

**Deanship of Graduate Studies
Al-Quds University**



Information Assurance Vulnerability Risk Assessment

Ayman Musbah Ali Al-bayya

M.Sc. Thesis

Jerusalem-Palestine

1436 H / 2015 C

Information Assurance Vulnerability Risk Assessment

Prepared by:
Ayman Musbah Ali Al-bayya

B.Sc. in Computer Engineering , Al-Quds University

Supervisor: Dr. Labib Arafah

A thesis submitted in partial fulfillment of requirements for
the degree of master in Computer and Electronic
Engineering, Najjad Zeenni Faculty of Engineering, Al-Quds
University

1436 H / 2015 C

Al-Quds University
Deanship of Graduate Studies
Najjad Zeenni Faculty of Engineering



Thesis Approval

Information Assurance Vulnerability Risk Assessment

Prepared by: Ayman Musbah Ali Al-bayya
Registration No:21111083

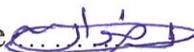
Supervisor: Dr. Labib Arafeh

Master Thesis submitted and accepted, Date: May 23, 2015

The Names and Signatures of the examining committee members are as follows:

1- Head of the committee: Dr. Labib Arafeh Signature.....

2- Internal Examiner: Dr. Rashid Jayousi Signature

3- External Examiner: Dr. Radwan Tahboub Signature 

Jerusalem-Palestine
1436 H / 2015 C

Dedication

My heartfelt gratefulness to my beloved country “Palestine especially the city of Jerusalem”, parents, wife and friends.

Ayman Musbah Ali Al-Bayya’

Declaration

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signature:

Ayman Musbah Ali Al-bayya

Date:.....

Acknowledgements

First and above all, I praise Allah, God the Almighty, for the completion of this master's thesis. Only due to His blessings I could finish my thesis.

This thesis appears in its current form due to the assistance and guidance of several people. I would therefore like to offer my sincere thanks to all of them.

My sincere thanks go to Al-Quds University, **Najjad Zeenni Faculty of Engineering**, and its academic and administrative staff.

Dr. Labib Arafeh, my esteemed promoter, my supervisor, whose expertise, understanding, and guidance, added considerably to my graduate experience. I appreciate his vast knowledge and skills in many areas. I want to express my cordial thanks for his trust, the insightful discussions, the valuable advices and spiritual support during the whole period of the study.

I would like to thank the members of my dissertation committee, Dr. Radwan Tahboub, and Dr. Rashid Jayousi for their excellent advices and detailed review of this study.

Finally, I can just say thanks for everyone and may Allah give you all the best in return.

Jerusalem, April 2015

Ayman Musbah Ali Al-Bayya'

Abstract

Recent events such as SQL Slammer and Blaster which spread across the Internet and infect more than 90% of the vulnerable systems have sparked a dramatic interest in Information Assurance (IA). Those events increase the number of high-profile organizational failures for inadequacy of data, information and intelligence available to decision making at key moments. Despite spending millions of dollars on firewalls, encryption technologies, and intrusion detection software, information infrastructure vulnerabilities and incidents continue to happen. Therefore, it is a necessity for organizations to provide confidence and certainty of its information and ensure certain levels of availability, integrity, authentication, confidentiality, and non-repudiation of their information assets against unacceptable risk. This is referred to as Information Assurance (IA) (Blyth and Kovacich 2001). It is assuring that the security mechanisms are actually effective and the system can be entrusted with the processing tasks on the critical information. This study presents a risk level estimation model that derives risk level as a conditional probability over frequency and impact estimates. The frequency and impact estimates are derived from a set of attributes specified in the Common Vulnerability Scoring System (CVSS). This model predicts the risk level of vulnerabilities based on service level and capability needs of this service from the organization to achieve the critical missions. Therefore, IA imposes on an organization a review of its mission and threats and the securing of its needed capabilities against the risks those threats pose. This study presents developed approaches depend on the Fuzzy-based techniques including adaptive Neuro-Fuzzy approaches. Historical data have been used from National Vulnerability Database (NVD) to develop and test the proposed models. Different models have been developed such as, Sugeno Fuzzy inference System (FIS) with hybrid optimization technique, Sugeno Model using Subtractive clustering, Sugeno cascaded model using subtractive clustering with hybrid optimization technique, and finally Mamdani models. All developed models have been checked

for adequacy. Different measures have been adapted such as Correlation Coefficient (CC), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and percentage of differences between the actual and predicted data. The IA risk level model performance has been improved by grid partition with hybrid optimization technique models which the CC obtained is equal to 0.993 (Maximum) and corresponding MAPE, RMSE, Percentage of difference are the minimum values which are 0.00354, 0.0110, 0.3546 respectively. These promising findings suggest the adequacy and potential of these mathematical techniques to address this type of problem. Although, we have demonstrated the potential of the Fuzzy-based approach, but still we need to extend this research with more data and different other types of Risk Analysis to state and conclude its promising approach.

Keywords: Information Assurance, Confidentiality, Integrity, Availability, CVSS, Sugeno, Hybrid Optimization, Subtractive Clustering, Cascaded Model, Cross Validation, Risk Level.

ان الأحداث الأخيرة مثل SQL Slammer و Blaster التي انتشرت عبر الانترنت وتصيب أكثر من 90% من الأنظمة التي تكون عرضة للاصابة أثارت اهتماما دراماتيكيا لتأمين المعلومات. تلك الأحداث تزيد من عدد الاخفاقات لمؤسسات رفيعة المستوى لعدم توافر البيانات، المعلومات والمعلومات الاستخبارية المتوفرة لاتخاذ القرارات في اللحظات الرئيسية.

وعلى الرغم من إنفاق ملايين الدولارات على جدران الحماية، وتقنيات التشفير، وبرامج كشف التسلسل الا ان ثغرات البنية التحتية للمعلومات وحوادث الاختراق مازالت تحدث.

لذلك، من الضروري للمؤسسات ان توفر الثقة واليقين لمعلوماتها، وضمان وجود مستويات معينة من توافرها، سلامتها، توثيقها، سريتها وعدم التنصل من أصول معلوماتها ضد مخاطر غير مقبولة. هذا ما يشار اليه بتأمين المعلومات (IA). فهي تؤكد أن التقنيات الأمنية المستخدمة هي في الواقع فعالة وان النظام يمكن أن يعهد إليها مهام المعالجة للمعلومات الهامة. تقدم هذه الدراسة نموذجا لتقدير مستوى المخاطر والتي تستمد باعتبار مستوى الخطر هو الاحتمال الشرطي من تقدير تردد حدوثه وتأثيره. وتستمد تقديرات تردد الحدوث وتأثيره من مجموعة من السمات المحددة من نظام تقدير الثغرات الشائعة (CVSS). هذا النموذج يتنبأ قيمة مستوى خطر الثغرات بناء على مستوى الخدمة وقدرة احتياج تلك الخدمة من المؤسسة لتحقيق مهامها الحرجة. لذلك فان تأمين المعلومات يفرض على المؤسسة مراجعة مهامها والتهديدات الأمنية وتأمين قدراتها اللازمة ضد المخاطر والتي تشكل تهديدات. تقدم هذه الدراسة نهج متطورة تعتمد على التقنيات الضبابية (Fuzzy) والتي تتضمن الأنظمة الاستنتاجية العصبية الضبابية القابلة للتكيف (Adaptive Neuro Fuzzy).

بيانات قديمة وحقيقية أستخدمت من قاعدة بيانات الثغرات العالمية لتطوير واختبار النماذج المقترحة. نماذج "سوجينو" (Sugeno) بتقنيات مختلفة تشمل "سوجينو" بواسطة تقنية التحسين الهجينة أو المركبة (Hybrid)، نموذج "سوجينو" باستخدام تقنية الكتل الطرحي (Subtractive Clustering)، نموذج "سوجينو" باستخدام الطريقة المتعاقبة (تقنية التكتل الطرحي ثم تحسين النموذج باستخدام تقنية التحسين الهجينة)، واخيرا نموذج "مامداني" (Mamdani).

كل النماذج المطورة تم فحص دقتها. مقاييس مختلفة تم ملاءمتها مثل معامل التوافق (Correlation Coefficient) لقياس مدى تطابق القيمة الحقيقية مع المتنبأ، متوسط نسبة القيمة المطلقة "ميب" (MAPE)،

ومتوسط مربع الجذر "أرمسي" (RMSE)، ونسبة الاختلاف بين البيانات الحقيقية والبيانات المتنبأة. وقد أشارت النتائج أن أداء نموذج مستوى الخطر لتأمين المعلومات نموذج "سوجينو" الذي بني باستخدام التحسين الهجينة حصل على أفضل النتائج حيث ان قيمة معامل درجة التوافق بلغت 0.993 وهي القيمة الأعلى بين بقية النماذج وبالمثل فان قيم "ميب"، "أرمسي" ونسبة الاختلاف وهي القيم الاصغر بين النماذج وهي 0.00354، 0.0110، 0.3546 على التوالي.

وتشير هذه النتائج المبشرة كفاية وإمكانات هذه التقنيات الرياضية لمعالجة هذا النوع من المشاكل، وعلى الرغم من أننا أظهرنا إمكانات النهج القائمة على الضبابية، لكننا لا زلنا بحاجة لتمديد هذا البحث مع المزيد من البيانات وأنواع أخرى مختلفة من تحليل المخاطر لتوضيح واستنتاج نهجها المبشر.

Table of Contents

Declaration	i
Acknowledgements	ii
Abstract	iii
ملخص الدراسة	v
Table of Contents	vii
List of Tables	xi
List of Figures	xii
List of Appendices	xiv
Acronyms and Abbreviations	xv
Chapter One : Introduction	1
1.1 Preface	1
1.2 IA Examples	4
1.2.1 SQL Slammer	4
1.2.2 Blaster	5
1.2.3 SOBIG.F	5
1.3 Motivation	6
1.4 Problem Statement	7
1.5 Thesis Contribution	9
1.6 Research Constraints	10
1.6 Thesis Methodology	10
1.7 Thesis Organization	14
Chapter Two : Literature Survey	15
2.1 Related Work	15

2.2	Related Work Analysis.....	29
Chapter Three : Common Vulnerability Scoring System (CVSS) V2		31
3.1	Introduction.....	31
3.2	Base Group.....	32
3.2.1	Access Vector.....	33
3.2.2	Access Complexity.....	34
3.2.3	Authentication	34
3.2.4	Confidentiality Impact.....	35
3.2.5	Integrity Impact	36
3.2.6	Availability Impact.....	36
3.3	Temporal Metrics.....	37
3.3.1	Exploitability.....	37
3.3.2	Remediation Level	38
3.3.3	Report Confidence.....	39
3.4	Environmental Metrics.....	40
3.4.1	Collateral Damage Potential.....	40
3.4.2	Target Distribution	41
3.4.3	Security Requirements	42
3.5	Summary	43
Chapter Four : Data Profiles.....		45
4.1	Introduction.....	45
4.2	Data Collection	46
4.3	Variable Selection.....	49
4.4	Data Formatting	52
4.5	Cross Validation.....	53

4.6	Summary	54
Chapter Five : Development of Models		55
5.1	Introduction.....	55
5.2	Sugeno Models with Hybrid Optimization Technique	56
5.2.2	Initial Impact Sugeno Models with Hybrid Optimization Technique.....	60
5.2.3	Updated Frequency Sugeno Models with Hybrid Optimization Technique	62
5.2.5	Vulnerability Risk Sugeno Models with Hybrid Optimization Technique.....	65
5.3	Sugeno Models with Subtractive Clustering.....	66
5.3.1	Initial Frequency Sugeno Models with Subtractive Clustering	67
5.3.2	Initial Impact Sugeno Models with Subtractive Clustering.	68
5.3.3	Updated Frequency Sugeno Models with Subtractive Clustering.	69
5.3.5	Updated Impact Sugeno Models with Subtractive Clustering.	69
5.3.4	Vulnerability Risk Sugeno Models with Subtractive Clustering.	70
5.4	Sugeno Cascaded Models with Subtractive Clustering and Hybrid Optimization.	70
5.4.1	Initial Frequency Sugeno Models with Subtractive Clustering and Hybrid Optimization technique.	71
5.4.2	Initial Impact Sugeno Models with Clustering and Hybrid Optimization technique.....	72
5.5	Mamdani Fuzzy Inference Method.....	72
5.5.1	Initial Frequency Mamdani Fuzzy Inference Model.....	73
5.6	Simulink Model: IA Vulnerability Risk Assessment.....	74
Chapter Six : Results and Discussions		78
6.1	Introduction.....	78
6.2	Results and Comparisons between the Developed Models.	78
6.3	Comparison with Other Studies	85
Chapter Seven : Conclusions and Suggestions for Future Works		89

7.1	Conclusions.....	89
7.2	Suggestions and Future Work.....	91
	Bibliography	92
	Appendix A: Information Assurance Definition from Different Perspectives	96

List of Tables

Table (2.1) Summary of related works	25
Table (3.1) Access Vector Scoring	33
Table (3.2) Access Complexity Scoring	34
Table (3.3) Authentication Scoring.....	35
Table (3.4) Confidentiality Impact Scoring	35
Table (3.5) Integrity Impact Scoring	36
Table (3.6) Availability Impact Scoring	37
Table (3.7) Exploitability Scoring Evaluation	38
Table (3.8) Remediation Level Scoring Evaluation	39
Table (3.9) Report Confidence Scoring Evaluation.....	40
Table (3.10) Collateral Damage Potential Scoring Evaluation.....	41
Table (3.11) Target Distribution Scoring Evaluation	42
Table (3.12) Security Requirements Scoring Evaluation	43
Table (4.1) Summery of the NVD Database Contents (NVD 2004)	47
Table (4.2) Vulnerabilities for All Services of Email Asset.....	49
Table (4.3) Available Datasets with Tr and Ts Datasets	54
Table (5.1) Sugeno Model Parameter for Initial Frequency Developed Model	58
Table (5.2) ANFIS Result for Initial Frequency Sugeno Developed Model	58
Table (5.3) Results for Initial Frequency Models with Hybrid Optimization	60
Table (5.4) Results for Initial Impact Model with Hybrid.....	61
Table (5.5) Results Error for Initial Frequency Model with Subtractive Clustering	67
Table (5.6) Results Error for Initial Impact Model with Subtractive Clustering.....	69
Table (5.7) Results Error for Initial Frequency Model with Subtractive Clustering and Hybrid Method	71
Table (5.8) Results Error for Mamdani Model of Initial Frequency	74
Table (6.1) The Correlation Measures for the Developed Initial Frequency Model	80
Table (6.2) The MAPE Measures Table for inital Frequency Models	81
Table (6.3) The RMSE Measures Table for Initial Frequency Models	82
Table (6.4) The Percentage of Differences Table for initial Frequency Models	83

List of Figures

Figure (1.1) The dimensions of the IA model.....	3
Figure (1.2) System internal and environmental fault produce a different risk level sources that affect the critical system behavior.	8
Figure (1.3) Impact and frequency estimation using subset of attributes from the CVSS	11
Figure (1.4) General Block Diagram for Predicting IA Risk Level	11
Figure (3.1) CVSS metric groups	31
Figure (3.2) CVSS calculation process	32
Figure (4.1) Data Preprocessing Procedure Stages.....	45
Figure (4.2) Openwebmail Email Asset.....	46
Figure (4.3) Some Vulnerabilities for Email Asset	49
Figure (4.4) Input-Output Variables for each Stage	50
Figure (4.5) Aggregated Vulnerability Risk for One Service.....	51
Figure (4.6) IA Risk Level Assessment for Email Asset.....	52
Figure (4.7) Frequency Estimation Rate and Numeric Values	52
Figure (4.8) Impact Estimation Rate and Numeric Values.....	52
Figure (4.9) Sample of Training Dataset File	53
Figure (5.1) General Developing Training Block Diagram for Sugno Model with Hybrid Technique.....	57
Figure (5.2) Initial Frequency FIS Model.....	58
Figure (5.3) Initial Frequency Inputs (Access Vector, Access Complexity, Authentication) MF's	59
Figure (5.4) Initial Impact FIS Model.....	60
Figure (5.5) Initial Impact Model Input (Confidentiality, Integrity, Availability) MF's	61
Figure (5.6) Initial Impact Sugneo FIS Model with Hybrid Optimization Technique	62
Figure (5.7) Updated Frequency FIS Model.....	63
Figure (5.8) Updated Frequency Inputs MF's Sugeno Model with Hybrid Method.....	63
Figure (5.9) Updated Impact Sugrno Model Inputs MF's.....	64
Figure (5.10) Updated Impact FIS Model.....	65

Figure (5.11) Frequency and Impact Input MF's for Risk Vulnerability Sugeno Model	65
Figure (5.12) Risk Level FIS Model.....	66
Figure (5.13) General Block Diagram for Developing/ Training Clustering Models	67
Figure (5.14) Initial Frequency Sugeno Model MF's with Clustering.....	68
Figure (5.15) Initial Impact sugeno model with clustering Inputs MF's	68
Figure 5.16 Updated Frequency Sugeno Model with Clustering Inputs MF's	69
Figure (5.17) Updated Frequency Sugeno Model with Clustering Inputs MF's.....	69
Figure (5.18) Vulnerability Risk Sugeno Model with Clustering Inputs MF's	70
Figure (5.19) A General Block Diagram for Developing/Training Cascaded Clustering with Hybrid Optimization	70
Figure (5.20) Initial Frequency Sugeno Model with Clustering and Hybrid inputs MF's.....	71
Figure (5.21) Initial Impact sugeno model with clustering and Hybrid Inputs MF's	72
Figure (5.22) Initial Frequency Mamdani Model Inputs MF's	74
Figure (5.23) Calculate Frequency Value used Simulink.....	75
Figure (5.24) Calculate Impact value using Simulink	76
Figure (5.25) IA Risk Level using Simulink	76
Figure (5.26) IA vulnerability Risk Assessment Simulink Model	77
Figure (6.1) The Correlations Measures for Results Obtained from Initial Frequency Models...	79
Figure (6.2) The MAPE Measures Chart for initial Frequency Models	81
Figure (6.3) The RMSE Measures Chart for Initial Frequency Models	82
Figure (6.4) The Percentage of Differences Chart for initial Frequency Models.....	83
Figure (6.5) All Measures Values (CC, MAPE, RMSE, Percentage of Differences) for Initial Frequency Models.....	84

List of Appendices

Appendix A	Information Assurance Definition from Different Perspectives	96
------------	--	----

Acronyms and Abbreviations

AC	Access Complexity
AHP	Analytic Hierarchy Process
ANFIS	Adaptive Neuro Fuzzy Inference System
AR	Availability Requirement
AU	Authentication
AV	Access Vector
BBN	Bayesian Belief Network
BP	Back Propagation
BR	Bayesian Regulation
CC	Correlation Coefficient
CDP	Collateral Damage Potential
CERT	Computer Emergency Response Team
CIA	Confidentiality, Integrity, Availability
COMSEC	Computer Security
CR	Confidentiality Requirement
CSIRT	Computer Security Incident Response Team
CVSS	Common Vulnerability Scoring System
DAG	Directed Acyclic Graph
DMZ	Demilitarized Zone

E	Exploitability
ECP	Evidence Collection Path
FEMRA	Fuzzy Expert Model for Risk Assessment
GNS	Graphical Network Simulator
IA	Information Assurance
INFOSEC	Information Security
INFOSSEC	Information System Security
IR	Integrity Requirement
IS	Information System
IT	Information Technology
KRI	Key Risk Indicator
LM	Levenberg Marquardt
MAPE	Mean Absolute Percentage Error
MC	Management Confidence
MF	Misuse Frequency
MF's	Membership Functions
MI	Misuse Impact
ML	Management Level
NETSEC	Network Security
NN	Neural Network
OCTAVE	Operationally Critical Threat, Asset, and Vulnerability Evaluation System

OWASP	Open Web Application Security Project
RMSE	Root Mean Square Error
SDLC	Software Development Life Cycle
TD	Target Distribution
ToE	Target of Evaluation
WASE	Web Application Security Evaluator
WSMS	Website Security Mining System

Chapter One : Introduction

1.1 Preface

In this digital era where the valuable information is being transferred, stored and processed electronically than any other forms and where the organizations recognize the information system importance to successfully carry out their missions, the ISO 27002 defines information as an asset, which, like other important business assets (Tajuddin, Olphert et al. 2015), is essential to an organization's business and consequently needs to ensure the security services (confidentiality, integrity, availability, authentication, non-repudiation) of modern information systems and applications. It is found that the number of large companies that suffered a security incident during 2008 - 2010 increased from 72% up to 92% (PWC 2014). Furthermore, the average cost of the worst security incident in large companies increased from 650 thousand dollars to 5.9 million dollars (PWC 2014). Installing firewalls, antiviruses and applying different security technologies and mechanisms are only addressing the protection of information and information systems against unauthorized activities such as disclosure, transfer, modification, or destruction. These activities affect the security needs of information systems and cannot deliver the level of information assurance that modern information systems require. Therefore, it is a necessity for organizations to provide confidence and certainty of its information and ensure a certain levels of availability, integrity, authentication, confidentiality, and non-repudiation of an organization's information assets against unacceptable risk. This is referred to as information assurance (IA). It is assuring that the security mechanisms are actually effective and the system can be entrusted with the processing tasks on the critical information. IA imposes on an organization a review of its mission and threats and the securing of its needed capabilities against the risks those threats pose. The complexity of IA represents a myriad of considerations and decisions that exceed technological

advancement, economic, social cultural, institutional, organizational, and educational dimensions. This may explain the difficulties associated with IA. Appendix A summarizes some of the IA definitions from different perspectives. It shows that the definitions mostly share five pillars of security goals (Confidentiality, Integrity, Availability, Authentication and Non-Repudiation). The interrelation between IA and Information Security (INFOSEC) has been presented as a survey among the IA and INFOSEC professionals (Cherdantseva and Hilton 2013). These professionals were invited to describe the relationship between the two disciplines. The summary of responses concludes that the IA is a part of INFOSEC. Whereas, Peng and coauthors (Peng Liu, Meng Yu et al. 2001) presented that the concept of IA is much broader than the INFOSEC. Whereas, the goal of INFOSEC technologies is to prevent attacks from happening and focus on technological tools, the goal of IA is to ensure that even if some attacks intrude into an Information System (IS), IS can still operate. Several available definitions in the literature are presented in Appendix A. In this research, we adopt the (Peng Liu, Meng Yu et al. 2001) definition as indicates about measures of an acceptable level of security services of an information to develop an approach that measures a quantifiable IA risks by conducting an IA vulnerability assessment which involves looking at the vulnerabilities in organization and understanding the mitigation of those vulnerabilities. IA is defined as information operations that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. This includes the provision for the restoration of information systems by incorporating protection, detection and reaction capabilities (IATF 1999). The adopted IA model is as shown in *Figure (1.1)*. IA model addresses five dimensions as follow:

1. **Availability:** The **availability** of information is a measure of how often information is capable of being accessed as needed. It involves information being where users need it when they need it. Availability is a key component to IA.
2. **Integrity:** **Integrity** of information is the probability that the information will be correct when accessed.
3. **Confidentiality:** **Confidentiality** is a measure of how well protected information is from being read, transmitted, viewed, or interpreted by unauthorized persons or organizations.
4. **Authentication:** **Authentication** is the process of verifying the identity or other attributes claimed by or assumed of an entity (user, process, or device), or to verify the source and integrity of data.
5. **Non-repudiation:** **Non-repudiation** is to assure that the sender of information is provided with proof of delivery and the recipient is provided with proof of the sender's identity, so neither can later deny having processed the information.

From the definitions of the five security goals as previously mentioned above and the IA, IA also assumes a security of authenticity, as it applies to confidentiality and integrity, and non-repudiation, as it applies to integrity (Matthews 2004).



Figure (1.1) The dimensions of the IA model
 Source (Maconachy, Schou et al. 2001)

1.2 IA Examples

The next three sections describe three IA cases and their impacts and illustrate the importance of risk assessment.

1.2.1 SQL Slammer

SQL slammer is the fastest self-propagation worm which spreads across the internet and infects more than 90 percent of the vulnerable systems within the first 10 minutes and takes many small networks offline with its scanning (Moore, Paxson et al. 2003). Slammer was doubling in size every 8.5 seconds and within 3 minutes it had scanned approximately 55 million IP addresses per second causing disruption of the internet. This worm targets the unpatched Microsoft SQL server which increases traffic on the UDP port 1434 and causes heavy network traffic that can slow down network performance and lead to denial of service and affect the availability of the server. Another way that can use the SQL slammer is to exploit the buffer overflow vulnerability in the SQL server which affects the availability of the server by generating a random IP addresses and sending itself to those addresses from the infected system. In this case the attacker use the exploitability tool (worm) to exploit the SQL server (asset) through the vulnerability (UDP port 1434) which make a loss in availability and reduce the IA of the asset. Therefore this threat will affect the mission of the organization. Thus, reviewing the IA vulnerabilities of the SQL server and securing it by applying the patch that prevent the attack to exploit the port. It will prevent the SQL slammer exploitation that cause denial-of-service attacks.

1.2.2 Blaster

The blaster worm is a virus program that targets the Microsoft operating system by exploiting the flow in Microsoft Remote Procedure Call (RPC) process using Transmission Control Protocol (TCP) port number 135 that shuts down the infected computers (Bailey, Cooke et al. 2005). It affected at least 500000 Microsoft computers around the worlds. Maryland motors vehicle administration shut its offices for a day. In eight days after the blaster worm propagated, the estimated cost of damages neared 2\$ billion (Bailey, Cooke et al. 2005). Exploited this vulnerability will affect the critical systems used by the infected computer which affect the availability capability of the systems. This vulnerability will reduce the IA and affect the availability security service of the information. This vulnerability was exposed by the Last Stage of Delirium (LSD) security group and later by Microsoft which released two different patches (MS03-026 and MS03-039) on its website.

