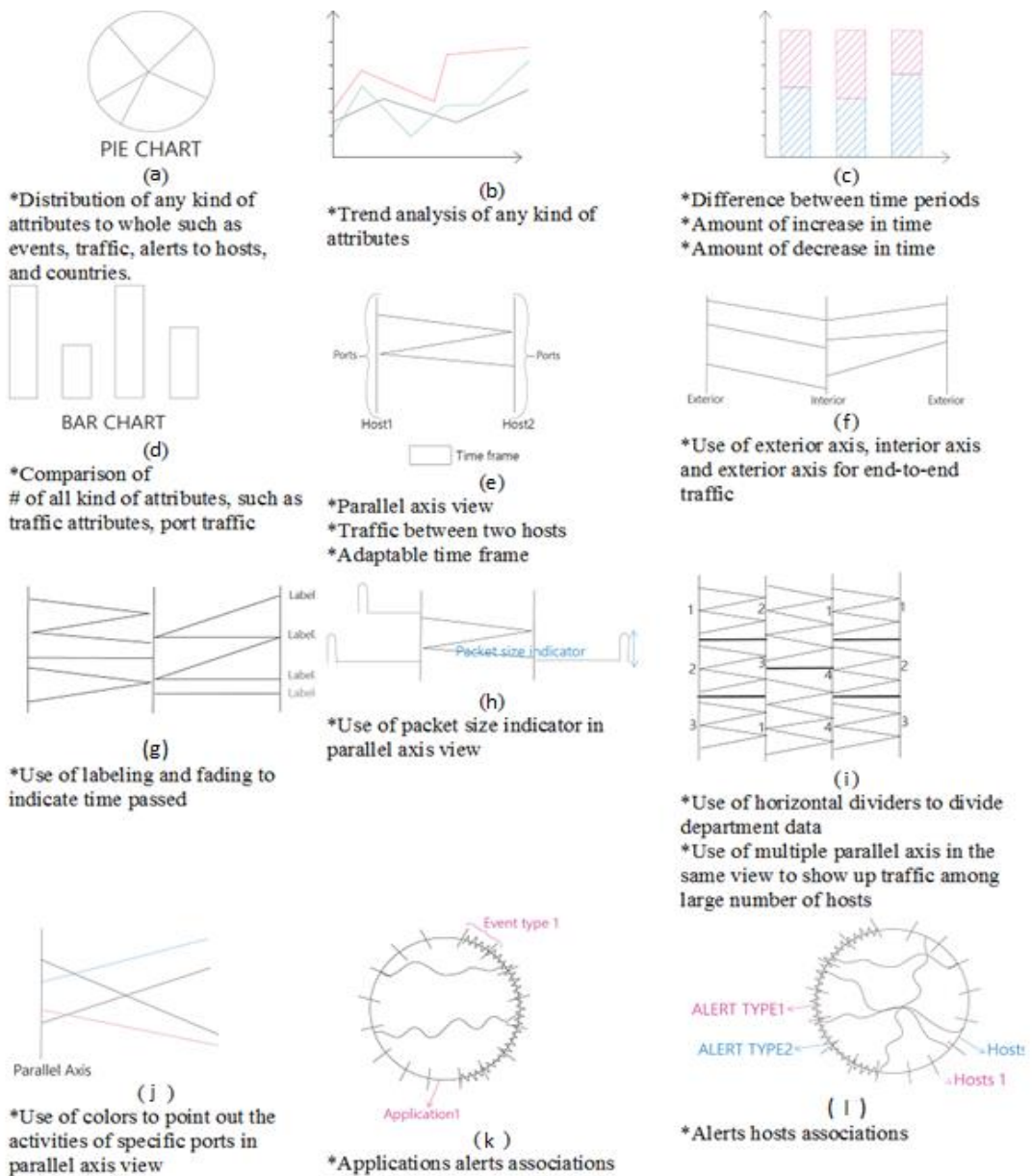


Figure 13 - Graphical illustrations of simplified display properties- Part 1

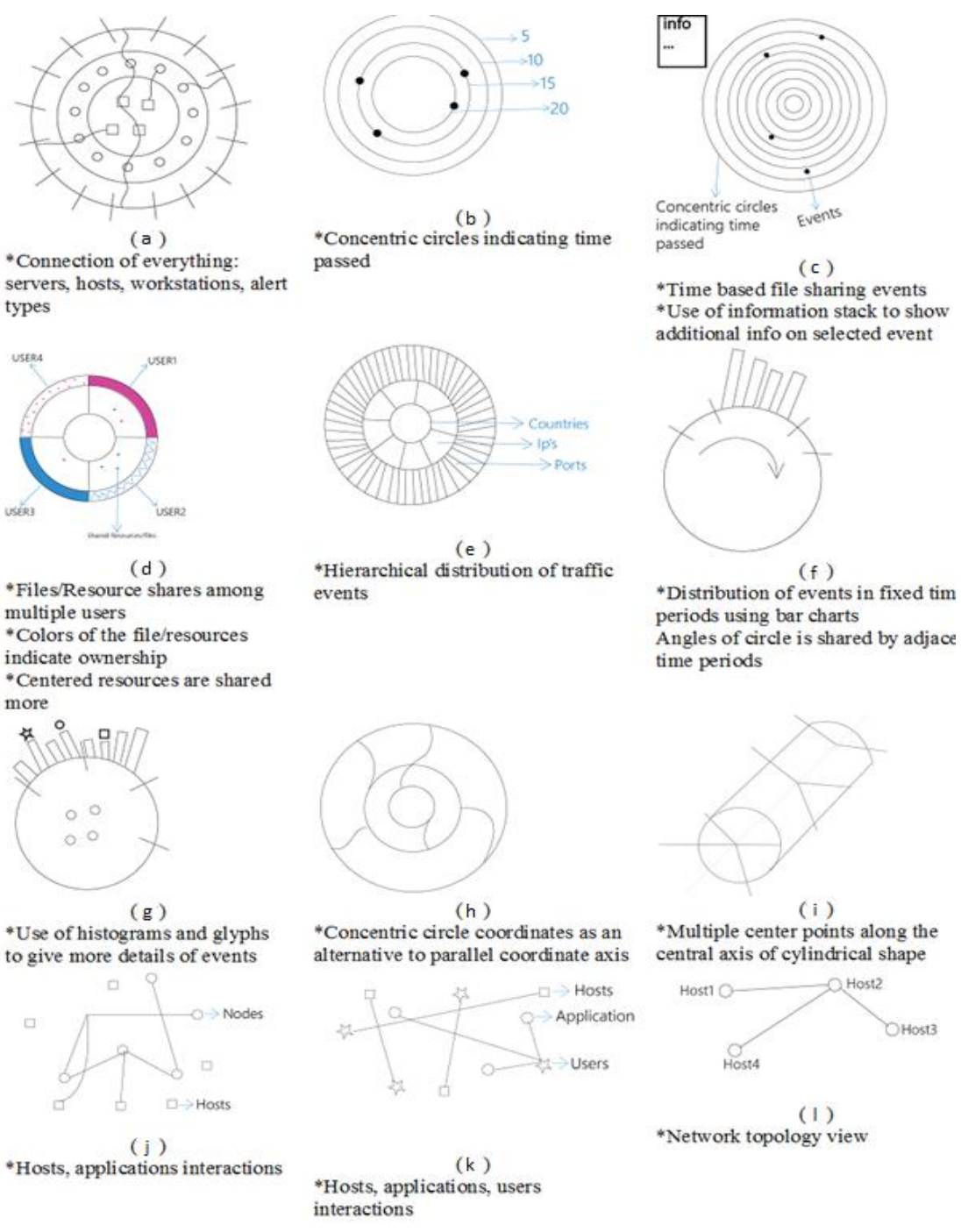


Figure 14 - Graphical illustrations of simplified display properties – Part 2

Radial (circular) design is another display type which is popularly used in the earlier works in various manners and for different purposes. Radial structures are, in general, proper to answer questions involving three terms; where, what, and when. Addition of other visual elements such as information stacks improve such designs by allowing reaching to details such as Avisa (Shiravi, Shiravi, & Ghorbani, 2010) and Avisa2 (Shiravi, Shiravi, & Ghorbani, 2011). Vizalert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005) is an IDS data visualization system having a radial shape, Figure 13 (k, l), Figure 14 (a,b). It gives answers to questions what, where, when, and how by placing the location of the alert on the map, the time on the concentric circles and the type of the alert to the angle of the circle. It allows multiple views simultaneously such as displaying alerts based on snort groups in one radial view and displaying alerts based on snort classifications in another view. Erbacher et al. (2005, October) uses radial display to present IDS data. The concentric rings in this design map to time units. As the time passes the intensity of the rings gets smaller to reduce the impact of older records. This design is also capable of animating the data and allows selection of various display parameters including the number of the rings.

Real-Time Visualization System for Network Security, NetsecRadar (Zhao, Zhou, & Shi, 2012) allows real-time visualization of intrusion detection alerts. The designers of NetsecRadar manages to visualize the hosts, alert type, and the histogram of attacks in one circular chart design. The colored arcs in this design show network security event types, the bars drawn on each arc show the number of events for each event type in the sampling time period, the colored nodes in the center of the graph show the servers or workstations in the selected corporate network, and the curves drawn between the central points and the arcs indicate the source and destination addresses of the selected security events, Figure 14 (a). Radial Traffic Analyzer (Keim, Mansmann, Schneidewind, & Schreck, 2006) uses a radial design for the monitoring of the current state of hosts and servers. In this design, each radial ring is mapped to one traffic attribute assigned by the user, Figure 14 (e). The relatively more important attributes are selected to be displayed in the inner rings. The hosts share the angle parts based on their amount of traffic for that specific attribute.

Impromptu (Rode, et al., 2006) uses a radial display to visualize file sharings. The angles of the radial are shared by the users which are also assigned different colors, Figure 14 (d). The file icons belonging to any user have the same color with the user. As a file is shared more and more, it gets closer to the center of the radial display. The files which are not shared at all, stay at the outer part of the radial shape. The file icon blinks with the color of the user who is actively working on itself. As the users get involved with the file in time, a ring is formed around the file icon having the users' color. This property allows identifying the users' history over the files. The angle parts of the users are marked by user characterization signs which correspond to unknown user, wireless user and wired user aiming to find out suspicious user activity. The thickness of the edges for each angle part corresponds to the level of user activity for that angle. This design "matches the system with the real world" (Nielsen, 1995). Thus, it is easy to understand. Being

able to show user history also increases “recognition of activity”. This design is aesthetically in well shape and seems to be effortless to use for even novice users. Tri and Dang (2009) uses a radial shape to visualize file sharings in a local area network, Figure 14 (c). This model does not include the user point of view. Instead, file events are included on an adjacent page to remove the need for checking them from the event viewer. Their approach minimizes the user’s memory load by making events visible in the same page. The visualization is extended by human readable explanations which appear on top of the design to reduce the learning curve of the users of the design.

Lu et al. (Lu, Zhang, Huang, & Fu, 2010) proposes concentric-circle display as an improvement to parallel coordinates in CCScanViewer, Figure 14 (h), (Lu, Zhang, Huang, & Fu, 2010) (Zhang, et al., 2009). Lu et al. uses CCScanViewer to demonstrate various types of network scans and DDOS attacks. Their use of circular view is different than the rest of the circular designs. They use concentric circles analogous to x, y, z axis from a 3D scatterplot.

Cylindrical coordinates security visualization, CCSVis (Seo, Lee, & Han, 2014) is a design based on cylindrical coordinates visualizing DNS queries. Cylindrical coordinates allow monitoring of multiple subjects, such as multiple hosts, multiple DNS servers simultaneously in one graph, Figure 14 (i), without totally overlapping the data by means of having multiple center points along the center line of the cylinder. This type of visualization allows also catching the interactions of multiple hosts with some exterior callers simultaneously.

Another group of visualization is based on node-link diagrams. An example of this category is by Mansman et al. (2008). Mansman et al. uses link analysis in which some particular applications are visualized as nodes. The behavior of hosts is visualized by showing their interaction with the nodes using a force directed graph, Figure 15 (f). Hviz (Gugelmann, Gasser, Ager, & Lenders, DFRWS 2015 Europe, 2015) is another example, which visualizes the web browsing activities by illustrating the visited web pages as nodes and links between them as links. Enterprise Network Activities Visualization, Enavis (Liao, Blauch, Striegel, & Thain, 2008) focuses on the enterprise security data and have similarities with Mansman et al. Both designs use node-link diagram to visualize the hosts, users, and applications in an enterprise, Figure 14 (k), aiming to show connections among them to answer the basic question of “who does what on where”. Nagios (Josephsen, 2007) uses a node link type of design to visualize the topology, Figure 14 (l), and determine the availability of topology items. Trust visualisation service for online communities, TrustVis (Peng, Chen , & Peng, 2012) is a design based on the network topology visualization. In this system, Figure 15 (c), nodes are the users and links are the interactions among them. This design aims visualization of trust management in a network. TrustVis allows unique profile drawings for each user rather than having a female and a male user type icon. Availability of unique profile drawings increases usability and users’ recognition level and decreases “recall”. It also ends up with a more aesthetic design compared to having single type of user icon for every user.

Dang and Dang (2014) visualizes the web site topology in a hierarchical manner using the node-link diagram, Figure 15 (d). Lai et al. (Lai, Zhou, Ma, Wu, & Chen, 2015) uses a node-link graph to visualize the most active IP addresses in three DNS servers, Figure 15 (b), and to point out most popular domain names, Figure 15 (e), in two adjacent graphs. Visualizing Packet-Process Correlation, Portall (Fink, Muessig, & North, 2005) visualized the selected set of client and server hosts as the nodes in a node-link type of display, Figure 14 (j). The hierarchies between the processes of the hosts are also shown in the same view. This design can visualize the end-to-end traffic in process level for a small number of processes, (around 40), due to display size limitations.

Node link type designs are suitable to represent all kinds of internet routing activities among devices and systems. Threshold and Merge Prefixes, Tamp (Wong, Jacobson, & Alaettinoglu, DSN 2005, 2005, June) combines node-link diagrams with the animation to simulate the internet routings and colors are used to represent packet pathways, Figure 15 (h). The size of the links get thicker as the number of prefixes using any link increases. This tool is designed to diagnose the internet routing algorithms either in real time or using historical data. BGPlay (Colitti, Di Battista, Mariani, Patrignani, & Pizzonia, 2005), and Linkrank (Lad, Massey, & Zhang, 2006), Figure 15(i), are other examples which use node link type of displays to present BGP data.

BGP Routing Visualization, BGPlay (Colitti, Di Battista, Mariani, Patrignani, & Pizzonia, 2005) does not only visualize the paths among autonomous systems, AS's, it also points out changed and unchanged paths in a time frame. In this design dashed lines are used to show the unchanged paths, Figure 15 (f). Information Visualization System for Monitoring and Auditing Computer Logs, Tudumi (Takada & Koike, 2002) extends an ordinary node-link diagram by including concentric disks, Figure 15 (g). The nodes which stay on the bottom disk represent user substitutions and nodes which stay on upper disks represent access to hosts and user log-in information. Positioning the nodes on concentric disks results in more compact appearance compared to arbitrarily laying out the nodes on display space. Network Intrusion Visualization Application, Niva (Nyarko, Capers, Scott, & Ladeji-Osias, 2002), which is a 3-D node link based intrusion detection visualization system, allows the user to navigate within the dataset with the haptics integration. Haptics integration results in the ability of touching and manipulating the computer generated objects. Thus, this sense of touch improves experiences of users.

Matrixes, grids or x-y (-z) plots are commonly used in the security visualization domain. Existence Plots (Janies, 2008) uses two x-y diagrams together to visualize the inbound and outbound port activity, Figure 15 (k). In this design, y axis is reserved for logarithmic scale of either 2^{16} inbound or outbound ports and x axis is reserved for the time dimension.

Use of logarithmic scale reduces the required space for the large range of port values. Correlation Layers for Information Query and Exploration, Clique (Best, Hafen, Olsen, & Pike, 2011) visualizes the time-aggregated network traffic data using an x-y axis plot.

It is assumed that the variance of counts around mean should be constant over time. However, as the number of traffic increases this assumption does not hold. So, the designers visualize square root of the aggregated values instead of aggregated values to reduce this effect for large values.

Sybil attack results due to malicious hosts which act as other hosts by impersonating their identities or using other fake identities. Lu et al. (2011) uses 2D matrixes to visualize the Sybil attacks, Figure 16 (b). Generally, this type of attack is demonstrated by topology diagrams. Lu et al. (2011) visualize time variant network topology and detect patterns which point to Sybil attacks in their work. PolicyVis (Tran, Al-Shaer, & Boutaba) uses the x-y axes to visualize the complicated allow/deny type rules of firewalls through the use of an easily readable design, Figure 15 (m). In general, matrix type of displays uses color of matrix cells which indicate the severity or number of the events. The x-y axes are commonly used for time-port, time-IP, source IP, destination IP, and port-IP respectively. Some of the matrix designs use additional lines to connect the matrix cells with other matrix cells, Time-based Network Visualizer, TNV (Goodall, Lutters, Rheingans, & Komlodi, 2005) or with other display elements, Figure 16 (e), Visual Information Security Utility for Administration Live, Visual (Ball, Fink, & North, 2004).

Securescope (Ferebee & Dasgupta, 2008) is another design which has a matrix type display. In this design, a matrix like 2-D grid is used visualizing the location of the hosts using the department names and the flooring number of the building, Figure 15 (n). Such a design is useful for enterprise network management purposes. Similar to TNV, the hosts which stay in the 2-D matrix are connected to nodes reserved for different communication protocols through lines. The size of the protocol nodes shows the amount of network traffic for that protocol. As the number of hosts or events increases, these type of designs combining node-link and matrices become complicated. If there are additional connecting lines between matrix cells, the understandability would decrease even more due to the overlapping lines. Visual analytics of firewall log events, Vafle (Ghoniem, Shurkhovetsky, Bahey, & Otjacques, 2014) adopts a 2D matrix display including custom heatmap view, magic lens interaction, clustering, multi-level navigation, and on demand details techniques, which increase its usability, Figure 15 (o). The author also thinks that it is mandatory to use vertical and horizontal scrollbars in these matrices, or grid type of displays as in the Vafle case.

2D and 3D scatterplot designs allow visualization of relatively higher size of data, because data points consume less space in these display types, Figure 16 (b). For example, Xiao et al. (2006) uses 2-D scatter plots to visualize network traffic attributes. Security Quad and Cube (Chang & Jeong, 2011) uses a cube structure to visualize network anomalies Figure 16 (c). The attributes used in this visualization are source IP and port and destination IP and port. Netbytes viewer (Taylor, Brooks, & McHugh, 2008) uses a 3-D impulse graph, which is similar to 3-D scatterplots using the port, time, and bytes attributes.

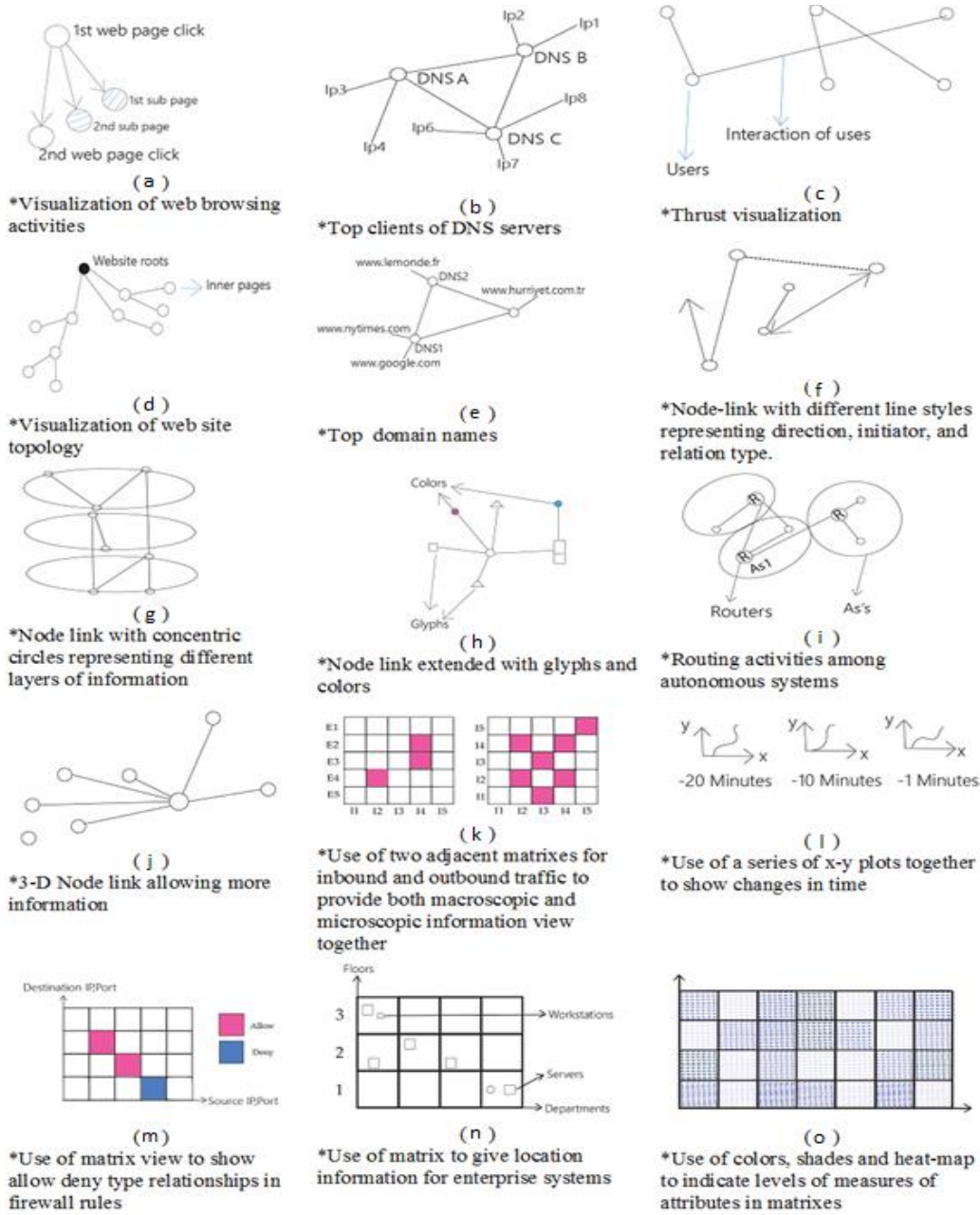
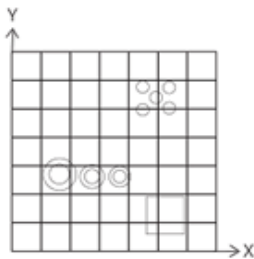
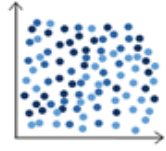


Figure 15 - Graphical illustrations of simplified display properties – Part 3



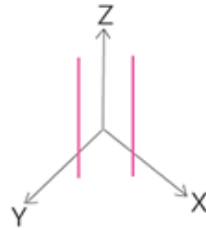
(a)

*x-y matrix with use of glyphs



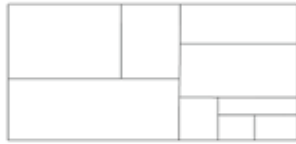
(b)

*Use of scatter plots to visualize large data in small space
 *Comparison of signatures of patterns through the use of x-y plots
 *Visualization of network topology to detect Sybill attacks



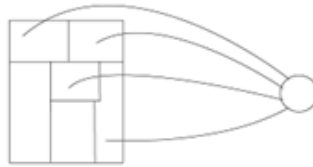
(c)

*3-D x-y-z scatter plots allows comparison of 3-D signatures



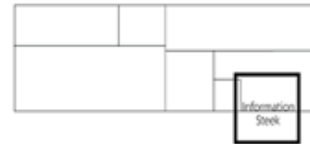
(d)

*Visualization of hierarchical data using tree-maps



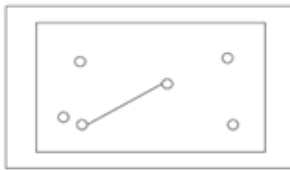
(e)

*Connecting parts of tree-maps to external objects such as external IP's



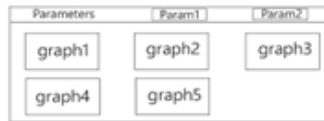
(f)

*Use of information stack for additional information in tree-maps



(g)

*GIS Based visualization
 *Connections between physical points of map



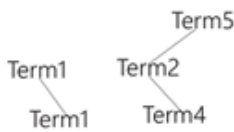
(h)

*Dashboards allowing multiple views



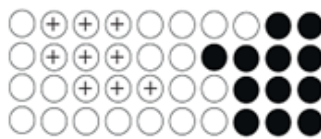
(i)

*Advanced 3-D Visualization and techniques
 *Camera view, stereoscopic view.



(j)

*Flying terms showing relations of terms in a log file



(k)

*Use of SOM diagram to classify traffic attributes



(l)

*Grayscale imaging of binary code for malware detection

Figure 16 - Graphical illustrations of simplified display properties –Part 4

Tool for Port-Based Detection of Security Events, Portvis (McPherson, Ma, Krystosk, Bartoletti, & Christensen, 2004) uses scatterplots to visualize port activity. It breaks the port number into two-byte x-y location on the plot. Compared to Abdullah et al. (2005)'s simple histogram based port activity monitoring design, these scatter plots require much more effort for data preparation and understanding phases. Independent of the designs, as the amount of the data increases the readability of these matrixes decreases. However, when reinforced with pattern evaluation and detection methods, as in the Lu et al. (Lu, Wang, Dnyate, & Hu, 2011), such 2-D, 3-D matrix type of displays become more appropriate for automatic and manual detection of attacks.

Intrusion Detection and Analysis Using Histograms, IDGraphs (Ren, Gao, Li, Chen, & Watson, 2005) is a design which has a scatter plot like display aiming to detect network intrusions using traffic data. Most of the scatterplot and matrix like designs use IP's and/or ports as the second dimension in addition to the time dimension. IDGraphs uses a different approach. It uses the ratio of number of SYN to number of SYN-ACK as a second dimension to detect attacks such as TCP SYN flooding, worm outbreaks, and port scanning. It is based on histogram technique including the brightness level of the cell based on the value of the data.

Although the data format is identical for both internal network traffic and network traffic with external nodes, two adjacent matrixes are used to visualize the IP traffic between internal hosts and external IP's in IPMatrix (Koike, Ohno, & Koizumi, VizSEC 05, 2005). It was necessary to find an approach which uses the display space effectively to show the 32 bit IP address information for a large dataset. Taking every bit individually would be complicated and unreadable. The information identifying the traffic differs for both traffic types. Therefore, x,y dimensions are selected as 3rd 8 bits, and 4th 8 bits for internal traffic matrix and 1st 8 bits, and 2nd 8 bits for the matrix showing the traffic with external hosts.

Information Visualization Tool for Intrusion Detection, IDtk (Komlodi, Rheingans, Ayachit, Goodall, & Joshi, 2005) is a design for intrusion detection based on a 3-D x,y,z plot using glyphs. It extends 3D capabilities by including size, shape, opacity, labels, and colors and visualizing more than three data attributes, such as priority, classification, source IP, source port, destination IP, destination port, and protocol, Figure 16 (a). Network Host-Centered Anomaly Visualization Technique, Svision (Onut & Ghorbani, 2007) is a 3-D design for the visualization of hosts to find out anomalies in their uses of services. While 2-D is enough to show the hosts' service usage, the third dimension is required to show the level of real traffic load for that host. This design uses only two colors to represent internal and external hosts, but it uses changing color intensity from dark to light representing the time of activities. Spinning Cube of Potential Doom (Lau, 2004) is a 3-D scatter plot design, Figure 16 (c), monitoring port activity. 3-D scatter plots are more difficult to understand using 2-D screens due to overlaying problems. 3-D full immersion visualization systems are superior compared to ordinary 3-D visualization systems, since they provide virtual reality where the user may get better

involved with the data. Ballora and Hall (2010) provides a 3-D full immersion visualization which allows the user to explore a huge amount of data, Figure 16 (i).

Visual comparison of patterns and/or signatures is important for detecting any kind of anomalies. Various features are offered to users for easy comparison of data patterns. For example, Muelder et al. (Muelder, Ma, & Bartoletti, 2006) uses side by side comparison by providing view spaces having same size side by side in the same display, visualizing 2-D scatter plot views of patterns, Figure 16 (b). Muelder et al. also uses wavelet scalogram which is a graphical representation type of data based on mathematical conversions. This type of graphs have sharper edges causing better comparison of patterns compared to scatterplot graphs. Muelder et al. shows that similar scans have similar wavelet scalograms, and dissimilar scans have totally different wavelet scalograms.

While some of the displays or visualization styles repeatedly emerge with some modifications in various works, some display types are unique and, thus, more original. Use of ternary plots to visualize the network traffic data is an innovative approach by Whitaker and Erbacher (2011). The ternary plot is a general purpose graph type which is suitable for data composed of three attributes. It has a triangular shape. A point is plotted on the triangular shape based on the percentage of each attribute value. Whitaker and Erbacher (Whitaker & Erbacher, 2011) take the port, size, and protocol as the three attributes for ternary visualization. They further extend standard ternary visualization by adding the time attribute. As time passes, the points on the triangle are animated. This addition provides a better understanding of network events.

Some type of visualization of malicious activity uses images of the codes, executables or execution log files as the display elements. Grayscale imaging of binaries is displayed to detect malwares in Nataraj et al. (2011), Figure 16 (l). The aim of this type of display is to enable a visual classification of software programs because malware programs have similarities with each other in terms of their binary structure which reflects their software architectures and implementations, and in terms of their execution log files which reflect the results of executed statements.

Treemap displays are commonly used to visualize hierarchical data, Figure 16 (d, e, f). Network Security Management Visualization Tool based on treemap, Netvis (Kan, Hu, Wang, Wang, & Huang, 2010) has a 2D treemap display that detects abnormal patterns in a network while supporting network management activities. NFlowViz (Fischer, Mansmann, Keim, Pietzko, & Waldvogel, 2008) visualizes the network traffic similar to Visual (Ball, Fink, & North, 2004), but unlike it, it uses treemap cells instead of matrix cells to represent hosts. Thus, it can show groups of hosts having same prefixes and the amount of network traffic for each host in the same view. HNmaps (Mansmann F., Keim, North, Rexroad, & Sheleheda, 2007) is an interactive treemap design which focuses on visualization of hosts' interaction with various AS's. This tool can visualize traffic coming from any kind of hierarchical network structure, such as country-wise traffic

including the ASs or a campus network including departments. In these designs, treemap cells are connected to other treemap cells via colored lines indicating the amount of network traffic. Using different tones of the same color indicates the amount of traffic, however, this results in a complicated graphic. Choosing totally random line colors ends up with less information, losing the traffic amount information, but leads to a more readable graph. Histomap (Mansmann F. , Keim, North, Rexroad, & Sheleheda, 2007) uses large scale network traffic data among multiple AS's to visualize continentwise network traffic. The number of IP's assigned to each country is represented by the size of treemap cells and the number of incoming traffic to each cell is represented by the cell color. Using the treemap display type this way, with large datasets, ends up with a visualization model which may be used for monitoring and network planning rather than detecting threats, and anomalies. While, the designs which visualize internet routing protocol data, adopt a node-link type of display in general, Experimental Visual Anomaly Detection, Elisha (Teoh, et al., 2003) has a totally different approach. It uses a quad tree approach, similar to treemap diagrams, dividing the display space according to IP address prefixes, and setting the size of the cells as the number of IP's reserved for that specific cell. It shows the internet routings by drawing lines between the cells. It is known that as the number of nodes increases, node link type of diagrams have scalability problems. The designers of Elisha (Teoh, et al., 2003) claim that one of the important benefits of their design is its scalability. Designs which aim to detect the anomalies in the Internet routing protocols are based on the fact that the user may detect the abnormal patterns by eye as the changes in the paths do not occur often. Since node link based designs better represent real life situations of this type of data, and are more user friendly, they would have much more shorter learning curves compared to Elisha (Teoh, et al., 2003). Treemaps are also used to visualize the vulnerability results. Nessus network level vulnerability results are visualized by NV (Harrison L. , Spahn, Iannacone, Downing, & Goodall, 2012), and application level (application code level) vulnerability results are visualized by Goodall et al. (Goodall, Radwan, & Halseth, VizSec '10, 2010). In the former case, workstation groups are taken as the higher level hierarchy elements and host IP's as the lower level hierarchy elements. The size and the color of the cells are available to associate with attributes, such as the number, and level of the vulnerabilities. In the latter case, vulnerability categories, such as input validation, encapsulation, encryption, and suspicious code are taken as the treemap top level hierarchy elements and application files are taken as the lower level hierarchy elements. The colors of the treemap cells indicate the severity of the vulnerabilities. The sizes of the cells indicate the number of vulnerabilities for the selected category.

Another group of visualization provides a set of views simultaneously in dashboard format, Figure 16 (h). An example of this type of design is Visual Monitoring of Network and System Security Sensors, Synema (Bousquet, Clemente, & Lalonde, 2011) visualized data from a distributed set of sensors such as Snort sensor, and SELinux sensor and encloses different types of visualizers for various types of sensors. Another example is from Alsaleh et al. (2015) visualization of an open source IDS (PhpIDS), and web server log data in various display types including attacker aggregation, bar view, attack

frequency view, IP aggregation, parallel coordinates, radial multiple source view, ring view, scatter plot, radial IP, treemap, treeview, and radial view. In terms of flexibility and efficiency of use (Nielsen, 1995), Synema (Bousquet, Clemente, & Lalande, 2011) has exceptional features. It provides the capability of creation of new frames as user requests in the display window, allowing simultaneous views of data coming from different sensors. Visual Firewall (Lee, Tros, Gibbs, Beyah, & Copeland, 2005) design is also a dashboard like design which is capable of displaying real-time traffic data, visual signatures, statistics information, and IDS alarms.

Another group of visualization designs is based on GIS displays. Li et al. (2012), uses geographic information systems, network topology graphs, bar charts, pie charts, dashboards, and attack patterns to provide an overall view for situational security (Teoh, Ranjan, Nucci, & Chuah, 2006). Dorothy project (Cremonini & Riccardi, 2009) uses Google Maps to visualize the locations of command and control, C&C, hosts, and satellites, Figure 16 (g). This is useful for some particular cases in which seeing the end-to-end traffic or global view using maps is more appropriate in terms of matching with the real world situations (Nielsen, 1995).

There are designs which use more advanced graphical models. Parallel 3D Coordinate Visualization, P3D is a design (Nunnally, et al., 2013) using stereoscopic 3D parallel visualization for network scans. Stereoscopic visualization models are superior compared to 2D and 3D models. 3D Stereoscopic Vulnerability Assessment Tool for Network Security, 3DSVAT (Nunnally, Uluagac, Copeland, & Beyah, 2012) is another stereoscopic design illustrating vulnerabilities of a group of nodes simultaneously in one display, Figure 16 (i). In this view, each host is represented by a cube. The pink, orange, and yellow regions represent host groups having different levels of vulnerabilities. The scatter plots show the highest CVE group number, such as level 3, level 5, which is assigned to each host group. The textures of the cubes represent various operating systems. Severity scores calculated for each host are shown using bar graphs. The hosts which stay in the stereoscopic region have the highest severity scores pointing out the vulnerabilities that may result in most severe actions. The design aims to provide a full perspective of vulnerabilities of the hosts in a network to the network managers.

There are some visualization designs which do not belong to these display type categories. For example, Flying Terms is one of them used in Ren et al.'s (2006) DNS traffic visualization design, Figure 16 (j). Flying Terms indicate the quantity of the traffic for each DNS query. It is a word cloud like text visualization technique which uses an x-y plot type background, capable of showing the change in queries in time. Ren et al. (2006) also adapted Chernoff Face Glyphs which are capable of showing 10 attributes in a 2D face display as part of a passive monitoring system. A series of glyphs is shown in this display type. If a face is quite different from previous faces, then it is a sign that an abnormal event may occur in the network.

Using various graphical filters over 2D or 3D visualization models to better represent abnormalities is being examined by some researchers (Alsaleh, Barrera, & Van Oorschot, 2008). In addition to improving the visual display, there are some techniques which rely on other senses. An interesting design related to network monitoring uses human aural, and visual pattern recognition ability simultaneously for higher rates of intrusion detection (Ballora & Hall, 2010).

The majority of display types require apriori knowledge on the data, such as the number of records, the number of dimensions etc. Girardin (1999) uses a totally different visualization system based on self-organizing map (SOM) diagrams which do not require prior knowledge of the data, Figure 16 (k). In this design, numerous attributes including time, packet size, flags, IPs, and ports are visualized in a rectangular shaped SOM diagram. The part of traffic data which point out some abnormal activities, such as high packet sizes, unacknowledged SYN requests, TCP connections which did not complete three hand shake communication protocol are grouped in the SOM diagram. Following this initial visualization, the user should investigate the details of the suspicious parts of the SOM to detect abnormal traffic activities.

Since security visualization designs have to be used, in general, both frequently and for long periods of time during the monitoring and analyses tasks, being visually appealing (Nielsen, 1995) would eventually enhance the usability of these designs. Among many other alternatives, Impromptu (Rode, et al., 2006), CCSVis (Seo, Lee, & Han, 2014), and Vafle (Ghoniem, Shurkhovetsky, Bahey, & Otjacques, 2014) stand out in this respect.

2.4.2.4 Findings Related to User Interaction

The interactivity of the selected studies is investigated theoretically rather than using experimental approach due to the difficulty of reaching an executable version to many of the designs. Few examples selected for this review study lack user interaction features, such as Abdullah et al. (2005), Security Quad and Cube (Chang & Jeong, 2011), and Dorothy Project (Cremonini & Riccardi, 2009). The interactivity of the other designs has various levels.

There may be different aims of the user system interactions, Figure 18. The first group of interactions act in getting user inputs. The most popular ways of interactions in this group enable selections of some aspects of visualizations by the users, as in NetsecRadar (Zhao, Zhou, & Shi, 2012). It is possible to select the time frame, or so called time window of the data in most of designs, such as Tamp (Wong, Jacobson, & Alaettinoglu, DSN 2005, 2005), and Inetvis (Riel & Irwin, 2006). In addition to time period, some designs which show discrete time intervals or which are based on aggregation of data over time, allow the users to change the time scale of the designs, such as in Inetvis (Riel & Irwin, 2006). Independent of the display type, majority of the designs, such as Alsaleh et al. (2015), allow selection of other parameters being continents, countries, sets of IP's, and ports, and alert types. Similar to parameter selection, 2-D or 3-D axis based designs

may allow definition of purposes for axes, such as PolicyViz (Tran, Al-Shaer, & Boutaba), Rumint (Conti G. , et al., 2006), and NIVA (Nyarko, Capers, Scott, & Ladeji-Osias, 2002). Parameter selection is a way of filtering the display data. Doing this more interactively, such as via mouse clicks is also possible, which is included as a feature in Radial Traffic Analyzer (Keim, Mansmann, Schneidewind, & Schreck, 2006) design.

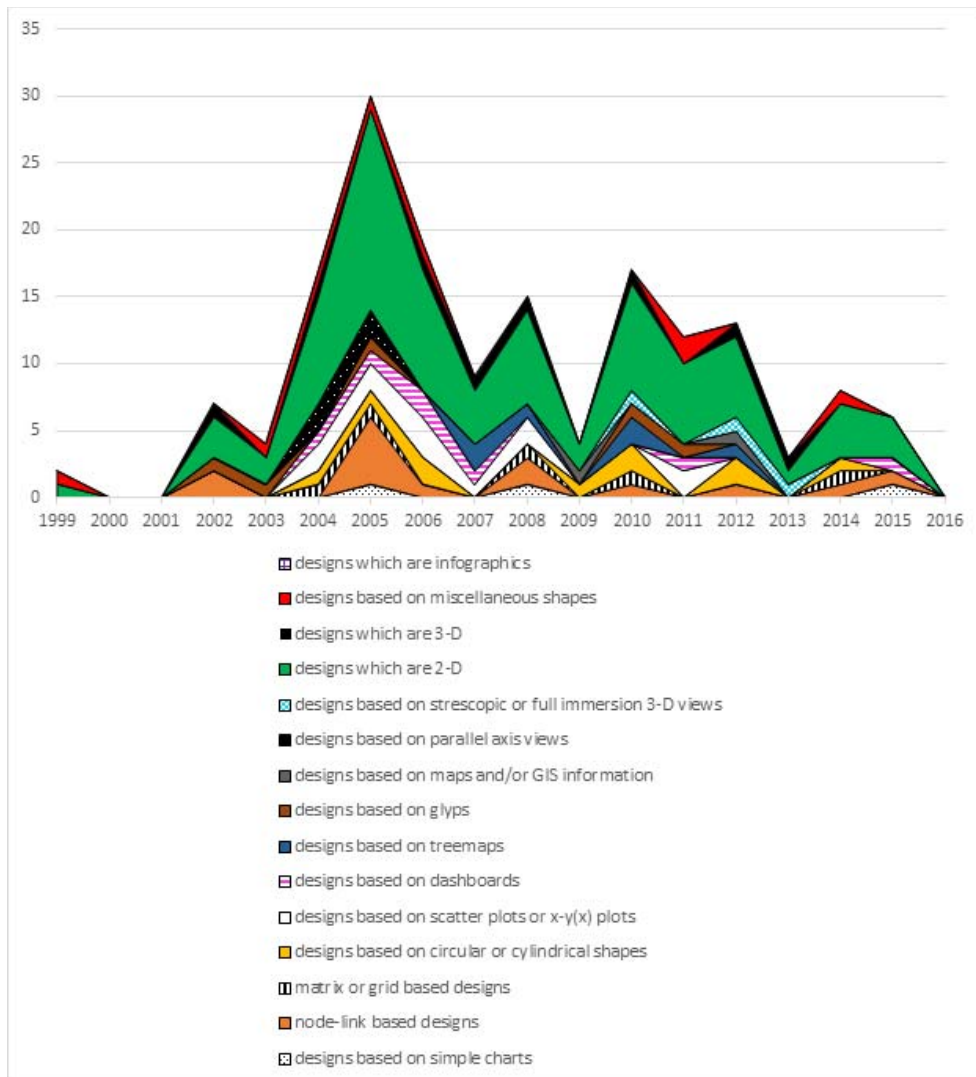


Figure 17 - Distribution of display types of security visualization designs over years

Generally, security visualization tools are designed to be used by advanced users. Adapting the visualization according to the viewers' expertise level is an extraordinary interaction design property which is proposed by Wong et al. (2010). This model takes

users' expertise level on network security, and adapts the security visualization views accordingly. In this model, as the user's expertise level increases, the number of attributes shown on the display increases, animation features are included, and level of interactivity also increases.

The second group of interactions help the user by providing better navigation. Since security visualization requires display of large sets of data in small space, navigation and zooming in and out are among the most necessary user interactions, which are provided for the users in designs, such as, IDSRainStorm (Abdullah, Lee, Conti, Copeland, & Stasko, 2005), Krasser et al. (2005), PortVis (McPherson, Ma, Krystosk, Bartoletti, & Christensen, 2004), Vizalert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005), Tri and Dang (Tri & Dang, 2009), and 3DSVAT (Nunnally, Uluagac, Copeland, & Beyah, 2012). In addition to other common interaction techniques, such as parameter selection, zooming in and out and navigation, both 2-D and 3-D scatter plot designs benefit from the drill down and drill up kind of user interactions, such as in NVisionIP (Lakkaraju, Yurcik, & Lee, 2004).

The third group of interactions aim improving the analysts' experiences by allowing searching or saving the data. While monitoring of attack patterns, IDGraphs (Ren, Gao, Li, Chen, & Watson, 2005) allows searching for similar traces in the overall data once a trace subset of network data is selected by the user through highlighting. This uncommon type of interaction which is called interactive query enables identifying distributed or recurring type of attacks, such as recurring traces from single or multiple sources. Sometimes interpretation of user findings using security visualization solutions may take time or may require further efforts such as comparison with other diagrams or recall of other data. Capability of saving discoveries and reusing those findings for later discoveries is a feature which may decrease this type of difficulty, increasing the overall usability of visualization designs. Such a property is offered by Xiao et al. (2006) allowing save of network patterns which are visualized using scatter plot diagrams.

Interactions focusing on giving feedback to the users form the fourth group. Some designs which encapsulate multiple display types allow users to select the display type, as in, Alsaleh et al. (2015). Some patterns of data would be more obvious in some displays. This interactivity feature provides the user the ability to use various display types for the same data. Majority of the node-link type of designs allow selection and replacement of nodes and links for better view, such as Mansman et al. (2008). Another interactivity related to giving feedback is not based on user actions, but is based on automatically highlighting some display elements. Examples to this type of interaction include blinking the file icons in the assigned color of the shared user in a file sharing application, as in Impromptu (Rode, et al., 2006) or highlighting an autonomous system (AS) as the number of network activity passes a threshold value in a network routing monitoring application as in Bgp Eye (Teoh, Ranjan, Nucci, & Chuah, 2006). Some of the designs combine interactivity with the animation by animating a part of the data in

time, such as Avisa (Shiravi, Shiravi, & Ghorbani, 2010), Tamp (Wong, Jacobson, & Alaettinoglu, DSN 2005, 2005), and Erbacher et al. (2005, October).



Figure 18 - Types of user interactions found in security visualization designs

Enabling a summary view of the data upon user request is a specific type of user interaction provided in some designs, such as IDGraphs (Ren, Gao, Li, Chen, & Watson, 2005). Tooltips is also used to display full label of information in case of displaying data in small segments of display area, such as in Radial Traffic Analyzer (Keim, Mansmann, Schneidewind, & Schreck, 2006). This design also offers popup menu based displaying of detailed information for a segment, which is accessible upon user request. Including human readable explanations of the discovered patterns also is an uncommon way of interaction with the user. This approach increases the understandability of the models, such as in (Tri & Dang, 2009), and reduce the learning curve of the users.

2.4.2.5 Other Notes

In this part there are review results related to encapsulation of classification or statistical analysis methods as part of visualization systems. The majority of the works, visualize the already classified data. However, some of the studies classify the data as a part of its workflow. For example, Security quad and cube (Chang & Jeong, 2011) makes a classification of patterns. When the number of hosts involved in the network traffic exceeds a limit, prioritization and selection of hosts is needed for effective visualization. Both Avisa (Shiravi, Shiravi, & Ghorbani, 2010) and Avisa2 (Shiravi, Shiravi, & Ghorbani, 2012) use heuristic host selection algorithms to make a prioritization among hosts. Based on this prioritization, the hosts which would actively be displayed on the view are selected. Lu et al. (2010) efficiently evaluates and classifies the host topology signatures to detect Sybil attacks.

Use of statistical value displays or information stacks is an important property which affects the usability of the overall design, because users are more used to interpret numerical results, rather than visual displays. Thus, including statistical numerical attributes as a part of visualization system, elevates its understandability. As mentioned earlier, Avisa2 (Shiravi, Shiravi, & Ghorbani, 2012) is better in terms of readability due

to the existence of information stack having graphics showing number of alerts for each alert type and their change over time.

Although Treemap display type is flexible in terms of the amount of data displayed and number of the hierarchies among them, it requires an information stack or detail window to explain what happens in some specific parts of the treemap as in Netvis (Kan, Hu, Wang, Wang, & Huang, 2010). In Dang and Dang's (2014), statistical outputs from multiple web site vulnerability scanners are at the heart of the visualization design. This main source of data for visualization is used in combination with hierarchical web site structure data by assigning pages as nodes and connections between them as links. Portall (Fink, Muessig, & North, 2005) is another design which includes the use of information stacks and popups for better understanding of network traffic between nodes of processes.

IDGraphs (Ren, Gao, Li, Chen, & Watson, 2005) includes a specific correlation analysis view based on the correlations of netflows within each other in a time frame. Positively and negatively correlated flows are shown in green and red colors. This type of display is useful to detect recurring or simultaneous patterns targeting multiple hosts or originating from multiple sources, in general, generated data from multiple points in a network.

NFlowviz (Fischer, Mansmann, Keim, Pietzko, & Waldvogel, 2008) provides an overall statistics of traffic data in addition to the treemap display of network traffic. One more interesting property of NFlowviz (Fischer, Mansmann, Keim, Pietzko, & Waldvogel, 2008) is, it allows making detailed analyses of hosts by enabling the use of popular query tools, such as, Whois, as a part of visualization solution. Thus, users do not have to leave the visualization tool to make further analysis of the hosts using these external query tools.

Correlation layers of information query and exploration is used in Clique (Best, Hafen, Olsen, & Pike, 2011). This is also a design based on a high usage of statistical calculations. This design finds out count of particular network traffic events in meaningful groupings, such as enterprise wise, department wise and protocol wise. Aggregation is made for 1 minute of intervals and the resulting values are defined as summary signals which are visualized instead of the raw network data.

2.5 Validation of Security Visualization Studies

Nearly every selected study includes a section for the presentation of design evaluation results. However, there is no systematic approach which may be taken as a standard for the validation of security visualization designs. This issue also makes it difficult to make a comparison of these designs. Every design selected for this chapter includes an implementation of the design either at prototype level or at product level. These prototypes and products are used to demonstrate several types of attacks for majority of

the designs. Naturally, each specific design is more powerful for demonstrating some type of scenarios or attacks and less powerful for some others. This issue results in the inclusion of around three to five different types of scenario demonstrations for each paper.

The data sources used for validation purposes for the selected studies are randomly generated data as in PolicyViz (Tran, Al-Shaer, & Boutaba), known data sets such as DARPA 99 (Laboratory, 1998-1999), as in Vafle (Ghoniem, Shurkhovetsky, Bahey, & Otjacques, 2014), and Svision (Onut & Ghorbani, 2007), data collected from laboratory conditions as in Vizalert (Livnat, Agutter, Moon, Erbacher, & Foresti, 2005), and data collected from the real world environments as in Security Quad and Cube (Chang & Jeong, 2011), and Whitaker and Erbacher (Whitaker & Erbacher, 2011). The laboratory generated data may include experimental attacks using various attack tools as in Security Quad and Cube (Chang & Jeong, 2011). Figure 19 illustrates a taxonomy of the validation data used for selected articles referenced in this text. As mentioned earlier, every design is more successful or more focused to visualize some group of scenarios. The success of the validation also depends on the demonstration data size, data quality, and data expressiveness. Unfortunately, this information is lacking for the majority of the designs.

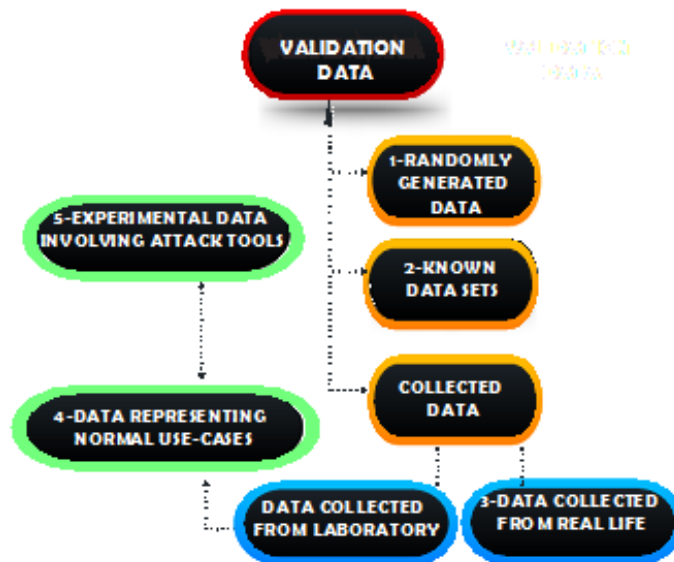


Figure 19 - Validation data sources

Although all of the designs aim to visualize some types of attack patterns, anomalies and misuses as part of their validation efforts, they have various approaches. The first group of validation approach is based on making experiments on the data for visualization purposes, commonly named as experimental validation. The second group of validation

approach follows a predefined scenario steps to achieve expected patterns commonly named as case study. The third group of validation approach is more systematic including a predefined set of attack tools and attack scenarios or vulnerabilities, and examines the results of combinations of them which may be named as a survey type validation. Although the use of known data sets enable visualization of attack patterns in a more extended manner due to the prior knowledge of attack data, the selected designs are either based on case studies or experiments.

Another issue related to the validation of the designs is the validation subjects. With few exceptions, most of the selected studies lack clear description of validation subjects, such as students, experts and their expertise level on information security issues.

2.6 Future Research Directions

The security visualization domain is still in its early stages. Thus, there are numerous new study subjects. In this part of the chapter, some of the trends gathered through the examination of the literature work, which do not directly aim to produce novel designs, but would improve the designs in some ways will be explained. The trending topics which are presented in this section forms a limited set of trends which may not include some of the studies and approaches since the security domain is evolving in many directions simultaneously and continuously.

Finding out novel designs is the continuing topic in the security visualization domain. There are also the design constraints mentioned in the previous sections. Primary topics of the trending studies are related to improving or solving those design constraints.

As the size of the data and number of dimensions increase, the readability of the diagrams becomes an important issue in visualization systems. Dimension reduction methods are one of the top trending topics in security visualization domain. For example, in Avisa (Shiravi, Shiravi, & Ghorbani, 2010) and Avisa2 (Shiravi, Shiravi, & Ghorbani, 2012) heuristic methods are used to reduce the number of hosts shown in the display.

Another way of dealing with large data with high dimensions is using incremental ways (Zhang, Liu, Nepal, & Chen, 2013), which is a trending topic itself. This incrementation can be used in earlier phases of the data visualization process, such as during data collection, preprocessing or in later phases, such as during classification and visualization phases.

Although, there is not much work specifically devoted to finding proper color schemes for security data visualization, this issue has been part of many design concerns so far, one sample work is from Mittelstädt et al. (Mittelstädt, Stoffel, & Keim, 2014).

Improvement of user interaction methods is a continuous trending topic. There are many researches on improved ways of user interaction. Including reasoning to user

interactions, such as semantic interaction is one of them. This requires sensing and capturing user interaction, inferring user models, adaptive interaction, and adaptive computation (Endert, North, Chang, & Zhou, 2014).

In general, there is more work to do for the validation, and evaluation of the security visualization designs. This includes the definition of evaluation metrics (Staheli, et al., 2014) and evaluation methods for the models. Evaluation of human computer interaction ways for security visualization designs is also an area which requires improvement in itself.

2.7 Concluding Remarks for the Literature Review

While researchers studying for the information security domain may have a better understanding of the attack tools, and platforms, ordinary system administrators and security analysts may not have the same level of knowledge about what types of tools and platforms the attackers use. Knowledge on attack types may result in expectations of predefined shapes in visualization displays, and may facilitate detection of patterns in the visualization designs. However, the designs should also be understandable without this knowledge for the users in the second group.

Security visualization designs should not depend on the assumption that the users of the systems already have such kind of knowledge. This requires the use of easily recognizable patterns, proper identification of all partitions of display shapes, and their purposes. Improving human computer interaction ways for all kinds of designs, reducing the complexity of the designs whenever applicable, decreasing the size of the data by proper filtering methods, and decreasing the number of data dimensions for the simplicity of the designs are also necessary for increased usability.

