CZECH UNIVERSITY OF LIFE SCIENCES

PRAGUE

Faculty of Economics and Management

Informatics

Department of Information Engineering

Diploma Thesis

USING FUZZY LOGIC IN THE EVALUATION OF USER PERCEPTION OF SECURITY RISK ON SOCIAL NETWORKING SITES

Author: Eric Afful-Dadzie

Supervisor: Doc. Ing. Arnošt Veselý, CSc.

VYUŽITÍ FUZZY LOGIKY K VYHODNOCENÍ UŽIVATELSKÉHO VNÍMÁNÍ
BEZPEČNOSTNÍCH RIZIK NA SOCIÁLNÍCH SÍTÍCH

USING FUZZY LOGIC IN THE EVALUATION OF USER PERCEPTION OF SECURITY RISK ON SOCIAL NETWORKING SITES

ABSTRAKT

Lidská povaha se často zapojuje nebo vzájemně působit na blizká cizi, ale online sociální sdělovací prostředky sití převazně opomíjeny. Je to nejaká otevřena interakce mezi dvěma znamými a neuvázanými spojenými uzivateli na sociální sdělovací prostředky sití. Výsledek toho, je že sociální normální překážky proti interakce s cizimi jsou nizké. Tato nedbalá otevřenost spíše vedla k přebujelému zvyšovaní světové cyber-zlocinnosti a odcizení totožnosti, očekávaní potenciálního soukromího neštěstí v blizké budoucnosti jestli nic není držet na uzdě. Tradicně, bezpečnost by byla považovana za subjektivní problematika kvůli vysoké úrovni nejistoty svých popisů a parametrů. Bezpečnost je zejména nejasná protože není to možné přesně určit hranice mezi co je bezpečné a co není bezpečné. Takže je to subjektivní problém ve vztahu ke individualnim povážovaní.

Zastřená(rozčepýřená) logic metoda je forma multihodnoceni logic metody odvozena z pojmu zastřený soubor teorie. Jeji metodika slouží k poskytnuti konečného rešení z informace, která může být chápat jako rozporulpná, nepřesná nebo hlučná jako je technologie bezpečnostní informace.

Tento projekt představuje techniku k hodnocení uzivatelského vnimaní o bezpečnostním rizku na sociální sdělovací prostředky sití používaní zastřene(rozčepýřené) logic metody . Vstupy do systému byly vhodné zastřené soubory, které zastoupi jazykové proměnné pro hodnocení cílů informační bezpečnosti o důvernosti, celistvosti a dostupnosti. IF-THEN pravidla byla sestavena používaním Mamdani zastřené vysvetlení techniky aby kvalitně provedl rozbor vstupy a také aby defuzzifikace technika byla udělana využitím techniky centroidu. Implementace designu byla provedena použitím MATLAB Fuzzy logic tool box. Výsledky tří používanějších sociálních sdělovacích prostředků sití Facebook, Twitter and LinkedIn ukazují systém, který muže být efektivně využití k hodnocení uživatelského vnímaní bezpečnostní informace.

Klicova slova: Zastřená(rozčepýřená) logic metoda, sociální sdělovací prostředky sití, fuzzificace, defuzzificace, fuzzy inference, bezpečnostní riziko a vázený průměr.

ABSTRACT

Human nature often frowns on engaging or interacting with near strangers but on online social media networks, this is largely ignored. There is an open interaction among both known users and loosely-connected users on social media networks, and as a result, the normal social barriers against interacting with strangers are lowered. This rather careless openness has resulted in rampant increase in cybercrime and identity theft worldwide, awaiting a potential privacy disaster in the near future if not curbed. Traditionally, security would be considered a subjective issue because of the high level of uncertainty with its descriptors and parameters. Security is vague mainly because it is improbable to define exact or sharp boundaries between what is secure and what is not, making it a subjective problem relative to the individual considering it.

Fuzzy logic is a form of multi-valued logic derived from the concept of fuzzy set theory. Its methodology aims at providing a definitive solution from information that may be construed as ambiguous, imprecise or noisy such as information technology security.

This project presents a technique for evaluating user perception of security risk on social networking sites using fuzzy logic. The inputs to the system were suitable fuzzy sets representing linguistic variables for information security evaluation goals of confidentiality, integrity and availability. The IF-THEN rules were constructed using the Mamdani fuzzy reasoning technique in order to adequately analyze the inputs and the defuzzification technique was done using the centroid technique. The implementation of the design was done using the MATLAB Fuzzy logic tool box. Using three of the popular online social networking sites namely, Facebook, Twitter and LinkedIn the results show a system that can be effectively employed to evaluate user perception of Information Security.

Keywords: Fuzzy Logic, Social Networking Sites (SNSs), fuzzification, defuzzification, fuzzy inference, security risk and weighted average.

DECLARATION

I certify that this diploma thesis, which I submit in partial fulfillment of the requirements of the MSc. Systems Engineering and Informatics Programme of the Czech University of Life Sciences, is entirely my own work and that any content that relates to the work of other individuals, published or otherwise, are acknowledged through appropriate referencing.

I also confirm that this work has not been submitted for assessment in whole or part for an award in any other University.

Signed: __________________________

Date: _____________________________

DEDICATION

To Naana Mensima, my best friend and wife,

for her immeasurable sacrifices,

and for all the good things that she

has brought into my life.

ACKNOWLEDGEMENTS

I would like to thank all individuals who one way or the other provided me with the valuable assistance, insight and time required in completing this study especially Lecturers of CULS- Prague.

I am extremely grateful for the support of my supervisor, Doc. Ing. Arnošt Veselý, CSc who spent valuable times with me to see to the completion of this work. Through his in-depth knowledge on the subject, I have come to adequately understand the concepts and principles of fuzzy logic. I gratefully appreciate the opportunity to have worked on such a fascinating subject matter within the field of Computational Intelligence.

I also thank the support from my wife, Naana and my son Jason. It is also to my mother, Mary Afful and all my siblings.

I finally dedicate this work to the memory of Prof. RNDr. Jiří Vaníček, CSc .

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

LIST OF APPENDICES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1. Introduction

Social Networking sites (SNSs) have come to stay and are now an integral part of our lives. In recent years, participation in social networking sites has dramatically increased. Online social media services such as Facebook, Twitter, and LinkedIn allow millions of people to create online profiles and share personal information with vast networks of friends and sometimes unknowingly with strangers.

In recent years whiles popularity is soaring for these SNSs and millions of users sign onto these sites on a daily basis, there have also been growing concerns about breach of privacy and identity theft on these sites. Privacy issues and identity theft in social media are a huge concern. This phenomenon is attracting the attention of academic and industry researchers who are intrigued not just by the affordances and wide reach of audiences for these social networking sites but the increasing concerns of security risks posed to users.

1.1 Overview of the Problem

As of June 2010, 22 percent of all time online or one in every four and half minutes spent online is social, i.e., sharing, messaging, commenting, and blogging [1]. It is also interesting to note that for the first time ever, social networks or micro-blogging sites are visited by three quarters of global consumers who go on-line [2]. Brazil leads the world chat with the highest percentage (86%) of internet consumers visiting a social networking site and in the U.S. the total minutes spent on social networking sites has increased eighty-three percent year-over-year [2]. Facebook alone as one of the major players of SNSs has over 800 million active users with each active user linked to an average of 130 other users making it the second most visited website on the Internet [3, 4].

However, interaction with strangers is naturally shunned by humans but on social networks, this is encouraged. There is an open interaction among both known users and loosely-connected users on SNSs, and as a result, the normal social barriers against interacting with strangers are lowered. This rather careless openness has resulted in the rampant increase in cyber-crime and identity theft worldwide, [5] awaiting a serious security and privacy disaster in the near future if not curbed.

In recent years, fuzzy logic has been found useful for handling uncertainty and subjectivities in data by evaluating certain range of vague variables using quantitative data with qualitative information. It has been found an effective alternative for measuring operational risks [6]. One advantage of fuzzy approach for evaluation or in measurement is its ability to process vaguely defined variables, and those variables whose relationships cannot be defined by mathematical relationships [7] by employing fuzzy IF-THEN rules to define those variables and their relationships. Security is conventionally considered a subjective issue because of the high level of uncertainty with its descriptors and parameters. Security is vague mainly because it is improbable to define exact or sharp boundaries between what is secure and what is not, making it a subjective problem relative to the individual considering it.

There have been several approaches to fuzzy evaluation as used in a number of different situations such as quality of service, educational measurement, management, e-commerce trust, information systems in public administration, homeland security among others. In risks management, fuzzy logic has been successfully applied to network vulnerability ranking system and operational risk evaluation [8, 9]. In the evaluation of security risks, a number of works have also been produced such as [10, 11] where a fuzzy logic approach based on Mamdani-inference style was used to detect potential threats to computer systems and evaluate security risks in software respectively and multi-criteria security system performance assessment using fuzzy logic [12]. There is also the cognitive fuzzy approach applied to analyzing risks in business information systems [13] and in enhancing risks assessment in health institutions [14].

So far, reviewing related literature has shown that few people have tried to apply fuzzy logic to the area of information technology security. This thesis develops a method of evaluating security risks on social networking sites using fuzzy set theory. This method could prove to be a good technique for this research area.

1.2 Objectives of study

The main objective of this research is to evaluate user perception of security risks on Social Networking Sites (SNSs) by the use of fuzzy logic. The research also seeks among other things the following sub-objectives:

- To determine which of the fuzzy methods can be applied for analyzing user perception of social networking sites security risk based on the review of user perceptions and fuzzy set theory.
- To formulate a general methodology for employing fuzzy sets to evaluate user perception of security risk in information technology applications.
- To apply the general methodology to three social networking sites (Facebook, Twitter, and LinkedIn) as application case studies and to understand the limitations of the developed methodology.

1.3 Methodology

The primary objective of this research is to evaluate user perception of security on social media or networking sites. The first step was to thoroughly review all available literature in the area so as to identify the best criteria to evaluate user perception of security risks on SNSs.

Based on the evaluation criteria, the author identified the factors (i.e input variables) which without them make social networking sites vulnerable or susceptible to security breaches and which are mainly used to measure or evaluate Information Security and used the inputs to determine the level of security on SNSs (output).

A well-constructed security related questions about SNSs sites mainly about Facebook, Twitter and LinkedIn were put online for users to give inputs. This helped in determining the membership functions for each linguistic variable with the extent of the range of responses from the questions. The fuzzy weighted average was used to find the overall user perception. MATLAB fuzzy tool kit was used for the implementation. The linguistic inputs to the system were supplied through the graphical user interface called the rule viewer. The Mamdani fuzzy reasoning was used to construct the fuzzy system so as to efficiently analyze the inputs. The centroid technique, also known as the center of gravity was also employed for the defuzzication technique after the inference step to achieve a final crisp output for the study.

1.4 Expected Contribution

Fuzzy set theory is a branch of artificial intelligence that has been applied successfully in many fields. This thesis is expected to make the following contributions:

- Proposing a fuzzy method that mirrors the decision-making process used in estimating security risks on social networking sites.
- Providing a reasonable framework, based on sound techniques of fuzzy set theory, that can be modified to apply to the many scenarios in information technology security.
- Defining data that needs to be collected in order to adequately provide the basis for evaluating security risks on online social media sites.

1.5 Thesis organization

Chapter two focuses on a literature review by introducing *fuzzy* set theory and its applications in the evaluation and measurement of security in the information technology industry. Security risks on social networking sites and the factors that make social networking sites vulnerable are discussed. The various models that have been used in the evaluation of security in information technology related issues so far are also reviewed.

Chapter three describes the design methodology and implementation of the fuzzy approach to evaluating security risks on social media.

In chapter four, a fuzzy inference engine is built using MATLAB fuzzy toolbox to combine the relevant fuzzy rules to infer level of security risk (output). This stage of the study involves approximate reasoning with several conditional fuzzy propositions.

Finally, defuzzification methods are discussed. The purpose of defuzzification is to obtain a crisp value for the level of security risk on social networking sites, derived from the fuzzy output. The level of security risk can then be estimated, based on the recommendations given by the model and the inputs fed into the MATLAB rule viewer.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

There is no doubt that SNS offer a new range of opportunities for communication and real-time exchange of all kinds of information, but in recent times, privacy and security have emerged as critical issues for concern [15]. In most of the SNSs, there is very little protection against copying of personal data from profiles and re-publishing the data elsewhere [16, 17] but one of the most important challenges of information sharing is how to assure its security [18]. So the question has always been how much of security is enough to safeguard personal data without compromising on how people interact and use social networking sites.

It is obviously not realistic to join a network of millions of people worldwide expecting to have the trust of all of them [19]. The question that arises therefore is that, does trust play any role in social networking given the surge in the number of people joining SNSs and of whom most of them are either ignorant of the security implications or rather trust that their data is kept confidential.

Trust as defined by [20] is the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trust or, irrespective of the ability to monitor or control that other party". When interacting face to face, trust is seen as a critical determinant of the willingness to sharing information and developing new relationships [21, 22]. Trust is also identified as an important factor for successful online interactions.

As millions of people join social networking sites, sharing, messaging, blogging and revealing their personal information, trust alone cannot be enough to protect them from potential predators. One solution through this study is to let users define their own perception of security risk so as to be incorporated into the design and maintenance of these SNSs sites.

In recent time, the reputation of social networking sites has been hit by a number of incidents as reported by the news media [23, 24]. It is therefore incumbent on SNSs to have clear policies regarding data protection so as to deliver the same level of social privacy that exists face to face. The concern is that, would legislation addressing privacy and information security safeguards for these sites affect how people interact and use social networking sites?

2.2 Information Security

Security may be defined as the state of being free from danger and not exposed to damage from accidents or attack, or defined as the process for achieving that desirable state. Good security assures that all information systems remain fully operational, robust, and accurate and that all data remain private and cannot be compromised [25].

Information security may therefore be defined as the process of keeping data safe and secure from the reach of unauthorized people or users. It must be ensured that as much as possible, data and information are not at all visible and disclosed to anyone. The goal of any information system security is to protect the integrity, confidentiality, and availability of information processed by the system as prescribed by many international information security bodies such as ITSEC, OECD, US Department of Defense etc. This goal is reached using identification, authentication, and authorization. Identification is a prerequisite, where each user is required to submit an identifier (ID) that is included in the authorization lists of the system to be accessed. *Authentication* is means of proving that the user is really the person to whom the ID has been assigned. *Authorization* consists of defining what a specific user ID, running specified programs, can legally do on the system. The security perimeter can be penetrated by compromising any of these functions [25].

2.2.1 Information Security Evaluation Criteria

Almost throughout the world, what has become the widely accepted model or criteria for evaluating information security is the basic CIA triad; standing for Confidentiality, Integrity and Availability. These three key criteria principles are deemed fundamental to guaranteeing security in any information system. These criteria have been applied across the whole subject of Security Analysis, from access to a user's internet history to security of encrypted data across the internet [26]. Therefore the universal classic definition of information security is brief and very simple: Information security is the confidentiality, integrity, and availability of information [27]. By extension, if any one of the three principles are violated or breached, it can have serious consequences for the parties concerned be it an organization or the individual user of an information system.

The Information Technology Security Evaluation Criteria (ITSEC), a consortium of Information Security experts from France, Germany, Holland and the United Kingdom, also employ confidentiality, integrity and availability as the yardstick for evaluation of Information Technology security [28]. The relationship among these factors however, has much ambiguity such that it is reasonable and scientific to apply fuzzy comprehensive evaluation method for evaluating security risk in an information technology system such as an online social networking site.

2.3 Confidentiality, Integrity, Availability (CIA)

The term CIA always quickly brings into mind the Central Intelligence Agency but for information security specialists, CIA would mean Confidentiality, Integrity and Availability of Information- the widely accepted benchmark for the evaluation of information systems security.

2.3.1 Data Confidentiality

According to the International Organization for Standardization (ISO) in ISO-17799 as cited by [29], confidentiality is "ensuring that information is accessible only to those authorized to have access” and is one of the key pillars of information security. Confidentiality is one of the design goals for many cryptosystems, made possible in practice by the techniques of modern cryptography.

In other words, confidentiality can be described as the act of limiting information access and disclosure to authorized users and preventing access by or disclosure to unauthorized users. Underpinning the goal of confidentiality are authentication methods like user-IDs and passwords that uniquely identify a data system's users, and supporting control methods that limit each identified user's access to the data system's resources [29].

2.3.2 Data Integrity

Largely, data integrity as one of the key criterion of information security evaluation is described as the trustworthiness of information resources. The concept is used to insist on the fact that data should not be changed inappropriately, whether by accident or deliberately malign activity. The concept also looks at where data originates from or the integrity of the source. That is to question whether data is actually coming from the person or entity you think it should come from, rather than an imposter. On a more restrictive view, however, integrity of an information system includes only preservation without corruption of whatever was transmitted or entered into the system, right or wrong.