1.2.3 SOBIG.F

SOBIG is another worm targets the Microsoft windows computer and infected millions of internet-connected systems. The CNN international news posted in august 2003 that the SoBig.F computer virus which has already overwhelmed hundreds of thousands of computers worldwide has become the fastest spreading virus ever with experts warning the worst is yet to come (CNN 2003). This worm transmit by email and arrives with various subject headers, such as: Your details, Thank you!, Re: Thank you!, Re: Details, Re: Re: My details, Re: Approved, Re: Your application, Re: Wicked screensaver or Re: That movie. The user triggers the worm when the user open the email which then flood the message to all other addresses in the email address book and infected more computers. This worm also affect the availability security service of information by reducing the

performance of the network and overwhelmed the network. Therefore, this worm will reduce the IA of the information.

1.3 Motivation

This motivation comes firstly from the necessary needs of the high-profile Palestinian organizations such as Hadarah Company. The information security manager in Hadarah Company announced to the researcher that the company are having efforts towards applying IA since four years. This was the motivation for research in this subject. Secondly from the background and experience of the researcher in information system, security, networks and developments, in addition to working in several organizations with different information systems which were based on critical and confidential information such as medical IS, accounting IS, university management system ...etc. In these information systems, there were no evidence of the effectiveness of the security features. As the importance of medical system, I chose the medical IS to mention the trends of hackers. Each medical IS record consists of different important fields of patients including names, identity number, diagnosis and medications code, billing information and insurance company. Therefore, the hackers discover a new way to make money by offering all data records of patients for sale to another (counterpart competitor medical center) or another insurance company. Also the fraudsters can use this data to create fake IDs to buy medical equipment or drugs that can be resold, or they combine a patient number with a false provider number and file made-up claims with insurers according to experts who have investigated cyber-attacks on healthcare organizations. According to an annual survey by the Ponemon Institute think tank on data protection policy, the percentage of healthcare organizations that have reported a criminal cyber-attack has risen to 40% in 2013 from 20% in 2009 (Luftman and Ben-Zvi 2010).

In addition, there was a lack of basic IA understanding in the Palestinian organizations. Worldwide, organizations spend thousands of dollars for a security software or consultation from different vendors, but these securities mostly have different vulnerabilities that make a crisis. Whereas, (Secunia 2014) showed the number of vulnerabilities detected was 13,073, discovered in 2,289 products from 539 vendors. According to (Secunia 2014), 45% of the vulnerabilities have increased in a five year trend, and a 32% increase from 2012 to 2013. Since 2012, the amount of vulnerable vendors have increased by 13% and the amount of vulnerable products has decreased by 6%. Also this thesis is important to promote the IA hot field and hopefully support the Palestinian vendors of the information systems whereas the vendor may lose around 0.6% from the value in stock price when a vulnerability is reported (Telang and Wattal 2005). Therefore, it is an imperative to research for this topic to publish and to be implemented by the large organizations in Palestine such as JAWWAL, PALTEL, Watania, JEDCO, Insurance Companies ...etc. to assessment the vulnerabilities and mitigate the risk.

1.4 Problem Statement

The complexity of the information assurance arises from the large number of assets, the connections among each asset (software, hardware, network, database, etc.), and from the interconnections between assets. Thus, each asset has a number of vulnerabilities from different sources which affect the system behavior of asset. *Figure (1.2)* illustrate the internal and environmental influence fault in a critical system such as an application failure; operating system failure and hardware failure that produce a risk level and affect the system behaviors.

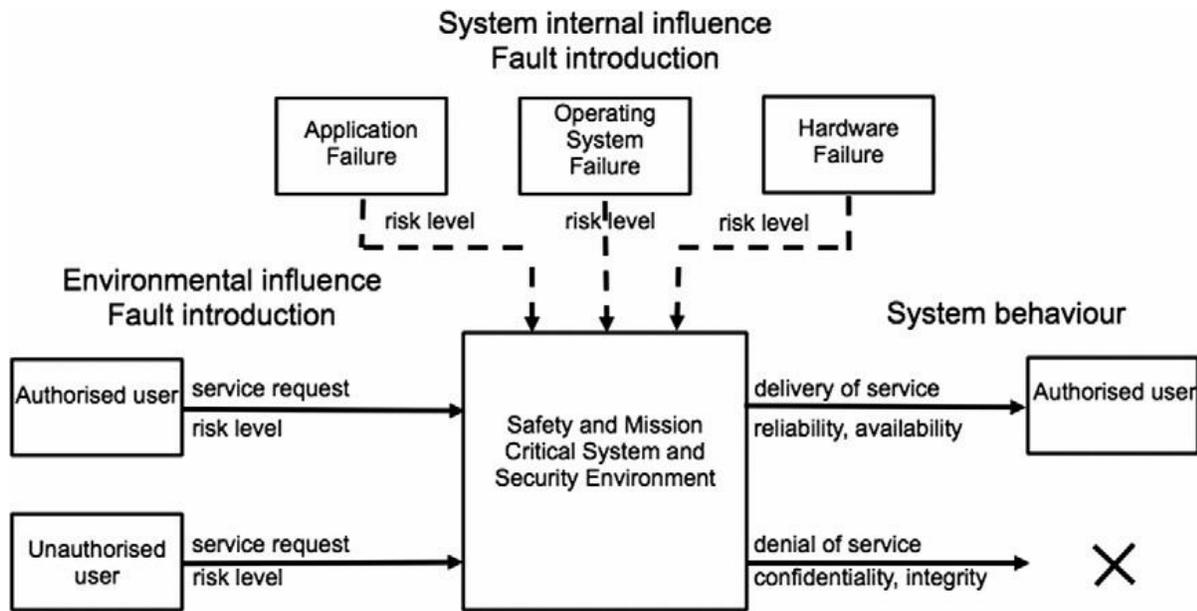


Figure (1.2) System internal and environmental fault produce a different risk level sources that affect the critical system behavior.
 (source:(Houmb, Nunes Leal Franqueira et al. 2008))

Organizations consider that using protection tools including firewalls, filtering routers, password protections, encryptions, access controls and file permissions are enough for protection. Yet a small vulnerability can be compromised in the confidentiality, integrity and availability of the information. For example, inserting a series of two dots “..” into a URL in a company’s web server application by a hacker can make the server navigate out of its document directories and retrieve a database of user names and encrypted passwords (Martin 2001).

Organizations must realize what they need to be able to do with their wealth of information, and what risks they are willing to take in order to maintain those capabilities. Mostly, the IA is determined primarily by offline evaluation processes such as analyses (CERT/CC Security Capability Model, NSA INFOSEC Assurance Capability Model), testing (penetration testing) and experimentation (red team experiments). Those models are largely qualitative.

This research will address the following issues:

1. How to measure an acceptable level of IA for each vulnerability?
2. How to find the total IA risk level for each asset?
3. What are the IA metrics applies across all systems?

1.5 Thesis Contribution

The ultimate goal of this thesis work is developing an IA vulnerability risk assessment approach that quantitatively evaluates the vulnerability in each asset and measures the IA level of acceptable risk. This major contribution will assist an organization to secure its information assets against unacceptable loss of availability, integrity, confidentiality, authenticity and non-repudiation. Conducting an IA vulnerability assessment involves looking at the vulnerabilities on the organization and understanding the mitigation of those vulnerabilities that could create unacceptable losses. In order to achieve the IA assessment for vulnerabilities, this thesis has contributed with the followings:

1. Building an IA vulnerability assessment model that combines the impact and frequency of the attack for each vulnerability. Before deciding which vulnerabilities to be concerned with, the organization needs to understand what are the capabilities' needs. Therefore, the organization understands what risks are acceptable. This thesis presents the developed models using NEUROFUZZY BASED MODELING TECHNIQUES. Several model have been developed such as: 1. Sugeno with Hybrid optimization techniques 2. Subtractive clustering 3. Cascaded model (clustering with hybrid optimization techniques) 4. Mamdani model also was developed but only for the first stage 5. Check the adequacy of the developed model to demonstrate their performance, four measurements have been used to effectively check the adequacy of results (CC, MAPE, RMSE, Percentage of differences).

2. The total risk for each asset can be calculated by aggregating all risks for all vulnerabilities by Mamdani system which takes the maximum risk value.
3. Calculation of risk security needs.
4. By observing the changes in the input values to figure out the effect of IA assessment.

1.6 Research Constraints

As this thesis discusses the topics in depth, it is also important to indicate the boundaries and limitations of this research. The restrictions are as follow:

Firstly, the information assurance domain at the organizational level is dynamic, highly connected to myriad of considerations and decisions that exceed technological advancement, and surpass legal, political, economic, social, cultural, institutional, organizational, and educational dimensions. This may explain the difficulties associated with the IA.

Secondly, the lack of information available, organizations did not accept the thesis request to collect some data from its Intrusion Detection System (IDS) or Intrusion Prevention System (IPS) due to the data confidentiality. Therefore, we have used a specific data that is available on the internet from National Vulnerability Database (NVD) (NVD 2004) which includes data for some vendors such as Microsoft, Symantec, Cisco, etc.

1.6 Thesis Methodology

In this thesis, the proposed approach is to quantify the IA risk level by conducting the vulnerabilities assessment for each critical asset. The risk level will be calculated by combining the frequency and impact of the attacks for each vulnerability in the asset by making use of the subset of attributes from CVSS as shown in *Figure (1.3)*. The CVSS attributes of each groups have rating terms and values (see Appendix A). This approach has been implemented using the

NEUROFUZZY BASED MODELING TECHNIQUES. Neurofuzzy approach combines the fuzzy inference system (FIS) and neural network (NN) making use of different learning procedure. Figure (1.4) shows the general block diagram of predicting IA risk level. The ANFIS approach was applied in modeling IA vulnerability assessment based on CVSS data that were derived from different security companies, software vendors, hardware vendors and researchers.

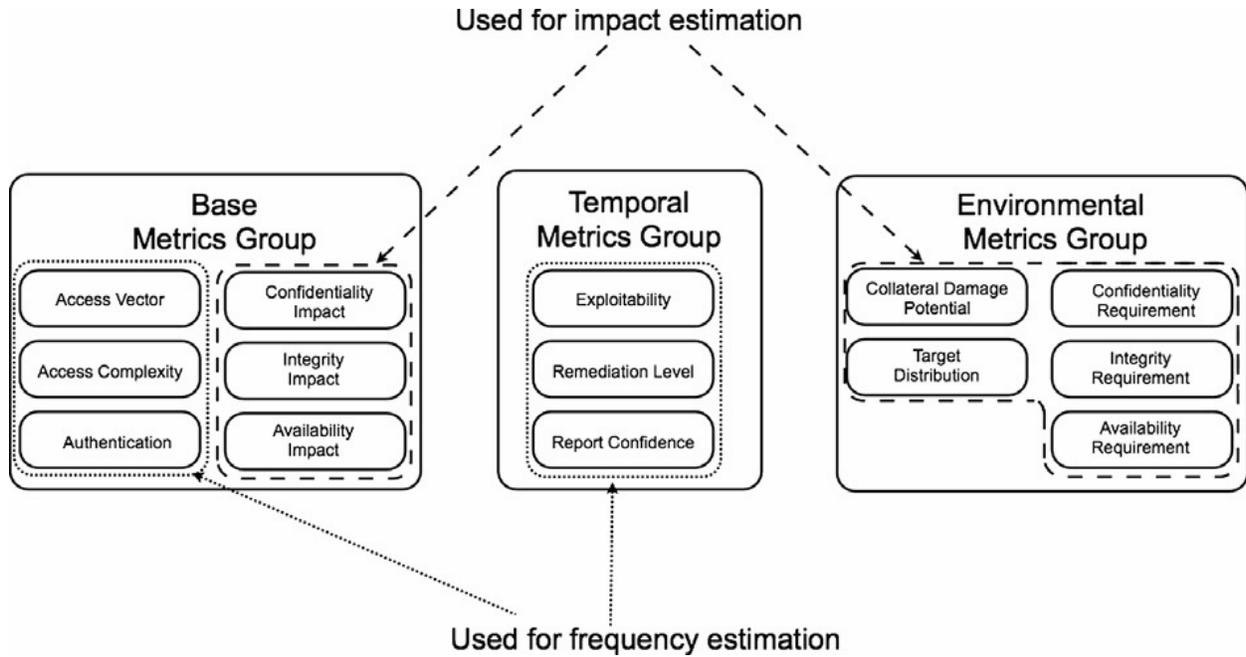


Figure (1.3) Impact and frequency estimation using subset of attributes from the CVSS (source:(Houmb, Nunes Leal Franqueira et al. 2008))

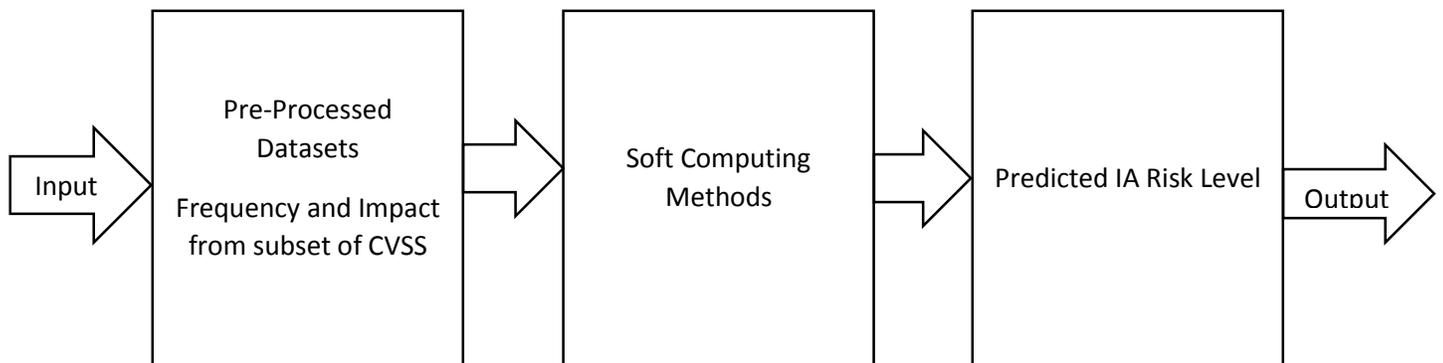


Figure (1.4) General Block Diagram for Predicting IA Risk Level (drawing by the author)

To achieve the contribution in this thesis, the approach can be summarized as follow:

1. Collecting some real vulnerabilities for available common IS application that is used in an organization retrieved from National Vulnerability Database (NVD).
2. Analyzing of the data and pre-processing the system input before the training stage.
3. Studying the initial and updated factors of CVSS metrics that affect the frequency and impact for each vulnerability.
4. Computing Initial Frequency: Identifying the input (Access Vector, Access Complexity, and Authentication) metrics from base group of CVSS v2 and the output (Exploitability). The three input and one output factor are used to develop a Sugeno Fuzzy Inference System (FIS) model using the Adaptive Neuro Fuzzy Inference System (ANFIS) with different optimization techniques to predict the initial frequency risk.
5. Computing Updated frequency: Identifying the input (Exploitability tools, Remediation Level, and Report Confidence) metrics from temporal group of CVSS v2 combined with the initial frequency and the output (Temporal Score). The four input and one output factor are used to develop a Sugeno Fuzzy Inference System (FIS) model using the Adaptive Neuro Fuzzy Inference System (ANFIS) with different optimization techniques to update the initial frequency risk.
6. Computing Initial Impact: Identifying the input (Confidentiality Impact, Integrity Impact, and Availability Impact) metrics from base group of CVSS v2 and the output (impact). The three input and one output factor are used to develop a Sugeno Fuzzy Inference System (FIS) model using the Adaptive Neuro Fuzzy Inference System (ANFIS) with different optimization techniques to predict the initial impact risk.

7. Computing Updated Impact: Identifying the input (Confidentiality Requirement, Integrity Requirement, Availability Requirement and collateral damage potential) metrics from environmental group of CVSS v2 and the output (updated impact) The four input and one output factor are used to develop a Sugeno Fuzzy Inference System (FIS) model using the Adaptive Neuro Fuzzy Inference System (ANFIS) with different optimization techniques to predict the updated impact risk.
8. The update frequency and updated impact are also combined to predict the final risk level for one vulnerability.
9. All final risks of vulnerabilities that are associated with one asset are also combined as an input to a Mamdani fuzzy system to calculate the IA risk level.
10. Evaluating the predicted output using the developed models and calculating several measures including the CC, MAPE, RMSE and Percentage of difference.
1. A simulation tool has been developed making use of the Simulink tools in matlab R2013a to compare the values at each stage.

Thus, this approach includes four stages to accomplish the IA assessment for one asset, starting with calculating the frequency, impact, risk value for one vulnerability and ending with The total risk for asset is aggregated the risks of all vulnerabilities in this asset as illustrated below:

Stage 1: computing the initial frequency and initial impact.

Stage 2: computing the updated frequency and updated impact.

Stage 3: calculates the risk level by combined the updated frequency and updated impact.

Stage 4: aggregate the vulnerabilities for one asset to Mamdani fuzzy system.

1.7 Thesis Organization

The research is mainly divided into six chapters, naming them: Literature Survey, Common Vulnerability Scoring System (CVSS), Data Profiles, Developing the Models, Results and Discussions and finally the Conclusions and Future Works. The search is organized as follows:

Chapter two gives an overview of the research approaches for the risk assessments. Also this chapter consists of different approaches accomplished by researchers that assess the risks with different data used. Chapter three provides a brief overview of the standard CVSS (Common Vulnerability Scoring System) and its metrics. Chapter four introduces the historical data that have been used, and the pre-processing and manipulation of data is discussed. Chapter five, presents in details, the various models that have been developed and make use of the metrics of the CVSS as inputs. These models include Sugeno models with hybrid optimization technique, Subtractive Clustering, cascading two techniques and Mamdani model. In addition, the error measures that were used to compare between the actual and predicted data was discussed in the same chapter. While chapter six, presents and discusses the results of the developed models, with brief comparisons with related findings published by other researchers chapter seven concludes and highlights the research results obtained and suggests future work for this important research work.

Chapter Two : Literature Survey

The purpose of this chapter is to give an overview of the research approach. This chapter is divided into two sections. The first sections shows related work which are general risk assessment approach in security and risk assessment in IA. The different approaches are listed and described below. There are a lot of methods and approaches make an offline assessment by analyzers according to guidelines. The second section is analysis of the related work.

2.1 Related Work

Chen and Tian (Chen and Tian 2015) have introduce a comprehensive introduction of condition of attack state and how they analysis it from massive xml file. They investigate the NIST National Vulnerability Database (NVD). Multi-attribute-based classification method supports mining privilege level of vulnerability. Owing to associated analysis, it combined directly and indirectly of vulnerability threats which makes evaluation more convinced. The initial approach was using model checking techniques A model checker could assist engineers to identify individual design flaws in a model of a system. Using model checkers, the researchers could get away of custom special purpose tools for attack graph generation. To check if the system has a bug or side effect, the model is completed looked after whether it meets a correctness specification. The model is a state staff defined by variables, initial values for the variables and a description of the 15 conditions under which variables may change value. When the variables change value they cause a state transition. The sum of all possible states of a state machine is the state space. The model can be automatically checked by a model checker against a correctness specification if the model has any flaws. The correctness specifications are expressed in propositional temporal logic. The model checker performs an exhaustive search through the state space to determine that each state satisfies

the correctness specification. If the correctness specification is not satisfied, the model checker will give a counterexample execution, showing the sequence of states that lead to the violation of the correctness specification.

Cho (Cho 2015) has implemented a system named the website security mining system, which leverages a web crawling algorithm to analyze web URL and e-mail address leaks through black-box testing of 20 well-known universities' websites. Based on their data, academic website maintainers can be clearly informed about what kind of danger they are exposed to, which URLs are highly in danger, and the need to patch the website to protect against vulnerabilities and prevent academic resources from attacks. The WSMS is designed to combine search engine technology with vulnerability testing to automatically spider and assess the security of a target website. This study present the Static and Dynamic Mining. Both of the Static Mining and Dynamic Scanning modules can lever the system's vulnerability inspecting function, which has two parts: known website vulnerability inspection and SQL injection inspection. The former compiles a database of open source website vulnerabilities into an XML file which is used to inspect the website to see whether it has the same vulnerability.

Al-Mahrouqi and coauthors (Al-Mahrouqi, Tobin et al. 2015) simulate an SQL injection attack scenarios in a complex network environment. They designed and simulated a typical Demilitarized Zone (DMZ) network environment using Graphical Network Simulator (GNS3), Virtual Box and VMware workstation. They collected the network logs by using Wireshark define an attack pathway prediction methodology that makes it possible to examine the network artifacts collected in case network attacks. This study use the Prediction Investigation Approach to predict and trace the source of the attack or illegal activities in the computer network. The idea behind this approach

is to identify the Evidence Collection Path (ECP) by using Evidence Collection Process Model (ECPM).

Prasad (Prasad 2014) has used Genetic Algorithm to generate dynamic IP for the network to avoid unauthorized data transfer and prevent from attack. The Intrusion Detection System can be viewed as a rule-based system (RBS) and Genetic Algorithm can be viewed as a tool to help generate knowledge for the RBS. This project shows how network connection information can be modeled as chromosomes and how the parameters in genetic algorithm can be defined in this respect. He implemented a server side interface which is solely under the control of the administrator. Any transaction in the network will be monitored by the Server. It receives the packet and reads the header information from the packet such as the Destination address, Source address, Port no. It sends each and every Inflowing packets header information's to the chromconvert module and then receives the converted real-time Chromosomes. The real time chromosomes are checked with the rule sets. If the particular chromosomes matches with the rules provided in the rule set, it takes the decision of whether allow or block depending on which rule set it matches.

Lee and coauthors (Lee 2014) has reviewed various models including AHP, fuzzy, neural network, group decision making, software computing and hybrid model. Lee has presented the quantitative and qualitative models in risk assessment and has concluded that researchers prefer the AHP approach. Furthermore, he proposed a new method by using a hybrid models such combined between AHP and fuzzy system.

In his study (Rani 2013), Rani has proposed a neuro-fuzzy approach to estimate the software risk in all stages of software development life cycle (SDLC). Firstly he used the fuzzy inference system with 17 input risk attribute. The input attributes were identified by a fuzzy terms, rules and output. After the Fuzzy Inference system he created then Neural Network based three different training

algorithms: BR (Bayesian Regulation), BP (Back Propagation) and LM (Levenberg-Marquardt) are used to train the neural network. He concluded that the Software Risk Estimation, BR (Bayesian Regulation) performs better than other algorithms.

In their study (Macedo and da Silva), Macedo and da Silva aimed at comparing and clarifying the different activities, inputs and outputs required by each information security risk assessment models and also aimed at analyzing which ones address information security risk effectively. The identified models are the following: OCTAVE, Mehari, MAGERIT, IT-Grundschtz, EBIOS, IRAM, SARA, SPRINT, ISO 27005, NIST SP800-30, CRAMM, MIGRA,MAR, ISAMM, GAO/AIMD-00-33, IT System Security Assessment, MG-2 and MG-3, Dutch A&K Analysis, MARION, Austrian IT Security Handbook, Microsoft's Security Risk Management Guide and Risk IT. The authors passed the selected risk assessment models into two selection iteration before ends up to the last stage. The first iteration is used to exclude some models based on some criteria (guidelines, model cannot identify the Information Security Risks, documentation is expensive or unavailable and if the model is discontinued, obsolete or not recently updated/reviewed). Six models were not excluded and compatible with all the criteria. These models are: OCTAVE, Mehari, MAGERIT, IT-Grundschtz, EBIOS and IRAM. Only these models will be continued to the second stage of selection iteration. The second selection iteration excludes the models based on five criteria. First criteria based on the complexity, effort and preparation this criterion tries to reflect the level of preparation, information, effort and skills needed to implement the model, and the level of detail and scope of the risk analysis results. The second criteria was based on approach of the model (the risk assessment approach each model advocates (e.g. self-assessment, interviews, workshops). Third criteria based on tools (if the model provides supporting tools and how can we obtain them). The fourth and fifth criteria describe the origin (academic, governmental) and

Geographical spread (countries in which the model is known to have been implemented). The final stage was compared between IT-Grundschutz, OCTAVE, IRAM under study were applied in a real organization. The findings of the study shows IRAM is the approach that better conciliates usability, complexity, flexibility and final results. OCTAVE, despite it is simple and quick, just produces the essential information with no great details. On the other hand, we have ITGrundschutz that calculates the IT security level of the organization and provides very detailed technical recommendations, but at a very high cost (time, expertise and resources). Most the models that are mentioned above were much more subjective and unquantifiable which make the risk assessment process complex and could not reflect the accurate risk.

Yu and coauthors has focused, in the study of (Yu, Liang et al. 2010), on the risk profile work sheet of OCTAVE, which is one of the risk assessment models. They proposed a method to give a numerical value to each of business impact along the definite threat path and its probability. The proposed method was based on CVSS which is one of the scoring methods to possible vulnerabilities in information network system. CVSS itself was discussed in this chapter as one of the approaches. This method used the CVSS indices and calculation formulas to give the impact and probability values on OCTAVE's threat work sheet by giving correspondence between each of their indices. This method matches between the CVSS metrics and the linguistic values for items on the threat path. Such as for "Access" in OCTAVE correspondence with "AV" value in CVSS because the OCTAVE does not have "adjacent network". Whereas the "Actor" correspondence with "AC, AU" and "Motive" correspondence with "L", "ML", "MC" and "Outcome" correspondence with "C", "I", "A". After determining the threat path, then go to scoring impact values for each of impact categories. This proposed to correspond each of them to a vector of values ("CDP", "CR", "IR", "AR"). In the next step, the preliminary scores are adjusted by values

of “Exploitability” and “Management Level”. Finally, the collateral damage potential (CDP) value is integrated to obtain the impact value for each of impact category which include “Reputation”, “Financial”, “Productive”, “Fines”, “Safety”, “Others”.