Some visualization properties are used repeatedly in the same way for more than one design. This repetition of some features is identical to each other in some designs, but there are some exceptions that use the same visualization attributes differently. Distance from a reference point is used to represent time past for an event repeatedly in designs. For example, distance from the radial center point is used in radial designs to represent different time periods, and distance from axis zero point is used as time passed for one axis of x-y-z charts. Color is most often used either to make a distinction between internal and external hosts, protocol types or to represent severity of the events. In some rare samples fading color or color intensity is used to represent the time past from a starting point in time. Another example of using a visual property in a different way is done by glyphs. Although, glyphs are most often used to point out the severity of events, they have been used to identify different groups of hosts and servers in some works.

There is much to say about the use of glyphs in security visualization designs. Use of glyphs puts into understandability of the designs. For example, the difficulty of showing the same data in a regular shaped matrix cells or cells filled with different shaped glyphs

indicating various types of attributes is nearly equal. However, the second one would have much more meaning to the user.

As the visualization system offers its own set of glyph designs, instead of using generic shapes like point, square, triangle, or star, it may describe more complicated situations, such as series of actions which take place in network traffic. This property looks like the main advantage of using design specific glyphs.

As the number of data dimension increase, the designs become more complex. Instead of visualizing all the dimensions in the same graph, visualizing graphs for different set of dimensions simultaneously is useful for dimension reduction. For some particular attack types, in order to understand the actual event which takes place within a time frame it is necessary to visualize a long period of time. Under such conditions, repeating the graph for subsequent small time periods would be useful and would result in better understandability specifically for 2-D, 3-D scatter plot type of visualizations.

One other issue related to the display types is the number of designs using advanced visualization techniques, such as animation or simulation. These types of displays are quite few. In the future, there should be more designs using these advanced interactivity and display features.

The existing visualization works commonly depend on data coming from a system and/or network analysis or monitoring tool. This results in a limited perspective of system monitoring results, because although there are quite large alternative data sources for security visualization systems, few of them are commonly used in existing designs. The number of works which collect its own data using command sets, such as operating system commands, network commands, and database commands is restricted. The reason for this seems to be that such designs require more effort for the data collection and consolidation phases. Specifically, if the design is related to multiple hosts and/or servers, then the data collection would require implementation of multiple sorts of agents which collect data from multiple points. An IDS system would automatically do this, but in order to use other commands these agents should be implemented as a part of the solution. These types of designs may also require the use of multiple sources of data simultaneously, which happens to be more complicated to interpret and visualize. Although difficult, such efforts may end up with novel designs with better usability levels.

Another issue related to data sources is the difficulties of using multiple sources at a time. These difficulties include different cardinalities, different time frames, normalized data versus unnormalized data, and different coding systems. Working to overcome these difficulties, especially in the case of using large data sets is a research topic in itself.

There is also the selection of visualized attributes arguments related to the data topic. The set of visualized attributes in the existing designs is not large. Although there are a few designs which use conceived metrics, the majority of the designs visualize common

networking attributes. In this domain, there is a requirement of well understood set of metrics. These metrics would help in seizing network events and trends and in diminishing the visualized data size and dimensions resulting in better visualization designs.

The validation parts of the existing work are not satisfactory. New validation methods are required, which give more idea on the level of the design constraints of the designs. Also, the existing evaluation system does not enable making any comparison between security visualization designs. There is a need for a framework which would help to define the design properties of the visualization system and enable evaluation and validation of the works. As a part of this framework a set of security visualization evaluation metrics should also be defined.

This review mainly focuses on display types, use cases and data sources of security visualization designs. There are some other categorization methods which are used in information visualization and/or security visualization domains and are not mentioned here. The way of handling anomalies is one type of categorization attribute used in the security visualization domain, such as being signature based or being anomaly based. Some particular data and use cases are more appropriate for signature based designs than others. Since a detailed analysis has been made on data types and use cases, no additional classification is done based on being signature or anomaly based.

Another classification method used in the information visualization domain is the level of interactivity of the designs. Although interactive properties of the designs are studied and noted throughout the study, only notable interactive features and a classification of them are included in this chapter. Assignment of an interactivity level for each selected design would require more experimental ways and platforms and is beyond the scope of this review study.

Considering the information visualization categorization attributes, as mentioned earlier the level of complexity is not determined throughout the study, because although the data sources are defined, the exact data attributes are not explicitly referenced in the majority of the selected designs.

There are some visualization categorization attributes, which are not specific to the security visualization domain. In the information visualization domain, one evaluation criteria makes a distinction between an infographics study and a visualization study. In terms of this evaluation criteria, all the works included in this review study are visualization studies. Another information visualization categorization attribute is the purpose of the visualization designs, which may be exploratory, explanatory or hybrid. All the works included in this review study are either exploratory or hybrid studies.

No review results are found on efficiency and performance of the selected designs. It is difficult to make any comments related to the performance and efficiency of the designs due to two reasons. The first reason is due to this review is not based on an experimental

approach. Therefore, actual trial and testing of this large set of designs is not possible. The main resource is the textual explanations and graphical definitions of design and display structures provided by the articles. Although the majority of the works are supported with case studies, the articles lack enough performance and efficiency related information. The second reason is the lack of standardization of the data. For example number of data points, length of time definitions do not exist for the diverse set of designs which are necessary to make a comparison of the performances, and efficiencies of the models. Yet, the usability of the works is discussed to some extent by thoroughly investigating the graphical images and corresponding data and design sections.

CHAPTER 3

SECURITY VISUALIZATION REQUIREMENTS SURVEY

3.1 Introduction to the Survey

Security visualization domain emerged at the beginning of the 21st century. Data has been the most authoritative element of the majority of the existing design decisions. Visualization designs might be due to seeking solutions to daily analytical problems. However, in order to make significant improvements, long-term researches are needed. While there are several security visualization designs, the number of use-cases and the case studies used in the academic studies are not as diverse as it should be.

Existing security visualization solutions (Özdemir Sönmez & Günel, Security Visualization Extended Review Issues, Classifications, Validation Methods, Trends, Extensions, 2018) are mostly focused on network security. Monitoring of intrusion detection systems, firewall logs, and configuration visualization are the most commonly implemented use-cases. Enterprise security visualization (Liao, Blaich, Striegel, & Thain, 2008) has been the subject of a small number of works so far. Host-server topology and host-server interaction visualizations form the most popular enterprise-focused security visualization subjects. To the authors' best knowledge there is no published earlier effort to gather user-centric requirements for enterprise security visualization solutions. Hence, in order to provide user-centric designs for the enterprise security visualization solutions, a security visualization requirements survey was carried out.

The survey's aim was to understand the existing situation regarding the use of security visualization solutions in the enterprises and to find out the requirements for new designs. It also aimed to find the answers related to the visual representation of different use cases in the security visualization domain. Thus, the survey consisted of questions related to the existing security analysis methods which encapsulate security visualization tools and techniques, the data sources which are collected and/or, stored and/or, analysed as part of the security analyses methods, the infrastructure elements of the enterprise including software, hardware and system components, the security analyses methods which may be extended by including security visualization methods and the user practices and expertise.

The survey contains both closed and open-ended questions. The participants are people with enterprise security expertise, from the academia and the industry. The qualitative and quantitative results coming from these users' responses are the subject of this chapter.

The rest of this chapter is structured as follows. Section 3.2 and Section 3.3 presents the need for the security visualization requirement analysis and the methodology, respectively. Section 3.4 provides the analysis and the results and Section 3.5 has the concluding remarks for this chapter.

3.2 The Need for Security Visualization Requirements Analysis

There have been numerous security visualization studies so far. Visualization designs are mainly affected by the data format, data type, size, and the use-cases. Generally, they are based on commonly known vulnerabilities and the threats. Available technologies also play an essential role in the design decisions. Although the number of existing studies is quite high, the number of user-centric designs is low. Limited coverage of user requirements is due to the restricted scope of client needs and planning perception. This issue is explained well in Frincke et al. (2009). In general, the researchers of the domain use conferences (Vis Sec, 2018) and domain-specific forum websites (Sec Viz, 2018) to share thoughts and information related to existing work, new design features, and future requirements. While these information sharing mechanisms contribute to the improvement of the domain, more effort is required.

Novel security visualization designs are scarce, as it requires composing a new way of data representation which is useful for the security domain. It requires knowledge of both security and visualization systems. If the target is to provide an enterprise security solution, the knowledge of enterprise security is also required.

Providing a successful design requires being more user-centric. There are studies which includes gathering user feedbacks in this domain. In these studies, users are incorporated as part of user experiments, and case studies for evaluation, and validation purposes. Although including users in these later steps is valuable for getting feedback to be used in subsequent studies, it is too late for users to influence the system requirements and design. Therefore, the authors decided to incorporate potential users in the requirements development phase.

Fry (2007) described the creation of the visualization process to be in seven steps including acquirement, parsing, filtering, mining, representation, refining, and interaction. The authors think that it will be more reasonable to give such an intense effort to design visualizations which correspond to real user security visualization requirements.

Lacking enough examination of security visualization requirements and not injecting this information into the security visualization studies results in:

- Rework for similar vulnerabilities or threats, which could have been examined together using the same data sources or same technologies, which further requires more effort to be spent on data collection and preparation, technology installation, education, and dissemination;
- Redesign of tools or multiple designs doing similar tasks, which could have been used to cover different situations, which causes late response to newly detected vulnerabilities and exposures besides wasting time and money;
- Design of tools which exhibit limited information or have only a few benefits, which further leads to the necessity of using multiple tools for visualization of security data for sufficient coverage.

3.3 Methodology of the Survey Study

Qualitative methods are commonly used for empirical studies of software engineering. Questionnaires including both qualitative and quantitative elements may be used to discover trends, generalizations, and new focus points. Collecting user requirements through qualitative and quantitative questionnaires might result in new and well-grounded security visualization hypotheses.

Security visualization requirements of the enterprises can be determined by

- asking questions related to the existing software, system and hardware infrastructure of the enterprises,
- reviewing commonly used security analysis techniques,
- determining the current level of security visualization usage in the enterprises,
- finding out the most popular security use cases for different types of enterprises,
- investigating the data sets which are collected and stored by enterprises, which would be taken as security visualization data sources,
- investigating the critical data attributes for the security analyzers,
- comparing various display types in terms of usability, and
- determining the staff awareness level on the infrastructure security data sources and their analysis techniques.

3.3.1 Survey

A detailed survey was prepared which consisted of questions related to the existing security analysis methods which encapsulate security visualization tools and techniques; data sources which are collected and/or, stored and/or, analysed as part of security analyses methods; the infrastructure of the enterprise including software, hardware and system components; security analyses methods which may be extended by including security visualization methods; and the user practices and expertise.

The survey contained 25 multiple-choice, seven grading scales and 14 open-ended questions. Participants were asked to complete the survey online.

Sections of the enterprise security visualization requirements survey are listed below.

- A. Participant Information Section
- B. Pre-survey Evaluation Quiz Section
- C. Security Visualization Use Cases
- D. Security Visualization Data
- E. Security Visualization Data Size
- F. Security Analysis Techniques
- G. Visualization Design and Display Properties
- H. Technical Infrastructure
- I. Organization and Domain Information
- J. User Information

The question set and the raw data of the requirement analysis survey study are published on GitHub under the name “Security Visualization Requirement Analysis Raw Results” for the interested audience who may want to refer to the components of the requirement analysis work and have more information related to the attendees' expertise levels and background. In this chapter, only the results of this study are explained in detail.

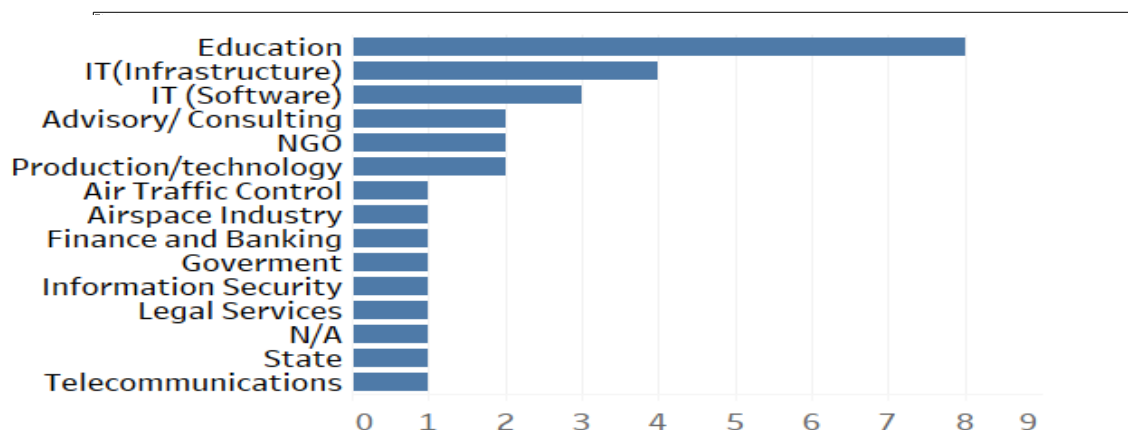


Figure 20 - Primary sectors of the attendees

3.3.2

The number of participants for the survey was 30. All had expertise in the security domain. Their primary sectors are shown in Figure 20. The security-related certificates that the attendees held were 6 CISSP certificates, 1 ISO27001:2013 lead auditor certificate, 2 CEH certificates, 1 ISO27005 Risk Manager certificate, 2 Security+ certificates, 2 CISM certificates, 1 TUBITAK SOME certificate, 1 Cisco Security certificate, 1 Cybersecurity certificate, 1 CCNA SECURITY certificate, and 1 PARTIAL CISA certificate.

3.4 Analysis and Results

The results extracted from the survey are grouped into three categories: quantitative results at a glance, further quantitative results and the qualitative results. In this section, together with the results, the facts and the topics that need to be examined in more detail which were determined by analyzing these results are also presented in the form of explanatory notes.

3.4.1 *Quantitative Results at a Glance*

When the existing studies are further examined, it is seen that the majority of the existing security visualization designs depend on a single type of data source, such as the network traffic data. Some of the visualization designs filter data sources according to the protocol types. TCP protocol data is the most commonly visualized data.

One of the main objectives of the requirement analysis survey was to determine what kind of security-related data is collected in the organizations, which of them are stored for future examination, and which of them are examined as part of security analysis methods. As a result of the questionnaire, 12 data sources were identified. In order to quantify and plot the importance of the data sources, the answers which state “*not collected at all*” were assigned the score of zero, the answers which state “*collected but not analysed*” were assigned the score one, and the answers which indicate “*analysed as part of security analyses*” were assigned the score two. The mean scores were then calculated for each data source. The resulting importance values for the data sources are shown in order in Figure 21. As expected, the network traffic data has the most noteworthy significance as a security perception information source. Router configuration log, on the other hand, has the least significance. For all the other questions, five-level Likert items were used with scales from one to five.

Considering that the security of shared resources is more critical than the security of non-shared ones, policies of sharing data, services, and infrastructure have been examined in the requirements analysis. It was found that enterprises routinely share such resources with customers (17 participants), suppliers (13 participants), partners (20 participants), and other stakeholders (17 participants).

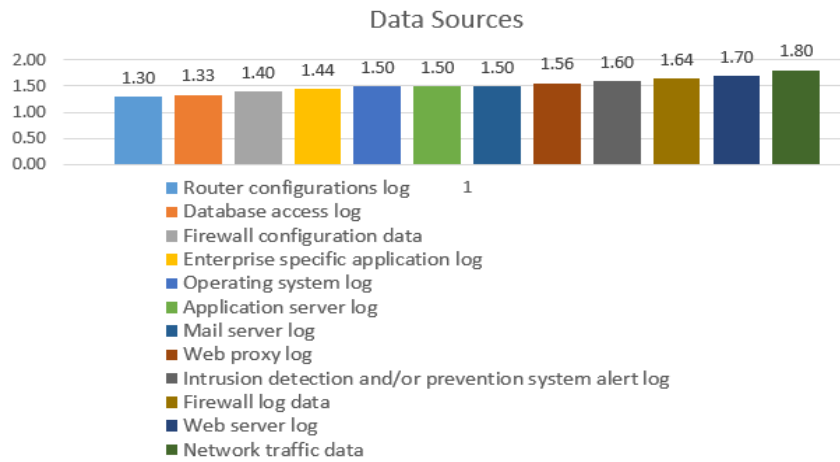


Figure 21 - Importance of data sources for the organizations

Another finding of the security visualization requirement survey was the list of popular security visualization use cases, which are most applicable and beneficial to the organizations. Figure 22 shows the summary information related to the adaptation of security visualization use cases in the organizations. Series 1 corresponds to the sum of answers either which the participants have no idea of the use case or think that it can not be applicable to their organization. Series 2 corresponds to the sum of the answers where it is stated that this use case has not been adopted yet, but would be moderately beneficial or very beneficial for their organization and that this use case has already been adopted in the organizations. It can be observed that the familiarity with and usefulness of the use cases do not vary much among 14 use-cases. However, enterprise users seem to be more familiar to enterprise data and enterprise asset related use-cases but less familiar to use cases related to core Internet protocols such as BGP and DNS.

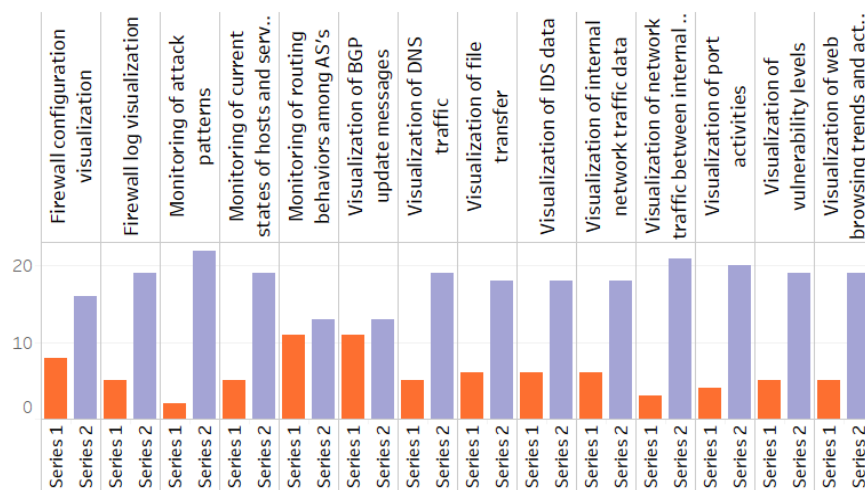


Figure 22 - Security visualization use-cases

		> 250	50 - 250	10-50	< 10
Firewall configuration visualization	li.	48.00	11.00	11.00	6.00
Firewall log visualization		56.00	12.00	11.00	7.00
Monitoring of attack patterns		55.00	15.00	13.00	8.00
Monitoring of routing behaviors among		43.00	11.00	6.00	6.00
Monitoring of the current state of hosts		54.00	16.00	7.00	7.00
Number of Records		19.00	5.00	4.00	2.00
Visualization of BGP update messages		42.00	7.00	6.00	6.00
Visualization of DNS traffic and lookup		52.00	16.00	10.00	7.00
Visualization of IDS data		55.00	13.00	12.00	6.00
Visualization of file transfers		46.00	13.00	6.00	7.00
Visualization of internal network traffic		50.00	13.00	11.00	6.00
Visualization of network traffic between internal hosts and..		53.00	16.00	12.00	7.00
Visualization of port activities		55.00	12.00	13.00	6.00
Visualization of vulnerability levels		49.00	15.00	10.00	6.00
Visualization of web browsing trends and activities		56.00	12.00	8.00	6.00

Figure 23 - Evaluation of security visualization use-cases according to the enterprise size (number of employees)

The evaluation of use cases according to the number of employees, which gives an indication of the enterprise size, is also presented in Figure 23. It can be observed that the familiarity with and usefulness of all of the use-cases increase as the number of employees (size of the enterprise) increases. The evaluation of use cases according to the primary sector of the enterprise has also been made and it was observed that the familiarity with and usefulness of the use-cases vary based on the primary sector of the participant. The education sector has the highest results, possibly due to increased awareness as a result of the graduate education.

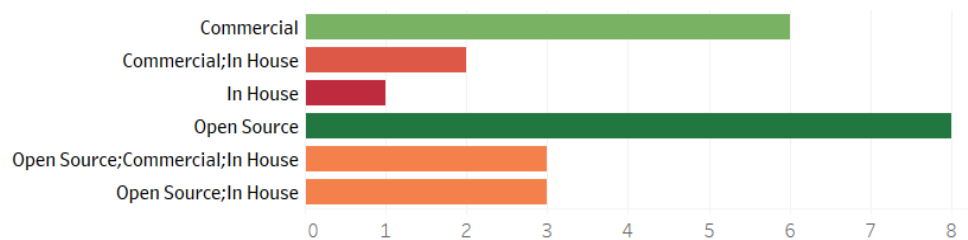


Figure 24 - Origin of existing security visualization solutions in the enterprises

The distribution of security visualization solutions used in the enterprises based on their origin as commercial, in-house, or opensource is shown in Figure 24. It can be seen that open source security visualization systems are more preferable among the attendees.

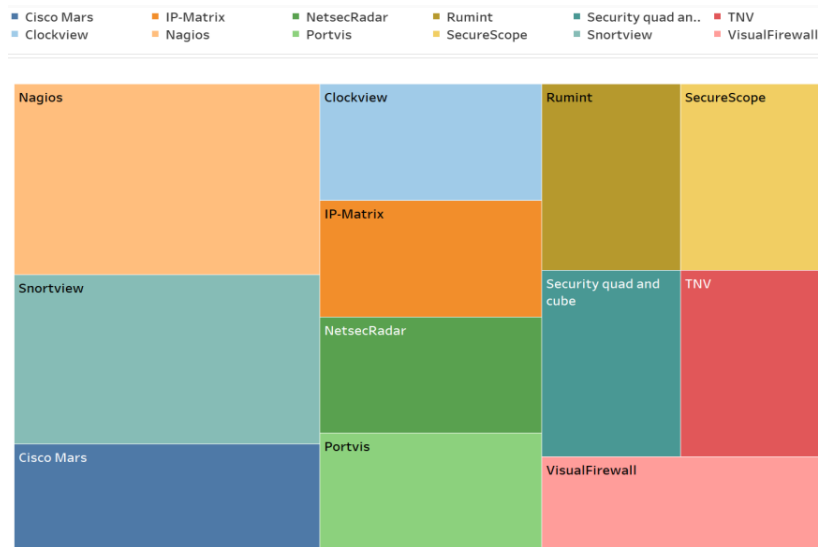


Figure 25 - Most popular security visualization solutions in the enterprises

The most popular security visualization solutions in the enterprises are shown in Figure 6. The most popular security visualization solutions are Nagios (Josephsen, 2007), SnortView (Koike & Ohno, SnortView: visualization system of snort logs., 2004), and CiscoMars (Halleen & Kellogg, 2007).

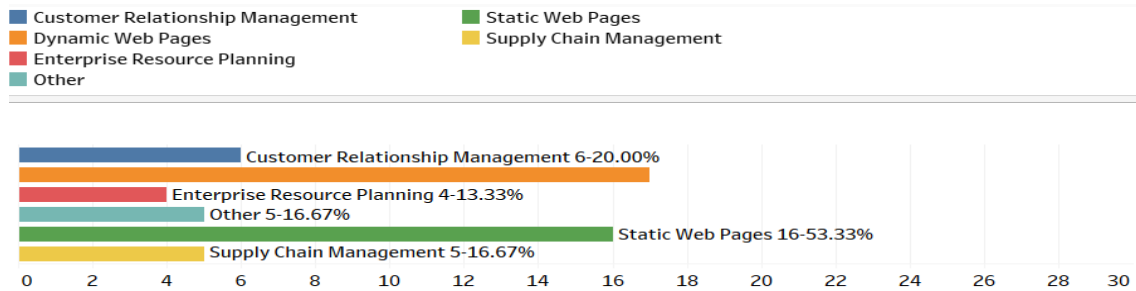


Figure 26 - Commonly used enterprise software solutions

During the requirement analysis survey, mostly used enterprise software systems, infrastructure components, and security systems were also questioned aiming to discover new security visualization areas for the enterprises. Figure 26 shows the usage of “Static Web Pages”, “Dynamic Web Application”, “Enterprise Resource Planning (ERP)”, “SCM”, “CRM” and “Other” systems in the organizations. It can be observed that most used software systems are static and dynamic web applications.

The use of different enterprise IT system components can also be considered as the subject of a security visualization study. The use of “File Sharing Server”, “Web Server”, “Mail Server (Internal)”, “Mail Server (External)”, “Application Server”, “Database Server”, “Cloud Storage”, “Other Cloud Services”, “External Router”, “Internal Switch or Router”, “Wireless Network”, Printer”, “E-Fax”, and “Other” systems along with security protection systems has been questioned during the security requirements analysis survey. The most popular systems are listed in Figure 27. It can be seen that network firewalls, printers, external mail servers, and web servers are the most commonly existing components in enterprise infrastructures.

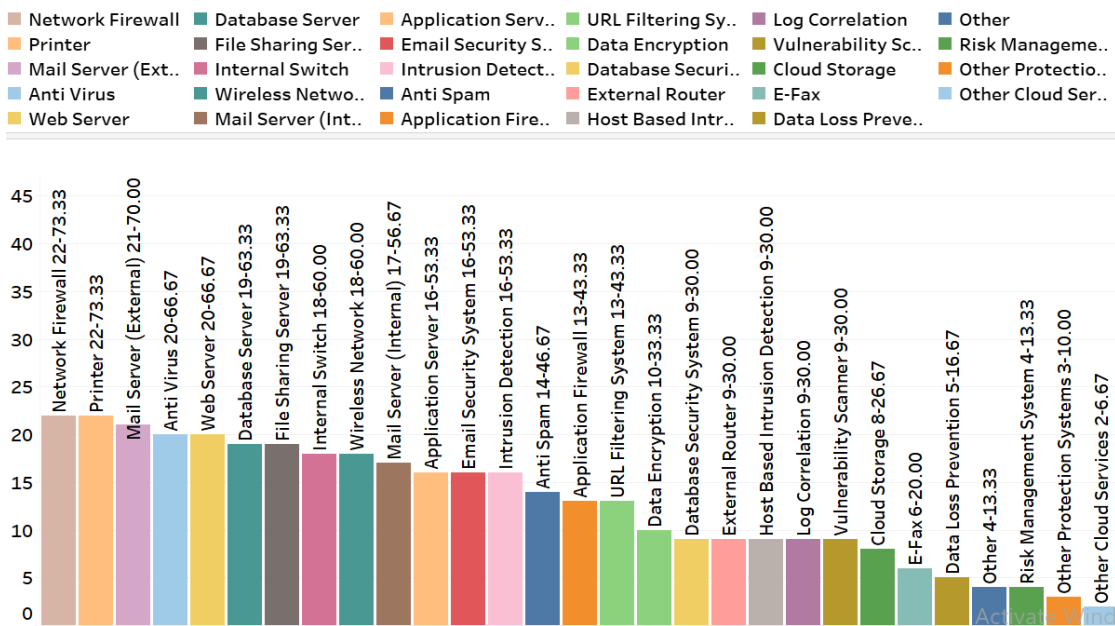


Figure 27 - Hardware, networking and system components that are part of the infrastructures

In the survey, in order to find new ideas to improve the existing threat analyses methods, the participants were asked to define analyses, mapping threats to security data sources and data attributes. As a result, 19128 tuples (threat, data source, data attribute) were identified. A portion of these association results is shown in Figure 28.

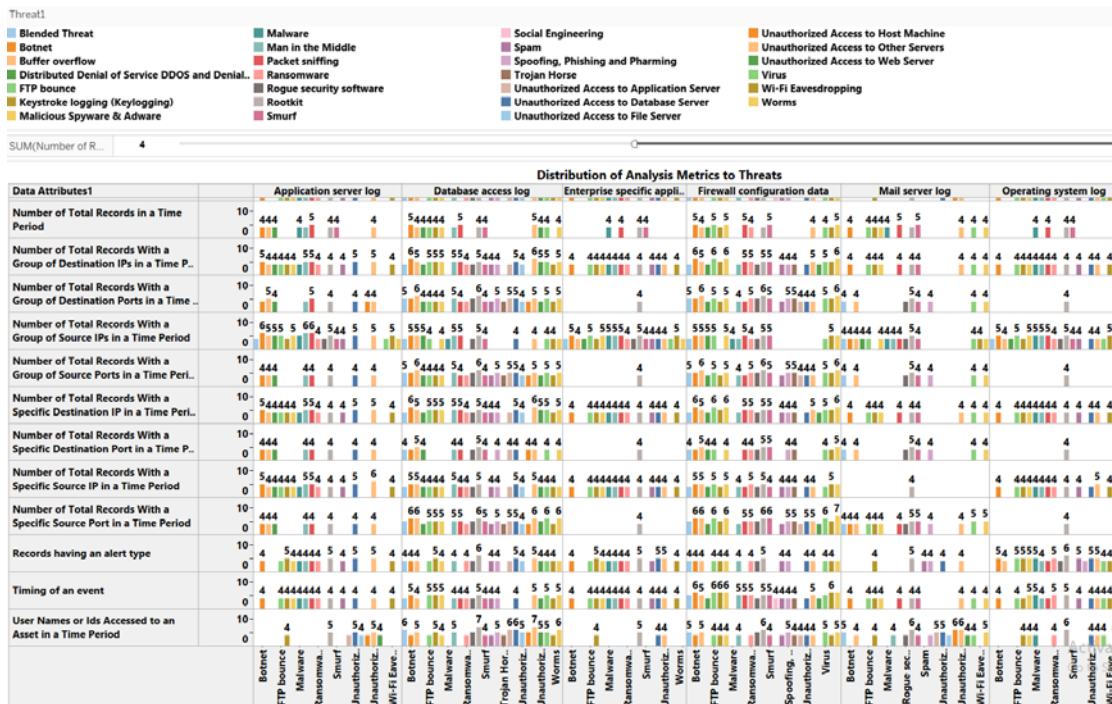


Figure 28 - Associations of threats to data sources and data attributes

Rootkit, botnet, and unauthorized access to other servers were the threats that were mostly associated to data sources and data attributes. Social engineering, unauthorized access to host machine and trojan horse threats were the least associated. The “Number of a specific type of error”, and the “Number of Total Records with a Group of Source IPs in a Time Period” were the data attributes which were mostly associated to the threats. The “Number of Total Records in a Time Period” was the least associated data attribute.

In order to contribute to the development of new designs, the users were also asked about the importance of design issues such as scalability, interactivity, searchability, and being zoomable, and the usability of display types, such as simple charts line charts, bar charts or complex charts with animation. The results obtained from these questions are shown in Figure 29 and Figure 30, respectively. The results do not allow making a sharp distinction between the importance of design properties. However, simple display types, such as line charts and bar charts are found more understandable by the users than the complex ones.

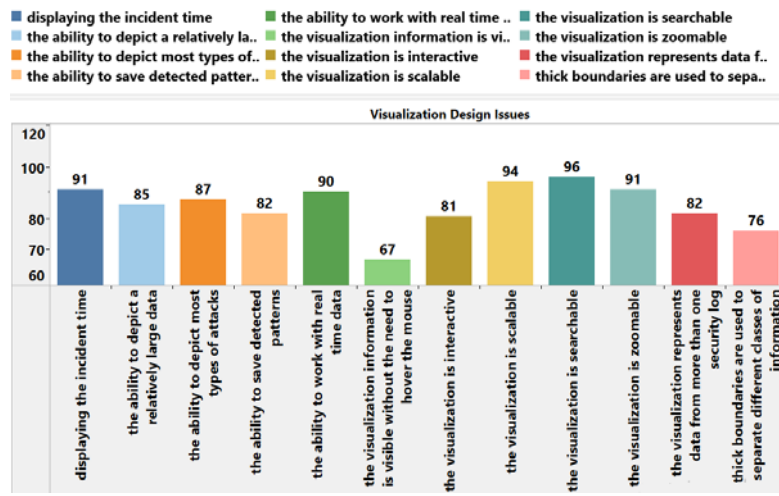


Figure 29 - Security visualization design issues

Popularity of Display Types

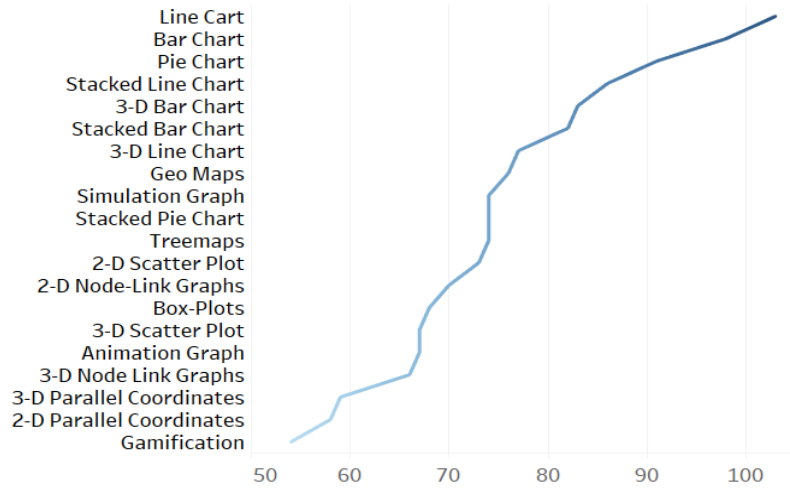


Figure 30 - Popular display types

Finally, the users were asked about their current security practices. Figure 31 shows the usage of correlation, escalation, forensic, incident response, threat, and triage type of analyses. While results do not allow making a sharp distinction between various security analyses types, the escalation analysis seems to be the least favourite one.

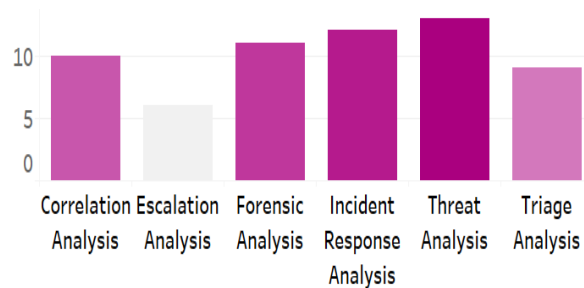


Figure 31 - Popular security analyses

Table 7 - Apriori Rule Generation for Enterprise Software Systems

=== Apriori Run information ===

Minimum support: 0.1 (3 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 9

Size of set of large itemsets L(3): 7

Size of set of large itemsets L(4): 2

Best rules found:

- 1.ERP, Static Web Pages, conf:(1) lift:(1.88) lev:(0.06) [1] conv:(1.87)
- 2.ERP, Dynamic Web Application, conf:(1) lift:(1.76) lev:(0.06) [1] conv:(1.73)
- 3.Dynamic Web Application, ERP, Static Web Pages conf:(1) lift:(1.88) lev:(0.06) [1] conv:(1.87)
- 4.Static Web Pages, ERP, Dynamic Web Application conf:(1) lift:(1.76) lev:(0.06) [1] conv:(1.73)
- 5.ERP, Static Web Pages, Dynamic Web Application, conf:(1) lift:(2.5) lev:(0.08) [2] conv:(2.4)
6. Dynamic Web Application, SCM, Static Web Pages, conf:(1) lift:(1.88) lev:(0.06) [1] conv:(1.87)
- 7.Static Web Pages, SCM, Dynamic Web Application, conf:(1) lift:(1.76) lev:(0.06) [1] conv:(1.73)
- 8.ERP, CRM, Static Web Pages, conf:(1) lift:(1.88) lev:(0.05) [1] conv:(1.4)
9. ERP, SCM, Static Web Pages, conf:(1) lift:(1.88) lev:(0.05) [1] conv:(1.4)
- .10. ERP, CRM, Dynamic Web Application conf:(1) lift:(1.76) lev:(0.04) [1] conv:(1.3)

Table 8 - Apriori Rule Generation for Enterprise Security Systems

Minimum support: 0.4 (12 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 11

Size of set of large itemsets L(3): 5

Best rules found:

1. Intrusion Detection and/or Prevention System, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.14) [4] conv:(4.27)
 2. Email Security System, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.14) [4] conv:(4.27)
 3. Email Security System, Anti Virus, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.12) [3] conv:(3.73)
 4. URL Filtering System, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.12) [3] conv:(3.47)
 5. Anti Spam, Anti Virus, conf:(1) lift:(1.58) lev:(0.16) [4] conv:(4.77)
 6. Intrusion Detection and/or Prevention System, Anti Virus, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.12) [3] conv:(3.47)
 7. Intrusion Detection and/or Prevention System, Email Security System, Network Level Firewalls, conf:(1) lift:(1.36) lev:(0.11) [3] conv:(3.2)
 8. URL Filtering System, Anti Virus, Network Level Firewalls conf:(1) lift:(1.36) lev:(0.11) [3] conv:(3.2)
 9. Network Level Firewalls, Anti Spam, Anti Virus, conf:(1) lift:(1.58) lev:(0.15) [4] conv:(4.4)
 10. Anti Virus, Network Level Firewalls, conf:(0.95) lift:(1.29) lev:(0.14) [4] conv:(2.53)
-

3.4.2 Further Quantitative Results

One of the most widely used instrument to mine association rules is Apriori (Agarwal & Srikant, 1994). As explained in the previous section, the participants were asked to detail their software systems, security systems, and other infrastructure elements. In order to find sets of software systems that are commonly used in the organizations of the participants, Weka Apriori algorithm was used (Hall, et al., 2009). The runtime information, and the results for software systems, security systems and other infrastructure elements are shown in Table 7, Table 8, and Table 9 respectively.

Table 9 - Apriori rule generation for other enterprise infrastructure elements

Minimum support: 0.1 (18 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 8

Size of set of large itemsets L(3): 1

Best rules found:

1. Database Server, Printer, conf:(1) lift:(1.05) lev:(0.04) [0] conv:(0.83)
 2. Internal Switch or Router, Printer, conf:(1) lift:(1.05) lev:(0.03) [0] conv:(0.78)
 3. Wireless Network, Printer, conf:(1) lift:(1.05) lev:(0.03) [0] conv:(0.78)
 4. Mail Server (External), Printer, conf:(0.95) lift:(1) lev:(-0) [0] conv:(0.46)
 5. Web Server, Mail Server (External), conf:(0.95) lift:(1.04) lev:(0.03) [0] conv:(0.87)
 6. Web Server, Printer, conf:(0.95) lift:(0.99) lev:(-0.01) [0] conv:(0.43)
 7. File Sharing Server, Web Server, conf:(0.95) lift:(1.09) lev:(0.06) [1] conv:(1.24)
 8. File Sharing Server, Printer, conf:(0.95) lift:(0.99) lev:(-0.01) [0] conv:(0.41)
 9. Web Server, Printer, Mail Server (External), conf:(0.95) lift:(1.04) lev:(0.03) [0] conv:(0.83)
 10. Web Server, Mail Server (External), Printer, conf:(0.95) lift:(0.99) lev:(-0.01) [0] conv:(0.41)
-

The sets formed by association mining might be useful while making various technical decisions. For example, multiple licensing options embracing sets of various infrastructure items/software systems/security systems can be offered by vendors or security visualization solutions. The corresponding data formats can be taken into consideration while designing visualization tools which would handle multiple data sources. Table 7, Table 8, and Table 9 are included to give an idea of how to choose the set of infrastructure elements to be visualized when building a holistic enterprise security visualization structure.

Table 10 includes the k-means clustering results for the association of threats, data sources, and data attributes. This data has three features; data source, data attribute, and threat name. K-means clustering is simply used to identify groups of threat, data source, data attribute triples without any feature extraction. The aim is to show a way to provide new associations of threats to the data attributes and data sources in order to handle the threats which are not commonly handled in the existing ones, with new designs. The results of k-means clustering show that the clusters numbered 0, 1, 2, 3, 4, 5, and 6 have good satisfaction levels and might be investigated further prior to making design decisions.

Table 10 - K-means clustering results of threat-data source, data attribute associations

Number of iterations: 3
 Within cluster sum of squared errors: 34339.0
 Initial starting points (random):
 Final cluster centroids:
 Cluster 0: 'Malicious Spyware & Adware','Operating system log','Number of a specific type of error'
 Cluster 1: 'Keystroke logging (Keylogging)','Web proxy log','Number of Total Records With a Specific Destination IP in a Time Period'
 Cluster 2: Botnet,'Operating system log','Records having an alert classification'
 Cluster 3: 'Unauthorized Access to Application Server','Mail server log','Timing of an event'
 Cluster 4: Ransomware,'Web server log','Number of Total Records With a Specific Source Port in a Time Period'
 Cluster 5: 'Unauthorized Access to File Server','Firewall log data','Number of Total Records With a Specific Destination Port in a Time Period'
 Cluster 6: 'Blended Threat','Firewall configuration data','Number of total errors'
 Cluster 7: 'Blended Threat','Firewall log data','Timing of an event'
 Cluster 8: 'Wi-Fi Eavesdropping','Router configurations log','Number of total errors'
 Cluster 9: 'FTP bounce','Intrusion detection and/or preventions system alert log','Number of Total Records With a Group of Destination Ports in a Time Period'
 Cluster 10: 'Unauthorized Access to File Server','Web server log','Number of Total Records With a Group of Destination Ports in a Time Period'
 Cluster 11: 'Unauthorized Access to Database Server','Enterprise specific application log','Number of a specific type of error'
 Cluster 12: 'Distributed Denial of Service DDOS and Denial of Service DOS','Firewall log data','Number of Total Records With a Group of Source IPs in a Time Period'

Cluster 13: 'Wi-Fi Eavesdropping','Network traffic data','Number of Total Records With a Group of Destination Ports in a Time Period'

Cluster 14: 'Unauthorized Access to Other Servers','Firewall log data','User Names or Ids Accessed to an Asset in a Time Period'

Cluster 15: 'Unauthorized Access to Database Server','Enterprise specific application log','Number of Total Records With a Group of Destination Ports in a Time Period'

Cluster 16: 'Unauthorized Access to Host Machine','Intrusion detection and/or preventions system alert log','Number of Total Records With a Group of Destination IPs in a Time Period'

Cluster 17: 'Buffer overflow','Application server log','Records having an alert classification'

Cluster 18: Virus,'Web server log','Timing of an event'

Cluster 19: Spam,'Web server log','Number of Total Records With a Specific Source Port in a Time Period'

Missing values globally replaced with mean/mode

Time taken to build model (full training data) : 0.14 seconds

Cluster	# of Tuples	%	Cluster	# of Tuples	%
0	3060	16%	10	174	1%
1	2894	15%	11	638	3%
2	2457	13%	12	417	2%
3	1659	9%	13	673	4%
4	1699	9%	14	276	1%
5	1222	6%	15	100	1%
6	1239	6%	16	247	1%
7	186	1%	17	352	2%
8	722	4%	18	192	1%
9	825	4%	19	96	1%

3.4.3 Qualitative Results

In the survey, the users were asked questions about their information levels on security-related log files. A few users were not very familiar with their log file types. In general, the participants were not very informed about their log file sizes. Only one user managed to enter numerical values for the average size of daily records generated in firewall log file, IDS alert file, application server access log file, application server error log file, web server access log file, web server error log file and mail server log file. Therefore, it can be said that the participants are not very knowledgeable about security log files.

There were some free format questions to collect strategies for different situations and new use cases which are applicable for the organizations. Table 11 summarizes these strategies and suggestions.

As a result, it can be said that in general, the participants propose solutions which are not directly related to the use-case asked, but general purpose solutions. The majority of the logical solutions that were offered by the participants are not novel. The strategies and proposed relevant metrics are better to be saved in a knowledge base structure.

Table 11 - Strategies and Suggestions

Strategies to reduce the size of logs
<ul style="list-style-type: none">• Archive in cloud and delete logs periodically• Check some features from other systems to filter important features• Use logs for specific traffic only• Filter useless entries and use compression• Use moar logs• Use security analytics
Strategies/methods to differentiate normal behavior of web browsing from abnormal behavior
<ul style="list-style-type: none">• Protecting the system under a firewall• Mod security implementation• Using next-generation firewalls• Exploring user agent strings passed by web browsers which may indicate known bad behavior, valid but forbidden by policy behavior or a covert channel• Investigating the malware command and controls via purported web browsing• Using baselining• Detection of anomalies by analyzing proxy logs, using darktrace etc.• Use of commercial and other whitelists• Checking for sudden changes• Visualization of firewall traffic log• Use of IPS features of the firewall• Monitoring the amount of abnormal web requests
Strategies/methods to differentiate normal activities of file sharing from suspicious activities
<ul style="list-style-type: none">• Using next-generation firewalls• Combining file sharing data with human resources data (ex. data of a person who is likely to be fired)• In-house tools• Sudden changes in volume/#connections• Block shadow IP's in the firewall• Check correlation of DLP logs• Use of Wireshark• Check times of download/upload processes
Strategies to differentiate normal behavior of social media usage from suspicious behavior using data
<ul style="list-style-type: none">• Controlling social media tools with the bare eye• Using social media sentiment analysis tools may be helpful.• N/A (Social Media is forbidden)• Block in L7 firewall
Any suggestions for security visualization usage scenarios which is beneficial for the organizations
<ul style="list-style-type: none">• "Log analysis and correlation applications would be good."• "I see visualization as a useful interface between the human and the machine. To me, the most interesting scenarios are when visualizations enable humans to find important things that machines can't, but then can enable the human to properly parameterize the insight so that the machine can do the heavy lifting in the future. This frees up the human to take on a new challenge the machine cannot yet handle. Then the cycle continues. This is the only approach that is scalable."• "Authentication success and failures."• "Do not restrict yourself to 2D visualization."

-
- “MS Baseline Analyzer for network analyses.” “Visualization of individual client's network traffic such as visualization of the clients DNS requests, file downloads via e-mail or web browsing, usage of unexpected ports could be correlated and visualized. In the visualization programs common information (IP addresses etc.) in different types of network traffic could be mapped in order to help drawing conclusions.t-SNE and Multidimensional scaling. Data visualizations such as in Kibana can be useful. With Kibana one can also do fraud analysis. Device information like OS, layer 3 protocol details and Tracert info belonging to attackers can be visualized. Use of Maltrail.”
 - “Use of Spice Works tool for IT helpdesk and system performance monitoring purposes.”
-

3.5 Concluding Remarks for the Survey

In an effort to determine user-based enterprise security visualization requirements, a survey was set up. Although the number of attendees was not very high, the experience and information level of the participants was at the desired level. This shows that the targeted audience was being reached during the survey.

From the survey, several results were obtained pointing out various observations related to the security visualization domain. Some of these were expected results. For example, web applications was the mostly used software applications; network traffic data was selected as the most important data source for security analysis; and users were more prone to select simple display types, such as bar charts and line charts compared to complex display types. There were also some unexpected results. For example, interactivity was claimed to be less important compared to some other design properties. There were some results which point out new visualization subjects. For example, more visualization studies were required focusing on printer usages and mail servers.

Further quantitative analysis results provide information which requires tdeep examinationd to improve existing security visualization designs and to form a novel design. For example, sets of infrastructure items which may be examined in groups in security visualization solutions, the clusters of threats and associated data sources and data attributes may point to new metrics for particular threats.

Majority of the results helped in distinguishing items among alternatives, or helped to understand new issues. A few of the results did not allow sharp distinctions among alternatives. During the scaling, multiplier sets (one to five) were used as mentioned before. Using a different multiplier set would end up with having more clear boundaries.

The reason for doing this kind of a survey was due to thinking that this type of survey might result in user-centric solutions with better designs. In this way, the designers can find out novel ideas which may contribute to creating holistic approaches for the enterprise security. These results should be reflected to the security visualization domain by novel designs which are not restricted to known data sources and known use-cases.

This survey may also be suitable for carrying out internally in the organizations. It may also be adapted for non-technical people. This effort may lead to other interesting results, such as the identification of new security sources, and new visualization use-cases.

One major limitation of this study was the limited number of attendees. They were all informed about the content of the survey prior to their participation, especially on the type of the questions, and the length of the survey. Some participants hesitated to contribute due to the length of the survey and some others hesitated due to the specific subject of the survey. A shorter survey involving similar concepts can be prepared as a future work, and new ways of survey distribution can be considered in order to get the maximum benefit.

The results have shown that the users are not familiar with the majority of security visualization solutions or have problems in using existing ones. More effort should be given to designing user-focused security visualization designs.

The results were recorded to be converted to functional and non-functional requirements during the subsequent thesis studies. The requirements should also be elicited accordingly and should be combined with the latest technological instruments to form an enterprise security visualization system design.

CHAPTER 4

DESIGN AND DEVELOPMENT OF A GENERIC SECURITY VISUALIZATION INFRASTRUCTURE PROTOTYPE

4.1 Introduction to Design and Development of a Generic Security Visualization Infrastructure Prototype Study

Since the 1940s the variety, and use-cases of the information and communication devices have grown highly. Initial computers have turned into PC's with network connections; PC's changed to laptops with wireless adapters; mobile phones, and other GPS enabled devices have taken the place of laptops from time to time; PDAs, and smartphones have dominated everything both in private and business life; and lastly the functionalities offered by wearable devices have increased day by day. The computational efforts changed over time from mainframe computing, desktop computing, ubiquitous computing, and cyber-physical systems. Cyber-physical systems are engineered systems which combines computing, communication, and data storage capabilities with the aim of coordinating, monitoring, and controlling of environmental entities.

There are many projects done so far which focus on providing new computational ways for various cyber-physical environments. National Institute of Standards and Technology, NIST, had a project named Smart Space (Fillinger, et al. 2009). As a part of this project, they developed a Meeting Room Recognition application (Stanford, et al. 2003) which forms a prototype for future's meeting rooms and command control centers. Future Computing Environments Group in Georgia Institute of Technology developed Classroom 2000 project (Abowd, et al. 1996) as a prototype to monitor the effects of the impact of ubiquitous computing in education. Interactive Workspaces Project (Johanson, Fox and Winograd 2004) by Stanford University investigated new technologies to form a multi-person, multi-device, collaborative working place.

It is a known fact that the cost of not giving enough importance to information security may be very high. To prevent these losses, each organization should provide enough significance to information security management. The primary sources of security management in an enterprise are the log files, alerts produced by security systems and devices, and network traffic data. In this work, an enterprise network with its information security related infrastructure elements will be the focus.

Security visualization outlines may have diverse purposes, for example, condensing the information, reenacting past occurrences, permitting design disclosure, location of malignant exercises, inconsistencies, misconfigurations, and anomalies. Security visualization may give various perspectives of similar information all the while or it might picture diverse information in a similar view. The aim of this study is to create a design which targets to use the data produced by environmental entities and allows the creation of visualizations. Existing security visualization systems are mostly attached to the log file type rigidly. The proposed system is based on generic parsers for each file format type (JSON, CSV, and TXT). The proposed system also enables feedback from the users, such as systems analysts, operation center users, heads of departments, or senior managers.

There are many data sources and many visualization tools. In order to use a visualization tool for a specific data source a considerable amount of effort is required. A generic visualization system which may visualize data in multiple forms with little preparation, based on metadata and selection of visualization type on the fly would be very beneficial.

The term generic visualization system is used in this study to point out two basic design features of the proposed system. The first feature is the availability of a parser which enables parsing of data based on data definitions. The second feature is the design structure which enables the use of multiple display libraries due to having content adapter boilerplate classes which extend an interface enabling a standard structure for various display libraries. These two features provide a level of genericness on two ends of the data visualization process; data input and data display. In general, non-generic visualization systems mainly rely on a fixed data format. Non-generic visualization systems also depend on a fixed number of display options, without allowing extensions with new display types. Some improvements to non-generic visualization systems are due to allowing visualization of data formats which may cover multiple security data formats, such as XML. This brings a level of genericness on the data input side, leaving the display type selection alternatives unchanged. Some of the proposed features are distinctive for a security visualization solution. A security visualization knowledgebase is aimed to be formed through the use of user feedbacks in the proposed system. These feedbacks will form an enterprise security visualization knowledgebase and may also be used for automatic processing for various purposes. This knowledgebase is expected to accelerate the learning and to help the creation of more successful visualizations in the future.

During this study, a set of requirements is prepared to specify the proposed system. The main source of the defined requirements is the results of a requirements analysis survey. The reason for making such an inquiry was to enable a user-oriented specification set for this domain. The size of the enterprise, set of business processes, and infrastructure of the organization would directly affect the visualization tasks. The target is to provide a set of security visualization requirements for all type of organizations. It does not point out a specific organization. However, the restricted scope depends on the outputs of the

survey and common design features from literature search results. Thirty participants with various levels of solid information security knowledge and experience attended this survey. The raw results of the survey are not included due to space limitations. However, for each requirement or groups of requirements, the rationality of the requirement is described briefly. This rationality may include reference to the parts of security visualization survey results or the known issues from literature. These requirements are elicited based on applicability, consistency with other needs, and compatibility with the overall structure of the projected enterprise visualization solution.

Being a cyber-physical system, the human entities of the selected case may be technical such as system operators, and security analysts, and non-technical such as non-technical managers or report writers. The environmental entities for the proposed system are the enterprise infrastructure elements, such as hardware or software firewalls, honey-nets, intrusion detection systems, and operating systems. The security-related outputs of these environmental entities are aimed to be monitored by the use of visualizations prepared by human entities.

One of the main concerns related to developing a visualization system for an enterprise is the existence of a variety of data formats and data types which are originated from various devices. Standardization is required to use the outputs of these devices. There have been some attempts to standardize the log files which are Extended Log File Format by W3C (Hallam-Baker ve Behlendorf 1996), Common Event Format by ArcSight (Arcsight 2009), Syslog by IETF (Gerhards 2009), IDMEF by IETF (Debar, Curry ve Feinstein 2007), SDEE/CIDEE by Cisco (Cisco 2009). While some of these standardization attempts are deprecated, the others are in their early stages. The lack of standardization for these files is a serious problem for the security visualization efforts (Marty 2009) (Chuvakin, Schmidt ve Phillips 2012). Another important concern is the scalability of these systems. As the time passes, the number of data accumulated can grow very high and may require specific storage and computational requirements. Creating a visualization depends on many factors. There are various issues, issues related to data preparation, issues related to selecting the correct display type. In this chapter, the concerns of the security visualization domain is not the main subject. The focus will be mostly the features of the proposed system, but while describing and discussing these features, the concerns of the domain will also be mentioned to some extent. In order to gain more knowledge on the security visualization domain please refer to Sonmez and Gunel's extended review (2018).

Similar to all CPS systems, the proposed system has some other (not domain specific) concerns, such as security, privacy, fault tolerance, safety, and reliability. Some of these concerns will be satisfied with the properties of the proposed system. Others will be fulfilled by the features of underlying infrastructure elements. The modular structure and service-based design of the proposed system allows further progress and adding new features. It has advantages regarding genericness and scalability. The safety of infrastructure elements is left to the enterprise policies. Physical protection mechanisms

may be included whenever possible to increase the safety and physical security of enterprise network infrastructure. Other security concerns, and reliability concerns such as timing and ordering of events in a distributed system, and fault tolerance are fulfilled by depending on industry standard technologies and architectures. These issues are depicted more in the design description.

The aim of this study can be briefly summarized as designing a generic security visualization system which is capable of visualizing data coming from multiple sources in an enterprise and which forms a knowledge base as a result of these visualization tasks. Thus, the enterprise is defined as a CPS system first. The provided design can be applied to any enterprise with varying types and sizes as long as it possesses the necessary infrastructure suitable for its data. The security-related data sources supported with the provided design is limited to structured and uncompressed CSV, TXT, and JSON files. Other data sources can be included using third-party parser tools. The visualization systems in the scope are the JavaScript-based visualization systems which are presented via the HTML pages.