2.3.3 Data Availability

Availability, as the word seems to imply, means making available requested information resources in time. An information system that is not available when you need it is at least as bad as none at all. It may be much worse, depending on how reliant the organization has become on a functioning computer and communications infrastructure. In today's world, we are witnessing a dramatic rise in popularity of online social networking services so much so that, peoples jobs, relationships, and even to a large extent lives depend on how available the sites are all year round. [30] emphasizes the point that public data such as social networking sites has to be always available and in real-time.

Also according to Microsoft [31], almost all modern organizations are highly dependent on functioning information systems and that many literally could not operate without them. Microsoft explains further that, one cannot however assume that data availability means having your data accessible and obtainable at all times. In the enterprise environment for example, there are quite a few factors that are considered when the issue of data availability comes up for discussion. These factors according to Microsoft TechNet include:

- Available bandwidth between devices and network connections of mediums
- Mechanisms for high availability and their own security and accessibility
- Prioritization and type of data to be made available

- Recovery roles and responsibilities
- Type of file system and level of access
- Type of storage/retrieval device or media including both hardware and software
- Service Level Agreements between responsible and affected entities
- Processing overhead of affected mechanisms
- Disaster Recovery/Business Resumption Plan (BRP) [31]

Availability may be affected by purely technical issues like the aspects of security. For example malfunctioning part of a computer or communications device, power failure, natural phenomena such as wind or water or human causes (accidental or deliberate) could all one way or the other affect data availability. But irrespective of the reasons above, it is incumbent on the organization providing the services to ensure that, the public data is accessible as when needed and in real-time.

2.4 Social Networking Sites Security

According to [32] an online social network site is a: web-based service that allows individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. Other names or phrases used to characterize such services are *social digital technologies* [33], *participatory media* [34] and *social media* [35]. Whereas this term “social network” site seems to reflect the fact that these sites represent existing social bonds, another term commonly used, “social networking” implies that people use these websites in order to forge new networks.

2.4.1 Social Media Security Concerns

Whiles there have been various attempts at producing high quality and very secure software systems, attackers still have their way round it by frequently breaking into these systems. Most of the recent works on risks of online social networks have had their focus on privacy concerns. According [36], many attacks on social networks have tried to exploit the trust that exists between users which tend to have users likely to click on fake links or fall prey to social engineering schemes. A case study by [37] as cited in the September 2011 issue of IEEE computer Society magazine titled “Security and Privacy in an Online World”, where they used a malicious app with dog pictures on Facebook. This app was able to phish out information such as users IP addresses, their browser versions, open ports and the version of their operating systems. The app though focused on users, it is claimed that it could have phished for lists of users’ friends and then send messages to them [37]. Elias et al looked at ways to turn social networks into a botnet through again a malicious Facebook app [38]. Andrew et al were confident that with powerful facebook apps, even when users do not consent to accessing their profile, the apps can still request such information from a user’s friend who mistakenly installed the app.[39]

In response to the numerous threats, software vendors over the last decade started incorporating security as a necessary feature for their products. Software security has always been critical consideration to information assurance and design-level vulnerabilities [40]. Previous research works in this area have presented or suggested different security measures and processes towards producing secure software. For example in [41], threat modeling was described as the basis for security requirements. The work explains that threat modeling as a process, consists of three high level steps: characterizing the system, identifying assets and access points, and identifying threats. The risks associated with the identified threats were then assessed. Another tool which has been found useful for identifying security risk is the attack trees. Attack trees are used to model a chosen set of attack via a finite state machine. Attack trees is used to model the decision making process of the attackers [42]. Attacks against a system are represented in a tree structure. The root of the tree represents the potential goal of an attacker (for example, to steal a credit card number). The nodes in the tree represent the actions the attacker can take and each path in the tree represents a unique attack to achieve the goal of the attacker. Attack trees can be used to answer question such as, what is the easiest attack, the cheapest attack, the attack that causes the most damage and so on. Attack trees are used for risk analysis, capturing of security knowledge in a reusable way and implementing counter-measures to attacks. However, the use of attack trees cannot replace the threat modeling process [43].

Over the years, fuzzy logic approach has been used for the evaluation of risk in different situations. A fuzzy logic technique based on the Mamdani-style inference engine was employed to identify the potential threats to computer-based systems [44]. The result showed an efficient way of undertaking threat modeling. Another work that focuses on fuzzy modeling was presented in [45]. A cognitive fuzzy modeling technique was designed for enhanced risk assessment in a health care institution. A paper which looks very similar was presented in [46]. The work outlined a methodology for assessing and analyzing risks in business information systems. It is worthy to note that in the last two papers, however, the risk assessment were carried out by considering what would happen if a particular decision was taken or if some information were lost in the process or considered unauthentic. When a fuzzy cognitive map is used, a clearer picture of the different phases at which risk can incur is seen. In this work, a rule based fuzzy logic system is used to analyze user’s perception of security risk in social networks.

2.4.2 Fuzzy Risk Evaluation Research

There have been several techniques of risk analysis used to help in managing uncertainty. Thorough risk analysis estimation and evaluation have provided valuable support for decision making. Currently, there are many risk analysis techniques in use that attempt to evaluate and estimate risk. These techniques are mainly qualitative or quantitative depending on the information available and the level of detail that is required [47]. The use of quantitative techniques rely heavily on statistical approaches, which include Monte Carlo Simulation [48], Fault and Event Tree Analysis [4,48], Sensitivity Analysis [48], Annual Loss Expectancy [49], Risk Exposure [50], Failure Mode and Effects Analysis [48], etc. On the other hand, qualitative techniques rely more on judgment than on statistical calculations such as Scenario Analysis [49], Fuzzy set theory [49], etc. Quantitative and qualitative techniques have their own advantages and disadvantages. Among these techniques, the application of Fuzzy set theory to risk evaluation seems appropriate; as such analysis is highly subjective and related to inexact and vague information. Since the introduction of Fuzzy set theory by Zadeh [51] to deal with problems in which vagueness was present, linguistic values have been widely used to approximate reasoning. Numerous studies of Fuzzy set theory in risk assessment have appeared in different areas, and some are summarized in **Table 2.1**. Fuzzy set theory has been effectively applied in such a variety of areas because it can handle inexact yet useful information.

Research Area	Description	References
Information Technology		
Information security	Presents a methodology for the modeling of the risk analysis process within a computing facility	[52]
Database gateway processor	Applies basic concepts of fuzzy logic modeling to risk analysis in database gateway systems	[53]
Software development	Applies Fuzzy set theory to evaluate the rate of aggregative risk in software development.	[54]
Engineering		
System Failure	Presents a fuzzy logic-based technique for prioritizing failures for corrective actions in a Failure Mode, Effects and Criticality Analysis. The method allows the analyst to evaluate the risks associated with item failure modes directly by using the linguistic terms employed in the criticality assessment. Ambiguous, qualitative, or imprecise information, as well as quantitative data, can be used in the assessment.	[55]
Construction	Outlines an approach to the assessment of construction project risk by linguistic analysis using Fuzzy Set Theory.	[56, 57]
Civil	Involves fuzzy set representations of structural damage and related safety analyzes in civil engineering	[58, 59, 60]
Environmental		
Natural Hazards	Employs fuzzy methods to calculate the risk of release, exposure to, and consequence of natural urban hazards.	[61]
Hazardous materials	Provides an application of fuzzy logic to the risk assessment of the transport of	[62]

	hazardous materials by road and pipeline to evaluate the uncertainties that affect both individual and societal risk estimates.	
Ground water nitrate risk management	A nitrate risk-management methodology using fuzzy sets in combination with a multi-criterion decision-making (MCDM) technique to assist decision makers in evaluating, with uncertain information, possible regulatory actions along with the various nitrate risk-management strategies to determine an appropriate strategy.	[63]
User Perception		
Health Care	Fuzzy logic method in the evaluation of customer perceived value on healthcare services.	[64]
Dweller perception	Introduced an uncertainty measure used to identify the strengths and weaknesses of slum upgrading projects from the using the dwellers' perception.	[65]
Transportation	A generalized approach for analyzing transportation User perception using fuzzy sets	[66]
Others		
Bank	Develops a fuzzy set approach in planning system for liquidity management in bank industry	[67]
Tourism	Applies a fuzzy multiple criteria decision-making method to conduct an evaluation of tourist risks.	[68]

Table 2.1: Applications of Fuzzy risk evaluation analysis

2.5 Fuzzy Set Theory Applications in Security Evaluation

The Fuzzy set theory approach, pioneered by Zadeh [69] was intended to be specifically used to deal with the issue of uncertainties that are not statistical in nature. The fuzzy set approach has been widely used to represent the uncertainties of real-life situations. The decade has witnessed a rapid growth in the number and variety of applications using fuzzy set theory. In the field of computer security, the applications in various areas of security have not been left out. For example, fuzzy set theory was employed by [70] for *Intelligent Quality Performance Assessment for E-Banking Security. The research dwelled on the* complex and dynamic nature of the many factors that are considered in E-banking security assessment and also how subjective and ambiguous the assessment of e-banking websites can be when considered. They were convinced that fuzzy logic (FI) model presents an effective tool in assessing and evaluating e-banking security performance and quality. [71] applied fuzzy set theory to assess online risk for distributed intrusion prediction and prevention systems. The research illustrated how the design of fuzzy logic based Distributed Intrusion Prediction and Prevention Systems (DIPPS) using DIPPS sensors, can be used to effectively assess online risk. Hierarchical Takagi-Sugeno Models is also used for Online Security Evaluation Systems [72] where the risk assessment problem was carried out using an evolutionary algorithm to automatically design a Hierarchical Takagi-Sugeno fuzzy inference system. The hierarchical structure is evolved using Probabilistic Incremental Program Evolution (PIPE) with specific instructions. The fine tuning of the *if-then* rule's parameters encoded in the structure was accomplished using Evolutionary Programming (EP). Authors [73] further on used a neuro-fuzzy learning method to optimize the performance of fuzzy risk models. The architecture of the developed hierarchical fuzzy inference system was however designed manually.

2.5.1 Fuzzy Set Theory

Ever since the concept of fuzzy logic was propounded by Prof. Lotfi Zadeh in 1965 [69], it has been found useful in several areas of discipline, providing a simple way for researchers to reach a definite conclusion that is based upon vague, imprecise, incomplete, randomness, noisy, ambiguous or missing input information. Until the discovery of fuzzy set by Prof. Zadeh, the world had only known the classical theory of sets also known as the traditional set theory or the crisp set.

The review of literature cannot be complete without some very good quotes that seem to be an embodiment of the whole idea of fuzziness. Below are some of the notable quotes about Fuzzy logic - the relative importance of precision.

As complexity rises, precise statements lose meaning and meaningful statements lose precision.

— ***Lotfi Zadeh***

Precision is not truth.

— ***Henri Matisse***

Sometimes the more measurable drives out the most important.

— ***René Dubos***

Vagueness is no more to be done away with in the world of logic than friction in mechanics.

— ***Charles Sanders Peirce***

I believe that nothing is unconditionally true, and hence I am opposed to every statement of positive truth and every man who makes it.

— ***H. L. Mencken***

So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.

— ***Albert Einstein***

All the above quotations are cited by [74]

2.5.2 Crisp Set Vs. Fuzzy Set

A crisp set is a set that can be considered as a container and the elements belonging to this set as the objects contained in it. By extension, an object will *either* be in the container *or* would not be in the container. A membership function μ_A of an element **x** for a crisp set **A** is defined as follows:

$$\mu_A(x) = \begin{cases} 1 & \text{If } x \,\varepsilon\, A. \\ 0 & \text{otherwise} \end{cases}$$

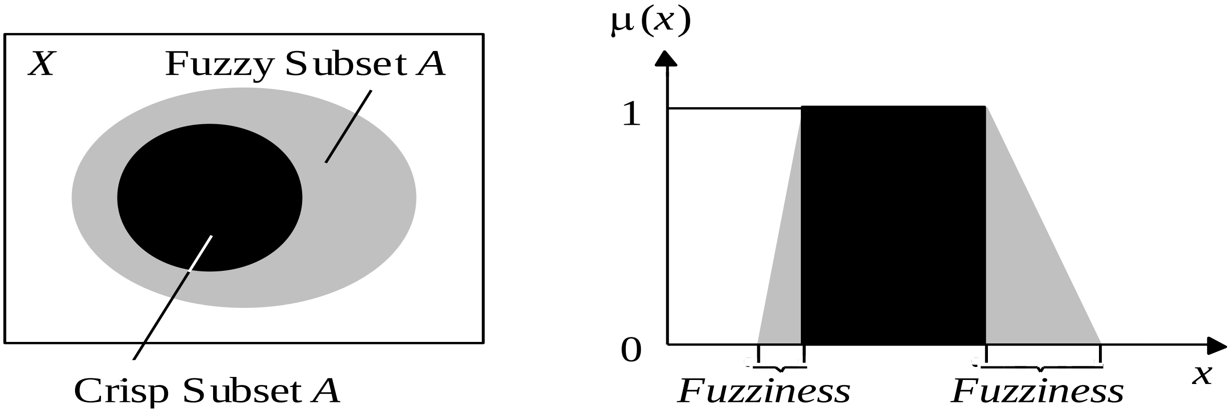


Figure 2.1: Fuzzy Set Representation

Similarly, the membership functions for the crisp set operations, like union, intersection and complement, can be expressed as follows:

$$\mu_{A \cup B}(x) = \max\,(\mu_A(x), \mu_B(x))$$

$$\mu_{A \cap B}(x) = \min\,(\mu_A(x), \mu_B(x)) \text{ and}$$

$$\mu_A(x) = 1 - \mu_A(x)$$

respectively. Therefore, a membership function for a crisp set **A** is defined as

$$\mu_A : X \longrightarrow \{0,1\}$$

Moreover, **A** is a subset of **B** if and only if $x \in \boldsymbol{A} \Rightarrow x \in \boldsymbol{B}\, \forall x$

In terms of membership function, **A** is a subset of **B** if and only if

$$\mu_A(x) \leq \mu_B(x),\ \forall x \in X$$

Example

For example: set A has elements a_1, a_2, a_3………….. a_{10}. If the set is a crisp set, it can be expressed as: $A = \{a_1, a_2, a_2 \ldots\ldots a_{10}\}$. (1.1)

It can also be written as:

$$\forall x \in \mathrm{X} : \mu_A(x) = 1 \text{ , if } x \in A$$

$$\mu_A(x) = 0 \text{ if otherwise.}$$

However, if it is a fuzzy set, it is defined as set of pairs $[\mu\,(a_i), a_i]$, where $\mu\,(a_i)$, is the membership value of element a_i. The fuzzy set is therefore expressed as:

$$\boldsymbol{A} = \{\, \mu(a_1) | \, a_1, \mu(a_2) | \, a_2, \mu(a_3) | \, a_{3,} \ldots\ldots \mu(a_{10}) | \, a_{10}\} \qquad (1.2)$$

Clearly, Figure 2.2 brings out the difference in membership value between the crisp set and the fuzzy set. The difference between the two concepts; *fuzzy* set and the conventional crisp set is largely the degree to which an object belongs to a set. In a crisp set, objects are either in or out of the set (container). A membership value of either 1 or O is assigned to each object in the universal set to discriminate between members and non-members of the crisp set under consideration. In a fuzzy set, however, a membership value between 1.0 and 0.0 can be assigned to each number in the universal set to indicate the degree to which the member belongs in the set under consideration, where zero means non membership and one signifies full membership.