Gallegos and Smith in their study (Gallegos and Smith 2006) proposed red team tactics which is composed of individuals skilled in performing ethical hacking—employing the same tactics malicious hackers may use against information systems, but instead of damaging systems or stealing information, the findings are reported back to the organization. The auditors of the IS can use this tactics in the organizations to assess the risk, but this tactics do not gain a wider acceptance in the organization. This is firstly because the auditors must increase their awareness of tactics used by hackers by training and collaborating with information security professionals. Secondly, the concept of “ethical hacking” is still a hard sell for organizations that might be wary of allowing someone to subvert their security without employees being given advance notice.

Houmb and coauthors presented a model in (Houmb, Nunes Leal Franqueira et al. 2008) that estimates risk level of known vulnerabilities as a combination of frequency and impact estimates derived from the Common Vulnerability Scoring System (CVSS). The model used the base and temporal metrics to estimate frequency and the base and environmental to estimate impact. The model was implemented as a Bayesian Belief Network (BBN). BBN is a directed acyclic graph (DAG) together with an associated set of probability tables. A DAG consists of nodes representing the variables involved and arcs representing the dependencies between these variables. Nodes were defined as stochastic or decision variables and multiple variables may be used to determine the state of a node. There were three types of nodes in a DAG: target node(s) which the objective of the network, intermediate nodes and observable nodes. Each state of a node was expressed using probability density functions. Probability density expresses the confidence in the various outcomes

of the set of variables connected to a node and depends conditionally on the status of the parent nodes at the incoming edges. The directed arcs between the nodes denote the causal relationship between the underlying variables. Evidence or information was entered at the observable nodes and propagated through the network using the causal relationships and a propagation algorithm based on the underlying computational model of BBN.

Dondo has presented in his study (Dondo 2008) a fuzzy system approach for assessing the individual asset by calculating the potential risk exposure for the vulnerabilities associated with these assets. Then the analyzer can rank the vulnerabilities associated with the asset. This model was based on CVSS attributes which defined the CVSS attributes as a Key Risk Indicators (KRIs). The proposed method models the KRIs as a fuzzy variables based on a combination of experience, expertise, or historical input and defined the MF for each variable. Then combine all the identified KRIs into FIS to come up with a final risk value. The FIS determine the fuzzy risk value represented by its CIA components. The combination between the impact and likelihood of the attack will produce the final risk value. Finally, defuzzify the result back into a crisp value and compare the results for each vulnerability in order to rank them.

Houmb and Franqueira have presented in their study (Houmb and Franqueira 2009) a Target of Evaluation (ToE) risk level estimation model that uses CVSS to estimate misuse frequency (MF) and misuse impact (MI), and from these derive the risk level of ToE. This is a general risk in which this model works on the level of vulnerabilities and is able to compose the vulnerabilities into service levels. The service levels define the potential risk levels and are modelled as a Markov process, which are then used to predict the risk level at a particular time. MF is estimated from attributes in the base and temporal metrics of CVSS and MI is estimated from attributes in the base

and environmental metrics of CVSS. The base metrics of CVSS is used to establish the initial estimates of both MF and MI. MF is then made attack specific by adding in factors concerning the attack tools available, the existing security measures and the report confidence. For MI, the initial MI of a potential vulnerability exploit (attack) derived from the base metrics is made ToE specific by taking the relevant security requirements into consideration. An important factor to note for MI is that there are no impacts of a potential vulnerability exploit (attack) if there are no relevant requirements.

Mell and coauthors have presented in their study (Mell, Scarfone et al. 2006; Mell, Scarfone et al. 2007) a CVSS which is an open framework that prioritize the vulnerabilities and remediate those that pose the greatest risk. This system assesses the vulnerabilities across many disparate hardware and software platform. CVSS is composed of three metric groups: Base, Temporal, and Environmental, each consisting of a set of metrics. The base metric is used to represent the intrinsic and fundamental characteristics of a vulnerability that are constant over time and user environments, whereas the temporal metrics represents the characteristics of a vulnerability that change over time but not among user environments and the Environmental metrics represents the characteristics of a vulnerability that are relevant and unique to a particular user's environment. The main aim of this approach is to defined the fundamental characteristics of the vulnerabilities and then the user can invoke the temporal and environmental metrics to provide more accurately risk to their critical asset. There are a number of other vulnerability "scoring" systems managed by both commercial and noncommercial organizations.

1. Security Bulletin Severity Rating System: which is intended to help the customers of Microsoft to decide which updates they should apply under their particular circumstances, and how rapidly they need to take action (Microsoft).
2. The SANS vulnerability analysis scale considers whether the weakness is found in default configurations or client or server systems.
3. CERT/CC produces a numeric score ranging from 0 to 180 but considers such factors as whether the Internet infrastructure is at risk and what sort of preconditions are required to exploit the vulnerability.
4. A New CVSS-Based Tool to Mitigate the Effects of Software Vulnerabilities (Ali, Zavorsky et al. 2011) .

In their study (Okereke and Osuagwu 2012), Okereke and Osuagwu have examined security metrics available to information systems and have proposed a metric model for web page vulnerability measurement and ranking. This proposed a model called web application security evaluator (WASE) which used a WASE software which crawls through the sites to extract the security vulnerability parameters and assign a specific value to each parameter.

Adebiyi and coauthors have proposed in their study (Adebiyi, Arreymbi et al. 2013) a new approach for assessing security during the design phase than implementation or testing phase by three-layered feed-forward back-propagation of neural network (NN). This method used the neural network approach in analyzing software design for security flaws which is based on the abstract and match technique through which software flaws in a software design can be identified when an attack pattern is matched to the design. This method collecting data from online vulnerability databases and identified the attack attributes which were used to abstract the data capturing the

attack scenario for training the neural network. The attack attributes identified in the paper which include 12 inputs. The data collected are converted in ASCII comma delimited format and then used in training the neural network. For the expected output from the neural network, the data used in training network is derived from the attack pattern which has been identified in each of the attack scenarios.

Shameli and Shajari have presented in their study (Shameli-Sendi, Shajari et al. 2012) a practical model for information security risk assessment. This model is based on multi-criteria decision-making and uses fuzzy logic. The proposed risk assessment is a qualitative approach according to ISO/IEC 27005 standard. In the proposed model, a fuzzy technique was used to connect expert opinion with linguistic variables. These linguistic variables reflect the expert opinions. In this model determined the likelihood and impact of each threat, effective criteria for their measurement have been considered.

Finally, Sendi and coauthors have presented in their study (Sendi, Jabbarifar et al. 2010) the FEMRA model which uses the fuzzy expert systems to assess the risk in organizations. The risk assessment varies considerably with the context, the metrics used as dependent variables, and the opinions of the persons involved. This model represents each risk with numerical values. The authors presented three steps to achieve the goal. The first step to identify the assets which uses a security cube to identify and classify the assets. Then list all potential threats applicable to these assets. The second step is to generate a list of asset vulnerabilities and risks. The final step is to calculate the effect risks which sing the fuzzy models. In this model the values for each asset is taken from three experts in terms of CIA triad and then calculate the average.

The above mentioned related work is summarized in *Table (2.1)*.

Table (2.1) Summary of related works

Reference	Methodology	Quantifiable	CVSS	Result
Cho 2015	Data Mining	No	No	used to inspect the website to see whether it has the same vulnerability
Al-Mahrouqi, Tobin et al. 2015	ECP	No	No	Use Prediction Investigation Approach to predict and trace the source of the attack or illegal activities in the computer network
Prasad 2014	Genetic Algorithm	No	No	The real time chromosomes are checked with the rule sets. If the particular chromosomes matches with the rules provided in the rule set, it takes the decision of whether allow or block depending on which rule set it matches.
Lee 2014	AHP	Yes	No	Represent a quantitative and qualitative risk assessment by using different methods and recommended using a hybrid method.
Macedo and da Silva	Guidelines	No	No	After comparing between different security guidelines, ITGrundschutz guideline can calculate the IT security level of the organization and provides very detailed technical recommendations, but at a very high cost (time, expertise and resources).

Reference	Methodology	Quantifiable	CVSS	Result
Cho 2015	Data Mining	No	No	used to inspect the website to see whether it has the same vulnerability
Al-Mahrouqi, Tobin et al. 2015	ECP	No	No	Use Prediction Investigation Approach to predict and trace the source of the attack or illegal activities in the computer network
Prasad 2014	Genetic Algorithm	No	No	The real time chromosomes are checked with the rule sets. If the particular chromosomes matches with the rules provided in the rule set, it takes the decision of whether allow or block depending on which rule set it matches.
Yu, Liang et al. 2010	Threat path on profile sheet of OCTAVE	Yes	CVSS formulae	Preliminary risk impact score on work categorization (Reputation, Financial, Productive, Fines, and Safety) by using CVSS indices and equations.
Gallegos and Smith 2006	Ethical hacking	No	No	Performing ethical hacking—employing the same tactics malicious hackers may use against information systems, but instead of damaging systems or stealing information, the findings are reported back to the organization.
Houmb, Nunes Leal Franqueira et al. 2008	BBN	Yes	Yes	Estimating risk level of known vulnerability by combining the impact and frequency of risks by using the CVSS metrics.

Reference	Methodology	Quantifiable	CVSS	Result
Cho 2015	Data Mining	No	No	used to inspect the website to see whether it has the same vulnerability
Al-Mahrouqi, Tobin et al. 2015	ECP	No	No	Use Prediction Investigation Approach to predict and trace the source of the attack or illegal activities in the computer network
Prasad 2014	Genetic Algorithm	No	No	The real time chromosomes are checked with the rule sets. If the particular chromosomes matches with the rules provided in the rule set, it takes the decision of whether allow or block depending on which rule set it matches.
Dondo 2008	Fuzzy	Yes	Yes	Calculating risk level of vulnerability by combining the CVSS metrics as a KRI's as input variable to a fuzzy system.
Houmb and Franqueira 2009	Markov process	Yes	Yes	Calculating the impact and frequency of vulnerabilities to produce the risk based on service level.
Adebiyi, Arreymbi et al. 2013	NN	No	Yes	This paper produce an attack pattern that used to check the software design flaws.
Shameli-Sendi, Shajari et al. 2012	Fuzzy	Yes	No	Risk level of the threats that associate with vulnerability by an expert decisions.
Sendi, Jabbarifar et al. 2010	Fuzzy	Yes	No	Risk level of the threats that associate with vulnerability by an expert decisions.

Reference	Methodology	Quantifiable	CVSS	Result
Cho 2015	Data Mining	No	No	used to inspect the website to see whether it has the same vulnerability
Al-Mahrouqi, Tobin et al. 2015	ECP	No	No	Use Prediction Investigation Approach to predict and trace the source of the attack or illegal activities in the computer network
Prasad 2014	Genetic Algorithm	No	No	The real time chromosomes are checked with the rule sets. If the particular chromosomes matches with the rules provided in the rule set, it takes the decision of whether allow or block depending on which rule set it matches.
Rani 2013	Neuro-Fuzzy	Yes	No	Estimate the software risk in all stages of software development life cycle (SDLC).

2.2 Related Work Analysis

Standards and guidelines provide tools for evaluating the security controls of systems. Examples of this tools were shown in (Macedo and da Silva) but most evaluations were qualitative and subjective activity biased by the evaluator (even though they follow a standard) and the other quantitative evaluations have provided very detailed technical recommendations and very high cost (time, expertise and resources) such as ITGrundschutz. (Shameli-Sendi, Shajari et al. 2012) and (Sendi, Jabbarifar et al. 2010) presented a Fuzzy system to calculate the risk based on experts opinion but the lack of quantitative data and the rapidly changing security environment makes it hard to derive accurate measures over such a long time-period and the risk value is expert specific. Yu, Liang et al. (2010) discussed an approach to measure security investment benefits for off the shelf software systems using CVSS. The authors proposed a threat path using OCTAVE profile sheet that focused on impact values such as productivity, reputation and privacy of the systems where the vulnerabilities are located. Our opinion is that, it is not easy to calculate the impact on productivity, reputation and privacy. Also it is better to use the environmental metrics as given in the CVSS, as it is easier to evaluate confidentiality, integrity and availability.

In Dondo (2008), an approach to vulnerability prioritization using fuzzy risk analysis was presented. Here, the construct asset value (AV) was used to derive the risk level or risks to a system. The asset value (AV) is assumed given. The approach derives risk level based on the CVSS base metrics variables, a measure of time from when the vulnerability was reported and the safeguards already in the system. The author applied fuzzy rules to compute impact (I) and likelihood (L) and derive risk level as: $AV \times I \times L$. This approach is similar to our model, but our model does not use fuzzy rules. Our model uses the temporal and environmental metric groups given in the CVSS to estimate the risk level rather than asset value and safeguard. Asset value is not always easy to evaluate and might be stakeholder specific. AV is not a generalizable variable,

but rather context and stakeholder specific. Our models is based on CVSS, which is an open standard that also reveals the details behind the scores provided. Furthermore, CVSS is regularly updated and several information sources is taken into consideration when calculating the CVSS score.

Chapter Three : Common Vulnerability Scoring System (CVSS) V2

3.1 Introduction

The Common Vulnerability Scoring System (CVSS) is a vulnerability scoring system which was created by NIAC (National Infrastructure Advisory Council) in 2004 and currently is maintained by the Forum of Incident Response and Security Teams (FIRST). This system is an effort of many companies involved including several vendors, vulnerability tools and bulletins such as hardware and software development companies like IBM, HP, Cisco Systems, Symantec, Microsoft, Internet Security Systems, vulnerability tools like Qualys and Nessus. CVSS provides a standard for communicating the characteristics and impacts of IT vulnerabilities. The CVSS score has a numeric value ranges from 0 to 10. The overall result of CVSS score resulted from three groups: Base, Temporal and Environmental group. Each group consists of metrics that are represented in terms of textual representation that reflect the values used to derive the score for each group. CVSS has more than one version, CVSS v1 and CVSS v2. This thesis adopt the CVSS version 2 (the latest version). *Figure (3.1)* shows the groups of CVSS v2 and metrics in each group (Mell, Scarfone et al. 2006).

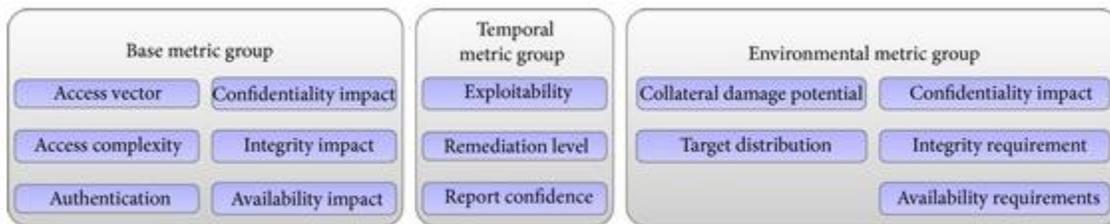


Figure (3.1) CVSS metric groups
(source (Mell, Scarfone et al.))

To validate scoring the CVSS vulnerability, the metric values of base group must be assigned. Thus, the base group is mandatory and must start with it to calculate the CVSS score whereas the two other groups (Temporal and environmental) are optional. If the temporal and environmental metrics are not assigned, the overall score is equal to the base score. Whereas, if the temporal metrics are assigned the base score will be combined with the temporal metrics to produce the temporal score which is equal to the overall CVSS score. Similarly, if an environmental score is needed, the environmental metrics are combined with the temporal score to produce the environmental score ranging from 0 to 10 which is equal to the overall CVSS score. *Figure (3.2)* describes the calculation of CVSS score by the three group metrics.

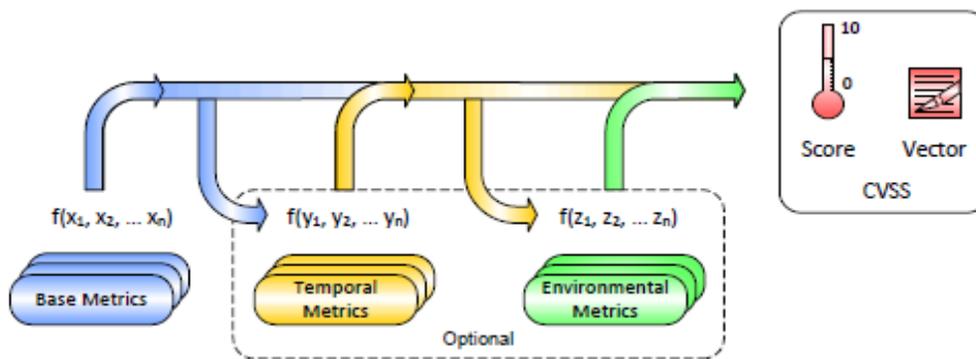


Figure (3.2) CVSS calculation process
(source (Mell, Scarfone et al.))

This chapter describes the groups and metric terms in details. Let's start with the three groups: Base, Temporal and Environmental group.

3.2 Base Group

The base group consists of the base metrics which reflects the characteristics of the vulnerability to produce the base score. The base score is combined from two sub score, exploitability sub scores

and impact sub score. The exploitability sub score is composed of the Access Vector (B_AV), Access Complexity (B_AC), and Authentication (B_AU) metrics. Those metrics of exploitability sub score measure how the vulnerability is accessed and whether or not extra conditions are required to exploit it. The second sub score of base score is the impact sub score which measure how a vulnerability, if exploited, will directly affect an IT asset, where the impacts are independently defined as the degree of loss of confidentiality (B_C), integrity (B_I), and availability (B_A). The next section describes the metrics of base group in more details.

3.2.1 Access Vector

The access vector metric measure how the vulnerability is exploited. This metrics have possible values which represented in string terms ranging from local, adjacent network and network (or remote). *Table (3.1)* lists the access vector metric values.

Table (3.1) Access Vector Scoring

Value	Description	Score
Local (L)	The attacker must either have physical access to the vulnerable system (e.g. firewire attacks) or a local account (e.g. a privilege escalation attack).	0.395
Adjacent Network (A)	The attacker must have access to the broadcast or collision domain of the vulnerable system (e.g. ARP spoofing, Bluetooth attacks).	0.646
Network (N)	The vulnerable interface is working at layer 3 or above of the OSI Network stack. These types of vulnerabilities are often described as remotely exploitable (e.g. a remote buffer overflow in a network service)	1.0

The metric values in *Table (3.1)* represented as linguistic terms and numeric values

3.2.2 Access Complexity

This metric measures the complexity of the attacks required to exploit the vulnerability of the system. Some attackers can exploit the system when gained some privilege such as buffer overflow and does not need additional step. Whereas, other attacks need additional step to exploit such as exploit the email which require the user to download and open a tainted attachment the possible values for this metrics are listed in *Table (3.2)*.

Table (3.2) Access Complexity Scoring

Value	Description	Score
High (H)	Specialised conditions exist, such as a race condition with a narrow window, or a requirement for social engineering methods that would be readily noticed by knowledgeable people.	0.35
Medium (M)	There are some additional requirements for access, such as a limit on the origin of the attacks, or a requirement for the vulnerable system to be running with an uncommon, non-default configuration.	0.61
Low (L)	There are no special conditions for access to the vulnerability, such as when the system is available to large numbers of users, or the vulnerable configuration is ubiquitous.	0.71

3.2.3 Authentication

This metric measures the number of times an attacker must authenticate to a target in order to exploit a vulnerability. This metric measure how many times the attacker authenticate before exploit the system and measure the gauge of authentication or complexity of authentication. The metric values of authentication listed in *Table (3.3)*.

Table (3.3) Authentication Scoring

Value	Description	Score
Multiple (M)	Exploitation of the vulnerability requires that the attacker authenticate two or more times, even if the same credentials are used each time.	0.45
Single (S)	The attacker must authenticate once in order to exploit the vulnerability.	0.56
None (N)	There is no requirement for the attacker to authenticate.	0.704

The combination of three above mentioned metrics in section (3.2.1, 3.2.2 and 3.2.3) will produce the exploitability score that range from 0 to 10. Now in the next three section will discover the three impact metrics which produce the impact score.

3.2.4 Confidentiality Impact

Confidentiality refers to limiting information access and disclosure to only authorized users. The confidentiality impact measures the impact on confidentiality of a successfully exploited vulnerability. The possible values for this metric listed in Table (3.4).

Table (3.4) Confidentiality Impact Scoring

Value	Description	Score
None (N)	There is no impact on the confidentiality of the system.	0.0
Partial (P)	There is considerable disclosure of information, but the scope of the loss is constrained such that not all of the data is available.	0.275

Value	Description	Score
Complete (C)	There is total information disclosure, providing access to any / all data on the system.	0.660

3.2.5 Integrity Impact

This metric measures the impact to integrity of a successfully exploited vulnerability. Integrity refers to the trustworthiness and guaranteed veracity of information. The possible values for this metric are listed in *Table (3.5)*. Increased integrity impact increases the vulnerability score.

Table (3.5) Integrity Impact Scoring

Value	Description	Score
None (N)	There is no impact on the integrity of the system.	0.0
Partial (P)	Modification of some data or system files is possible, but the scope of the modification is limited.	0.275
Complete (C)	There is total loss of integrity; the attacker can modify any files or information on the target system.	0.660

2.2.6 Availability Impact

This metric measures the impact to availability of a successfully exploited vulnerability. Availability refers to the accessibility of information resources. Attacks that consume network bandwidth, processor cycles, or disk space all impact the availability of a system. The possible values for this metric are listed in

Table (3.6). Increased availability impact increases the vulnerability score.

Table (3.6) Availability Impact Scoring

Value	Description	Score
None (N)	There is no impact on the availability of the system.	0.0
Partial (P)	There is reduced performance or loss of some functionality.	0.275
Complete (C)	There is total loss of availability of the attacked resource.	0.660

3.3 Temporal Metrics

The threat posed by a vulnerability may change over time. Three such factors that CVSS captures are: confirmation of the technical details of a vulnerability, the remediation status of the vulnerability, and the availability of exploit code or techniques. Since temporal metrics are optional they each include a metric value that has no effect on the score (Not defined).

3.3.1 Exploitability

This metric measures the current state of exploit techniques or code availability. Whereas the Public availability of exploit code will increase the number of potential attackers and can be easy-to-use those who are unskilled. Increasing the exploitability code will increase the severity of the attack. The possible value of the exploit tools or codes ranges from easy to use and availability by unskilled, can be executed by a skilled hackers or this vulnerability can be exploited theoretically.

The possible value listed in Table (3.7). The more easily a vulnerability can be exploited, the higher the vulnerability score.

Table (3.7) Exploitability Scoring Evaluation

Value	Description	Score
Unproven (U)	No exploit code is available, or the exploit is theoretical	0.85
Proof-of-concept (P)	Proof-of-concept exploit code or demonstration attacks are available, but not practical for widespread use. Not functional against all instances of the vulnerability.	0.9
Functional (F)	Functional exploit code is available, and works in most situations where the vulnerability is present.	0.95
High (H)	The vulnerability can be exploited by automated code, including mobile code (such as a worm or virus).	1.0
Not Defined (ND)	This is a signal to ignore this score.	1.0

3.3.2 Remediation Level

The remediation level of a vulnerability is an important factor for prioritization. The typical vulnerability is unpatched when initially published. Workarounds or hotfixes may offer interim remediation until an official patch or upgrade is issued. Each of these respective stages adjusts the temporal score downwards, reflecting the decreasing urgency as remediation becomes final.

The possible values for this metric are listed in Table (3.8). The less official and permanent a fix, the higher the vulnerability score is.

Table (3.8) Remediation Level Scoring Evaluation

Value	Description	Score
Official Fix (O)	A complete vendor solution is available - either a patch or an upgrade.	0.87
Temporary Fix (T)	There is an official but temporary fix / mitigation available from the vendor.	0.90
Workaround (W)	There is an unofficial, non-vendor solution or mitigation available - perhaps developed or suggested by users of the affected product or another third party.	0.95
Unavailable (U)	There is no solution available, or it is impossible to apply a suggested solution. This is the usual initial state of the remediation level when a vulnerability is identified.	1.0
Not Defined (ND)	This is a signal to ignore this score.	1.0

3.3.3 Report Confidence

This metric measures the degree of confidence in the existence of the vulnerability and the credibility of the known technical details. Sometimes, only the existence of vulnerabilities are publicized, but without specific details. The vulnerability may later be corroborated and then confirmed through acknowledgement by the author or vendor of the affected technology. The urgency of a vulnerability is higher when a vulnerability is known to exist with certainty. This

metric also suggests the level of technical knowledge available to would-be attackers. The possible values for this metric are listed in *Table (3.9)*. The more a vulnerability is validated by the vendor or other reputable sources, the higher the score.

Table (3.9) Report Confidence Scoring Evaluation

Value	Description	Score
Unconfirmed (UC)	A single unconfirmed source, or multiple conflicting sources. Rumored vulnerability.	0.9
Uncorroborated (UR)	Multiple sources that broadly agree - there may be a level of remaining uncertainty about the vulnerability	0.95
Confirmed (C)	Acknowledged and confirmed by the vendor or manufacturer of the affected product.	1.0
Not Defined (ND)	This is a signal to ignore this score.	1.0

3.4 Environmental Metrics

The CVSS environmental metric group captures the characteristics of a vulnerability that are associated with a user's IT environment. Since environmental metrics are optional they each include a metric value that has no effect on the score. This value is used when the user feels the particular metric does not apply and wishes to "skip over" it.