The rest of the chapter is constructed as follows. Section 4.2 includes the description of an enterprise network as a ubiquitous environment. Section 4.3 has the functional requirements for the proposed system. Section 4.4 has the initial design features, and section 4.5 has the improvements made by integrating the initial design with big data technologies. Section 4.6 includes the results achieved. Section 4.7 and 4.8 are devoted to the discussion and concluding remarks, respectively.

4.2 Description of an Enterprise Network as a Ubiquitous Environment

Enterprise security is defined by Sherwood, as protecting business goals and assets for an enterprise (Sherwood, Clark, & Lynas, 2005). Enterprises have substantial differences in their , since enterprise security depends on many factors, such as the criticality of business models, number and types of internal and external users, size of the data stored, types of business and/or infrastructure protection software, hardware used and network architecture. The approaches to information security management also vary from organization to organization.

The current trends of the enterprises lead to the growth of potential risks, such as moving to e-business, increased mobility, fast and flexible change management and, cloud computing. While the majority of the threats originate from the Internet for an organization, there may also be malicious actions that originate from insiders. Vendors of systems and devices offer specific precautions. These precautions have a variable set of behaviors and produce log files which are far from being processed using standard procedures in different organizations.

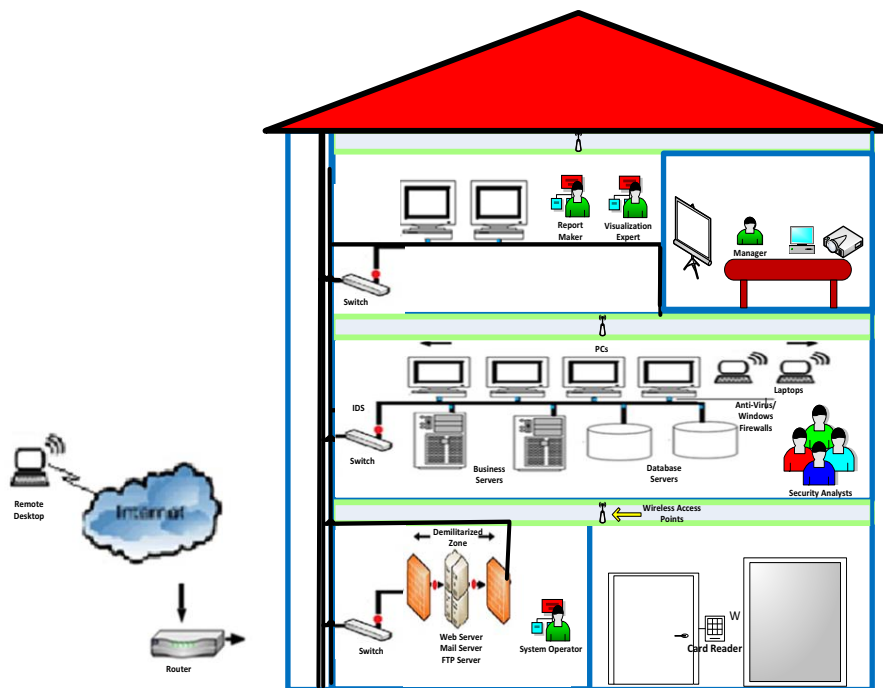


Figure 32 - Typical IT infrastructure of an organization

The owners of the IT investments in an enterprise may assume that everything looks perfect after building up the IT infrastructure by spending a non-ignorable amount of money. However, they will eventually understand that the vendors of the hardware and software elements of the IT infrastructure usually undertake the security issue. The best security solution is as good as how it is administered. Enterprise network and security management may include many devices and tools layered in various layers of the network. Enterprise infrastructure and devices have been elaborated in detail in Shin (2017). In order to have a base to examine enterprise-level security further, a possible set of hardware and software solutions is illustrated in Figure 32, which forms a sample IT infrastructure model for an enterprise.

In a cyber-physical environment, close coupling of cyber and physical devices is required. In the proposed system, the data produced by physical and computerized systems are aimed to be processed by cyber systems. To put it concretely, data to be treated may be as the data from intrusion detection system, data from card reader device, and the data from firewall logs. Although in the picture mostly familiar structures are shown, the limit is the networking and processing capability of the enterprise, thus, other devices which have other cyber capabilities, such as active RFID tags to protect business-

critical assets, sensors to monitor the heat, and humidity sensors in the system operation room, biometric access control systems may as well be in the picture.

A generic visualization system which is capable of storing and visualizing data coming from multiple sources will result in many benefits in such an enterprise. Due to the increase of data in information technologies, visualization has become a popular technique for analyzing, communicating and decision making for big data. Using visualization in the security domain is a relatively new research area. The first published work was in 2004. The major reason for the emergence of security visualization is the necessity of analyzing security-related vast data on time. Security visualizations enable human assessment of large size log files efficiently, which results in rapid and improved decision making. Marty (2009) described the benefits of using visualization in the security domain as “it answers questions, it poses new questions, it allows exploration and discovery, it supports decisions, it communicates information, it increases efficiency, and it inspires”.

Characteristics of cyber-physical systems include large-scale wired and wireless networking, cyber capability in most of the physical components, the networking speed in extreme scales, existence of high number of systems with various complexity varying from simple to too complex, existence of some unconventional systems with cyber physical capabilities (for example embedded authentication systems based on biometrics in an enterprise), existence of non-technical people in the control loop.

So far, the explanation of the enterprise as a cyber-physical system has been made. In the next section, the functional requirements of the proposed visualization system will be explained and, the non-functional requirements of the proposed system will be depicted together with design features.

4.3 Functional Requirements for an Enterprise Visualization System with Feedback from Users

Security analyses tasks are divided into three consecutive groups of activities. The first group of activities focuses on data collection, the second group of activities focuses on data preparation, such as filtering, normalization, and sampling, and the third group of activities focuses on the data analysis. Visualization is an effective way of data analysis. As mentioned before, prior to the system design, requirements are identified through the use of examination of enterprise needs, literature search, and inquiry results. During the preparation of these requirements, data preparation tasks such as cleansing, conversion, formatting, and normalization have been excluded. These are earlier undertakings which require manual reasoning, and hence they do not fit in as a part of an automated structure.

The requirements captured can be grouped into two categories. The first group is predominantly specific to the proposed design, and the second group is either already implemented by other studies in different ways or are known design issues. A few of the

requirements are included for the sake of completeness of the design. In the text below, for each requirement, a number is provided in parenthesis for traceability during design and validation.

A display type library is the first feature detected (Requirement 1). It is designed to be in a form to store information related to display types aiming the proper use and selection of various display types. Most of the enterprise users are not experts of display types or visualization technologies. The level of visualization knowledge has not been questioned explicitly during the survey. Nevertheless, a set of display type thumbnail views were included in the survey content. Some participants asked simple questions regarding these display types which shows that although they have expertise in the security domain, they need more support on display types. Each display type is powerful to exhibit some data classes. For example, scatter plots are more effectual to display large datasets. The reason is each point occupies a small space, and this allows visualization of extensive data in a small space in scatter plots. Treemaps are more suitable to display hierarchical data. Departmental data, data coming from hierarchical network devices may be more appropriate to be visualized with this kind of displays. Circular display types allow visualization of data including what, how, when, and where forms of information, such as events, alerts occurred in specific devices/hosts. Therefore, a dictionary-like platform including such information is necessary.

Ability to read data in multiple various formats (Req. 2) such as JSON, TXT, and PCAP files are required due to the high variety of security data sources. During the survey, the usefulness of examining twelve independent data sources was asked. The results show that all twelve data sources were nearly equally crucial to the enterprises. It is not desired that security analysts should give importance to one or two data sources and leave the others unanalyzed. Each of these sources has specific formats. Formats also change based on the brand of the security systems. Being ready for such a diverse set of security data sources is difficult for most of the enterprises. The system should facilitate the addition of a new type of security data sources for examination and visualization purposes (Req. 3).

The survey results also show that users are not familiar enough about their security data resources for their organizations which may affect the data preparation tasks. Survey results show that a platform regarding sharing such information among users may be very beneficial. In this platform, the information such as file locations, file access information, excepted file sizes, the frequencies of renewal for looped files, data formats, responsibilities, and tasks of staff regarding analyses of these log files, and, experiences can be shared (Req. 4). These feedbacks from the users can also be used for the automatic creation of visualizations (as a future work). During the design phase, how these feedbacks from users can be used to form a closed control loop in an enterprise which is essential for a cyber-physical environment will also be discussed.

A generic metric definition system is also an inherent requirement (Req. 5). Not all security visualization solutions enable the creation of data queries during runtime. Most of them run on predefined metrics. On the other hand, there are many attributes for each mentioned data source. These attributes are meaningful in specific ways. For example, some attributes between intervals, count of some qualities, min or max of some traits, and the set of some characteristics might be meaningful for various purposes. Allowing the user to define the queries for these attributes easily, during runtime would result in more user-centric metrics, rather than the predefined ones. Some of the general purpose visualization tools have excellent properties of forming generic user queries, and some enable selection of display types on the fly using very sophisticated user interactions such as drag and drop. However, these tools lack security perspective and do not help to form an enterprise security knowledgebase while visualizing the data.

Threats definition system (Req. 6) is another requirement detected by the survey. During the requirements analysis survey, the participants were asked to group analyses that they make using security data sources to monitor or detect a set of threats. As a continuation of this question, they were also asked to associate the attributes such as "number of events in a time duration", "types of alerts ", "list of source IPs " to the threat sets for each analysis group. Using right associations of visualizations to threats and other visualization purposes will be the key to success for the analysts. If people know what to do or what to look for, they are more likely to succeed. However, the survey showed that people have issues related to making these decisions. Although all of the survey participants were familiar with well-known threats, some other risks were not very well-known. Admitting that they knew the threat mechanism the majority of the participants had difficulty in associating a threat with a particular data source or with a particular data source attribute, resulting with many illogical associations of data sources to the threats or data attributes to the threats.

Similar to the feedback from the users related to security data sources recognized before, storing associations of threats to the visualizations (Req. 7) and associations of threats to data sources (Req. 8) in the knowledge base is useful. The problem is to find the correct associations. When asked theoretically as in the survey, people have difficulty in accurately giving answers to these questions. Showing something concrete, as the created visualizations, may allow getting more effective feedbacks with correct essence.

Examining the data through the use of visualizations to seek possible threats is one purpose in itself. Purpose definition system (Req.9) is also included in the requirements to enable the definition of other intents. Even some infrastructure elements may have different intents when installed differently. For example, a web server may serve to Intranet users or Internet users; a firewall may protect the overall organization or a department; a honey network may be used for protection or educational purposes. As the size of the enterprise increases, there may be multiple installations of the same hardware and software elements which have different purposes. Knowing the goals of these infrastructure elements may result in a better analysis of the generated data. Similar to

the threats purposes can be associated with data sources (Req. 10) or visualizations (Req.11). This requirement arose due to the investigation of data sources. Besides associating the threats and purposes to the visualizations, users are free to make other feedbacks to the viewings (Req. 12). These feedbacks can also be automatically processed. (as a future work).

Use of the various type of displays with different complexity levels (Req.13) is another requirement. The literature study showed that existing studies used various display types. Some display types are more mainstream contrasted with others. Survey results point that although users have an interest in more complicated charts, such as 3-D charts, results indicate that simple charts are easier to understand and have higher usability for the majority of the participants. However, some of the more complex display types, such as parallel coordinates are more proper for some specific cases. Association of display types to the generic data definitions on the fly is required (Req.14). Selection of the display type during runtime is a feature which exists mostly in some of the dashboard designs. For the sake of completeness, it was necessary to combine this feature with the generic data file and generic data metric definitions in the proposed design.

Assigning difficulty levels to the display types is also necessary (Req. 15). Not every display type has the similar difficulty level. This requirement arose due to the same rationality as in display type descriptions. Majority of the security experts are not visualization experts so the association of difficulty levels to the display types may allow better selection of displays during the visualization tasks for new users.

Easy access to external visualization tools is required (Req.16). During the survey, the participants were asked about their familiarity with a long list of security visualization solutions. Unfortunately, the participants were only familiar with only a few of these visualization tools. These are the visualization tools which are used in conjunction with other tools such as scanner tools. Due to this reason, encapsulating access information for such products, such as links to websites to download or use (for online tools) may be beneficial to increase familiarity to these tools for the enterprise users.

There are requirements which are captured through the literature search. These were also questioned during the survey. Having the ability to depict a relatively large data (Req.17), the ability to represent data from more than one security log simultaneously (Req.18), the ability to save detected patterns (Req.19), the ability to work with real-time data (Req.20), the ability to depict most types of attacks (Req.21) are some significant requirements among them. A few of the requirements from literature search are mainly related to display type technology selection. Displaying incident time (Req.22), having thick boundaries to separate different classes of information (Req.23), ability to use the visualization without mouse (Req.24), being interactive (Req.25), being searchable (Req.26), being zoomable (Req.27), being scalable in terms of the amount of data displayed (Req.28) are in this group..

Besides the listed ones, some design decisions are made based on technical background, industrial development standards, and the properties of existing display libraries which provide a high level of diversity in form and difficulty. Due to the variety and richness of JavaScript-based display type technologies, using JavaScript-based display type technology is decided. Java development language is selected due to background knowledge and its high compatibility with the mentioned display type libraries, and a web-based design is to enable easy access from any computing device for an enterprise.

4.4 Initial Design Features

There are many concerns related to the design and development of cyber-physical models. While deterministic models allow the creation of more robust models, it is not possible or feasible to create a deterministic model consisting of many parts in a distributed manner at all times (Lee E. A., 2015). Thus, the authors decided to implement a prototype which may be used for the materialization of some design problems through trial and error.

In top-down design, the modeler starts with the domain at large and starts with the design of the upper-level application modules and divides the top level structure into smaller pieces, generally in the form of classes. This results with classes with either complex or coercion relations. In the bottom-up approach, the design of the necessary data structures is completed first. Later, first simple, then complex functions are added to form an integrated structure to process the anticipated design structure and to fulfill the requirements. A bottom-up approach has been taken during the prototype design. The data entities have been extracted from the requirements and the data model shown in Figure 33 has been created as a first step. The detailed structure and Java code of these data entities will be shared on Git-Hub under the name of “Data Entities for Generic Security Visualization Solution with Knowledgebase”. As a next step, the functions and data structures are mapped to modules. For each module, the front end and back end classes have been implemented.

The overall functionality including relations to data structures is described in the state diagram shown in Figure 34. The solid lines correspond to service calls which transform the data into another state. The dashed lines are used to show the relations of the data structures from the proposed design. The development of blue colored transformations is straightforward. The orange colored parts consist of specific structures which form the genericness of the system. The green colored part includes algorithms specifically designed for this application which converts non-generic file formats to the defined generic file definitions. The red colored parts consist of the combination of both custom data structures and algorithm. Finally, the pink colored part includes XHTML files adapted from third-party display libraries and display content adapters which both form a boilerplate for a display type. These boilerplates are based on external JavaScript based visualization libraries.

The brief explanation of how each part of the data structure is associated with the features of the proposed system is as follows. *BasicEntity* is the abstract class which is on top of the hierarchy for all entity classes. Instances of *GenericDataFile* and *GenericFileElement* classes are used to define various types of security data sources which possibly exist in an enterprise network. The list of *GenericFileElements* is used in order to parse the element values from the data. Two separate parsers are required one for JSON formatted files and one for the TXT formatted files. For PCAP files, third-party PCAP to TXT converters might be integrated while a PCAP parser is implemented. Regardless of the file format and number of attributes, the element values are stored in a data structure named *DataStore* in tuple formats. In order to define and execute User Requested Queries on Generic Data Store “SQL” query like queries are defined involving selected fields, group by fields and query conditions. These queries are converted to hibernate queries by using associations of *GenericDataFileElements* to *DataStore* elements which select from generic *DataStore*. Each query is given a name and called as a *GenericDataFileMetric*. Each query result is stored in a structure named *DataStoreQueryResult*. Selection of destination IPs and the ports initiated from a specific Source IP, selection of the number of source IPs grouped by honeynets, min port number accessed between a time interval are sample queries that are defined in the prototype system using test data. The prototype allows association of query results which are called as data metrics to display types. A query result may be associated with more than one display type, or one display type can be associated with more than one query result. It is important to make correct associations of *DataStoreQueryResult* structure data types to display type data fields. For example, in the bar chart visualization script data array d1 can be any numeric type such as Integer, Double or Long, and data array d2 can be any categorical or numeric type.

As the *DataStoreQueryResult* instances encapsulate a list of result data types, the correctness of user-defined association can be made by the *DisplayType* specific *ContentAdapter*. Associations of these metrics to display types are stored in the system, *VisCode*. When the display is actually associated with data an instance of *VisCodeWithValue* is created. The visualization in the proposed system are designed to be in dashboard style. This will allow visualization of selected metrics of the data coming from multiple sources simultaneously encapsulating subviews in the same screen. The classes named *Threat*, *Purpose*, *ExternalVisualizationSystem*, and *UserFeedback* are not directly related to visualization generation but part of enterprise visualization knowledgebase. The states marked with colors are associated to phases with data entrance to the Visualization Knowledgebase. Yellow marked states are relevantly static, in big data terminology having low velocity, and low volume, medium variety; green color indicates points that will grow when the associations are made to threats or other purposes having low volume, low velocity, and medium variety, the red color points out user feedbacks with topics which are expected to be the most dynamic part that is growing fast in time, having high volume, high velocity and high variety.

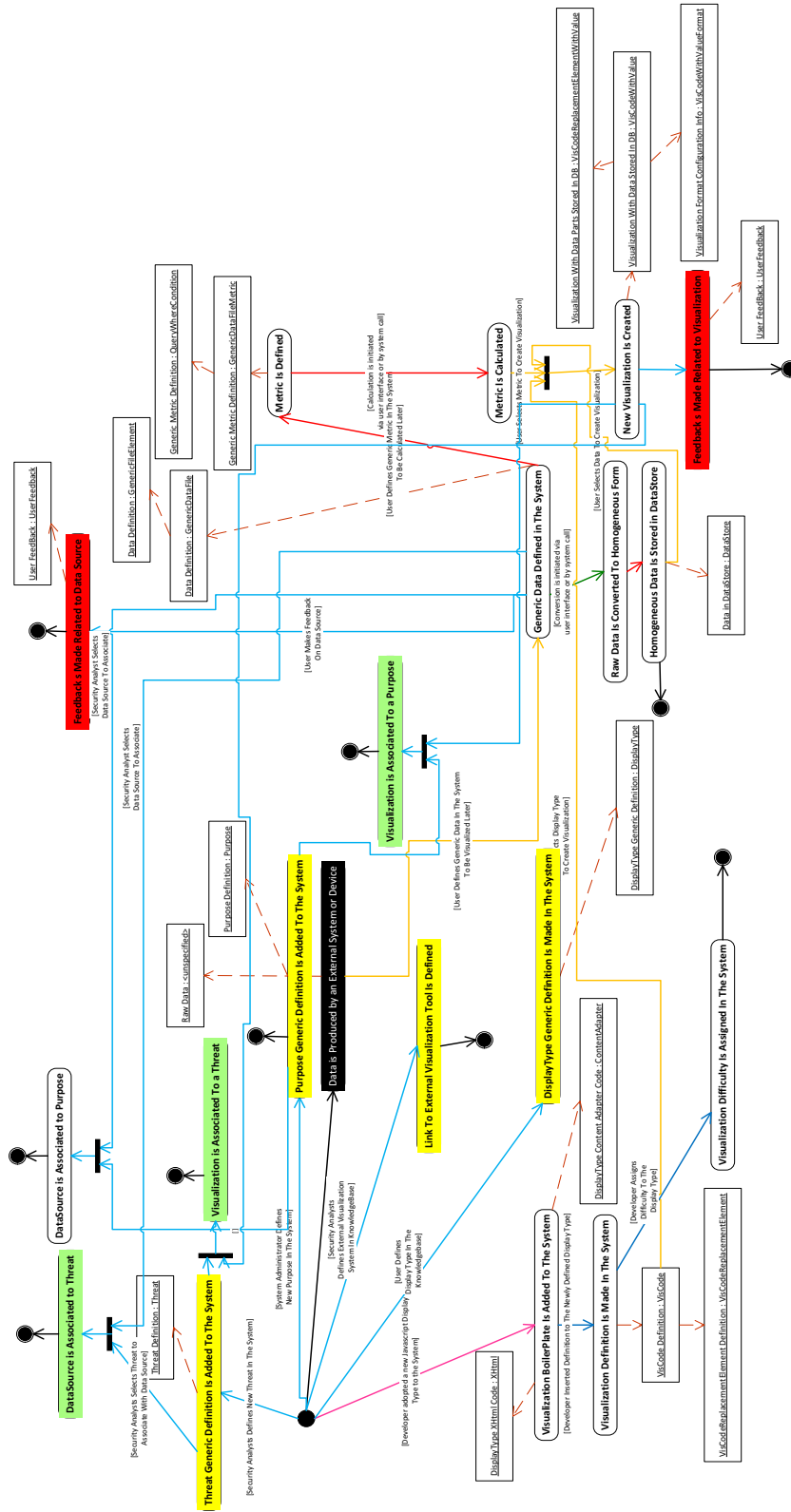


Figure 34 - States and activities of the proposed system



Figure 35 - a) XHTML content for a JavaScript-based display b) ContentAdapter structure for Flotr JavaScript library based Bar Chart visualization c) Sequence diagram for visualization display in dashboard form

4.4.1 Extensible Display Type Library and Dashboard Design

In order to gain knowledge on display type technologies, some JavaScript-based visualization libraries have been explored. The libraries which are examined so far are Flotr (Humble Software, 2018), FlotCharts (Laursen, 2018), Data Driven Documents, D3, (Bostock, 2018), and Sparklines (Splunk, 2013). Flotr allows drawing simple static charts such as bar chart, line chart, pie chart. FlotCharts allows more interaction such as zooming in and out. D3 allows custom visualization of data. Sparklines allows better integration of text and data by using inline charts and visualizing more data by encapsulating sparklines in a table. The JavaScript-based visualization libraries are not limited to this list. Surely, there are other alternatives.

The existence of a high number of available display type libraries with different properties and difficulty levels resulted in the decision of using these third-party scripts in the proposed design. In the proposed design, for each display type, an XHTML file which embeds necessary display container and the corresponding reference to the JavaScript library and required JavaScript source has to be prepared. In order to fill the XHTML page correctly with data, a Java class which is called as *ContentAdapter* has to be implemented for each display. The use of Java interfaces standardizes this part of the implementation. XHTML page structure, a sample hierarchy of *ContentAdapter* for bar chart visualization based on Flotr JavaScript library, and a sequence diagram showing the visualization generation and exhibition via dashboard are shown in Figure 35 (a), (b), and (c) respectively.

This design allows the use of JavaScript display libraries with various designs. The standardized structure allows easy integration of new libraries. A significant portion of the requirements, such as zoomable design due to interactive display type library, incident time display due to proper chart design selection, are fulfilled due to this extensible display library design structure.

4.5 Evolving the Initial Design Using Big Data Technologies

In the previous section, an enterprise security visualization solution design was introduced. Basically, it allows the dynamic definition of various types of data files, which happen to correspond to log files. It again allows the definition of metrics on these data files. It has JavaScript based visualization boilerplates. Using these boilerplates visualizations are created by associating the displays with the metric calculation results.

Distinguishable features of the proposed prototype structure are its genericness which enables the use of different log files without pre-knowing the data structure, its visualization boilerplates which enable easy adaptation of available JavaScript based displays, its permanent structure which enables storing of metric values, and corresponding visualizations created earlier, and the knowledge base formed through the use of some static information with some associations, and user feedback.

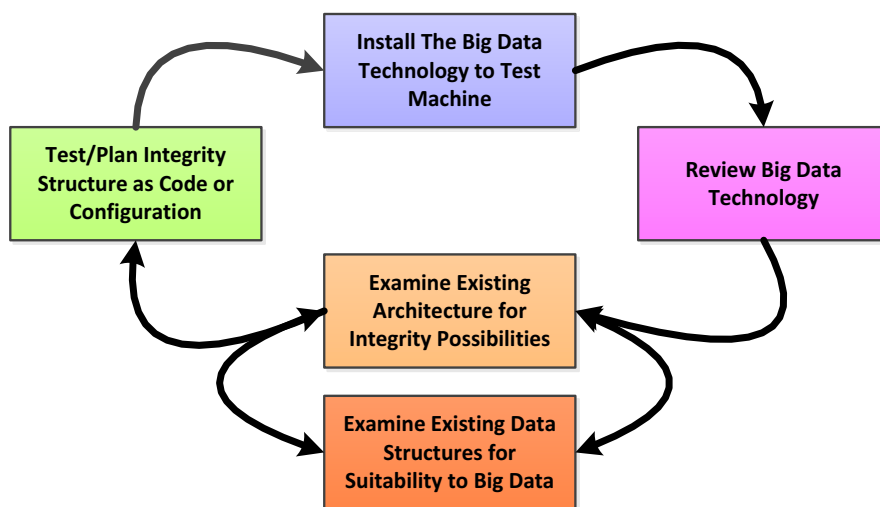


Figure 36 - Methodology to find integration points with big data technology

However, it lacks some features such as the ability to process real-time data, the ability to store and process extensive log files for very large enterprises, encapsulating horizontal scalability, high data reliability, and parallel processing abilities. Thus, another step has been taken in the study. Big data technologies have been examined, and the prototype design has been improved by integrating the design features with the appropriate big data technologies. First, the big data technologies which are selected to

be integrated into the design will be introduced shortly. Following this introduction, the intersections of the enterprise security visualization solution and the related design decisions will be described.

4.5.1 Big Data Technologies

Apache Hadoop has been selected, because it is the most popular big data ecosystem at the present day. This ecosystem envelopes big data technologies for various purposes. Technologies related to storing of big data in a distributed manner, related to analyzing big data, running queries and algorithms in a distributed manner, related to streaming big data from multiple sources possibly distributed in multiple machines, and related to collecting big data in real-time from multiple nodes are part of Hadoop. Some of the technologies which are demonstrated as a part of Hadoop also have standalone designs which allow them to work independently. The methodology used during the evolving of the initial design with big data technologies is shown in Figure 36. First, Hortonworks Hadoop Sandbox (Hortonworks, 2018) has been installed on a virtual machine. All of the technologies available in the sandbox have been reviewed. All along this revision, besides responsibilities, available API's, compatibility with Java language, structures (structure, semi-structured, non-structured) have been examined and tested to some extent. In order to integrate the existing structure with the big data technologies, the data structures which are part of the first design have been reviewed. During this work, for the selected technology, the integration is planned either as code or configuration structure. As a result of this effort, two new designs which are called second design and third design for security visualization solution have been prepared. In this part, the technologies which are part of the evolved security visualization solution will be briefly reminded. Later, in the next section how these technologies are integrated into the new design structures will be explained.

Some of the big data technologies are used inherently due to their roles in Hadoop (Zikopoulos & Eaton, 2011) ecosystem. Hadoop Distributed File System (HDFS) (Cohen & Acharya, 2013) is a file system for large volumes of data. It consists of name nodes and data nodes providing a distributed file system encapsulating multiple nodes in possibly multiple Hadoop clusters. It is optimized for handling large files, but it can also handle small files. It stores files by breaking them into blocks. These blocks are distributed among several computers. In order to handle failures, it stores multiple copies of any block. Name node keeps track of where each copy of each block is stored.

The role of Apache Zookeeper (Haloi, 2015) is to maintain the Hadoop configuration information and act as a name lookup service. It is the prime building block providing distributiveness of the Hadoop environment. Apache Hadoop Yarn (Vavilapalli, et al., 2013) is responsible for separating the resource management and processing tasks. The Yarn is also responsible for arranging that the processes using some data from one HDFS block runs on the same node with the HDFS block. Even if it not explicitly initiated by any application, it is there managing the Hadoop cluster's resources.

Apache Ambari (Wadkar & Siddalingaiah, 2014) is an open source management platform which provides management, securing and monitoring of Hadoop clusters. Ambari is not directly part of the proposed solution, however, when the flat files, and some part of the database is moved to Hadoop, Ambari will be a valuable tool to manage the Hadoop and will act as a file manager and a database client. Moreover, using Ambari will provide users new interfaces to run custom analyses based on other distributed big data processing methods such as running Apache Map-Reduce (Dean & Ghemawat, 2008) tasks, Apache Spark MLlib (Meng, et al., 2016) machine learning libraries or Apache Pig Latin (Olston, Reed, Srivastava, Kumar, & Tomkins, 2008) scripts on the data stored in the Hadoop clusters.

So far the big data technologies which are inherently or seamlessly used are mentioned. Some of the technologies are used for specific purposes in the evolved architecture. Apache HBase (Vora, 2011) is a column-oriented database running on Apache HDFS. It depends on the Google's Big Table (Chang, et al., 2008) architecture. It is designed for low latency operations. It is a non-relational NoSQL database. Since the rest of the solution is relational and runs via SQL queries generated through hibernate, in order to make minimum changes on the architecture some additional big data technology which converts ordinary SQL to HBase queries is encapsulated.

Apache Phoenix (Apache, 2018) is an open source relational database driver tool for Hadoop. It makes a bridge between HBase's low latency world and applications using OLTP. It enables benefiting with standard SQL queries and JDBC API, and it also enables ACID (Atomicity, Consistency, Isolation, Durability) transactions over non-ACID compliant HBase database. It takes standard SQL and converts it to a series of HBase scans which is later transformed into standard JDBC ResultSet.

Apache Kafka (Garg, 2013) is a general purpose publish/subscribe type messaging system based data streaming tool. Streaming technology allows processing new data as it is generated. Kafka servers store all incoming messages from publishers for some period and publish them into a data structure called topic. Kafka consumers subscribe to one/or more topics. In the proposed design Kafka will act as a data collector for the visualization system.

The proposed system has its intrinsic features. However, there are also highly enhanced commercial or open source big data visualization systems available in the ecosystem. Some of these big data visualization systems are developed as a part of the Hadoop ecosystem such as Apache Zeppelin (Apache, 2018). Apache Zeppelin is a web-based notebook which enables data exploration and supports technologies like Apache Spark (Apache, 2018), SQL or Python. Some of the external visualization tools are developed independently from Hadoop but provide ways to be integrated with big data stored in Hadoop clusters, such as Tableau (Tableau, 2018) and Qlik (Qlik, 2018). There are also common purpose visualization systems based on technologies such as D3.js (D3.js, 2018) or R (R Foundation, 2018) which are also commonly used for the visualization of

big data. In the evolved version of the proposed visualization system, some new ways will be offered to associate visualizations created with these enhanced visualization systems.

Apache Sqoop (Jain, 2013) is a data transfer tool which is used to transfer data from structured resources such as relational databases to Hadoop. Use of this tool is anticipated as a part of the third design. Similar to Apache Sqoop, Apache Spark (Kane, 2017) (Zaharia, et al., 2016) is part of the third design structure. Apache Spark is a technology which allows processing of massive amounts of data in various ways. It has features including data streaming, machine learning, and graph analysis. Spark streaming receives a stream of data, divides it into batches, and processes to generate a final stream of data.

4.5.2 Big Data Technologies Related Design Decisions For Generic Enterprise Security Visualization Solution

There is a variety of big data technologies. Conceding that the target is to adopt a web-based application to big data technologies, the application can evolve into many different structures embracing various integration items. There is not one right structure for a big data related visualization design. In this design study, the main principle is to make the change while sticking to the original service methods and data structure substantially.

The first version of the security visualization prototype is a Java web application using Spring (Johnson, Höller, Arendsen, Risberg, & Sampaleanu, 2009) and Hibernate (Konda, 2014) frameworks, and a relational database structure running on MySQL database for data storage and processing purposes. The input data was expected to be flat files stored in the operating system disk space in different formats such as TXT, CSV, or JSON. Users make a request for various phases of the visualization generation from the web interface. The service layer fulfills these requests. During the evolving of the first design, besides the technical work, existing data structures are examined mainly for their suitability to be saved and to be processed by big data technologies. Expected size, processing requirements, level of being relational regarding existing relations to other data structures, level of usability by third-party analysis tools were the focus points during this examination. As a result, Table 12 was formed. This table does not include the data enumeration structures and the abstract data structures.

The CRUD (create, read, update, delete) like operations initiated from the user interfaces are used to create and modify the data structures related to file structure definition, metric definition, visualization boilerplate definition. ORM (Object-Relational Mapping) (Myerson, 2002) technologies were selected due to its ability to provide increased performance and scalability while providing protection against SQL injections. This part of the data mainly is the metadata required for the visualization. In all three designs, this metadata stays in the relational database.

Table 12 - Examination of Data Structures for Suitability to Big Data Technologies

Data Structure	Expected Size	Being Relational In Nature	Processing Difficulty	Useful for Third-Party Tools
Flat Files (Text File, JSON File)	Small to very large	N/A	Easy to Difficult based on file size and format	Yes
Generic Data File Definition (Generic_Data_File, Generic_Data_File_Element)	Small	Has complex relations with other data structures	Easy	No
Generic Data File Metric (Generic_Data_File_Metric, Query_Where_Condition)	Small	Has complex relations within each other and with other data structures	Easy	No
Metrics Calculation Result (Data_Store_Query_Result, Data_Store_Column_Data, Data_Store_Row_Data)	Small to medium	Has complex relations within each other and with other data structures	Easy to Medium based on Visualization Boiler Plate	No
Visualization BoilerPlate Definition (Vis_Code, Vis_Code_Replacement_Element)	Small to medium	Has complex relations within each other and with other data structures	Various difficulty based on Chosen Javascript Library	No
Visualization (Vis_Code_With_Value, Vis_Code_Replacement_Element_With_Value, Vis_Code_With_Value_Format)	Medium to large	Has complex relations within each other and with other data structures	Easy to Medium Based on Visualization Dashboard Features	No
KnowledgeBase-Direct Feedback(UserFeedBack)	Medium to very large	Has simple relations with other data structures	Medium for Automated Processing	Yes
KnowledgeBase-Static(Threat, DisplayType, ExternalVisualizationSystem)	Low to medium	Has simple relations with other data structures	Easy	Yes
KnowledgeBase-Associations (Vis_Code_With_Value_Threats, Vis_Code_With_Value_Threats, Generic_Data_File_Threats, Generic_Data_File_Purposes)	Small to medium	Has simple relations with other data structures	Medium for Automated Processing	Yes
DataStore	Small to very large	Has one way relation to Generic Data File	Medium to Difficult Based on Chosen Metric	Yes

In the first design, flat data files are static files which are defined and accessed through their operating system path and file name. In the second design, these files are collected through the use of Apache Kafka file connectors and stored in the Hadoop HDFS structure. This time, the service layer reaches these files via HDFS URL value assigned for each. The third design includes streaming via Apache Spark technology. By this way, the data from flat files collected by Apache Kafka is directed to Apache Spark for streaming. In order to associate the file structure with the segmented streamed data Apache Spark SQL is used. Through this, a fundamental part of data processing tasks is moved to Hadoop in the third design. This design will also result in real-time processing of the input data.

DataStore is a structure used to store data in various forms in a homogenous structure after parsing of input data. This *DataStore* is used together with the *GenericDataFileElements* while producing and executing user-defined queries (metrics). In the first version, this *DataStore* was part of the relational database. Due to its expected size and usability from external analysis tools, in the second design, this data structure was moved to Apache HBase non-relational database. Since Apache HBase can not be directly reached from Java-SQL environment, Apache Phoenix which is called SQL skin for HBase is used as a layer. This layer allows conversion of normal Java Database Connectivity (JDBC) calls to HBase scans.

In the first design, users give feedback using web user interfaces. The user interfaces are additionally used to share the previous feedbacks. Each feedback comprises of a topic, which is chosen from a predefined set and a text-based feedback content. The reason for including a structured feedback mechanism in the visualization system is to permit automatic processing of these feedbacks later on. Automatic processing of the feedbacks may cause a timely response to events and increment the speed of information sharing. Users of the visualization system can enter remarks/assessment results/report entries/commands based on the data analysis results immediately. Each topic may relate to a particular purpose, such as the automatic communication of data in various ways such as e-mail or reports (topic:WEEKLY_REPORT, content:check IDS alerts in detail), or the automatic command execution, such as commands for firewalls, commands for active directory (topic:ACTIVE_DIRECTORY_COMMAND, content:NET USER loginname newpassword /DOMAIN).

The knowledgebase data is created with CRUD activities through the user interface and supposed to be viewed from there combined with other metadata information. The amount of feedback as a part of knowledgebase is likewise expected to grow quick in time. Thus, to permit automated processing of these feedbacks in the future, in the third design, user feedbacks are decided to be replicated in a denormalized form in HBase using Apache Sqoop. Sqoop can be configured to replicate the views having joins of multiple tables on the relational database to the HBase as the new tuples of data entered in the knowledgebase.

The last change made to the first design is encapsulating a visualization boiler-plate for external URL based visualization systems. In the first design, a visualization boilerplate was required to use a JavaScript-based visualization. For each boilerplate, there is a content adapter class and an XHTML based display type code definition which include the JavaScript code for that specific display type, as mentioned before. There are various external visualization systems associated with visualization of big data. Most of these visualization systems are accessible via web URLs. In the second version of the enterprise visualization system, in order to benefit from these external visualization systems, an XHTML page which merely includes embedded external content and a content adapter which is not responsible to any data processing task are prescribed. Utilizing a boilerplate which allows integrating URL based visualizations created by external visualization systems will cause the additional benefits, such as using existing URL based visualizations, integrating with third party tools. Figure 37, a, b, c, and d shows all three versions of the enterprise security visualization design structure and Spark streaming details. As the design improved the amount of processing and data storage increased in the big data environment, as illustrated in Figure 37, e.

4.5.3 *Security Concerns*

As the ubiquity of the designs rises, the privacy and security requirements of the system will increase eventually. A multi-layered Internet of Things, IoT, system has the following layers: perception layer, network layer, and application layer. In some architectures, a service layer which is responsible for service management and service discovery is added to the other three layers (Lin, et al., 2017). Each layer has its security concerns. Security concerns of an IoT application which has a web-based interface were previously depicted in Özdemir Sönmez's IoT case study running in IBM BlueMix platform. (2016). The security concerns of the proposed system have similarities to that list. However, in the proposed system these concerns are fulfilled in different ways, mainly depending on the security features of industry-standard platform choices.

The security of the environment running the proposed design is the first concern. The proposed system is designed to be used in an enterprise environment which would have a firewall, an intrusion prevention system, and network security controls. These protections may be supported by server level hardening mechanisms. The second concern for the cloud IoT system is the genuity of the cloud platform which does not apply to this design. The third concern is the authentication of the web application which will be fulfilled using a standard way to implement user authentication, Spring Security (Mularien, 2010).

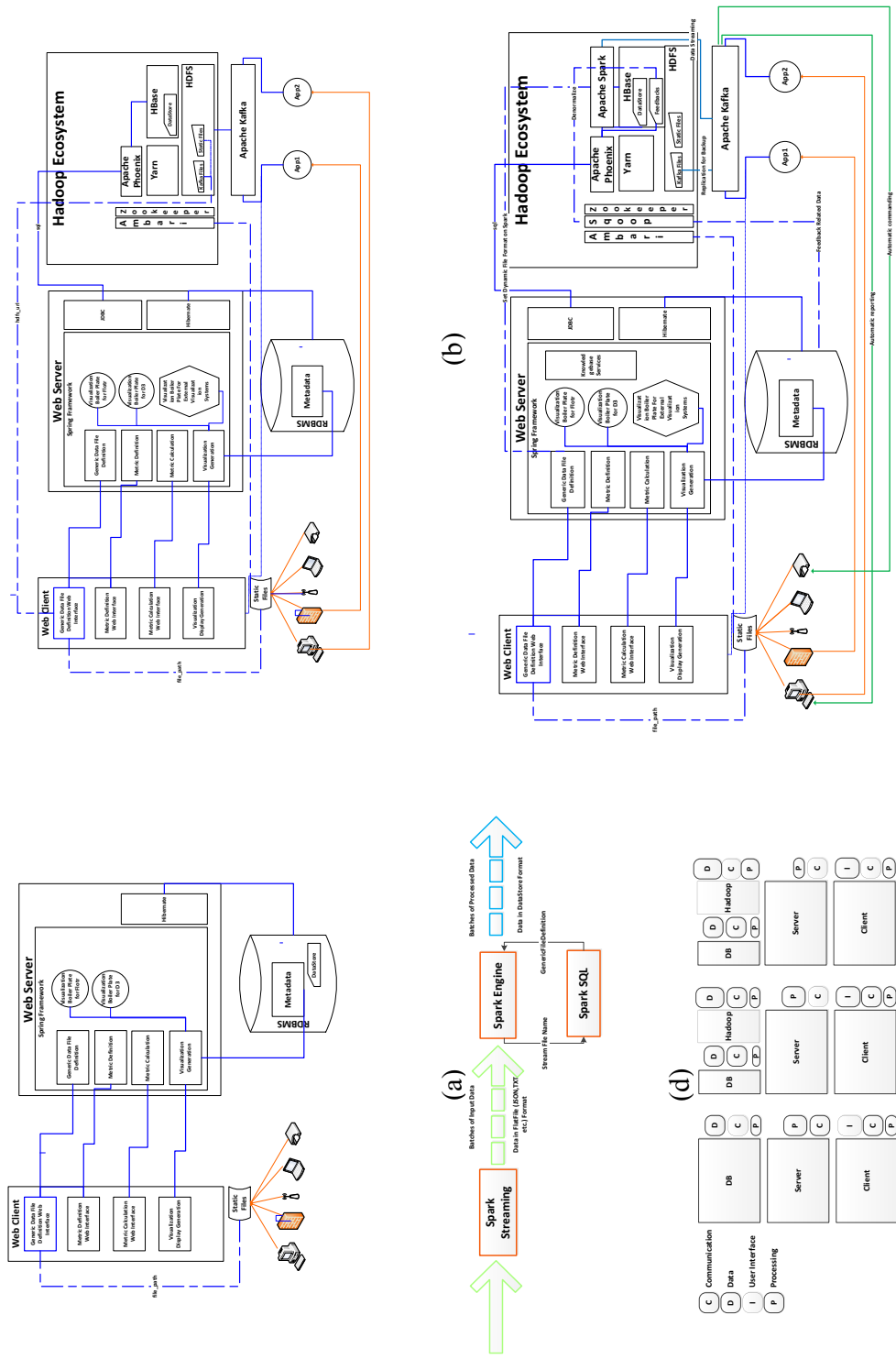


Figure 37 - a) The first design architecture, b) The second design architecture, c) The third design architecture, d) Streaming details for third design e) Evolution summary

The fourth concern is the authentication of the devices to the application. In a standard Kafka installation, any user can write any messages to any topic. However, in a more advanced setup, Kafka provides authentication of Kafka clients via SSL or SASL. The fifth concern is the security of the authentication data stored in the cloud which does not apply to the current solution. However, this time the security of data stored in the relational database and the Hadoop is the concern for which the system will rely on the protections of the underlying technologies. Each Hadoop component has its authentication, authorization, encryption of data at rest, and encryption of data in transit (Sharma & Navdeti, 2014). Similarly, contemporary database systems have advanced protections systems (Basharat, Azam, & Muzaffar, 2012). The sixth concern is a secure gateway between various platforms, in the distributed system. Although not included in the proposed design Hadoop has multiple gateway structures. The seventh concern is secure messaging between devices and the application. In the proposed design, the devices/applications are not expected to send direct messages to the application, but Kafka file connectors are in charge of reading device/application log files. The eighth concern is preventing the data leakage between devices, and the ninth concern is to prevent the data leakage between devices and the application. The devices are not expected to communicate as a part of the proposed solution. As mentioned above, all issues related to the communication of devices and application depends on Kafka security.

4.6 Results

In order to test critical implementations and the proposed data structure, a web-based prototype has been developed as a part of this study. The coverage of the requirements has been self evaluated based on this prototype. Table 13 is a traceability matrix which shows whether the requirements are met, not met or requires modification by the decided design features. The modifications are thought of as future work which may include implementation of new modules, new adaptors, and parsers for new file formats.

Table 13- Traceability of the Requirements & Detailed Design Features

#.	Short Description	Generic Data File	Metric Definition	Metric Calculation	Visualization Gen.	Standardized Display Boiler	Dashboard Design	Known Gebas	Use of HDF S and HBas e	Generic Display Boiler Plate	Kafka Data Collection	Use of Other Structures	Avg . Review Results
1	Display type library							(+)					3.75
2	Read data in various formats	(+)											4.50
3	Addition of new data sources easily	(+)											4.50
4	Feedbacks for data files	(+)						(+)					2.25

5	Non-predefined metrics	(+)	(+)						4.50
6	Threat definition					(+)			3.75
7	Associating threats to visualizations			(+)		(+)			3.75
8	Associating threats to data sources	(+)				(+)			3.75
9	Purpose definition					(+)			3.75
10	Associating purposes to data sources	(+)				(+)			3.75
11	Associating purposes to visualizations			(+)		(+)			3.75
12	Feedback for generate visualizations			(+)		(+)			2.25
13	Used of display types with various complexity								4.75
14	Display type on the fly								4.50
15	Visualization display difficulty			(+)		(+)			3.75
16	Access to external visualization			(+)				(!)	3.25
17	Depict large data					(+)		(+)	4.50
18	Simultaneous display					(+)		(+)	4.50
19	Save detected pattern			(*)		(+)			3.50
20	Work with real-time data							(+)	4.00
21	Depict most type of attacks	(+)	(+)						4.50
22	Displaying incident time					(#)			4.00
23	Thick boundaries between classes					(#)			4.00
24	Visualization without mouse					(#)			4.00
25	Being interactive					(#)			4.00
26	Being searchable					(#)			3.00
27	Being zoomable					(#)			4.00
28	Being scalable ITO data display					(#)			4.50

(+) This requirement is met through this design feature

(#) This requirement can be met through integration of proper visualization libraries

(*) This requirement requires updates to this design feature

(!) This requirement is allowed using primitive ways

One of the critical concerns is the provided level of genericness of the system. The system was evaluated with sample data, and the suitable examples are provided in Table 14. Although, there are known issues with the parsers, such as allowing single separators,

encapsulating nested collections as a part of row data, these missing points can be improved without any side effect to the overall structure. The proposed design also does not handle the compressed file inputs as is.

A significant concern to be tested is the storing of generic data in a non-generic format to be queried. No issue has been detected related to this concern. However, there are limitations to the current design which may be extended easily. The first limitation is the ninety-four generic queries for conditions identified in the current design, which may be extended. The second limitation is the maximum number of fields for each data type (present limit is ten) in a log file.

Table 14 - Sample Data Sources and Their Representation in the Proposed Design

Modern Honey Network Alert Log: Type=JSON, , Separator = “-”, Elements: oid, destination_ip, protocol, hp_feed_id_oid, timestamp_date, source_ip, source_port, destination_port, identifier, honeypot

Line 409755: { "_id" : { "\$oid" : "58c1d06f58e5cf04aff99ea3" }, "destination_ip" : "200.200.200.201", "protocol" : "pcap", "hpfeed_id" : { "\$oid" : "58c1d06e58e5cf04aff99ea0" }, "timestamp" : { "\$date" : "2017-03-10T00:00:14.147+0200" }, "source_ip" : "221.229.162.121", "source_port" : 4405, "destination_port" : 22, "identifier" : "fea0bde0-5d6d-11e6-9709-000c297e338e", "honeypot" : "p0f" }

Web Access Log: Type =TXT, Separator = “-”, Elements: IP, Dummy1, DateAndTime, Dummy2, MethodAndURL, SystemInfo,

117.201.11.139 - - 02/Jan/2017:02:35:43 -0800 "GET /wp-login.php HTTP/1.1" 404 295 "-" "Mozilla/5.0 (Windows NT 6.1; WOW64; rv:40.0) Gecko/20100101 Firefox/40.1"

117.201.11.139 - - 02/Jan/2017:02:35:49 -0800 "GET /wp-login.php HTTP/1.1" 404 295 "-" "Mozilla/5.0 (Windows NT 6.1; WOW64; rv:40.0) Gecko/20100101 Firefox/40.1"

Hardware Firewall Log: Type =TXT, Separator = “|”, Elements: Count, fw1src, fw1service , fw1proto, fw1action, fw1tcpflags

| 2 | 192.168.184.5 | 80 | tcp | accept | NULL |

| 2 | 172.16.224.16 | 80 | tcp | accept | NULL |

| 1 | 172.16.100.38 | 80 | tcp | accept | NULL |

The proposed system uses the JavaScript-based displays as is. Thus, any advanced JavaScript-based display type having advanced display properties, such as interactivity, zooming, or having proper display designs, such thick boundaries, and displaying incident time can be integrated to the proposed design based on two conditions. The first condition is the display code should be represented as shown in Figure 35-a. The second condition is developing display type specific ContentAdapter Java code should be both probable and feasible.

Another significant concern was to display the generated visualizations in dashboard form. For this purpose, the PrimeFaces (Çalışkan & Varaksin, 2013) Dashboard control was used in the prototype. This component has built-in dashboard features such as drag and drop, resize, and reorder of dashboard parts.

Attempts to integrate the initial design with big data technologies resulted in small changes in the original design. However, these changes resulted in an extensive list of benefits. For example, moving the input files to HDFS resulted in a more scalable structure for storage of large files, and, low latency due to Hadoop's vast data file processing capabilities. After moving the input files to shared storage, they became reachable by other data analysis and data visualization tools. Moving the data, DataStore and Feedbacks to the HBase increased the scalability and performance of data processing and data storage. Besides, this allows execution of other big data analyses available in the Hadoop. Original web interface, controller and service structure was based on building dynamic queries on DataStore based on SQL capabilities including the mathematical SQL functions as shown in Figure 37. Since the original analysis methods are protected, the same queries are arranged to be run on Hadoop HBase non-relational database. At this step, Apache Phoenix is the main catalyzer of the overall process. As mentioned in the previous section, it allows running SQL queries on HBase by converting SQL to HBase Scans resulting in low latency queries. It provides ACID properties which allow OLTP over data, and it returns a standard JDBC structure which can be further converted to Hibernate objects so that metric calculation results can be stored and can be associated to the selected display types in the relational database. In the first design, static files stored in the operating system disks are used for visualization purposes. In the improved design, Apache Kafka is integrated with the design to enable automatic data collection from multiple points for visualization purposes. Integrating Apache Kafka to the original design resulted in near real-time examination of raw data files, painless collection of visualization data from multiple points/nodes, and the standardization. As a next step, Apache Spark improved the near real-time design to be real time.

So far, the benefits of big data technologies which are mainly related to the initial requirement set are mentioned. There may be some additional benefits. For example, in this design Apache Sqoop was used to move data from relational data storage to Hadoop. It may as well be used to export data from Hadoop to other data storages with different formats if required in the future. Another example comes due to having the data in Hadoop storage. This allows reaching data from third-party tools. One last example is being able to run machine learning algorithms from Apache Spark MLlib library for security analysis purposes.

It is necessary to depict the critical points when processing large data which depends on several factors. The benefits of Hadoop for processing and storing big data is already mentioned. Other issues include the capability of the visualization script to display large data, and the processing algorithms. Authors found out some of the mentioned display

libraries are already known for their proper performance with big data. Some of these, for example, D3, provides online performance test tools. There are three main algorithms; parsing and storing task algorithm, querying task algorithm, and content adapter task algorithm. In general, these have the following algorithmic complexities in Big O (Abu Naser, 1999) notation respectively, $O(n)$, $O(\log(n))$, and $O(n)$ where n is the number of rows in the dataset. However, if the complexity of the content adapter algorithm for a specific display increases the latest complexity may change.

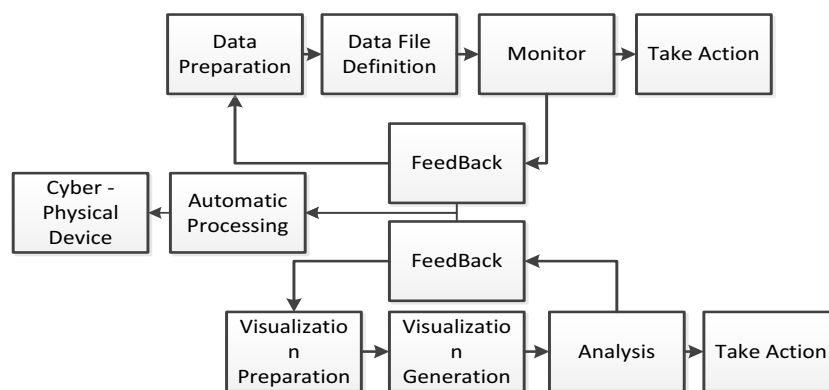


Figure 38 - Feedback loops

Feedbacks are included in the system for increasing learning and, for information sharing. Moving the feedbacks to the HBase allows further examination of these feedbacks. Making queries to ask the number of feedbacks for each topic, and for each topic and feedback item pair is a straightforward processing way of the feedbacks. These feedbacks may help users during the data preparation and visualization generation tasks. User feedbacks may also trigger certain conditions and events in cyberspace when appropriately processed. These feedbacks may form closed control loops as shown in Figure 38. Identification of feedback topic types and other ways of feedback processing is left as future work.

Benefits of generic visualization boilerplate include displaying of same raw data with external visualization systems and displaying the visualizations created by external visualization systems without leaving the enterprise visualization system. Simultaneous view of data by external systems and the proposed system in the same dashboard will enable comparison and may result in showing different aspects of the same data.

As a part of the validation effort, a series of semi-structured expert judgment interviews have been done. The last column of Table 13 corresponds to the average reviewer scores in five scale (1-5). The validation efforts and results is further discussed in the next section.

4.7 Discussion (Including validation efforts)

Difficulties of the validating cyber-physical systems and conceptual systems are known concepts. The issues that can be or can not be verified vary based on the validation subject. The proposed design is partly conceptual. For the conceptual parts, tests were made to check the interoperability for most of the parts. For the implemented elements, the prototype was used to test the functionalities and coverage. Hence, some of the issues were verified conceptually, some parts were verified through the run of the prototype, and by inspection. During the evolution of the design, two presentations were made to the academics, and feedbacks from these meetings resulted in the improvement of the design in stages.

The soundness of the requirement set and soundness of the design were two factors that were examined in phases during this study. Talking about the soundness of the requirements, the requirements were tested against survey results and the literature findings. As a result of these checks, the rationality for each item was provided based on the combined results.

Qualitative methods are commonly used for the cases when it is difficult or expensive to run the experiments such as in distributed systems, or systems with big data, when there are conceptual design issues, when the sample data sets are not adequate to show all aspects of the design, and when the number of evaluators is low (Seaman, 1999). The soundness of the design and selected technologies are checked through the use of a series of semi-structured interviews with the experts. These interviews also included questions to check the ability of design to fulfill the initial requirement set. These four participants include one faculty member who have long-term experience and position in the information systems field with particular focuses on software architecture, computer networks, and internet of things, a second faculty member with long-term experience and position in information systems particular focus on technologies, design patterns, and software testing, one senior manager who has 30+ years of experience in the IT, 20+ years of experience in information security and risk management supported with certificates (CISA, CISM, CGEIT, CRISC- ISACA Chapter Founder), and finally, one current chief researcher in public research center, past Microsoft engineer who have 20+ years of experience in software development, software project management, and 4+ years in information security.