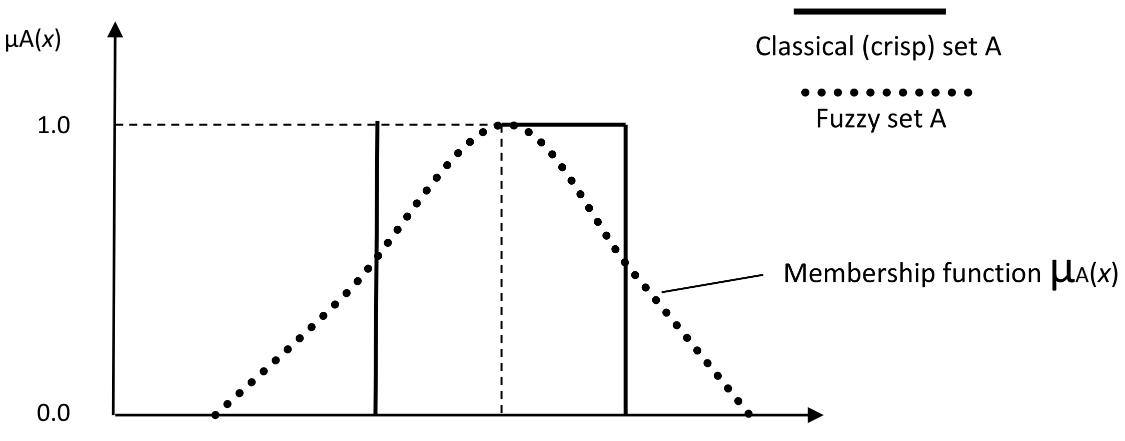


Figure 2.2**:** Crisp Set Vrs Fuzzy Set

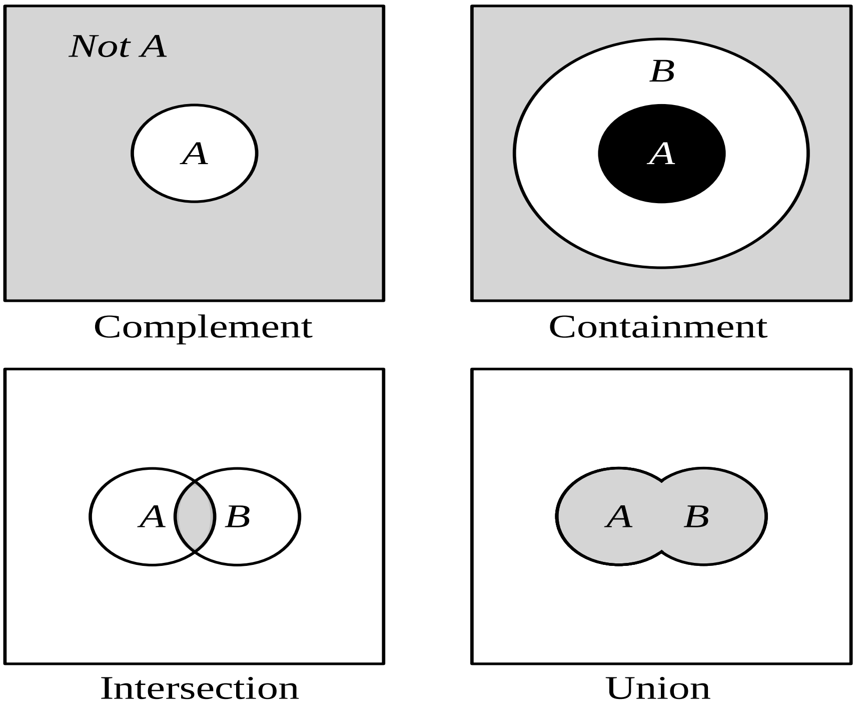
2.5.3 Fuzzy Sets Properties and Operations

The fuzzy properties and operations are the basis on which the fuzzy sets have been successfully used to deal with uncertainty on the one hand and to represent knowledge on the other. Fuzzy sets operations are defined much the same way as the classical sets operations are defined. In classical sets, operations like intersections and union of two sets and complement of a set are commonly used. Set theoretic operations like intersection, union and complement are uniquely defined for classical sets and are shown in table below.

2.5.3.1 Operations of Fuzzy Sets

A	**B**	**A⋂B**	**A∪B**	**Ā**
0	0	0	0	1
0	1	1	0	1
1	0	1	0	0
1	1	1	1	0

Table 2.2: Set theoretical operations



Complement

Containment

Intersection

Union

Figure 2.3: Fuzzy operations on Set A and B

2.5.3.2 Complement

The membership function of the Complement of a Fuzzy set A with membership function μ_A is defined as

$$\mu_{\neg A} = 1 - \mu_A$$

- Crisp Sets: Who does not belong to the set A?
- Fuzzy Sets: How much do elements not belong to the set A?
- The complement of a set is an opposite of this set. For example, if we have the set of tall men, its complement is the set of NOT tall men. When we remove the tall men set from the universe of discourse, we obtain the complement.
- If A is the fuzzy set, its complement $\neg A$ can be found as follows: [75]

$$\mu_{\neg A}(x) = 1 - \mu_A(x).$$

2.5.3.3 Containment

Containment begins on the premise that a set can contain other sets

- Crisp Sets: Which sets belong to which other sets?
- Fuzzy Sets: How much sets belong to other sets?

- Similar to a Chinese box, a set can contain other sets. The smaller set is called the subset. For example, the set of tall men contains all tall men; very tall men is a subset of tall men. However, the tall men set is just a subset of the set of men.
- In crisp sets, all elements of a subset entirely belong to a larger set.
- In fuzzy sets, however, each element can belong less to the subset than to the larger set. Elements of the fuzzy subset have smaller memberships in the subset than in the larger set. [75]

2.5.3.4 Intersection

- Crisp Sets: Which element belongs to both sets?
- Fuzzy Sets: How much of the element is in both sets?
- In classical set theory, an intersection between two sets contains the elements shared by these sets. For example, the intersection of the set of tall men and the set of fat men is the area where these sets overlap.
- In fuzzy sets, an element may partly belong to both sets with different memberships.
- A fuzzy intersection is the lower membership in both sets of each element. The fuzzy intersection of two fuzzy sets *A* and *B* on universe of discourse X: [75]

$$\mu_{A\cap B}(x) = \min\, [\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x),$$

where $x \in X$.

2.5.3.5 Union

- Crisp Sets: Which element belongs to either set?
- Fuzzy Sets: How much of the element is in either set?
- The union of two crisp sets consists of every element that falls into either set. For example, the union of tall men and fat men contains all men who are tall OR fat.
- In fuzzy sets, the union is the reverse of the intersection. That is, the union is the largest membership value of the element in either set. The fuzzy operation for forming the union of two fuzzy sets A and B on universe X can be given as: [75]

$$\mu_{A\cup B}(x) = \max\, [\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x), \quad \text{where } x \in X.$$

The extension of the intersection and union of two classical sets to the intersection and union of two fuzzy sets is not uniquely defined. It is clear that union and intersection operations for fuzzy sets should subject to the intersection and union of classical sets, because a classical set can be seen as special case of a fuzzy set. Zadeh (1965) proposed the following definition: [75]

$\mu_{A \cap B}(x) = \min [\mu_A(x), \mu_B(x)] = \mu_A(x) \cap \mu_B(x)$, intersection ....Eq no (1)

$\mu_{A \cup B}(x) = \max [\mu_A(x), \mu_B(x)] = \mu_A(x) \cup \mu_B(x)$, union ....Eq no (2)

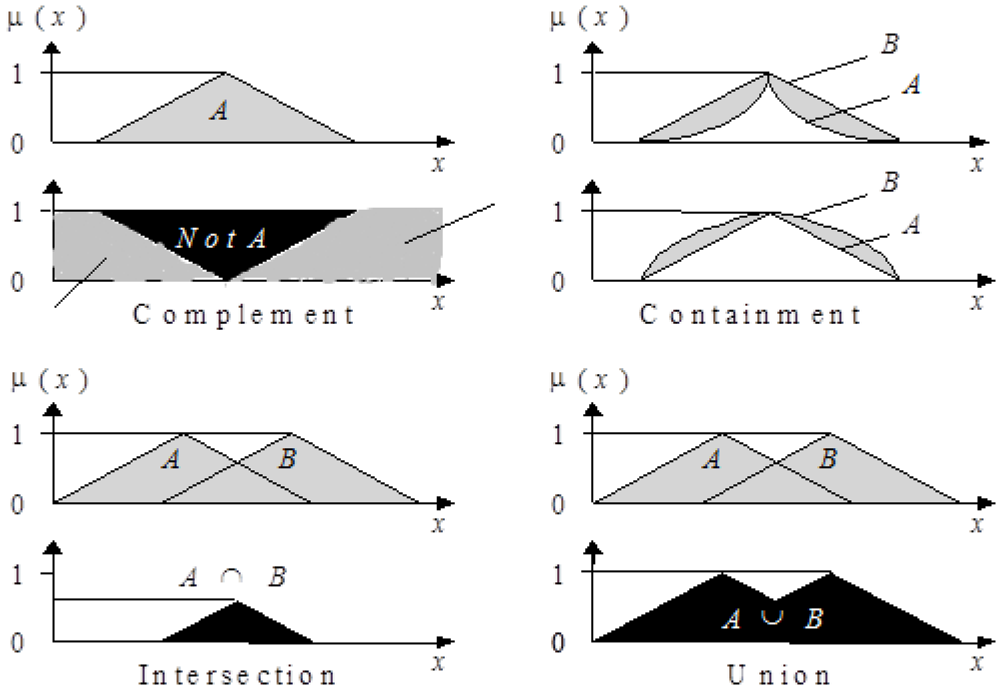


Figure 2.4: Summary of fuzzy set operations

2.5.4 Properties of Fuzzy Sets

2.5.4.1 Equality

Fuzzy set *A* is considered equal to a fuzzy set *B*, IF AND ONLY IF:

$$\mu_A(x) = \mu_B(x),\ \forall x \in X$$

Example: *A* = 0.3/1 + 0.5/2 + 1/3, *B* = 0.3/1 + 0.5/2 + 1/3, therefore *A* = *B*.

2.5.4.2 Inclusion

Inclusion of one fuzzy set into another fuzzy set. Fuzzy set $A \subseteq X$ is included in (is a subset of) another fuzzy set, $B \subseteq X$:

$$\mu_A(x) \leq \mu_B(x),\ \forall x \in X \qquad [75]$$

Example: Consider $X = \{1, 2, 3\}$ and sets A and B

$A = 0.3/1 + 0.5/2 + 1/3$;

$B = 0.5/1 + 0.55/2 + 1/3$

then A is a subset of B, *or* $A \subseteq B$

2.5.4.3 Empty Fuzzy Set

A fuzzy set A is empty, IF AND ONLY IF:

$$\mu_A(x) = 0,\ \forall x \in X$$

Example: Consider $X = \{1, 2, 3\}$ and fuzzy set

$A = 0/1 + 0/2 + 0/3$,

A is then described as *empty*.

2.5.4.4 Alpha-Cut

An alpha-cut (α-cut) or α-level set of a fuzzy set $A \subseteq X$ is defined as an ORDINARY SET $A_\alpha \subseteq X$, such that:

$$A_\alpha = \{\mu_A(x) \geq \alpha,\ \forall x \in X\}.$$

Example: Consider $X = \{1, 2, 3\}$ and set $A = 0.3/1 + 0.5/2 + 1/3$

then: $A_{0.5} = \{2, 3\}$, $A_{0.1} = \{1, 2, 3\}$, $A_1 = \{3\}$.

Example: Consider continuous universe of discourse $X = [a, b]$ and fuzzy set A with the membership function $\mu_A(x)$. α-cuts for some α_1 *and* α_2 are:

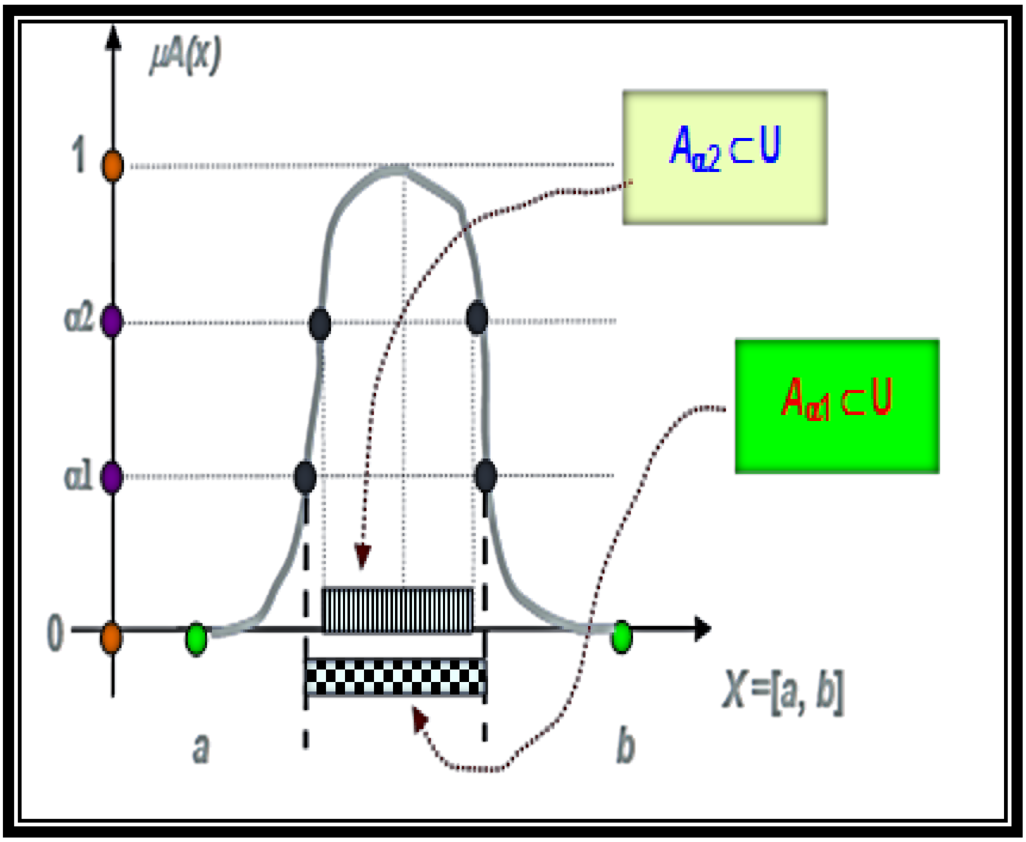


Figure 2.5: Alpha-cut

2.5.4.5 Fuzzy Set Normality

- A fuzzy subset of X is called normal if there exists at least one element $x \in X$ such that $\mu_A(x) = 1$.
- A fuzzy subset that is not normal is called subnormal.
- All crisp subsets except for the null set are normal. In fuzzy set theory, the concept of nullness essentially generalises to sub-normality.
- The height of a fuzzy set *A* is the largest membership grade of an element in *A*

$$height(A) = \max_x(\mu_A(x)).$$

- Fuzzy set is called normal if and only if: [75]

$$height(A) = 1$$

2.5.4.6 Fuzzy Sets Core and Support

Core: The core of a membership function of a fuzzy set A can be defined as the region of the universe that is characterized by complete and full membership in the fuzzy set A [75]. Thus the core comprises of those elements „*x*‟ within the universe; such that $\mu_A(x) = 1$.

Support: The support of a membership function of a fuzzy set A can be defined as the region of the universe that is characterized by all non-zero memberships in the fuzzy set A [75]. That is to say that, the support comprises of those elements „*x*‟ within the universe; such that $\mu_A(x) > 0$

Assume *A* is a fuzzy set over universe of discourse *X*.

- The support of *A* is the crisp subset of *X* consisting of all elements with membership grade:

$$supp(A) = \{x \mid \mu_A(x) > 0 \text{ and } x \in X\}$$

- The core of *A* is the crisp subset of *X* consisting of all elements with membership grade:

$$core(A) = \{x \mid \mu_A(x) = 1 \text{ and } x \in X\}$$

Example:

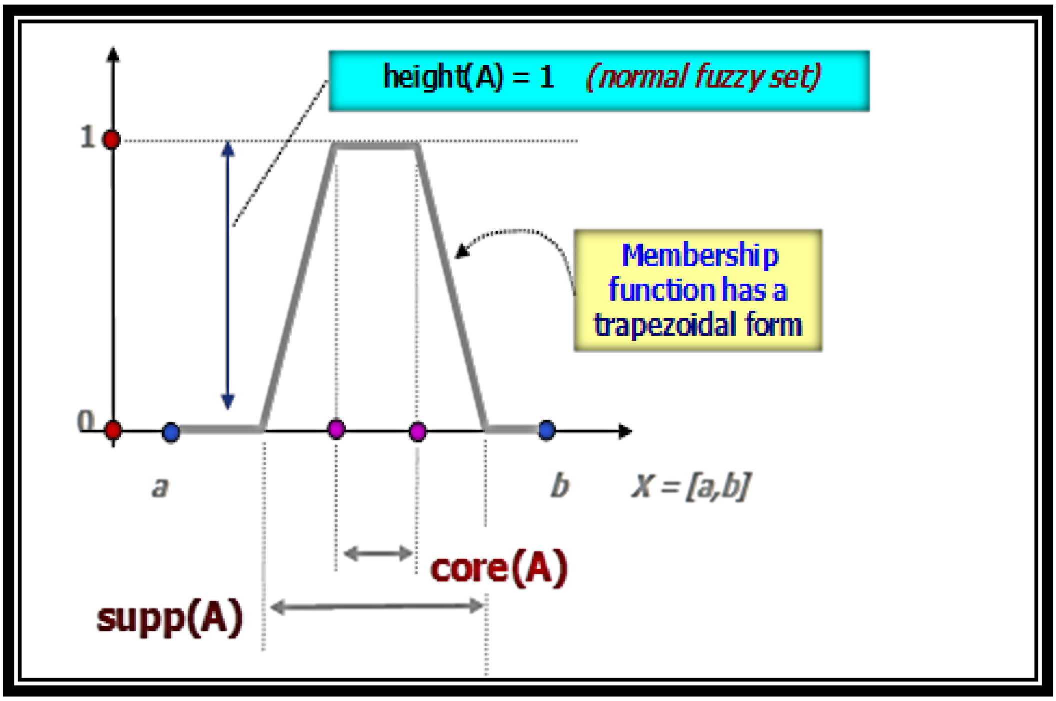

height(A) = 1 (normal fuzzy set)
1
0
Membership function has a trapezoidal form
a
b
X = [a,b]
core(A)
supp(A)

Figure 2.6: Fuzzy Core and Support.