3.4.1 Collateral Damage Potential

This metric measures the potential for loss of life or physical assets through damage or theft of property or equipment. The metric may also measure economic loss of productivity or revenue.

The possible values for this metric are listed in *Table (3.10)*. Naturally, the greater the damage potential, the higher the vulnerability score.

Table (3.10) Collateral Damage Potential Scoring Evaluation

Value	Description	Score
None (N)	No potential for loss of property, revenue or productivity	0
Low (L)	Slight damage to assets, or minor loss of revenue or productivity	0.1
Low-Medium (LM)	Moderate damage or loss	0.3
Medium-High (MH)	Significant damage or loss	0.4
High (H)	Catastrophic damage or loss	0.5
Not Defined (ND)	This is a signal to ignore this score.	0

Clearly, each organization must determine for themselves the precise meaning of "slight, moderate, significant, and catastrophic."

3.4.2 Target Distribution

This metric measures the proportion of vulnerable systems. It is meant as an environment-specific indicator in order to approximate the percentage of systems that could be affected by the vulnerability. The possible values for this metric are listed in

Table (3.11). The greater the proportion of vulnerable systems, the higher the score.

Table (3.11) Target Distribution Scoring Evaluation

Value	Description	Score
None (N)	No target systems exist, or they only exist in laboratory settings	0
Low (L)	1%-25% of systems at risk	0.25
Medium (M)	26%-75% of systems at risk	0.75
High (H)	76%-100% of systems at risk	1.0
Not Defined (ND)	This is a signal to ignore this score.	1.0

3.4.3 Security Requirements

These metrics enable the analyst to customize the CVSS score depending on the importance of the affected IT asset to a user’s organization, measured in terms of confidentiality, integrity, and availability. That is, if an IT asset supports a business function for which availability is most important, the analyst can assign a greater value to availability, relative to confidentiality and integrity. Each security requirement has three possible values: low, medium, or high.

The possible values for the security requirements are listed in Table (3.12). For brevity, the same table is used for all three metrics. The greater the security requirement, the higher the score

Table (3.12) Security Requirements Scoring Evaluation

Value	Description	Score
Low (L)	Loss of (confidentiality / integrity / availability) is likely to have only a limited effect on the organization.	0.5
Medium (M)	Loss of (confidentiality / integrity / availability) is likely to have a serious effect on the organization.	1.0
High (H)	Loss of (confidentiality / integrity / availability) is likely to have a catastrophic effect on the organization.	1.51
Not Defined (ND)	This is a signal to ignore this score.	1.0

Next chapter will introduce the thesis approach that uses the CVSS metrics with some rearrangement to be used as IA metrics.

3.5 Summary

CVSS is an open standard for communicating the characteristics and impacts of IT vulnerabilities. The CVSS score has a numeric value ranges from 0 to 10. The overall result of CVSS score resulted from three groups: Base, Temporal and Environmental group. Each group consists of metrics that are represented in terms of textual representation that reflect the values used to derive the score for each group. It is platform and technology independent; in practice. There are a lot of vulnerabilities affecting a very wide range of software products: operating systems, web and legacy applications, security products (firewalls, antivirus software, etc.), databases, etc.

An overall CVSS Score is actually composed of three sub-scores (the "Metric Groups"): the Base Score, the Temporal Score, and the Environmental Score.

The Base Score reflects "the intrinsic and fundamental characteristics of a vulnerability that are constant over time and user environments."

The Temporal Score reflects "the characteristics of a vulnerability that change over time but not among user environments."

The Environmental Score reflects "the characteristics of a vulnerability that are relevant and unique to a particular user's environment."

Chapter Four : Data Profiles

4.1 Introduction

As mentioned in chapter one, the major contribution is to develop a reliable IA risk level using a neurofuzzy based modeling techniques. Therefore, developing any supervised-based soft computing model needs pairs of data (inputs and outputs), and in order to have a reliable model, we need reasonable actual sets of data composed of vulnerabilities as an output for a specific systems. In this thesis we collect the vulnerabilities related to webmail system and its services (Perl , Apache) and Internet Explorer (IE). The selected variables are the same as we used the CVSS metrics. . This chapter illustrates the data profile and preprocessing which is summarized in *Figure (4.1)* and described in detailed, in this chapter.

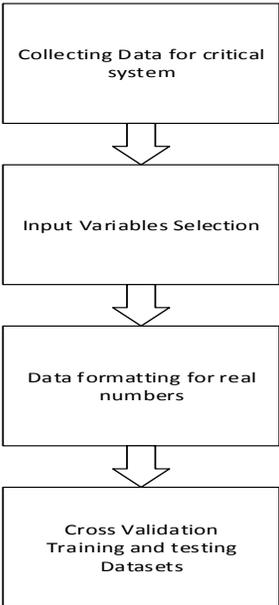


Figure (4.1) Data Preprocessing Procedure Stages

4.2 Data Collection

The first step for assessment of the IA risk level for a critical IS (system failure can have severe human or economic consequence) (Rushby 1994) is to collect data about all vulnerabilities associated with it. NVD is a huge database of vulnerabilities for most vendors, software, applications and services. It is an aggregation from different databases. **Table (4.1)** illustrates and summarizes the NVD databases contents. Therefore, this thesis adopt the NVD database as the source of the vulnerability dataset for critical IS. For the complexity of obtainment the database of critical systems, therefore this thesis assumes an email asset and make an assessment for all vulnerabilities of this asset after collecting data. This asset is a client-server program and consist of different service which all of them can affect the capability of asset and impact the assurance of the organization as shown in **Figure (4.2)**.

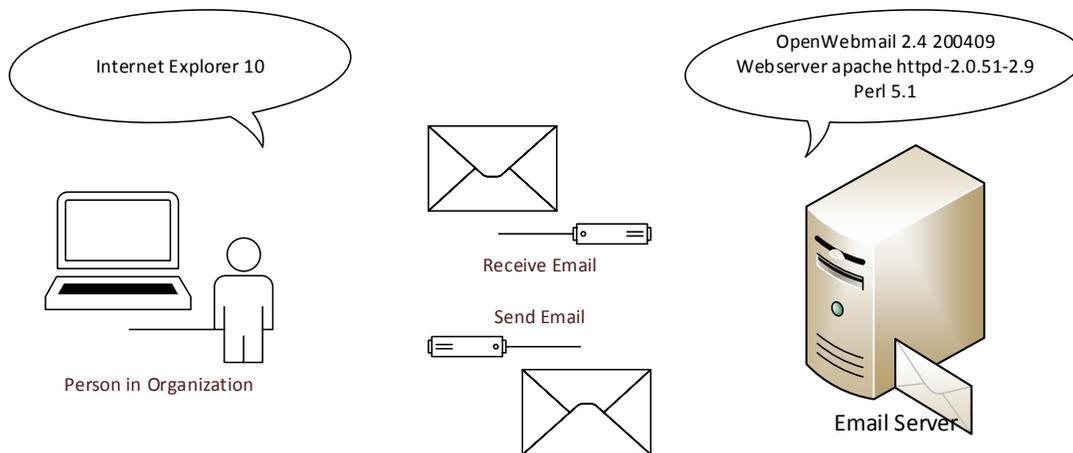


Figure (4.2) Openwebmail Email Asset

Table (4.1) Summary of the NVD Database Contents (NVD 2004)

Database	Full name	No. records	Notes
CVE	Common Vulnerabilities and Exposures	69164	Maintained by MITRE corporation which founds the vulnerability in services such as (cross site scripting, buffer overflow, denial of services) in software products such as internet browsers product (IE, chrome, Mozilla), multimedia software (Quick time), emails software (openWebmail)
NCP	National Checklist Program	285	Defined by the NIST SP 800-70 Rev. 2 that provide detailed low level guidance on setting the security configuration of operating systems and applications
US-CERT Alert	United States Computer Emergency Readiness Team Alert	249	Alerts provide timely information about current security issues, vulnerabilities, and exploits

Database	Full name	No. records	Notes
US-CERT Vulnerability Notes	Vulnerability Notes Database	4335	The Vulnerability Notes Database provides timely information about software vulnerabilities. Vulnerability notes include summaries, technical details, remediation information, and lists of affected vendors.
OVAL Queries	Open Vulnerability and Assessment Language	10286	Determining vulnerability and configurations issues in computer system
CPE Names	Common Platform Enumeration	102000	describing and identifying classes of applications, operating systems, and hardware devices present among an enterprise's computing assets

Figure (4.3) shows a sample the vulnerability attributes. To assess the vulnerability of the email asset, we need to collect the vulnerabilities for all its services. Therefore, we need to collect the vulnerability for the client side (IE) and the server side (Openwebmail, Apache, Perl). *Table (4.2)* illustrates the number of vulnerabilities for each service and the total vulnerabilities for email asset is 1129 vulnerabilities.

	B	C	D	F	G	H	I	J	K	L	M	N	O
1	#	CVE ID	CWE ID	Vulnerability Type(s)	Publish Date	Update Date	Score	Gained Access	Access	Complexity	Authentication	Conf.	Integ.
2	1	CVE-2015-0072	79	XSS Bypass	2/7/2015	2/13/2015	4.3	None	Remote	Medium	Not required	None	Partial
3	2	CVE-2015-0071	264	Bypass	2/10/2015	2/18/2015	4.3	None	Remote	Medium	Not required	Partial	None
4	3	CVE-2015-0070	200	+Info	2/10/2015	2/18/2015	4.3	None	Remote	Medium	Not required	Partial	None
5	4	CVE-2015-0069	264	Bypass	2/10/2015	2/18/2015	4.3	None	Remote	Medium	Not required	Partial	None
6	5	CVE-2015-0068	399	DoS Exec Code Mem. Corr.	2/10/2015	2/18/2015	9.3	None	Remote	Medium	Not required	Complete	Comple
7	6	CVE-2015-0067	399	DoS Exec Code Mem. Corr.	2/10/2015	2/18/2015	9.3	None	Remote	Medium	Not required	Complete	Comple
8	7	CVE-2015-0066	399	DoS Exec Code Mem. Corr.	2/10/2015	2/18/2015	9.3	None	Remote	Medium	Not required	Complete	Comple
9	8	CVE-2015-0055	264	+Priv	2/10/2015	2/18/2015	4.3	None	Remote	Medium	Not required	None	Partial

Figure (4.3) Some Vulnerabilities for Email Asset

Table (4.2) Vulnerabilities for All Services of Email Asset

Type	Service	Number of Vulnerability
Client side	IE	505
Server Side	OpenWebmail	13
	Apache	587
	Perl	24
Total		1129

4.3 Variable Selection

As mentioned earlier in chapter one in methodology section, the risk assessment is predicted from frequency and impact of vulnerability exploited by attacks as briefly shown in Figure (1.3). In this section, we will describe in more details the input-output variables for each stage. In the first stage,

the initial frequency are predicted from the attributes of base group of CVSS (AV, AC, AU) and the initial impact from base group (C, I, A). The output from the first stage will be combined with inputs of the second stage as shown in *Figure (4.4)*. Changing factors will update the initial values in the second stage. If there is no change the initial values will go to the third stage.

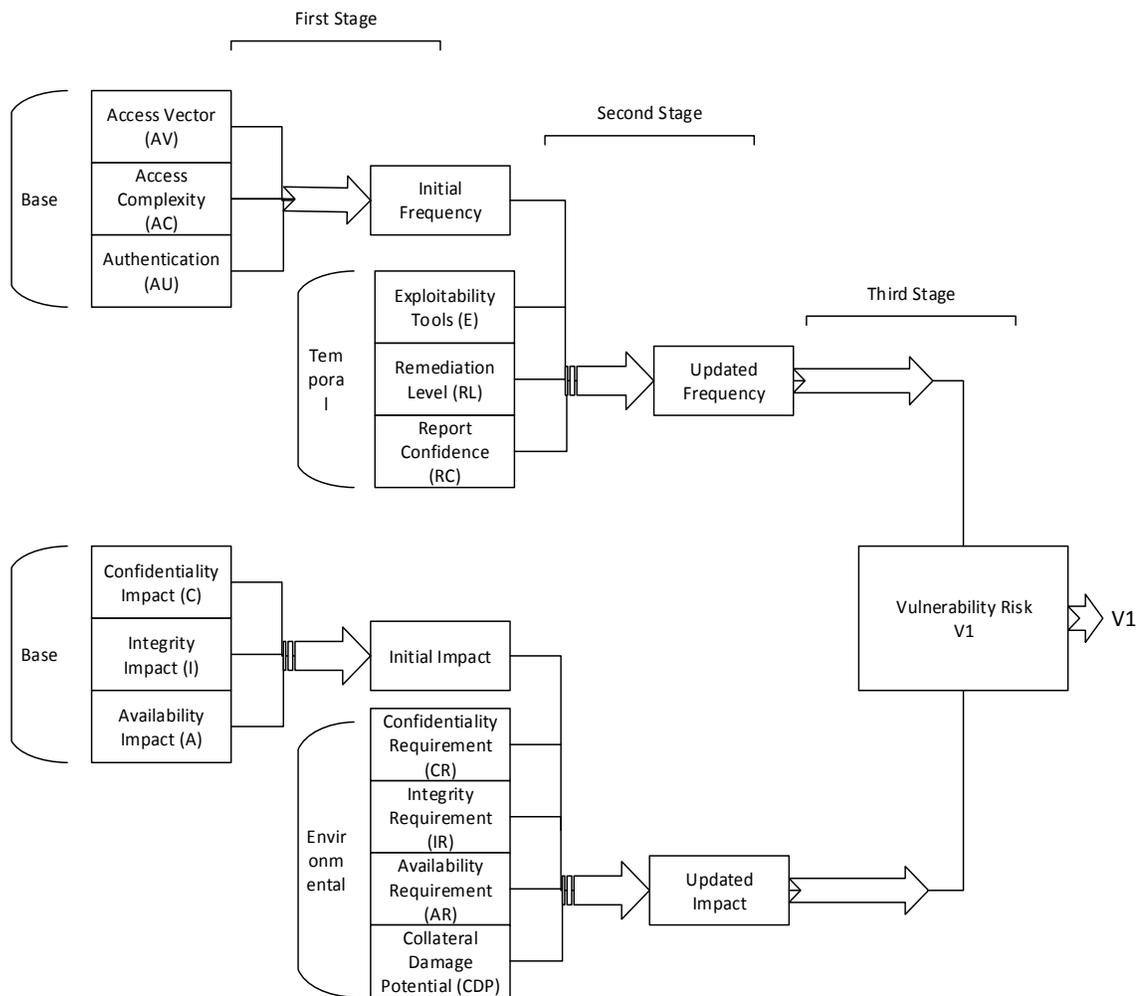


Figure (4.4) Input-Output Variables for each Stage

In Figure (4.4) shows the risk level for one vulnerability. Therefore to predict the total vulnerability risk for one service, all predicted risk are aggregated to predict the total risk for one service as shown in Figure (4.5). In final stage, will aggregated all totals vulnerability for each service to produce the IA risk assessment for an asset. In our case, the (Email asset), will combine the total risk of IE, openwebmail, apache and perl as shown in Figure (4.6).

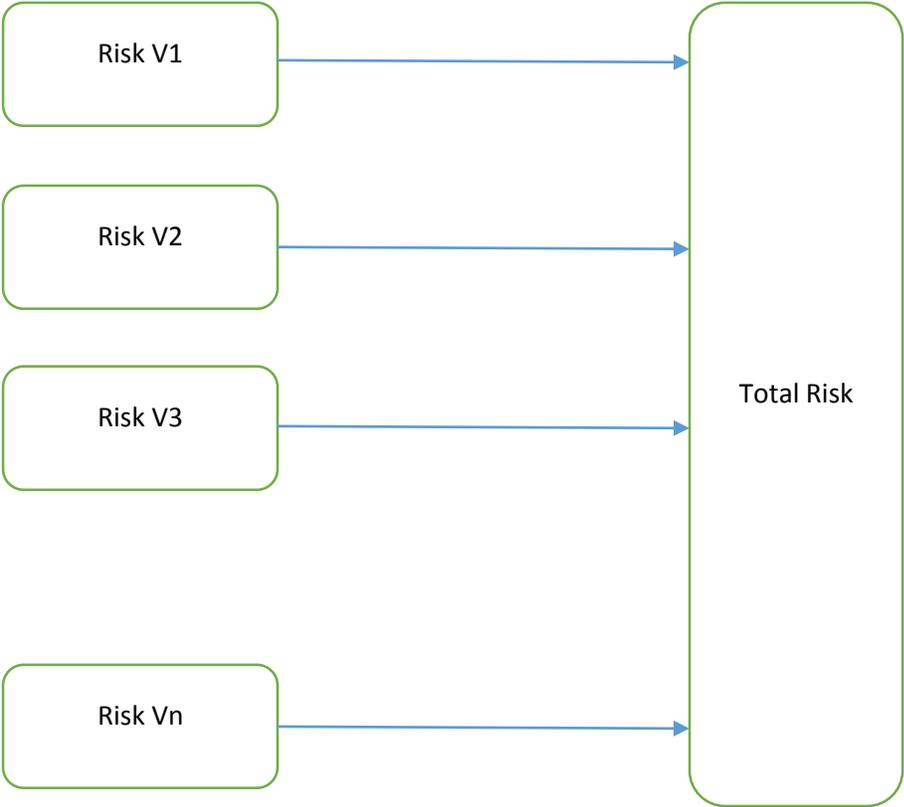


Figure (4.5) Aggregated Vulnerability Risk for One Service

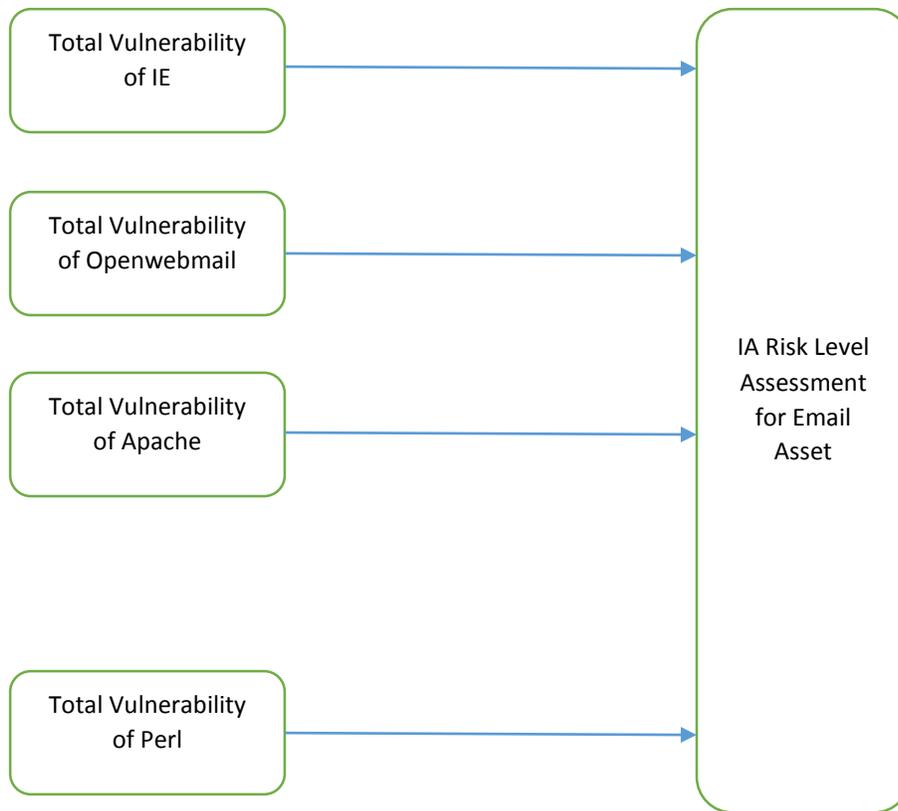


Figure (4.6) IA Risk Level Assessment for Email Asset

4.4 Data Formatting

To build the models, the inputs and outputs for the model should be in real numbers format. The data were collected for vulnerabilities, it was in linguistic terms and will be converted into real numbers. *Figure (4.7)* and *Figure (4.8)* illustrate the numeric values for data formatting.

CVSS metric group	CVSS attribute	Rating	Rating value
Base metric	Access vector (B_AV)	Local (L) adjacent	0.395
		Network (A)	0.646
	Access complexity (B_AC)	Network (N)	1.0
		High (H)	0.35
	Authentication instances (B_Au)	Medium (M)	0.61
		Low (L)	0.71
Multiple (M)		0.45	
Single (S)		0.56	
Temporal metric	Exploitability tools & techniques (T_E)	None (N)	0.704
		Unproved (U)	0.85
		Proof-of-concept (POC)	0.9
	Remediation level (T_RL)	Functional (F)	0.95
		High (H)	1.0
		Official fix (OF)	0.87
		Temporary fix (TF)	0.90
	Report confidence (T_RC)	Workaround (W)	0.95
		Unavailable (U)	1.0
		Unconfirmed (UC)	0.90
Uncorroborative (UR) confirmed (C)		0.95	

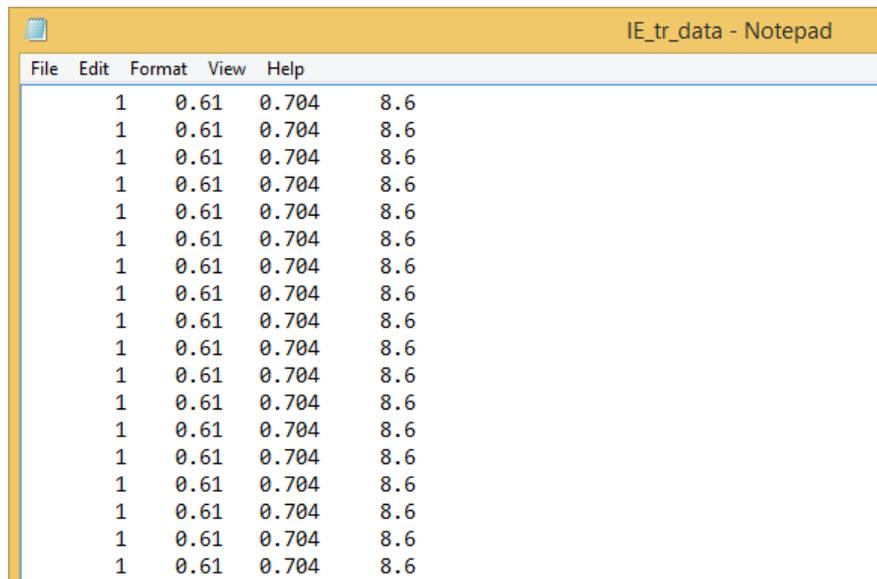
Figure (4.7) Frequency Estimation Rate and Numeric Values

CVSS metric group	CVSS attribute	Rating	Rating value
Base metric	Confidentiality impact (B_C)	None (N)	0.0
		Partial (P)	0.275
		Complete (C)	0.660
	Integrity impact (B_I)	None (N)	0.0
		Partial (P)	0.275
	Availability impact (B_A)	Complete (C)	0.660
None (N)		0.0	
Environmental metric	Confidentiality requirement (E_CR)	Partial (P)	0.275
		Complete (C)	0.660
		None (N)	0.0
	Integrity requirement (E_IR)	Low (L)	0.5
		Medium (M)	1.0
		High (H)	1.51
	Availability requirement (E_AR)	Low (L)	0.5
		Medium (M)	1.0
		High (H)	1.51
	Collateral damage potential (E_CDP)	Low (L)	0.1
		Low medium (LM)	0.3
		Medium high (MH)	0.4
	High (H)	0.5	

Figure (4.8) Impact Estimation Rate and Numeric Values

4.5 Cross Validation

The basic idea of using a cross validation algorithm is to avoid the over fitting problem (Error on the dataset used to fit the model can be misleading and perform poorly in predicting out-of-sample cases) (Hawkins 2004) and to construct from the available dataset two datasets, training (Tr) and testing (Ts) datasets. The cross validation algorithm that was used by initiating two matrices, the first one is used to store the training datasets and the second one is to store the testing datasets. This works by scanning the available datasets and selects recursively three elements for training and moving them to Tr datasets matrix, and then moving the fourth one to the Ts datasets matrix. The algorithm repeats the process until reach to the end of the available file. The available file that containing the training and testing dataset must be in DAT format as shown in *Figure (4.9)*.



File	Edit	Format	View	Help
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	
1	0.61	0.704	8.6	

Figure (4.9) Sample of Training Dataset File

Table (4.3) shows the summary of available datasets which have been used in developing the models and containing the training and testing the datasets.

Table (4.3) Available Datasets with Tr and Ts Datasets

Service	Total Number of Datasets	No of Training Datasets (Tr)	No of Testing Datasets (Ts)
IE	505	379	126
Openwebmail	13	10	3
Apache	587	441	146
Perl	24	18	6

4.6 Summary

This chapter discuss the data profile and preprocessing approach before developing the soft computing model. The preprocessing data consists from four stages. The first stage is to collect the actual data of vulnerabilities for critical system. Those data consists from input-output datasets which are a webmail system and some related services such as Perl, Apache server and Internet Explorer. The next stage is to select the variables of the available data. In this stage we use the CVSS V2 standard to calculate the frequency and impact of each vulnerability as mentioned in methodology section in chapter one. The selected variables shown in Figure (4.4). Third stage is to formatting the selected variables to numbers. Those values conducting from CVSS metrics score as mentioned in chapter three. The last stage is used to avoid the over fitting problem by using cross validation algorithm. Computing the frequency and impact for each vulnerability conducting to IA risk level value of each vulnerability. Each system has more than one vulnerability. Therefore, the maximum IA risk level of the vulnerability it will be the IA risk level for the system.