During these interviews, IEEE 1471-2000 (Hilliard, 2000) standard is used which is a recommended practice for Architectural Description of Software-Intensive Systems. This standard requires that the system should be introduced systematically by means of a system definition, including environment description, mission and stakeholders identification, and architecture descriptions. These architecture descriptions include a series of architectural views and model definitions each having viewpoints, and concerns. In order to demonstrate the proposed system, a presentation based on IEEE 1471-2000 is made prior to each interview. This presentation included conceptual graphs, top-level

architectural views, class diagrams, detailed views for critical parts, data structures, definitions, scenarios, user interface screenshots, demonstrative information for the available display type libraries and their integration to the proposed system, and code parts to describe various interoperability or algorithmic details. While other standard information is presented in the slides, the rationality of each presented item is explained verbally to the reviewer. Adequateness of the development infrastructure is a significant issue that is questioned for cyber-physical systems, since wrong development infrastructure will result with unscalable solutions with low performances or bad security. Numerical scores given related to each requirement item are shown in Table 13. Other information is in Table 15.

One more issue, which is verified for some cyber-physical systems is the compactness of the overall design. Compactness of the proposed system can be specified based on the compactness of the model-view-controller-with service layer architecture, compactness of the Java Spring framework, and the compactness of Hadoop sandbox. Although the reviewers made no explicit evaluation for compactness, the author can claim that depending on the industry standard technologies, and the architectures will add on to the level of compactness of the proposed design.

Reliability of CPS systems is another significant issue. Reliability of a CPS will depend on the reliability of the system structure and the reliability of the underlying network. Reliability of the system is contingent on the reliability of the selected technologies and underlying structure. Reliability of the algorithms was tested for correctness using sample data. No other experiments were done to test the reliability of the overall system due to conceptual parts.

The main focuses were the interoperability of various systems and technologies, and the scalability and the reliability of the overall system. The main limitation of this study was the proposed design being partly conceptual. Thus, it could not provide quantitative outputs to compare the overall performance of the final design. Certification is a way which helps in the judgment of a design's adequateness, safety and, reliability in a specified environment. CPS systems may be subjected to legal assurance and the certification in real life, commonly before the production phase, which does not apply for this specific case. How this CPS system can be certified and, what type of certificate is more proper, is left as future work. Another future work detected is improving the structure of free text feedbacks from the users. This improvement may allow further ways of automatic processing other than automatic reporting of these feedbacks. The reviewers also mentioned the shortcomings of the feedback system, and suggested to create a taxonomy of probable feedback topics, and improve the free text feedback structure for proper automatic processing in the future.

In general, although a systematic introduction of the study objective and design is made based on a recommended standard, and the review material included a considerable amount of details, there had been times when the reviewers had difficulties in

handicaps. A secondary integration item may be included. Some of the validation questions are vague.

Reviewer 4: Former Microsoft Engineer, Security Researcher Review Duration: Two hours

Predefined metrics such as those synthesized from CVE databases can be incorporated. Threat definition structures may be improved. With real-time visualization, scalability might be an issue for high –volumes of data.

4.8 Concluding Remarks for The Design and Development of a Generic Enterprise Security Visualization Prototype

In this study, an enterprise security visualization system which targets generic processing of log data files, non-predefined metrics, and a knowledgebase design has been presented. The first contribution is its unique scope identified using requirement analysis survey. The second contribution is its generic and standardized design which allows adaptation and extension of new files and new display types. The third contribution is the methodology and design adopting a web-based application to Hadoop big data technologies which may also be exemplary for the integration of legacy applications with big data technologies. It is believed that the final version of the design results in a more scalable system regarding raw data storage place, and corresponding processed data stores, and higher performance due to low latency. The big data technology integrations included easy to implement changes which resulted in extensive benefits. The data collection mechanism, which is included in the second and third design shifts the initial design to a better spot and provides better usability. Finally, enabling integration with external visualization systems increases the overall quality and usability of the proposed system by making simple changes in the original design.

The final design has an easily implementable structure with enhanced qualities mainly in terms of performance, scalability, interoperability, and security due to the design structure and the underlying technologies. This design also has good abstraction and multiplicity features due to generic and/or standardized definitions. Including a knowledgebase in the enterprise security visualization system may help learning and may allow creation of better visualizations easier.

The proposed design may likewise be expanded effectively with future work. A mechanism for automated processing of the user feedbacks given for the visualizations can be formed. Taxonomy of these feedbacks can be done as a part of the visualization knowledgebase. It is constantly conceivable to add new kind of parsers or adapters to support other file types. The current system is primarily intended for JavaScript-based visualizations. It might be tried with other visualization libraries which are applicable to

web-based applications. Other, significant limitation is the allowed file types for the generic parsers. Current data structure did not allow to move all the data parts to NO-SQL database. A more detailed data denormalization study may allow further usage of the NO-SQL database.

CHAPTER 5

APPLICATION SECURITY VISUALIZATION

5.1 Introduction to Application Security Visualization Study

The number of web-based applications is increasing each year. In authors' best knowledge, there are no statistics for the number of existing web-based applications in the world, but as of first quarter of 2017, the number of domain names was around 330 million. From one point of view, each of these domains might be considered as a web application, either static or dynamic.

Risk control and risk assessment are constant challenges in the software project management domain. Demir (Demir, 2009) surveyed on project management challenges with 78 participants. The results show that approximately one of every four projects has problems in the security and risk control area. Web-based applications have been mainstream in the enterprises. Many of the web applications are integrated with each other and with other enterprise systems. Web-based applications are also prone to continuous update due to continuous change in the requirements and new functionalities added. On the other hand, these applications are commonly developed in an ad-hoc manner, without properly understanding the reliability and security requirements (Murugesan, 2008).

The reason for focusing on web-based applications in this visualization study is the increasing amount of attacks and vulnerabilities for them and the small number of visualization studies on this subject. Figure 39 shows the trend of the Open Web Application Security Project (OWASP) (OWASP, 2018) the top ten vulnerabilities between 2014 and 2017.

Fortunately, the security analyses and protection techniques are also improving. Doing only manual analyses and traditional tests are not sufficient. Thus, there had been many tools to make automated analyses for security checks. Some of the tools are prepared to make white-box analyses. These are called static code analysis tools. These tools use application code and configuration files as the analysis source and are more suitable to be used during the development phase. These tools aim analysis of application code, resources and configuration files to find code smells, bugs, and vulnerabilities for continuous code quality without the necessity of code execution. Examples of this group

of vulnerability analysis tools are SonarQube (Campbell & Papapetrou, 2013) and Parasoft (Parasoft, 2018). In order to make tool based static code analysis, the tool which suit the development language and development framework have to be selected. Thus, in order to make a white-box test, the technology used in the development of the web application should be depended. The number of development languages, platforms, and technologies is very high. Thus, being dependent on the web application technologies and languages is a limitation for a security analyst.

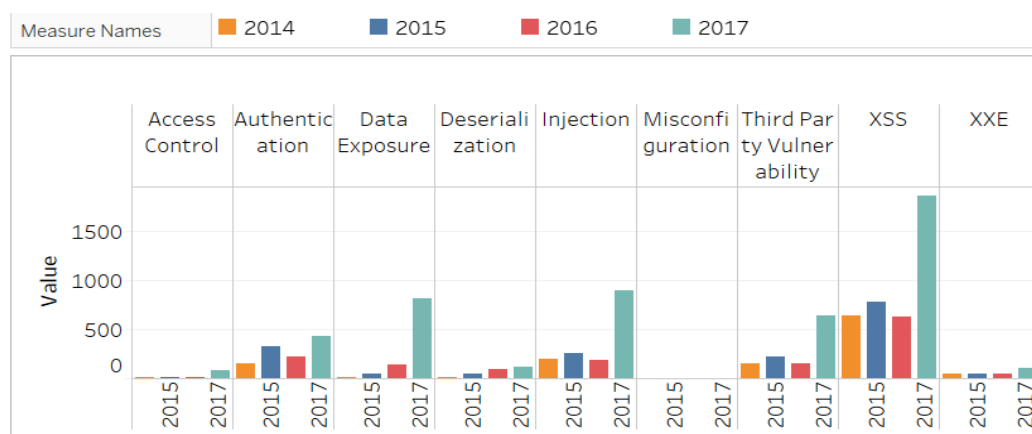


Figure 39 - The state of web application vulnerabilities between 2014 and 2017 (Imperva, 2018)

The second group of vulnerability analysis tools focuses on black-box tests/analyses, and they do not require a dependency on the selected technologies. They use standard HTTP requests to make controls and attacks on the web applications and are more suitable to be used after the deployment of the application either to test servers or the production environment. The tools in this group are also useful during the whole lifetime of the application and may help while taking the security-related design decisions after the first deployment.

In the security visualization domain, security-related data such as log files, network traffic data, operating system data, data from security protection systems such as firewall, IDS systems or vulnerability scanners are visualized to provide more efficient and effective ways of security analyses. The focus of this study is to visualize the outputs of the second group of automated security analyses tools, black-box tests, which are typically the scan results, and the identified alerts. Web application security black-box test tools are called vulnerability scanner tools, in general. The types of vulnerability scanner tools vary. Port scanner tools, web server scanner tools, and web application scanner tools are the most well-known types. The scans may be host based or network based. There are both commercial and open source alternatives. Well-known vulnerability scanners tools include Acunetix (Acunetix, 2019), Netsparker (Netsparker | Web Application Security Scanner, 2019), Retina (OWASP, 2018), Whitehat Sentinel (OWASP, 2018), Burp Suite (Portswigger, 2019), Grendel Scan (OWASP, 2018),

Grabber (OWASP, 2018), Nikto (OWASP, 2018), and Zed Attack Proxy (Romania, 2018) (ZAP)..

This black-box vulnerability scanner tools have various usability problems. Although, they have reporting systems, these reporting systems are not adequate to monitor the vulnerability status of software projects. When the focus is more than one project, then usability problems increase. More information related to existing reporting and visualization capabilities of these tools are provided in the next section.

The primary contribution of this chapter is a new dashboard tool for visualizing vulnerability scan results coming from black-box vulnerability scan tools. In order to achieve this, data attributes of these type of tools are examined and a data structure is formed which includes the attributes commonly found in the vulnerability scan results. A secondary contribution of this work is the list of metrics/measures that the tool presents. The chapter also describes a validation study in which the participants attended a user quiz using the tool prototype. The proposed system has similarities to SIEM systems. In order to clarify these similarities and its unique features an evaluation of the SIEM systems was made. While summary of this evaluation results is provided in this chapter, details are located in Chapter 6.

The proposed solution allows for dynamically inspecting and comparing the characteristics of vulnerabilities on multiple software projects or different versions of the same software project. It aims to allow having a quick understanding of vulnerability levels, types, association of these vulnerabilities to the standards, and the trend in security related development and security related bug fixes for software projects.

The rest of the chapter is organized as follows. First, the related work is described followed with the detailed explanation of holistic web application security vulnerabilities visualization approach. Next, a case study which is developed using the proposed model including the validation efforts is demonstrated. The chapter continues with a discussion of theoretical and managerial implications and research directions. Finally, there is the concluding remarks section.

5.2 Related Work

When the literature is examined in detail, it will be seen that web application visualization studies form a tiny part of the overall security visualization studies (Özdemir Sönmez & Günel, Security Visualization Extended Review Issues, Classifications, Validation Methods, Trends, Extensions, 2018). Dang and Dang (2014) proposed a web application security model to be used for security evaluators. Alsaleh et al. (2015) proposed a study which visualizes the security logs of PHP based web applications. Attacks made to web applications and web-based attack scenarios are also visualization subjects related to web applications. Dang and Dang (2014) in another work, proposed a system which visualizes web attack scenarios. This system is based on

exploiting the links of web application pages and aims to understand intrusion detections. A PHP based web application study was made by Alsaleh et al. (Alsaleh, Alarifi, Alqahtani, & Al-Salman, 2015). This application visualizes web application logs to help security analysts. HVIZ (Gugelmann, Gasser, Ager, & Lenders, DFRWS 2015 Europe, 2015) is a system which does not directly aim web application security, but it visualizes web browser activities. This design might be used for evidence gathering when an incident occurs.

There are also some web application security related studies which do not directly or only visualize security-related data such as security logs, vulnerability scan results or client access logs. However, they may help security analysts in various ways. For example, Dang and Dang (2014) used website hierarchical structure along with website vulnerability scan results. Other studies which deal with application related vulnerability scan results are Goodall et al. (Goodall, Radwan, & Halseth, VizSec '10, 2010) which visualized application code vulnerability scan results and Harrison et al. (2012) which visualized vulnerability scanner results on NV, Nessus Vulnerability Visualization for the Web.

The number of security visualization studies which focus on web application vulnerability scan results is low. Cesar is a prototype proposed by Assai et al. (Assal, Chiasson, & Biddle, 2016) which aims to promote the usability of static code analyzer tools by increasing the collaboration among software developers. Static code analyzer tools are not specific for web applications but all kind of softwares. The authors of Cesar claim that contemporary static code analyzer tools do not provide enough collaboration and for this reason software developers are reluctant to use them. In authors' opinion, collaboration is not a feature which should be primarily expected from a vulnerability scanner tool. Definitely, any property which would enhance collaboration would be valuable for any tool. When we look at software development domain we see that there are already many tools which provide collaboration such as task management tools (e.g. JIRA). Best practice development environments includes the use of vulnerability scanners as a part of a more collaborative tool which supports continuous development and integration process. This way the outputs of vulnerability scanners can be shared among all the project team members continuously.

When Cesar is examined intensely, it can be seen that it focuses a relatively different topic, vulnerability scan results from static code analyzer tools. These results are visualized using the treemap visualization technique. Cesar also provides a way to jump to the code from the vulnerability result.

Dang and Dang's (Dang & Dang, 2014) study is a single example forming the group of visualization studies focusing web application vulnerabilities. This study aggregates data from multiple scanners and provides statistical information for each web page, URL based on the alerts gathered from these scanner tools. There are usability issues for developers using re the static code analyzer tools (Johnson, Song, Murphy-Hill, &

Bowdidge, 2013). The reporting systems of black-box vulnerability scanners are also not very convenient for various reasons (Bingham, Skillen, & Somayaji, 2014). The number of abstraction levels, the need for more fine-grained abstraction, the difficulty of exporting scan results, the lack of definitions for vulnerability types, the lack of boundaries between different classes are commonly reported usability issues. It is thought that the problems for developers occur due to the difficulty of information transfer from security domain to the software development domain by the authors. The reporting systems of black-box vulnerability scanner tools, commonly, serve the detailed scan results data and aggregated data based on alert types. They do not offer any other calculation or metrics, and they do not offer a way to make a transition between aggregated data and the detailed data.

The focus of this chapter is related to the presentation of data rather than scan data generation. In order to understand the presentation features of the existing black box web application vulnerability scanners, part of the popular tools are examined in detail. Blackbox vulnerability scanners may allow you to choose groups of vulnerabilities to scan such as in Acunetix (Acunetix, 2019), or they may allow limiting the scanned domain URLs by the use of some techniques such as Regex mechanism as in Netsparker (Netsparker | Web Application Security Scanner, 2019). They may allow queuing multiple scans, such as in Burp Suite (Portswigger, 2019), or scheduling scans for the future as in Acunetix (Acunetix, 2019). In general, commercial black box vulnerability scan tools are more professional with a higher number of metrics and reports. Open source and free tools are simpler, mainly serving the scan data and the alerts. There are also some commercial tools which have community versions with fewer features.

The metrics/measures commonly presented by these tools are scan duration, number of requests, avg. response time, the number of locations, latest alerts, list of discovered hosts, list of vulnerabilities including URL, type and parameters, detail of the selected vulnerability including the vulnerability description, attack details, and HTTP request detail. Standard metrics are the number of vulnerabilities by severity (such as high, medium, low, or informational) and the number of vulnerabilities by alert/vulnerability type as in Acunetix (Acunetix, 2019) which is one of the higher level tools of this type. Netsparker (Netsparker | Web Application Security Scanner, 2019) presents scan results/scan logs, URL, scan duration, attack type, number of total requests, number of requests per second (speed), and average speed, number of failed requests, total time elapsed, head requests, alerts found and vulnerability description. Some tools such as Burp Suite (Portswigger, 2019) and Zed Attack Proxy (Zaproxy, 2019) have simpler presentation styles with less number of metrics and measures. Burp Suite shows host, method, URL, params, status, length, MIME, title, IP, cookies, details of HTTP requests. In the alerts list (issue activity), status, issue type, URL, (host and path), and issue time is presented. Zed Attack Proxy (Zaproxy, 2019) has similar output fields. These tools with simpler presentation designs also involve alert/vulnerability descriptions either short as in ZAP (Zaproxy, 2019) or long as in Burp Suite (Portswigger, 2019).

While most of the data is presented through the tool screen views, few of the more advanced tools provide one or more type of reports, such as executive summary report as in Acunetix (Acunetix, 2019), developer, auditor, and administrator reports including OWASP top ten report, PCI compliance report, and knowledgebase report as in Netsparker (Netsparker | Web Application Security Scanner, 2019).

Most of the web application scanner tools provide raw alert and scan data. A minimal number of metrics/measures related to the comparison of subsequent scans are included in the provided reports and screen views. Among the examined tools, it is seen that few tools allow manually marking the detected issues as “resolved”. This helps to track the status of security-related updates using the vulnerability scanner interfaces to some extent. Automatic comparison of multiple scans together with the integration of application and project data may provide better monitoring of security updates while examining current security issues for the web projects.

Although coloring and small icons for different levels of vulnerabilities are generally used both at vulnerability screen views and in the generated reports, nearly no other visual element exists as an output in the web application security vulnerability scanners. Only a few of the tools allow filtering of the output data, as in the drill down feature of the Acunetix (Acunetix, 2019), which allows filtering based on severity, target, business criticality, status, and CVSS values.

The low number of studies focusing web application vulnerability visualization and the usability problems of vulnerability scanner tools provide challenges to the researchers. These challenges include definition or identification of new metrics and providing new visualization designs which will enable monitoring of these newly identified metrics, enable monitoring of changes between multiple security checks and effects of new developments and bug fixes.

Moreover, associating the alert data to the standards will be useful to show at what level the analyzed web application is compatible with the security standards. The Dang and Dang’s study focuses on statistical metrics. It does not provide a view which shows repeated alerts, fixed alerts between phases, and it does not relate the alert data to the standards. Software project managers responsibilities include review the current status, and progress against intermediate and final development targets and to identify the obstacles. Managers also should monitor if the security implementation meets the standards and technical requirements. Including repeated alerts, and fixed alerts between periods the may be valuable for the managers besides security analysts, and would end up with a new visualization perspective for the same type of data.

Existing web application vulnerability visualization projects does not have a holistic approach which combines vulnerability scanner results with the environment properties and timeline of activities which affect web application security. Although some of the visualization studies have clear metric definitions, the majority of the web application

vulnerability studies also lack clear web application security metric definitions. Lacking metric definitions and having an inadequate number of metrics result in an incomplete picture of security statuses of web applications.

Against this background, the purpose of this study is to propose an alternative visualization tool which visualizes the data attributes commonly available for web applications, combine these data attributes with common outputs of web application vulnerability scan results, namely, scanner results and alerts, and to find out measures and metrics based on the proposed data structure.

5.3 Holistic Web Application Security Vulnerability Visualization, HWAS-V

In general, the users of web application vulnerability scanners are expected to have technical competency having positions such as security analysts, and system admins. In order to provide a system which gives a broader perspective of the security statuses, a new approach is introduced in this study. In this approach, the project level details, such as earlier security analysis results, application level details, such as size of application, number of -modules, the number of external libraries and the standards related information are integrated with the vulnerability scan results. The proposed security visualization solution is expected to be valuable for the software project managers as well as the security analysts. Thus, as a first step, a data structure is formed as an outcome of examining the typical results of vulnerability scanners and selected secondary data sources.

An essential part of this study is defining quantitative metrics. Following the data source structure formation, a large set of metrics/measures have been identified using the proposed data structure. Encapsulating a large set of metrics is not very common in security visualization solutions. Similarly, explicitly marking each metric in the security visualization solutions is also not accomplished yet. Although the statistical measures are distinctly pointed out in security visualization solutions, other measures are left to the users' understanding.

Prototyping is the most widely taken approach in the security visualization domain to illustrate the novel visualization designs. For this purpose, to illustrate the proposed approach, a visualization prototype tool has been developed in dashboard form, called Holistic Web Application Security Vulnerability Visualization, HWAS-V. Dashboard style is selected due to its ability to encapsulate numerous metrics in one design easily. During the preparation of the dashboard, several combinations of the metrics are built experimentally, and among them, most appropriate combinations are selected to form a legitimate dashboard design.

To investigate the similarities and differences of SIEM tools and the proposed tool six SIEM tools (Manage Engine Event Log Analyzer (ManageEngine, 2018), Splunk (Splunk, 2013), Rapid7 InsightIDR (Rapid7, 2018), Solar Winds Log and Event Manager (Solarwinds, 2018), Micro Focus ArcSight (MicroFocus, 2018), AlienVault (AlienVault, 2018)) located in four quadrants of the Gartner analysis were installed on a test machine as an extension to this study. Findings of this investigation is depicted in the Discussion section.

In the following section, the metric definition process is described. The primary and secondary data sources are also explained in this section. Later, the offered metrics are depicted in a classified manner. In the final part of this section, the visualization process is revealed in detail.

5.3.1 Metric Definition Process and Data Measures

Vulnerability scanners may have a variety of different focuses, still they have similar working mechanisms. These tools make scans based on defined rules, i.e., “scanner rules”. During the scans, they generate the “scan data” which include requests and responses and the resulting “alert”s. It is difficult to form a generic data structure which supports all web application scanners, because each scanner would have its own attributes and data types. Thus, during the design of the data structure, the selection of mandatory data attributes has been made by using a minimum set which commonly exists in the web application vulnerability scanner results. There are also some optional data attributes. These attributes take part in relating the minimum data set to some other data. For example, `CWE_id` in alerts and `scan_rules` relates alerts and scan rules to some existing standards information. The prototype is designs so that, if there is no association information for the selected automated vulnerability scanner tool, this does not affect the overall visualization system. If there is a known relationship, this relationship is used by the visualization system to provide some additional metrics.

The scan rules defined in a scan tool is the first primary data source used for visualization purposes in this study. These rule definitions may include a “rule name”, and a “rule id”. There may be some additional categorical measures such as a “version” information. Optionally, there may be some information which relates “rule” to the security standards. Although this type of data does not change frequently, it provides information related to the coverage efficiency (scope) of the vulnerability scans, and be used to relate scan outputs to the common standards, such as Open Web Application Security Project (OWASP), Common Weakness Enumeration (CWE), and the Web Application Security Consortium (WASC).

At the start of a vulnerability scanning activity, a base URL is needed. Once the base URL is identified, automated scanner tools check for all available pages in that domain, thereby forming a list of all available pages for that web application. Forming such a list is called spidering or crawling in the web terminology. The result of spidering is called

“scan results” in this study. These results include information related to scanning the base URL and related pages which are found during the crawling. The scan results may include information such as “scan id”, “process result”, “request timestamp”, “method”, “URL”, “response code”, “reason”, “RTT”, “request header size”, request body size”, “response header size”, “response body size”, “highest alert”, and “tags”. Tags are optional, and they may not be available for all scanner types. However, similar to the relationship to standards, the existence of this information may provide some additional metrics. The second primary data source of this study is the “scan results”.

Once a URL is provided to the vulnerability scanner tool, and all the pages are crawled, the tool filters the pages which do not belong to the target domain. Later, it checks the results of applying the scanner rule for each page. The checking mechanism may depend on passive controls or active attacks. In the end, it provides a list of alerts associated with a list of instances where each instance correspond to a URL. These results include the “alert name”s, and, “URL”’s of the related pages. Alert names may be equal or similar to the corresponding “scan rule name.” Unfortunately, there is no standard for naming the scan rules and the corresponding alerts. Mapping of scan rules to alerts can be made, which is done only once, by the tool users to be benefited during the subsequent scans. The alerts list is the third primary data source of this study.

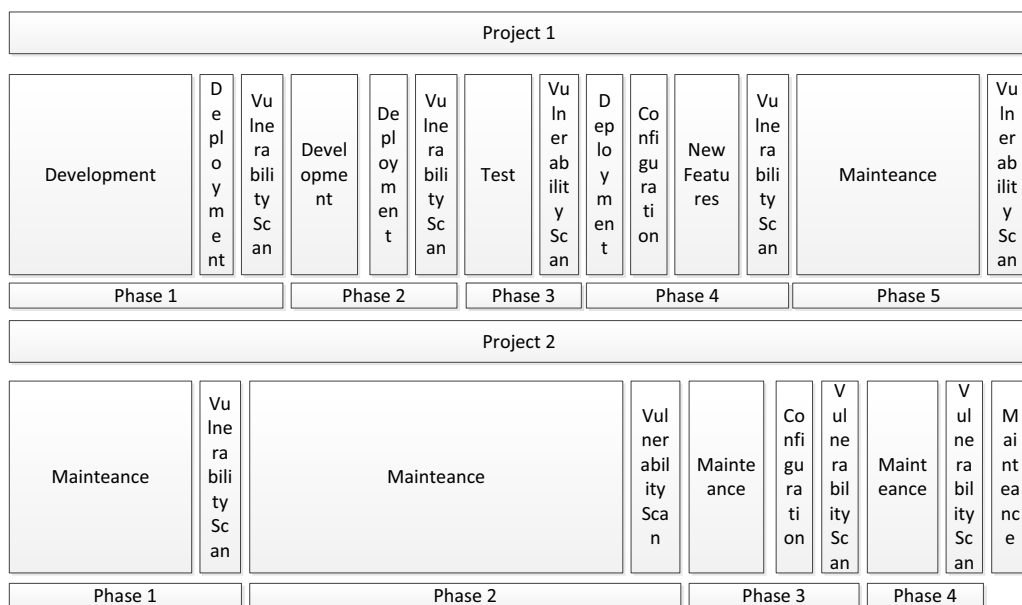


Figure 40 - Multi-project multi-phase vulnerability scan results

During the design of visualization prototype, first, the available data attributes have been examined. Following this, a data structure has been formed. In order to form the measures/metrics list, earlier academic or commercial published material which points to the web application security metrics have been examined. Later, this initial set of metrics has been enlarged by including the measures and proposed metrics which may

be generated using the proposed data structure. Thus, a few of the proposed metrics were mainly designed, because they were convenient to measure using the available data. However, since the aim is not to find a solution to solve all problems of security analysts, but to improve the ways of examining the web application vulnerability data. Knowing this, and the earlier complaints of lacking enough measures and abstractions to examine the vulnerability data, it can be claimed that proposed metrics and measures were most suitable for the security analysts providing various abstractions of vulnerability data in different levels.

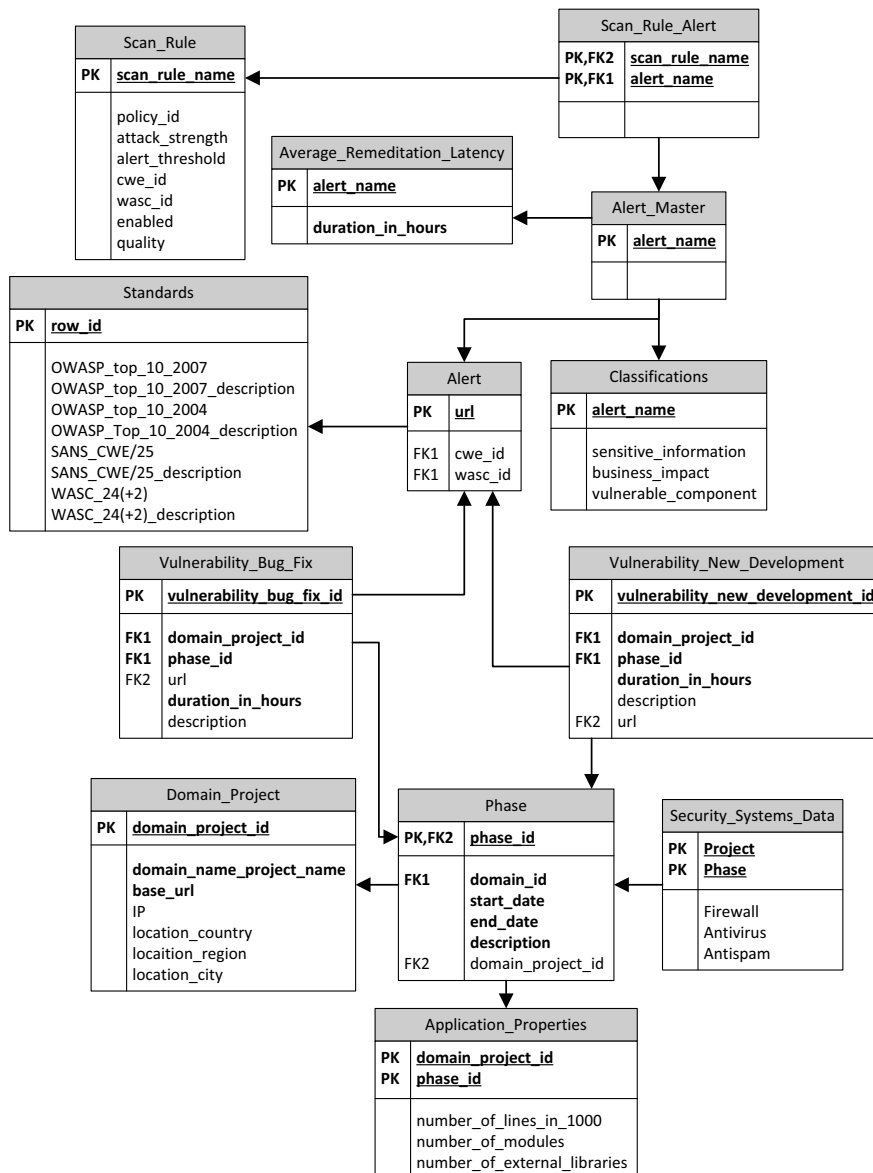


Figure 41 - Data structure and attributes of the proposed model

Before examining the proposed measures and metrics, it is necessary to define the “project/domain” and “phase” terminology used in this study. “Project/domain” corresponds to either a web application which is in the development state and subject to the security tests, a project, or an already deployed application which is the subject of a security analysis/test task, or a domain. Since such security analyses/tests are repeated actions, a “phase” definition was required to measure the effects of repeated vulnerability scanner results and the changes made after each security analysis. Hence, whenever a new automated test is recorded for a project/domain, a new phase is started for a project. These phases do not necessarily correspond to the phases in a project life cycle. This phase concept was necessary to enable measures related to changes due to vulnerability fixes and bug fixes or new developments in a defined phase in the proposed framework. Figure 40 shows a series of vulnerability scan results achieved for two independent projects. Throughout the lifetime of project 1, the vulnerability scanner is executed five times. For project 2, the vulnerability scanner is executed only in the maintenance phase, four different times.

During the metric design phase, application properties such as application size, the number of modules, and the number of external libraries are used in conjunction with vulnerability scan results. If the development continues for the application, then this information will change from phase to phase. Data coming from security protection systems are another type of secondary measures used in the proposed framework. These measures are combined with security standards related information to form a holistic data structure which enables monitoring security statuses of the web applications for multi-project, multi-phase platform combined with an association to the standards and other information related to the project structure and the environment. In summary, measures related to the application properties, text/analyses phases, security protection systems, and standards are included as secondary data sources to this study. Figure 41 shows the data structure of the proposed model.

Table 16 - Base Measures/Metrics Based on Vulnerability Scanner Tools

	Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
1	Alerts(t_n)	Alerts/Vulnerabilities $t=t_n$, {a1, a2, ax}	Measure	Figure 50
2	VT(t_n)	# of vulnerabilities $t=t_n$, Alerts(t_n)	Metric	Figure 50
3	URLsProcessedSet(t_n)	{ url1, url 2, ..urlp}	Measure	Figure 49
4	URLsProcessed(t_n)	# of URLs processed at $t=t_n$, URLsProcessedSet(t_n)	Metric	Figure 48
5	URLsScannedSet(t_n)	{ url1, url 2, ..urls}	Measure	Figure 49
6	URLsScanned(t_n)	# of URLs scanned(t_n), URLsScannedSet(t_n)	Metric	Figure 48
7	URLsWithAlertSet(t_n)	{ url1, url 2, ..urla}	Measure	Figure 49
8	URLsWithAlert(t_n)	# of URLsWithAlert(t_n)	Metric	Figure 49

9	RptdAlerts	Repeated Alerts, \cap URLsWithAlert(t_n) URLsWithAlert(t_{n-1})	Metric	Figure 48
10	NVulnerabilityLevelChange	Change in Vulnerability Level, $VT(t_n) - VT(t_{n-1})$	Metric	Figure 48
11	PagesWithFixedVulnerabilities	Pages whose vulnerabilities are fixed between phases, $URLsWithAlert(t_{n-1}) \cap URLsWithAlert(t_n)$	Metric	-
12	NVulnerabilityLevelChangePerPage	# of Pages whose vulnerabilities are fixed between phases, $ URLsWithAlertSet(t_{n-1}) - URLsWithAlertSet(t_n) $	Metric	Figure 48
13	NNonProcessedPages(t_n)	Number of NonProcessedPages(t_n), $URLsScanned(t_n) - URLsWithAlert(t_n)$	Metric	Figure 50
14	PctScannedPages(t_n)	Percentage of ScannedPages(t_n), $URLsProcessed(t_n) / URLsScanned(t_n) * 100$	Metric	-
15	NNonAlertedPages(t_n)	# of Non-AlertedPages, $URLsScanned(t_n) - URLsWithAlert(t_n)$	Metric	-
16	PctAlertedPages(t_n)	Percentage of AlertedPages(t_n), $URLsAlerted(t_n) / URLsScanned(t_n) * 100$	Metric	-
17	TotalRTT	Total RTT in One Scan, $\sum_{i=0}^{URLsScanned(t_n)} RTT$	Metric	Figure 48
18	AvgRTT	Average RTT for One Page, $TotalRTT / URLsScanned(t_n)$	Metric	Figure 48

The proposed measures/metrics are the results of primary vulnerability scan results and related factors such as information related new developments, bug fixes, maintenance effect, time effect, classifications, application properties, protection systems, and the standards. They will be presented using this classification for better understandability.

5.3.2 Web Application Visualization Measures/Metrics Based on Common Vulnerability Scan Outputs and Related Data

Measure and metric are two terms which sometimes are used interchangeably in some contexts. The main difference between them is, the measure is the direct result of a measurement activity. Metric, on the other hand, is the result of a calculation made using one or more measures. In this study, these terms are used complying these definitions. The measures and metrics that are available using the proposed data structure and sources mentioned in the previous section are described in this section. While simple

measures/metrics are listed by name in a grouped manner, textual definitions of more complex metrics are provided individually.

1) *Base Measures/Metrics - Measures Based on Vulnerability Scanner Tools*

The measures in this group are originated from the web application security vulnerability scanner tools. In this study, these measures are called base measures because they are based on measurements from the primary data source, i.e., the web application vulnerability scanner results. In the subsequent sections, this base list is enlarged by the potential effects of secondary data sources. The measures of: alerts set, set of URLs scanned through vulnerability scanner, set of URLs processed, set of URLs associated with an alert, round trip time (RTT) for an HTTP request and response for each alert check are provided by the scanner tools.. The metrics of: number of vulnerabilities/alerts, number of URLs scanned, number of URLs processed, number of URLs with an alert, repeated alerts, change in vulnerability level, pages with fixed vulnerabilities, number of non-processed pages, number of non-alerted pages, percentage of processed pages, percentage of alerted pages, total RTT, and average RTT are the results of using simple arithmetical or set operations on the measures listed in this group. Definitions of metrics which have less straightforward calculations or which are difficult to understand are using the “Definition” keyword for each group of metrics/measures. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 16.

Definition 1: Repeated alerts is the set of alerts for a project which is detected in a scan and a subsequent previous scan.

Definition 2: Change in the vulnerability level is the difference of the number of alerts detected in one scan session and in a subsequent scan session independent of the scanned URLs set and the detected alerts in each session.

Definition 3: Pages with fixed vulnerabilities is the result of set subtraction operation from the alerted pages detected in one session and a subsequent session.

Definition 4: Non-processed pages in one scan are equal to URLs which are detected in one scan, yet, which are not further processed for alert checks due to either permissions or similar reasons, calculated through the use of set subtraction.

Definition 5: Non alerted pages in a scan is equal to the set of URLs which are scanned, processed and resulted in no alerts, calculated with set operations.

Definition 6: Percentage of processed pages is the fraction of processed pages over the total number of scanned pages in one scan divided by 100.

Definition 7: Percentage of alerted pages is the fraction of alerted pages over the total number of processed pages in one scan divided by 100.

Definition 8: Total RTT is the sum of the durations for all round trips between the web application and scanner made in one scan.

Definition 9: Average RTT is the average of the durations for all round trips between web application and scanner made in one scan calculated by dividing total RTT to the total number of scanned pages in one scan.

Table 17 - Metrics-Measures/New Developments-Bug Fix Maintenance Effect

Given Name	Meaning	Measure/Metric	Prototype Fig.Num.
19 VFixedDueBugFixSet (t_n)	Alerts Fixed Due to Bug Fix $t=t_n$, { a1, a2, ..a _b }	Measure	
20 VFixedDueToNewDevSet (t_n)	Alerts Fixed Due to New Development $t=t_n$, { a1, a2, ..a _d }	Measure	
21 TTFx(t_n)	Time to Fix Vulnerability x at $t=t_n$, Vx(t_n)	Measure	
22 NFixedDueBugFix(t_n-t_{n-1})	# of security related bugs fixed in time period, VFixedDueBugFixSet (t_n)	Metric	
23 NFixedDueNewDevSec(t_n-t_{n-1})	# of security related new developments in time period, VFixedDueBugFixSet (t_n)	Metric	

2) *Metrics/ New Development-Bug Fix-Maintenance Effect*

During the lifecycle of a web application, both new developments and bug fixes might exist in time. In order to monitor the effects of these tasks, these efforts might be associated with the previous findings of the vulnerability scans. Once these associations are made, it is possible to include the following measures and metrics to the proposed system.

The measures set of alerts/vulnerabilities fixed due to bug fix related tasks, set of alerts/vulnerabilities fixed due to new developments and time to fix for each vulnerability can be measured through the use of task management systems properly. Arithmetic operations calculate the metrics of the number of related bugs fixed in a period and the number of security-related new developments. These metrics are used in conjunction with the metrics presented in the base measures/metrics part in the provided design.

The metrics and measures proposed in this part would form an internally developed remediation latency database for the known alert types for that specific application in the long term and may be used for planning purposes. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 17.

Table 18 - Metrics/Measures Based on Effects of Previous Measurements and Time

	Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
24	AvRL	Average Remediation Latency for a Vulnerability	Measure	-

3) *Effects of Previous Measurements and Time*

As time passes, a system/software would eventually undergo some changes. The effects of new developments and bug fixes are already mentioned above. In this part, the internal and external measurements, namely the remediation latency indicator values collected by IT companies in time for common alert types are included in the proposed metric/measure list. Remediation latency is an indicator which measures the security update performance of the developer organization. The IT companies periodically announce such information related to common security issues. Integrating such information with the vulnerability scanner results would be beneficial for both planning purposes of new developments/ bug fixes and for the monitoring purposes of the progress of the ongoing projects. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 18.

Definition 10: Average remediation latency for a vulnerability is either a measure gathered from vendor report or a measure which is ascertained during the tasks described in the metrics/ new development-bug fix-maintenance effects part

Table 19 - Metrics-Measures/Application Properties Effect

	Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
25	CLOC(t_n)	# of code length at time t_n , LOC	Measure	Figure 50
26	CMdl(t_n)	# of modules at time t_n , modules	Measure	Figure 50
27	NWP(t_n)	# total number of web pages in web app	Measure	-

26	VTPP(t_n)	Vulnerabilities per page at time t_n , VT(t_n)/ NWP(t_n)	Metric	Figure 49
27	CLOCPP(t_n)	Vulnerabilities per line of codes at time t_n , VT(t_n)/ CLOC(t_n)	Metric	Figure 50
28	MdIPP(t_n)	Vulnerabilities per modules at time t_n , VT(t_n)/ CMdl(t_n)	Metric	Figure 50
29	PctURLsScanned(t_n)	Percentage of scanned pages, URLsScanned(t_n)*100/ NWP(t_n)	Metric	-
30	BaseURL		Measure	Figure 47
31	BaseURLAlerts(t_n)	# of alerts for base URL at time t_n , VT(t_n) and url = BaseURL	Metric	-

4) Metrics- Measures/Application Properties Effect

The size of the application, the use of internal or external libraries, and integrations made with third-party tools would affect the security status of the web application. The measures and metrics related to this information are also integrated with vulnerability scanner results to provide a more holistic view of the web application security.

The measures related to application properties used in this study are the elements which indicate application size in various ways, such as the numbers of line of codes (LOC), modules, and web pages at time t , and the elements which show information related to application deployment structure, base URL and the geographic deployment location. The latter parameters are used for monitoring the application status in a map. The metrics designed by the authors using these measures are the number of alerts for base URL, the percentage of scanned pages to total pages, the number of vulnerabilities per LOC, vulnerabilities per module, and vulnerabilities per web page. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 19.

Definition 11: Alerts for base URL is calculated through the use of string operations on the scan results. These are the alerts related to base URL regardless of the extension of the URL with the alert.

Definition 12: Percentage of scanned page is the fraction of the scanned pages over the total number of web pages provided in the application properties divided by 100.

Table 20 - Metrics/ Classification Effect

Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
32 VTSIL(t_n)	# of Vulnerabilities/Alerts Related to Immeditate Sensitive Information Lost	Metric	-
33 VTHBI(t_n)	# of Vulnerabilities/Alerts Related to High Business Impact	Metric	-

34	VTVC(t_n)	# of Vulnerabilities/Alerts Related to VulnerableComponents	Metric	-
35	VTH(t_n)	# of High Vulnerabilities/Alerts	Metric	Figure 49
36	VTM(t_n)	# of Medium Vulnerabilities/Alerts	Metric	Figure 49
37	VTL(t_n)	# of Low Vulnerabilities/Alerts	Metric	Figure 49

5) Metrics/Classification Effect

The majority of vulnerability scanner tools provides information which would help categorization of the alerts/vulnerabilities. The most common categorization criteria is the severity/importance level such as high, medium, and low. If the alerts/vulnerabilities can be categorized based on their effect to sensitive information as “sensitive information risk exists”, “sensitive information risk does not exist”, their effect to business “high business impact”, “low business impact” and their origins “a vulnerable component exist”, “a vulnerable component does not exist vulnerability is due to other effects” then this information can be converted to the measures and metrics. Making these categorizations is commonly the responsibility of the scanner developers, because they are the people who know the inner mechanisms and targets of the attacks/scans. The associations of the proposed metrics to the available standards can be used to make these classifications further. However, the scope of this work does not include the classification of the vulnerabilities/alerts based on the proposed classifications. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 20.

Definition 13: Number of vulnerabilities/alerts related to immediate sensitive information forms a group of vulnerabilities which are directly related to loss of any sensitive information. Not all vulnerabilities result in information loss, so the vulnerabilities in this group should be directly associated with the information loss.

Definition 14: Number of vulnerabilities/alerts related to high business impact would change from business to business and may be difficult to detect. The vulnerabilities of this group need not be related to information loss. For example, a DDOS attack which has a business impact, but no information loss would be in this category.

Definition 15: Number of vulnerabilities/alerts related to vulnerable components are found out during an examination of application structure and its possible relations to the alerts. Some alerts may not be directly related to application structure but may be due to external effects. Such vulnerabilities are not counted in this group.

Table 21- Metrics/Standards -Lists Effect

	Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
38	VTOWASP10(t_n)	# of vulnerabilities related to OWASP top ten list at time t_n	Metric	Figure 50
39	VTWASC25(t_n)	# of vulnerabilities related to WASC top twenty five list at time t_n	Metric	Figure 50

6) *Metrics/Standards- Lists Effect*

The majority of the alerts have associations to available security standards, such as OWASP, WASC or CWE. These association values may be used to calculate new metrics as a part of the web application security monitoring dashboard. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 21.

Definition 16: Number of vulnerabilities/alerts related to OWASP top ten list is found out using published associations of scan rules to the OWASP standard.

Definition 17: Number of vulnerabilities/alerts related to WASC top twenty-five list is found out using published associations of scan rules to the WASC standard.

Definition 18: Number of vulnerabilities/alerts covered from the CWE standard is found out using published associations of scan rules to the CWE standard.

Table 22- Metrics/Protection Systems Effect

	Given Name	Meaning	Measure/Metric	Prototype Fig. Num.
40	SecSys(t_n)	# of vulnerabilities prevented/blocked by security protection systems or other layers	Metric	Figure 47
41	RatioVTSecSys(t_n)	Ratio of scanned vulnerabilities to detected ones, $\text{SecSys}(t_n) * 100 / \text{VT}(t_n)$	Metric	Figure 47

7) *Metrics/Protection Systems Effect*

In an environment where continuous monitoring of web application security exists, it is commonplace that there might be other security protection systems. The measured number of vulnerabilities prevented/blocked by external security protection systems and the metric ratio of scanned vulnerabilities to detected ones are also included in the

proposed web application security monitoring system. The associations of the metrics/measures in this group to the dashboard parts are shown in Table 22.

Definition 19: Number of vulnerabilities prevented/blocked by external protection systems is the numerical quantity of data that can be gathered from systems such as firewall, antivirus, antispam, and IDS.

Definition 20: Ratio of scanned vulnerabilities to the detected ones is the fraction of vulnerabilities prevented/blocked by external protection systems to the total number of vulnerabilities detected by scanner systems divided by 100.

5.3.3 *Visualization of Metrics*

The motivation for visualizing the proposed metrics resulted in a dashboard design which integrates automated vulnerability scanner results with other related data sources providing a summary of the vulnerabilities and its relations to application, system, and environment. By this way, the design presents security related highlights and eases the monitoring and tracking of security statuses of one or more projects in multiple phases. The outstanding features of the proposed system are its practicality, its efficiency in analyzing the data, and its decision informing and difference detection capabilities.

As explained along with tool description, in this concept, the “project” refers to the actual web application which is either in development or production state. Sometimes the keyword “domain” is also used to refer to the “project” throughout the text. The phase refers to a duration which ends with a tool based security analysis for a web project. Thus, the duration of the phases will be very variable. If frequent automated security analyses are done for a domain, then there will be more phases. If a new analysis is made and recorded, then a phase is finished. Thus, the analysis end time defines the phase end time, and phase start time is identified by the previous phase end time or the project start time. As a result, the design enables to analyze the repeated alerts, the effects of bug fixes, new developments, and the environmental changes between the phases. The automation “tool” refers to a vulnerability scanner which provides data close to the model of the proposed model specification meaning “providing a list of scan data and alert data associated with URL’s and alert types”.

Tableau desktop software (Murray, 2013) is used for the creation of dashboard type of visualization prototype. The aim of visualization design is to enable visualizing the automated vulnerability scan results as is. Thus, the vulnerability scan results are visualized without any cleansing or modification operation, for quick responses. However, during the data preparation phase, in order to merge multiple data sources and differentiate subsequent executions of the automated scans, some numerical id values are generated. For example, “phase id” is generated to identify each execution of

vulnerability scans and determine the duration passed between each automated test/analyses period. “Project id” is generated to be used in the creation of sets for the dynamic calculation of sets among projects and phases. Some part of the original data consisted of united information which is better to be examined separately. For example, the original URL strings included the method (e.g. Get, Post). These strings are converted to method-URL pairs. Other text operations are done dynamically through the use of text operations of the visualization tool. For example, original URL information appeared in scan results, and alert results included base URLs, additional path information and request parameters. In order to identify the actual paths for some visualization parts, such as visualizing the unique number of alerted pages, the URL parameters are eliminated by text operations.

Although some of the data is used to make associations or classification among other data parts, such as the “Classifications” table and “Standards” table, some dynamic classifications are also made using set operations of the tool over data. The reasons for using set theory were the necessity of grouping the alert data based on multiple fields and the necessity of using set operations such as union, intersection, and set minus. For example, in order to find “repeated alerts”, and “new alerts” in different phases set intersections and set minus operations are used over URL datasets respectively.

New numerical values for some categorical values are created through calculated fields. For example, the “numerical alert level” is created from the categorical alert level attribute. Aggregation is often used for many purposes, such as aggregation of data based on scan rules, projects, and project phases. Besides aggregated data, the proposed visualization system also includes page level data visualizations.

Several dashboards are created which focus on different aspects of the data. These dashboards are aggregated in a “story” which is a feature of the tool that allows easy navigation among multiple dashboards. The dashboards consist of various type of simple 2-D charts.

Using simple 2-D charts, such as a bar chart or a pie chart, makes it easier to comprehend and monitor multiple measures/metrics simultaneously. 2-D charts result in fewer false readings and enable better comparison of items. The overlapping of data points in 3-D charts commonly results in a misunderstanding of patterns. As the number of measures/metrics increase, interpreting these measures/metrics via the 3-D charts or complex charts become even more difficult. Due to the high number of proposed measures/ metrics, this web application vulnerability visualization study uses a dashboard type of design consisting of simple 2-D charts and tables.

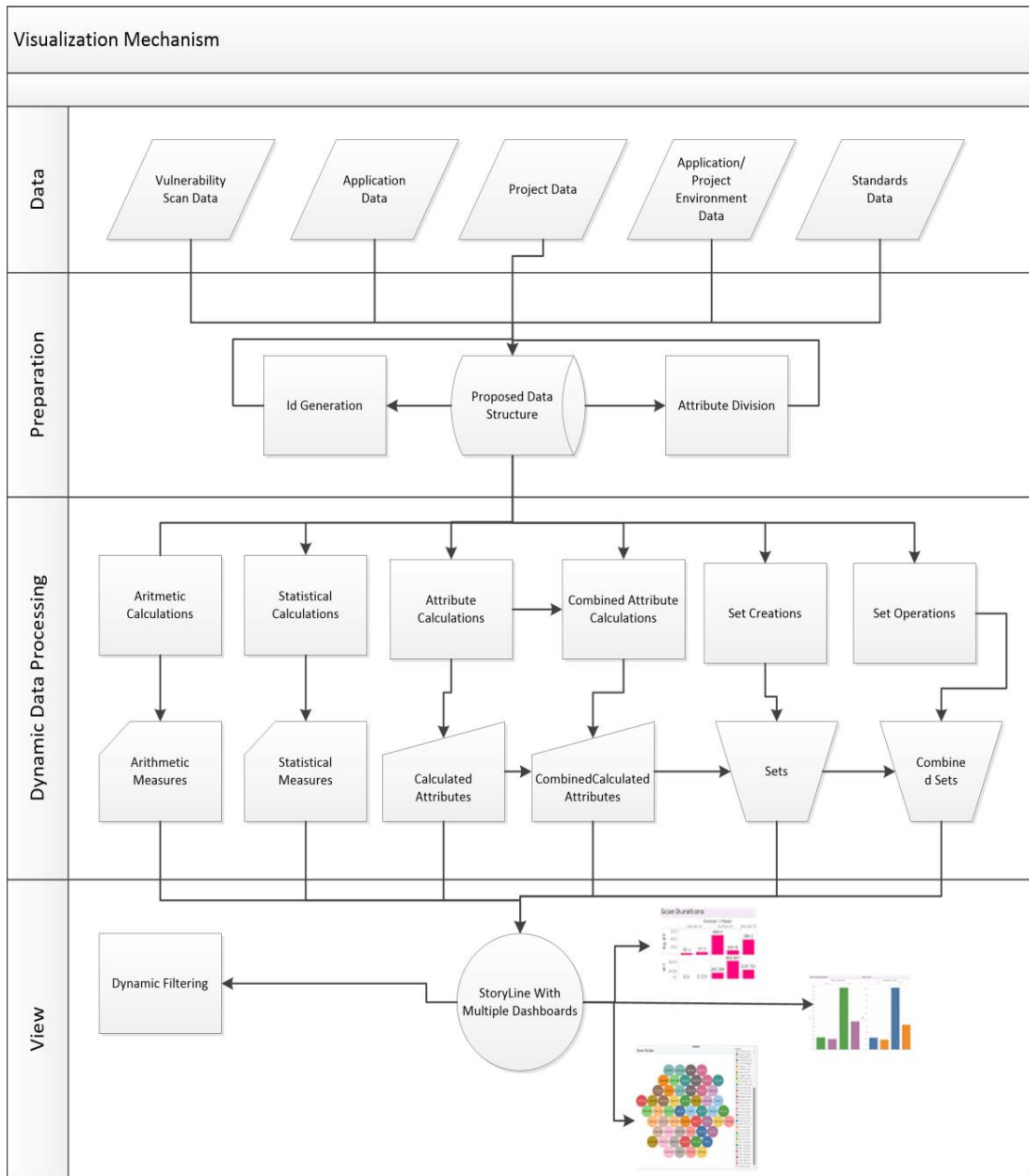


Figure 42 - Visualization mechanism

Some of the visualizations are very straightforward. These may be the results of some arithmetic calculations on a single data source. Results which are single numeric values are often visualized using “Formatted Text”s. Results which include a series of numeric values are visualized using charts, such as “Bar Chart”, and “Pie Chart”. Some other visualizations require joining or blending data from multiple data sources, which were presented using novel visualizations. These visualizations were presented through a case

study in the results section. Tables are used to show some details of data such as URL based detailed information. Bubble charts are used to show some grouping effects such as grouping based on standards. In some charts, both the percentile and actual values are shown. In these charts, the axes are overlapped using the dual axis property of Tableau, to allow simultaneous monitoring of both values.

Encapsulating a large set of metrics in a dashboard design requires using the space effectively. Sometimes a small portion of the view area has to be used to show some metrics. Sometimes, the resulting points shown in a chart may have very close values, causing overlapping of the data points. On several charts, logarithmic scales are preferred to normal scales on data distribution on the axes to overcome this difficulty.

A novel visualization property in the proposed dashboard design is the explicit association of the proposed metrics to the charts. Showing tooltips for charts automatically or on demand is a feature of the design tool. These tooltips are used to show additional information, such as detail data, values of related attributes, etc. for all charts. Besides this information, each named metric/measure information is also shown in a formatted and colored manner via the tooltips. This format makes it easier to understand and interpret the values for the users when navigating through the dashboards.

The resulting visualization system consists of multiple dashboards which are fragmented based on the logical grouping of metrics and ordered based on some logical flow of information. In Tableau, this demonstration form is called a “story”. The story allows navigation among multiple dashboards easily and allows an explicit description of each dashboard. The aggregation of these definitions indeed forms the story itself. Serving the metrics in grouped manner based on several titles would help to have more information and to make a transition between aggregated data and the detailed data without getting lost in details, and reveals the relation of vulnerability data with other data parts in a clear manner. In this way, the user of the tool can examine the relationship of vulnerability data with other data sources in a systematic manner.

Very few of the proposed metrics were excluded from the dashboard prototype based on multiple reasons. The first reason is the lack of corresponding data for the case study. For example, collecting average remediation latencies of known vulnerabilities was out of the scope of this study. Thus, due to the lack of corresponding data, no visualization was included related to remediation latencies in the prototype. Another reason is the limited space available in the proposed dashboard. Although some charts and, tables include such detailed information, some of the measures/metrics which include sets of URLs were not visualized in the prototype on purpose. Such detailed information may be served on demand using other ways such as tooltips connected to blank sheets which do not take much space in a real product.

5.4 Case Study and HWAS-V

OWASP Zed Attack Proxy (OWASP, n.d.) (ZAP) automatic web application vulnerability scanner tool was used to generate the vulnerability scanner data for this case study. OWASP ZAP tool is a proxy application which combines a number of features including spider tool, active scan, passive scan, port scan, rest API, and the reporting functions. For each scan type, rules are defined by the community users (contributors), and independent evaluators evaluate these scan rules, prior to integration with the tool.