2.55 Fuzzy Rules

Lotfi Zadeh published his second most influential paper in 1973. In the paper he outlined a new approach to analyzing complex systems by capturing human knowledge in through fuzzy rules [76].

In order to reason with fuzzy logic, fuzzy rules have to be represented by an implication. Such a fuzzy implication has the same truth value as the truth table of the classical implication in classical logic, but in fuzzy logic these types of statements are often referred to as fuzzy (*if then)* statements or fuzzy rules [77]. A typical rule-based system consists of *if-then rules*, a bunch of *facts*, and an *interpreter* controlling the application of the rules, given the facts. These *if-then* rule statements are used to formulate the conditional statements that comprise the complete knowledge base. A single *if-then* rule assumes the form 'if *x* is *A* then *y* is *B*' and the if-part of the rule '*x* is *A*' is called the *antecedent* or *premise*, while the then-part of the rule '*y* is *B*' is called the *consequent* or *conclusion.* [77]

A fuzzy rule can be defined as a conditional statement in the form:

$$IF \quad x \;\; is \;\; A \qquad THEN \;\; y \;\; is \;\; B$$

where x and y are linguistic variables; and A and B are linguistic values determined by fuzzy sets on the universe of discourses X and Y, respectively.

2.56 Fuzzification of Inputs

For any fuzzy rule to be implemented it has to be first fuzzified, implying that the numerical inputs are converted to fuzzy inputs and then a translation from fuzzified output to numerical output. The first translation is known as fuzzification and second is known as defuzzification. The fuzzification of the fuzzy input is the construction of fuzzy relation and compositional rule of inference [78].

There are generally three types of fuzzifiers, which are used for the fuzzification process; they are:

1. Singleton fuzzifier,
2. Gaussian fuzzifier, and
3. Trapezoidal or triangular fuzzifier.

2.57 Defuzzification of Outputs

Defuzzification is used to translate the fuzzy output of a fuzzy system to a numerical representation to be used for controlling the output of a fuzzy behavior of a control system. A number of defuzzification methods [78] are available for defuzzifying the output. These include

1. Center of Gravity Defuzzification.
2. Mean of Maxima Defuzzification.
3. Indexed Defuzzification Method.
4. Center of Area Defuzzification.

2.58 Centroid Defuzzification Technique

The centroid method is also popularly known as center of gravity or center of area defuzzification. This technique was developed by Sugeno in 1985. It is the most commonly used technique and is very accurate. The CoG, method is a technique for finding a crisp value (*u*) from the mid-point of the output fuzzy set using a weighted average of the membership grades. Suppose, there exists a fuzzy set within a discrete universe, and μ (xi) is its membership value in the membership function [79]. The following expression can be used to represent the weighted average of the elements in the support set.

$$x^{*} = \frac{\int \mu_i(x)\, x\, dx}{\int \mu_i(x)\, dx}$$ [78]

The researcher will use the center of gravity for the defuzzification process through the MATLAB fuzzy tool kit.

2.59 Fuzzy Inference System

The Fuzzy inference process is used to map a given input to an output using fuzzy logic. The mapping then provides the basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves Membership Functions, Logical Operations, and If-Then Rules. The two main types of fuzzy inference systems used are the

Mamdani-type and Sugeno-type. These two types of inference systems vary somewhat in the way outputs are determined. [80, 81, 82],

2.6 Mamdani Method

Mamdani's fuzzy inference method is the most commonly used fuzzy methodology. Mamdani's method was among the first control systems built using fuzzy set theory. It is commonly used in applications, due to its simple structure of 'min-max' operations [83]. The mamdani fuzzy inference process involves four steps namely:

- Fuzzification of the input variables
- Rule evaluation
- Aggregation of the rule outputs and
- Defuzzification

The research uses the mamdani method.

2.7 Fuzzy Aggregation

Fuzzy aggregation is a method used to aggregate subjective data often based on algebraic operations with fuzzy numbers. This method usually applies by the α-cuts concept in representing fuzzy numbers [51]. Subjective public or expert opinions regarding certain alternatives or services are more easily represented by linguistic terms than by a numerical value. In the fuzzy aggregation method, the subjective opinions represented by linguistic terms are transformed into fuzzy membership functions to be used in extended fuzzy operation algebras. Usually the aggregation is conducted by using many criteria and their different fuzzified weights.

The strength of this method in evaluating user perception of security is that it allows for the use of a linguistic “value,” which is known as the most efficient way to represent a person’s perception, and it aggregates the perceptions without losing the variety of each individual’s decision making characteristics. Another advantage of this method is that there is no need to have numerical inputs that would need to be fuzzified.

CHAPTER 3
METHODOLOGY

3.1 Introduction

The aim of this chapter is to provide a general overview of the methodological approach and design implementation selected for the evaluation of users' perception of security on online social networking sites by fuzzy set theory. In this chapter, the methodology applied to the evaluation of users' perception consists of the following stages:

1. Method of delivery
2. Data Collection
3. Design model
4. Analysing data

3.2 Method of Delivery

The design implementation of this fuzzy logic system was done using an interactive environment known as MATLAB which enables the user to perform computationally intensive tasks and to implement numerical algorithms for a variety of process applications. MATLAB is a technical computing environment that incorporates its own programming language similar to *C* or *C++*. MATLAB also has toolboxes for almost all the in-built applications including fuzzy logic.

3.2.1 MATLAB Fuzzy Logic Toolbox

MATLAB has several in-built tool boxes and one of them is the Fuzzy Logic Toolbox. The Fuzzy Logic Toolbox provides functions, graphical tools, and a Simulink block for analyzing, designing, and simulating systems based on fuzzy logic. The product takes the user through the steps of designing fuzzy inference systems. The toolbox enables the user to model complex system behaviors using simple logic rules and then implement these rules in a fuzzy inference system [84].

3.3 Data Collection

In most research based on fuzzy approaches, it is common to have surveys designed for experts where their opinions are used to construct fuzzy membership functions. In other surveys there is a combination of both users' and experts' opinions in constructing membership functions but premium weight is giving to the experts' opinions. [85]. However, there is a significant issue regarding the use of small number of experts rather against using a large number of people from the public especially when the research is user-centered.

The issue is whether the experts' opinions truly represent users' perceptions or not. In other words, are the experts' opinions different from the public's opinions? In the case of security risk on social networking sites, sampling the opinions of users would be far more effective than using experts who may even not be using social media sites.

3.3.1 Online questionnaire

In order to construct a fuzzy rule-based assessment for the evaluation of users' perception of security risk on social networking sites, an online questionnaire was designed mainly based on the CIA triad of confidentiality, Integrity and Availability. In all there were seven social network security based questions under each Linguistic variable to help form users' perception of security on each of the selected social networking sites which were Facebook, Twitter and LinkedIn respectively. The link for the survey was pasted on all the three social networking sites for users to respond. Other users were emailed with the link to respond to the survey. The survey questionnaire was online for approximately two months. In **Appendix A,** is a sample of the questionnaires.

In representing users' perceptions as a fuzzy membership function, the interval estimation method was used. The interval estimation generates more suitable results for continuous measurements. Participants understand and represent their opinions more easily using interval estimation. Often time an interval estimation method used for constructing fuzzy membership functions representing the respondents' perception level for each linguistic scale is most appropriate and is commonly used [86, 87]. Also membership functions constructed using interval estimation is more precise as compared to those developed using direct rating or polling methods. In **Appendix B** is a sample of the questionnaire used for the interval estimation method.

3.4 Design Model of the Linguistic Variables

The inputs to the system as mentioned in chapter two were criteria commonly used in the evaluation of Information technology security. The same criteria were found appropriate to be used in the evaluation of user perception of security risk on social networking site security risk and these were confidentiality, integrity and availability. These criteria or linguistic variables are assumed to be of the same weight and a particular value is determined for each of them based on questions that are answered about a specific social networking site. The values determined for each of the input were defined as a fuzzy number instead of crisp numbers by using suitable fuzzy sets. Designing the fuzzy system requires that the different inputs (that is, confidentiality, integrity, and availability) are represented by fuzzy sets. The fuzzy sets are in turn represented by a membership function. The membership function used in this research is the triangular membership function which is a three point function defined by minimum (α), maximum (β) and modal (m) values where ($\alpha \leq m \leq \beta$).

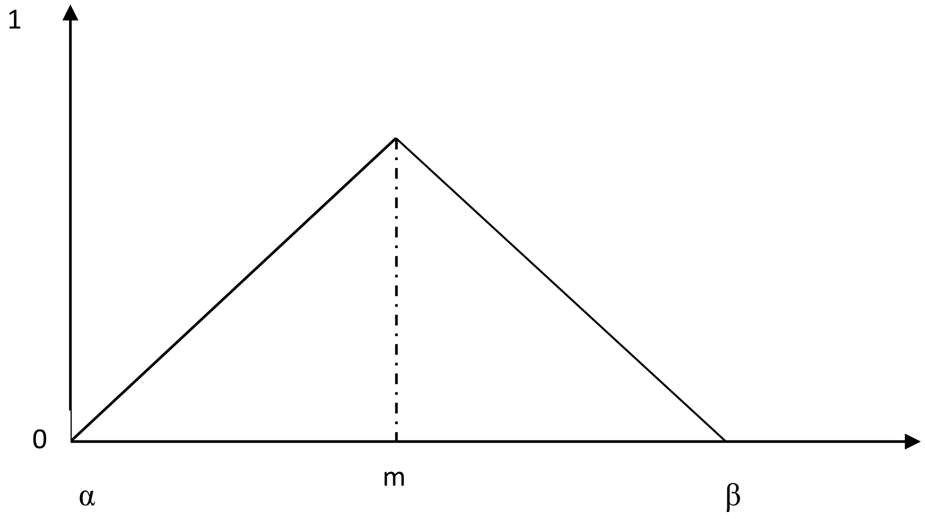


Figure 3.1: Triangular membership function

3.4.1 The Fuzzy Sets

The level of confidentiality as a linguistic variable was defined on a set of membership functions of *not confidential*, *slightly confidential*, *very confidential* and *extremely confidential*. The level of integrity was also defined based on the scales of *very low*, *low*, *high*, *very high*, and *extremely high* whiles the level of availability was defined by the scales of *not often, rarely often, often, very often*, and *always available*. The levels defined above were based on a range definition with an estimated interval of [0-10].

The output, which is, the level of security risk is similarly designed and also represented by fuzzy sets and then a membership function. The level of security risk, the output, is defined based on the scales: *not secure*, *slightly secure*, *secure, very secure*, and *extremely secure* within the range of [0 - 30].

Now based on the results from the survey which were based on the interval estimation method, the following range were defined for each of the membership functions belonging to each of the three inputs.

Not Confidential	**(0, 0, 2.5)**
Slightly Confidential	**(0, 2.5, 5)**
Confidential	**(2.5, 5, 7.5)**
Very Confidential	**(5, 7.5, 10)**
Extremely Confidential	**(7.5, 10, 10)**

Table 3.1: Range of Inputs for confidentiality

Very Low	**(0, 0, 2.5)**
Low	**(0, 2.5, 5)**
High	**(2.5, 5, 7.5)**
Very High	**(5, 7.5, 10)**
Extremely High	**(7.5, 10, 10)**

Table 3.2: Range of Inputs for Data Integrity

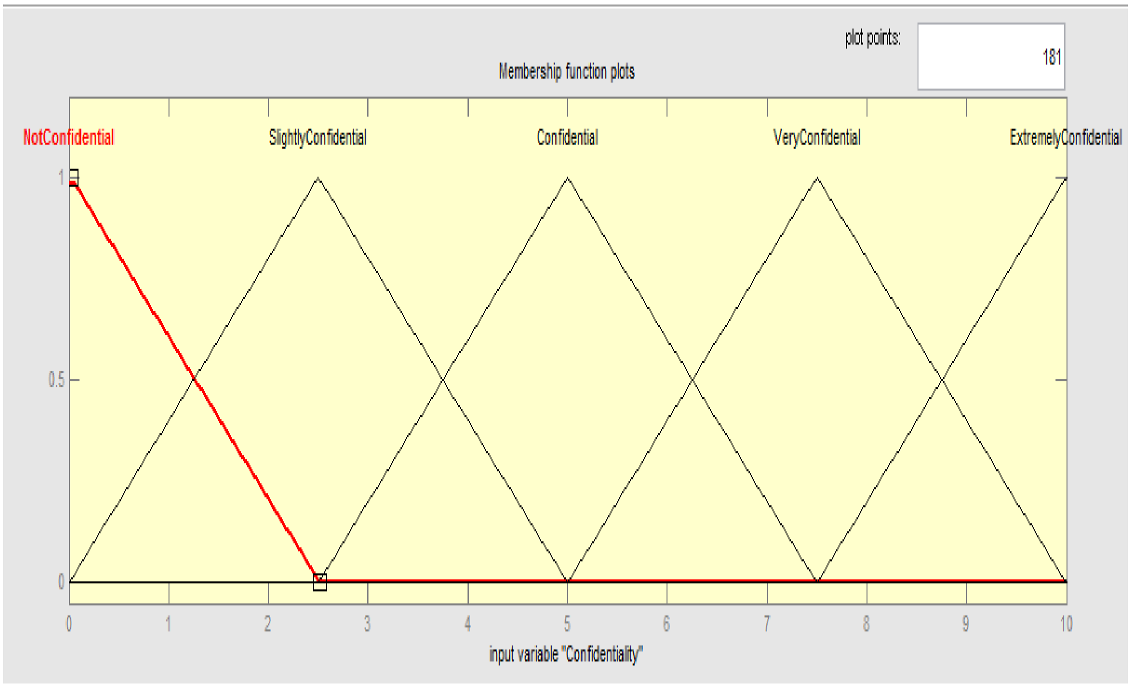
Not Often	**[0, 0, 2.5]**
Slightly often	**[0, 2.5, 5]**
Often	**[2.5, 5, 7.5]**
Very Often	**[5, 7.5, 10]**
Always Available	**[7.5, 10, 10]**

Table 3.3: Range of Inputs for Availability

Not Secure	**[0, 0, 7.5]**
Slightly Secure	**[0, 7.5, 15]**
Secure	**[7.5, 15, 22.5]**
Very Secure	**[15, 22.5, 30]**
Extremely Secure	**[22.5, 30, 30]**

Table 3.4: Range of output for Level of Security

The following exhibits are the membership functions for the three inputs and the output as designed in the MATLAB fuzzy tool box.