Chapter Five : Development of Models

5.1 Introduction

The major contribution of this thesis is developing a model that assess the IA risk level based on the input-output historical available data. This chapter is concerned with NEUROFUZZY BASED MODELING TECHNIQUES. Several models have been developed such as, Sugeno with Hybrid optimization techniques, Subtractive Slustering, Subtractive Slustering with Hybrid optimization techniques and check the adequacy of the developed model to demonstrate their performance. Mamdani model was also developed for the first stage.

Three measurement have been used to effectively check the adequacy of results, these measures are as illustrated below:

1. Correlation Coefficient (CC) which measure the correlation between the actual and predicted risk. This measure will be between the actual values and between the predicted outputs from developed model and is calculated by (Rae'd Basbous and Arafeh):

$$CC_{xy} = \sqrt{1 - \frac{\sum_{i=1}^N (y_i - x_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5-1)$$

Where y_i : is the i th actual data,

\bar{y} : is the average of all actual data,

x_i : is the i th predicted data,

N : is the number of data points under consideration.

2. The Mean Absolute Percentage Error (MAPE), which is expressed error as percentage. MAPE is the average of the absolute difference between the actual and forecasted divided by actual. The MAPE is calculated by (McSharry,2006 as cited in Rae'd Basbous and Arafeh,2009):

$$MAPE = \sum_{i=1}^N \left| \frac{y_i - x_i}{y_i} \right| * \frac{100}{N} \% \quad (5-2)$$

3. The Root Mean Square Error (RMSE), which is used to evaluate the error (differences) between the forecasted and actual loads. The general form of the RMSE equation for the actual risk (Y) and the predicted ones (X) is given by (Oriqat,2007 as cited in Rae'd Basbous and Arafeh,2009):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - x_i)^2}{(N-1)}} \quad (5-3)$$

4. Percentage of differences.

$$Percentage\ of\ Differences = \sum \left| \frac{y_i - x_i}{y_i} \right| * \frac{100}{N} \quad (5-4)$$

5.2 Sugeno Models with Hybrid Optimization Technique

The models are to be trained with historical data before testing them. The first step for training model is obtaining an accurate historical data and should be relevant to the model. The following section under section 5.2 have several Sugeno-based models with hybrid optimization techniques were developed for each stage to produce the overall IA risk model. *Figure (5.1)* illustrates a general developing training block diagram for each stage of our models. It consist of two main steps:

1. The first stage is pre-processing the input datasets for the system. For each stage we defined the inputs-outputs datasets that are used in the second step.
2. This stage is concerned with sugeno models using hybrid techniques. The processed datasets have been used in developing all models.

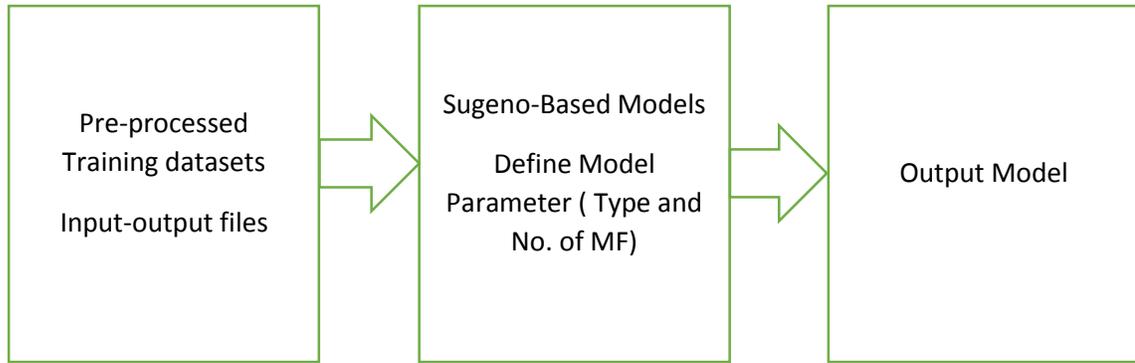


Figure (5.1) General Developing Training Block Diagram for Sugeno Model with Hybrid Technique

5.2.1 Initial Frequency Sugeno Models with Hybrid Optimization Technique

Hybrid learning algorithm (Jang 1993) and (Jang and Mizutani 1996) combines the Gradient Descent and the Least-Squares algorithm and it is the most widely used algorithm in literature to identify the parameters of the ANFIS. In this section we will import the data collection both the training and testing datasets to produce the Sugeno model for initial factor of frequency. In the initial stage we notice that each metrics in the initial frequency have a limited change in values such as the metrics access vector in base group have just three values (1,0.395,0.646) the same are for the other metrics, the access complexity and authentication. For this purpose we took all available dataset for the initial stage. All available data were 1129 record, after we applied the cross validation algorithm, two matrices were produced one for training with 847 records and the other for testing with 282 records and both of them were loaded into ANFIS Editor. *Figure (5.2)* illustrates the FIS model.

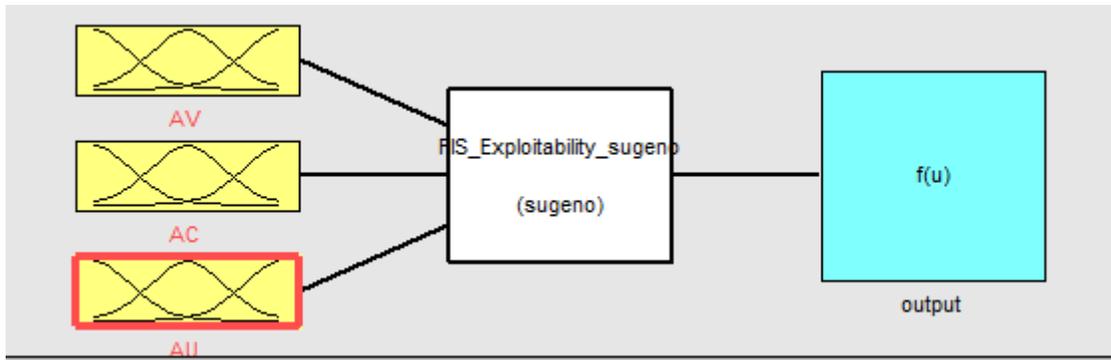


Figure (5.2) Initial Frequency FIS Model

In order to obtain the best results from the developed models, the model parameters (type and number of membership functions) need to be updated and determined manually. We will set the number of the MF's to 3 for each variable in the proposed model as the limited three values only. The type of MF's will be defined as triangular shapes according to (Bouchon-Meunier, Dotoli et al. 1996) recommendation for a singleton. *Table (5.1)* shows the ANFIS parameter for initial frequency developed model and *Table (5.2)* shows the output from the ANFIS process.

Table (5.1) Sugeno Model Parameter for Initial Frequency Developed Model

Generate FIS	No. Input	No. output	No MF's	Optimization technique	Epochs
Grid	3	1	3 3 3	Hybrid	100

Table (5.2) ANFIS Result for Initial Frequency Sugeno Developed Model

No. nodes	No. linear parameters	No. nonlinear parameters	Total number of parameters	No. training data pairs	No. fuzzy rules
78	108	27	135	847	27

Figure (5.3) represents the inputs MF's that have been used in building initial frequency developed model. For the initial frequency sugeno model a typical rule with three inputs (access vector, access

complexity, authentication) and one output (initial frequency), has the form (Arafah, Singh et al. 1999):

If AV is AV_j and AC is AC_k and AU is AU_l , then

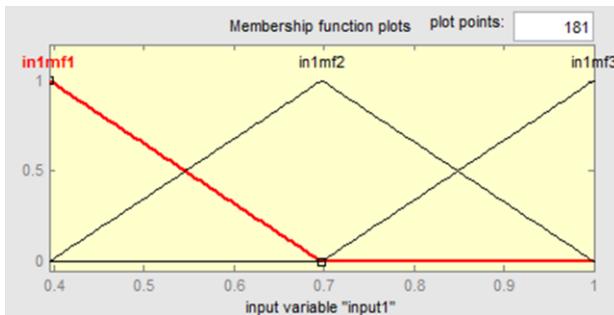
$$\text{The initial frequency} = p_i AV_j + q_i AC_k + r_i AU_l + s_i, \quad (5-5)$$

Where (j) represents the AV (Access Vector) input MF, (k) represent the AC (Access Complexity) input MF, and the (l) represents the AU (Authentication) MF. The term p_i , q_i , r_i , s_i , indicate the consequent parameters. For a zero-order sugeno model, the output level of initial frequency is a constant. The output level of initial frequency₁ of each rule is weighted by the firing strength w_i of the rule. For example, for an AND rule with $AV=AV_j$ and $AC=AC_k$ and $AU=AU_l$, the firing strength is (MathWorks 2008):

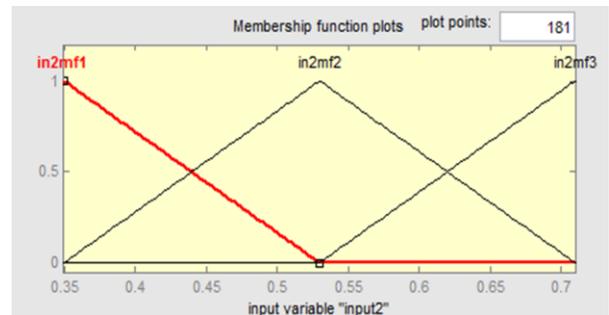
$$W_i = \text{AndMethod}(F1(AV_j), F2(AC_k), F3(AU_l)), \quad (5-6)$$

Where $F_{1,2,3}(\cdot)$ are the membership functions for AV, AC and AU. The final output for this stage is weighted average for all rule outputs (MathWorks 2008), as the following

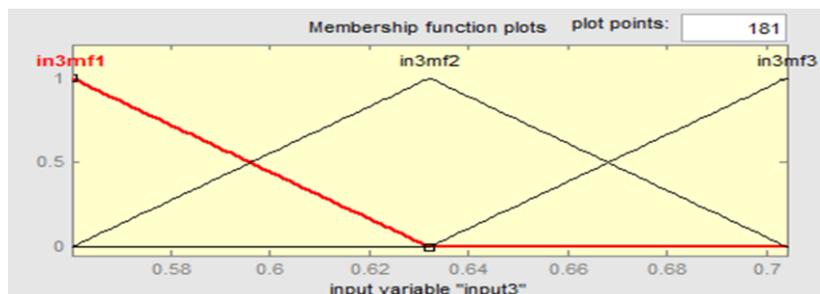
$$\text{Final Output} = \frac{\sum_{i=1}^N w_i \text{initialFreq}_i}{\sum_{i=1}^N w_i} \quad (5-7)$$



a Access Vector MF's



b Access Complexity MF's



c Authentication MF's

Figure (5.3) Initial Frequency Inputs (Access Vector, Access Complexity, Authentication) MF's

Table (5.3) Results for Initial Frequency Models with Hybrid Optimization

No. of MF	No. of Dataset	Testing Dataset		
		CC	MAPE	RMSE
3	283	0.9937	0.0035	0.0110

As shown in table above, very good results obtained from the sugeno model with hybrid optimization technique. A CC of 0.9937 describes the agreement between the actual and the predicted values for the initial frequency model. In addition, small values for the two error measures (MAPE and RMSE) show the error using two different formulas.

5.2.2 Initial Impact Sugeno Models with Hybrid Optimization Technique

The same as the section above (initial Sugeno model) we used the available dataset and input-output variable to obtain the initial impact Sugeno model using ANFIS with Hybrid optimization technique. We apply the same process as we applied above to produce the models. The dataset has 1129 input-output record, then we apply the cross validation record to produce the training and testing matrices. We used the same parameter as listed in *Table (5.1)* and the result as listed in *Table (5.2)*. *Figure (5.4)* shows the FIS model for the initial impact model.

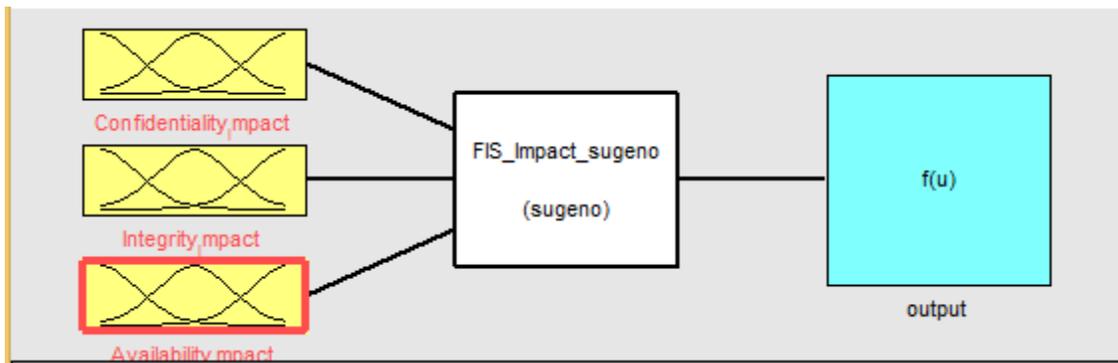


Figure (5.4) Initial Impact FIS Model

As the inputs of initial impact model (C, I, A) have the same range of values which consist from three numbers (0, 0.275, 0.66), the developed model propose three MF's for each as shown in

Figure (5.5) and Table (5.4) shows the results error for this model. The Figure (5.6) shows the initial impact FIS model.

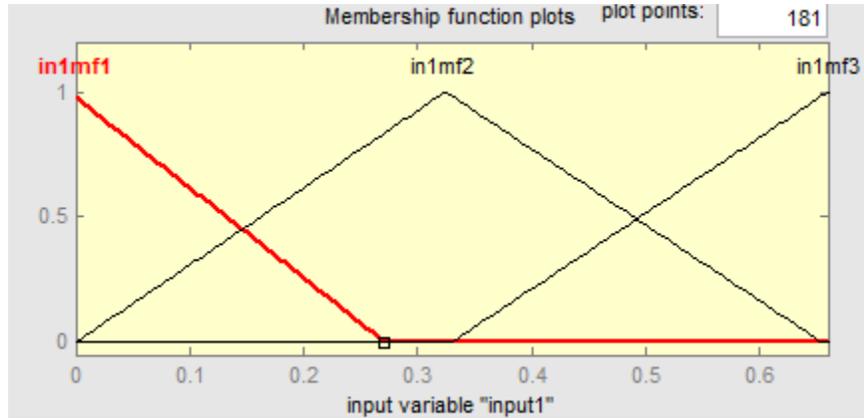


Figure (5.5) Initial Impact Model Input (Confidentiality, Integrity, Availability) MF's

Table (5.4) Results for Initial Impact Model with Hybrid

No. of MF	No. of Datasets	Testing Dataset		
		CC	MAPE	RMSE
3	283	1	9.33E-09	7.03E-09

The table above shown, the CC value are equal to 1 which mean the value of predicted and actual values are equal and this represent the strength of this model. Despite of the CC value are equal to 1, the error measures (MAPE and RMSE) have a slight error which can be neglected.

Figure (5.5) represents the inputs MF's that have been used in building initial impact developed model. For the initial impact sugeno model a typical rule with three inputs (Confidentiality, Integrity, Availability) and one output (initial impact), has the form (Arafah, Singh et al. 1999):

If C is C_j and I is I_k and A is A_l , then

$$\text{The initial impact} = p_i C_j + q_i I_k + r_i A_l + s_i, \quad (5-8)$$

Where (j) represents the C (Confidentiality Impact) input MF, (k) represent the I (Integrity Impact) input MF, and the (i) represents the A (Availability Impact) MF. The term p_i , q_i , r_i , s_i , indicate the consequent parameters. For a zero-order sugeno model, the output level of initial impact is a

constant. The output level of $initialimpact_1$ of each rule is weighted by the firing strength w_i of the rule. For example, for an AND rule with $C=C_j$ and $I=I_k$ and $A=A_l$, the firing strength is (MathWorks 2008):

$$W_i = AndMethod(F1(C_j), F2(I_k), F3(A_l)), \tag{5-9}$$

Where $F_{1,2,3}(\cdot)$ are the membership functions for C, I and A. The final output for this stage is weighted average for all rule outputs (MathWorks 2008), as the following

$$Final\ Output = \frac{\sum_{i=1}^N w_i\ initialimpact}{\sum_{i=1}^N w_i} \tag{5-10}$$

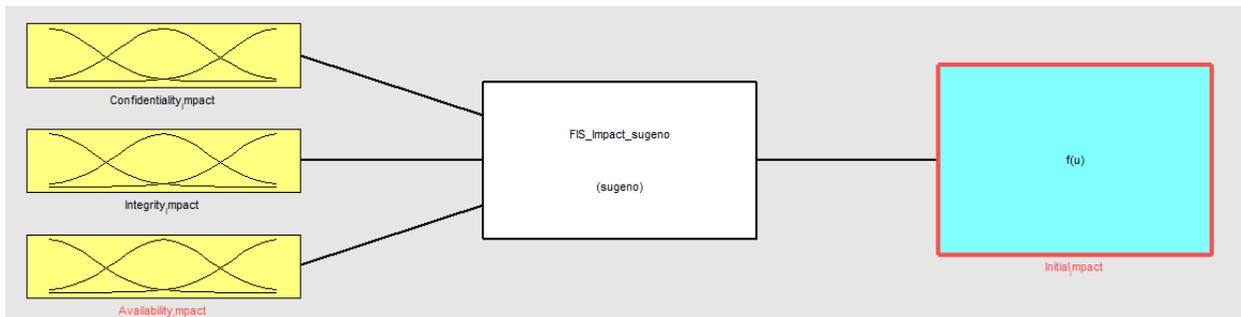


Figure (5.6) Initial Impact Sugeno FIS Model with Hybrid Optimization Technique

5.2.3 Updated Frequency Sugeno Models with Hybrid Optimization Technique

This model used the temporal metric which measures the state of exploit techniques or code availability that may increase or decrease the severity of vulnerabilities. The temporal metrics (Exploitability tools, Remediation Level, Report Confidence) with the output from initial frequency model are combined in this model to produce the updated frequency model. The dataset used in this model consist from four inputs and one output, this datasets are divided into two matrices, training dataset with 848 records and testing dataset with 283 records. Those matrices are imported into ANFIS editor tools to produce the model with hybrid optimization technique.

To generate the FIS we selected the 3 MF's for each input of temporal metrics with triangular shape as each of them limited with three values only. The fourth input has a range from 1 to 10

with different values not limited like the temporal metrics with some values so we change the number of MF's manually for this input to five to provide the best results. *Figure (5.7)* illustrates the updated frequency FIS Model. Whereas, *Figure (5.8)* shows the input MF's with numbers and types, the initial frequency input have 5 MF's and the 3 others (E, RL, CR) with 3 MF's.

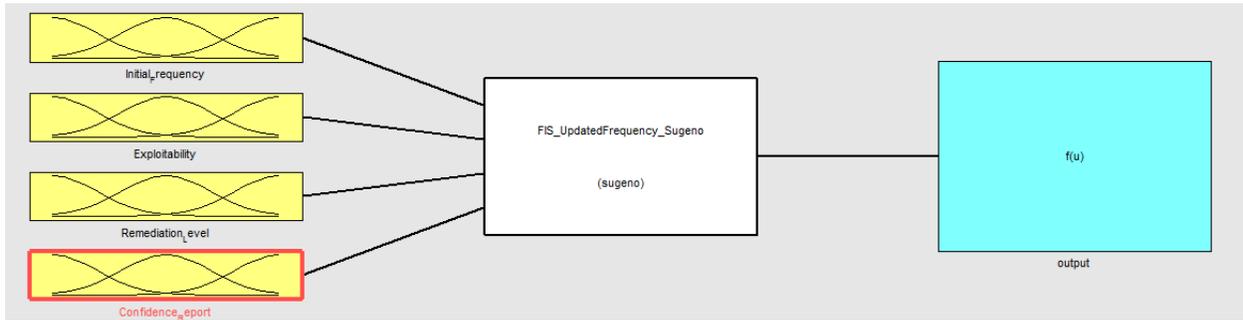
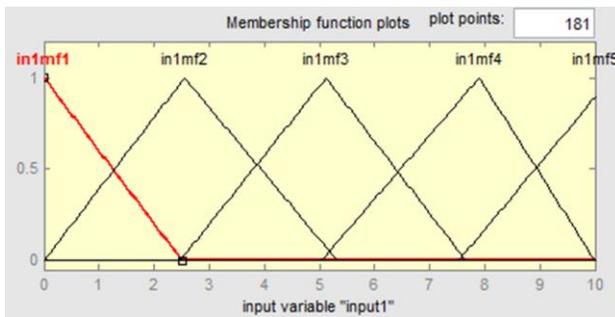
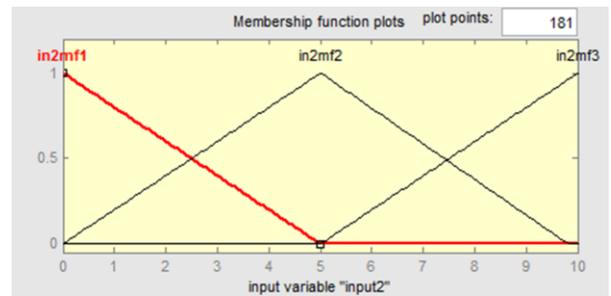


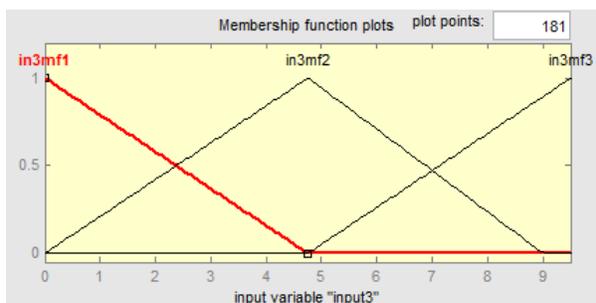
Figure (5.7) Updated Frequency FIS Model



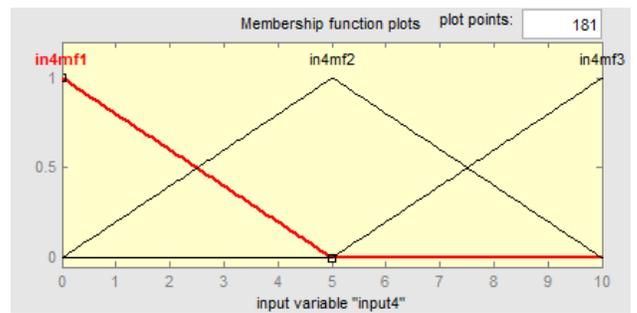
a Initial Frequency MF's



b Exploitability MF's



c Remediation Level MF's

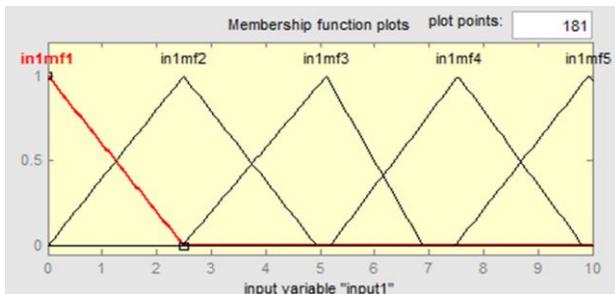


d Confidence Report MF's

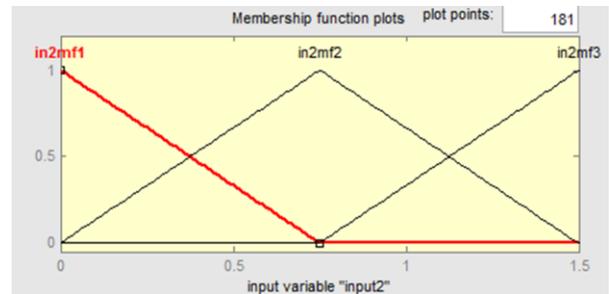
Figure (5.8) Updated Frequency Inputs MF's Sugeno Model with Hybrid Method

5.2.4 Updated Impact Sugeno Models with Hybrid Optimization Technique.

This model used the environmental metrics with the initial impact from base group. This model is asset-based which mean it depends on the needs capabilities for each asset in the organization. Suppose an email asset has a highly demand on availability and confidentiality but low demand in integrity on an organization so it needs to increase the availability and confidentiality requirements (AR, CR) and set those values with high and integrity requirement (IR) to low in environmental metrics. Whereas in other organization it is completely dependent on an email asset which mean the needs capabilities for email asset is highly demand in availability, confidentiality and integrity which needs to increase all requirements on to high. In this proposed model we assume a highly demand in C, A and low demand in I. The inputs MF's illustrated in Figure (5.9). The environmental metrics limited with three values. Therefore the number MF's are 3 for each with triangular shape as mentioned above for temporal metrics. Whereas the input of initial impact has a values ranging from 0 to 10 and not limited as the metrics values. Therefore we set the number of MF's to this input to 5 to be more accurate. *Figure (5.10)* illustrates the updated impact FIS model.