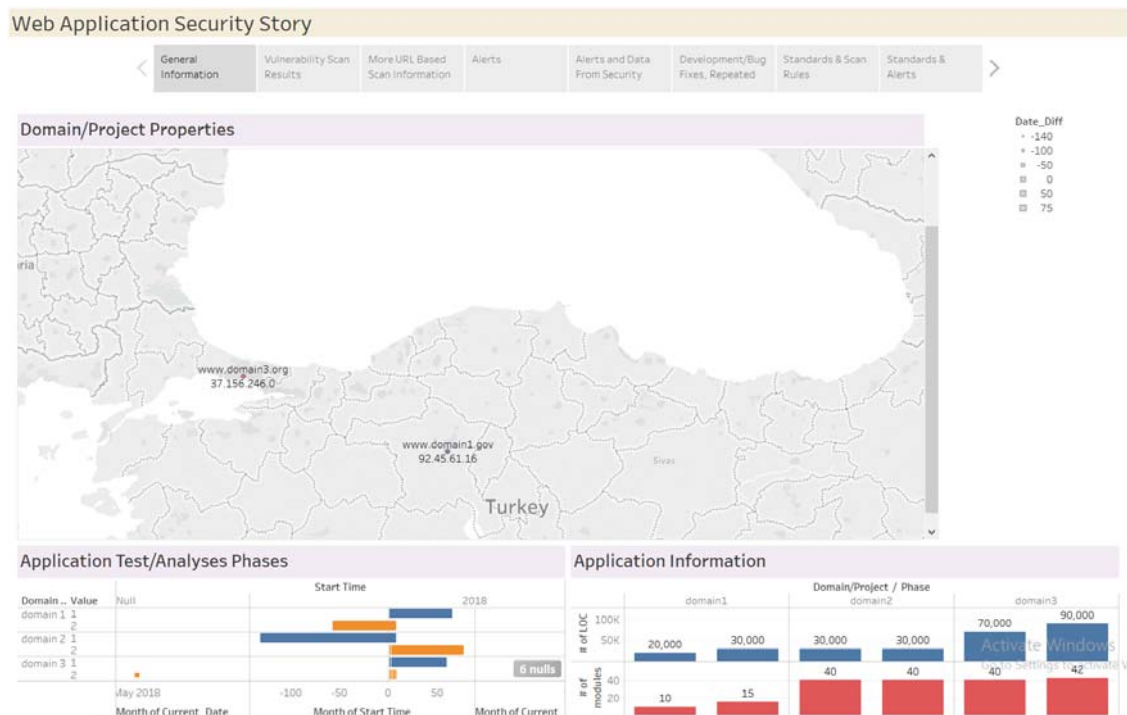


Figure 43 - General information dashboard

OWASP ZAP has various modes; standard mode, safe mode, protected mode, and attack mode. In this study, “attack mode” was selected, because it provides a higher level of information related to the targeted domains. Once a URL is provided to the tool; first, the tool crawls all the URLs in that domain. Later, it filters the URLs which do not belong to the target domain. Then, it attacks to the selected pages. In the end, it provides a list of alerts associated with a list of instances where each instance corresponds to a URL.

This OWASP ZAP attack tool was utilized on three independent domains several times to provide data related to scanning results and the alerts. Later, this initial data was anonymized to some extent and combined with some data related to other aspects of the proposed visualization system to form a mockup dataset for demonstration purposes. The

resulting mock-up dataset includes all the data attributes shown in Figure 41 for three independent domains for two analyses phases.

Figure 43 to Figure 50 illustrate different parts of the proposed dashboard design. Figure 43 shows the first dashboard design which provides a top-level view to three web application's security statuses. The locations where the web pages installed are shown on a map. This location information may be useful for security analysts who monitor security statuses distributed in large regions. Showing the location information may point out hosting place-based problems for some projects. In this view, besides location, IP, base URL, information related to web application size, and the earlier testing efforts shown on a Gantt chart are also included for each project.

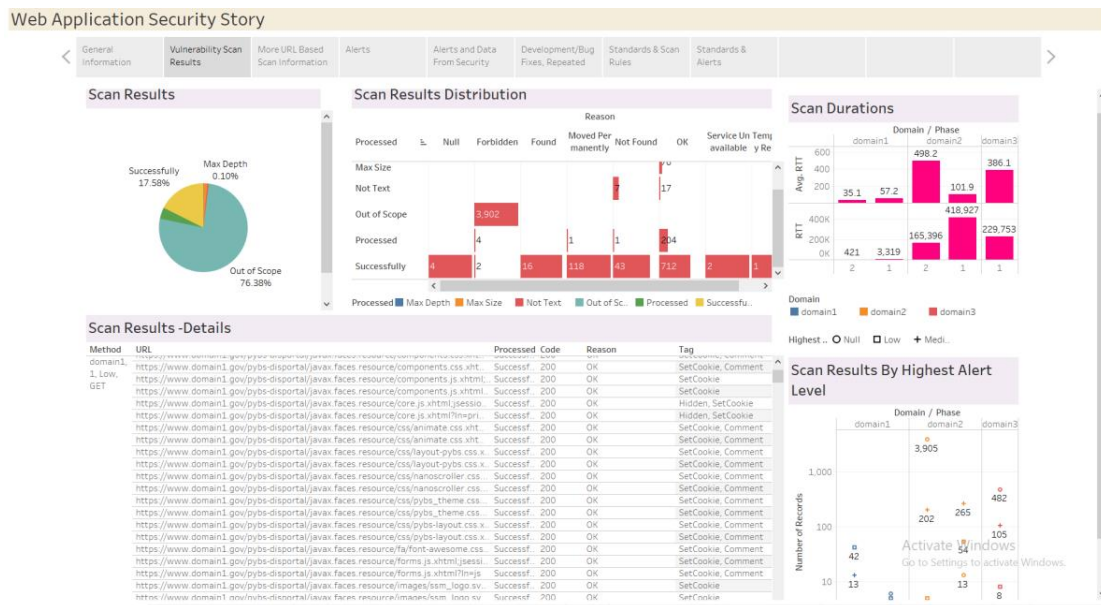


Figure 44 - Vulnerability scan results dashboard

In Figure 44, the basic information related to scan results is shown. The responses which were given to each scan effort, the distribution of successfully scanned and unscanned pages, the scan durations for each scan, detailed scan results, scan results by each domain and phase are shown in this view. Figure 45 provides a sample view of the tooltip showing the related metric names and URL based detailed scan information.

Some charts are repeated among multiple dashboards. For example, application size related information is also included in Figure 47, alerts dashboard. In this view, it is possible to see the number of alerts for each project, the distinct number of alerted URL's. Numerical information based on the number of modules, the number of external libraries and the number of lines of codes are also available in this dashboard. Dual axis property is used to overlap the charts showing the total number of and percentile information in this view.

In Figure 47, alerts information is combined with the data coming from other existing security systems. In the environment where the web application is installed, there may be other protection systems such as firewalls, anti-spam, and anti-virus systems. In this view, the percentile of detected alerts by the vulnerability scanner tool and other security system are provided. The alerts classified by scan rules are also included in this view.

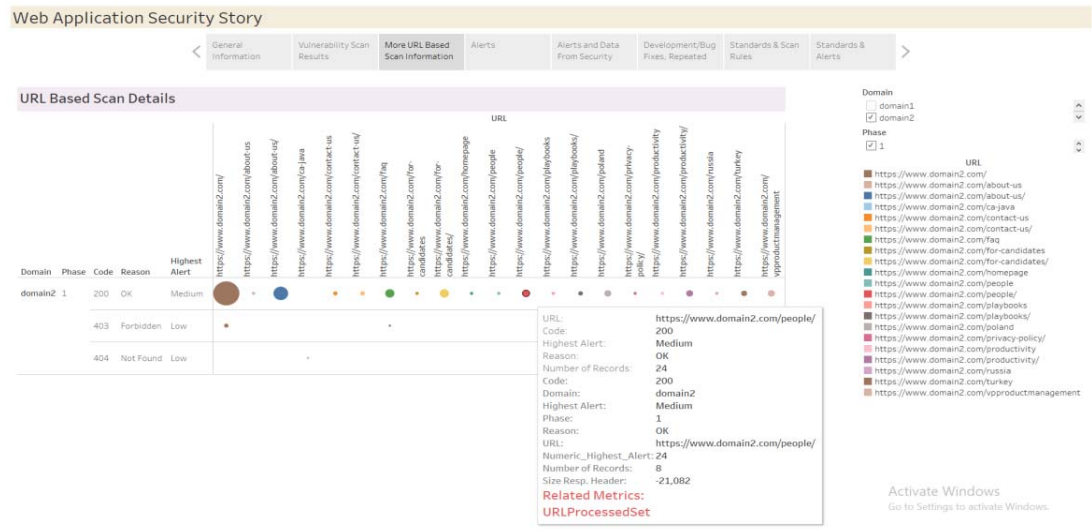


Figure 45 - URL based scan details dashboard

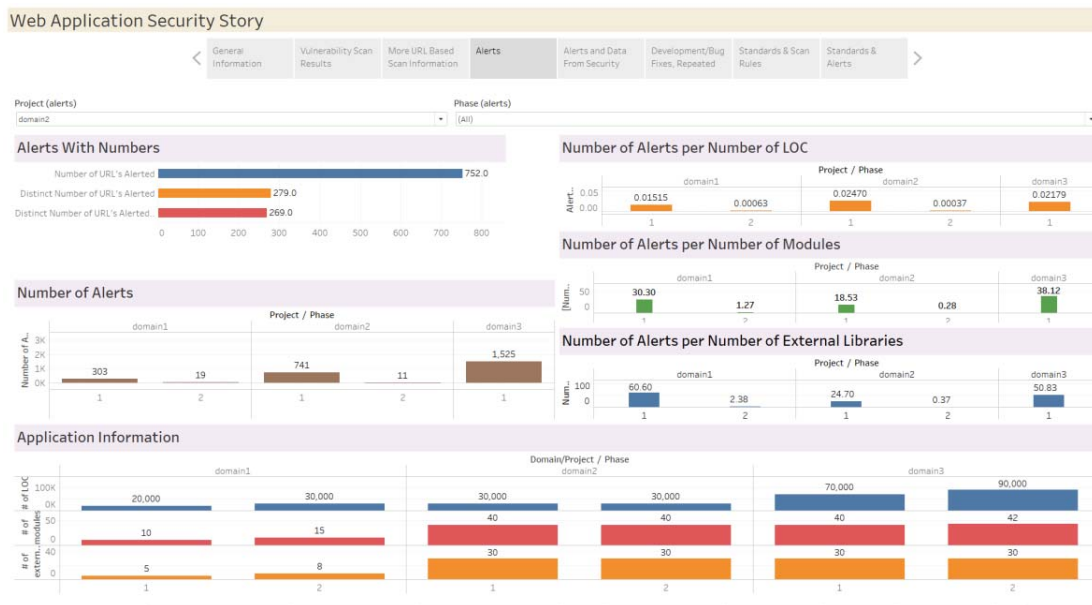


Figure 46 - Alerts dashboard

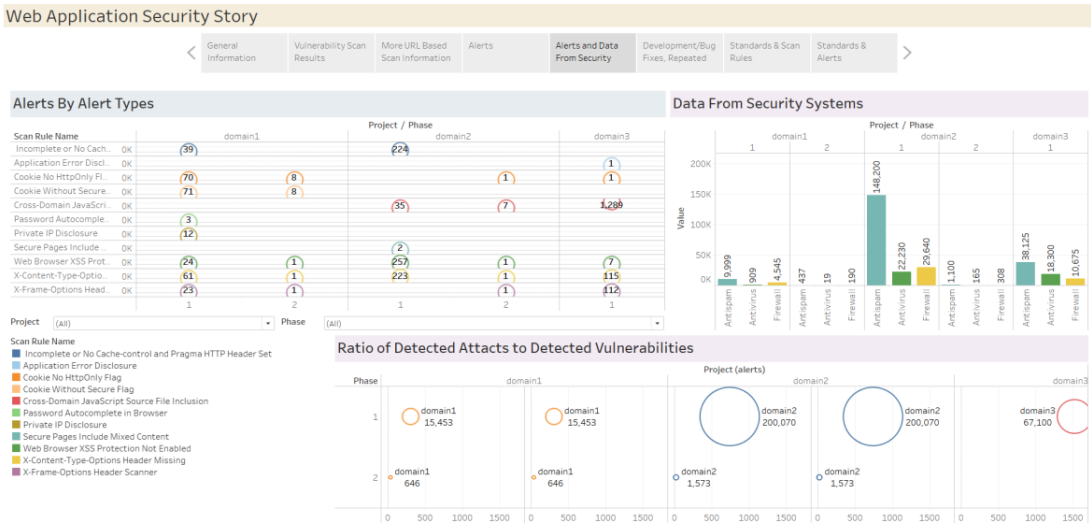


Figure 47 - Alerts and data from security protection systems

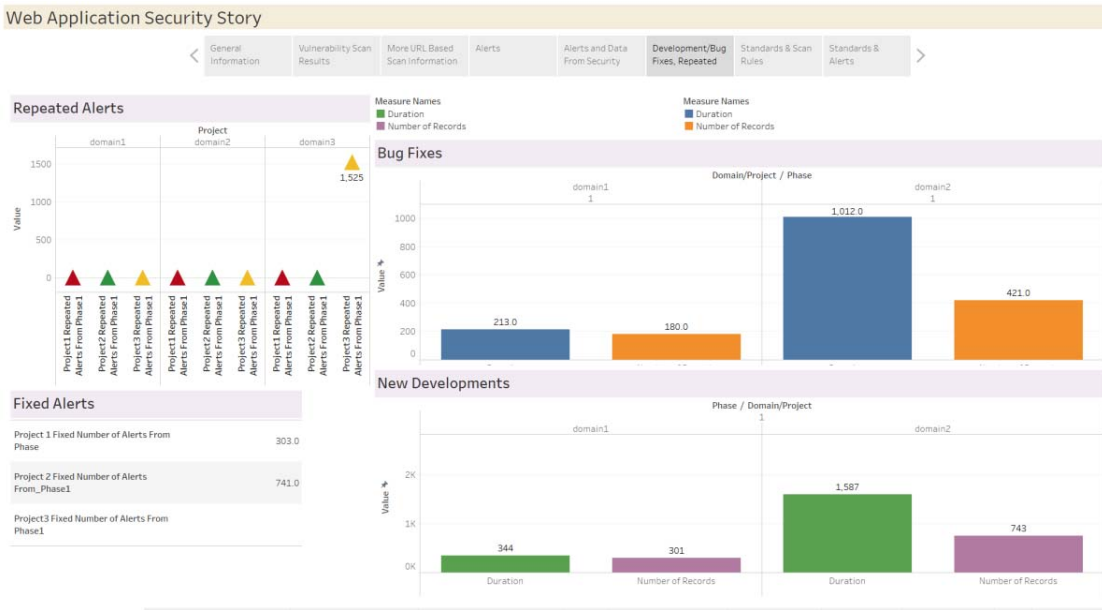


Figure 48 - New developments, bug fixes, repeated alerts, fixed alerts

Once the sets of alerts from one phase were compared with sets of alerts belonging other phases through set operations, it was possible to find out sets of repeated alerts from the previous phase and fixed alerts from an earlier phase for each project/domain. This would provide meaningful information related to the the overall security status and efforts given for the web application concerning security. To empower the dashboard, the number of security-related new developments and bug fixes also included in this view as shown in

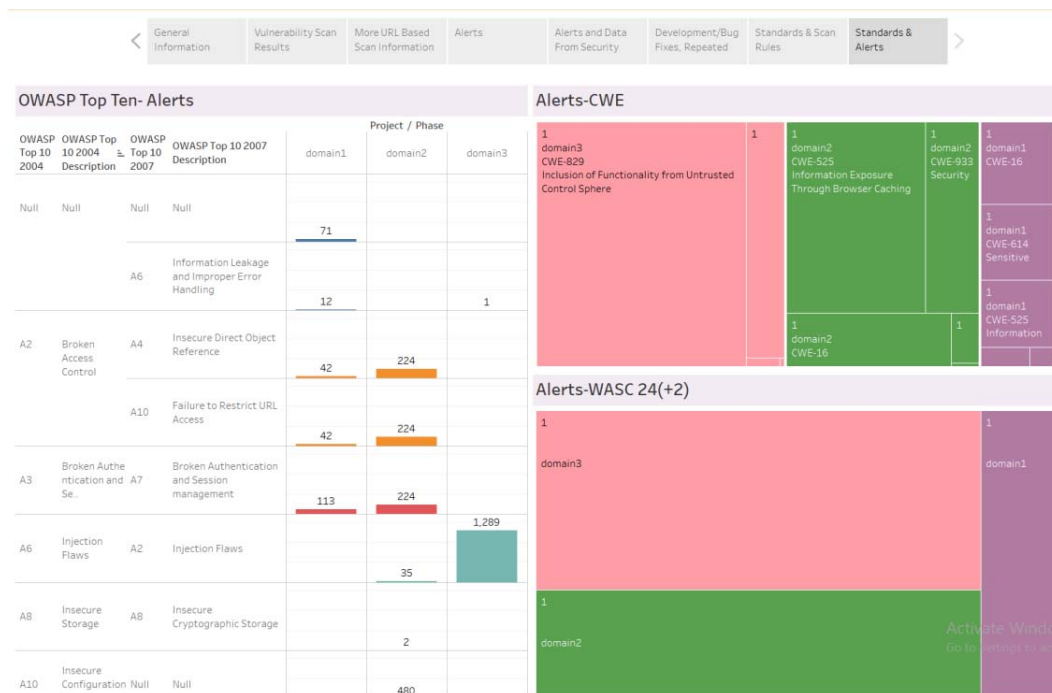


Figure 50 - Standards and Alerts

Figure 50 shows the distribution of alerts to the OWASP, CWE, and WASC standards. Compatibility with the standards may be valuable for some projects due to regulations or other obligations. Knowing the alerts related to well-known vulnerabilities and working on them using recommended best practices would eventually end up with better security for web applications.

Using the case study data and outputs, a validation study is conducted. The participants of this validation study is recruited using authors' social contacts. The characteristics of the participants are shown in Table 23. The aim of the study is described briefly to each participant while sharing a paper copy of the quiz and survey to enable them to quickly determine if they are able or willing to assist the study. Later, electronic versions of the quiz and survey were also shared to collect data. This validation study included 15 quiz questions which should be answered using the HWAS-V. The quiz questions prepared were created to allow the user to use every part of the HWAS-V tool while trying not to exaggerate the number of questions. After this quiz a set of four questions were asked the participants to measure the practicality, efficiency, decision-informing, and difference-detection attributes of the proposed system. These attributes of HWAS-V were questioned using a five-point Likert scale mechanism (1-Not At All Helpful, 2-Slightly Helpful, 3-Helpful, 4-Very Helpful, 5-Extremely Helpful). The quiz questions and the number of correct and incorrect answers are presented in Table 24. Numerical evaluation results achieved related to HWAS-V features are provided in Table 25.

Table 23 – Case Study Participant Characteristics

Characteristic	Value
Number of participants	14
Profession	Software Development, design and Test
Average age	30.7857
Average number of experience in software development and design	7.214286

One sample t-test has been used to test the significance of the survey outputs. The hypothesis was “The applicants found that the proposed tool was more than “Helpful” (numeric value =3 in Likert scale)” for the chosen measurements, namely practicality, efficiency, decision informing, and difference detection. The results indicate that there is only %16 probability that efficiency of this tool can be lower than Helpful, these probabilities are , %77, %1, and %1 for efficiency, decision informing, and difference detection properties. The evaluation results show that there is enough evidence to infer that "Decision Informing" and "Difference Detection" properties of the proposed design is significantly greater than “Helpful” according to the users. The other properties “Practicality”, and “Efficiency” are over the “Slightly Helpful” region. However, they lack enough evidence to be “Helpful” for the users during their analyses. According to the authors, the participants were quite successful in answering a relatively complicated set of questions with numerous comparisons and high level of decision information in a reasonable time.

Table 24 - Evaluation Questions and Results

Question	T	F
1-In which city does the web application corresponding to "domain 3" locates? a) Ankara, b) Istanbul, c) New York, d) Other	14	0
2-For all three projects two vulnerability measurements were done. Select the most vulnerable software project based on the number of vulnerabilities per project size measured as LOC (line of code). a) Domain1, b) Domain2, c) Domain3, d) All are equal	2	12
3-For all three projects two vulnerability measurements were done. Select the project for which no new development or bug fix was done between two analyses phases? a) Domain1, b) Domain2, c) Domain3, d) All are equal	11	3
4-What is the number of pages which have highest associated alert "Low" for “domain1” web application project? a) 105, b) 42, c) 265, d) 13	7	7
5-Vulnerability scanners can not process all the pages for web applications due to several reasons. One of these reasons is reaching the "Max Depth". Looking at the overall results for all three projects, what is the percentage of pages which are unprocessed due to reaching max depth. a) 10%, b) 0.1%, c) 20%, d) Other	12	2
6-Which project has the highest round trip time taken for a vulnerability scan session? a) Domain1, b) Domain2, c) Domain3, d) All are equal	14	0

7-What is the "metric name" shown in the tool tip box for the previous dashboard a) OWASPTopTen2007Vulnerabilities, b) NumberOfVulnerabilities, c) URLProcessedSet, d) URLsWithAlert	12	2
8-What is the number of alerts per modules per "Phase 2" of "Domain 2"? a) 0.28, b) 38.12, c) 100, d) 60.60	14	0
9-For which project, the project size did not change between two independent vulnerability scans? a) Domain 1, b) Domain 2, c) Domain 3, d) None	11	3
10-For "Domain 1" in "Phase 1", what is the number of vulnerabilities of the type "Web Browser XSS Protection Not Enabled"? a) 24, b) 1, c) 7, d) 35	14	0
11-What is the number of repeated alerts for "Domain 3" in "Phase 2"? a) 0, b) 1525, c) 303, d) 741	8	6
12-What is the number of fixed alerts for "Domain 2" in "Phase 2" from "Phase 1"? a) 0, b) 1525, c) 303, d) 741	11	3
13- Based on the scanner rules used in this tool, how many vulnerabilities in the CWE database were covered? a) 50, b) 27, c) 4, d) 0	9	5
14-Is "A10 - Failure to Restrict URL Access" of OWASP top ten 2007 vulnerabilities list is covered in the existing scanner rules? a) Yes, b) No, c) I don't know, d) N/A	7	7
15- For "Domain 2", what is the number of vulnerabilities of the type "A6 Injection Flaws"? a) 35, b) 2, c) 224, d) 71	14	0

Table 25 - Summary of Feature Evaluation T-test Results

	<i>Practicality</i>	<i>Efficiency</i>	<i>Decision Informing</i>	<i>Difference Detection</i>
Mean	2.71	2.92	3.64	3.71
Observations	14	14	14	14
Hypothesized Mean	3.0	3.0	3.0	3.0
Df	13	13	13	13
sd	0.726273039	0.916875	0.841897	0.913874
SE	0.194104634	0.245045	0.225007	0.244243
t-stat	1.471960144	0.291492	2.85706	2.924488
P	0.164823445	0.775275	0.013473	0.011838

5.5 Discussion

At first sight, due to having a dashboard design, and including a quite a large amount of metrics, HWAS-V can be considered as a Security Information and Event Management (SIEM) tool and can be compared to them. Gartner (Nicolett & Kavanagh, 2013) divided the SIEM products into four quadrants; Leaders, Challengers, Visionaries, and Niche Players. Based on this categorization HWAS-V can be located in the Niche players region due to its niche focus area, vulnerability scan results of black-box vulnerability scanners.

Although the aim of the study was not to provide a SIEM tool but had a web application specific focus, the similarities and differences have been examined. SIEM tools are more successful in working with continuous real-time data. The data structure proposed in this study is not continuous data but is collected intermittently based on the project schedules including the maintenance phases. The evaluation results indicate that SIEM tools do not have a specific focus on the web application vulnerability scan results. Although a few of them can integrate to some vulnerability scanners (mostly network vulnerability scanners), they do not provide a built-in data structure that will fit for most of the web application scanner results. Prebuilt metrics specific to web application vulnerability scan results are very low compared to HWAS-V or do not exist at all. A few of the SIEM systems allow importing custom data, which may allow the creation of part of the visualizations presented in this study. However, SIEM tools do not have comparable data joining, data blending, and set operation features comparable to HWAS-V, which relies on Tableau business intelligence tool. The detailed information regarding evaluating results for custom visualization generation capabilities of SIEM systems and available metrics related to web application security domain can be found in Özdemir Sönmez's and Günel's work (Özdemir Sönmez & Günel, Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation, 2018).

Another critical factor which differentiates the HWAS-V with the SIEM tools is the decision of combining project life cycle related information with the vulnerability scan results. This differentiation may also be seen less commonly for the combination of vulnerability scan results with multiple security standards, and a built-in structure for comparison of vulnerability scan results in multiple phases collected intermittently. HWAS-V has a project management perspective. It aims to provide a way to monitor the security-related progress such as new developments, bug fixes associated with previous alerts which do not commonly exist in the SIEM systems.

Investigation results indicate that SIEM tools and proposed web application vulnerability visualization tool are entirely different, both intended purposes and features do not overlap. The intended purpose of SIEM tools is generally to provide ways to collect data, to analyze data in real-time, generate compliance and regulatory reports, to correlate data and to find out indicators of events, to present these findings. However, the intended purpose of HWAS-V is to provide a practical, efficient way to examine present, past, and

recent vulnerability scan results that are coming from black-box tests for one or more projects for the decision informing and difference detection purposes.

This study examined common outputs of web application security vulnerability scanner tools and provided a data structure which is further used during the definition of a set of metrics and measures. New metrics are defined by combining the initial set of standard web application security metrics with possible effects of secondary data sources which may be originated from the development efforts, web application properties or the dynamics of the system environment. A case study is presented showing the visualizations based on data generated using OWASP Zed Attack Proxy (ZAP) tool together with some user-generated sample data as an improvement to the existing web application security vulnerability reporting systems.

In authors' best knowledge, few studies are focusing on the visualization of web application vulnerability scan results in the security visualization domain. The only existing study which targets web application security black-box test results enables visualization of statistical measures. The measures/metrics proposed in this study would enable a broader perspective of the security status for various stakeholders. More studies are required in this area to empower an extensive comparative analysis for this field.

Contributions of this study include a new dashboard tool for visualizing vulnerability scan results based on a unique data structure formed by combining multiple data sources. Using the phasing structure allows combining these multiple data sources. The design is developed through the use of Tableau software. Tableau dashboard designs can be viewed through Tableau Public, Tableau Desktop, and Tableau Reader applications. They may also be integrated with any web application which allows frame based HTML pages. Using Tableau Public will not be appropriate for viewing the web vulnerability scanner results for security reasons. However, Tableau Reader and Tableau Desktop may be more adequate to access HWAS-V. The secondary contribution of this work is the list of metrics/measures that the tool presents. The chapter also presents a case study and the validation efforts.

Some of the proposed metrics/measures are left out of scope during visualization design. Collecting information related to the average remediation latencies for each alert type is left as future work. Similarly, the classification of the alerts based on their effects to sensitive information, their impacts on business and their relation to the existing vulnerable components is also left out of the scope of this work.

The main limitation of this study is the use of OWASP ZED Attack tool for the case study. The proposed system is not tested with other vulnerability scanner outputs. The list of available attributes may change slightly using other vulnerability scanners.

The proposed metrics provide information related to the different aspects of web application security. It enables monitoring and comparing independent analyses for multiple projects. It is not limited to the raw outputs of the vulnerability scanner. On the

contrary, it serves a quite large amount of metrics and measures. However, there are some issues which the proposed metrics are not directly related.

The web application security related factors which are not measured with the proposed metrics/measures are:

- Tool efficiency
 - Number of false positives
 - Number of false negatives
- Security economics
 - Cost to fix
- The success of security education/certification
 - Defect injection ratio
- The success of code analysis
 - Defect detection ratio during code analysis
- The success of the test
 - Defect detection ratio by testers

In order to measure tool efficiency, the results of ZAP should be compared to manual inspection or similar results or should be compared with the results of other tools. This comparison, thus, measuring the tool efficiency is not in the scope of this work. In order to measure security economics, various other data types such as precaution costs, personnel costs, education costs should be associated with ZAP data. This association, and, thus, security economics of web applications is not in the scope of this work. The success of security education is totally out of the scope of this topic. The success of the users of the tool requires a comparison among multiple users, or a test project with known defects, which allows measuring the success of the users.

An analysis can be made using a broader set of vulnerability scan tools to enable a set of attributes which are available for all of them. This analysis may result in a light version of the dashboard. Similarly, using the available attributes with most vulnerability scanners may result in a more comprehensive version of the tool.

The Tableau software was also evaluated to some extent in this study for its suitability to develop a dashboard based on data coming from multiple sources and showing a large

number of metrics in association with each other. The results show that the software is a proper choice to design and develop complex security dashboards in feasible effort and time.

5.6 Concluding Remarks for Application Security Visualization

This chapter presented a new visualization study focusing on web application vulnerability scanner results. The visualization supports a large set of measures/metric. It integrates the vulnerability scanner data with some secondary data sources. By this way, it provides both a technical view and a managerial view.

The results indicate that the proposed design can help analysts and managers due to its “decision informing” and “difference detection” capabilities. Its level of “Efficiency” and “Practicability” are on the other hand questionable. These are in the “Slightly Helpful” range based on the validation results. These results are logical indeed. The large number of metrics and the large number of charts distributed to eight interconnected dashboards might affect the level of “Efficiency” and “Practicality” of the proposed design. A light version of the proposed design with less number of charts and metrics might have different results.

In the authors’ best knowledge, dashboard security visualization systems with a set of metrics/measures involving a specific user interaction: precise identification of each metric in the displays via tooltips is unique to this study, which has not been used in security visualization systems before. Precise identification of the metrics will increase the usability of the proposed system.

The system will provide a broad perspective of the security status of one or more projects. It also allows presenting results from subsequent analyses made by automatic web security analysis tools and comparison among them. Automatic comparison of subsequent scans will also enable to understand if there is a barrier which prevents proper scanning in a specific vulnerability analysis session. The proposed system is demonstrated using data generated by ZAP tool. The authors believe that incorporating other automated web vulnerability scanner results is also technically feasible and may be a logical direction for future research.

CHAPTER 6

EVALUATION OF SIEM SYSTEMS FOR CUSTOM DATA VISUALIZATION

6.1 Introduction to the Evaluation of SIEM Systems for Custom Data Visualization Study

Security Information and Event Management (SIEM) systems (Dimitrios, 2014) is the current trend for the examination of big data related to cybersecurity or information security. The rapid evolution of big data technologies and the existence of a considerable amount of data sources resulted in the development of many SIEM systems.

SIEM systems commonly include the tasks of data collection, data aggregation, data normalization, event correlation, reporting, and alerting. A few of the SIEM systems have capabilities to give information related to compliance with well-known security standards.

Visualization is one efficient way of data analysis which may aim (Sinar, 2018) data summary, comparison of values across groups, displaying connections/ relationships between variables, showing hierarchical or part to whole structures, illustrating change over time, and exhibiting data patterns.

SIEM systems commonly include built-in visualizations as part of reporting tasks. These visualizations ordinarily happen to be in dashboard formats. Some of the SIEM systems also allow visualization of custom data. Thus, the visualization capabilities of custom-made security visualization dashboard designs are commonly compared to other security-related dashboard designs prepared by using business intelligence (BI) tools such as Tableau (Tableau, 2018) or dashboard designs encapsulated in the SIEM tools.

All three groups of tools have specific characteristics and pros and cons. Thus, lack of detailed examination of custom visualization generation capabilities of SIEM tools results in incorrect or missing perceptions. For instance, the capabilities of custom security visualization systems designed in dashboard format are perceived as lower than they are because they do not have other common SIEM features. Another example is unnecessarily increasing the expectations for the capabilities of designing custom

visualizations using custom data in SIEM systems which may compete with visualization focused tools.

These drawbacks are the results of unique features served by business intelligence tools, custom-made security visualization solutions, and the SIEM tools. Some examples of these incomparable features are:

- advanced interactivity through drag and drop type of user actions for visualization generation and a large number of display types which exist commonly in visualization focused tools and BI tools,
- correlation analyses, easy enterprise integration, a large number of use-cases, and advanced data collection features which commonly exist in the SIEM tools,
- data or use-case specific design details which may exist in custom-made security visualization studies.

SIEM systems have many comparable features which may be used during the evaluation of these systems, such as the number of platforms supported, scalability, latency, number of built-in metrics, number of built-in dashboards and the, number of integration ways with third-party tools. In this study, the evaluation is limited to current capabilities related to the generation of custom visualizations using popular SIEM systems.

Since the ability to use the SIEM tools and achieving the correct results may be related to the experience and knowledge of the users, before starting the evaluation, it should be stated that the author has more experience in business intelligence (BI) tools (specifically Tableau (Tableau, 2018)) and visualization tools compared to SIEM systems. The author is at an equal distance to all the SIEM systems and did not receive formal education on any of them.

The rest of this chapter is structured as follows. In Section 6.2, the methodology of the study will be described. Section 6.3 includes the results of the study. Section 6.4 and Section 6.5 are the discussion and the concluding remarks for the chapter sections.

6.2 Methodology

6.2.1 Selecting the SIEM Systems to be Evaluated

Gartner Report 2017 (Nicolett & Kavanagh, 2013) divided the SIEM products into four quadrants as shown in Figure 51; Leaders: IBM Q1 Labs (IBM, 2018), LogRhythm (LogRhythm, 2018), Splunk (Splunk, 2018), McAfee (McAfee, 2018), Challengers: Micro Focus ArcSight (MicroFocus, 2018), Dell RSA (Dell, 2018), Visionaries: Rapid7 (Rapid7, 2018), Exabeam (Exabeam, 2018), Securonix (Most Visionary Next-Gen SIEM Platform, 2018), and Niche Players: AlienVault (AlienVault, 2018), Micro Focus NetIQ (MicroFocus, 2018), FireEye (FireEye, 2018), FortiNet (Fortinet, 2018), VenusTech

(Venustech, 2018), Trustwave (Trustwave: Smart Security On Demand, 2019), EventTracker (EventTracker, 2018). SolarWinds (Solarwinds, 2018), ManageEngine (ManageEngine, 2018), BlackStratus (BlackStratus, 2018). This categorization is based on two factors: the ability to execute and the completeness of vision. The ability to execute includes product/service properties, overall viability, sales execution and pricing, market responsiveness and record, market execution, customer experience, and operations factors. The completeness of vision includes market understanding, marketing strategy, sales strategy, product offering strategy, business model, vertical/industry strategy, innovation, and geographic strategy factors. See Gartner report for further explanation of the quadrants.

Due to limited time, and non-existence of trial versions for some SIEM systems, the authors decided to select one or two systems from each quadrant based on the count of systems in each quadrant.



Figure 51 - Gartner magic quadrant for SIEM systems (Adapted from Gartner 2017) (Nicolett & Kavanagh, 2013) Systems marked in red color were included in the evaluation.

Accessibility of trial versions or existence of cloud demo platforms for evaluation also affected the selection. As a result, AlienVault, Micro Focus ArcSight, Manage Engine

Event Log Analyzer, Splunk, Rapid7 InsightIDR, and Solar Winds Log and Event Manager were selected for evaluation, which are marked with red color in Figure 51.

6.2.2 Evaluation Scenario

The majority of the SIEM tools are very complex and would require specialized training to achieve complex tasks. In order to allow an interpretation of the capabilities and make a comparison with each other, a simple scenario was needed. This scenario should point out the steps required to build a custom visualization using custom data.

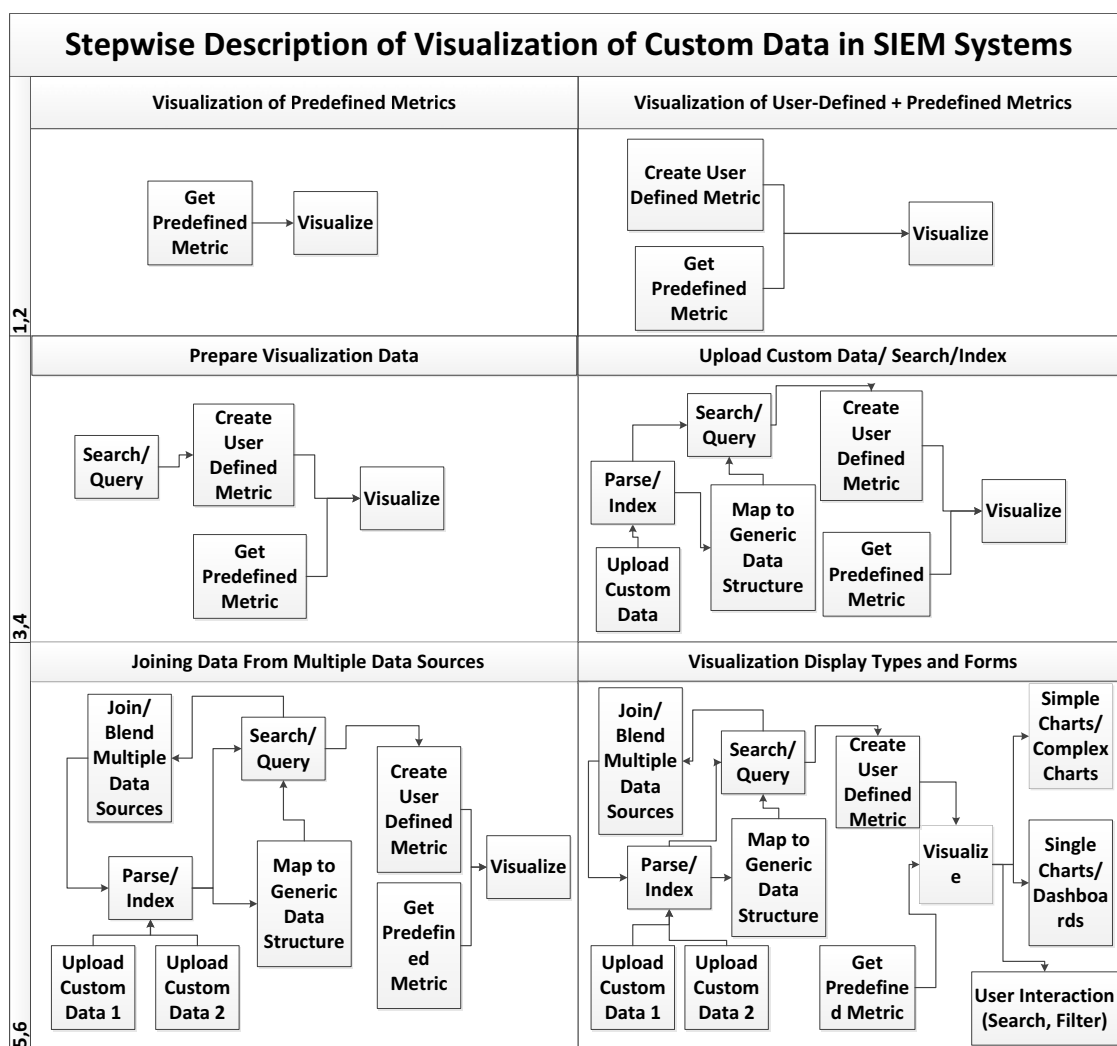


Figure 52 - Stepwise description of visualization of custom data using SIEM tools

Each SIEM system eventually has its own predefined metrics and visualizations. However, the creation of custom visualization would require additional features and tasks. In order to provide a basis for the evaluation which mainly targets checking the availability of these necessary features and tasks on each SIEM, an evaluation scenario was selected. Figure 52 shows a stepwise description of the examination steps for the selected scenario. This scenario includes the examination of:

- predefined metrics for a selected use-case,
- data import options for suitable data for the use-case,
- the existence of built-in data structures which may be used to map the imported log files for the selected use-case,
- the ability to load custom data files,
- the ability to form custom searches,
- the abilities of data joining and data blending for visualizations combining multiple data sources,
- built-in visualization capabilities to display the selected data results.

In the evaluation scenario, all the steps, except the first, are directly related to custom visualization generation. The first step, examination of predefined metrics, was included to help the researchers get accustomed with the SIEM systems before trying more complex steps.

As the target, the use-case "visualization of web application vulnerabilities" was selected by the authors due to its extensive usage and high recognition properties. Thus, during this examination, primarily the predefined metrics, existing log file types and data structures, and predefined visualizations related to the web application vulnerability scan results were investigated. However, other remarkable findings which are not directly related to the use-case, but which may be useful during visualization of other custom data, are also mentioned.

6.3 Results

In this section, first, the selected SIEM systems are introduced briefly. After the introduction, the evaluation platform is identified for each SIEM. Following this, each SIEM system is examined in seriatim using the evaluation scenario. Later, in the Discussion Section, overall comparison results are explained.

6.3.1 Manage Engine Event Log Analyzer

Manage Engine Event Log Analyzer stays in the Niche quadrant in 2017 report. The Event Log Analyzer free edition was installed as a desktop application for evaluation and the user guide was used to find out solutions for complicated tasks. Event Log Analyzer has a rich set of predefined metrics including application logs, operating systems logs,

firewall logs, antivirus, and Hyperware management information. It does not include any predefined metric for web application vulnerability scan results, explicitly. Event Log Analyzer has built-in alert definitions, but these point out general purpose alerts not vulnerability scan related alerts. Built-in alert data structure does not include scan information but only alerts. The tool is undoubtedly prepared to have advantages to search for indicators, but using custom log files for different purposes has some usability issues. It allows loading custom files. It allows searching the log files through the use of search expressions which mainly consist of a series of search criteria groups (key value pairs) concatenated with AND and OR keywords. Field values can also be directly used as search criteria for a quick search experience. The author could not manage to join multiple log files using the described type of search expressions. The application has a considerable amount of latency for basic search queries. Lastly, the author observed the existence of visualizations of custom data through line chart, area chart, and vertical bar chart.

6.3.2 *Splunk*

Splunk stays in the Leaders quadrant in the 2017 report. In order to make an evaluation, Splunk server was installed in a Windows machine, and Splunk Universal Forwarder was installed in a Linux virtual machine. Both the server and the universal forwarder applications are easy to install, execute, and configure. Splunk base application does not come with predefined metrics. Splunk has a large number of add-ons called applications. These applications provide ways to integrate with other tools and include predefined metrics and visualizations for these integrations. The authors searched for an add-on specifically for web application vulnerability scans using “web application” search term, but could not associate any application with this topic. It has a large number of add-ons related to vulnerabilities, VulnDB is one of them.

Splunk, SPL query language allows the joining of multiple data sources. However, forming any query with or without joining, requires specialized training and is much more complicated than using BI tools. Splunk has two types of join operations; left join and inner join. The user typically makes a query from a data source and assigns a table name to the result using the SPL language and this table can be joined to another table which is formed in a similar way. The difference from normal SQL queries is, in SQL query the query fields, joins and constraints are designed all in once, in SPL they are like separate and sequential operations piped to each other. Both approaches may have its own pros and cons. Preparation of queries to build the dashboard with a large number of metrics may be an issue for a user who does not have experience with the SPL language. Splunk is more successful in automatic field extraction; it even assigns new fields, such as index time. Splunk has good time facilities to investigate events.

Splunk has display options which are comparable to BI tools in look-and feel. The authors observed the existence of line, area, column, bar, pie, and scatter charts, and radial, filler and marker gauge type displays. Splunk add on applications may provide other display types, which have not been observed and tested in this study. However, the

design phase of these displays is more complicated compared to BI tools due to the complexity of the search statements. Other dashboard creation steps are straightforward. Each metric should be prepared as a table or as a visualization which should then be saved as dashboard panels for reuse. The created dashboards cannot be accessed via external applications. However, reports can be accessed. So, dashboards should run and be converted to a report before access from external applications.

6.3.3 *Rapid7 InsightIDR*

Rapid7 InsightIDR stays in the Visionaries section of the magic quadrant of Gartner 2017 report. Cloud Trial Platform was used for evaluation. Rapid7 InsightIDR Collector has been installed on a Windows machine for data collection. Rapid7 InsightIDR has predefined metrics for various subjects including firewall activity, ingress authentication, active directory admin activity, compliance, asset authentication, DNS queries, IDS alerts, virus alerts, and file access activity. It does not have predefined metrics for application vulnerability scan results.

It has built-in integration with Rapid7 Nexpose vulnerability scanner system, and thus built-in data structure is compatible with the Rapid7 Nexpose vulnerability scanner. It allows importing custom data. However, this process is not very straightforward. It requires that every source machine has a fully qualified domain name which may not be possible for all cases.

Rapid7 has a unique language, Log Entry Query Language (LEQL) for data query which follows SQL syntax. It allows building queries based on multiple data sources easily with the use of joins. Saving queries is possible. No information was found related to set operations.

It has good dashboard building features which resemble BI tools. Dashboards are designed as a composition of cards which may be either built-in cards or user-defined cards. The authors observed the availability of timeline area chart, horizontal bar chart, bar chart, calculated number, gauge chart, timeline line chart, timeline multi-area chart, horizontal multi-bar chart, multi-bar chart, timeline multi-line chart, pie chart, table data, in the trial platform.

6.3.4 *Solar Winds Log and Event Manager*

SolarWinds Log and Event Manager (LEM) stays in the Niche quadrant of 2017 report. Log and Event Manager server application was installed as a virtual machine on VMWare, and SolarWinds reporting tool was installed on a Windows machine for evaluation. The server application is primarily responsible for data collection, data correlation, and alerting tasks. Reporting application has around 300 built-in reports encapsulating a large number of metrics. These metrics are related to agent status, authentication, change management, event summary, file audit, incident alert, machine audit, malicious code, network events, network traffic audit, registry audit, resource configuration, and tool maintenance. The authors did not find an existing data structure or a generic data structure suitable for application vulnerability scan imports.

SolarWinds has a number of connectors for various devices or formats. It does not have a generic connector which reads custom log files. The company offers that if the third party tools can be generated in Syslog format, then it may be indexed and searched using LEM. One other solution suggested in the user forums is forming a new user-defined connector. This suggestion depends on the fact that each connector is actually an XML file which defines the mapping of log file attributes to LEM items.

In SolarWinds, available built-in reports can be modified by the users by adding user-defined filters based on report attributes. These reports can then be saved in Crystal reporting format. SolarWinds uses “custom reports” term for these user-defined reports.

LEM reports application depends on Crystal Reports third-party visualization tool. Thus, it encapsulates various types of table and display formats available in crystal reports.

6.3.5 *Micro Focus ArcSight*

The authors made an effort to evaluate SIEM systems from all four quadrants. However, it was not possible to find and access an evaluation setup for some of the SIEM systems, either permanently due to test platform maintenance (RSA) or indefinitely (Microfocus ArcSight). Micro Focus ArcSight stays in the Challengers quadrant of 2017 Gartner report. No trial version was available at the time of evaluation. Thus, a series of workshop video tutorials and product documents have been used to understand the critical features related to custom visualization generation.

ArcSight comes with a standard structure which involves a series of coordinated resources. This structure involves built-in metrics and dashboards related to configuration monitoring, such as undesired actions to systems, devices, and applications, intrusion monitoring, network monitoring, incident response tracking and ArcSight system monitoring.

ArcSight uses ArcSight Common Event Format data structure. Thus, even a custom log file can be loaded to the system. Then, the data attributes which are identified after parsing can be mapped to Common Event Format Data Fields by the user. This mapping is done as a continuation of the Regex definition for the particular file.

ArcSight has a large number of Flex Connectors. ArcSight Regex connector is one of them. ArcSight Regex connector allows making a definition of the log file by using Regex format. Using Regex format allows parsing and indexing of complex log files. ArcSight Regex connector breaks the log statement into tokens using the declared Regex statement. The system provides a helper tool, ArcSight Regex Tester, which can be used to generate the necessary Regex statement to parse the custom file.

ArcSight has viewer panels which can include HTML based reports and several charts. Although the product includes other chart types, such as hierarchy maps (treemaps), custom query results can be visualized as a table, pie chart, bar chart and horizontal bar chart according to user documents in ArcSight.

6.3.6 *AlienVault*

AlienVault stays in the Niche quadrant of Gartner 2017 report. The tool cannot be installed on Windows OS directly. It is designed to be installed on VirtualBox. Since the author had some problems with this installation, the AlienVault online demo version was examined for evaluation.

The online version has a number of prebuilt dashboards encapsulating a large set of metrics. However, it does not have a specific dashboard for web application vulnerabilities scan results.

The authors examined generic vulnerabilities dashboard which is designed to be used by various assets including software programs. Generic vulnerabilities dashboard has very few metrics which includes more vulnerable assets, mostly detected vulnerabilities, vulnerabilities by type, and a number of scan jobs.

The online demo version did not allow uploading custom data due to restrictions. However, the author contacted the customer support and found out that this restriction is only applied to the online demo version.

The querying mechanism of AlienVault is based on search strings consisting of key-value pairs. These keys can be both built-in fields such as IP, src_port and user-defined fields. The filename can also be used as a search parameter which shows that data source specific search can be made using file names. It looks like there is no straightforward method to join multiple data files. However, AlienVault has connectivity to several databases and allows complex database queries including joining of multiple tables. The author observed the creation of visualization using simple charts including line chart, area chart, and vertical bar chart in this tool.

6.4 Discussion

SIEM systems are generally expensive systems, which require specific installation platforms to be installed. For this reason, the author thinks that the majority of the SIEM users are familiar with only a few of these systems. Independent evaluations, such as this study, would help to get familiar with these systems. This familiarity would eventually help to make better selections in the long term. Table 26 contains information related to the configuration of the selected SIEM systems in this study. There may be multiple interfaces for a few of the SIEM systems. For example, ArcSight has Console, Web, and Command Center interfaces. The table includes the access type which was evaluated during this review. A few of the SIEM systems are suitable to be installed on different operating systems. The data connectors or data collectors have different mechanisms with the same purpose, gathering data for the SIEM systems. The SIEM systems may have various types of data collectors. The collectors listed in the table are the ones which are tested during this study. It is important to note that this table is prepared based on authors' own experiences and limited with the configurations tested in this study.

SIEM systems are focused on threat capture, gathering network intelligence and detecting malicious activities. In general, they have very advanced features to accomplish these targets. Although most of the SIEM tools are very handy and have useful features, when the objective is working on custom log files, they have different approaches which result in several difficulties.

Table 26 - SIEM Configuration Table

	Quadrant	Inst. Platform	Data Collector/Connector App.	Reporting App.	Access Type
Manage Engine Event Log Analyzer	Niche	Windows machine	-	-	Web Based Access
Splunk	Leaders	Windows machine, Universal Forwarder - Linux machine	Universal Forwarder	-	Web Based Access
Rapid7 InsightIDR	Visionaries	Cloud Platform, Rapid7 InsightIDR Collector- Windows machine	Rapid7 InsightIDR Collector	-	Web Based Access
Solar Winds Log and Event Manager	Niche	Server App - VMWare Virtual Machine, SolarWinds Reporting App - Windows machine	-	SolarWinds Reporting Tool	Web Based Access
Micro Focus ArcSight	Challengers	-	ArcSight Regex Connector	-	Web Based Access
AlienVault	Niche	VirtualBox, Online Demo Version	-	-	Cloud Access

Comparison of these systems is also challenging due to the existence of different data flows as a result of different sequence of actions which end up with user interfaces that are difficult to compare.

The author could not complete some steps of generating visualization for each SIEM, such as importing custom data, joining multiple data sources, building a visualization encapsulating multiple displays, designing a visualization by a drag and drop type user interactivity. These difficulties or inability were interpreted as either not having this feature or not having a straightforward way to achieve this step.

Table 27 - Evaluation Summary

	Visualization of Predefined Metrics	Creation of User Defined Metrics	Search/Query Mechanism	Upload Custom Data	Join/Blend Multiple Custom Data	Display Types
Manage Engine Event Log Analyzer	Yes	Yes	Key-value paired search expressions	Hard	No	Line, area, and vertical bar charts
Splunk	Yes	Yes	Search Processing Language (SPL)	Very Easy	Yes	Line, area, column, bar, pie, scatter charts and radial, filler and marker gauges Timeline area, horizontal bar, bar, gauge, timeline line, timeline multi-area, horizontal multi-bar, multi-bar, timeline multi-line, and pie charts, table data, and calculated number
Rapid7 InsightIDR	Yes	Yes	Log Entry Query Language (LEQL)	Hard	Yes	horizontal multi-bar, multi-bar, timeline multi-line, and pie charts, table data, and calculated number

Solar Winds Log and Event Manager	Yes	No	Update Reports	Built-in	Hard	No	Display formats available in crystal reports Table data, and pie, bar and horizontal bar charts Line, area, and vertical bar charts
Micro Focus ArcSight	Yes	Yes	Regex Based Query		Easy	Yes	
AlienVault	Yes	Yes	Search strings consisting of key-value pairs		Neutral	No	

Although each SIEM system has its own outstanding features, one obvious result of this examination is, it was not possible to complete the planned scenario for the majority of the selected SIEMs. This result points out the known differences in BI tools and SIEM tools.

Installation and file upload difficulties were the most common difficulties during this study. Different ways of mapping the available custom data to the product fields have different results. Some tools make an automatic mapping of provided custom data to an available standard data structure. While this automatic mapping is faster and less tedious, the author felt that mapping the fields manually as in the ArcSight example allows more correct mappings of the fields and helps to manage the data better in subsequent sections, such as search and display. Otherwise, the tool has all the control, and the user may end up with visualizations that he/she did not plan. The background of the user is also important. Having prior knowledge on some technologies such as Regex syntax makes things easier. Otherwise, a long preprocessing step for some tools may be a burden for some users.

The author thinks that, it was prudent to choose visualization of web application vulnerability scanner results as the custom use-case. The reason is it never existed in a built-in manner in the evaluated tools. Otherwise, the comparison would be biased, and the target of generation of custom visualization would have been strayed by the author, unintentionally.

The observations described in this evaluation study apply only to the custom data visualization. In general, the tools behave entirely differently in data parsing, indexing, and querying and even in data display tasks for a data source which has a familiar data structure such as sys log or built-in integrator with the SIEM. In that condition, most of the manual tasks may turn into automatic tasks, and the displays are generated quickly in real time with the data occasionally. The known data structure will also enable other tasks, such as automatic correlation of data with other data sources, automatic threat/vulnerability detection with known metrics and, automatic display of prebuilt

dashboards. Table 27 provides a summary of the comparisons of the selected SIEM systems.

The vendors for the majority of the evaluated products have other security analyzer tools along with SIEM systems. A few of those products may be more suitable for custom log file visualizations.

One significant contribution of this study is better decision making. The author aims that the potential users of SIEM systems may benefit from this study when choosing a SIEM for their needs and when designing their custom log management systems.

6.5 Concluding Remarks for Evaluation of SIEM Systems

This chapter presents the evaluation results for SIEM systems focused on the creation of custom visualizations. The evaluation results demonstrated custom visualization generation related features/functions which are powerful or open for improvement for six well-known SIEM systems.

The provided evaluation method points out a practical scenario to test the effectiveness of SIEM tools regarding the targeted objectives. This scenario may as well be used for other use-cases which may have other impressive results.

Generally, these SIEM systems are compared according to their feature lists. The authors claim that scenario based comparisons as in the provided case would provide better information for these SIEM systems.

The SIEM systems which stay in different quadrants of the Gartner report change annually due to changes in the SIEM systems. Related to this issue, the results achieved during this study would eventually be affected as existing features are modified and new features are added to the current SIEM systems. Thus, this kind of scenario based evaluations should be repeated in short periods.