plot points: 181

Figure 3.2: Membership function for Confidentiality

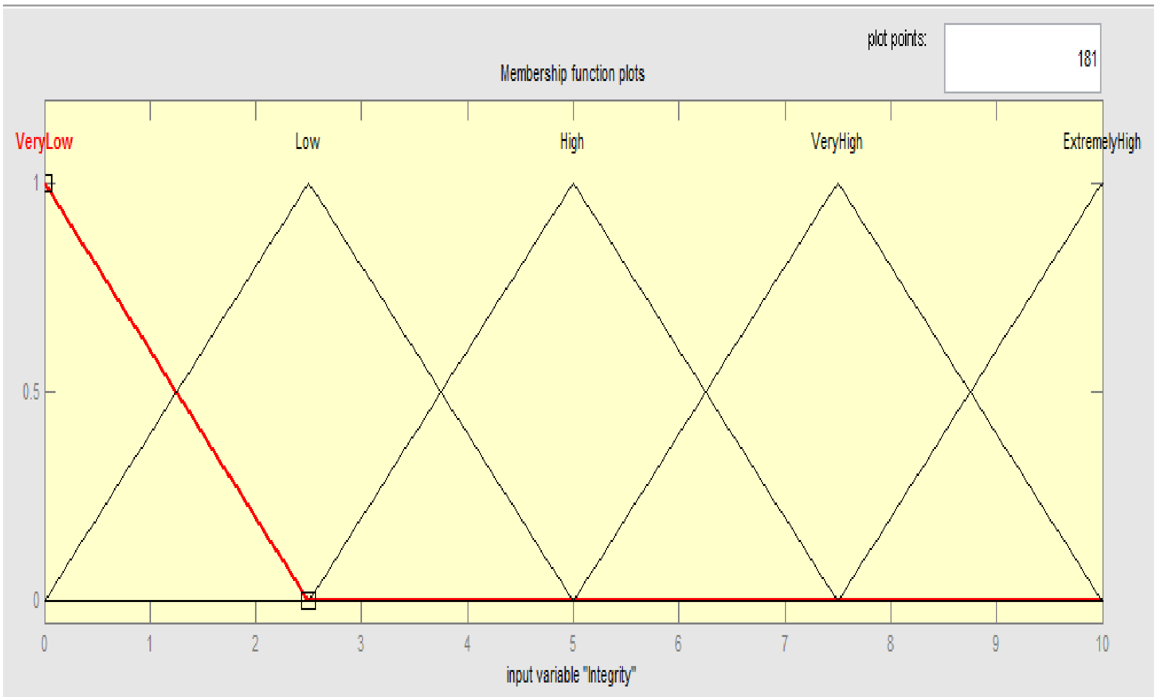


Figure 3.3: Membership function for Integrity

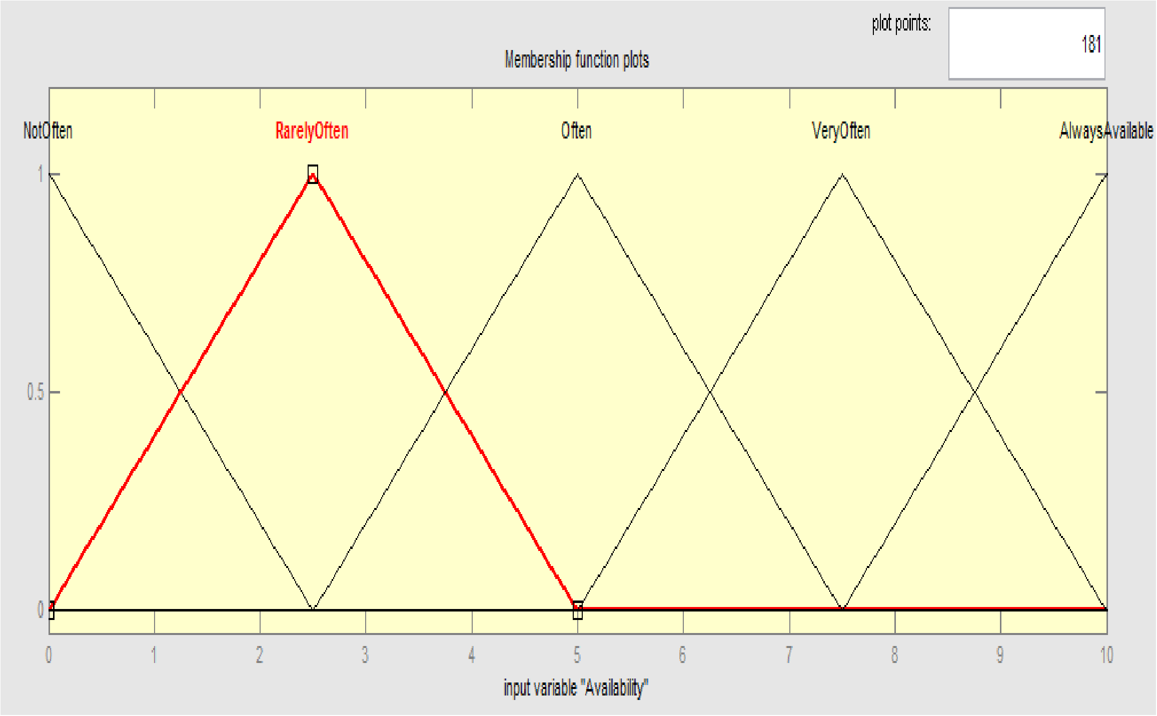


Figure 3.4: Membership function for Availability

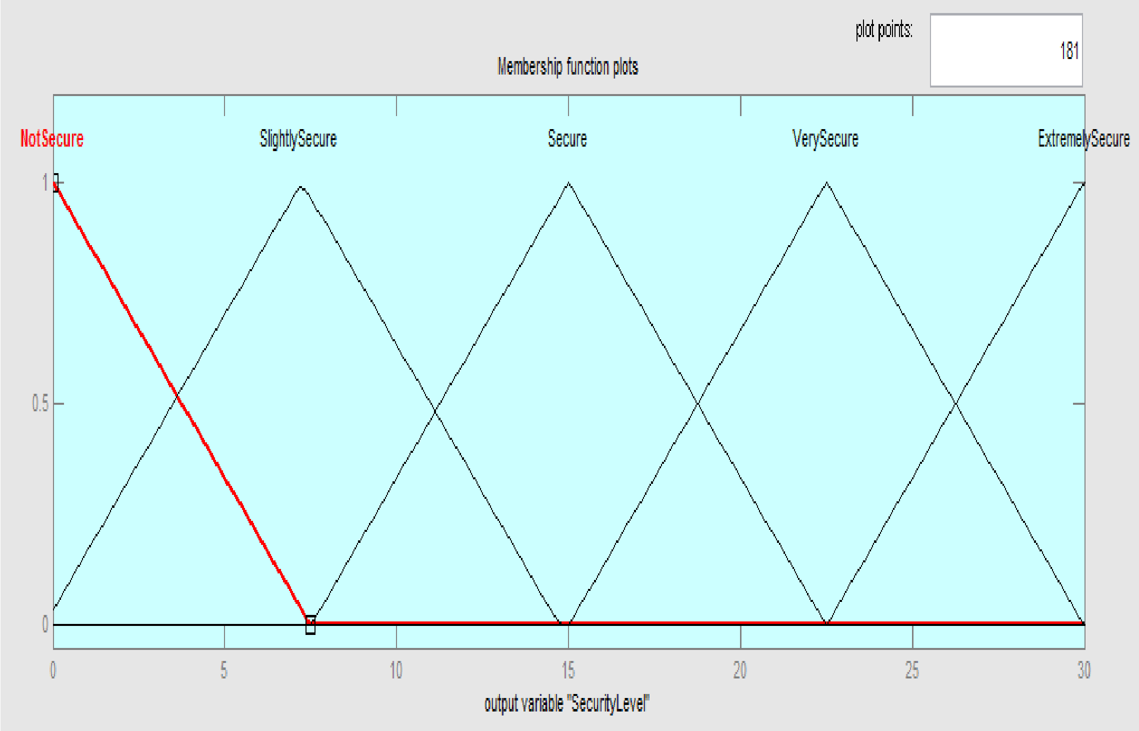

plot points: 181

Figure 3.5: Membership function for Level of Security risk

3.4.2 Design of the fuzzy inference system

Stage 1: **How the input variables were determined**

The input variables were determined as mentioned above based on the popular CIA TRIAD of confidentiality, integrity and availability that are used in the evaluation of information technology security. The same CIA criteria were deemed appropriate to be incorporated into a secure social networking site application system.

Stage 2: **Defining the input variable membership function**

Each of the inputs were defined on a domain interval of [0 - 10], based on the results from the survey using the interval estimation method. Again based on the results, the domain was then divided into 2N + 1 regions and each region is attached a fuzzy membership function. In this research, the domain was divided into 5 regions (N =2). The regions were represented by triangular membership functions as exhibited in MATLAB above.

Stage 3: Defining the output variable membership function.

The output domain interval was estimated to be [0 - 30]. The domain interval was again divided into 2N + 1 region for each region and a membership function attached for each. For example level of security risk (the output) is divided into 5 regions (N = 2) represented by not secure, slightly secure, secure, very secure, and extremely secure as the fuzzy sets.

Stage 4: **Formulating rules and populating the rule base.**

The rules were built based on knowledge of the relationships between the variables. The rules were formulated so as to reflect the relationships between any possible relations of the input variables to the output variable. The rules in this work reflected the relationships between the levels of confidentiality, integrity and availability and also the level of security risk. Thus, there were $(2N + 1)^3$ fuzzy rules in the rule base of the fuzzy system. To determine the overall security risk level for each networking site, the rule base needs $5^3 = 125$ rules since there were five linguistic values and three linguistic variables (confidentiality, Integrity and availability). The complete rules base used to construct the overall knowledge base are summarized in **Table 3.5** for different linguistic values

The levels of confidentiality, integrity, and availability were used in the antecedent of rules and the level of security risk as the consequent of rules. A fuzzy rule is conditional statement in the form: IF x is A THEN y is B where x and y are linguistic variables and A and B are linguistic values determined by fuzzy sets on universe of discourses X and Y, respectively. Both the antecedent and consequent of a fuzzy rule can have multiple parts. All parts of the antecedent are calculated simultaneously and resolved in a single number and the antecedent affects all parts of the consequent equally.

Rule	**IF**	**AND**	**AND**	**THEN**
r	**Confidentiality is**	**Integrity is**	**Availability is**	**Security Level is**
1	*Not Confidential*	*Very Low*	*Not Often*	*Not Secure*
2	*Not Confidential*	*Very Low*	*Rarely Often*	*Not Secure*
3	*Not Confidential*	*Very Low*	*Often*	*Not Secure*
4	*Not Confidential*	*Very Low*	*Very Often*	*Slightly Secure*
5	*Not Confidential*	*Very Low*	*Always Available*	*Slightly Secure*
6	*Not Confidential*	*Low*	*Not Often*	*Slightly Secure*
7	*Not Confidential*	*Low*	*Rarely Often*	*Slightly Secure*
8	*Not Confidential*	*Low*	*Often*	*Slightly Secure*
9	*Not Confidential*	*Low*	*Very Often*	*Slightly Secure*
10	*Not Confidential*	*Low*	*Always Available*	*Slightly Secure*
11	*Not Confidential*	*High*	*Not Often*	*Slightly Secure*
12	*Not Confidential*	*High*	*Rarely Often*	*Secure*
13	*Not Confidential*	*High*	*Often*	*Secure*
14	*Not Confidential*	*High*	*Very Often*	*Secure*
15	*Not Confidential*	*High*	*Always Available*	*Secure*
16	*Not Confidential*	*Very High*	*Not Often*	*Slightly Secure*
17	*Not Confidential*	*Very High*	*Rarely Often*	*Slightly Secure*
18	*Not Confidential*	*Very High*	*Often*	*Slightly Secure*
19	*Not Confidential*	*Very High*	*Very Often*	*Secure*
20	*Not Confidential*	*Very High*	*Always Available*	*Secure*

21	*Not Confidential*	*Extremely High*	*Not Often*	*Slightly Secure*
22	*Not Confidential*	*Extremely High*	*Rarely Often*	*Slightly Secure*
23	*Not Confidential*	*Extremely High*	*Often*	*Slightly Secure*
24	*Not Confidential*	*Extremely High*	*Very Often*	*Secure*
25	*Not Confidential*	*Extremely High*	*Always Available*	*Secure*
26	*Slightly Confidential*	*Very Low*	*Not Often*	*Slightly Secure*
27	*Slightly Confidential*	*Very Low*	*Rarely Often*	*Slightly Secure*
28	*Slightly Confidential*	*Very Low*	*Often*	*Slightly Secure*
29	*Slightly Confidential*	*Very Low*	*Very Often*	*Slightly Secure*
30	*Slightly Confidential*	*Very Low*	*Always Available*	*Secure*
31	*Slightly Confidential*	*Low*	*Not Often*	*Slightly Secure*
32	*Slightly Confidential*	*Low*	*Rarely Often*	*Slightly Secure*
33	*Slightly Confidential*	*Low*	*Often*	*Slightly Secure*
34	*Slightly Confidential*	*Low*	*Very Often*	*Secure*
35	*Slightly Confidential*	*Low*	*Always Available*	*Secure*
36	*Slightly Confidential*	*High*	*Not Often*	*Slightly Secure*
37	*Slightly Confidential*	*High*	*Rarely Often*	*Slightly Secure*
38	*Slightly Confidential*	*High*	*Often*	*Secure*
39	*Slightly Confidential*	*High*	*Very Often*	*Secure*
40	*Slightly Confidential*	*High*	*Always Available*	*Secure*
41	*Slightly Confidential*	*Very High*	*Not Often*	*Slightly Secure*
42	*Slightly Confidential*	*Very High*	*Rarely Often*	*Slightly Secure*
43	*Slightly Confidential*	*Very High*	*Often*	*Secure*

44	*Slightly Confidential*	*Very High*	*Very Often*	*Secure*
45	*Slightly Confidential*	*Very High*	*Always Available*	*Secure*
46	*Slightly Confidential*	*Extremely High*	*Not Often*	*Slightly Secure*
47	*Slightly Confidential*	*Extremely High*	*Rarely Often*	*Slightly Secure*
48	*Slightly Confidential*	*Extremely High*	*Often*	*Secure*
49	*Slightly Confidential*	*Extremely High*	*Very Often*	*Secure*
50	*Slightly Confidential*	*Extremely High*	*Always Available*	*Secure*
51	*Confidential*	*Very Low*	*Not Often*	*Slightly Secure*
52	*Confidential*	*Very Low*	*Rarely Often*	*Slightly Secure*
53	*Confidential*	*Very Low*	*Often*	*Slightly Secure*
54	*Confidential*	*Very Low*	*Very Often*	*Secure*
55	*Confidential*	*Very Low*	*Always Available*	*Secure*
56	*Confidential*	*Low*	*Not Often*	*Slightly Secure*
57	*Confidential*	*Low*	*Rarely Often*	*Slightly Secure*
58	*Confidential*	*Low*	*Often*	*Slightly Secure*
59	*Confidential*	*Low*	*Very Often*	*Secure*
60	*Confidential*	*Low*	*Always Available*	*Secure*
61	*Confidential*	*High*	*Not Often*	*Slightly Secure*
62	*Confidential*	*High*	*Rarely Often*	*Secure*
63	*Confidential*	*High*	*Often*	*Secure*
64	*Confidential*	*High*	*Very Often*	*Secure*
65	*Confidential*	*High*	*Always Available*	*Secure*
66	*Confidential*	*Very High*	*Not Often*	*Slightly Secure*

67	*Confidential*	*Very High*	*Rarely Often*	*Slightly Secure*
68	*Confidential*	*Very High*	*Often*	*Secure*
69	*Confidential*	*Very High*	*Very Often*	*Very Secure*
70	*Confidential*	*Very High*	*Always Available*	*Very Secure*
71	*Confidential*	*Extremely High*	*Not Often*	*Slightly Secure*
72	*Confidential*	*Extremely High*	*Rarely Often*	*Secure*
73	*Confidential*	*Extremely High*	*Often*	*Secure*
74	*Confidential*	*Extremely High*	*Very Often*	*Very Secure*
75	*Confidential*	*Extremely High*	*Always Available*	*Very Secure*
76	*Very Confidential*	*Very Low*	*Not Often*	*Slightly Secure*
77	*Very Confidential*	*Very Low*	*Rarely Often*	*Slightly Secure*
78	*Very Confidential*	*Very Low*	*Often*	*Slightly Secure*
79	*Very Confidential*	*Very Low*	*Very Often*	*Secure*
80	*Very Confidential*	*Very Low*	*Always Available*	*Secure*
81	*Very Confidential*	*Low*	*Not Often*	*Slightly Secure*
82	*Very Confidential*	*Low*	*Rarely Often*	*Slightly Secure*
83	*Very Confidential*	*Low*	*Often*	*Secure*
84	*Very Confidential*	*Low*	*Very Often*	*Secure*
85	*Very Confidential*	*Low*	*Always Available*	*Secure*
86	*Very Confidential*	*High*	*Not Often*	*Slightly Secure*
87	*Very Confidential*	*High*	*Rarely Often*	*Secure*
88	*Very Confidential*	*High*	*Often*	*Secure*
89	*Very Confidential*	*High*	*Very Often*	*Very Secure*