a Initial Impact MF's



b CR, AR, IR MF's

Figure (5.9) Updated Impact Sugeno Model Inputs MF's

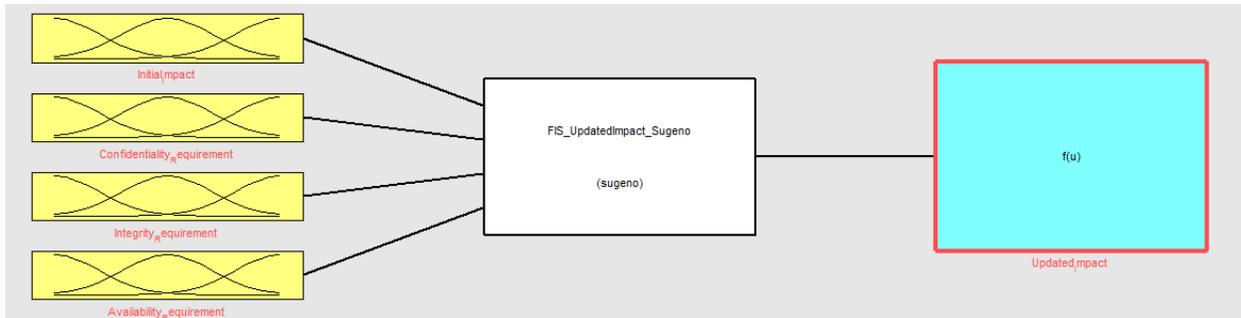
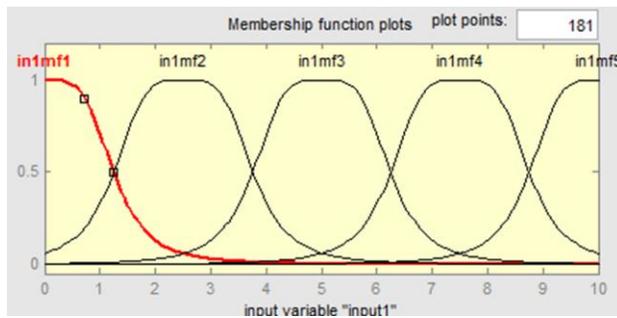


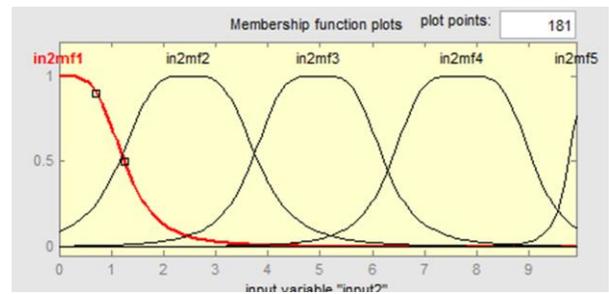
Figure (5.10) Updated Impact FIS Model

5.2.5 Vulnerability Risk Sugeno Models with Hybrid Optimization Technique.

In this model we combined the Updated frequency and impact to produce the total risk for each vulnerability. This model consists from 2 inputs and 1 outputs. The MF's for each inputs are set to 5 as the values are ranged and limited with some values as shown in *Figure (5.11)*. The *Figure (5.12)* shows the Risk Level FIS Model.



a Frequency MF's



b Impact MF's

Figure (5.11) Frequency and Impact Input MF's for Risk Vulnerability Sugeno Model

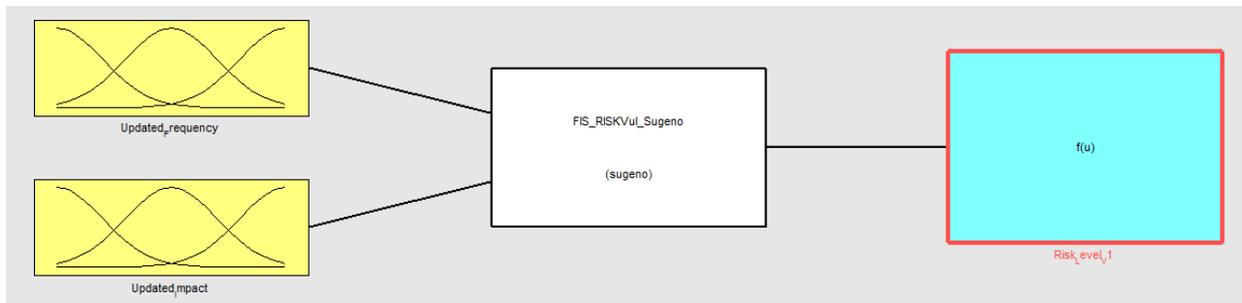


Figure (5.12) Risk Level FIS Model

5.3 Sugeno Models with Subtractive Clustering

The concept of data clustering is partitioning the dataset into several groups such that the similarity within a group is larger than that among groups. Clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction. By finding similarities in data, one can represent similar data with fewer symbols for example. Also if we can find groups of data, we can build a model of the problem based on those groupings. The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior. According to Chiu in (Chiu 1996) subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster center in a set of data. In this section we used the ANFIS tools to find clusters in input-output training data and to generate a Sugeno-type fuzzy inference system that best models the data behavior using a minimum number of rules. The same steps that have been followed in developing the models in the previous sections applied here. This model is a radius cluster-based that depends on the radius of the cluster. *Figure (5.13)* illustrates a general developing “training” block diagram of a subtractive model.

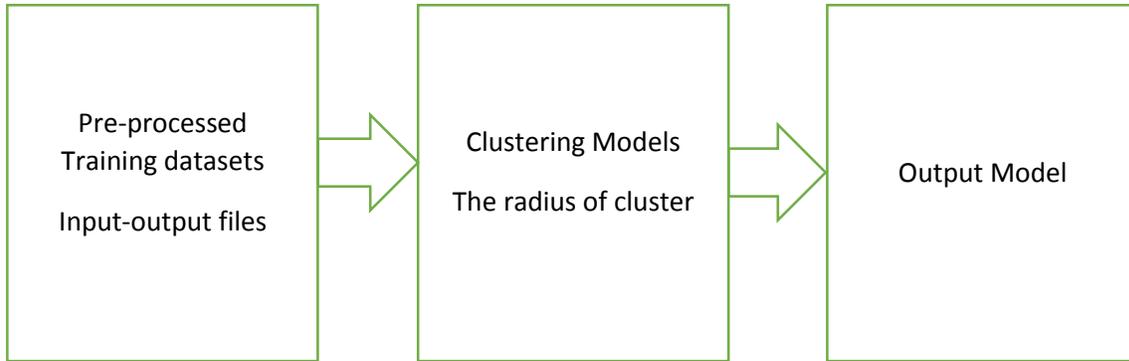


Figure (5.13) General Block Diagram for Developing/ Training Clustering Models

This model depends on the radius of cluster, therefore some of models in the next section may change this radius manual to produce the best results instead of finding the best number and type of MF's as in the previous section in Sugeno models with Hybrid optimization.

5.3.1 Initial Frequency Sugeno Models with Subtractive Clustering

The same dataset that were used in Sugeno model with Hybrid techniques are used here. In this experiment, the radius ranging from 0.1 to 0.5 which is the default value used by matlab fuzzy toolbox. It was noticed that increasing the value of the radius leads to decreasing the number of rules and to decrease the accuracy of developed model. As the limited values of dataset that we have, the number of rules ranging from 1 rule when the radius is 0.5 and to 2 rules when the radius ranged from 0.4 to 0.1. Also we notice that average testing error of FIS when the radius is 0.2 is 0.17248 which is the smallest error. The FIS model is same as shown in *Figure (5.2)*.

Table (5.5) Results Error for Initial Frequency Model with Subtractive Clustering

Value of radius	No. of Rules	Testing Dataset		
		CC	MAPE	RMSE
0.2	3	0.9927	0.0141	0.0118

The CC for the subtractive model is equal to 0.9927 (agreement between the actual and predicted values) whereas the error measure of MAPE is equal to 0.0141 and RMSE is equal to 0.0118.

Figure (5.14) shows the inputs MF's for each inputs, whereas each input has two MF' (AV, AC, AU).

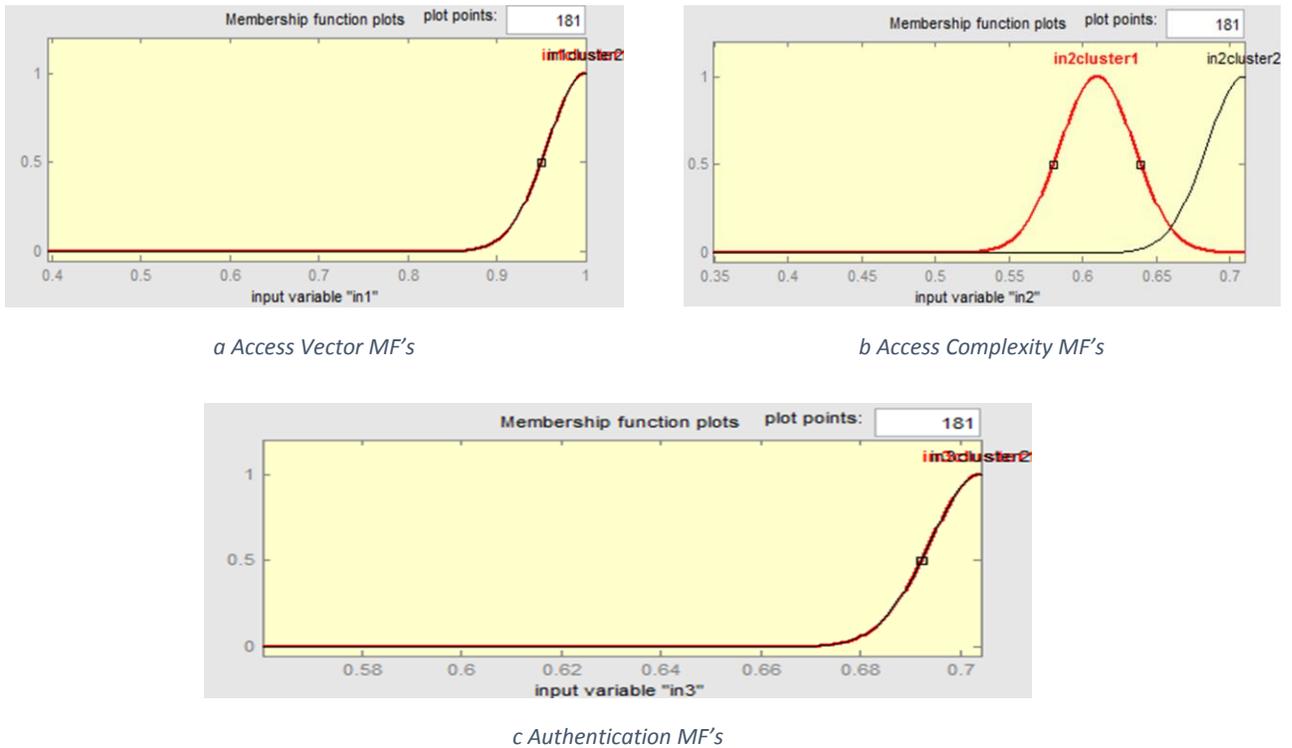


Figure (5.14) Initial Frequency Sugeno Model MF's with Clustering

5.3.2 Initial Impact Sugeno Models with Subtractive Clustering.

For this model we apply the same dataset applied before in initial impact Sugeno model with hybrid. In this model the radius from 0.1 to 0.5 has the same number of rules which is 5 rules. In this model we choose the radius to 0.5. The MF of the input variables (C, I, A) is shown in. *Figure (5.15).*

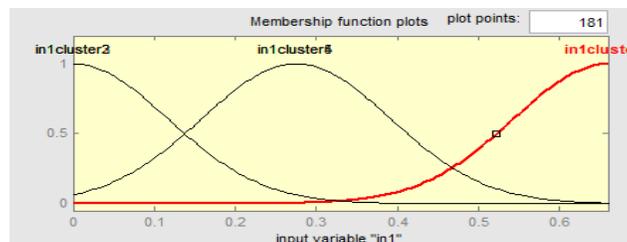


Figure (5.15) Initial Impact sugeno model with clustering Inputs MF's

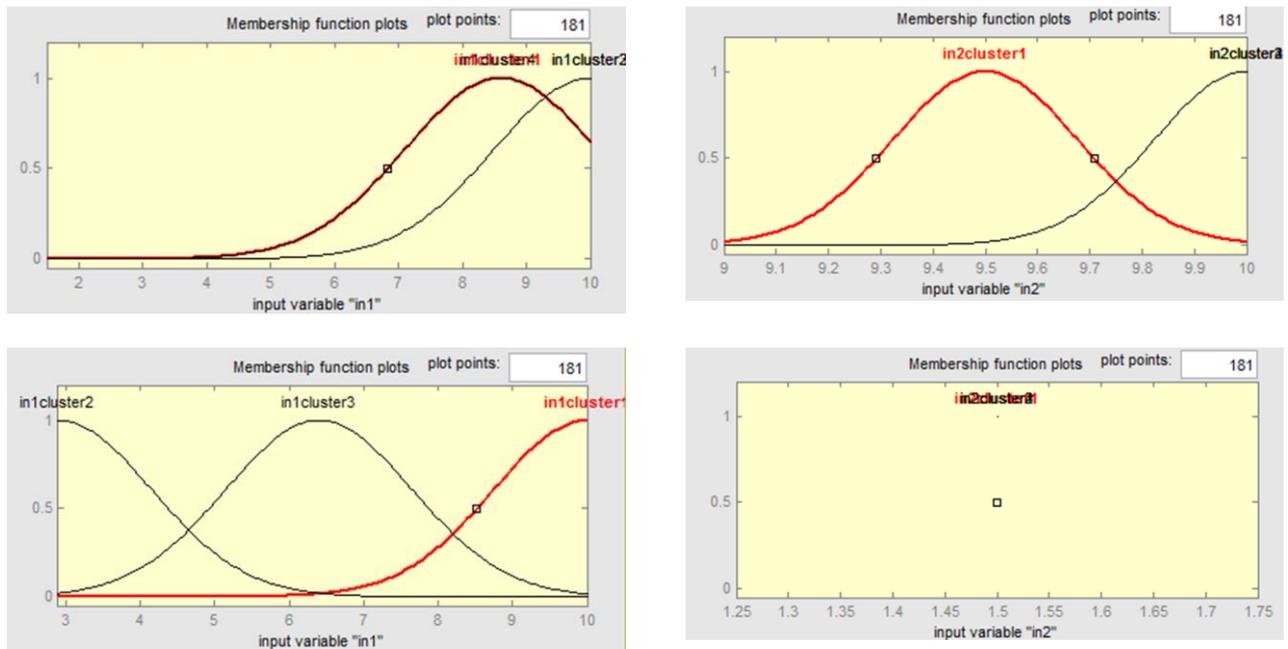
As listed in Table (5.6), there is no major difference between the real values and the predicted outputs.

Table (5.6) Results Error for Initial Impact Model with Subtractive Clustering

Value of radius	No. of Rules	Testing Dataset		
		CC	MAPE	RMSE
0.2	3	1	1.06876E-14	5.48058E-15

5.3.3 Updated Frequency Sugeno Models with Subtractive Clustering.

The output for this model are 4 rules when the radius of cluster is 0.5. The inputs MF's of this model are illustrate in *Figure (5.17)*.



A Initial Impact MF's

b. CR, IR, AR MF's

Figure (5.17) Updated Frequency Sugeno Model with Clustering Inputs MF's

5.3.4 Vulnerability Risk Sugeno Models with Subtractive Clustering.

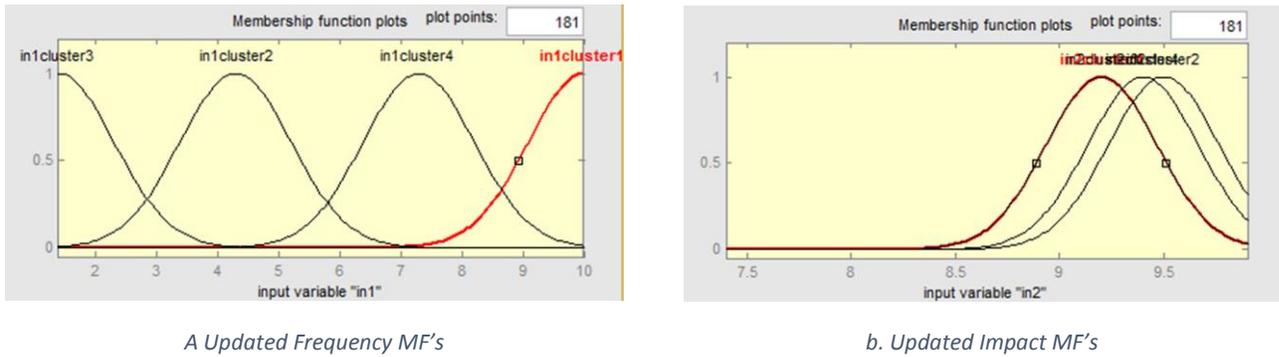


Figure (5.18) Vulnerability Risk Sugeno Model with Clustering Inputs MF's

5.4 Sugeno Cascaded Models with Subtractive Clustering and Hybrid Optimization.

The purpose of cascaded model is to achieve a more accurate model. This model the same as above section (cluster model) but with using a hybrid optimization technique. Therefore the same steps that mentioned in the previous sections should be followed in building cascaded model. At first a sugeno model using clustering should be developed in the same way as mentioned in the previous section. After that the constructed model should be enhanced using the hybrid optimization technique as shown in *Figure (5.19)*.

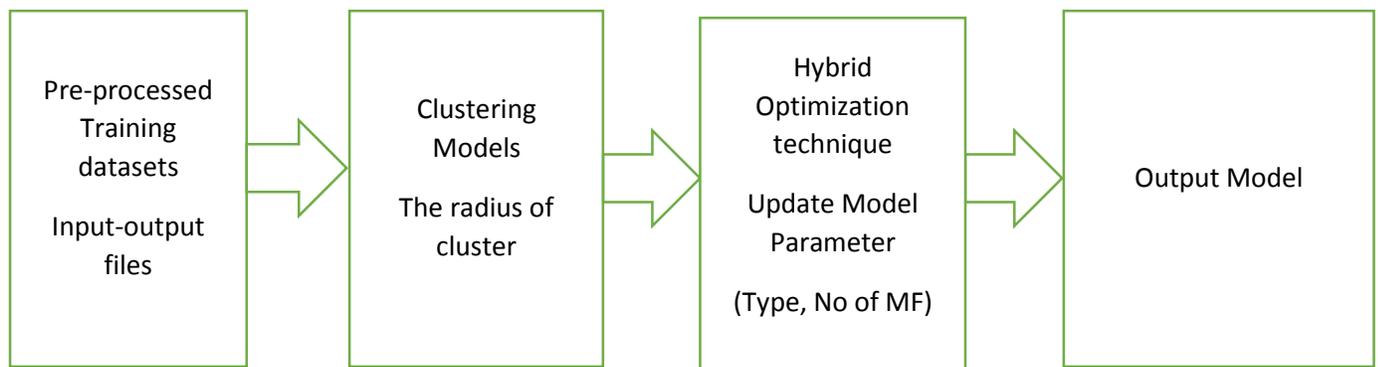


Figure (5.19) A General Block Diagram for Developing/Training Cascaded Clustering with Hybrid Optimization

As shown above in Figure (5.19), developing a Cascaded model consists of two main stages. At first, the same training datasets have been used to construct a Sugeno model using Clustering.

Then, Hybrid optimization technique has been applied to fine tuning the constructed model parameters to achieve a more accurate model.

5.4.1 Initial Frequency Sugeno Models with Subtractive Clustering and Hybrid Optimization technique.

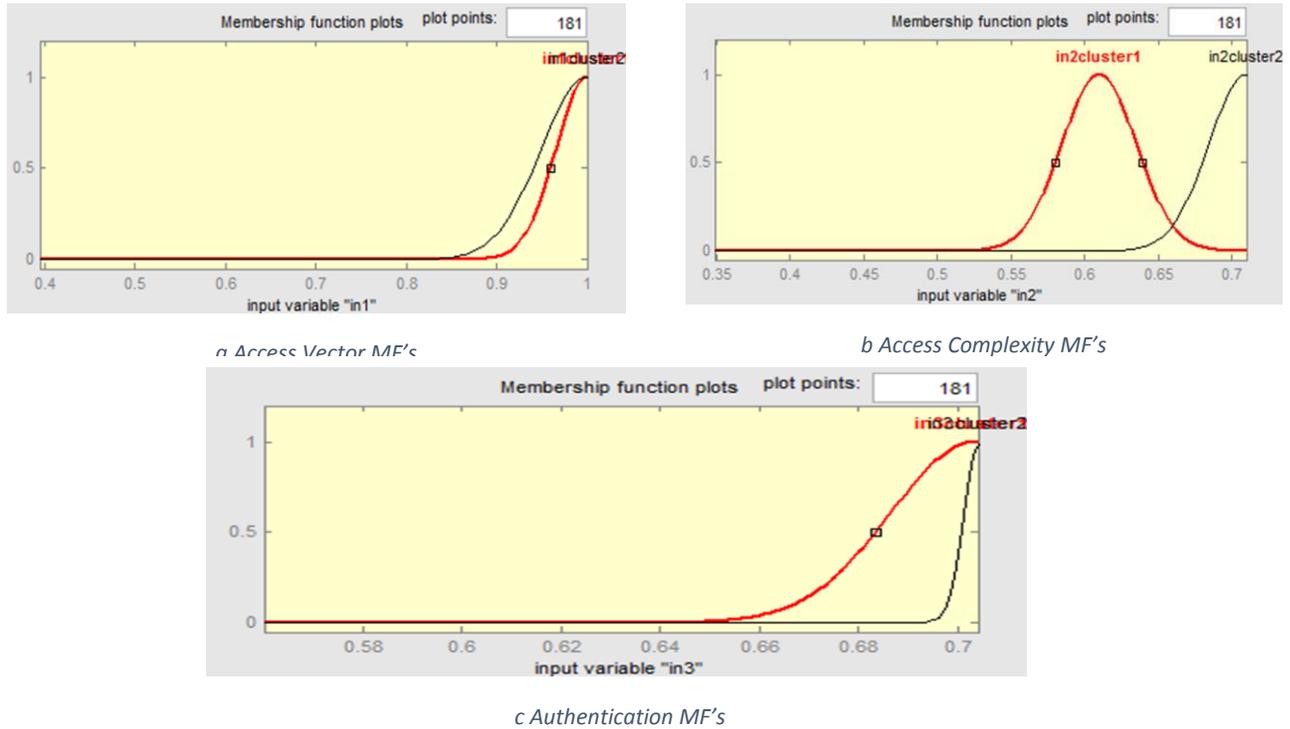


Figure (5.20) Initial Frequency Sugeno Model with Clustering and Hybrid inputs MF's

Table (5.7) Results Error for Initial Frequency Model with Subtractive Clustering and Hybrid Method

Value of radius	No. of Rules	Testing Dataset		
		CC	MAPE	RMSE
0.2	3	0.9880	0.008	0.0146

5.4.2 Initial Impact Sugeno Models with Clustering and Hybrid Optimization technique.

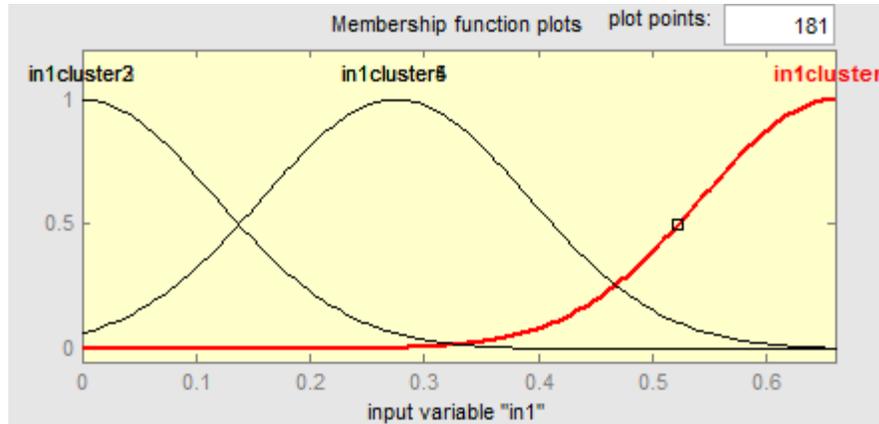


Figure (5.21) Initial Impact sugeno model with clustering and Hybrid Inputs MF's

The same models are applying for updated frequency, impact and risk vulnerability using Subtractive Clustering with Hybrid optimization technique.

5.5 Mamdani Fuzzy Inference Method.

The second type of FIS method is using Mamdani FIS method which is the most commonly seen fuzzy methodologies. Mamdani's method was among the first control systems built using fuzzy set theory. It was proposed by (Mamdani and Assilian 1975) as an attempt to control a steam engine and boiler combination by synthesizing a set of linguistic control rules obtained from experienced human operators Mamdani's effort was based on (Zadeh 1973) paper on fuzzy algorithms for complex systems and decision processes. Mamdani's method was among the first control systems built using fuzzy set theory. Another thesis contribution is to using the Mamdani method to present the vulnerability risk assessment model. To produce the Mamdani model of our system, steps must be defined before ranging from determining a set of fuzzy rules fuzzifying the inputs using the input membership function, combining the fuzzified inputs according to the fuzzy

rules and combining the rule strength and the output membership function. By using this method we apply the first model “Initial Frequency” by defined the inputs, outputs, and fuzzy rules, input output MF’s and compares with the other method we mentioned above.