CHAPTER 7

A DECISION SUPPORT SYSTEM FOR OPTIMAL SELECTION OF ENTERPRISE INFORMATION SECURITY PREVENTATIVE ACTIONS ALONG VISUALIZATION

7.1 Introduction to Decision Support System For Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization Study

Enterprise security is characterized as the protection of business assets and goals (Sherwood N. A., 2005). All the familiar, well-known security targets such as confidentiality, integrity, availability, non-repudiation, and accountability also apply to the enterprise security domain. Enterprise security is a broad area comprising of numerous subdomains, such as application security, network security, database security, web server security and wireless security. Huge numbers of these subdomains might be considered based on the enterprise's needs. Actual security requirements of an enterprise would be dependent on the criticality of the data it stores, its business functions, assets, architecture, and physical locations. Subsequently, prior to making a security establishment for an improvement plan, a requirement analysis specific to the enterprise should be made.

Information security is an extremely complicated issue which should be dealt with deliberately and in a timely manner. Security experts should act smarter than the owners of malicious actions. Information security depends on many factors. Installing the selected precautions and leaving the system as such is not a robust arrangement which will solve every aspect of the security problems consistently. Even the newly installed solutions may bring new problems (Finne, 1998). Other fundamental issues should be considered before continuing. Security plans should be enterprise-wide and should be supported by the top level management to be successful (Solms R. V., 1996). In the event that these conditions are not fulfilled, efforts for better security management may result in failures easily.

Different scales of enterprises would eventually have diverse security requirements and diverse solution approaches. As the enterprise scale increases, the number and sophistication level of possible malicious actions and the number and sophistication of corresponding precautions also increase. Security personnel number and their roles

would also change along with the solutions implemented. In large enterprises, there may be more dedicated security personnel, such as the chief security officer and security analysts. In smaller enterprises, usually, system admins are in charge of dealing with the enterprise security.

Regardless of the scale of the enterprise, the risks for the enterprise should be identified. Accepting, avoiding, transferring risks and applying appropriate security solutions to handle risks are alternative actions that can be taken after the identification of the risks. Even if the enterprise policy results in the utilization of freeware or shareware protection software, there would still be a need for a security budget to cover the costs of hardware components, human resources, and other security-related services. Crossler et al. (2017) asserted that the response cost and the response efficacy explain why certain responses are repeatedly selected for specific threats. In spite of the fact that there are numerous studies which focus on definition and prioritization of the information security risks, limited research has guided decision makers on how to determine the adequate actions for preventing the risks while considering the budget limits. As almost all enterprises have financial limits for security spending, consideration of enterprise resource constraints is crucial for making rational security management decisions.

As business roles, functions and procedures change, the importance of security targets also changes. For example, if non-availability of the business functions is costly to the business, then availability should have higher priority. Similarly, if the enterprise has tasks which are sensitive to the tasks' owner and denial of such tasks is costly, then non-repudiation should be critical. Therefore, prioritization of security targets is essential. AHP has been used in numerous earlier works for the prioritization of security targets for an enterprise. Their focus was on the evaluation of security targets. However, evaluation of threatening actions and precautions has not been considered. Without knowing threatening actions or situations which cause vulnerabilities for an enterprise, a right security expenditure plan cannot be made. Prioritization of security targets without evaluating the vulnerabilities and possible threatening situations does not help in finding the optimal set of precautions for an enterprise. In the best conditions, choosing a subset of precautions may be based on adopting industry best practices. Still, the resulting set of precautions would be far from being optimal for that specific enterprise. Existing security prioritization studies also do not take into account a fixed amount of budget criteria. In the real world, enterprises should stick to an initial security cost budget. Similarly, during periodical system improvements, which is necessary due to the changeable nature of security status, enterprises would have a fixed amount of maintenance budget. Having an assigned total budget and attempting to maximize the total amount of risk prevented is an optimization decision-making problem.

Beginning from the last decade information visualization has been used in the visualization of network security related events. However, general security management techniques and enterprise security management related methodologies do not commonly benefit from the capabilities of information visualization. None of the earlier studies

focusing on the prioritization of security targets provide a graphical view which may further assist the decision-making process of managerial staff. Enterprises require an efficient, verifiable computation mechanism which is easy to apply, easy to repeat, and which provides an end-to-end result, starting from a «threat model» ending with «set of precautions» for a «predefined budget.»

Hence, within this context, the main focus of this chapter is to fill the gap in the literature by presenting a new decision support system (DSS). This DSS will not only provide guidance on assessment and prioritization of risks but also will enable optimal allocation of the limited enterprise budgets to take the adequate preventive actions for maximizing the enterprise information security and allow visualization of results for better decision making. The proposed decision support system is based on the Analytical Hierarchical Process (AHP) (Saaty, 1988), Mixed Integer Programming (MIP), and Treemap visualization techniques (Shneiderman & Plaisant, 1998).

In this novel decision support system, decision makers may find answers to the questions “For a given set of threats, what is the minimum security budget to attain a desired level of risk?”, alternatively, “For a given set of threats and a security budget limit, what should be the optimal preventative actions to minimize the risks?”. The DSS also enables different visualizations of the results to assist further the parties involved in the decision-making process (Shneiderman & Plaisant, 1998) including technical managers, financial managers, and security analysts to improve comprehensibility and communication of the data.

The remainder of this chapter is divided into seven sections. Section 7.2 is devoted to the literature review which describes analytical methods for security domain . In Section 7.3 the methodology of the study is described. In Section 7.4, the proposed DSS is described and in Section 7.5 a case study is provided. Section 7.6 includes the discussion and finally in Section 7.7 concluding remarks are made.

7.2 Literature Review for Analytical Methods for Security Domain

In recent years, growing security needs have led to researchers to focus on the role of management to improve information security (Soomro, Shah & Ahmed, 2016). One particular area of interest has been the information security risks assessment and prioritization. Security risk assessment involves both qualitative and quantitative comparison. AHP (Saaty, 1988) has been used commonly in security risk assessment studies as it enables qualitative and quantitative comparison simultaneously. AHP is a popular structural technique used for complex decision making. The decision problems which AHP can be applied include choice, ranking, prioritization, resource allocation, benchmarking, quality management, and conflict resolution (Forman & Gass, 2001).

In an early study, Wang and Wulf (1997) proposed the AHP use to measure and compare computer security. Kim and Lee (Kim & Lee, 2007) provided a methodology which

consisted of a process model, a criteria selection stage and AHP to support the security controls selections for information security management systems. Taha et al. (2014) proposed a method based on AHP to compare the security level of cloud service providers in which service level agreement parameters of cloud service providers were used. These parameters do not always consist of quantitative values which feature the ability to combine quantitative and non-quantitative values of AHP modeling. Breier (2014) evaluated the security properties such as availability and confidentiality using AHP towards ISO/IEC 27002:2005 standard (Disterer G. , 2013) security clauses including “security policy (SP)” and “asset management (AS).” This approach provided a different viewpoint such as “In order to provide availability in an environment Asset Management is nine times more important than Information Security Incident Management.” The approach may be useful in order to prioritize actions that should be taken, but would not be equally useful when making decisions for spending a fixed amount of budget for security. Bodin et al. (2005) proposed an AHP model for comparison of security investments which included the expected security properties, such as confidentiality, availability, integrity, but did not consider any threats or specific actions for decreasing the risks of threats. Siddiqui et al. (2011) used the AHP model to evaluate commercial and non-commercial enterprise service buses. The evaluation criteria included information security, interoperability, and high availability. The AHP based models mentioned so far, focus on the prioritization of security properties and targets, they do not have a specific focus on the prioritization of security threats.

Hybrid models are developed using two or more existing models. They are beneficial in security risk assessment because some techniques are superior to handle some aspects of security better than the others. The use of AHP (Saaty, 1988) and Linear Programming Model (Gass, 1958) in combination to prioritize threats and spend a fixed amount of budget optimally, also makes the proposed methodology a hybrid methodology. Lee (2014) combined AHP and Fuzzy methods for the evaluation of security risk assessment. In order to reach the desired level of security, measuring information security progress and maturity is essential. Kyung et al. (2011) used the Fuzzy AHP model to evaluate 31 information security measurement indicators which were developed through questionnaire results. Syamsuddin used the Fuzzy AHP model (Syamsuddin, 2012) to evaluate decisions related to information security, particularly in governmental organizations. Although this work took security economy as one of the evaluation criteria among other criteria, it did not consider spending a fixed amount of budget for an enterprise. Syamsuddin used the Ternary AHP model in another work (Syamsuddin, 2013) in which ten evaluation criteria including management, technology, economy and culture categories were evaluated towards availability, integrity and confidentiality capabilities. Unlike classical AHP, Ternary AHP includes only three options, win, loss, and tie, which minimizes the efforts, and decreases the inconsistencies. Cuihua and Jiajun (2009) used a combination of AHP and Grey Relational Analytic Process (GRAP) for the evaluation of system security. Lai et al. (2016) used AHP for the prioritization of the risks caused by threats to a website. In this model, Information Entropy (Shannon, 1948) and Game Theory (Gibbons, 1992) techniques have been utilized following the AHP to

reduce the subjectivity of pairwise comparisons for the prioritization values for the detected threats. Dogu and Celikoglu (2012) also used AHP for risk prioritization together with Bayesian prioritization procedure to handle missing values in AHP pairwise comparison matrix. This approach allows the security risk evaluators to make some of the initial judgments and leave the rest of the decisions, such as the decisions requiring additional information, for future processing during the prioritization of risk items using AHP. These studies have a focus on the classification of the AHP results in general, with no intention for the optimization of the budget.

In recent years, alternative techniques have been proposed for information security risk assessment. Yang et al. (2013) presented a multicriteria optimization and compromise solution, VIKOR model based on the analytic network process (ANP) and trial and evaluation laboratory techniques for information security risk-control assessment. The VIKOR model enabled independence among the criteria/variables and unequal weights for each cluster. Yang et al. also mentioned that future studies should focus on achieving the lowest cost and the least resources to establish controls for reducing the risks to an acceptable level. Feng et al. (2014) proposed a security risk analysis model which defined the risk factors and their causal relationships using Bayesian networks and performed security vulnerability propagation analysis by ant colony optimization. Bayesian networks were also used by Shin et al. (2015) for cybersecurity risk analysis of nuclear facilities. Shameli-Sendi et al. (2016) provided a taxonomy of information security risk assessment to help organizations to conduct the risk assessment properly. Game theory was used to model a security investment DSS system (Fielder, Panaousis, Malacaria, Hankin, & Smeraldi, 2016). Utility theory was used to provide a model which makes optimization of security cost while considering the dynamic nature of vulnerabilities (Miaoui & Boudriga, Enterprise security investment through time when facing different types of vulnerabilities, 2017).

AHP is commonly used in the literature for information security risk assessment as it is a practical method and can also be used to assess qualitative information, however, alternative methods exist which are used for enterprise information security risk assessment. A well-known formal enterprise risk management technique is the COSO Enterprise Risk Management Technique (Moeller, 2007). The COSO procedure requires the definition of the specific control environment which serves as a foundation for the rest of the risk management activities. The techniques which are used as part of COSO procedure include, but are not limited to, following techniques: surveys, interviews, industry or event focused benchmarking, scenario analysis, risk interaction matrixes, causal at-risk models, fault trees, event trees, bow-tie diagrams, risk hierarchies, risk maps, and MARCI charts (Curtis & Carey, 2012). CORAS (Aagedal, et al., 2002), is another risk assessment technique developed for enterprise security. CORAS encapsulates and evaluates several techniques for different steps of enterprise risk assessment procedure. It includes a graphical modeling language which is used to model the risky situations for the enterprise. Hazard and operability study (HazOp) (Kletz, 1997) technique is another method proposed for the identification and early analysis of

risks, during the identification of treatment options. Fault tree analysis (FTA) (Ericson & Li, 1999) technique is proposed for the top-down evaluation of failure modes of particular applications in an enterprise. Failure mode effect criticality analysis (FCMECA) (Arunraj & Maiti, 2007) is another technique which is used for bottom-up analysis of risks for critical sub-parts and the addressing of barriers and countermeasures. Markov analysis (Markov) (Littlewood, 1975) is also applied for addressing system states and likelihoods of events for the evaluation and treating of risks. CCTA Risk Analysis and Management Methodology (CRAMM) (Eloff, Labuschagne, & Badenhorst, 1993) can also be used for the valuation of assets, focusing on data groups and for the identification of countermeasures (Aagedal, et al., 2002). Other well-known risk assessment techniques include Delphi Method (Moeller, 2007) and Monte Carlo Simulation (Moeller, 2007).

Few research has focused on the use of exact methods such as linear programming (LP) (Gass, 1958) or Mixed Integer Programming (MIP) (Gomory, 1963) for the information security optimization. Sun et al. (2013) used LP modeling to evaluate the security of block ciphers. Sun et al. (2013) applied LP modeling to calculate the number of active S-boxes of the PRESENT cipher. Vigo et al. (2013) presented a LP model to make quantitative analyses on attacker behaviors to improve the trustworthiness of the cyber systems. Qui et al. (2012) used LP to optimize the security of cryptographic algorithms used for encryption and integrity of ubiquitous computing systems. The attributes related to the computational overhead of algorithms were included in optimization.

One of the main limitations of the majority of security risk assessment studies is that they did not consider the costs of actions for eliminating or decreasing the risks of threats to improve information security. Few studies have focused on the security costs and budgets. Gordon and Loeb (2002) provided an economic modeling framework for information security investment decisions. Anderson and Choobineh (2008) presented information security strategies for enterprises and mentioned that the security budget depends on the risk tolerances of decision makers and the information available about threats, vulnerabilities, potential damages, and likelihoods. Fessi, Benabdallah, Boudriga & Hamdi (2014) proposed a multi-attribute genetic algorithm approach for automatic intrusion response based on a cost-benefit model for assessing the quality of responses and a multi-attribute value model to assess different attributes including financial costs, reputation loss, and processing resources. Miaoui and Boudriga (2017) proposed a method which relates security investments to the evaluation of vulnerabilities using utility theory. Although few research considered information security costs, to authors' best knowledge none of the existing studies presented a method that integrates information security risk analysis, security investments, and optimization to achieve optimal information security risk management and investment decisions, which is the main objective of this study.

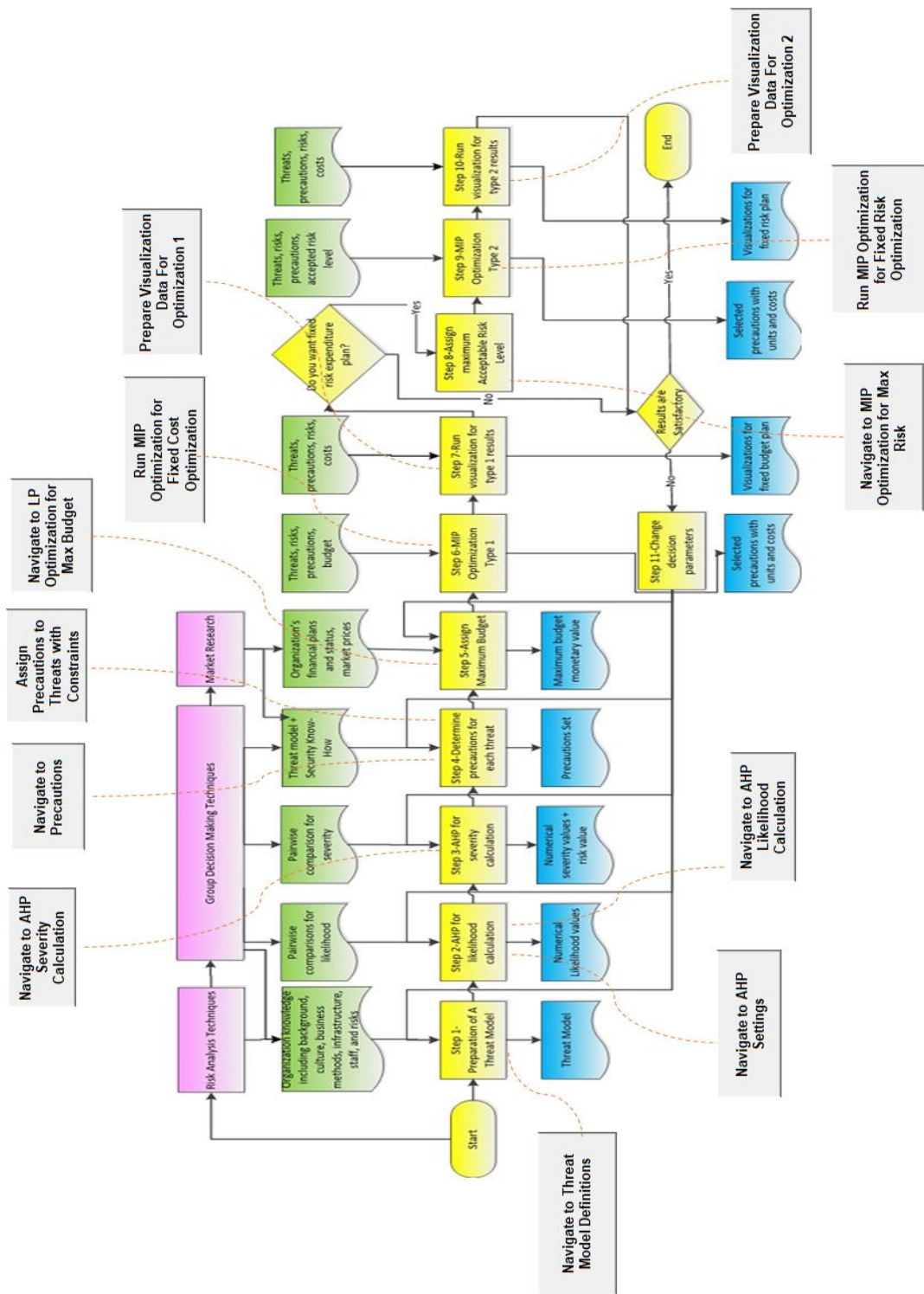


Figure 53 - Flow of activities in the proposed model

7.3 Methodology

The proposed DSS which is named Optimal Information Security Preventative Actions Along Visualization (OPISPA-V) integrates analytical hierarchical process, mixed integer programming and visualization techniques for optimal selection of enterprise information security preventative actions. Selection of AHP as a prioritization method in ratio scale over other risk evaluation techniques was sound to authors due to its ability to provide numerical prioritization values distributed in a wide range which can be input to MIP calculations later. Methods which provide values in nominal or ordinal scale would not be equally helpful in the later steps. However, other risk analysis techniques, such as graphical risk modeling techniques can be incorporated to comprehend and assess the hazard related points of interest.

AHP is a method which may be subjective in its design. In this study, it is used for expert judgment purposes which contain subjectiveness in nature. However, the consistency of the AHP pairwise comparisons is endured by the utilization of consistency checks. Ordinarily, the suggested value of consistency ratio for AHP pairwise comparisons is 0.1. MIP method is also included in the new DSS to determine the optimal investment solutions to the risks quantified based on the AHP method. Application of the proposed decision support system would require the necessary steps described in the flow diagram shown in Figure 53.

The size of the AHP matrix is identified with the number of threats defined in the threat model. This size has a direct effect on the number of pairwise comparisons made for AHP. The number of pairwise comparisons is equal to the combination of the matrix size. Nonetheless, having a layered structure decreases the number of comparisons. For example, a one level matrix with size seven requires 21 comparisons. If these seven items are grouped with an upper layer as three, two, two items, then the total number of comparisons required would be five plus three, five comparisons for lower level items and three comparisons for the upper level items. The size of the AHP matrix has an indirect effect on MIP calculations. For each precaution that may be assigned to a threat in a matrix row at most four constraints can be defined by the user. If the user decides to associate more than one precaution to a threat, then the number of constraints will increase accordingly. In general, the number of constraints which should be handled in a MIP session has a non-linear relation with the matrix size. A large matrix with sparse constraints may have the same runtime with a small matrix with many constraints for the precautions.

The proposed DSS requires identification of threatening risks for an enterprise and selection of alternative solutions. This would include prioritization of several risk items usually by multiple decision makers, and optimal allocation of the information security budgets. AHP method is integrated in the proposed DSS as it is a practical method, and can easily be implemented for prioritization of information security risks when there are

multiple decision makers. Utilizing group decision making (Dyer & Forman, 1992) techniques for the evaluation of threatening actions table may result in decisions that may be taken in phases. The proposed model allows changing of the decisions and rerunning of the model again and again or in subsequent phases due to low execution costs. If the decision makers are not satisfied with the results they may make changes, such as change of threat model, change of pairwise comparisons, change of precautions, change of precautions' constraints, change of budget, change of acceptable risk level. Each decision change can be made independently from each other and the subsequent steps should be executed accordingly.

The proposed system requires preparation of the precaution model. Usage of defense trees is quite common for analyzing security risks and representing alternative precautions (Bistarelli, Peretti, & Trubitsyna, 2007). The precautions model used in the proposed model is similar to defense trees in functionality but it is in tabular form, and it includes cost and unit fields for each alternative precaution.

The assignment of a predefined budget is required for the proposed model. The assignment of a predefined budget for the expenditure of the security costs is a decision problem which depends both on security-related factors, such as threats and costs of precautions, and external factors, such as financial plans and status of the enterprise. This assignment may be a result of an analytical process or intuitive thoughts. The assignment process has a highly interchangeable nature from enterprise to enterprise and thus was excluded from this study.

The emergence of big data in the last decade has increased the importance of information visualization. Information visualization simply forms an association between two information processing devices, human mind and computing devices (Gershon & Eick, 1997). Information visualization encapsulates numerous display techniques. Using the right display depends on the number of data points, number of data dimensions and applicability of the targeted use case for a particular display. The tree-node link diagrams grow too large and become useless when the depth of hierarchy and number of nodes are over some limit. A Treemap visualization is an alternative visualization technique which uses the whole visualization space without leaving blank areas and thus allows showing a large number of hierarchical nodes in a relatively small space. Basically, the Treemap diagrams which are used for visualization purposes in this study, consist of nested rectangular shapes. Borders, colors, and labels are used for a more comprehensible visual view of the nested elements.

OPISPA-V consists of 11 steps which are explained below:

Step 1: Preparation of a multi-level threat model is required which will be used in subsequent decision-making activities. Different kinds of enterprises having diverse missions, strategies, and environments would eventually result in different threat models. Either formal risk management techniques or a combination of general information

collection and open source intelligence techniques with brainstorming techniques could be used for the identification of security problems and corresponding risks. Threats may be grouped based on numerous factors including physical locations, security targets, departmental or functional divisions, and business processes. Different leveling criteria may end up with the identification of a diverse set of actions. Several attempts can be made in order to reach a threatening actions model covering the majority of possible risks that may cause harm to enterprise's assets and business functionalities. OPISPA-V also requires identification of alternative solutions for the detected risks.

Step 2: Pairwise comparisons of the likelihood of threats are made next. Likelihood and impact prioritization of security threats include both qualitative factors, such as survey results, user interviews, expert evaluation results, and quantitative factors, such the number of users, number of hosts, number of transactions and the monetary cost of exposure to the threat.

Step 3: After the pairwise comparison of the likelihood of threats, pairwise comparison of the severity of threats should be made for risk calculation. Application of AHP for severity prioritization is similar to the application of AHP for likelihood prioritization. The likelihood and severity of threats are determined independently by applying AHP twice.

Step 4: Once the likelihood and severity of threats are quantified, the precautions for each threat are determined, and the costs of precautions are estimated. During this identification of precautions, similar to Step 1, the enterprise characteristics affect the resulting precautions set. The precaution costs may involve a license cost, installation cost, or annual maintenance cost. Some precautions could be costless, such as a control activity which may not consume additional human resources. In such cases, the cost of precaution should be taken as zero. For all precautions, a unit should be identified. For most of the precautions, the unit would be one. An example from this group is a firewall device which would protect the enterprise's network. For some other risks, some particular precautions should be applied for more than once, such as installing a virus protection software on each host and server machine. The risk diminish factor in percentage for the precautions is also assigned in this step.

Step 5: Next, a budget constraint is specified for security expenditures to maximize the amount of risk prevented for a given fixed amount of budget by identifying the optimal set of precautions.

Step 6: Once the budget is specified, optimization Type 1, optimization of risk in fixed budget, is executed to determine the optimal precautions for the fixed budget option: The inputs of the MIP model are a fixed amount of budget, set of all threats, set of all precautions and their costs, sets of likelihood and severity values calculated using AHP. The objective of the first MIP model is the maximization of the amount of risk prevented without exceeding the amount of budget.

Step 7: The visualization for Type 1 is performed next, to examine the distributions of the results in the fixed budget option. In OPISPA-V, the Treemap visualization technique is used to visualize the distribution of costs or risks to threats or precautions. In step 7, the visualizations 1, 2, 3, and four which are given in Table 28 are utilized. Prior to using an information visualization technique in an enterprise, standardization should be made.

Step 8: In order to evaluate alternative expenditure plans a maximum risk level can be set. OPISPA-V enables the decision makers to determine the cost of eliminating the risks to a fixed level, such as 50% of the existing risks.

Step 9: Once the budget is specified, optimization Type 2, optimization of the budget in fixed risk level, can be executed to determine the optimal precautions and the necessary budget for a fixed risk level. The objective of the second MIP model is to determine the minimum enterprise information security budget for a given level of risk.

Step 10: Visualization Type 2 can be used next to examine the distributions in the fixed risk option. The visualizations 5, 6, 7, and eight that are shown in Table 28 are utilized in the visualization Type 2.

Step 11: The analysis can be repeated for different decision parameters. OPISPA-V enables easy identification and visualization of the impact of changes in the decision parameters.

Table 28 -Types of Visualizations Available in OPISPA-V

Fixed Budget Optimization		Fixed Risk Optimization	
1.	Distribution of Cost to Precautions Based on Fixed Budget Optimization	5.	Distribution of Cost to Threats Based on Fixed Risk Optimization
2.	Distribution of Risks to Precautions Based on Fixed Budget Optimization	6.	Distribution of Risks to Threats Based on Fixed Risk Optimization
3.	Distribution of Cost to Threats Based on Fixed Budget Optimization	7.	Distribution of Cost to Precautions Based on Fixed Risk Optimization
4.	Distribution of Risks to Threats Based on Fixed Budget Optimization	8.	Distribution of Risks to Precautions Based on Fixed Risk Optimization

7.4 Optimal Information Security Preventative Actions Along Visualization

The proposed decision support system is named Optimal Information Security Preventative Actions Along Visualization (OPISPA-V). Although the proposed methodology included 11 steps, external approaches which are suitable for the organizations' threat modeling activities are welcome and will not be further scrutinized in this section. This section aims to explain the main focuses of the proposed decision-making tool which are risk assessment of threats, optimal selection of enterprise information security preventive actions and the visualization of the outputs in detail.

7.4.1 Risk Assessment of Threats

In OPISPA-V the likelihood and severity of threats are evaluated by AHP for risk assessment. The security threats are considered as the starting point and are assessed in groups. Threats may be grouped based on physical locations, security targets, departmental or functional divisions, and business processes depending on the characteristics of the enterprise. A multi-level threat model including three threat groups and eight threats is illustrated in Figure 54. Once the threat groups and threats are defined, the likelihood and severity of threat groups and threats are evaluated in a hierarchical structure to perform the risk assessment based on the AHP (Saaty, 1988) method. In OPISPA-V, the security experts are asked to make a pairwise comparison of the severity and likelihood of threat groups, followed by the pairwise severity and likelihood comparisons of the threats within each group. During pairwise comparison, a 1-9 scale is used in which one indicated the equal significance of the two groups (or threats), and nine indicated the absolute dominance of one group (or threat) over another. In enterprises that have more than one security experts, if a consensus is not reached on pairwise comparisons, the comparisons can be performed by experts individually, and the individual judgments can be aggregated into a single representative judgment by using the geometric mean (Saaty, 2008).

The pairwise comparisons of the experts are used to quantify the level of severity and likelihood of the threats. OPISPA-V builds the matrix of pairwise comparisons of severity (*PSG*) for the threat groups as in Eq. (1):

$$PSG = \begin{bmatrix} 1 & sg_{12} & \dots & sg_{1n} \\ \frac{1}{sg_{12}} & 1 & \dots & sg_{2n} \\ \dots & \dots & \dots & \dots \\ \frac{1}{sg_{1n}} & \frac{1}{sg_{2n}} & \dots & 1 \end{bmatrix} \quad (1)$$

where; sg_{ij} is the pairwise comparison of severity among the Group- i and the Group- j , and n is the number of groups. The matrix of pairwise comparisons of likelihood (*PLG*) is built similarly. The normalized pairwise comparison matrices of severity (*NSG*) and likelihood (*NLG*) for groups are determined by Eq. (2).

$$NSG_{ij} = \frac{PSG_{ij}}{\sum_{i=1}^n PSG_{ij}}, \quad \forall i = \{1, 2, \dots, n\} \quad \text{and} \quad \forall j = \{1, 2, \dots, n\} \quad (2)$$

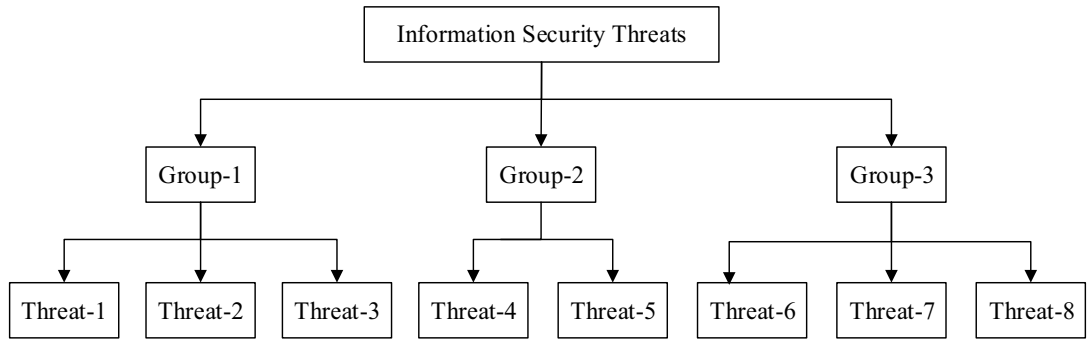


Figure 54 - Multi-level threat model

The weights (eigenvectors) for the severity (ESG) and likelihood (ELG) for groups are determined by Eq. (3).

$$ESG_i = \frac{\sum_{j=1}^n NSG_{ij}}{n}, \quad \forall i$$

$$= \{1, 2, \dots, n\} \quad (3)$$

In AHP method, since the comparisons are performed through subjective judgments, their consistency should be checked. Consistency ratio (CR) is used commonly to evaluate the consistency of the comparisons. A CR 0.1 or less is acceptable to continue the AHP analysis (Saaty, 1980). CR is calculated by Eq. 4, Eq. 5, and Eq. 6.

$$\begin{bmatrix} 1 & sg_{12} & \dots & sg_{1n} \\ \frac{1}{sg_{12}} & 1 & \dots & sg_{2n} \\ \dots & \dots & \dots & \dots \\ \frac{1}{sg_{1n}} & \frac{1}{sg_{2n}} & \dots & 1 \end{bmatrix} \begin{bmatrix} esg_1 \\ esg_2 \\ \dots \\ esg_n \end{bmatrix} = \lambda_{max} \begin{bmatrix} esg_1 \\ esg_2 \\ \dots \\ esg_n \end{bmatrix} \quad (4)$$

where; λ_{max} is the eigenvalue of NSG with the corresponding eigenvector ESG , and Consistency Index (CI) and CR are calculated as follows:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (5)$$

$$CR = \frac{CI}{RI} \quad (6)$$

where; RI is a random index that represents the judgments which have been entered randomly and are expected to be highly inconsistent. The RI values for matrices of different sizes (n) are as shown in Table 29 (Saaty, 1980).

Table 29 - Random Index(RI)

Matrix size(\	1	2	3	4	5	6	7	8	9	10
0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	

OPISPA-V checks the consistency of the pairwise comparisons of the experts for the severity and likelihood of the threat groups using the *CR*. In case of an inconsistency, the security experts are asked to revise the comparisons. Once the eigenvectors for the severity (*ESG*) and likelihood (*ELG*) for groups are determined, and consistency checks are made, the same procedure is repeated for the severity and likelihood of threats within groups. The eigenvectors for the severity (*ESW*) and likelihood (*ELW*) for threats within groups are determined similarly by using Eqs. (2, 3, 4). The overall severity (*ST*) and likelihood (*LT*) levels for the threats are calculated by multiplying the severity (likelihood) weight of the group of the threat with the severity (likelihood) weight of the threat within the group, Eq.(7):

$$ST_k = (ESG_{U(k)})(ESW_k), \quad \forall k = \{1, 2, \dots, m\} \quad (7)$$

in which; U(k) is the group number of threat k, and m is the number of threats. Finally, the magnitude of risks of threats (MT) is determined by multiplying their severity levels with their likelihood levels, Eq.(8):

$$MT_k = (ST_k)(LT_k), \quad \forall k = \{1, 2, \dots, m\} \quad (8)$$

7.4.2 Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization

In OPISPA, along with the security threats, the preventative actions (countermeasures) for decreasing or eliminating the risks of threats are also considered. OPISPA-V determines the optimal enterprise information security preventative actions using Mixed Integer Programming (MIP) technique. The security experts are asked to enter the information of preventative actions for each threat including their cost (*CP*) per one unit, their expected impact (*IP*) on the threat (decreased amount of risk for the threat by the preventative action) per one unit, and their maximum amount (*AP*) as shown in Table 30. Majority of the preventative actions will have a cost involving a license cost, installation cost or annual maintenance cost. Few preventative actions, such as a control activity may not require significant additional human and financial resources. For such cases, the cost of these actions can be taken as zero. For the majority of the preventative actions, the maximum quantity is one, but some preventative actions could be applied

for more than once, such as installing a virus protection software on each host and the server machine. For some threats, there may be several preventative actions. In this case, all actions can be included as shown in Table 30. For example, for Threat5, and Threat6 two preventative actions are included.

Table 30 - Preventative Actions

Threat	Preventative Action1(PA1)	Cost PA1	Impact of PA1	Maximum Amount1	Preventative Action2(PA2)	Cost PA2	Impact of PA2	Maximum Amount2
T1	PA11	100	0.40	1				
T2	PA21	100	0.40	1				
T3	PA31	200	0.30	1				
T4	PA41	100	0.30	2				
T5	PA51	50	0.05	10	PA52	150	0.30	2
T6	PA61	500	0.50	1	PA62	400	0.25	2
T7	PA71	100	0.35	2				
T8	PA81	600	0.20	3				

OPISPA-V first reports the magnitude of information security risks for the enterprise if no preventative action is taken based on the risk assessment results. OPISPA-V also reports the magnitude of risks when all of the preventative actions are considered along with the budget to achieve the minimum risk level. In OPISPA-V two different optimization models are available for selection of preventative actions. In the first model, the decision maker is asked to enter a budget limit. The optimal preventative actions which minimize the overall risk magnitude for a given budget limit are determined as in Eq (9):

$$\text{minimize } \sum_{k=1}^m \sum_{r=1}^s (MT_k)(IP_{kr}) x_{kr} \quad (9)$$

subject to:

$$\sum_{k=1}^m \sum_{r=1}^s (CP_{kr}) x_{kr} \leq B, \quad (10)$$

$$x_{kr} \leq AP_{kr}, \quad \forall k = \{1,2, \dots, m\} \quad \text{and} \quad \forall r = \{1,2, \dots, s\} \quad (11)$$

$$x_{kr} \geq 0, \quad \forall k = \{1,2, \dots, m\} \quad \text{and} \quad \forall r = \{1,2, \dots, s\} \quad (12)$$

where; IP_{kr} is the expected impact of r^{th} preventative action for threat k per one unit, x_{kr} is the amount of r^{th} preventative action for threat k , CP_{kr} is the unit cost of r^{th} preventative action for threat k , B is the budget limit, AP_{kr} is the maximum amount of r^{th} preventative action for threat k , and s is the number of preventative actions for threat k .

The second optimization model enables selection of preventative actions that minimize the costs for a target magnitude of information security risks. The optimal preventative actions which minimize the overall budget for a target risk magnitude are determined as follows:

$$\text{minimize } \sum_{k=1}^m \sum_{r=1}^s (CP_{kr}) x_{kr} \quad (13)$$

subject to:

$$\sum_{k=1}^m \sum_{r=1}^s (MT_k)(IP_{kr})x_{kr} \leq V, \quad (14)$$

$$\begin{aligned} x_{kr} &\leq AP_{kr}, \\ &= \{1, 2, \dots, s\} \end{aligned} \quad (15) \quad \forall k = \{1, 2, \dots, m\} \quad \text{and} \quad \forall r$$

$$\begin{aligned} x_{kr} &\geq 0, \\ &= \{1, 2, \dots, s\} \end{aligned} \quad (16) \quad \forall k = \{1, 2, \dots, m\} \quad \text{and} \quad \forall r$$

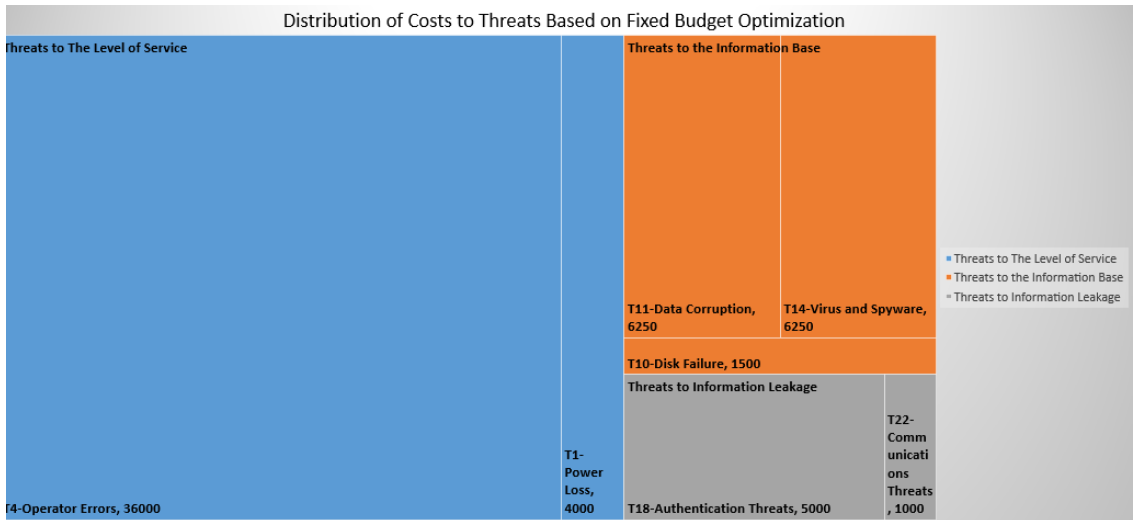
in which; V is the target magnitude for information security risks.

7.4.3 Visualization of Outputs

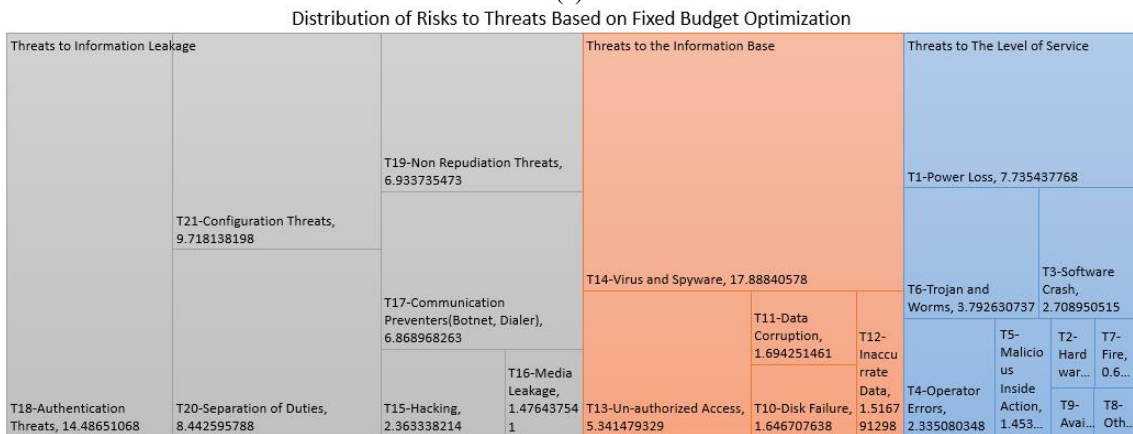
In OPISPA-V treemap technique is used to visualize the distribution of budget and risk to the selected (during the optimization process) optimal precautions. Figure 55 (a) shows a sample visualization of the distribution of cost to threats. In this view, the label of the diagonal shapes indicates the threats, the size of the diagonal shapes represents the monetary amount spent for a threat and color represents the threat grouping attribute. Figure 55 (b) shows a distribution of risks to threats where the size represents the level of risks calculated for each risk. Figure 55 (a) and Figure 55 (b) would help the decision makers to visualize the magnitude of risks assigned to different threatening actions, the value of investments made to prevent different risks. In Figure 55 for example, Group2/Threat 6 has the highest risk value, 0.5, but it has no cost (due to using free or opensource software and services). Group 1/Threat 3 has the second highest risk, and 2500 units of money were planned to be used to overcome that risk. Looking at these charts, it is also possible to calculate the total amount of risk and the total amount of money associated with each threat group. For example, it is possible to see that the risks associated with group 1 are higher than the risks associated with group 2 and group 3.

In OPISPA-V, treemap technique is also used to visualize the distribution of budget and risk to the selected (during the optimization process) precautions. In Figure 56 (a) the label of the diagonal shapes represents the precaution name and the budget, and the size

of the diagonal shapes represents the budget allocated for that precaution. Each precaution is associated with different color, and there is no grouping. In Figure 56 (b) the labels indicate the precaution names and the risk covered by that specific precaution. Figure 56 (a) and Figure 56 (b) would enable the decision makers to visualize the magnitude of the budget invested in different preventative actions and to visualize the preventative actions to prevent the risks.

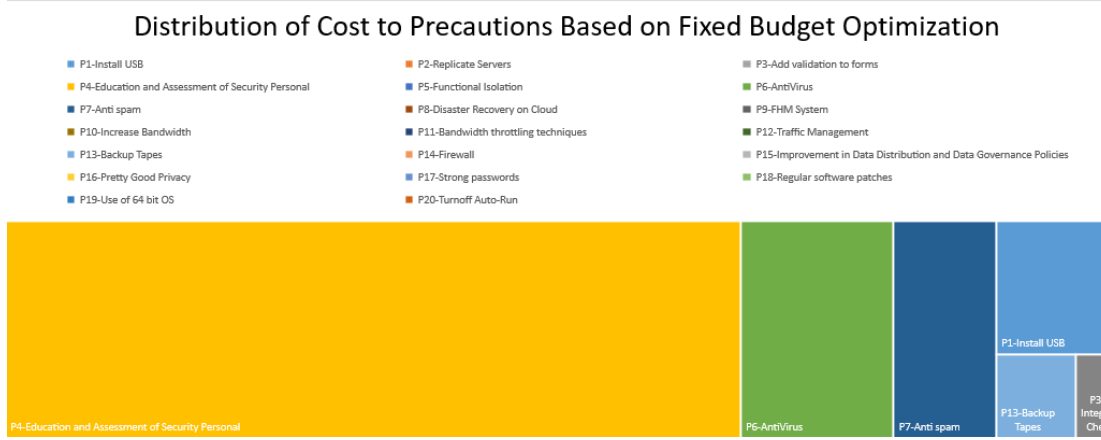


(a)

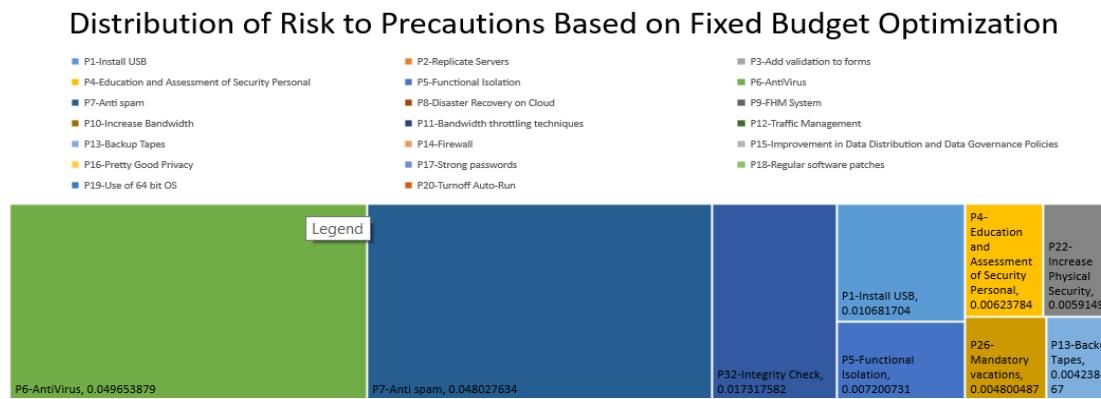


(b)

Figure 55 - Sample visualization of costs (a) and risks (b) distributions to threats using OPISPA-V



(a)



(b)

Figure 56 - Sample visualization of cost (a) and risks (b) distributions to precautions

7.5 OPISPA-V and Case Study

Developed OPISPA-V tool has been made publicly available on Github under the “OPISPA-V” project name to enable enterprises to determine the optimal information security preventative actions for their organization. In this section, the OPISPA-V tool is presented along with a case study. The case study is used to illustrate the benefits of the proposed decision support system. OPISPA-V tool consists of Excel sheets using embedded Excel formulas and, Visual Basic (VB) modules. Users need to activate the Solver add-in of Excel and input the threat, and the corresponding precaution sets to run the model for their specific goals.

OPISPA-V has 14 parts. The first part consists of the license information, a how-to-use guide and a Visual Menu Panel. In the second part, definitions of threats and threat groups are included. Parts 3, 4 and 5 comprise of AHP settings, AHP likelihood calculations, and AHP severity calculations, respectively. Parts 6, 7, and 8 include the precautions, MIP models for optimization Type 1, and summary of results of optimization Type 1. Parts 9 and part 10 compromise of visualizations of costs and risks

for the threats and precautions in optimization Type 1. Parts 11, and 12 include MIP models for optimization Type 2 and summary of results of optimization Type 2. Finally, Parts 13 and part 14 comprise of visualizations of costs and risks for the threats and precautions in optimization Type 2.

Table 31 - Threatening Actions for the Case Study Based on (Hunter, 2012)

	Power Loss	Power Loss	T1
	Hardware Failure	Hardware Failure	T2
	Software Crash	Software Crash	T3
	System Operators		
	Authorized Users		
	Programmers	Operator Errors	T4
	The Rest of The World		
Threats to The Level of Service		Malicious Inside Action	T5
	Viruses		
	Trojan	Trojan and Worms	T6
	Worms		
		Fire	T7
	Explosions		
	Floods	Other Environmental Disasters	T8
	Earthquakes		
		Availability Threats(DDOS)	T9
		Disk Failure	T10
		Data Corruption	T11
		Inaccurrate Data	T12
Threats to the Information Base	Covert Channel		
	Mandatory Access Control		
	Discretionary Access Control	Un-authorized Access	T13
	Physical Access Control		
		Virus and Spyware	T14
	Hacking		
	Keylogger	Hacking	T15
	Backdoor		
	Media Leakage		
	Theft of Media	Media Leakage	T16
Threats to Information Leakage	Tempest		
	Botnet	Communication Preventers(Botnet, Dialer)	T17
	Dialer		
		Authentication Threats	T18
		Non Repudiation Threats	T19
		Separation of Duties	T20
		Configuration Threats	T21
		Communications Threats	T22

A visual menu panel, as shown in Figure 53, is provided for the decision makers which shows the sequential order of actions and navigation to appropriate sheets along with triggering appropriate actions (macros) during this navigation. In the case study, an

enterprise which plans to grow and install a new IT system was selected. The current and planned business procedures, informational and technical assets besides technological know-how and economic strategies of the enterprise were gathered first. This enterprise had some primitive IT systems initially. However, as the enterprise was growing, it was moving to a new building and needed to increase its service capacity significantly. Hence, as part of this installment, a new acquirement plan for the security systems along with other software and hardware acquirements was necessary.

Hunter's (2012) threat classification was taken as a base and modified based on the organization's requirements. The security threats were identified under three categories: threats to the level of service, threats to the information base and threats to the information leakage, as shown in Table 31. Threats section of OPISPA-V which includes the threats of Table 31 is used for the definition of threats and threat groups for the case study. The RI values (Alonso & Lamata, 2006) included in the AHP settings sheet, show the acceptable consistency level and are used for AHP pairwise comparison consistency checks. In AHP sections both consistency checks and AHP calculations for threats and threat groups are done separately by OPISPA-V. The AHP results for the threat groups are shown in Figure 57.

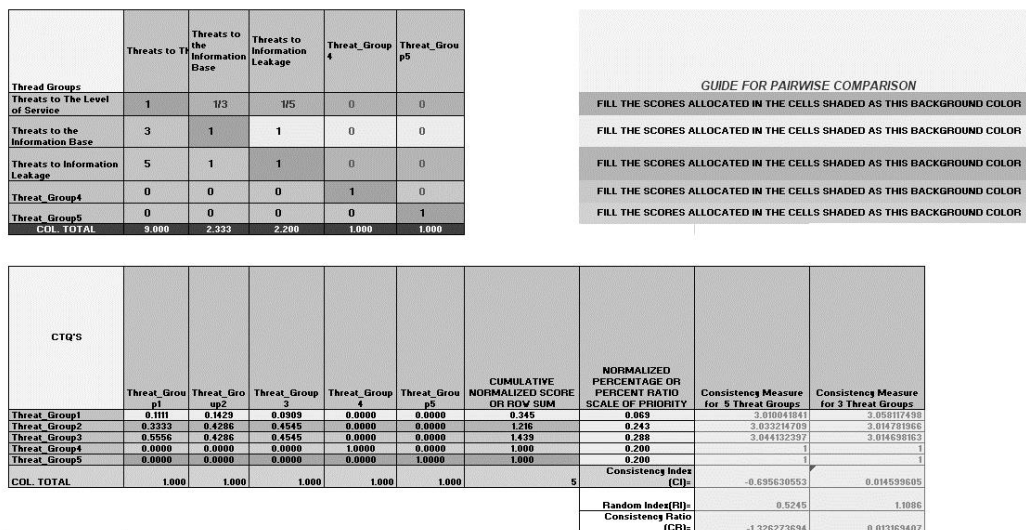


Figure 57 - AHP calculation and consistency check for threat groups

Pairwise comparison of threats was also made similarly for the threat groups. AHP calculations were also performed automatically by OPISPA-V tool for the threats and threat groups, as shown in Figure 58 OPISPA-V calculated the risk amount associated with each threat by multiplying the threat probability and threat severity (impact). At this step, a list of precautions was prepared for the case enterprise. This precautions list included 33 precautions (preventative actions), as shown in Table 33. For each threat, one or more precautions were selected for the case study. The estimated costs of precautions were also inputted into OPISPA-V as shown in Figure 59. Precautions can

also be new risk sources. For example, including backup tapes as a precaution may have its own security vulnerabilities. In this model, possible risks that may arise due to precautions were not included. However, periodically running the proposed model would enable handling such newly emerged risks. Some of the countermeasures may prevent more than one threat. In such a case, the precaution can be inserted into the model multiple times by dividing the cost of the precaution. Using mixed integer programming allows the definition of additional case-based constraints easily, without changing the overall model. For example, a precaution which requires the existence of another particular precaution can be defined easily by creating an additional constraint.

Table 34 and Table 35 show the results of OPISPA-V for the same threats, precautions, and precaution constraints for various options of budget and acceptable risk levels for the case study. All the inputs, outputs, and resulting visualizations for the case study are also available in the OPISPA-V Github site. Executing this DSS multiple times by endeavoring to distribute different budget amounts among different precaution sets or deciding on the appropriate risk demonstrates that this method can be utilized to determine the required budget amount for different acceptable risk levels.

Table 32 - Acceptable Comparison Values for AHP Severity and Likelihood Comparison

A	Extreme	B	9
A	Very Strong	B	7
A	Strong	B	5
A	Moderate	B	3
A	Equal	B	1
B	Moderate	A	1/3
B	Strong	A	1/5
B	Very Strong	A	1/7
B	Extreme	A	1/9
A	Not Related To	B	0

CUMULATIVE NORMALIZED SCORE OR ROW SUM	NORMALIZED PERCENTAGE OR PERCENT RATIO SCALE OF PRIORITY	Corresponding Threat Group	Threat Group Likelihood/Priority Calculation Result	Threat Likelihood Overall Result	Consistency Measure For Threats	Threat Names
3.522	0.160	ats to The Level of Ser	0.069	0.243	12.47310566	T1-Power Loss
0.341	0.015	ats to The Level of Ser	0.069	0.023	9.853549067	T2-Hardware Failure
1.304	0.059	ats to The Level of Ser	0.069	0.090	11.53007974	T3-Software Crash
0.975	0.044	ats to The Level of Ser	0.069	0.067	10.69107535	T4-Operator Errors
0.645	0.029	ats to The Level of Ser	0.069	0.044	9.071005251	T5-Malicious Inside Action
1.509	0.068	ats to The Level of Ser	0.069	0.104	11.36949722	T6-Trojan and Worms
0.242	0.011	ats to The Level of Ser	0.069	0.017	9.60521557	T7-Fire
0.175	0.008	ats to The Level of Ser	0.069	0.012	10.05027006	T8-Other Enviromental Disasters
0.341	0.015	ats to The Level of Ser	0.069	0.024	10.2219783	T9-Availability Threats(DDOS)
0.300	0.014	ats to the Information E	0.243	0.073	5.152014739	T10-Disk Failure
0.396	0.017	ats to the Information E	0.243	0.094	4.390662009	T11-Data Corruption
0.303	0.014	ats to the Information E	0.243	0.074	5.197480376	T12-Inaccurate Data
1.109	0.050	ats to the Information E	0.243	0.270	5.474856125	T13-Un-authorized Access
2.903	0.132	ats to the Information E	0.243	0.706	6.819690505	T14-Virus and Spyware
0.333	0.015	ats to Information Leak	0.288	0.096	8.241989245	T15-Hacking
0.204	0.009	ats to Information Leak	0.288	0.059	8.244852643	T16-Media Leakage
0.943	0.043	ats to Information Leak	0.288	0.271	8.277156096	T17-Communication Preventers(Botnet, Dialer)
2.322	0.092	ats to Information Leak	0.288	0.592	8.461704533	T18-Authentication Threats
0.995	0.045	ats to Information Leak	0.288	0.286	8.199100835	T19-Non Repudiation Threats
1.153	0.052	ats to Information Leak	0.288	0.332	8.406091634	T20-Separation of Duties
1.332	0.060	ats to Information Leak	0.288	0.383	8.479118091	T21-Configuration Threats
1.019	0.046	ats to Information Leak	0.288	0.293	8.396167946	T22- Communications Threats

Figure 58 - AHP calculation for threats

Table 33 - Counter Measures Selected for the Case Study

Name	Description	Name	Description	Name	Description
P1	Install UPS	P12	Traffic Management	P23	Service level software agreement, PAAS
P2	Replicate Servers	P13	Backup Tapes	P24	Auditing
P3	Add validation to forms in enterprise software	P14	Firewall	P25	Pre-employment screening
P4	Education and Assessment of Security Personal	P15	Improvement in Data Distribution and Data Governance Policies	P26	Mandatory vacations
P5	Functional Isolation	P16	Pretty Good Privacy	P27	Identity and Access Management (IAM) Systems
P6	Anti-Virus	P17	Strong passwords	P28	Encryption Solutions
P7	Anti-Spam	P18	Regular software patches	P29	Intrusion Detection and Prevention Systems
P8	Disaster Recovery Center on Cloud Systems	P19	Use of 64 bit OS	P30	Digital Signatures
P9	FHM System	P20	Turnoff Auto-Run	P31	Trusted Communication Channels
P10	Increasing bandwidth	P21	Load – Un-Load Drivers	P32	Integrity Check
P11	Bandwidth throttling techniques	P22	Increase Physical Security	P33	Limit Transfer of Executables

Table 34 - The Results of Running OPISPA-V for Fixed Budget Optimization for Two Budget Options.