90	*Very Confidential*	*High*	*Always Available*	*Very Secure*
91	*Very Confidential*	*Very High*	*Not Often*	*Slightly Secure*
92	*Very Confidential*	*Very High*	*Rarely Often*	*Secure*
93	*Very Confidential*	*Very High*	*Often*	*Secure*
94	*Very Confidential*	*Very High*	*Very Often*	*Very Secure*
95	*Very Confidential*	*Very High*	*Always Available*	*Extremely Secure*
96	*Very Confidential*	*Extremely High*	*Not Often*	*Slightly Secure*
97	*Very Confidential*	*Extremely High*	*Rarely Often*	*Secure*
98	*Very Confidential*	*Extremely High*	*Often*	*Very Secure*
99	*Very Confidential*	*Extremely High*	*Very Often*	*Very Secure*
100	*Very Confidential*	*Extremely High*	*Always Available*	*Extremely Secure*
101	*Extremely Confidential*	*Very Low*	*Not Often*	*Slightly Secure*
102	*Extremely Confidential*	*Very Low*	*Rarely Often*	*Slightly Secure*
103	*Extremely Confidential*	*Very Low*	*Often*	*Slightly Secure*
104	*Extremely Confidential*	*Very Low*	*Very Often*	*Secure*
105	*Extremely Confidential*	*Very Low*	*Always Available*	*Secure*
106	*Extremely Confidential*	*Low*	*Not Often*	*Slightly Secure*
107	*Extremely Confidential*	*Low*	*Rarely Often*	*Slightly Secure*
108	*Extremely Confidential*	*Low*	*Often*	*Secure*
109	*Extremely Confidential*	*Low*	*Very Often*	*Secure*
110	*Extremely Confidential*	*Low*	*Always Available*	*Secure*
111	*Extremely Confidential*	*High*	*Not Often*	*Slightly Secure*
112	*Extremely Confidential*	*High*	*Rarely Often*	*Slightly Secure*

113	*Extremely Confidential*	*High*	*Often*	*Secure*
114	*Extremely Confidential*	*High*	*Very Often*	*Very Secure*
115	*Extremely Confidential*	*High*	*Always Available*	*Very Secure*
116	*Extremely Confidential*	*Very High*	*Not Often*	*Slightly Secure*
117	*Extremely Confidential*	*Very High*	*Rarely Often*	*Secure*
118	*Extremely Confidential*	*Very High*	*Often*	*Very Secure*
119	*Extremely Confidential*	*Very High*	*Very Often*	*Very Secure*
120	*Extremely Confidential*	*Very High*	*Always Available*	*Extremely Secure*
121	*Extremely Confidential*	*Extremely High*	*Not Often*	*Slightly Secure*
122	*Extremely Confidential*	*Extremely High*	*Rarely Often*	*Secure*
123	*Extremely Confidential*	*Extremely High*	*Often*	*Very Secure*
124	*Extremely Confidential*	*Extremely High*	*Very Often*	*Extremely Secure*
125	*Extremely Confidential*	*Extremely High*	*Always Available*	*Extremely Secure*

Table 3.5: Complete set of IF-THEN Fuzzy Rule

3.5 Analyzing the data

The questionnaire was structured into two sections. Section A contains the personal data of the respondents such as gender, age, the kind of social networking site they use, the number of hours they spend on their respective social networking sites in a week, the people they interact with most on social networking sites among others.

The Section B concentrated on sampling users' views or perception about security on social networking sites by answering questions mainly on the three criteria used for the evaluation of security risk. These were confidentiality, Integrity and Availability. Respondents were asked to fill the questionnaire each for every social networking site they subscribe to using the appropriate listed membership functions.

3.5.1 Analysis of Survey Responses Using Fuzzy Aggregation Method

In many research studies pertaining to user perceptions, the criteria under review are evaluated using linguistic scales with various numbers of descriptors. For example, a five descriptor linguistic scale could probably comprise of the terms: *very low, low, high, very high*, and *extremely high*. However, most studies calculate only simple descriptive statistics, such as percentage or frequency, to analyze these types of responses. The proposed method, based on the fuzzy aggregation method, enables one to analyze and aggregate the subjective responses without losing the variety of individual decision making characteristics.

3.5.2 Fuzzy Aggregation using Weighted Average

One of the most common aggregation operator often found in literature is the weighted average (WA). Also known as the weighted mean, it is similar to an arithmetic mean where instead of each of data points contributing equally to the final average, some data points contribute more than others. There are weighted versions of other means such as the weighted geometric mean (WGM) and the weighted harmonic mean (WHM). There is also the ordering weighted average (OWA).

3.5.3 Mathematical definition of weighted average

Formally, the weighted average of a non-empty set of data

$\{x_1, x_2, \dots, x_n\}$, with non-negative weights

$\{w_1, w_2, \dots, w_n\}$, is the quantity

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

which means:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}.$$

Therefore data elements with a high weight contribute more to the weighted mean than elements with low weights. The weights cannot be negative. Some may be zero, but not all of them (since division by zero is not allowed).

The formulas are simplified when the weights are normalized such that they sum up to 1,

i.e. $\sum_{i=1}^{n} w_i = 1$

For such normalized weights the weighted mean is simply. $\bar{x} = \sum_{i=1}^{n} w_i x_i$

The common mean $\frac{1}{n}\sum_{i=1}^{n} x_i$ is a special case of the weighted mean where all data have

equal weights, $w_i = w$

When the weights are normalized then $w_i = \frac{1}{n}$.

Another interesting concept that can be used as an aggregation operator is the probability. These two concepts have been used in a lot of applications concerning statistics, economics, engineering, physics, etc. Probably, these two concepts are the most relevant in statistics. However, there are a lot of other aggregation operators such as the ordered weighted averaging (OWA) operator and others.

3.5.4 Triangular Fuzzy Numbers

For a triangular fuzzy number X (TFN X) having minimum value X1, kernel value X2 and maximum value X3 the number is written as (X1, X2, X3).

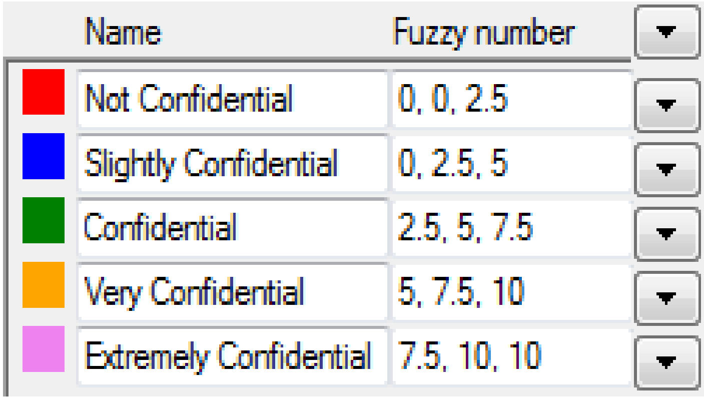
Let TFN X = (X1, X2, X3)

Let TFN Y = (Y1, Y2, Y3)

The sum of X and Y is (X1+Y1, X2+Y2, X3+Y3)

The difference of X and Y would also be (X1-Y3, X2-Y2, X3-Y1)

In this research, respondents were to choose between a series of statements on the ordinal/interval scale the one they judge most appropriate and it is argued that the choice of score is, in effect, a judgement between 3 indicator statements. Thus, for example, respondents rate the level of confidentiality of the following information such dating history or intimate secrets they submit with friends on Facebook on the following scale.



	Name	Fuzzy number	▾
🟥	Not Confidential	0, 0, 2.5	▾
🟦	Slightly Confidential	0, 2.5, 5	▾
🟩	Confidential	2.5, 5, 7.5	▾
🟧	Very Confidential	5, 7.5, 10	▾
🟪	Extremely Confidential	7.5, 10, 10	▾

Figure 3.6: Range of inputs of confidentiality

In this interpretation, a respondent who judges "*very confidential*" to be the appropriate score makes a constrained choice in the range where 5 is the minimum value, 7.5 is the modal value and 10 the maximum. (To think of it another way, the respondent must consider which of the three hypotheses, *confidential*, *Very confidential and extremely confidential* best represents their judgement of the situation.)

In extracting the fuzzy scores on a range of 10, the descriptor "*Not confidential"* corresponds to a triangular fuzzy number (0, 0, 2.5). Similarly, the descriptor "*very confidential" also* corresponds to (5, 7.5, 10), and so on. The full scoring correspondence is taken to be as follows.

Not confidential = TFN (0, 0, 2.5)
Slightly confidential = TFN (0, 2.5, 5)
Confidential = TFN (2.5, 5, 7.5)
Very confidential = TFN (5, 7.5, 10)
Extremely confidential = TFN (7.5, 10, 10)

Now with a results table such as the following from

Total Number of Respondents 42

Dating History	Membership function	Not confidential	Slightly confidential	Confidential	Very confidential	Extremely confidential
	Frequency	7	5	11	12	7

Taking the average weighted score for each Triangular fuzzy number (TFN) representing the appropriate membership function or descriptor yields the TFN (2.17, 2.95, 3.73) when carried out with appropriate attention to arithmetic rules for Triangular fuzzy numbers (TFNs) thus;

(0,0,2.5)x7+(0,2.5,5)x5+(2.5, 5, 7.5)x11+(5, 7.5, 10)x12+(7.5, 10, 10)x7)/42 **= (3.33, 5.416, 7.5)**

Financial Information	Membership function	Not confidential	Slightly confidential	Confidential	Very confidential	Extremely confidential
	Frequency	3	5	15	14	5

(0,0,2.5)x3+(0,2.5,5)x5+(2.5, 5, 7.5)x15+(5, 7.5, 10)x14+(7.5, 10, 10)x5)/45 = **(3.45, 5.77, 7.98)**

Now to find the overall perception of the two questions under confidentiality, we find the aggregate sum as the following:

[(3.33, 5.416, 7.5)*1 + (3.45, 5.77, 7.98)*1]/1+1 = 5.57

This method of aggregation would then be replicated in Chapter 4 for all the three linguistic variables (confidentiality, integrity and availability). The three inputs representing the three variables would then be fed into the fuzzy logic tool box to generate the appropriate output of the level of security risk on each of three social networking sites.

CHAPTER 4

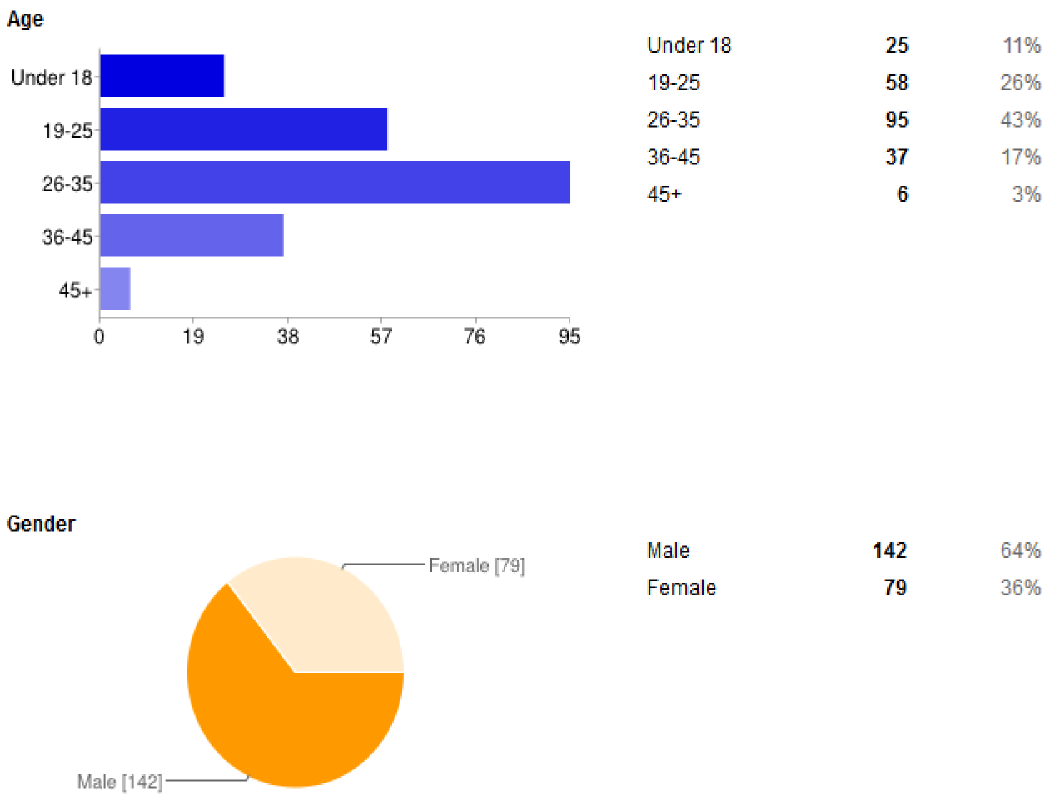
EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Introduction

A survey was designed to collect the data needed to evaluate users' perception and then test the design methodology model described in Chapter 3. The main goals of this survey are:

- To collect actual data from the completed online questionnaire for analysis.
- To use appropriate fuzzy aggregation technique as identified in Chapter 3 to find the overall user perception of security risk for each of the three social networking sites namely Facebook, Twitter and LinkedIn.

4.2 Descriptive Statistics of Respondents

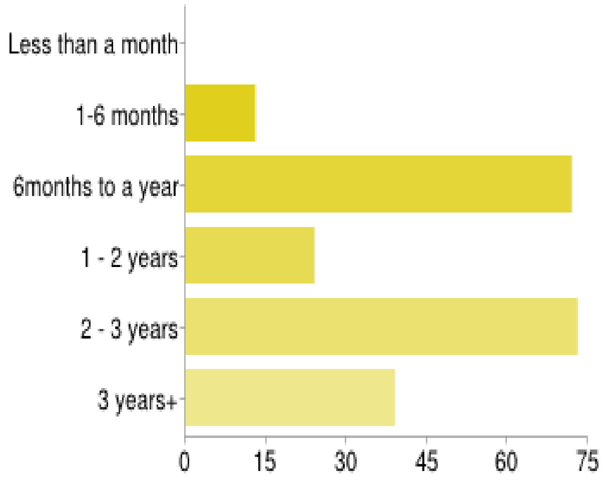


Under 18	25	11%
19-25	58	26%
26-35	95	43%
36-45	37	17%
45+	6	3%

Male	142	64%
Female	79	36%

Figure 4.1: Age and Gender statistics of Respondents

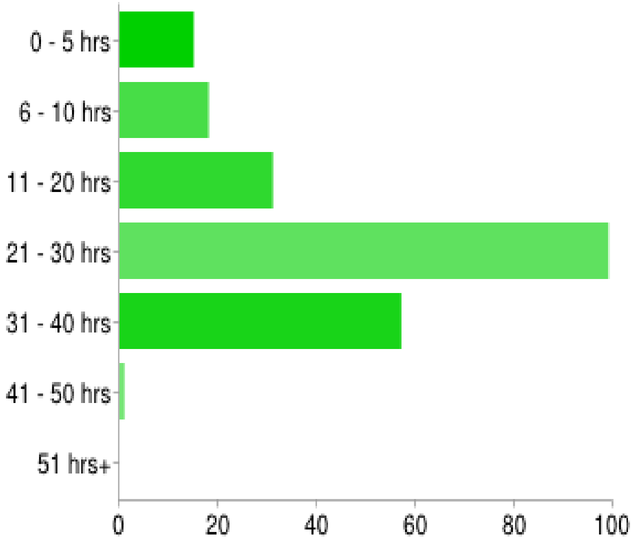
Estimate how long you have been using social networking sites?



Less than a month	**0**	0%
1-6 months	**13**	6%
6months to a year	**72**	33%
1 - 2 years	**24**	11%
2 - 3 years	**73**	33%
3 years+	**39**	18%

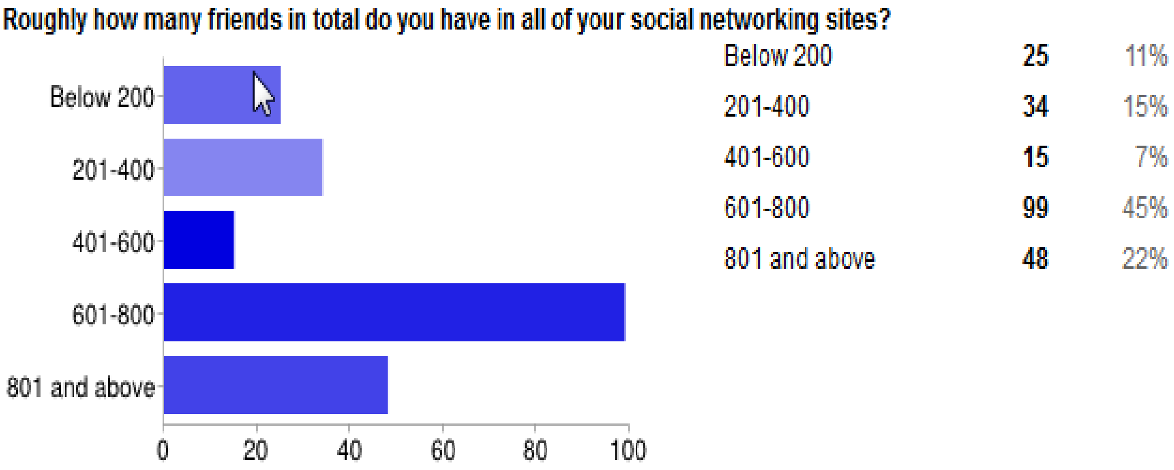
Figure 4.2: How long respondents have been using SNSs

Estimate how many hours a week you spend on social networking sites?