5.5.1 Initial Frequency Mamdani Fuzzy Inference Model

We model an initial frequency model as a set of fuzzy attributes. The first attributes we will look at are the access vector (AV), access complexity (AC), authentication (AU). Some of the value ranges used to fuzzify them in this work correspond to the value definitions used in CVSS. This values are used to simplify the task of choosing appropriate values of attribute ranges, and also to capitalize on the expertise put into establishing these values. *Figure (5.22)* shows the input-output attributes membership functions. The fuzzy AV attribute is shown in *Figure (5.22) (a)*. The “Local” attribute that lies between 0.36 and 0.41, but never exceeds 0.52. Similarly, the “Adjacent” access represents a linguistic value that is never below 0.47, but is most certainly between 0.604 and 0.656 and never exceed 0.74. Similarly, the “Network” access represents a linguistic value that is never below 0.68, but is most certainly between 0.94 and 1. *Figure (5.22) (b)* shows the fuzzy AC attribute. The “High” terms represents a linguistic value that lies between 0.3 and 0.374, but never exceeds 0.4167. The same as for “Medium” and “Low”. The same as for the AU input attributes. For every fuzzy inference system (FIS), a fuzzy output variable has to be defined before any inference is performed. The above input fuzzy attributes are combined using fuzzy rules to give a fuzzy output values. The *Figure (5.22) (d)* shows the fuzzy output “Initial frequency” we define the output fuzzy as a smooth Gaussian MF in order to be able to distinguish between small inference differences.

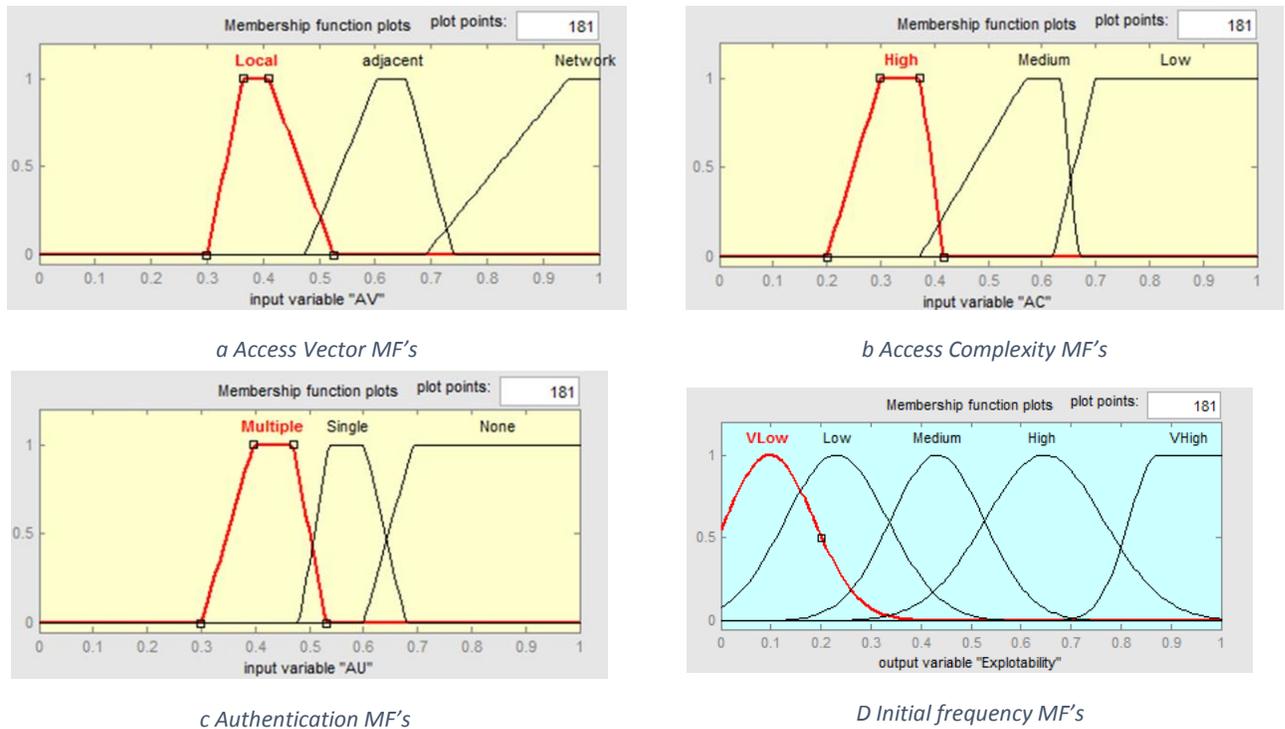


Figure (5.22) Initial Frequency Mamdani Model Inputs MF's

We use if- then- rules to combine the attributes based on the linguistic declarations about the attributes. Rules can be given weights depending on the importance of a rule over others. In this model we defined 27 rules with equal weight.

Table (5.8) Results Error for Mamdani Model of Initial Frequency

No. of inputs	Testing Dataset		
	CC	MAPE	RMSE
3	0.91373	0.080646	0.042078

5.6 Simulink Model: IA Vulnerability Risk Assessment.

We use the Simulink library browser to build our own Simulink systems that present the FIS model that we mentioned before. To simulate the four stage IA Risk assessment, a Simulink-based model has been developed as shown below. The overall Simulink model is shown in *Figure (5.26)*. To

be clear the figure we divided it in more figures. *Figure (5.23)* shows the Simulink of frequency model which consist from two stages (initial and updated frequency). The initial frequency calculated from AV, AC, AU, whereas, the updated frequency calculated from initial frequency from the first stage combined with other three metrics E, RL, RC.

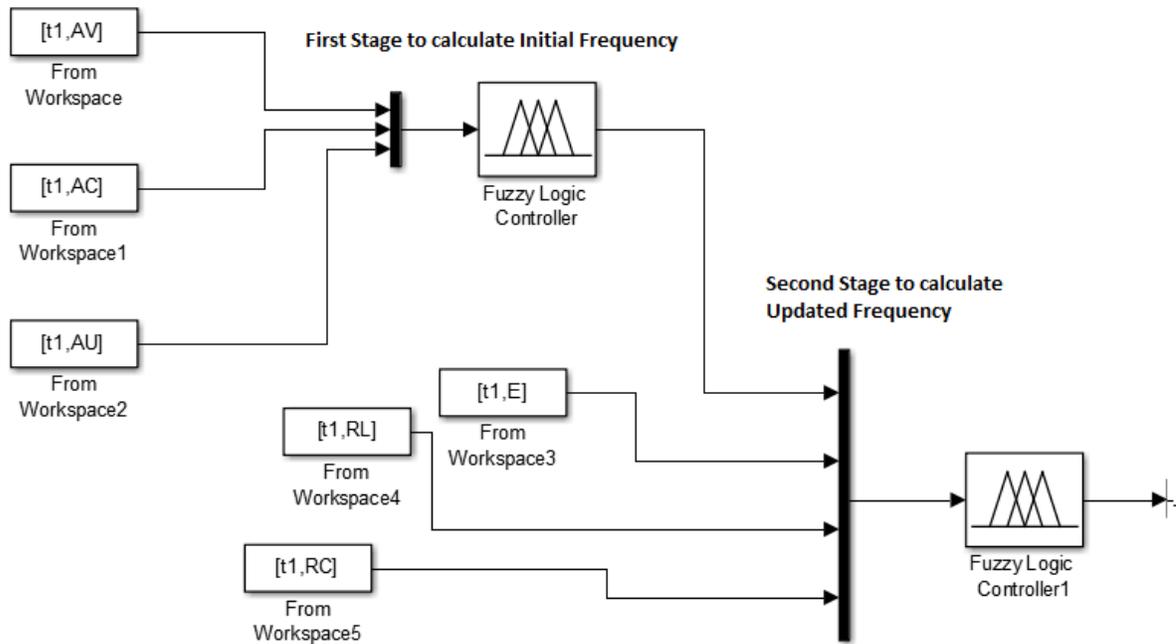


Figure (5.23) Calculate Frequency Value used Simulink

Figure (5.24) shows the Simulink of Impact model which consist from two stages (initial and updated Impact). The initial Impact calculated from C, I, A, whereas, the updated impact calculated from initial impact from the first stage combined with other three metrics CR, IR, AR.

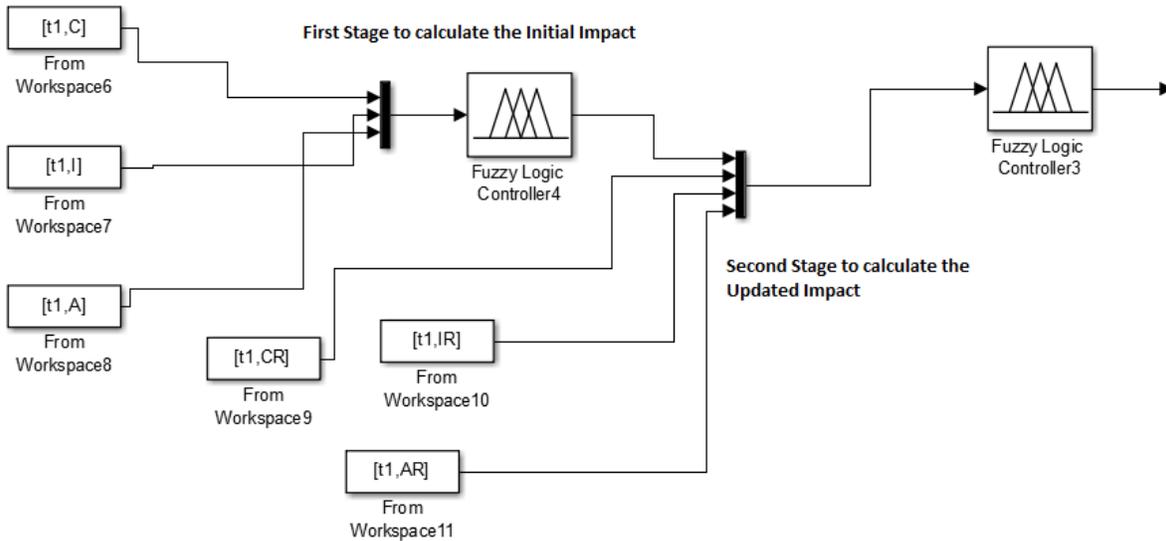


Figure (5.24) Calculate Impact value using Simulink

Figure (5.25) shows how to calculate the IA Risk level for one asset. The updated frequency and impact are combined to calculate the risk level for the vulnerability. All vulnerabilities are combined as input to Mamdani fuzzy system to calculate the IA Risk level for the asset. Figure (5.26) shows the overall Simulink model.

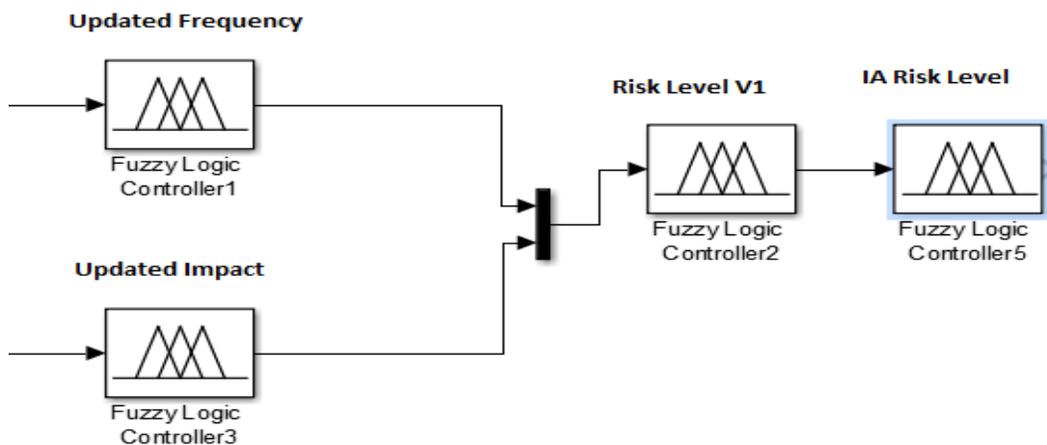


Figure (5.25) IA Risk Level using Simulink

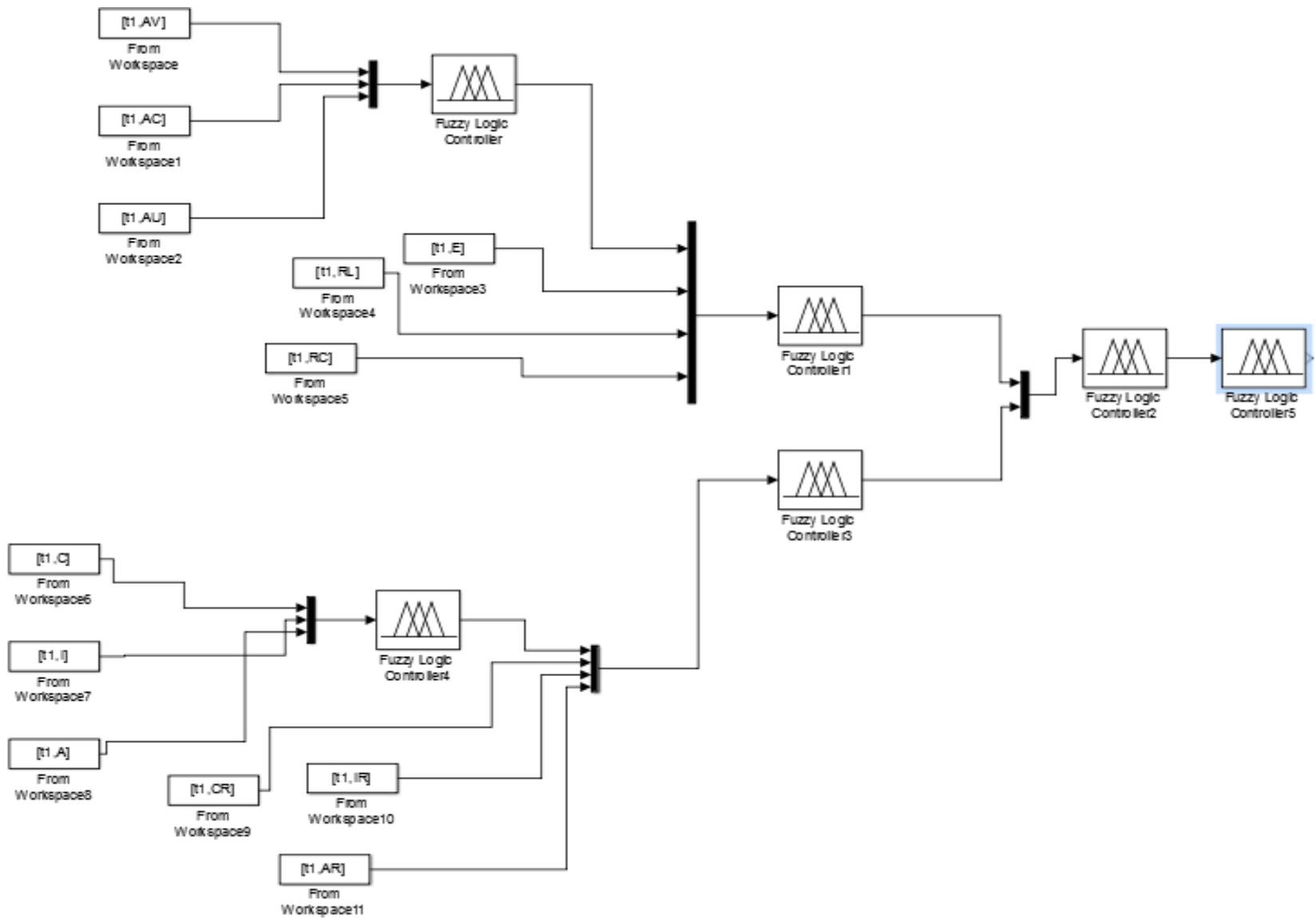


Figure (5.26) IA vulnerability Risk Assessment Simulink Model

Chapter Six : Results and Discussions

6.1 Introduction

The overall model consists from three stages to produce a risk value for each vulnerability. These stages models have been developed based on FIS using different methods (Sugeno and Mamdani). Sugeno Models have been developed using ANFIS with different methods. Sugeno Model using Grid with Hybrid optimization technique have been developed, and finally Sugeno model with subtractive clustering with and without hybrid optimization technique have been developed.

Different measures have been used to check the adequacy of the developed models for each stage. These measures including the CC, MAPE, RMSE and Percentage of Differences.

For each stage, the inputs were defined and were used to develop the models as mentioned in chapter four. Different models using different methods and techniques have been developed for each model to obtain best results. The next section illustrates a detailed comparison and discussion about the developed models.

6.2 Results and Comparisons between the Developed Models.

In this thesis and using a sample of historical datasets profiles presented in chapter four, we have started by developing five sugeno models with hybrid optimization technique. These five models are used together to produce the overall system model. The overall model is used to predict the risk value for one vulnerability. Then all risks values of vulnerabilities for one asset are combined together as an input to final Mamdani model to produce the final IA risk level for those asset. Another five models have been developed by using the same datasets with a subtractive clustering

methods. Then the subtractive clustering with the hybrid optimization technique have been used to construct a cascade models.

As mentioned in chapter five, the training datasets and models parameters (numbers and type of MF's, number of rules, and cluster radius) have been fixed and used for proposed models. For example the initial frequency and impact model with hybrid have only three values, therefore each model has 3 MF to represent the values. Whereas the updated impact and frequency have a range in values, therefore each model of them have 5 MF to obtain the best results. By using subtractive clustering, the radius cluster in initial frequency model is 0.2 which obtain the best results. Whereas the updated frequency model with subtractive clustering method using 0.5 values of radius cluster. *Figure (6.1)* show the CC measures for results obtained from the developed initial frequency models. The figure shows the CC for testing datasets. It is clear from the figure that the best results obtained from the models that have been developed to predict the initial frequency using the grid partition.

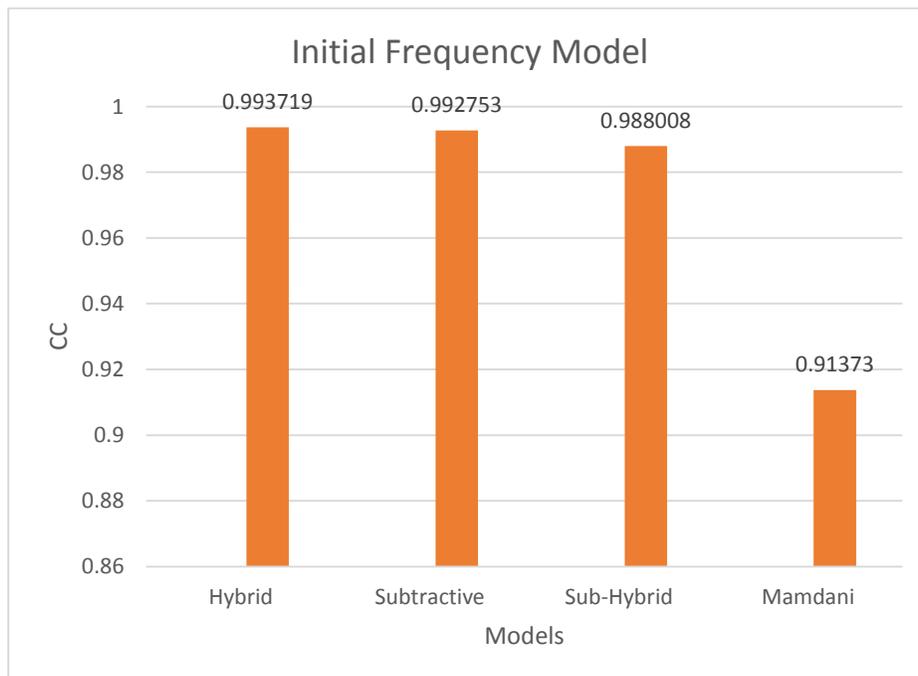


Figure (6.1) The Correlations Measures for Results Obtained from Initial Frequency Models

With hybrid optimization technique. These high CC values that have been obtained refer to the agreements between the predicted data and the original data. The CC for the developed model with hybrid equal to 0.993 which is close to 1. Table (6.1) lists the correlation measures for the results obtained from the developed initial frequency model. The results from the above figure can be noticed in the table.

Table (6.1) The Correlation Measures for the Developed Initial Frequency Model

Models	Hybrid	Subtractive	Sub-Hybrid	Mamdani
Initial Frequency	0.993719	0.992753	0.988008	0.91373

To have a solid conclusion, the other three measures have been used. We notice that when the CC value increased the value of MAPE, RMSE and percentage of differences decreased. The MAPE and RMSE have been used to examine and show the adequacy of the developed model and its outcome. The error measures RMSE and MAPE give an indication how the performance of the developed models are. *Figure (6.2)* and *Table (6.2)* below represents a summary chart graph for the MAPE values calculated for results obtained from all developed models with different optimization techniques. As shown in figure below the MAPE values for hybrid model is the

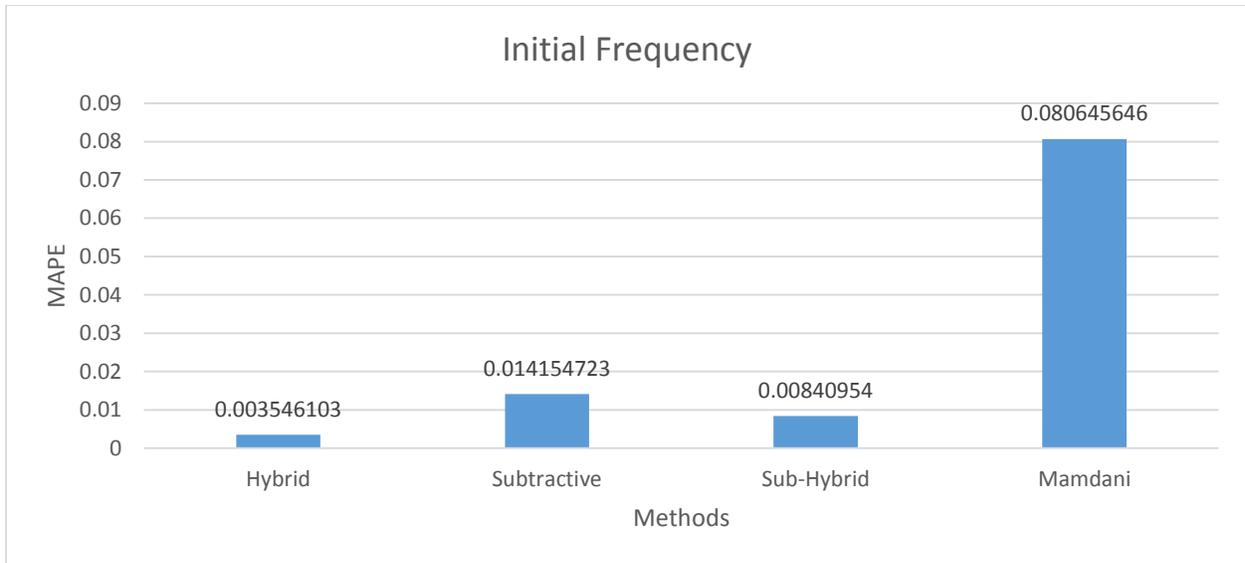


Figure (6.2) The MAPE Measures Chart for initial Frequency Models

Lowest one and this reflect the highest CC that achieved from these models as shown in the table listed below. Also we notice that the MAPE value 0.014 of Subtractive models has been furtherly reduced to 0.008 by cascading the subtractive clustering and hybrid optimization technique.

Table (6.2) The MAPE Measures Table for initial Frequency Models

Models	Hybrid	Subtractive	Sub-Hybrid	Mamdani
Initial Frequency	0.003546	0.014155	0.00841	0.080646

The other error measure values (RMSE) shown in *Figure (6.3)*. This measures shows the adequacy of the developed models in addition to the MAPE measures. The same thing for the RMSE results as in the MAPE results achieved where the grid partition with hybrid optimization technique has the best results (the lowest RMSE values which equal to 0.011). *Table (6.3)* shows the results of the error measures RMSE for all models. These results are shown in figure below.

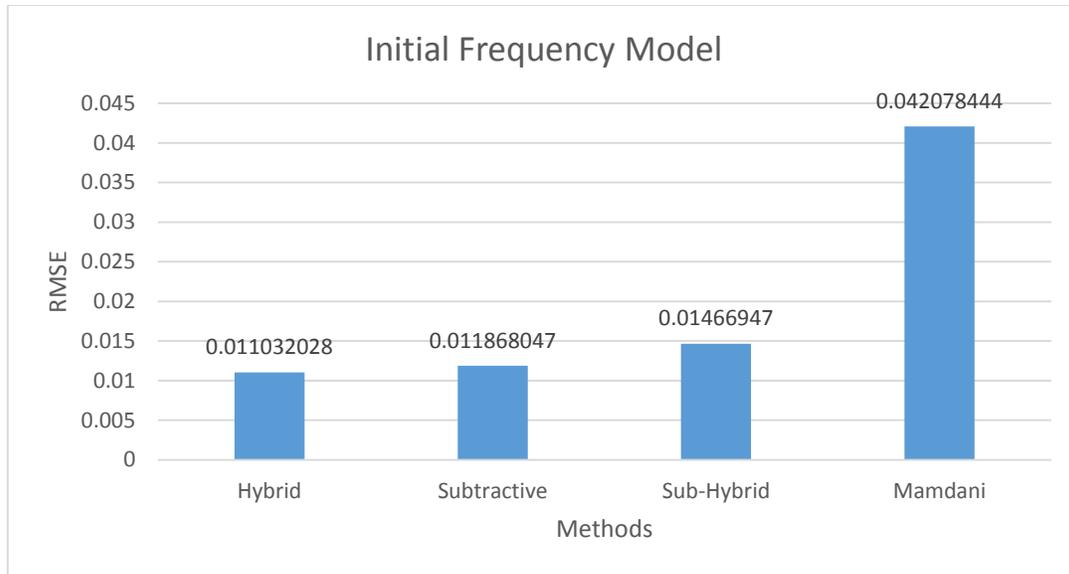


Figure (6.3) The RMSE Measures Chart for Initial Frequency Models

Table (6.3) The RMSE Measures Table for Initial Frequency Models

Models	Hybrid	Subtractive	Sub-Hybrid	Mamdani
Initial Frequency	0.011032	0.011868	0.014669	0.042078

Another error measures values (Percentage of differences) shown graphically in Figure (6.4) and listed the values in Table (6.4). As noticed from table and figure below the smallest difference error values for initial frequency model by using hybrid technique which equal to 0.354.