	For 100000 budget	For 60000 budget
Budget Spent	99500	60000
Number of Threats Handled	16	13
Eliminated Risks(Tx)	T1,T3,T4,T5,T6,T9, T10,T11, T13, T14, T15, T16, T18, T19, T20, T22	T1,T3,T4,T5,T6,T10,T11,T14,T16,T18,T19,T20, T22
Percentage of Eliminated Risk over Total Risk	%75.29	%47.28
Selected Precautions by Optimization(Px(unit))	P1(1), P4(3), P5(1), P6(50), P7(50), P13(6), P14(1), P17(1), P18(1), P19(1), P20(1), P21(1), P22(1), P24(1), P25(1), P26(1), P27(2), P30(1), P32(1), P33(2)	P1(1), P4(3), P5(1), P6(50), P7(50), P13(3), P18(1), P19(1), P20(1), P21(1), P22(1), P25(1), P26(1), P27(1), P32(1), P33(2)

Table 35- The Results of Running OPISPA-V for the Case Study for Fixed Acceptable Risk Level Optimization for Two Risk Level Options

	Minimum Acceptable Risk Level % 100	Minimum Acceptable Risk Level % 70
Budget Spent	270700	89000
Number of Threats Handled	22	16
Eliminated Risks(Tx)	T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, T12, T13, T14, T15, T16, T17, T18, T19, T20, T21, T22	T1,T3,T4,T5,T6,T10,T11,T13,T14,T15,T16,T18,T19,T20,T21, T22
Percentage of Eliminated Risk over Total Risk	%100	%70
Selected Precautions by Optimization(Px(unit))	P1(1), P2(5), P4(3), P5(1), P6(50), P7(50), P8(2), P9(2), P10(6), P11(3), P12(1), P13(6), P14(1), P15(4), P16(3), P17(1), P18(1), P19(1), P20(1), P21(1), P22(2), P23(1), P24(2), P25(1), P26(1), P27(2), P28(1), P28(2), P29(1), P30(1), P31(5), P32(1), P33(2)	P1(1), P4(3), P5(1), P6(50), P7(50), P13(2), P17(1), P18(1), P19(1), P20(1), P21(1), P22(1), P24(1), P25(1), P26(1), P27(2), P30(1), P32(1), P33(2)

7.6 Discussion

Despite the fact that enterprises have limited budgets for information security investments and the majority of the actions for eliminating or decreasing the risks of threats to improve information security are costly, few studies have focused on the security costs and budgets for assessing and preventing security risks. This study presented a DSS based on AHP, MIP, and Treemap visualization methods. The DSS provides a new approach for information security risk assessment and prevention and enables consideration of costs of preventative actions and budget constraints.

7.6.1 Contributions

Contributions of this study to research and practice are four fold. First, the proposed DSS considers the preventative actions and budget constraints enabling determination of optimal precautions for a fixed budget. Earlier AHP models provide prioritization of items, but not consider the preventive actions or budget limits. The proposed DSS provides an approach to determine the optimal information security preventative actions for a given budget constraint. The DSS also determines the minimum enterprise information security budget for a given level of risk. Hence, the new DSS assists the enterprise decision makers not only for assessing the information security risks, but also for the prevention of risks, and enables optimal information security investment decisions.

In general, the studies which use AHP for information security related decision-making, take security properties such as privacy, integrity, and confidentiality as the comparison criteria. Hence, existing approaches do not handle the security requirements adequately due to the lack of a threat analysis method. In this DSS, security threats are taken as the starting point and investigated at three levels, which are threats to the level of service, threats to the information base and threats leading to information leakage. Hence, the second contribution of this DSS is that it integrates threat analysis to the AHP for management of the information security risks.

The third contribution is the inclusion of visualization as part of the decision making process. Visualization helps to identify the magnitude and distribution of information security investments and risks eliminated.

Finally, the DSS was demonstrated by a real-world case study and made publicly available on Github under the “OPISPA-V” project name to enable its practical use. The implementation of the DSS was aggregated in Excel using embedded Excel formulas, VB modules, and Excel graphics. OPISPA-V could be used by small and medium-size enterprises to manage information security risks, resources, and budgets. This DSS may be used independently or as a part of an existing security-risk management system.

7.6.2 Limitations and Future Research

Although in the case study a sample organization’s data was used and the tool was demonstrated using this data, the evaluation of a tool which deals with the set of threats and corresponding precautions for an enterprise is problematic based on two main reasons. The first reason is privacy. No enterprise would accept to share such critical information with third parties. Thus, it was not possible to evaluate the case study results with independent experts. The second reason is the time-consuming stages of the process, such as threat modeling. In order to get feedback from the users, the OPISPA-V was made available in Github. Volunteered and interested organizations can experiment with the tool to check the effectiveness after the publication of the study.

The proposed DSS comprises of some time-consuming stages which rely on analysis, inspection, and evaluation specifically during threat modeling, precaution modeling and market price research. However, the tasks related to threat modeling and selection of corresponding precautions should exist in any security management system. Therefore, time and effort given to those practices are unavoidable when the objective is to provide an optimal security solution for an enterprise.

In the AHP, each element in the hierarchy is considered to be independent of all the other elements. However, in real life conditions, there are many situations where the existence of one risk factor will affect the likelihood or impact of other risk factors. Hence, the proposed DSS has certain limitations in representing the severity and likelihood of the threats, in particular when there is a dependency between the threats. Analytic network process (ANP), (Saaty, 1999) may provide an alternative approach for future information security studies, which would allow considering the dependencies among elements of

the threat model and would permit including real-life dependencies among the risk factors.

The OPISPA-V is restricted to 22 threats, and five threat groups. MIP method which was integrated into the proposed DSS can obtain the optimal security preventative actions for the specified restrictions. Large size enterprises may need to consider more threats, and threat groups for information security management, which would make the optimization process complicated and may require a significant amount of time if MIP is used. Heuristic and meta-heuristic methods appears to be another promising area for future research to achieve optimal information security risk management solutions and decisions.

7.7 Concluding Remarks for a Decision Support System for Optimal Selection of Enterprise Information Security Preventative Actions along Visualization

This chapter presented a new DSS for information security risk management which not only focused on the prioritization of vulnerabilities but also considered preventative actions and budget constraints. The DSS supports enterprise information security decision-making activities related to risk identification, prioritization, and prevention. The DSS enables identifying the optimal precautions for a given budget, and also provides the minimum budget for the desired risk level. Hence, the new DSS provides an effective approach for information security decision makers to achieve the optimal combination of preventative actions with limited enterprise security budgets.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 Conclusions

In this thesis five goals have been accomplished: The existing literature work has been reviewed in detail. A survey has been prepared and conducted to gather security visualization requirements of the enterprises. Software design and implementation has been made which aims to provide an infrastructure for the visualization of enterprise security data in a generic and standardized manner. The available metrics have been examined as a part of literature work. Knowing the fact that the software applications using the web based access dominate in the enterprises, and were being used for diverse internal and external enterprise users, the metric set used for the monitoring of vulnerability scan results of these applications have been enlarged by offering new metrics. A prototype using dashboard display has been prepared for the evaluation of the proposed metrics. Later, selected SIEM systems have been evaluated for their custom visualization generation capabilities, due to their common containment of dashboard type of displays. In the last part of the study, a decision support system which aims to find out the optimum cost for the security expenditures for the enterprises has been modeled.

The notion of enterprise security visualization has appendages both to enterprise information security and enterprise cybersecurity concepts. The basic idea of it is to improve enterprise security analysis methods through the use of visualization. In this thesis, an in-depth study of enterprise security visualization requirements, data sources, and use-cases, tools, and techniques are included.

So far, enterprise security visualization concept was limited with the visualization of enterprise computing systems. Mainly, this corresponded to visualizations showing the topology of enterprise network elements, and users' interactions with these elements. The studies related to enterprise security visualization concept did not include an extensive analysis of enterprise security visualization requirements, and the domain literature was disorganized. In general, the security visualization designs proposed up to now presented their narrowed benefits for specific use-cases. These use-cases are not depicted for their gains for the enterprise users. Enterprise security visualization should not be bounded to a limited number of use-cases and data sources.

There are a number of security-related data sources which may be associated with various use-cases. The difficulty is these data sources have various types and formats. The variety of display techniques is high. Frequently, incorporating multiple data source in a new display technique for enterprise visualization system requires implementing new software. Motivated by the difficulties of implementing a new software for security data sources with various type and formats, a generic security visualization design was developed. This design provides a way to generically define data sources and a structure to integrate a large number of display types in a standardized manner.

Notably, there were gaps in security visualization domain. One of these gaps pointed out the topic of visualization of black-box vulnerability scan results for the web-based applications. In order to fill this gap, available metrics for this kind of data was enlarged by providing a data structure which combines alert, scan, application, project, and standards data. As a result of using this structure, about forty measures and metrics were visualized through a series of dashboard type prototypes. Users from the software development domain evaluated these prototypes. The results show that the proposed prototype was found highly useful regarding its decision informing, and difference detection capabilities.

Another gap was related to security management issues. The final aim of the thesis was to improve enterprise security management practices by offering a decision support system for the optimization of security costs. Security management was neglected in the security visualization domain. Security systems are costly in general. To install an enterprise security system a considerable amount of money is spent. These systems are also living systems which require regular updates. An analytical instrument which provides optimum expenditure for security costs would be very beneficial for the enterprises. To fill the gap in the security domain and deliver an instrument for optimum security costs a decision support system was designed. This system was validated through a real-world case study. Although analytical methods have been used to prioritize risks for the security domain, an end to end system from threats to security costs is an entirely novel approach. Using visualization on top of this concept increased its decision informing capability which would be beneficial for the security managers.

This thesis has methodological, systemic, and practical results. The methodological results include the systematic process proposed for the optimal selection of precautions, the systematic evaluation method used for the decision of migrating data to big data environment, and the SIEM evaluation scenario specifically created for this thesis. Literally systemic results deal with putting items/parts into a concept and establishing the relationships among them. The classifications provided in the literature review, graphical security visualization library, various evaluation results are systemic outputs gathered during this thesis. These systemic results would have effects on the overall process of enterprise visualization creation. The tools, metrics, and prototypes provided throughout the study are among the practical results. While the literature study and the survey outputs would aid corporate managers to understand the requirements and

difficulties related to enterprise security visualization and realize opportunities, the presented visualization infrastructure, and the visualization prototypes would lead to the discovery of their specific application areas.

8.2 Future Work

The thesis has several concepts which point out a series of future work topics. Firstly, as a part of the literature review, a graphical security visualization library was created. This library elements may be carried to Web, and may be extended by periodical examination of novel security visualization studies as a future work.

Second part of the study was the security visualization requirements survey. Although, the survey provided a good insight for the security visualization requirements of the enterprises, due to its length, the number of participants was limited. A shorter survey involving similar concepts can be prepared as a future work, and new ways of survey distribution can be considered in order to get the maximum benefit.

During the design and development part of the thesis, a generic security visualization infrastructure was presented. This system allows getting feedbacks from the users as part of an enterprise cyber-physical system. A future work topic which was foreseen related to this issue was automatic creation of visualizations using the feedbacks made by the users for the data sources and display types. This future study may take the burden of associating display types with the data elements from the users and provides the most suitable and probable associations based on the previous feedback information.

Automatic processing of the feedbacks was also suggested by the expert reviewers as a future work to provide automatic reporting functionalities. For this purpose a taxonomy of probable feedback topics related to enterprise security visualization elements, such as threats, vulnerabilities, data sources, and display types might be made as a future work. Besides a more established feedback structure, one of the reviewers suggested forming a more structured threat definition as a part of enterprise security visualization knowledgebase structure.

The reviews also included some suggestions for the generic enterprise security visualization study such as working with compressed files, including predefined metrics besides allowing user-defined metrics, and relying on multiple integration items for Spark. These design elements might be injected into the system smoothly as future work. The design may be extended with new parsers and other visualization libraries to form a commercial visualization product.

Another design and implementation topic was related to web application security visualization. A prototype was implemented using the Tableau software. This prototype excluded visualization of some suggested metrics/measures due to time and scope limitations. For example, average remediation latencies for the known alert types is left

as a future work due to high effort required to form these data from multiple vendors. Similarly, forming a classification mechanism for alerts based on their effects to sensitive information, their impacts on business, and relation to existing vulnerable components is left out of scope of this study.

Scenario based SIEM evaluation is offered to be repeated both periodically and for other use-cases. Scenario based SIEM evaluations provide better insight compared to feature based evaluations. Comparison of Business Intelligence tools, SIEM tools, and custom security visualization studies will cause improvements to security visualization studies.

The decision support system provided under the topic of security management visualization. It uses AHP for the prioritization of threats, and threat groups. AHP assumes that each element in a hierarchy is independent. Analytical Network Process (ANP) on the other hand (Saaty, 1999) allows considering dependencies among elements of the threat model. Thus, implementing a similar DSS system in the future based on ANP may provide an alternative which resembles real life constraints better. This DSS prototype was restricted to 22 threats and five threat groups. A larger version of this prototype may also be implemented to serve companies which have to consider more threats. The DSS integrated MIP method for optimization and treemap display type. Other optimization techniques and display types may be considered later.

REFERENCES

- Aagedal, J. O., Braber, F. D., Dimitrakos, T., Gran, B. A., Raptis, D., & Stolen, K. (2002). Model-Based Risk Assessment to Improve Enterprise Security. *Proceedings of the 6th International Enterprise Distributed Object Computing Conference* (p. 51). Washington, DC, USA: IEEE Computer Society.
- Abdullah, K., Lee, C., Conti, G., & Copeland, J. A. (2005). Visualizing network data for intrusion detection. *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop* (pp. 100-108). IEEE.
- Abdullah, K., Lee, C., Conti, G., Copeland, J. A., & Stasko, J. (2005). Ids rainstorm: Visualizing ids alarms.
- Abu Naser, S. S. (1999). Big O Notation for Measuring Expert Systems complexity. *Islamic University Journal*, 57. Retrieved 8 17, 2018, from <http://www.scs.ryerson.ca/~mth110/Handouts/PD/bigO.pdf>
- Acunetix. (2018). *Acunetix*. Retrieved 01 08, 2019, from <https://www.acunetix.com/>
- Acunetix. (2019, 8 1). *Acunetix*. Retrieved 01 08, 2019, from [Acunetix: https://www.acunetix.com/](https://www.acunetix.com/)
- Agarwal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th Very Large Data Bases Conference* (pp. 487-499). Burlington, MA, USA: Morgan Kaufmann.
- Alberts, C. J., Behrens, S. G., Pethia, R. D., & Wilson, W. R. (1999). *Operationally critical threat, asset, and vulnerability evaluation (OCTAVE) Framework Version 1.0*. Pittsburgh, USA: Carnegie Mellon Software Engineering Institute.
- AlienVault. (2018, 9 19). *AlienVault Unified Security Management*. (AlienVault) Retrieved 9 19, 2018, from <https://www.alienvault.com/products>
- Allen, D. (2003). *Ready for Anything: 52 Productivity Principles for Getting Things Done*. Newyork, USA: Penguin Books.
- Allen, D. (2015). *Getting things done: The art of stress-free productivity*. Newyork,USA: Penguin Books.

- Allen, J. H. (2007). *Governing for Enterprise Security (GES) Implementation Guide*. Pittsburgh, USA: Carnegie Mellon University.
- Alonso, J. A., & Lamata, T. M. (2006). Consistency in the analytic hierarchy process: a new approach. *International journal of uncertainty, fuzziness and knowledge-based systems*, 445-459.
- Alsaleh, M., Alarifi, A., Alqahtani, A., & Al-Salman, A. (2015). Visualizing web server attacks: patterns in PHPIDS logs. *Security and Communications Networks*, 8(11), 1991-2003.
- Alsaleh, M., Barrera, D., & Van Oorschot, P. C. (2008). Improving security visualization with exposure map filtering. *Annual Computer Security Applications Conference, 2008* (pp. 205-214). IEEE.
- Anderson, E. E., & Choobineh, J. (2008). Enterprise information security strategies. *Computers & Security*, 27, 22-29.
- Apache. (2018, 6 12). *OLTP and operational Analytics for Apache Hadoop*. Retrieved from Apache Phoenix : <https://phoenix.apache.org>
- Apache. (2018, 6 12). *Unified analytics engine for large-scale data processing*. Retrieved from Apache Spark: <https://spark.apache.org>
- Apache. (2018, 6 12). *Web-based notebook taht enables dta-driven, interactive data analytics and collaborative documents with SQL, Scala and more*. Retrieved from Apache Zeppelin: <https://zeppelin.apache.org>
- Arunraj, N. S., & Maiti, J. (2007). Techniques and methodologies of risk analysis. *Journal of Hazardous Materials*, 11(4), 653–661.
- Assal, H., Chiasson, S., & Biddle, R. (2016). Cesar: Visual Representation of Source Code Vulnerabilities. *IEEE Symposium on Visualization fro Cyber Security*. Baltimore, MD, USA.
- Ball, R., Fink, G. A., & North, C. (2004). Home-centric visualization of network traffic for security administration. *Proceedings of the 2004 ACM workshop on Visualization and Data Mining for Computer Security* (pp. 55-64). ACM.
- Ballora, M., & Hall, D. L. (2010). Do you see what I hear: experiments in multi-channel sound and 3D visualization for network monitoring? *Proceedings Volume 7709, Cyber Security, Situation Management, and Impact Assessment II; and Visual Analytics for Homeland Defense and Security II*. 7709, p. 0J. Orlando, Florida, United States: SPIE. doi:10.1117/12.850319

- Basharat, I., Azam, F., & Muzaffar, A. W. (2012). Database Security and Encryption: A Survey Study. *International Journal of Computer Applications*, 47(12), 0975 – 888.
- Best, D. M., Hafen, R. P., Olsen, B. K., & Pike, W. A. (2011). Atypical behavior identification in large-scale network traffic. *IEEE Symposium on Large Data Analysis and Visualization* (pp. 15-22). IEEE.
- Bingham, M., Skillen, A., & Somayaji, A. (2014). Even Hackers Deserve Usability: An Expert Evaluation of Penetration Testing Tools. *9th Annual Symposium on Information Assurance*. Albany, NY, USA.
- Bistarelli, S., Peretti, P., & Trubitsyna, I. (2007). Analyzing Security Scenarios Using Defence Trees and Answer Set Programming. *Proceedings of the 3rd International Workshop on Security and Trust Management*. 197, pp. 121-129. Dresden, Germany: Elsevier.
- BlackStratus. (2018, 9 19). *BlackStratus Managed Security Services*. (BlackStratus) Retrieved 9 19, 2018, from <https://www.blackstratus.com/>
- Bodin, L. D., Gordon, L. A., & Loeb, M. P. (2005). Evaluating information security investments using the analytic hierarchy process. *Communications of the ACM*, 48(2), 78-83.
- Bogen, A. C., Dampier, D. A., & Carver, J. C. (2007). Support for computer forensics examination planning with domain: a report of one experiment trial. *40th. Annual Hawaii International Conference on System Sciences, 2007* (pp. 267b-267b). IEEE.
- Bostock, M. (2018, 8 7). *Data Driven Documents*. Retrieved 8 7, 2018, from <https://d3js.org/>
- Bousquet, A., Clemente, P., & Lalande, J. F. (2011). SYNEMA: Visual monitoring of network and system security sensors. *Proceedings of the International Conference on Security and Cryptography* (pp. 375-378). IEEE.
- Breier, J. (2014). Security evaluation model based on the score of security mechanisms. *Information Sciences and Technologies*, 6(1), 19.
- Brotby, K. W. (2009). *Information Security Management Metrics A Definitive Guide to Effective Security Monitoring and Measurement*. Auerbach Publications.
- Brunk, C., Kelly, J., & Kohavi, R. (1997). MineSet: An Integrated System for Data Mining. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, (pp. 135-138).

- Çalışkan, M., & Varaksin, O. (2013). *PrimeFaces Cookbook*. Birmingham, UK: Packt Publishing Ltd.
- Campbell, G. A., & Papapetrou, P. P. (2013). *Sonarqube in Action*. Greenwich, CT, USA: Manning Publications Co.
- Chang, B. H., & Jeong, C. Y. (2011). An efficient network attack visualization using security quad and cube. *Electronics and Telecommunications Research Institute Journal*, 33(5), 770-779.
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., . . . Gruber, R. E. (2008). Bigtable: A Distributed Storage System for Structured Data. *ACM Transactions on Computer Systems* , 26(2), 4.
- Coelli, T. J., Rao, D. P., O'Donnell, C. J., & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Newyork, USA: Springer.
- Cohen, J., & Acharya, S. (2013). Towards a more secure apache hadoop hdfs infrastructure. *International Conference on Network and System Security* (pp. 735-741). Berlin, Heidelberg: Springer.
- Colitti, L., Di Battista, G., Mariani, F., Patrignani, M., & Pizzonia, M. (2005). Visualizing Interdomain Routing with BGPlay. *Journal of Graph Algorithms and Applications*, 9(1), 117-148.
- Conner, B., Noonan, T., & Holleyman, R. (2003). *Information security governance: Toward a framework for action*. International: Business Software Alliance.
- Conti, G., & Abdullah, K. (2004). Passive visual fingerprinting of network attack tools. *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security* (pp. 45-54). ACM.
- Conti, G., Abdullah, K., Grizzard, J., Stasko, J., Copeland, J. A., Ahamad, M., . . . Lee, C. (2006). Countering Security Analyst and Network Administrator Overload Through Alert and Packet Visualization. *IEEE Computer Graphics and Applications*, 26(2), 60-70.
- Cremonini, M., & Riccardi, M. (2009). The Dorothy Project: An Open Botnet Analysis Framework for Automatic Tracking and Activity Visualization. *2009 European Conference on Computer Network Defense (EC2ND)*. Milano, Italy. doi:10.1109/EC2ND.2009.15
- Crossler, R. E., Bélanger, F., & Ormond, D. (2017). The quest for complete security: An empirical analysis of users' multi-layered protection from security threats. *Information Systems Frontiers* , 1-15.

- Cuihua, X., & Jiajun, L. (2009). An Information System Security Evaluation Model Based on AHP and GRAP. *International Conference on Web Information Systems and Mining* (pp. 493 - 496). Shanghai: IEEE.
- Curtis, P., & Carey, M. (2012). *Risk Assessment in Practice*. Deloitte & Touche LLP, Durham, NC.
- D3.js. (2018, 6 12). *D3*. Retrieved from Dara Driven Documents: <https://d3js.org>
- Dang, T. T., & Dang, T. K. (2014). An Extensible Framework for Web Application Vulnerabilities Visualization and Analysis. In *Future Data and Security Engineering* (pp. 86-96). doi:10.1007/978-3-319-12778-1_7
- Dang, T. T., & Dang, T. K. (2014). Visualizing Web Attack Scenarios in Space and Time Coordinate Systems. *Transactions on Large-Scale Data and Knowledge Centered Systems XVI*, 1-14.
- de Oliveira Alves, G. A., da Costa Carmo, L. R., & de Almeida, A. (2006). Enterprise Security Governance; A practical guide to implement and control Information Security Governance. *IEEE/IFIP Business Driven IT Management*, 71-80.
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 107-113.
- Dell. (2018, 9 19). *RSA At a Glance*. (Dell) Retrieved 9 18, 2018, from <https://www.rsa.com/en-us/customers/dell-technologies>
- Demir, K. A. (2009). A Survey on Challenges of Software Project Management. *Software Engineering Research and Practice*, (pp. 579-585). Las Vegas, Nevada, USA.
- Dhillon, G., & Backhouse, J. (2000). Information System Security Management in The New Millenium. *Communications of the ACM*, 43(7).
- Dimitrios, K. (2014). *Security Information and Event Management Systems: Benefits and Inefficiencies*. Piraeus, Greece: University of Piraeus.
- Disterer, G. (2013). Iso/iec 27000, 27001 and 27002 for information security management. *Journal of Information Security*, 4(2). doi:10.4236/jis.2013.42011
- Disterer, G. (2013). ISO/IEC 27000, 27001 and 27002 for information security management. *Journal of Information Security*, 92-100.
- Dyer, R. F., & Forman, E. H. (1992). Group decision support with the Analytic Hierarchy Process. *Decision Support Systems*, 8(2), 99-124.

- Eloff, J. H., Labuschagne, L., & Badenhorst, K. P. (1993). A comparative framework for risk analysis methods. *Computers & Security*, 12(6), 597-603.
- Endert, A., North, C., Chang, R., & Zhou, M. (2014). Toward usable interactive analytics: Coupling cognition and computation. *KDD 2014 Workshop on Interactive Data Exploration and Analytics*. New York City, NY US.
- Erbacher, R. F. (2003). Intrusion behavior detection through visualization. *IEEE International Conference on Systems, Man and Cybernetics*. 3, pp. 2507-2513. IEEE.
- Erbacher, R. F., Christensen, K., & Sundberg, A. (2005, October). Designing visualization capabilities for ids challenges. *IEEE Workshop on Visualization for Computer Security* (pp. 121-127). IEEE.
- Erbacher, R. F., Walker, K. L., & Frincke, D. A. (2002). Intrusion and misuse detection in large-scale systems. *Computer Graphics and Applications*, 22(1), 38-47.
- Eren Doğu, Z., & Çelikoğlu, C. C. (2012). Information Security Risk Assessment: Bayesian Prioritization for AHP Group Decision Making. *International Journal of Innovative Computing, Information and Control*, 8(11), 8019-8032.
- Ericson, C. A., & Li, C. (1999). Fault tree analysis. *System Safety Conference*, (pp. 1-9). Orlando, Florida.
- EventTracker. (2018, 9 17). *Event Tracker*. (EventTracker) Retrieved 9 17, 2018, from <https://www.eventtracker.com/>
- Exabeam. (2018, 9 19). *The Exabeam Security Management Platform*. (Exabeam) Retrieved 9 19, 2018, from <https://www.exabeam.com/product/>
- Feng, N., Wang, H. J., & Li, M. (2014). A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis. *Information sciences*, 57-73.
- Ferebee, D., & Dasgupta, D. (2008). Security Visualization Survey. *Proceedings of the 12th Colloquium for Information Systems Security Education*. Dallas, TX.
- Fessi, B. A., Benabdallah, S., Boudriga, N., & Abd Shukor, M. H. (2014). A multi-attribute decision model for intrusion response system. *Information Sciences*, 237-254.
- Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2016). Decision support approaches for cyber security investment. *Decision support systems*, 13-23.

- Fink, G. A., Muessig, P., & North, C. (2005). Visual correlation of host processes and network traffic. *IEEE Workshop on Visualization for Computer Security* (pp. 11-19). IEEE.
- Finne, T. (1998). A Conceptual Framework for Information Security Management. *Computers and Security*, 303-307.
- FireEye. (2018, 9 19). *FireEye Leading The Way* . (FireEye) Retrieved 9 19, 2018, from <https://www.fireeye.com/>
- Fischer, F., Mansmann, F., Keim, D. A., Pietzko, S., & Waldvogel, M. (2008). Large-scale network monitoring for visual analysis of attacks. In *Visualization for Computer Security* (pp. 111-118). Berlin Heidelberg: Springer.
- Forman, E. H., & Gass, S. I. (2001). The Analytic Hierarchy Process—An Exposition. *Operations Research*, 49(4), 469 - 486.
- Fortinet. (2018, 9 19). *Fortinet Featured Security Insights & Information*. (Fortinet) Retrieved 9 19, 2018, from <https://www.fortinet.com/>
- Fox, D. (2006). Open web application security project. *Datenschutz und Datensicherheit*, 30(10), 636-636.
- Frincke, D. A., Gates, C. E., & Goodall, J. R. (2009). Message from the Workshop Chairs. *6th International Workshop on Visualization for Cyber Security* (pp. iv-v). Atlanta, GA, USA: IEEE.
- Fry, B. (2007). *Visualizing data: exploring and explaining data with the Processing environment*. O'Reilly Media, Inc.
- Garg, N. (2013). *Apache Kafka*. Birmingham, UK: Packt Publishing.
- Gass, S. I. (1958). *Linear programming*. John Wiley & Sons.
- Geolocation Map*. (3, July 2009). (SecViz- Security Visualization) Retrieved May 2016, 19, from <http://secviz.org/content/geolocation-map>
- Gershon, N., & Eick, S. G. (1997). Information Visualization. *IEEE Computer Graphics and Applications*, 4, 29-31.
- Getting Started*. (2018, 6 12). Retrieved from R Project for Statistical Computing: <https://www.r-project.org>
- Ghoniem, M., Shurkhovetsky, G., Bahey, A., & Otjacques, B. (2014). VAFLE: visual analytics of firewall log events. *Published in SPIE Proceedings. 9017*. International Society for Optics and Photonics. doi:10.1117/12.2037790

- Gibbons, R. (1992). *A primer in game theory*. Prentice Hill.
- Girardin, L. (1999). An Eye on Network Intruder-Administrator Shootouts. *Workshop on Intrusion Detection and Network Monitoring*, (pp. 19-28).
- Gomory, R. E. (1963). An algorithm for integer solutions to linear programs. *Recent advances in mathematical programming*, 64, 260-302.
- Goodall, J. R., Lutters, W. G., Rheingans, W. G., & Komlodi, A. (2005). Preserving the big picture: Visual network traffic analysis with tnv. *IEEE Workshop on InVisualization for Computer Security* (pp. 47-54). IEEE.
- Goodall, J. R., Radwan, H., & Halseth, L. (2010). Visual Analysis of Code Security. *Proceedings of the Seventh International Symposium on Visualization for Cyber Security* (pp. 46-51). New York, NY, USA: ACM. doi:10.1145/1850795.1850800
- Goodall, J. R., Radwan, H., & Halseth, L. (2010). Visual Analysis of Code Security. *Proceedings of the Seventh International Symposium on Visualization for Cyber Security* (pp. 46-51). New York, NY, USA: ACM. doi:10.1145/1850795.1850800
- Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security*, 438-457.
- Gugelmann, D., Gasser, F., Ager, B., & Lenders, V. (2015). Hviz: HTTP(S) traffic aggregation and visualization for network forensics. *Proceedings of the Second Annual DFRWS Europe. 12*, pp. 1-11. Elsevier. doi:10.1016/j.diin.2015.01.005
- Gugelmann, D., Gasser, F., Ager, B., & Lenders, V. (2015). Hviz: HTTP(S) traffic aggregation and visualization for network forensics. *Proceedings of the Second Annual DFRWS Europe. 12*, pp. 1-11. Elsevier. doi:10.1016/j.diin.2015.01.005
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *11*(1), 10-18.
- Halleen, G., & Kellogg, G. (2007). *Security monitoring with cisco security mars*. Boston, MA, USA: Pearson Education.
- Haloi, S. (2015). *Apache Zookeeper Essentials*. Birmingham, UK: Packt Publishing.
- Harrison, L., & Lu, A. (2012). The future of security visualization: Lessons from network visualization. *Network*, 26(6), 6-11.

- Harrison, L., Spahn, R., Iannacone, M., Downing, ., & Goodall, J. R. (2012). NV: Nessus vulnerability visualization for the web. *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, (pp. 25-32). Seattle, Washington, USA. doi:10.1145/2379690.2379694
- Harrison, L., Spahn, R., Iannacone, M., Downing, E., & Goodall, J. R. (2012). NV: Nessus vulnerability visualization for the web. *Proceedings of the ninth international symposium on visualization for cyber security* (pp. 25-32). ACM.
- Harrison, L., Spahn, R., Iannacone, M., Downing, E., & Goodall, J. R. (2012, October). NV: Nessus vulnerability visualization for the web. *Proceedings of the ninth international symposium on visualization for cyber security* (pp. 25-32). ACM.
- Herrmann, D. S. (2002). *Using the Common Criteria for IT Security Evaluation*. New York,USA: Taylor & Francis.
- Hilliard, R. (2000). Ieee-std-1471-2000 recommended practice for architectural description of software-intensive systems. *IEEE Standards*, pp. 16-20.
- Hofstadter, D. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid*.
- Hortonworks. (2018). Hortonworks Sandbox on a VM.
- Humble Software. (2018, 8 7). *Flotr2 JavaScript visualization library*. Retrieved 8 7, 2018, from <http://www.humblesoftware.com/flotr2/>
- Hunter, J. M. (2012). *An information security handbook* (Science & Business Media ed.). Springer.
- Hwang, C.-L., & Lin, M.-J. (2012). *Group Decision Making under Multiple Criteria: Methods and Applications*. Newyork,USA: Springer-Verlag.
- IBM. (2018, 9 17). *IBM QRadar SIEM*. Retrieved from IBM: <https://www.ibm.com/tr-tr/marketplace/ibm-qradar-siem>
- Imperva. (2018, 5 19). *The State of Web Application Vulnerabilities in 2017*. Retrieved from Imperva: <https://www.imperva.com/blog/2017/12/the-state-of-web-application-vulnerabilities-in-2017>
- Istvan, R. L. (1992). A new productivity paradigm for competitive advantage. *Strategic Management Journal*, 13(7), 525-537.
- ITGI. (2006). *Information Security Governance: Guidance for Boards of Directors and Executive Management 2nd Edition*. Rolling Meadows, USA: ITGI.
- Jain, A. (2013). *Instant Apache Sqoop*. Birmingham/UK: Packt Publishing Ltd.

- Janies, J. (2008). Existence plots: A low-resolution time series for port behavior analysis. *Visualization for Computer Security*, 161-18.
- Johnson, B., & Shneiderman, B. (1991). Tree-maps: A space-filling approach to the visualization of hierarchical information structures. *IEEE Conference on Visualization Proceedings* (pp. 284-291). IEEE.
- Johnson, B., Song, Y., Murphy-Hill, E., & Bowdidge, R. (2013). Why don't software developers use static analysis tools to find bugs? *International Conference on Software Engineering*. San Francisco, CA, USA.
- Johnson, R., Höller, J., Arendsen, A., Risberg, T., & Sampaleanu, C. (2009). *Professional Java development with the Spring framework*. John Wiley & Sons.
- Josephsen, D. (2007). *Building a monitoring infrastructure with Nagios*. Upper Saddle River, NJ, USA: Prentice Hall.
- Kan, Z., Hu, C., Wang, Z., Wang, G., & Huang, X. (2010). NetVis: A network security management visualization tool based on treemap. *2010 2nd. International Conference on Advanced Computer Control*. 4, pp. 18-21. IEEE.
- Kane, F. (2017). *Frank Kane's Taming Big Data with Apache Spark and Python*. Birmingham, UK: Packt Ltd.
- Kavanagh, K., & Bussa, T. (2017). *Magic Quadrant for Security Information and Event Management*. Stamford, CT, USA: Gartner.
- Keim, D. A., Mansmann, F., Schneidewind, J., & Schreck, T. (2006). Monitoring network traffic with radial traffic analyzer. *IEEE Symposium On Visual Analytics Science And Technology* (pp. 123-128). IEEE.
- Kim, S., & Lee, H. J. (2007). A study on decision consolidation methods using analytic models for security systems. *Computers & Security*, 26(2), 145-153.
- Kintzel, C., Fuchs, J., & Mansmann, F. (2011). Monitoring large ip spaces with clockview. *Proceedings of the 8th international symposium on visualization for cyber security* (p. 2). ACM.
- Kintzel, C., Fuchs, J., & Mansmann, F. (2011, July). Monitoring large ip spaces with clockview. *Proceedings of the 8th international symposium on visualization for cyber security* (p. 2). ACM.
- Kletz, T. A. (1997). Hazop—past and future. *Reliability Engineering & System Safety*, 55(3), 263-266.

- Koike, H., & Ohno, K. (2004). SnortView: visualization system of snort logs. *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security* (pp. 143-147). ACM.
- Koike, H., Ohno, K., & Koizumi, K. (2005). Visualizing cyber attacks using IP matrix. *IEEE Workshop on Visualization for Computer Security* (pp. 91-98). IEEE.
- Komlodi, A., Rheingans, P., Ayachit, U., Goodall, J. R., & Joshi, A. (2005). A user-centered look at glyph-based security visualization. *IEEE Workshop on Visualization for Computer Security*. Minneapolis, MN, USA. doi:10.1109/VIZSEC.2005.1532062
- Konda, M. (2014). *Just Hibernate: A Lightweight Introduction to the Hibernate Framework*. CA,USA: O'Reilly.
- Krasser, S., Conti, G., Grizzard, J., Gribschaw, J., & Owen, H. (2005). Real-time and forensic network data analysis using animated and coordinated visualization. *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop*, (pp. 42-49).
- Kyung, T., Kim, K., Kim, S., & Song, Y. (2011). A Study on Information Security Level Evaluation Using Fuzzy AHP. *International Conference on Information Science and Applications* (pp. 1-6). Jeju Island: IEEE. doi:10.1109/ICISA.2011.5772350
- Laboratory, M. L. (1998-1999). Darpha Intrusion Detection Data Sets. Lincoln.
- Lad, M., Massey, D., & Zhang, L. (2006). Visualizing internet routing changes. *IEEE Transactions on Visualization and Computer Graphics*, 12(6), 1450-1460.
- Lai, Q., Zhou, C., Ma, H., Wu, Z., & Chen, S. (2015). Visualizing and characterizing DNS lookup behaviors via log-mining. *Neurocomputing*, 169, 100-109.
- Lai, Z., Shen, Y., & Zhang, G. (2016). A Security Risk Assessment Method of Website Based on Threat Analysis Combined with AHP and Entropy Weight. *IEEE International Conference on Software Engineering and Service Science* (pp. 481-484). Beijing, China: IEEE.
- Lakkaraju, K., Yurcik, W., & Lee, A. J. (2004). NVisionIP: netflow visualizations of system state for security situational awareness. *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security* (pp. 65-72). ACM.
- Langton, J. T., & Newey, B. (2010, April). Evaluation of current visualization tools for cyber security. *SPIE Defense, Security, and Sensing*(International Society for Optics and Photonics), 770910-770910.

- Lau, S. (2004). The Spinning Cube of Potential Doom. *Communications of the ACM Commun*, 47(6), 25-26. doi:10.1145/990680.990699
- Laursen, O. (2018, 8 7). *FlotCharts*. Retrieved 8 7, 2018, from <http://www.flotcharts.org>
- Lee, C. P., Tros, J., Gibbs, N., Beyah, R., & Copeland, J. A. (2005). Visual firewall: real-time network security monitor. *IEEE Workshop on Visualization for Computer Security* (pp. 129-136). IEEE.
- Lee, E. A. (2015). The Past, Present and Future of Cyber-Physical Systems: A Focus on Models. *Sensors*, 4837-4869.
- Lee, M.-C. (2014). Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method. *International Journal of Computer Science & Information Technology*, 6(1).
- Li, X., Wang, Q., Yang, L., & Luo, X. (2012). The research on network security visualization key technology. *2012 Fourth International Conference on Multimedia Information Networking and Security*, (pp. 983-988).
- Liao, Q., Blaich, A., Striegel, A., & Thain, D. (2008). ENAVis: Enterprise Network Activities Visualization. *Proceedings of the 22nd conference on Large installation system administration conference*, (pp. 59-74).
- Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Wei, Z. (2017). A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications. *IEEE Internet of Things Journal*, 4(5), 1125 - 1142.
- Littlewood, B. (1975). A reliability model for systems with Markov structure. *Applied Statistics*, 172-177.
- Livnat, Y., Agutter, J., Moon, S., Erbacher, R. F., & Foresti, S. (2005). A visualization paradigm for network intrusion detection. *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop*, (pp. 92-99).
- LogRhythm. (2018, 9 17). *LogRhythm Security Made Smarter*. Retrieved 9 17, 2018, from <https://logrhythm.com/>
- Lu, A., Wang, W., Dnyate, A., & Hu, X. (2011). Sybil Attack Detection through Global Topology Pattern Visualization. *Information Visualization*, 10(1), 32-46. doi:10.1057/ivs.2010.1
- Lu, L. F., Zhang, J. W., Huang, M. L., & Fu, L. (2010). A new concentric-circle visualization of multi-dimensional data and its application in network security. *Journal of Visual Languages & Computing*, 21(4), 194-208.

- Luse, A. (2009). *Exploring utilization of visualization for computer and network security*. Iowa State University.
- ManageEngine. (2018, 9 19). *ManageEngine*. (ManageEngine) Retrieved 9 19, 2018, from <https://www.manageengine.com/>
- Mansman, F., Meier, L., & Keim, D. A. (2008). Visualization of Host Behavior for Network Security. *Proceedings of the Workshop on Visualization for Computer Security* (pp. 187-202). Berlin Heidelberg: Springer.
- Mansmann, F., Keim, D. A., North, S. C., Rexroad, B., & Sheleheda, D. (2007). Visual Analysis of Network Traffic for Resource Planning, Interactive Monitoring, and Interpretation of Security Threats. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1105 - 1112. doi:10.1109/TVCG.2007.70522
- Mansmann, F., Keim, D. A., North, S. C., Rexroad, B., & Sheleheda, D. (2007). Visual analysis of network traffic for resource planning, interactive monitoring, and interpretation of security threats. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1105-1112.
- Marty, R. (2009). *Applied security visualization*. Addison Wesley Professional.
- Marty, R. (2009). *Applied Security Visualization*. Boston,USA: Addison Wesley Professional.
- McAfee. (2018, 9 17). *McAfee*. Retrieved from Security Information and Event Management (SIEM): <https://www.mcafee.com/enterprise/en-us/products/siem-products.html>
- McPherson, J., Ma, K. L., Krystosk, P., Bartoletti, T., & Christensen, M. (2004). Portvis: a tool for port-based detection of security events. *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, (pp. 73-81).
- McRee, R. (2008). Security Visualization: What you don't see can hurt you. *Information Systems Security Association*.
- Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkatamaran, S., Liu, D., . . . Talwalkar, A. (2016). MLlib: Machine Learning in Apache Spark. *Journal of Machine Learning Research*, 1-7.
- Miaoui, Y., & Boudriga, N. (2017). Enterprise security investment through time when facing different types of vulnerabilities. *Information Systems Frontiers* , 1-40.
- Miaoui, Y., & Boudriga, N. (2017). Enterprise security investment through time when facing different types of vulnerabilities. *Information System Frontiers*, 1-40.

- MicroFocus. (2018, 9 19). *ArcSight Enterprise Security Manager* . (MicroFocus) Retrieved 9 19, 2018, from <https://software.microfocus.com/en-us/products/siem-security-information-event-management/overview>
- Microfocus. (2018, 0 17). *Microfocus*. Retrieved from ArcSight Enterprise Security Manager (ESM): <https://software.microfocus.com/en-us/products/siem-security-information-event-management/overview>
- Mittelstädt, S., Stoffel, A., & Keim, D. A. (2014). Methods for Compensating Contrast Effects in Information Visualization. *Computer Graphics Forum*, 33(3), 231–240.
- Moeller, R. R. (2007). *COSO Enterprise Risk Management: Understanding the New Integrated ERM Framework*. John Wiley & Sons.
- Most Visionary Next-Gen SIEM Platform*. (2018, 9 19). (Securonix) Retrieved 9 19, 2018, from <https://www.securonix.com/>
- Muelder, C., Ma, K.-L., & Bartoletti, T. (2006). Interactive Visualization for Network and Port Scan Detection. In *Recent Advances in Intrusion Detection* (Vol. 3858 , pp. 265-283). Springer Berlin Heidelberg. doi:10.1007/11663812_14
- Mularien, P. (2010). *Spring Security 3*. Packt Publishing Ltd.
- Murray, D. G. (2013). *Tableau your data!: fast and easy visual analysis with tableau software*. Indianapolis, Indiana: John Wiley & Sons.
- Murugesan, S. (2008). Web Application Development: Challenges and the Role of Web Engineering. In *Web Engineering: Modeling and Implementing Web Applications* (pp. 7-32). London: Springer.
- Myerson, J. M. (2002). *The complete book of middleware*. Taylor and Francis.
- Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. S. (2011). Malware images: visualization ad automatic classification. *Proceedings of the 8th international symposium on visualization for cyber security* (p. 4). ACM.
- Netsparker | Web Application Security Scanner*. (2018). (Netsparker) Retrieved 01 08, 2019, from <https://www.netsparker.com/>
- Netsparker | Web Application Security Scanner*. (2019, 8 1). (Netsparker) Retrieved 01 08, 2019, from <https://www.netsparker.com/>
- Nicolett, M., & Kavanagh, K. M. (2013). *Magic Quadrant for Security Information and Event Management*. Stamford, CT, USA: Gartner.

- Nielsen, J. (1995). *10 usability heuristics for user interface design*. Fermont Ca: Nielsen Norman Group.
- Nunnally, T., Chi, P., Abdullah, K., Uluagac, A. S., Copeland, J. A., & Beyah, R. (2013). P3D: a parallel 3D coordinate visualization for advanced network scans. *2013 IEEE International Conference on Communications* (pp. 2052-2057). IEEE.
- Nunnally, T., Uluagac, A. S., Copeland, J. A., & Beyah, R. (2012). 3DSVAT: a 3D stereoscopic vulnerability assessment tool for network security. *2012 IEEE 37th Conference on Local Computer Networks* (pp. 111-118). IEEE.
- Nyarko, K., Capers, T., Scott, C., & Ladeji-Osias, K. (2002). Network intrusion visualization with NIVA, an intrusion detection visual analyzer with haptic integration. *10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 277-284). IEEE.
- Oberheide, J., Goff, M., & Karir, M. (2006). Flamingo: Visualizing internet traffic. *10th IEEE/IFIP Network Operations and Management Symposium* (pp. 150-161). IEEE.
- Olston, C., Reed, B., Srivastava, U., Kumar, R., & Tomkins, A. (2008). Pig Latin: a not so-foreign language for data processing. *Proceedings of the 2008 ACM SIGMOD international conference on management of data* (pp. 1099-1110). Newyork, NY, USA: ACM.
- Onut, I. V., & Ghorbani, A. A. (2007). Svision: A novel visual network-anomaly identification technique. *Computers & Security*, 26(3), 201-212.
- OWASP. (2018, 5 20). *Category: Vulnerability Scanning Tools*. Retrieved from OWASP: https://www.owasp.org/index.php/Category:Vulnerability_Scanning_Tools
- OWASP. (2018, 11 5). *The OWASP Foundation*. Retrieved from OWASP: https://www.owasp.org/index.php/Main_Page
- OWASP. (n.d.). *OWASP Zed Attack Proxy Project*. (OWASP) Retrieved 9 2018, 2018, from https://www.owasp.org/index.php/OWASP_Zed_Attack_Proxy_Project
- Özdemir Sönmez, F. (2016). Case Study: Development of Automated Remote Security Prototype Based on IOT Technologies on IBM Bluemix Platform. *International Conference on Information Security* (pp. 88-97). Ankara: ISCTURKEY.
- Özdemir Sönmez, F. (2019). A Decision Support System for Optimal Selection of Enterprise Information Security Preventative Actions Along Visualization. *Information System Frontiers*.

- Özdemir Sönmez, F., & Günel, B. (2018). Evaluation of Security Information and Event Management Systems for Custom Security Visualization Generation. *International Conference on Big Data, Deep Learning and Fighting with Cyber Terrorism*. Ankara.
- Özdemir Sönmez, F., & Günel, B. (2018). Security Visualization Extended Review Issues, Classifications, Validation Methods, Trends, Extensions. In Y. Maleh, *Security and Privacy Management, Techniques, and Protocols* (pp. 152-197). IGI Global.
- Özdemir Sönmez, F., & Günel, B. (2019). Holistic Web Application Vulnerabilities Visualization for Multi-Project and Multi-Phase Vulnerability Scan Results. *Computer and Security*.
- Parasoft. (2018, 5 20). *Parasoft*. Retrieved from Parasoft: <http://www.parasoft.com/>
- Parkinson, N. C. (1993). *Parkinson's Law: Or The Pursuit Of Progress*. Newyork,USA: Buccaneer Books.
- Peng, D., Chen , W., & Peng, Q. (2012). TrustVis: Visualizing Trust towards Attack Identification in Distributed Computing Environments. *Security and Communication Networks*, 6(12), 1445-1459.
- Portswigger. (2018). *Portswigger Web Security- BurpSuite*. Retrieved 01 08, 2019, from <https://portswigger.net/>
- Portswigger. (2019, 8 1). *Portswigger Web Security- BurpSuite*. Retrieved 01 08, 2019, from <https://portswigger.net/>
- Qiu, M., Zhang, L., Ming, Z., Chen, Z., Qin, X., & Yang, L. T. (2012). Security-aware optimization of ubiquitous computing systems with SEAT graph approach. *Journal of Computer and System Sciences*, 518-529.
- Qlik. (2018, 6 12). *Data analytics for modern business intelligence*. Retrieved from Qlik: <https://www.qlik.com>
- R Foundation. (2018, 6 12). *The R Project for Statistical Computing*. Retrieved from R Project for Statistical Computing: <https://www.r-project.org>
- Rapid7. (2018, 9 19). *InsightIDR*. (Rapid7) Retrieved 9 19, 2018, from <https://www.rapid7.com/products/InsightIDR>
- Ren, P., Gao, Y., Li, Z., Chen, Y., & Watson, B. (2005). IDGraphs: intrusion detection and analysis using histograms. *IEEE Workshop on Visualization for Computer Security* (pp. 39-46). IEEE.

- Ren, P., Kristoff, J., & Gooch, B. (2006). Visualizing DNS traffic. *Proceedings of the 3rd international workshop on Visualization for computer security* (pp. 23-30). ACM.
- Riel, J.-P., & Irwin, B. (2006). InetVis, a visual tool for network telescope traffic analysis. *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, (pp. 85-89). Cape Town, South Africa. doi:10.1145/1108590.1108604
- Rode, J., Johansson, C., DiGioia, P., Filho, R. S., Nies, K., Nguyen, D. H., . . . Redmiles, D. (2006). Seeing Further: Extending Visualization as a Basis for Usable Security. *Proceedings of the second symposium on Usable privacy and security* (pp. 145-155). ACM.
- Romania, O. (2018, 5 20). *OWASP Zed Attack Proxy*. Retrieved from OWASP: https://www.owasp.org/images/9/96/OWASP_2014_OWASP_ROMANIA.pdf
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New-York: McGraw-Hill.
- Saaty, T. L. (1988). What is the Analytic Hierarchy Process? In G. Mitra, H. J. Greenberg, F. A. Lootsma, M. J. Rijkaert, & H. J. Zimmermann, *Mathematical Models for Decision Support* (Vol. 48, pp. 109-121). Berlin & Heidelberg: Springer-Verlag.
- Saaty, T. L. (1999). Fundamentals of Analytical Network Process. *Proceedings of the 5th international symposium on the analytic hierarchy process*, (pp. 12-14).
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 83-98.
- Saint-Germain, R. (2005). Information Security Management Best Practice Based on ISO/IEC 17799. *The Information Management Journal*, 39(4).
- Seaman, C. B. (1999). Qualitative methods in empirical studies of software engineering. *IEEE Transactions on software engineering*, 4, 557-572.
- Sec Viz. (2018, 9 10). *SecViz Security Visualization*. Retrieved 9 8, 2018, from <https://secviz.org/>
- Seo, I., Lee, H., & Han, S. C. (2014). Cylindrical Coordinates Security Visualization for multiple domain and control botnet detection. *Computers & Security*, 46, 141-153.
- Shameli-Sendi, A., Aghababaei-Barzegar, R., & Cheriet, M. (2016). Taxonomy of Information Security Risk Assessment. *Computers & Security*, 14-30.

- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(4), 623 - 656.
- Sharma, P. P., & Navdeti, C. P. (2014). Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution. *International Journal of Computer Science and Information Technologies*, 2126-2131.
- Sherwood, J., Clark, A., & Lynas, D. (2005). *Enterprise security architecture: a business-driven approach*. CRC Press.
- Sherwood, N. A. (2005). *Enterprise security architecture: a business-driven approach*. CRC Press.
- Shin, B. (2017). *A Practical Introduction to Enterprise Network and Security Management*. Florida, USA: CRC Press-Taylor and Francis.
- Shin, J., Son, H., Rahman, K., & Heo, G. (2015). Development of a cyber security risk model using Bayesian networks. *Reliability Engineering*, 208-217.
- Shiravi, H., Shiravi, A., & Ghorbani, A. A. (2010). IDS alert visualization and monitoring through heuristic host selection. *Information and Communications Security*, 445-458.
- Shiravi, H., Shiravi, A., & Ghorbani, A. A. (2011). Situational assessment of intrusion alerts: a multi attack scenario evaluation. *Information and Communications Security*, 399-413.
- Shiravi, H., Shiravi, A., & Ghorbani, A. A. (2012). A survey of visualization systems for network security. *18*(8), 1313-1329.
- Shneiderman, B., & Plaisant, C. (1998). *Treemaps for space-constrained visualization of hierarchies*.
- Siddiqui, Z., Abdullah, A. H., Khan, M. K., & Alghathbar, K. (2011). Analysis of enterprise service buses based on information security, interoperability and high-availability using Analytical Hierarchy Process (AHP) method. *International Journal of Physical Sciences*, 6(1), 35-42.
- Sinar, E. F. (2018). *Data Visualization: Get Visual to Drive HR's Impact and Influence*. Bowling Green, OH, USA: Society for Human Resource Management (SHRM)-Society for Industrial Organizational Psychology (SIOP) Science of HR White Paper Series. .
- Solarwinds. (2018, 9 19). *Solve your toughest IT management problem, today*. (Solarwinds) Retrieved 9 19, 2018, from <https://www.solarwinds.com/>

- Solms, R. V. (1996). Information Security Management: The Second Generation. *Computers and Security*, 281-288.
- Solms, V. R. (1998). Information security management (2): guidelines to the management of information technology security (GMITS). *Information Management & Computer Security*, 221-223.
- Splunk. (2013, 7 15). *JQuery Sparklines*. Retrieved 8 7, 2018, from <https://omnipotent.net/jquery.sparkline>
- Splunk. (2018, 9 17). *Splunk and AWS provides visibility into U.S. stock and options market transactions*. (Splunk) Retrieved 9 17, 2018, from <https://www.splunk.com/>
- SSE-CMM. (2019, 01 22). *The Systems Security Engineering Capability Maturity Model (SSE-CMM)*. Retrieved from SSE-CMM Systems Security Engineering: <http://www.sse-cmm.org/model.htm>
- Staheli, D., Yu, T., Crouser, J. R., Damodaran, S., Nam, K., O’Gwynn, D., . . . Harrison, L. (2014). Visualization Evaluation for Cyber Security: Trends and Future Directions. *Proceedings of the Eleventh Workshop on Visualization for Cyber Security* (pp. 49-56). New York, NY, USA: ACM. doi:10.1145/2671491.2671492
- Sun, S., Hu, L., Song, L., Xie, Y., & Wang, P. (2013). Automatic Security Evaluation of Block Ciphers with S-bP Structures against Related-key Differential Attacks. *Information Security and Cryptology*, 39-51.
- Syamsuddin, I. (2012). Evaluation of Strategic Information Security with Fuzzy AHP Method. *American Journal of Intelligent Systems*, 2(1), 9-13.
- Syamsuddin, I. (2013). Multicriteria evaluation and sensitivity analysis on information security. *Journal of Computer Applications*, 69(24), 22-25.
- Symantec. (2014). *Symantec Internet Security Threat Report 2014*. Retrieved April 5, 2016, from http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf
- Tableau. (2018, 6 12). *Make your data make an impact*. Retrieved from Tableau: <https://www.tableau.com>
- Taha, A., Trapero, R., Luna, J., & Suri, N. (2014). AHP-Based Quantitative Approach for Assessing and Comparing Cloud Security. *13th International Conference on*

Trust, Security and Privacy in Computing and Communications. IEEE.
doi:10.1109/TrustCom.2014.39

- Takada, T., & Koike, H. (2002). Tudumi: Information visualization system for monitoring and auditing computer logs. *Sixth International Conference on Information Visualization* (pp. 570-576). IEEE.
- Taylor, T., Brooks, S., & McHugh, J. (2008). NetBytes viewer: An entity-based netflow visualization utility for identifying intrusive behavior. *Proceedings of the Workshop on Visualization for Computer Security* (pp. 101-114). Berlin Heidelberg: Springer.
- Teoh, S. T., Ma, K. L., Wu, S. F., & Jankun-Kelly, T. J. (2004). Detecting flaws and intruders with visual data analysis. *Computer Graphics and Applications*, 24(5), 27-35.
- Teoh, S. T., Ma, K. L., Wu, S. F., & Zhao, X. (2002). Case study: Interactive visualization for internet security. *Proceedings of the conference on Visualization'02* (pp. 505-508). IEEE Computer Society.
- Teoh, S. T., Ma, K. L., Wu, S. F., Mankin, A., Massey, D., Zhao, X., . . . Bush, R. (2003). ELISHA: A Visual-Based Anomaly Detection System for the BGP Routing Protocol. *IFIP/IEEE Distributed Systems: Operations and Management*, 155-168.
- Teoh, S. T., Ranjan, S., Nucci, A., & Chuah, C. N. (2006). BGP eye: a new visualization tool for real-time detection and analysis of BGP anomalies. *Proceedings of the 3rd international workshop on Visualization for computer security* (pp. 81-90). ACM.
- Teoh, S. T., Zhang, K., Tseng, S. M., Ma, K. L., & Wu, S. F. (2004). Combining visual and automated data mining for near-real time anomaly detection and analysis in BGP. *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security* (pp. 35-44). ACM.
- Tran, T., Al-Shaer, E., & Boutaba, R. (n.d.). PolicyVis: Firewall Security Policy Visualization and Inspection. *21st Large Installation System Administration Conference*, 7, pp. 1-16.
- Tri, D. T., & Dang, T. K. (2009). Security Visualization For Peer-To-Peer Resource. *International Journal on Computer Science and Engineering*, 1(2), 47-55.
- Tricaud, S. (2008). Picviz: Finding a needle in a Haystack. *Proceedings of the First USENIX conference on Analysis of system logs*, (pp. 3-3).