0 - 5 hrs	**15**	7%
6 - 10 hrs	**18**	8%
11 - 20 hrs	**31**	14%
21 - 30 hrs	**99**	45%
31 - 40 hrs	**57**	26%
41 - 50 hrs	**1**	0%
51 hrs+	**0**	0%

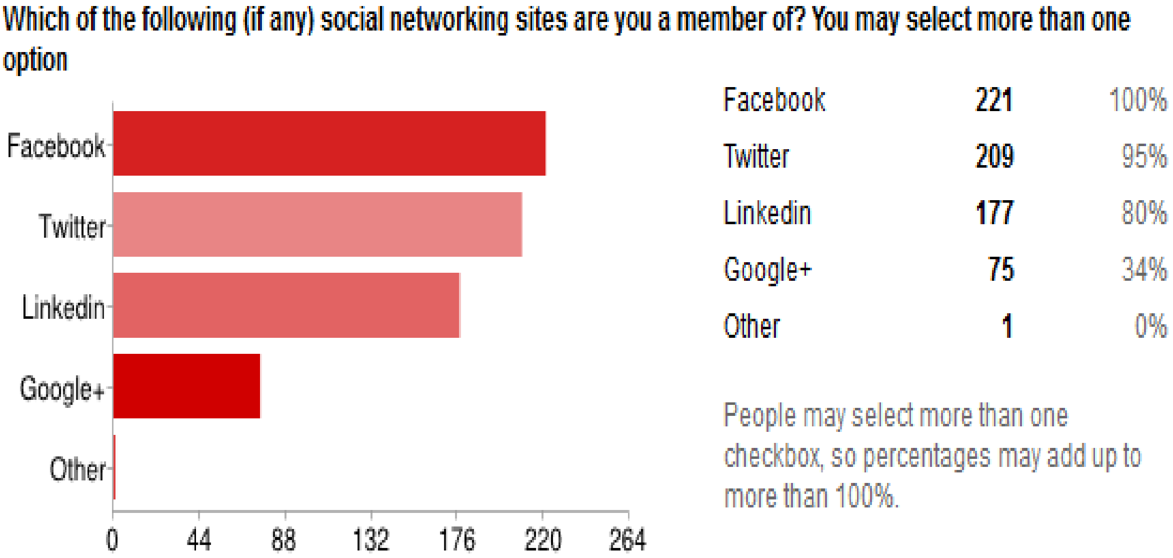
Figure 4.3: Time in a week spent on SNSs



Roughly how many friends in total do you have in all of your social networking sites?

Below 200	**25**	11%
201-400	**34**	15%
401-600	**15**	7%
601-800	**99**	45%
801 and above	**48**	22%

Figure 4.4: Total number of friends on SNSs



Which of the following (if any) social networking sites are you a member of? You may select more than one option

Facebook	**221**	100%
Twitter	**209**	95%
Linkedin	**177**	80%
Google+	**75**	34%
Other	**1**	0%

People may select more than one checkbox, so percentages may add up to more than 100%.

Figure 4.5: Respondents choice of SNSs

Facebook: Confidentiality

Question: *In general, how would you rate the level of confidentiality of the following information you submit with friends on Facebook.*

Total Number of Respondents: 221

Membership functions							
		Not confidential	***Slightly confidential***	***Confidential***	***Very confidential***	***Extremely confidential***	***Weighted Average***
Dating History	Frequency	0	2	135	81	3	(3.46, 5.96, 8.43)
Financial information (eg. info on things you buy, where you buy from, etc)	Frequency	4	4	94	118	1	(3.77, 6.22, 8.71)
Gossip between friends	Frequency	4	156	59	1	1	(0.72, 3.18, 5.67)
Intimate secrets	Frequency	3	42	72	103	1	(3.18, 5.64, 8.13)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	36	105	55	24	1	(1.20, 3.30, 5.78)
Professional / work related information	Frequency	33	80	79	28	1	(1.56, 3.70, 6.18)
Religious / political beliefs	Frequency	77	76	3	64	1	(1.51, 3.14, 5.63)

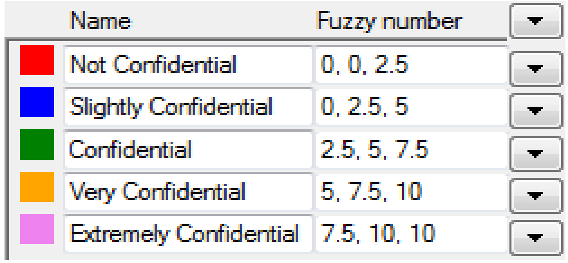
Table 4.1: Aggregation of responses of confidentiality on Facebook

For Example Dating history was calculated as below:

= (0, 0, 2.5) *0 + (0, 2.5, 5) *2 + (2.5, 5, 7.5) * 135 + (5, 7.5, 10) * 81+ (7.5, 10, 10)*3)/221

Final sum of weighted averages: **[2.2, 4.5, 6.93]**

Center of gravity: **4.52**



	Name	Fuzzy number
🟥	Not Confidential	0, 0, 2.5
🟦	Slightly Confidential	0, 2.5, 5
🟩	Confidential	2.5, 5, 7.5
🟧	Very Confidential	5, 7.5, 10
🟪	Extremely Confidential	7.5, 10, 10

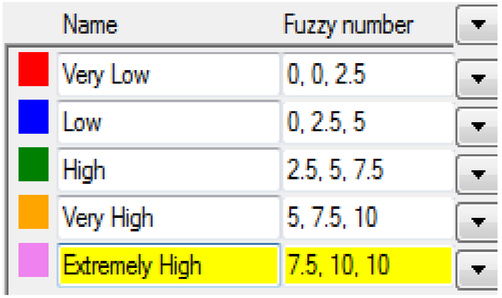
Facebook: Integrity

Question: *Rate the level of authenticity of the origin of the message and files you receive from the following people on Facebook*

Total Number of Respondents: 221

Membership functions							
		Very Low	***Low***	***High***	***Very High***	***Extremely High***	***Weighted Average***
Close Friends	Frequency	2	2	52	160	5	(4.37, 6.85, 9.30)
Co-workers	Frequency	1	56	10	148	6	(3.67, 6.23, 8.58)
Family members	Frequency	1	16	75	126	3	(3.80, 6.28, 8.76)
Friends	Frequency	1	48	27	139	6	(3.65, 6.14, 8.57)
People who live far away from you	Frequency	1	52	109	46	13	(2.71, 5.20, 7.55)
Strangers (people you have never met in person)	Frequency	3	60	10	127	21	(3.70, 6.16, 8.42)
Friends of your friends	Frequency	2	2	114	102	1	(3.63, 6.10, 8.60)

Table 4.2: Aggregation of responses of Integrity on Facebook



	Name	Fuzzy number
🟥	Very Low	0, 0, 2.5
🟦	Low	0, 2.5, 5
🟩	High	2.5, 5, 7.5
🟧	Very High	5, 7.5, 10
🟪	Extremely High	7.5, 10, 10

Final sum of weighted averages: [3.65, 6.14, 8.54]

Center of Gravity: **6.10**

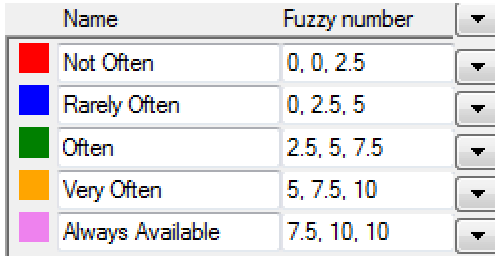
Facebook: Availability

Question: *How often do you have ready access to the following information on your chosen Facebook?*

Total Number of Respondents: 221

Membership functions							
		Not often	***Rarely often***	***Often***	***Very often***	***Always Available***	***Weighted Average***
Message history	Frequency	1	3	6	43	168	(6.74, 9.23, 9.76)
Chat history	Frequency	2	2	36	102	79	(7.02, 9.51, 9.93)
The website itself	Frequency	1	1	1	34	184	(6.74, 9.23, 9.76)
Intimate secrets	Frequency	1	2	1	1	216	(7.36, 9.85, 9.90)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	1	1	34	71	114	(5.85, 8.35, 9.56)
Professional / work related information	Frequency	3	2	10	105	101	(5.91, 8.38, 9.74)
Profile information	Frequency	1	1	1	41	177	(6.95, 9.43, 9.93)

Table 4.3: Aggregation of responses of availability on Facebook



	Name	Fuzzy number
🟥	Not Often	0, 0, 2.5
🟦	Rarely Often	0, 2.5, 5
🟩	Often	2.5, 5, 7.5
🟧	Very Often	5, 7.5, 10
🟪	Always Available	7.5, 10, 10

Final sum of weighted averages: [6.65, 9.14, 9.80]

Center of gravity: **8.53**

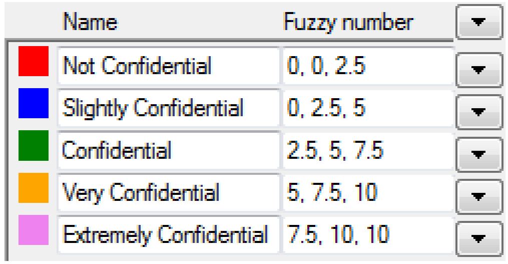
Twitter: Confidentiality

Question: *How often do you have ready access to the following information on your chosen Facebook?*

Total Number of Respondents: 209

		Membership functions					
		Not confidential	*Slightly confidential*	*Confidential*	*Very confidential*	*Extremely confidential*	*Weighted Average*
Dating History	Frequency	1	31	76	101	0	(3.32, 5.81, 8.31)
Financial information (eg. info on things you buy, where you buy from, etc)	Frequency	0	0	13	196	0	(4.84, 7.34, 9.84)
Gossip between friends	Frequency	0	1	106	102	0	(3.70, 6.21, 8.71)
Intimate secrets	Frequency	0	1	162	46	0	(3.04, 5.53, 8.04)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	0	0	82	98	29	(4.36, 6.87, 9.02)
Professional / work related information	Frequency	3	21	103	62	20	(3.43, 5.90, 8.15)
Religious / political beliefs	Frequency	1	0	109	61	38	(4.12, 6.61, 8.70)

Table 4.4: Aggregation of responses of confidentiality on Twitter



	Name	Fuzzy number
🟥	Not Confidential	0, 0, 2.5
🟦	Slightly Confidential	0, 2.5, 5
🟩	Confidential	2.5, 5, 7.5
🟧	Very Confidential	5, 7.5, 10
🟪	Extremely Confidential	7.5, 10, 10

Final sum of weighted averages: [3.83, 6.32, 8.68]

Center of gravity: **6.30**

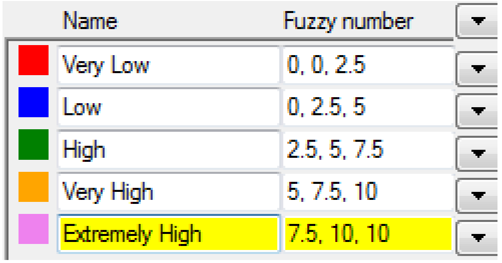
Twitter: Integrity

Question: *How often do you have ready access to the following information on your chosen Facebook?*

Total Number of Respondents: 209

Membership functions							
		Very Low	***Low***	***High***	***Very High***	***Extremely High***	***Weighted Average***
Close Friends	Frequency	1	1	1	32	174	(7.02, 9.51, 9.92)
Co-workers	Frequency	1	3	3	141	61	(5.60, 7.36, 9.86)
Family members	Frequency	1	2	69	133	4	(4.15, 6.64, 9.10)
Friends	Frequency	1	1	33	172	2	(4.58, 6.97, 9.52)
People who live far away from you	Frequency	1	3	1	143	61	(5.62, 8.11, 9.88)
Strangers (people you have never met in person)	Frequency	1	4	1	132	71	(5.72, 8.20, 9.85)
Friends of your friends	Frequency	7	14	101	45	42	(3.80, 6.21, 8.21)

Table 4.5: Aggregation of responses of integrity on Twitter



	Name	Fuzzy number
🟥	Very Low	0, 0, 2.5
🟦	Low	0, 2.5, 5
🟩	High	2.5, 5, 7.5
🟧	Very High	5, 7.5, 10
🟪	Extremely High	7.5, 10, 10

Final sum of weighted averages: [5.21, 7.57, 9.50]

Center of gravity: **7.42**

Twitter: Availability

Question: *How often do you have ready access to the following information on your chosen Facebook?*

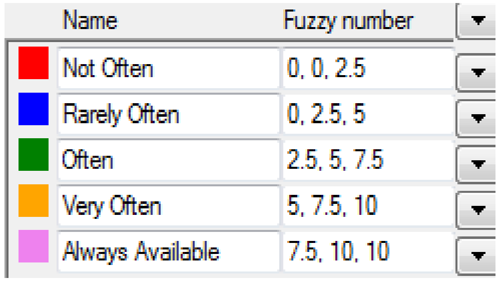
Total Number of Respondents: 209

Membership functions							
		Not often	***Rarely often***	***Often***	***Very often***	***Always Available***	***Weighted Average***
Message history	Frequency	1	2	1	93	112	(6.25, 8.74, 9.90)
Chat history	Frequency	1	4	1	146	57	(5.55, 8.03, 9.85)
The website itself	Frequency	1	7	5	33	163	(6.70, 9.18, 9.73)
Intimate secrets	Frequency	1	1	1	94	112	(6.30, 8.77, 9.93)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	1	1	2	162	43	(5.44, 7.93, 9.92)
Professional / work related information	Frequency	0	3	1	104	101	(6.12, 8.62, 9.92)
Profile information	Frequency	0	1	1	72	135	(6.57, 9.07, 9.96)

Table 4.6: Aggregation of responses of availability on Twitter

Final sum of weighted averages: [6.13, 8.62, 9.88]

Center of gravity: **8.21**



	Name	Fuzzy number
🟥	Not Often	0, 0, 2.5
🟦	Rarely Often	0, 2.5, 5
🟩	Often	2.5, 5, 7.5
🟧	Very Often	5, 7.5, 10
🟪	Always Available	7.5, 10, 10

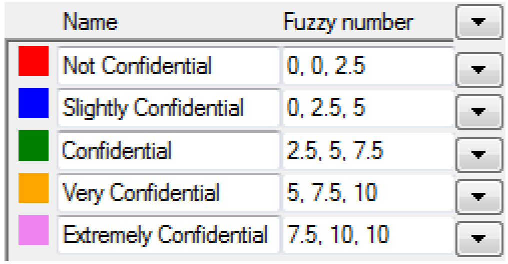
LinkedIn: Confidentiality

Question: *How often do you have ready access to the following information on your chosen Facebook?*

Total Number of Respondents: 177

Membership functions							
		Not confidential	***Slightly confidential***	***Confidential***	***Very confidential***	***Extremely confidential***	***Weighted Average***
Dating History	Frequency	2	4	3	162	6	(4.87, 7.34, 9.76)
Financial information (eg. info on things you buy, where you buy from, etc)	Frequency	1	143	33	0	0	(0.45, 2.95, 5.45)
Gossip between friends	Frequency	41	36	100	0	0	(1.41, 3.33, 8.82)
Intimate secrets	Frequency	2	173	2	0	0	(0.02, 2.5, 5.00)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	40	45	64	28	0	(1.70, 3.62, 6.12)
Professional / work related information	Frequency	65	43	69	0	0	(0.97, 2.55, 5.10)
Religious / political beliefs	Frequency	1	75	101	0	0	(1.43, 3.91, 6.41)

Table 4.7: Aggregation of responses of confidentiality on LinkedIn



	Name	Fuzzy number
🟥	Not Confidential	0, 0, 2.5
🟦	Slightly Confidential	0, 2.5, 5
🟩	Confidential	2.5, 5, 7.5
🟧	Very Confidential	5, 7.5, 10
🟪	Extremely Confidential	7.5, 10, 10

Final sum of Weighted averages: [1.55, 3.74, 6.67]

Center of gravity: **3.98**

LinkedIn: Integrity

Question: *How often do you have ready access to the following information on your chosen Facebook?*