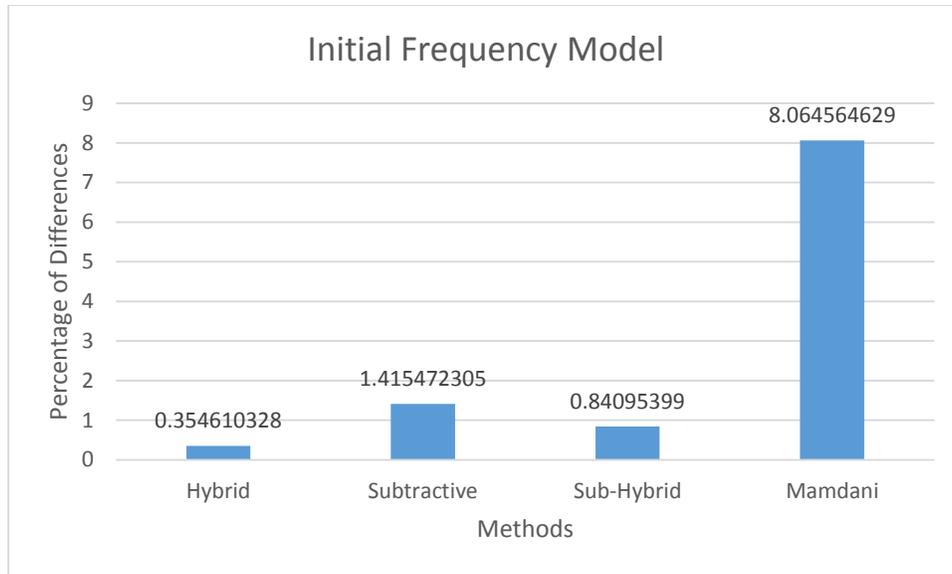


Figure (6.4) The Percentage of Differences Chart for initial Frequency Models

As we notice from the figure above that the measures of percentage of difference values between the real values and predicted values is above 1 where the other errors values below the 1.

Table (6.4) The Percentage of Differences Table for initial Frequency Models

Models	Hybrid	Subtractive	Sub-Hybrid	Mamdani
Initial Frequency Model	0.35461	1.415472	0.840954	8.064565

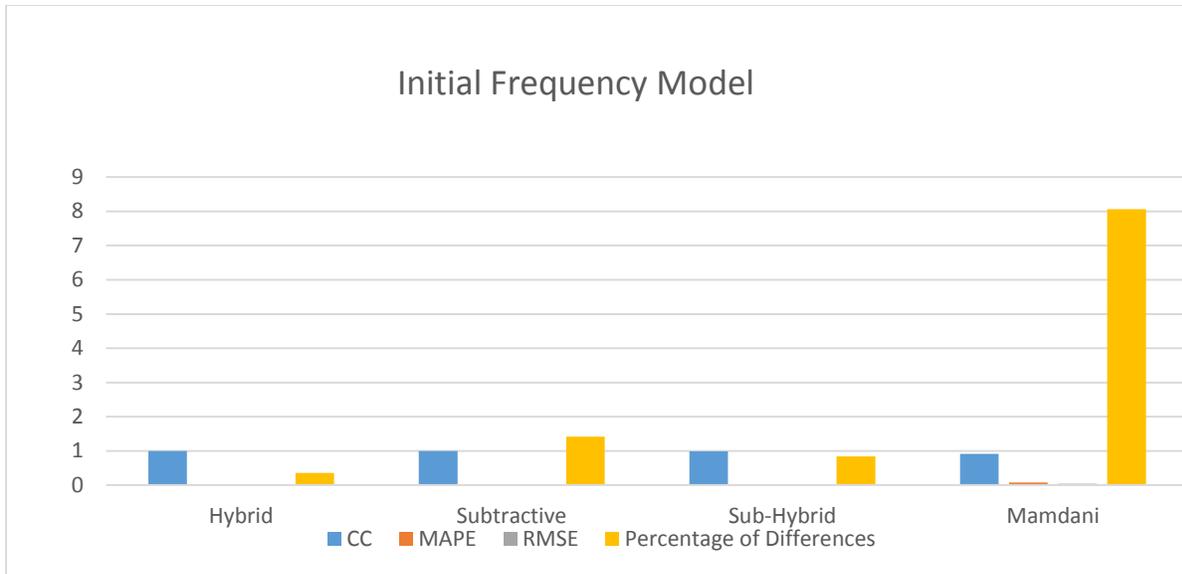


Figure (6.5) All Measures Values (CC, MAPE, RMSE, Percentage of Differences) for Initial Frequency Models

A graphical representation for the CC and error measures (MAPE, RMASE, percentage of difference) are shown in figures above. A relation can be conclude from charts above which is: an increasing in the CC leads to decrease in the error measures (MAPE, RMSE and percentage of differences). This chapter can be summarized by the following points:

1. The developed grid partition model with hybrid optimization technique produced the highest results. The value of CC between the actual and predicted values equal to 0.993719.
2. The developed Mamdani model has the lowest CC compared to the other models and similarly for the two error measures. The CC value for this model is equal to 0.91373.
3. Finally, we can notice that the highest the CC the lowest the MAPE, RMSE and percentage of differences.

6.3 Comparison with Other Studies

As mentioned in chapter two, plenty works can be found in the risk assessment field, but it is important to mention that different datasets and different approaches have been used in those works which make it difficult to compare our models and findings with other studies. The lack of data that we have obtained and the denied our request for real data from different organization will make it difficult to check the adequacy of our models and compare it with other studies. But we can compare as models and techniques that used.

Rani (Rani 2013) has proposed a neuro-fuzzy approach to estimate the software risk in all stages of software development life cycle (SDLC). Firstly he used the fuzzy inference system with 17 input risk attribute. The input attributes were identified by a fuzzy terms, rules and output. After the Fuzzy Inference system he created then Neural Network based three different training algorithms: BR (Bayesian Regulation), BP (Back Propagation) and LM (Levenberg-Marquardt) are used to train the neural network. This model is applicable only to software design during the software development life cycle. Whereas our model can be used for hardware and software vulnerabilities.

Shameli and Shajari have presented in their study (Shameli-Sendi, Shajari et al. 2012) a practical model for information security risk assessment. This model is based on multi-criteria decision-making and uses fuzzy logic. The proposed risk assessment is a qualitative approach according to ISO/IEC 27005 standard. In the proposed model, a fuzzy technique was used to connect expert opinion with linguistic variables. These linguistic variables reflect the expert opinions. In this model determined the likelihood and impact of each threat, effective criteria for their measurement have been considered. But the lack of quantitative data and the rapidly changing security environment makes it hard to derive accurate measures over such a long time-period and the risk value is expert specific.

Sendi and coauthors have presented in their study (Sendi, Jabbarifar et al. 2010) the FEMRA model which uses the fuzzy expert systems to assess the risk in organizations. The risk assessment varies considerably with the context, the metrics used as dependent variables, and the opinions of the persons involved. This model represents each risk with numerical values. The authors presented three steps to achieve the goal. The first step to identify the assets which uses a security cube to identify and classify the assets. Then list all potential threats applicable to these assets. The second step is to generate a list of asset vulnerabilities and risks. The final step is to calculate the effect risks which sing the fuzzy models. In this model the values for each asset is taken from three experts in terms of CIA triad and then calculate the average. But the lack of quantitative data and the rapidly changing security environment makes it hard to derive accurate measures over such a long time-period and the risk value is expert specific.

Dondo has presented in his study (Dondo 2008) a fuzzy system approach for assessing the individual asset by calculating the potential risk exposure for the vulnerabilities associated with these assets. Then the analyzer can rank the vulnerabilities associated with the asset. This model was based on CVSS attributes which defined the CVSS attributes as a Key Risk Indicators (KRIs). The proposed method models the KRIs as a fuzzy variables based on a combination of experience, expertise, or historical input and defined the MF for each variable. Then combine all the identified KRIs into FIS to come up with a final risk value. The FIS determine the fuzzy risk value represented by its CIA components. The combination between the impact and likelihood of the attack will produce the final risk value. Finally, defuzzify the result back into a crisp value and compare the results for each vulnerability in order to rank them. The construct asset value (AV) was used to derive the risk level or risks to a system. The asset value (AV) is assumed given. The approach derives risk level based on the CVSS base metrics variables, a measure of time from

when the vulnerability was reported and the safeguards already in the system. The author applied fuzzy rules to compute impact (I) and likelihood (L) and derive risk level as: $AV \times I \times L$. This approach is similar to our model, but our model does not use fuzzy rules. Our model uses the temporal and environmental metric groups given in the CVSS to estimate the risk level rather than asset value and safeguard. Asset value is not always easy to evaluate and might be stakeholder specific. AV is not a generalizable variable, but rather context and stakeholder specific. Our models is based on CVSS, which is an open standard that also reveals the details behind the scores provided. Furthermore, CVSS is regularly updated and several information sources is taken into consideration when calculating the CVSS score.

Houmb and Franqueira have presented in their study (Houmb and Franqueira 2009) a Target of Evaluation (ToE) risk level estimation model that uses CVSS to estimate misuse frequency (MF) and misuse impact (MI), and from these derive the risk level of ToE. This is a general risk in which this model works on the level of vulnerabilities and is able to compose the vulnerabilities into service levels. The service levels define the potential risk levels and are modelled as a Markov process, which are then used to predict the risk level at a particular time. MF is estimated from attributes in the base and temporal metrics of CVSS and MI is estimated from attributes in the base and environmental metrics of CVSS. The base metrics of CVSS is used to establish the initial estimates of both MF and MI. MF is then made attack specific by adding in factors concerning the attack tools available, the existing security measures and the report confidence. For MI, the initial MI of a potential vulnerability exploit (attack) derived from the base metrics is made ToE specific by taking the relevant security requirements into consideration. An important factor to note for MI

is that there are no impacts of a potential vulnerability exploit (attack) if there are no relevant requirements.

Chapter Seven: Conclusions and Suggestions for Future Works

7.1 Conclusions

The general objective of this work is to calculate the IA risk level for the vulnerabilities that can be exploited by an attack and then affect the IA of the asset. To achieve this work we have adopted the soft computing and artificial intelligence approach to develop a system that predict the risk for each vulnerability that can affect the assurance and ability of the asset. After calculating all risk values of vulnerabilities, then the organization can maximize the security requirements to minimize the risk based on capability asset needs.

In conclusion, it is mentioned that soft computing is an emerging approach which parallels the remarkable ability of the human mind to reason and learn in an environment of uncertainty and imprecision. While artificial intelligence approaches imitate human beings way of thinking and reasoning to get knowledge from the past experience and predict the future risk.

A state of arts about IA risks has been presented in this thesis. The approaches of IA risk level can be mainly divided into two categories: Guidelines approaches and artificial approaches. In the guidelines approaches which are used by a checklist guidelines or a third party company. These approaches are used to assess the organization based on checklist guidelines which are very expensive and take a long time. While the artificial intelligence approaches try to imitate human beings' way of thinking and reasoning to get knowledge from the past experience. Expert system, ANN and Fuzzy inference approaches belong to the artificial intelligence category.

In our review of literature survey in risk assessment approaches we have found that various variables could be considered for risk assessment such as access vector, authentication, access complexity, remediation level, the availability and easy to use of exploitability tools...etc. other approaches used the experts opinion to assessment the vulnerabilities. This thesis used the

variables metrics of CVSS v2. This thesis is composed of two parts: historical data treatment, individual approaches proposed for risk assessment. The overall system composed of five models to produce the vulnerability risk value. Then combined all values of risk for each vulnerability for one asset to the final Mamdani model.

The sample of historical data have been collected form NVD and used to develop and test the various models. The data collected are divided into two datasets one for training datasets and other for testing datasets. The training datasets are used to developing the models. The same training datasets are used to develop four models.

Fuzzy Inference System (FIS) with different optimization techniques have been used to develop our models. Firstly we started by developing five models using ANFIS with hybrid optimization technique. Then we apply the same datasets to develop five models using ANFIS with Subtractive Clustering and cascaded model using the subtractive clustering with hybrid optimization technique. We also apply a Mamdani fuzzy inference system to compare all models with each other's.

The adequacy of the developed models has been checked using the Correlation Coefficient (CC) to measure the agreements between the actual and predicted risk values. In addition three error measures were used namely, Mean Absolute Performance Error (MAPE), the Root Mean Square Error (RMSE), and the Percentage of Differences to indicate the accuracy and the performance of developed models.

While testing these models using the testing datasets that has been isolated before the training stage using the developed cross validation algorithm, the obtained CC between the actual and predicted values of risk for all developed models ranges between 0.91 and 0.993. The corresponding MAPE that ranges between 0.003546 and 0.080646 and RMSE that ranges between 0.011032 and

0.042078 and the Percentage of differences that ranges between 0.35461 and 8.064565. This demonstrate the adequacy of adopting these types of approaches to IA risk levels.

It was noticed that the performance of the developed models has been improved by grid partition with hybrid optimization technique. This improvement is notices as an improvement in the obtained CC that results from the hybrid models which equal to 0.993 and compared to the CC of other models that ranges from 0.91 to 0.992 when developing the models using other optimization techniques. Also the hybrid developed model has the minimum value error of MAPE, RMSE and Percentage of differences than the other models that developing with other optimization techniques.

7.2 Suggestions and Future Work

Although we have obtained preliminary results, yet still the following recommendation proposes further contributions to researchers:

1. Further investigation using more historical data and updated parameters of the models need to be performed to conclude the adequacy of these approaches.
2. Build a stand-alone application that automatically discover the critical assets in the organization and collect all historical data for these assets from an online NVD by middleware interface that connect the stand-alone application with the NVD.
3. Additional enhancement for this model is to develop this model to reduce the IA risk level for the critical asset.
4. Trying to find more parameter that affect the IA level and applying those parameters in our models to find the adequacy of our approaches.
5. Finally, it is worthy to explore the use of other different soft computing modeling techniques such as Genetic Algorithm.

Bibliography

- Adebiyi, A., J. Arreymbi, et al. (2013). "Security Assessment of Software Design using Neural Network." arXiv preprint arXiv:1303.2017.
- Al-Mahrouqi, A., P. Tobin, et al. (2015). Simulating SQL-Injection Cyber-attacks using GNS3. The 6th International Conference on Computer Modeling and Simulation (ICCMS 2015), Amsterdam Netherlands, 12-13 February 2015.
- Ali, A., P. Zavarsky, et al. (2011). "A new CVSS-based tool to mitigate the effects of software vulnerabilities." International Journal for Information Security Research (IJISR) **1**(4): 178-182.
- Arafeh, L., H. Singh, et al. (1999). "A neuro fuzzy logic approach to material processing." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **29**(3): 362-370.
- Bailey, M., E. Cooke, et al. (2005). "The blaster worm: Then and now." Security & Privacy, IEEE **3**(4): 26-31.
- Blyth, A. and G. L. Kovacich (2001). Information assurance: surviving in the information environment, Springer-Verlag New York, Inc.
- Bouchon-Meunier, B., M. Dotoli, et al. (1996). "On the choice of membership functions in a mamdani-type fuzzy controller."
- Boyce, J. and D. Jennings (2002). Information assurance: managing organizational IT security risks, Butterworth-Heinemann.
- Chen, J. and Z. Tian (2015). Statistical Analysis of The Privilege Level of Vulnerability. 2015 International Symposium on Computers & Informatics, Atlantis Press.
- Cherdantseva, Y. and J. Hilton (2013). "Information Security and Information Assurance: Discussion about the Meaning." Organizational, Legal, and Technological Dimensions of Information System Administration: 167.
- Chiu, S. (1996). Method and software for extracting fuzzy classification rules by subtractive clustering. Fuzzy Information Processing Society, 1996. NAFIPS., 1996 Biennial Conference of the North American, IEEE.
- Cho, Y.-C. (2015). "Implementation and analysis of website security mining system, applied to universities' academic networks." Tehnički vjesnik **22**(2): 279-287.
- CNN (2003). "SoBig.F breaks virus speed records."
- CSIA (2007). A National Information Assurance Strategy. Crown.

DOD (2002). Directive Number 8500.01E, U.S. Department of Defense.

Dondo, M. G. (2008). A vulnerability prioritization system using a fuzzy risk analysis approach. Proceedings of The Ifip Tc 11 23rd International Information Security Conference, Springer.

Gallegos, F. and M. Smith (2006). "Red teams: An audit tool, technique and methodology for information assurance." Information Systems Control Journal **2**: 51.

Hawkins, D. M. (2004). "The problem of overfitting." Journal of chemical information and computer sciences **44**(1): 1-12.

Houmb, S. H. and V. N. Franqueira (2009). Estimating ToE risk level using CVSS. Availability, Reliability and Security, 2009. ARES'09. International Conference on, IEEE.

Houmb, S. H., V. Nunes Leal Franqueira, et al. (2008). "Estimating impact and frequency of risks to safety and mission critical systems using CVSS."

IATF (1999). "Information Assurance Technical Framework (IATF)." U.S. National Security Agency (NSA)(Rel 2.0.1. Ft. Meade).

Jang, J.-S. (1993). "ANFIS: adaptive-network-based fuzzy inference system." Systems, Man and Cybernetics, IEEE Transactions on **23**(3): 665-685.

Jang, J.-S. and E. Mizutani (1996). Levenberg-Marquardt method for ANFIS learning. Fuzzy Information Processing Society, 1996. NAFIPS., 1996 Biennial Conference of the North American, IEEE.

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**.

Luftman, J. and T. Ben-Zvi (2010). "Key issues for IT executives 2009: Difficult economy's impact on IT." MIS Quarterly Executive **9**(1): 203-213.

Macedo, F. and M. M. da Silva "Comparative Study of Information Security Risk Assessment Models."

Maconachy, W. V., C. D. Schou, et al. (2001). "integrated approach to information assurance modeling."

Mamdani, E. H. and S. Assilian (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." International journal of man-machine studies **7**(1): 1-13.

Martin, R. A. (2001). "Managing vulnerabilities in networked systems." Computer **34**(11): 32-38.

- MathWorks (2008). Fuzzy Logic Toolbox User's Guide, Inc.
- Matthews, S. W. (2004). "Assessing Information Assurance Posture: Key Steps to IA Assessment Methodology."
- McCumber, J. (1991). Information systems security: A comprehensive model. Proceedings of the 14th National Computer Security Conference.
- Mell, P., K. Scarfone, et al. "A complete guide to the common vulnerability scoring system version 2.0 (2007)." Διαθέσιμο On-line στο: <http://www.first.org/cvss/cvss-guide.pdf> [Accessed 20 March 2012].
- Mell, P., K. Scarfone, et al. (2006). "Common vulnerability scoring system." Security & Privacy, IEEE **4**(6): 85-89.
- Mell, P., K. Scarfone, et al. (2007). A complete guide to the common vulnerability scoring system (CVSS), version 2.0, forum of incident response and security teams.
- Microsoft. "Security Bulletin Severity Rating System." from <https://technet.microsoft.com/en-us/security/gg309177.aspx>.
- Moore, D., V. Paxson, et al. (2003). "Inside the slammer worm." IEEE Security & Privacy **1**(4): 33-39.
- NVD (2004). "National Vulnerability Database." from <https://nvd.nist.gov/>.
- Okereke, G. and C. Osuagwu (2012). "Security metrics model for web page vulnerability classification and ranking." African journal of Computing and ICT, IEEE Nigeria **5**(5).
- Peng Liu, P. S. U., University Park, M. U. Meng Yu, et al. (2001). "Information assurance." BT technology journal **19**(3): 107-114.
- Prasad, T. R. (2014). "Network Intrusion Detection Systems Using Genetic Algorithm." IJSEAT **2**(3): 107-111.
- Pub, J. (1998). "Pub 3-13." Joint Doctrine for Information Operations **9**: 1-9.
- PWC (2014). "Global Information Security Survey: 2015 Results by Industry." 2015, from <http://www.pwc.com/gx/en/consulting-services/information-security-survey/>.
- Rae'd Basbous, D. and L. Arafeh "Electric Power Load Short Term Forecasting."
- Rani, D. S. S. (2013). "Neuro-Fuzzy based Software Risk Estimation Tool." Global Journal of Computer Science and Technology **13**(6).

Rushby, J. (1994). "Critical system properties: Survey and taxonomy." Reliability Engineering & System Safety **43**(2): 189-219.

Secunia (2014). Key figures and facts from a global IT-Security perspective. Denmark. **20**.

Sendi, A. S., M. Jabbarifar, et al. (2010). FEMRA: Fuzzy expert model for risk assessment. Internet Monitoring and Protection (ICIMP), 2010 Fifth International Conference on, IEEE.

Shameli-Sendi, A., M. Shajari, et al. (2012). "Fuzzy multi-criteria decision-making for information security risk assessment." The Open Cybernetics and Systemics Journal **6**: 26-37.

Sherwood, N. A. (2005). Enterprise security architecture: a business-driven approach, CRC Press.

Tajuddin, S., W. Olphert, et al. (2015). Relationship between stakeholders' information value perception and information security behaviour. INTERNATIONAL CONFERENCE ON INTEGRATED INFORMATION (IC-ININFO 2014): Proceedings of the 4th International Conference on Integrated Information, AIP Publishing.

Tawileh, A. and S. McIntosh (2012). "Understanding Information Assurance: A Soft System Approach."

Telang, R. and S. Wattal (2005). "Impact of software vulnerability announcements on the market value of software vendors-an empirical investigation." Available at SSRN 677427.

Yu, Q. H., G. Y. Liang, et al. (2010). "Risk Scoring Method on Business Information Management System."

Zadeh, L. A. (1973). "Outline of a new approach to the analysis of complex systems and decision processes." Systems, Man and Cybernetics, IEEE Transactions on(1): 28-44.

Appendix A: Information Assurance Definition from Different Perspectives

Reference	IA definition	Organization Type	Security Goals
(McCumber 1991)	Presented a model of IA with confidentiality, integrity, availability (CIA) triad.		confidentiality, integrity, availability
(Pub 1998)	Five Pillars of IA: availability, integrity, authentication, confidentiality and non-repudiation		availability, integrity, authentication, confidentiality and non-repudiation
(DOD 2002)	Through a defense-in-depth approach that integrates the capabilities of personnel, operations, and technology, and supports the evolution to network centric warfare	Government	
(Maconachy, Schou et al. 2001)	Replace the CIA triad with five pillars		availability, integrity, authentication, confidentiality and non-repudiation
(Boyce and Jennings 2002)	"Information Assurance is one of the newly refined processes of information protection	Academic	

Reference	IA definition	Organization Type	Security Goals
	that has evolved from computer security and information system security.		
(Sherwood 2005)	Information Assurance is a discipline the main aim of which is to give confidence or certainty in information; to give belief that one can rely on data, knowledge, facts, and its meaning.		
(CSIA 2007)	Information Assurance is the term given to management of risk to information. Effective IA ensures that the opportunities provided by new technology can be exploited to maximum benefit.	Military	Confidentiality, Integrity, Availability, Non-repudiation, Accountability, Possession, Utility, Authenticity, Auditability, Transparency, Cost-effectiveness, Efficiency
(Peng Liu, Meng Yu et al. 2001)	Information operations that protect and defend information and Information systems by ensuring their availability, integrity,		availability, integrity, authentication, confidentiality and non-repudiation

Reference	IA definition	Organization Type	Security Goals
	<p>authentication, confidentiality, and non-repudiation. This includes providing for restoration of information systems by Incorporating protection, detection, and reaction capabilities.</p>		
(Tawileh and McIntosh 2012)	<p>a system to prevent the improper user, corruption or theft of the business's information and information systems by internal users or employees through the derivation and implementation of appropriate policy, technical and disciplinary measures and trust mechanisms.</p>	Business	<p>availability, integrity, authentication, confidentiality and non-repudiation</p>
(Tawileh and McIntosh 2012)	<p>A system to reap benefits through the interception, theft, misuse, corruption or</p>	Hacker	

Reference	IA definition	Organization Type	Security Goals
	<p>manipulation of information stored in, transmitted through and processed by information systems by exploiting weaknesses and vulnerabilities in these systems while avoiding traceability and chances of being caught.</p>		
<p>(Tawileh and McIntosh 2012)</p>	<p>A system to protect the integrity and privacy of private information stored in, transmitted through and processed by information system in order to avoid negative consequences and legal liability by implementing appropriate measures and practices.</p>	<p>End user</p>	<p>Integrity, privacy</p>
<p>(Tawileh and McIntosh 2012)</p>	<p>A system to assure information systems users that these systems will function as expected with regards to the protection of the</p>	<p>Information system developers</p>	

Reference	IA definition	Organization Type	Security Goals
	<p>information stored, transmitted and processed</p> <p>by these systems by providing acceptable</p> <p>evidence that the system are built by</p> <p>knowledgeable people using sound</p> <p>development processes in addition to testing</p> <p>results that confirm the claims made.</p>		