- Trustwave: Smart Security On Demand*. (2019, 9 19). (Trustwave) Retrieved 9 19, 2019, from <https://www.trustwave.com/home/>
- Tuttle, T. C. (1983). Organizational productivity: A challenge for psychologists. *American Psychologist*, 38(4), 479.
- Vavilapalli, V. K., Murty, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., . . . Baldeschwieler, E. (2013). Apache Hadoop YARN: yet another resource negotiator. *Proceedings of the 4th annual Symposium on Cloud Computing* (p. 5). Newyork, NY, USA: ACM.
- Venustech. (2018, 9 19). *Venusense UTM* . (Venustech) Retrieved 9 19, 2018, from <http://www.venusense.com/product/view/11168.html>
- Vigo, R., Bruni, A., & Yüksel, E. (2013). Security Games for Cyber-Physical Systems. *Secure IT Systems*, 17-32.
- Vis Sec. (2018, 9 10). *IEEE Symposium on Visualization for Cyber Security*. Retrieved 9 10, 2018, from <https://vizsec.org/>
- Vora, M. N. (2011). Hadoop - HBase for large-scale data. *Proceedings of 2011 International Conference on Computer Science and Network Technology* (pp. 601-605). Harbin, China: IEEE.
- Wadkar, S., & Siddalingaiah, M. (2014). Apache Ambari. In S. Wadkar, M. Siddalingaiah, & J. Venner, *Pro Apache Hadoop* (pp. 399-401). Berkeley, CA: Apress.
- Wang, C., & Wulf, W. A. (1997). Towards a Framework for Security Measurement. *20th National Information Systems Security Conference*, (pp. 522-533). Baltimore, MD.
- Whitaker, R. B., & Erbacher, R. F. (2011). A tri-linear visualization for network anomaly detection. *SPIE Proceedings of Visualization and Data Analysis 2011*. 7868. The Society for Imaging Science and Technology. doi:10.1117/12.872697
- Wong, D. H., Chai, K. S., Ramadass, S., & Vavasseur, N. (2010). Expert-Aware Approach: A New Approach to Improve Network Security Visualization Tool. *2010 Second International Conference on Computational Intelligence, Communication Systems and Networks* (pp. 227-231). IEEE.
- Wong, T., Jacobson, V., & Alaettinoglu, C. (2005). Internet routing anomaly detection and visualization. *Proceedings of International Conference on Dependable Systems and Networks*, (pp. 172-181).

- Wong, T., Jacobson, V., & Alaettinoglu, C. (2005, June). Internet routing anomaly detection and visualization. *Proceedings of International Conference on Dependable Systems and Networks*, (pp. 172-181).
- Xiao, L., Gerth, J., & Hanrahan, P. (2006). Enhancing visual analysis of network traffic using a knowledge representation. *2006 IEEE Symposium On Visual Analytics Science And Technology* (pp. 107-114). IEEE.
- Yang, Y.-P., Shieh, H.-M., & Tzeng, G.-H. (2013). A VIKOR technique based on DEMATEL and ANP for information security risk control assessment. *Information Sciences*, 482-500.
- Yin, X., Yurcik, W., Treaster, M., Li, Y., & Lakkaraju, K. (2004). VisFlowConnect: netflow visualizations of link relationships for security situational awareness. *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security* (pp. 26-34). ACM.
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., . . . Stoica, I. (2016). Apache Spark: A Unified Engine For Big Data Processing. *Communications of the ACM*, 56-65.
- Zaproxy. (2019, 8 1). *ZAP OWASP Zed Attack Proxy*. Retrieved 01 08, 2019, from <https://www.zaproxy.org/>
- Zhang, J., Wen, Y., Nguyen, Q. V., Lu, L., Huang, M., Yang, J., & Sun, J. (2009). Multi-dimensional Data Visualization using Concentric Coordinates. *Visual Information Communication*, 95-118.
- Zhang, X., Liu, C., Nepal, S., & Chen, J. (2013). An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. *Journal of Computer and System Sciences*, 79(5), 542–555.
- Zhang, Y., Xiao, Y., Chen, M., Zhang, J., & Deng, H. (2012). A survey of security visualization for computer network logs. *Security and Communication Networks*, 5(4), 404-421.
- Zhao, Y., Zhou, F., & Shi, R. (2012). NetSecRadar: A real-time visualization system for network security: VAST 2012 Mini Challenge Award: Honorable mention for interesting use of radial visualization technique. *IEEE Conference on Visual Analytics Science and Technology*. IEEE.
- Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. Emeryville, CA: McGraw-Hill Osborne Media.

APPENDIX

A-Security Visualization Requirements Survey

A. Volunteering Information Section

This research, is a study conducted by Dr. Banu Gunel, Professor of the Department of Informatics faculty at METU. This form is intended to inform you about the research conditions.

* Required

What is the purpose of this study:

The aim of the study is to collect information about the methods used for the visual analysis of data which can be used to improve the security of the organizations and the requirements on the topic in participants' institutions.

How do we want you to help us:

If you agree to participate in the study, we expect you to answer a survey consisting of a set of questions involving 25 multiple-choice, 7 grading scales and 14 open-ended questions. It takes an average of 25 minutes on average.

How we use information we collect from you:

Your participation is entirely voluntary. The survey does not include any questions related to determining the identity of you or your organization. Your answers will be kept completely confidential. They will be assessed by the researchers. Information obtained from participants will be evaluated in batches and will be used in scientific publications. The data you provide will not be matched with the identity information collected in the form of voluntary participation (The last sentence is related to the volunteers who participate in an interview section only).

What you should know about your participation:

The survey in general does not include questions that give personal discomfort. However, if the participant feels unwell during the participation he/she is free to quit answering the survey. In such a case, it is suffice to tell it to the person performing the survey that you do not want to complete the survey (The last sentence is related to the volunteers who participate in an interview section only).

If you would like more information about this research:

At the end of the survey, your questions on the research, if you have any, will be answered. Thank you in advance for your participation in this study. In order to get more information about the study you can communicate with Informatics Institute faculty member Dr. Banu Gunel (e-mail: bgunel@metu.edu.tr) or researcher Ferda Özdemir Sönmez (e-mail: ferda.ozdemir@metu.edu.tr).

A.1. I have read the above information and I totally agree to volunteer to answer this survey. *

A-Yes

B-No

B. Presurvey Evaluation Quiz Section

This section consists of questions aiming to find out your level of security information.

B.1. What can a firewall protect against? *

0 points

A-unauthenticated interactive logins from the "outside" world

B-viruses

C-fire

D-misuse of passwords

This is a required question

B.2. What is the main purpose of access control? *

0 points

A-to authorise full access to authorised users

B-to limit the actions or operations that a legitimate user can perform

C-to stop unauthorised users accessing resources

D-to protect computers from viral infections

This is a required question

B.3. Which of the following is NOT a good property of a firewall? *

0 points

A-only authorised traffic must be allowed to pass through it

B-the firewall itself, should be immune to penetration

C-it should allow for easy modification by authorised users

D-traffic must only be allowed to pass from inside to outside the firewall

This is a required question

B.4. A false positive can be defined as... *

0 points

A-an alert that turns out to represent legitimate activity upon further investigation.

B-an alert that indicates nefarious activity on a system that is not running on the network.

C-the lack of an alert for nefarious activity.

D-Both a. and b.

This is a required question

B.5. When discussing IDS/IPS, what is a signature? *

A-An electronic signature used to authenticate the identity of a user on the network

B-Attack-definition file

C-It refers to "normal," baseline network behavior

D-None of the above

This is a required question

C. Security Visualization Use Cases

In this section there are questions related to both usability of visualization solutions and level of adoption to such solutions in your organization.

C.1. Please select the origin of data visualization systems which are part of system monitoring and analyses tasks in your organization.

Open Source

Commercial

In House

C.2. For the security visualization use cases below, please select applicable choices. *

I haven't heard
the use case
before

I am familiar to
use case, but it
is not applicable
for my
organization

Application of
use case is
possible and
moderately
beneficial

Application of
use case would
be very
beneficial for
my organization

This use case is
already part of
my organization
procedures

None of
the above

visualization of port activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of internal network traffic data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of network traffic between internal hosts and external IP's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of IDS data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of web browsing trends and activities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
monitoring of current state of hosts and servers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
firewall log visualization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
firewall configuration visualization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of DNS traffic and lookup behavior	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of vulnerability levels	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of file transfers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
monitoring of routing behaviors among AS's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
visualization of BGP update messages	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

C.3. Do you have any strategies/methods to differentiate normal behaviour of web browsing from abnormal behavior? If yes, please describe any specific tool you use, procedure, checklist etc.

Your answer

C.4. Do you have any strategies/methods to differentiate normal activities of file sharing from suspicious activities? If yes, please describe any specific tool you use, procedure, checklist etc.

Your answer

C.5. Do you have any strategies to differentiate normal behaviour of social media usage from suspicious behavior using data? If yes, please describe any specific tool you use, procedure, checklist etc.

Your answer



C.6. If you have any suggestions for other types of security visualization usage scenarios which is beneficial for your organization, please explain (Any data visualization effort which may be beneficial during system security checks would count, such as trend analysis of some particular data, use of Excel graphs to visualize some particular log data or a complicated visualization system would count.)

Your answer



C.7. Which of the following security visualization tools and prototypes are you familiar with?

Security quad and cube

- Avisa
- Avisa2
- NetsecRadar
- P3D
- CCScanviewer
- Hviz
- Enavis
- PolicyVis
- CCSvis
- Vafle
- TrustVis
- Netvis
- NV
- Synema
- Visflowconnect
- IDsRainStorm
- Rumint
- Dorothy Project
- SecureScope
- NetIQ Security Manager
- Monitoring and Response Security Console
- Cisco Mars
- HnMap
- Histomap
- Visual
- TNV
- Portvis

- Snortview
- IDGRaphs
- IP-Matrix
- Vizalert
- Tamp
- LinkRank
- IDTK
- Davast
- PHPIDS
- 3DSVat
- InetVis
- Tudumi
- Spinning cube of Potential Doom
- NetBytesViewer
- NIVA
- Impromptu
- NV
- Hone
- Existence plots
- Nflowviz
- Svision
- NvisionIP
- Clique
- Clockview
- Nagios
- Portall
- Radial Traffic Analyzer
- VisualFirewall
- Bgplay
- Elisha
- Bgp Eye
- RNA Visualization Module of SourceFire
- IDSRadar
- SeeNet
- PCAV
- Other:

BACK

NEXT

D. Security Visualization Data

Majority of the security analysis methods depend on analysis of big security data. Working with this type of data has inherent difficulties. In this section there are questions related to the data sources commonly used in security analyses and, your level of familiarity with those data sources.

D.1. Select the data types that are collected, stored, analyzed in your organization. *

Not collected
at all

Collected and stored but
not analysed

Analysed as part of
security analyses

- Network traffic data
- Firewall configuration data
- Firewall log data
- Intrusion detection and/or prevention system alert log
- Operating system log
- Web server log
- Application server log
- Web proxy log
- Database access log
- Router configurations log
- Enterprise specific application log

D.2. How often security log files are manually analyzed in your organization?

- Never
- Daily
- Weekly
- Monthly
- Only when a security incident occurs
- Other :

D.3. How many personnel are responsible for analyzing such log data?

Your answer

D.4. Do you have a strategy to reduce the size of any of your logs? If so please explain.

Your answer

D.5. Are you knowledgeable on approximate log file sizes? *

Yes

No

BACK

NEXT

E. Security Visualization Data Size

Majority of the security analysis methods depend on analysis of big security data. Working with this type of data has inherent difficulties. In this section there are questions related to the data sources commonly used in security analyses and, your level of familiarity with those data sources.

E.1. How many daily records are generated in your firewall log file approximately?

Your answer

E.2. How many daily records are generated in your IDS alert file approximately?

Your answer

E.3. How many daily records are generated in your application server access log file approximately?

Your answer

E.4. How many daily records are generated in your application server error log file approximately?

Your answer

E.5. How many daily records are generated in your web server access log file approximately?

Your answer

E.6. How many daily records are generated in your web server error log file approximately?

Your answer

E.7. How many daily records are generated in your mail server log file approximately?

Your answer

BACK

NEXT

F. Security Analysis Techniques

This section consists of questions related to security analysis techniques adopted in the participant's organization.

As an answer to the below three questions we want you to describe up to 6 independent ways of doing security checks in your organization.

First you have to decide on the threat types which are attempted to be detected for each analysis.

F.1. Threats which are subject of the analyses.

	Analysis 1	Analysis 2	Analysis 3	Analysis 4	Analysis 5	Analysis 6
Botnet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Distributed Denial of Service DDOS and Denial of Service DOS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorized Access to Web Server	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorized Access to File Server	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorized Access to Application Server	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Unauthorized Access to Database Server	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorized Access to Other Servers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unauthorized Access to Host Machine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trojan Horse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ransomware	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Spoofing, Phishing and Pharming	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Social Engineering	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rootkit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Malicious Spyware & Adware	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Virus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Malware	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rogue security software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wi-Fi Eavesdropping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Buffer overflow	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
FTP bounce	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Smurf	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Packet sniffing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Blended Threat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Keystroke logging (Keylogging)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Worms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Man in the Middle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Then select the data sources which are required for each particular analysis type

F.2. Data Sources used as a part of analyses.

	Analysis 1	Analysis 2	Analysis 3	Analysis 4	Analysis 5	Analysis 6
Network Traffic Data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Firewall configuration data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Firewall log data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intrusion detection and/or preventions system alert log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Operating system log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Web server log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Application server log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Web proxy log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mail server log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Database access log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Router configurations log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Enterprise specific application log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Later select the data attributes from the list below which are examined during each particular analysis

F.3. Data Attributes which are controlled during the analyses. *

	Analysis 1	Analysis 2	Analysis 3	Analysis 4	Analysis 5	Analysis 6
--	------------	------------	------------	------------	------------	------------

Number of Total Records in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Specific Source IP in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Specific Source Port in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Group of Source IPs in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Group of Source Ports in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Specific Destination IP in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Specific Destination Port in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Group of Destination IPs in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of Total Records With a Group of Destination Ports in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Records having an alert type	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Records having an alert classification	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of total errors	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Number of a specific type of error	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Timing of an event	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
User Names or Ids Accessed to an Asset in a Time Period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

F.4. Which of the common security analyses methods are handled in your organization? (Short explanation of the methods are provided in the bottom of the question.) *

- Triage Analysis
- Escalation Analysis
- Correlation Analysis
- Threat Analysis
- Incident Response Analysis
- Forensic Analysis

1- Triage Analysis is an analysis type that aims to make a prioritization of the data based on urgency. In triage analysis a quick first look at the data is taken and false positives are

eliminated in order to capture the suspicious activities which require further analysis. 2- Escalation Analysis can take from hours to weeks. In this analysis, potential incidents are examined by taking into account the tips from colleagues and cooperating organizations. 3- Correlation Analysis searches for patterns in the current and historical data. This task may take from weeks to months. 4- Threat Analysis uses external information such as information from hacker web sites to identify the attackers' true identity and motivation. 5- Incident Response Analysis involves decision of possible sets of actions that should be taken in case of an incident. 6- Forensic Analysis searches for evidence to enable law enforcement.

BACK

NEXT

G. Visualization Design and Display Properties

In this part, there are questions related to expected design properties of the security visualization tools as well as usability of various display types in security visualization solutions.

G.1. Can you rate the importance of the visualization properties listed? *

	Not important	Slightly Important	Moderately Important	Very Important	Extremely important
the ability to depict a relatively large data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the ability to save detected patterns	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the ability to work with real time data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the ability to depict most types of attacks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the visualization information is visible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

without the need to hover
the mouse

the visualization

represents data from
more than one security
log

displaying the incident

time

thick boundaries are

used to separate different
classes of information

the visualization is

interactive

the visualization is

searchable

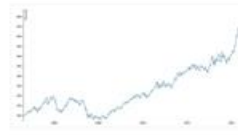
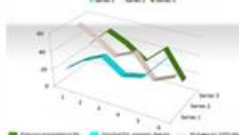
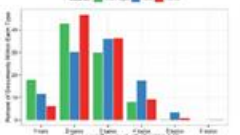
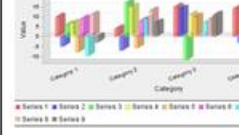
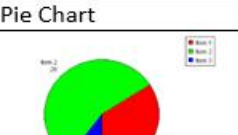

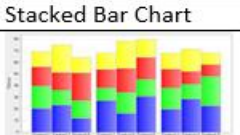


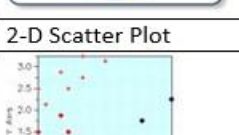
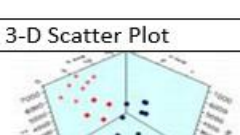
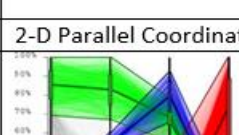
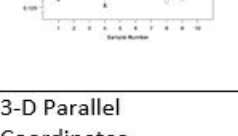
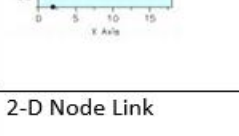
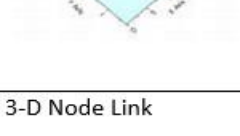
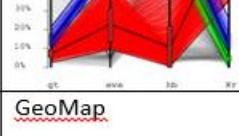
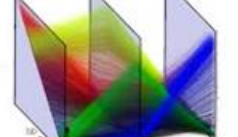



the visualization is

zoomable

the visualization is

scalable

Sample Graphs for Various Display Types

Line Chart 	3D Line Chart 	Bar Chart 	3D Bar Chart 
Pie Chart 	Stacked Pie Chart 	Stacked Bar Chart 	Stacked Line Chart 
Box Plot 	2-D Scatter Plot 	3-D Scatter Plot 	2-D Parallel Coordinates 
3-D Parallel Coordinates 	2-D Node Link 	3-D Node Link 	GeoMap 
Treemap 	Animation 	Simulation 	Gamification 

G.2. Can you rate the display types according to their usability? *

	Not at all	Not really	Neutral	Somewhat ok	Almost always
Line Cart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3-D Line Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bar Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3-D Bar Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pie Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stacked Pie Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stacked Bar Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stacked Line Chart	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Box-Plots	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2-D Scatter Plot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3-D Scatter Plot	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2-D Parallel Coordinates	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3-D Parallel Coordinates	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2-D Node-Link Graphs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3-D Node Link Graphs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Geo Maps	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Treemaps	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Animation Graph	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Simulation Graph	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gamification	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

H. Technical Infrastructure

This section consists of questions related to software and hardware infrastructure of your organization.

H.1. Approximate number of hosts : *

Your answer

H.2. Approximate number of servers : *

Your answer

H.3. Select the software systems that are used in your organization.

Select all that apply. *

- Static Web Pages
- Dynamic Web Application
- ERP
- SCM
- CRM
- Other:

H.4. Select the hardware, networking and system components that are part of your infrastructure? Select all that apply. *

- File Sharing Server
- Web Server
- Mail Server (Internal)
- Mail Server (External)
- Application Server
- Database Server
- Cloud Storage
- Other Cloud Services
- External Router
- Internal Switch or Router
- Wireless Network
- Printer
- E-Fax
- Other:

H.5. Select the security systems that are part of your organization infrastructure? Select all that apply. *

- Network Level Firewalls
- Application Level Firewalls
- Intrusion Detection and/or Prevention System
- Database Security System
- Email Security System

- Log Correlation System
- Host Based Intrusion Prevention System
- Data Loss Prevention System
- Data Encryption Software
- Vulnerability Scanner
- Risk Management System
- URL Filtering System
- Anti Virus
- Anti Spam
- Other:

BACK

NEXT

I. Organization and Domain Information

This section consists of questions aiming to gather information related to participant's organization and its domain.

I.1. What is the primary business area of your organization (regardless of your position)? *

- Advisory/ consulting
- Agriculture
- Commerce
- Construction
- Education
- Finance and banking
- Insurance
- IT (software)
- IT(infrastructure)
- Logistics
- Media/ advertising
- Municipal services
- NGO
- Power engineering
- Production/technology
- Public administration
- Telecommunications
- Tourist services and sports
- Other :

I.2. Number of employees? *

- < 10
- 10- 50
- 50 – 250
- > 250

I.3. Do you share infrastructure, services and/or data with ... *

	Yes	No
your customers	<input type="radio"/>	<input type="radio"/>
your suppliers	<input type="radio"/>	<input type="radio"/>
your partners	<input type="radio"/>	<input type="radio"/>
other stakeholders	<input type="radio"/>	<input type="radio"/>

J. User Information

This section consists of questions aiming to gather information related to your experience on information technologies and on information security area.

J.1. Level at work place?

- Junior
- Senior
- Manager
- Senior Manager
- Owner or Partner

J.2. Relation to Technology?

- Non-technical
- Technical

J.3. How long have you been working in IT sector in years? *

Choose

J.4. How long have you been working in information security area? *

Choose

J.5. Select the tasks that are carried out in your organization? Which ones are you responsible for? *

	Not done.	Done. I do not take part.	Done. I do take part.
Software design and development	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Project management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Manage web server
- Manage file server
- Manage end user workstations
- Set security policy
- Manage firewall(s)
- Manage corporate network
- Manage IDS system(s)
- Manage customer site
- Manage database server
- Manage application server

J.6. Did you take any formal education related to security analysis methods ? *

- Yes
- No

J.7. If you took any education, did it make any changes to your security analysis methods? *

- Yes
- No
- N/A

J.8. Please specify the security related certificates you hold.

Your answer

B-Survey Permission Form

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER



DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr
Sayı: 28620816/409

24 EKİM 2016

Konu: Değerlendirme Sonucu

Gönderilen: Doç.Dr. Banu GÜNEL;
Enformatik Enstitüsü

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Doç.Dr. Banu GÜNEL;

Danışmanlığını yaptığınız doktora öğrencisi Ferda ÖZDEMİR SÖNMEZ'in "Kurum/Organizasyon Düzeyinde Veri Güvenliği Durumu Görselleştirme" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülerek gerekli onay 2016-FEN-058 protokol numarası ile 10.10.2016-30.03.2017 tarihleri arasında geçerli olmak üzere verilmiştir.

Bilgilerinize saygılarımızla sunarız.

Prof. Dr. Canan SÜMER

İnsan Araştırmaları Etik Kurulu Başkanı

Prof. Dr. Meliha ALTUNIŞIK

İAEK Üyesi

Prof. Dr. Mehmet UTKU

İAEK Üyesi

Yrd. Doç. Dr. Pınar KAYGAN

İAEK Üyesi

Prof. Dr. Ayhan SOL

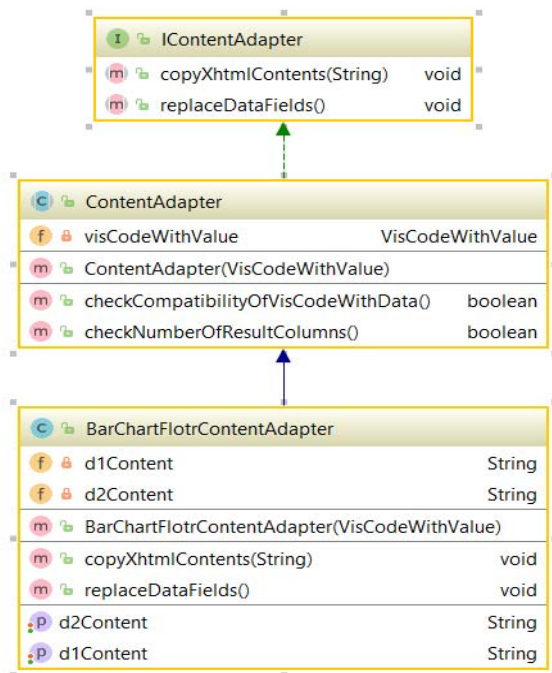
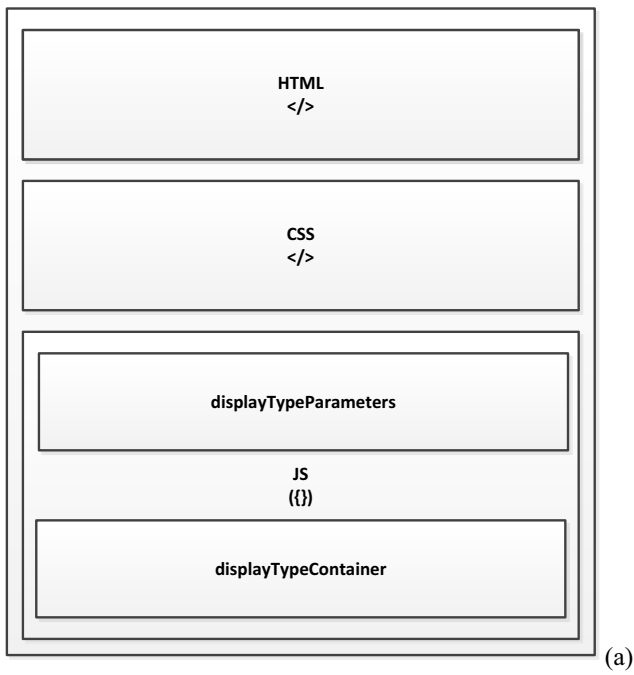
İAEK Üyesi

Prof. Dr. Ayhan Gürbüz DEMİR

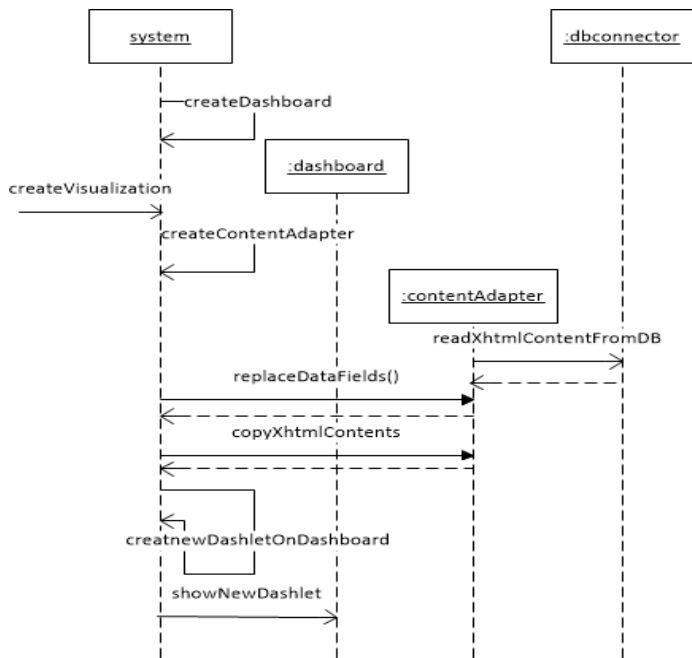
İAEK Üyesi

Yrd. Doç. Dr. Emre SELÇUK

İAEK Üyesi

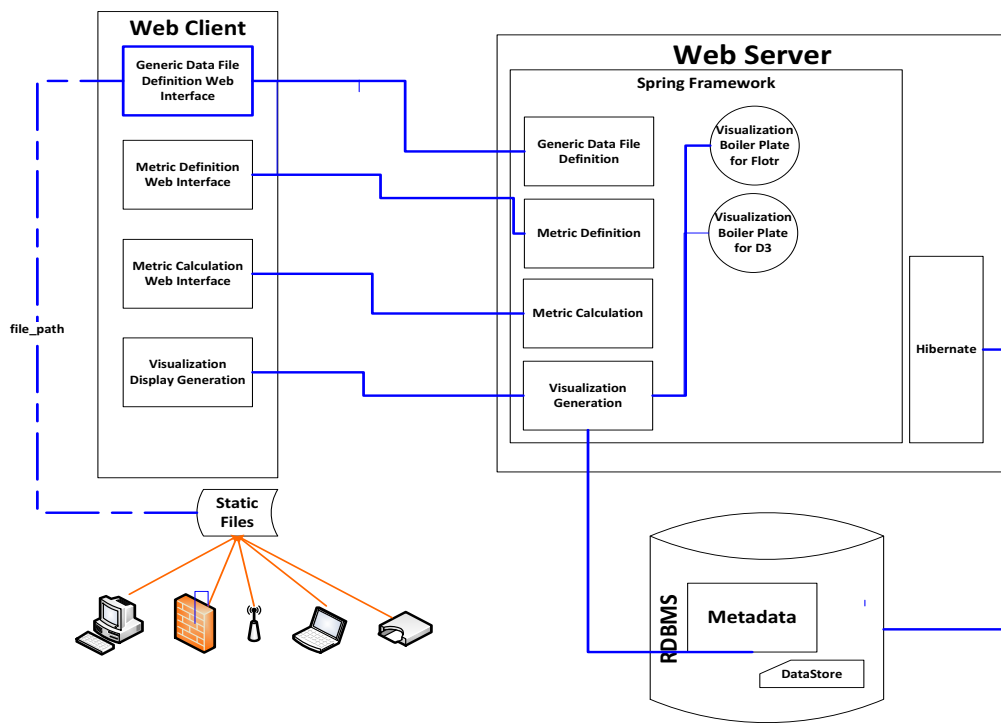


(b)

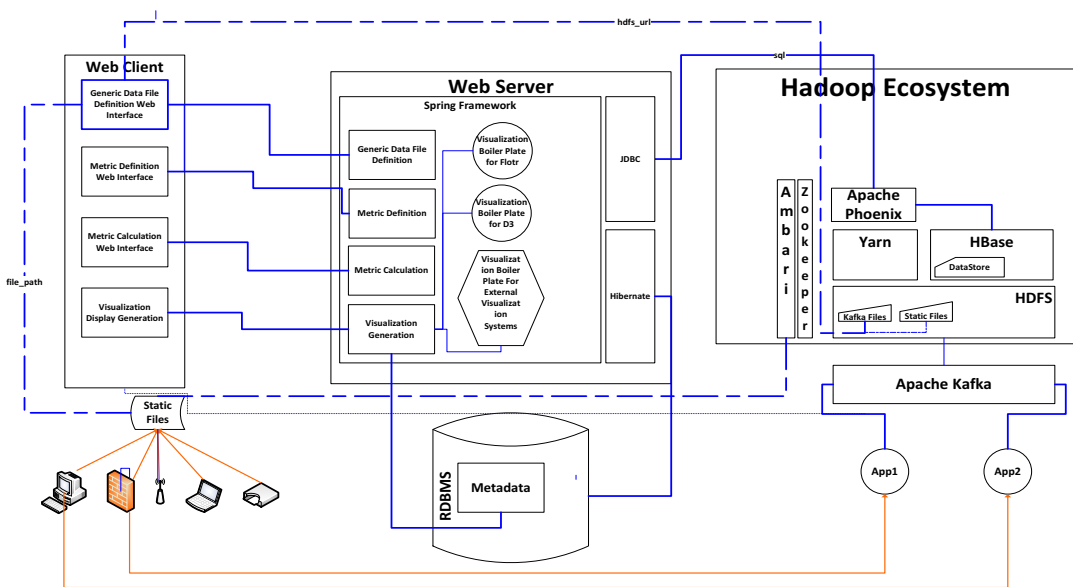


(c)

Figure 35 - a) XHTML content for a JavaScript-based display b) ContentAdapter structure for Flotr JavaScript library based Bar Chart visualization c) Sequence diagram for visualization display in dashboard form (Large Scale)

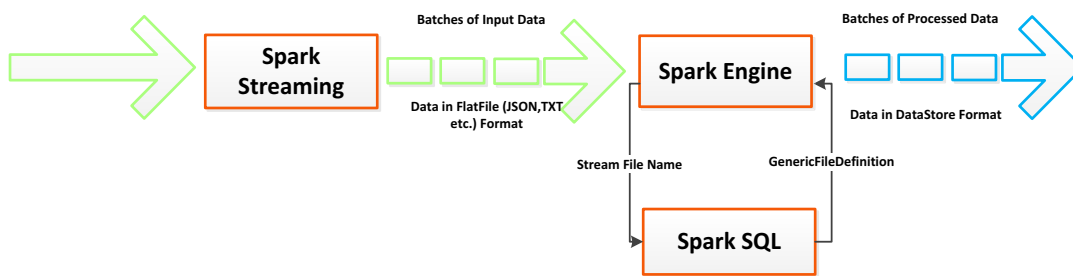


(a)

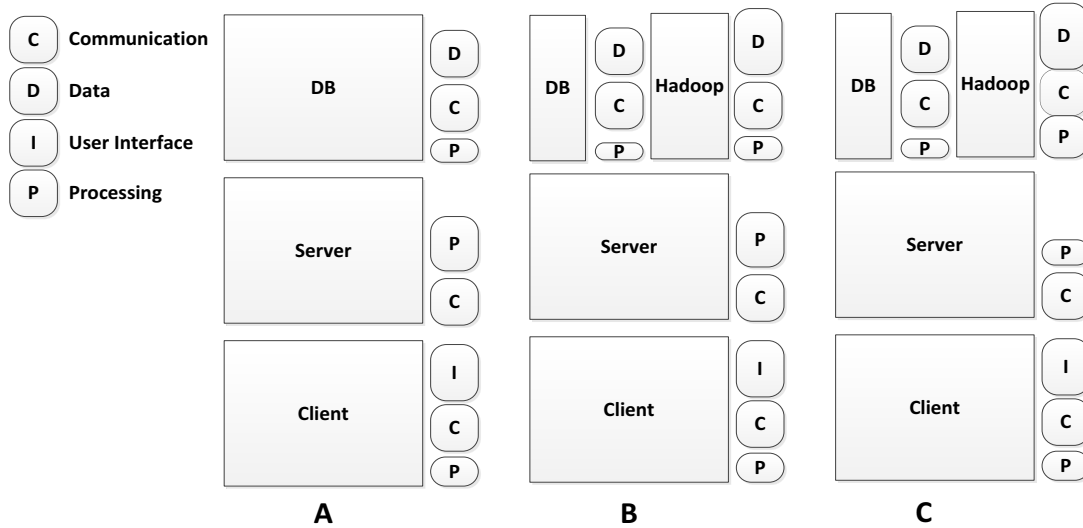


(b)

(c)



(d)



(e)

Figure 37 - a) The first design architecture, b) The second design architecture, c) The third design architecture, d) Streaming details for third design e) Evolution summary (Large Scale)

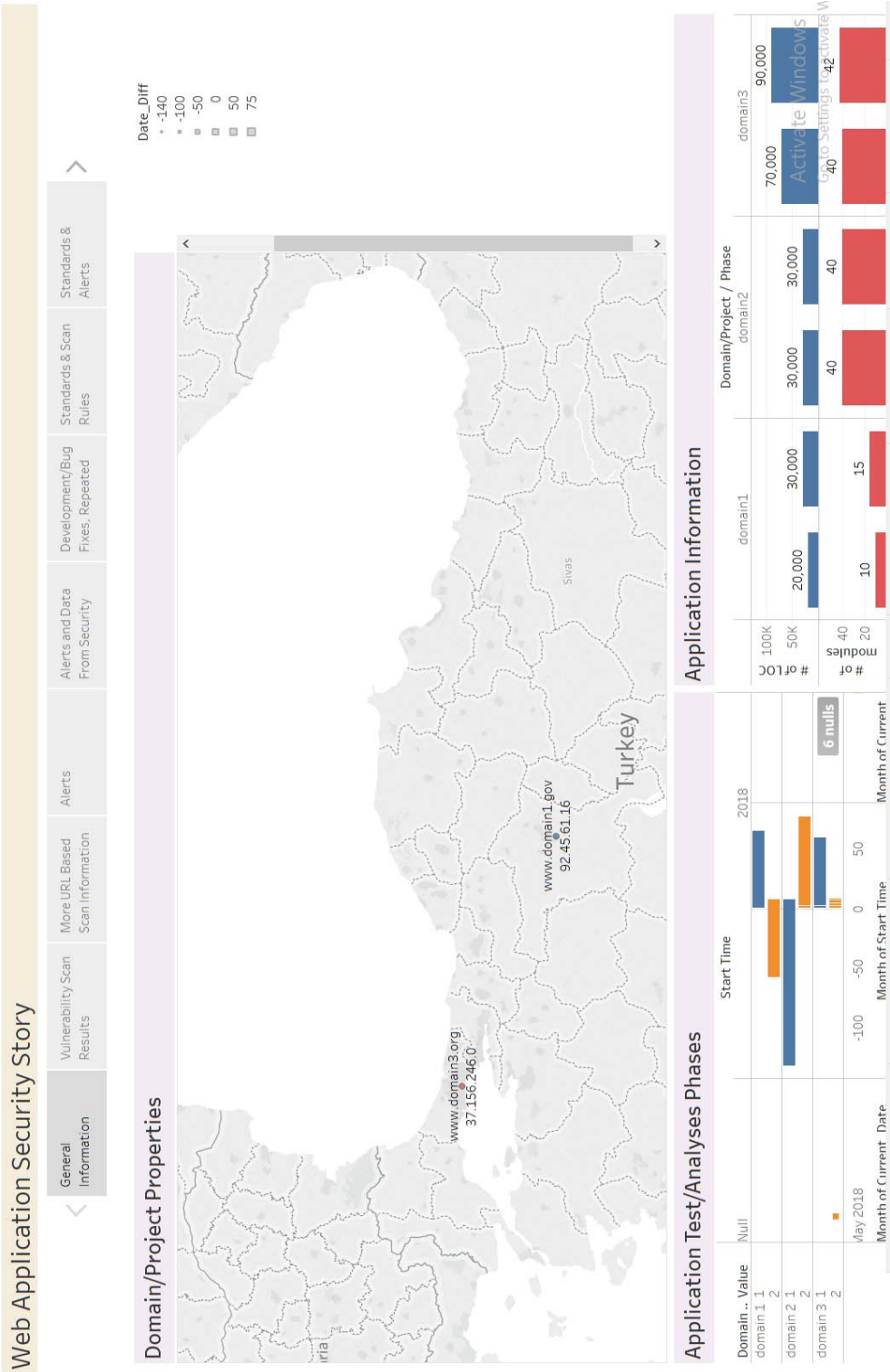


Figure 43 - General information dashboard (Large Scale)

Web Application Security Story

General Information
Vulnerability Scan Results
More URL Based Scan Information
Alerts
Alerts and Data From Security
Development/Bug Fixes, Repeated
Standards & Scan Rules
Standards & Alerts

Scan Results

Category	Percentage
Successfulity	17.58%
Max Depth	0.10%
Out of Scope	76.38%

Scan Results Distribution

Reason	Count
Processed	4
Max Size	3,902
Not Text	4
Out of Scope	2
Successfully	4

Scan Durations

Domain / Phase	domain1	domain2	domain3
Avg. RTT	95.1	57.2	386.1
RTT	421	3,319	229,753

Scan Results By Highest Alert Level

Number of Records	domain1	domain2	domain3
1,000	3,905	482	105
100	202	265	105
10	13	13	8

Go to Settings to activate Windows.

Scan Results -Details

Method	URL	Processed	Code	Reason	Tag
domain1	https://www.domain1.gov/pybs-disportal/javax.faces.resource/components.css.vht...	Successf...	200	Ok	SetCookie, Comment
1, Low	https://www.domain1.gov/pybs-disportal/javax.faces.resource/components.js.vht...	Successf...	200	Ok	SetCookie
GET	https://www.domain1.gov/pybs-disportal/javax.faces.resource/components.js.vht...	Successf...	200	Ok	SetCookie
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/core.js.vhtmi/jseessi...	Successf...	200	Ok	Hidden, SetCookie
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/core.js.vhtmi/jseessi...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/animate.css.vht...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/layout-pybs.css.x...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/layout-pybs.css.x...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/nanoscroll.css...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/pyb_theme.css...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/pyb_theme.css...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/pybs-layout.css.x...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/css/pybs-layout.css.x...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/ta/font-awesome.css...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/forms.js.vhtmi/jseessi...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/forms.js.vhtmi/jseessi...	Successf...	200	Ok	SetCookie, Comment
	https://www.domain1.gov/pybs-disportal/javax.faces.resource/images/ism_logo.svg...	Successf...	200	Ok	SetCookie
	https://www.domain1.gov/nvhs-dioportal/javax.faces.resource/images/ism_logo.svg...	Successf...	200	Ok	SetCookie

Figure 44 - Vulnerability scan results dashboard (Large Scale)

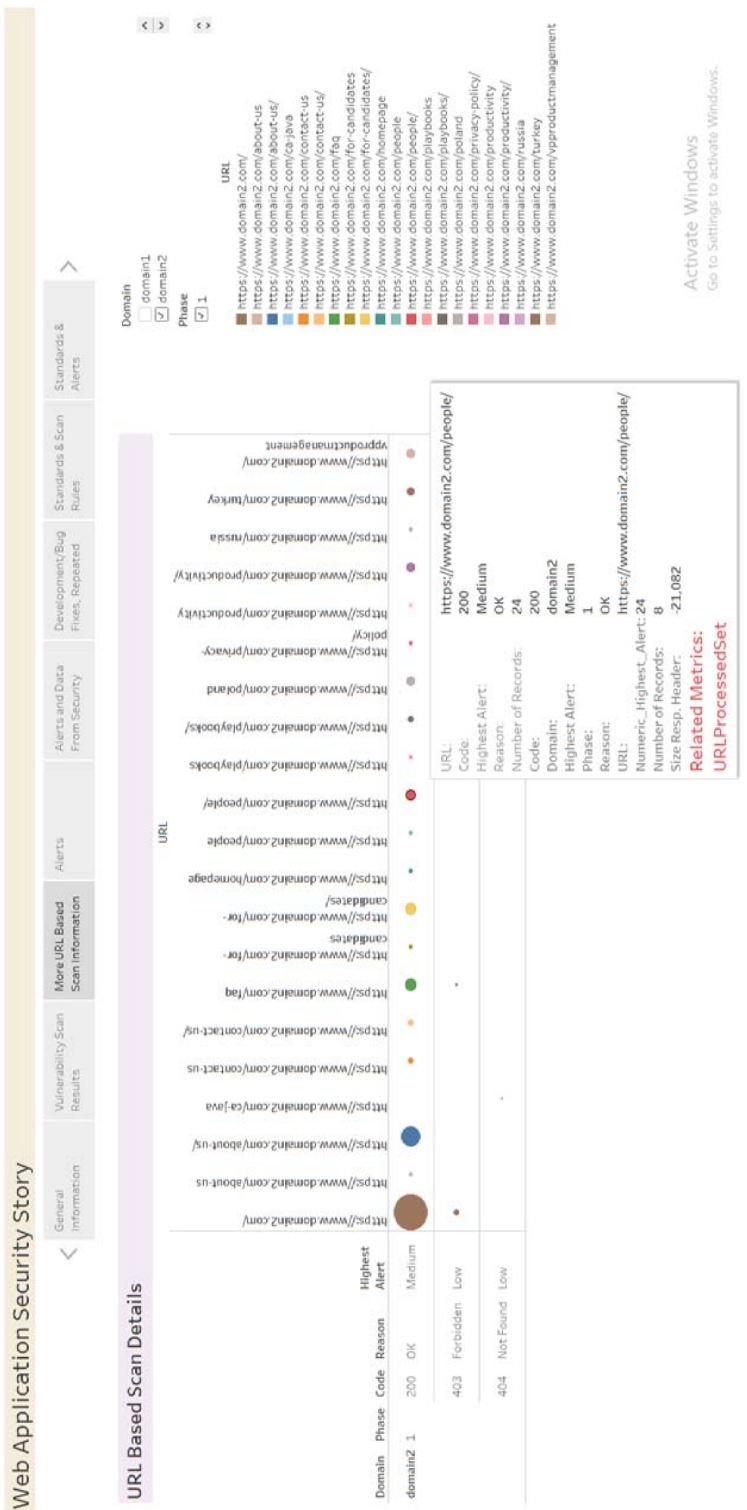


Figure 45 - URL based scan details dashboard (Large Scale)

Web Application Security Story

- General Information
- Vulnerability Scan Results
- More URL Based Scan Information
- Alerts
- Alerts and Data From Security
- Development/Bug Fixes, Repeated
- Standards & Scan Rules
- Standards & Alerts



Figure 46 - Alerts dashboard (Large Scale)



Figure 47 - Alerts and data from security protection systems (Large Scale)

- General Information
- Vulnerability Scan Results
- More URL Based Scan Information
- Alerts
- Alerts and Data From Security
- Development/Bug Fixes, Repeated
- Standards & Scan Rules
- Standards & Alerts

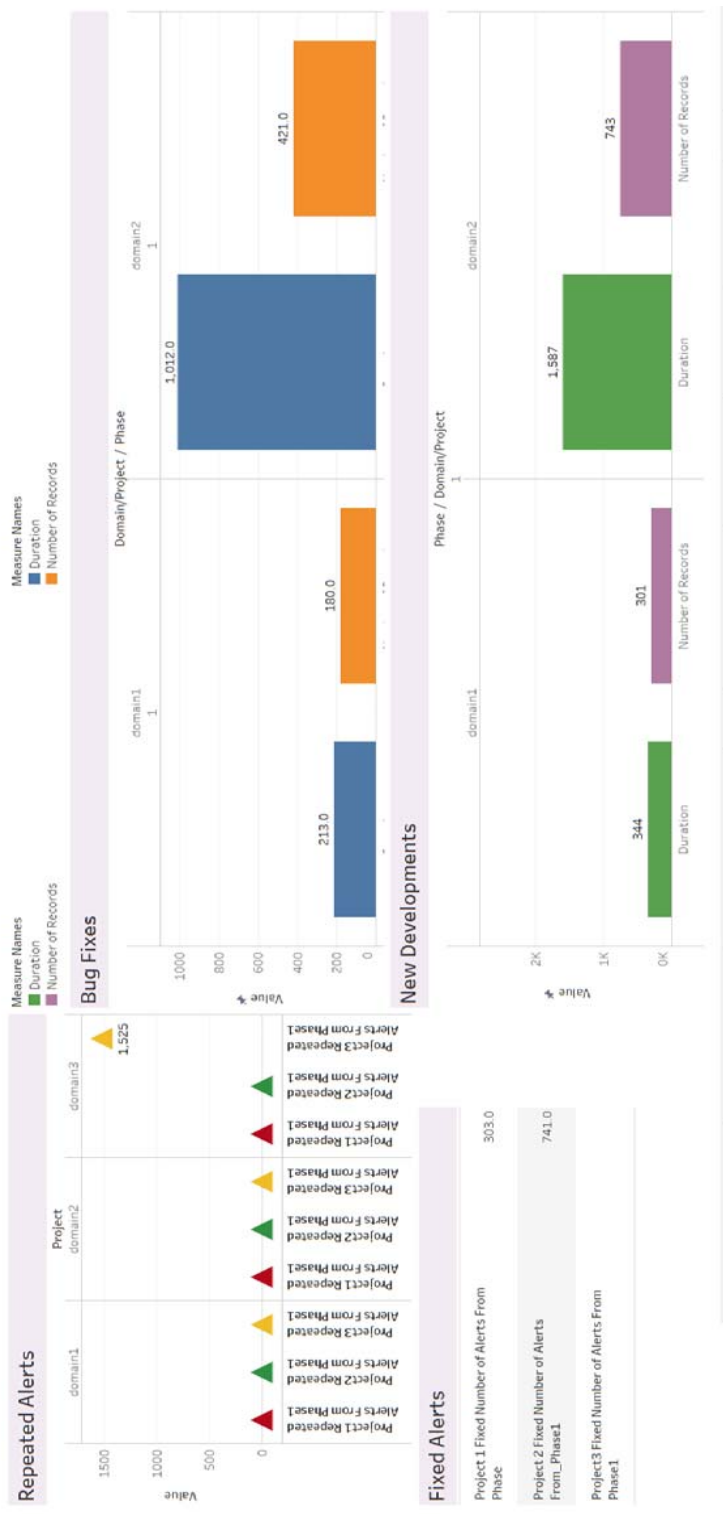
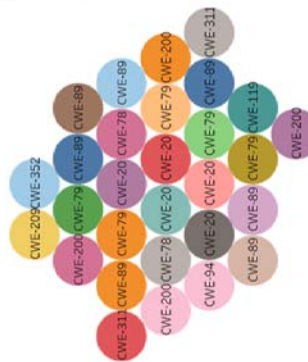
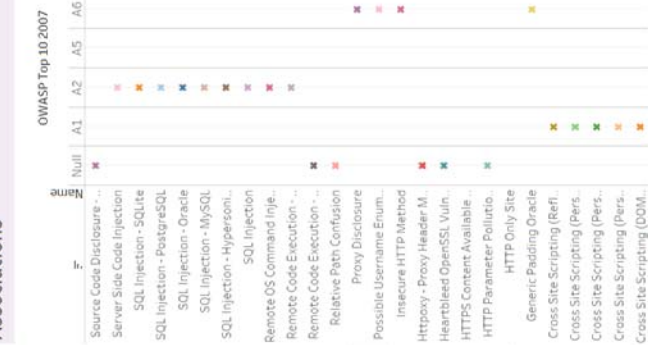


Figure 48 - New developments, bug fixes, repeated alerts, fixed alerts (Large Scale)

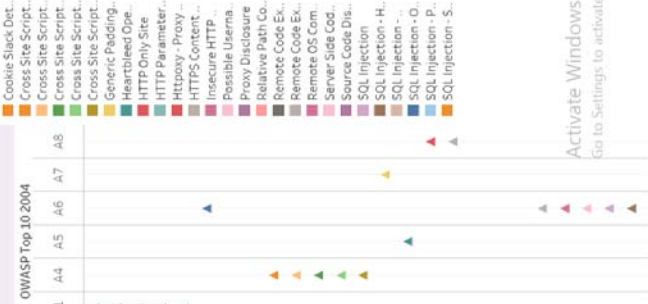
Scan-Rule - CWE Associations



OWASP Top 10 2007- Scan Rules Associations



OWASP Top 10 2004- Scan Rules Associations



WASC 24 (+2)-Scan Rule Associations



Activate Windows
Go to Settings to activate Windows.

Figure 49 - Standards and Scan Rules (Large Scale)



Figure 50 - Standards and Alerts (Large Scale)

**BU BÖLÜM, İLGİLİ BÖLÜMLERİ TEMSİL EDEN İNSAN ARAŞTIRMALARI
ETİK ALT KURULU TARAFINDAN DOLDURULACAKTIR.**

Protokol No: 2016-FEK-075

İAEK DEĞERLENDİRME SONUCU

Sayın Hakem,

Aşağıda yer alan üç seçenektan birini işaretleyerek değerlendirmenizi tamamlayınız. Lütfen **“Revizyon Gereklidir”** ve **“Ret”** değerlendirmeleri için gerekli açıklamaları yapınız.

Değerlendirme Tarihi: 2016-02-26

Ad Soyad: Metin girmek için tıklayın

Herhangi bir değişikliğe gerek yoktur. Veri toplama/uygulama başlatılabilir.

Revizyon gereklidir

Gönüllü Katılım Formu yoktur.

Gönüllü Katılım Formu eksiktir.

Gerekçenizi ayrıntılı olarak açıklayınız: Metin girmek için tıklayın

Katılım Sonrası Bilgilendirme Formu yoktur.

Katılım Sonrası Bilgilendirme Formu eksiktir.

Gerekçenizi ayrıntılı olarak açıklayınız: Metin girmek için tıklayın

Rahatsızlık kaynağı olabilecek sorular/maddeler ya da prosedürler içerilmektedir.

Gerekçenizi ayrıntılı olarak açıklayınız: Metin girmek için tıklayın

Diğer.

Gerekçenizi ayrıntılı olarak açıklayınız: Metin girmek için tıklayın.

Ret

Ret gerekçenizi ayrıntılı olarak açıklayınız: Metin girmek için tıklayın

TEZ İZİN FORMU / THESIS PERMISSION FORM

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics

Enformatik Enstitüsü / Graduate School of Informatics

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences

YAZARIN / AUTHOR

Soyadı / Surname :

Adı / Name :

Bölümü / Department :

TEZİN ADI / TITLE OF THE THESIS (İngilizce / English) :

.....
.....
.....
.....

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans / Master**

Doktora / PhD

1. **Tezin tamamı dünya çapında erişime açılacaktır. / Release the entire work immediately for access worldwide.**

2. **Tez iki yıl süreyle erişime kapalı olacaktır. / Secure the entire work for patent and/or proprietary purposes for a period of two year. ***

3. **Tez altı ay süreyle erişime kapalı olacaktır. / Secure the entire work for period of six months. ***

** Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.
A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature

Tarih / Date