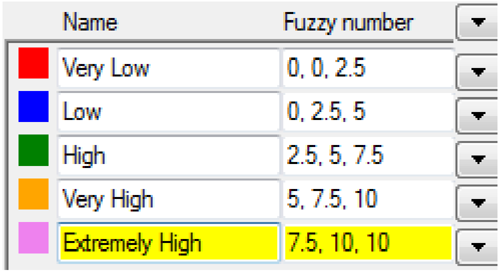
Total Number of Respondents: 177

Membership functions							
		Very Low	***Low***	***High***	***Very High***	***Extremely High***	***Weighted Average***
Close Friends	Frequency	1	138	33	5	0	(0.60, 3.09, 5.60)
Co-workers		43	69	65	0	0	(0.92, 2.81, 5.31)
Family members	Frequency	90	1	80	6	0	(1.30, 2.52, 5.02)
Friends	Frequency	132	21	61	0	0	(0.86, 2.01, 5.04)
People who live far away from you	Frequency	95	188	1	0	0	(0.01, 2.68, 6.70)
Strangers (people you have never met in person)	Frequency	75	58	34	10	0	(0.50, 1.78, 4.13)
Friends of your friends	Frequency	104	45	28	0	0	(0.40, 1.43, 3.92)

Table 4.8: Aggregation of responses of integrity on LinkedIn

Final sum of Weighted averages: [0.65, 2.33, 5.1]

Center of gravity: **2.70**



	Name	Fuzzy number
🟥	Very Low	0, 0, 2.5
🟦	Low	0, 2.5, 5
🟩	High	2.5, 5, 7.5
🟧	Very High	5, 7.5, 10
🟪	Extremely High	7.5, 10, 10

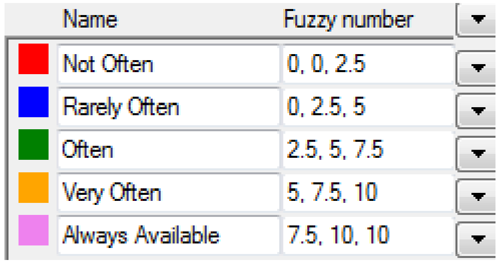
LinkedIn: Availability

Question: *How often do you have ready access to the following information on your chosen Facebook?*

Total Number of Respondents: 177

		Membership functions					
		Not often	***Rarely often***	***Often***	***Very often***	***Always Available***	***Weighted Average***
Message history	Frequency	1	0	53	76	47	(4.88, 7.37, 9.20)
Chat history		1	0	0	129	47	(5.64, 8.12, 9.96)
The website itself	Frequency	1	0	25	93	58	(5.44, 7.92, 9.60)
Intimate secrets	Frequency	0	0	9	77	91	(6.20, 8.67, 9.87)
Lifestyle related (eg. photos, blogs, history etc)	Frequency	0	0	54	35	88	(5.50, 7.98, 9.24)
Professional / work related information	Frequency	0	0	57	78	42	(4.80, 7.30, 9.20)
Profile information	Frequency	1	0	10	1	165	(7.16, 9.65, 9.81)

Table 4.9: Aggregation of responses of availability on LinkedIn



	Name	Fuzzy number
🟥	Not Often	0, 0, 2.5
🟦	Rarely Often	0, 2.5, 5
🟩	Often	2.5, 5, 7.5
🟧	Very Often	5, 7.5, 10
🟪	Always Available	7.5, 10, 10

Final sum of Weighted averages: [5.66, 8.14, 9.55]

Center of gravity: **7.78**

Summary of Results

	Confidentiality	*Integrity*	*Availability*
Facebook	4.52	6.10	8.53
Twitter	6.30	7.42	8.21
LinkedIn	3.98	2.70	7.70

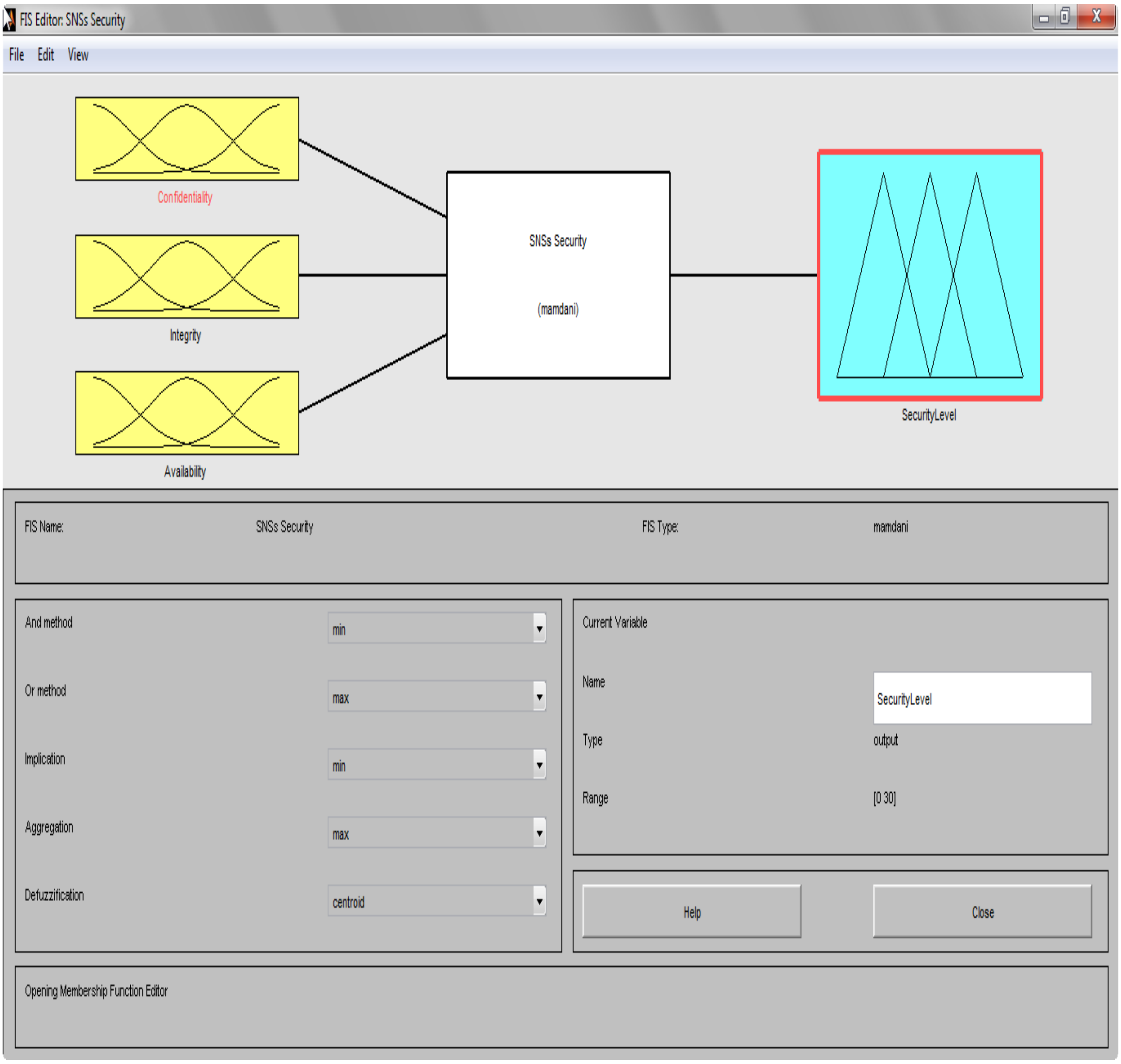
Table 4.10: Final value of aggregated responses

4.3 Implementation Procedure in MATLAB

The final result for each of the linguistic variables (the inputs) which were derived after aggregating the responses from the well-constructed online social networking sites security questions were fed into MATLAB to derive the final output (security risk level) for each of the three selected social media sites. The linguistic inputs were supplied through the graphical user interface called rule viewer. Once the rule viewer has been opened, the input variables are supplied in the text box captioned *input* with each of them separated with a space.

4.3.1 The MATLAB Fuzzy Inference System (FIS) Editor

The fuzzy inference system editor below shows the summary of the fuzzy inference system. In the editor, there is shown the mapping of the inputs to the system type and to the output. The input variables were respectively *confidentiality*, *integrity* and *availability*. The output was *security level* whiles the rules were constructed using the Mamdani fuzzy reasoning and the defuzzification technique was done using the centroid technique.



FIS Name: SNSs Security

FIS Type: mamdani

And method: min

Or method: max

Implication: min

Aggregation: max

Defuzzification: centroid

Current Variable

Name: SecurityLevel

Type: output

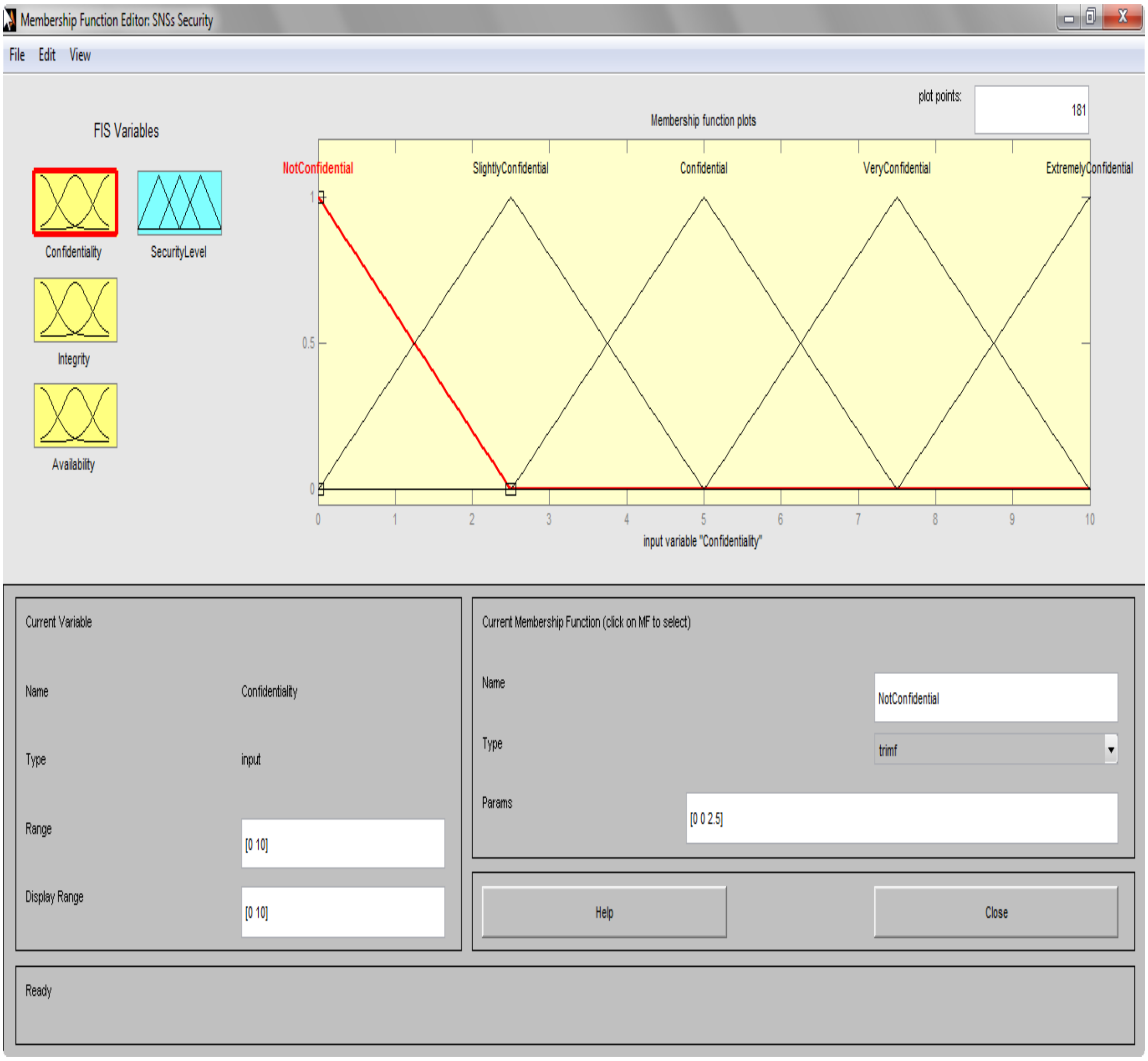
Range: [0 30]

Opening Membership Function Editor

Figure 4.6: MATLAB FIS Editor

4.3.2 The MATLAB Membership Function Editor

The membership function editor shows a plot of highlighted input or output variable along their possible ranges and against the probability of occurrence. Figure 4.7 shows the linguistic variable, *confidentiality* highlighted and displaying its membership functions.



plot points: 181

Current Variable

Name: Confidentiality

Type: input

Range: [0 10]

Display Range: [0 10]

Current Membership Function (click on MF to select)

Name: NotConfidential

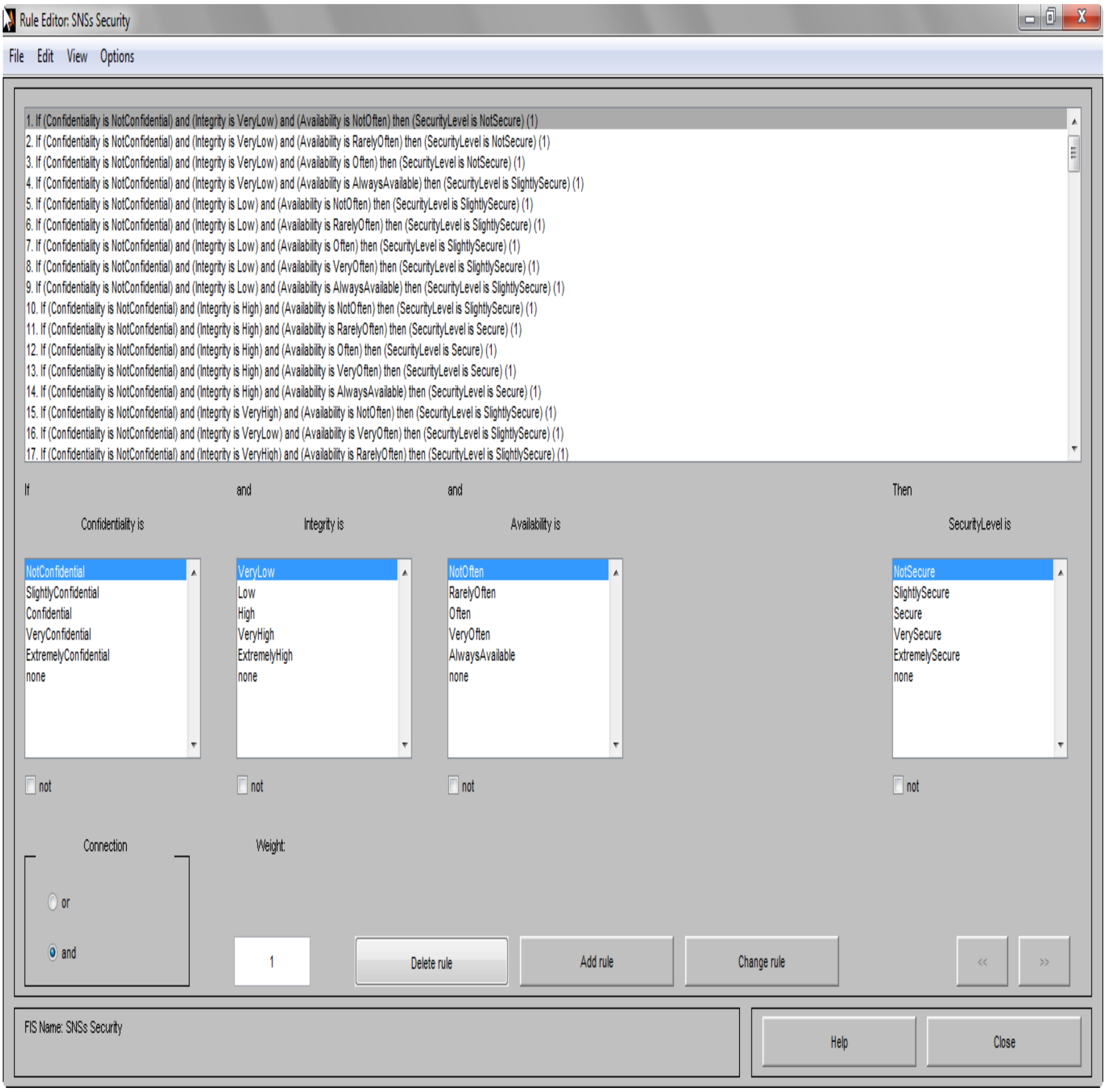
Type: trimf

Params: [0 0 2.5]

Figure 4.7: MATLAB Membership function editor

4.3.3 The MATLAB Rule Editor

The rule editor was used to add, delete and change the rules. It could also be used to change the connection type and the weight of a rule. The rule editor for the application as shown in figure 4.8 had 125 rules in the rule base of the fuzzy system.

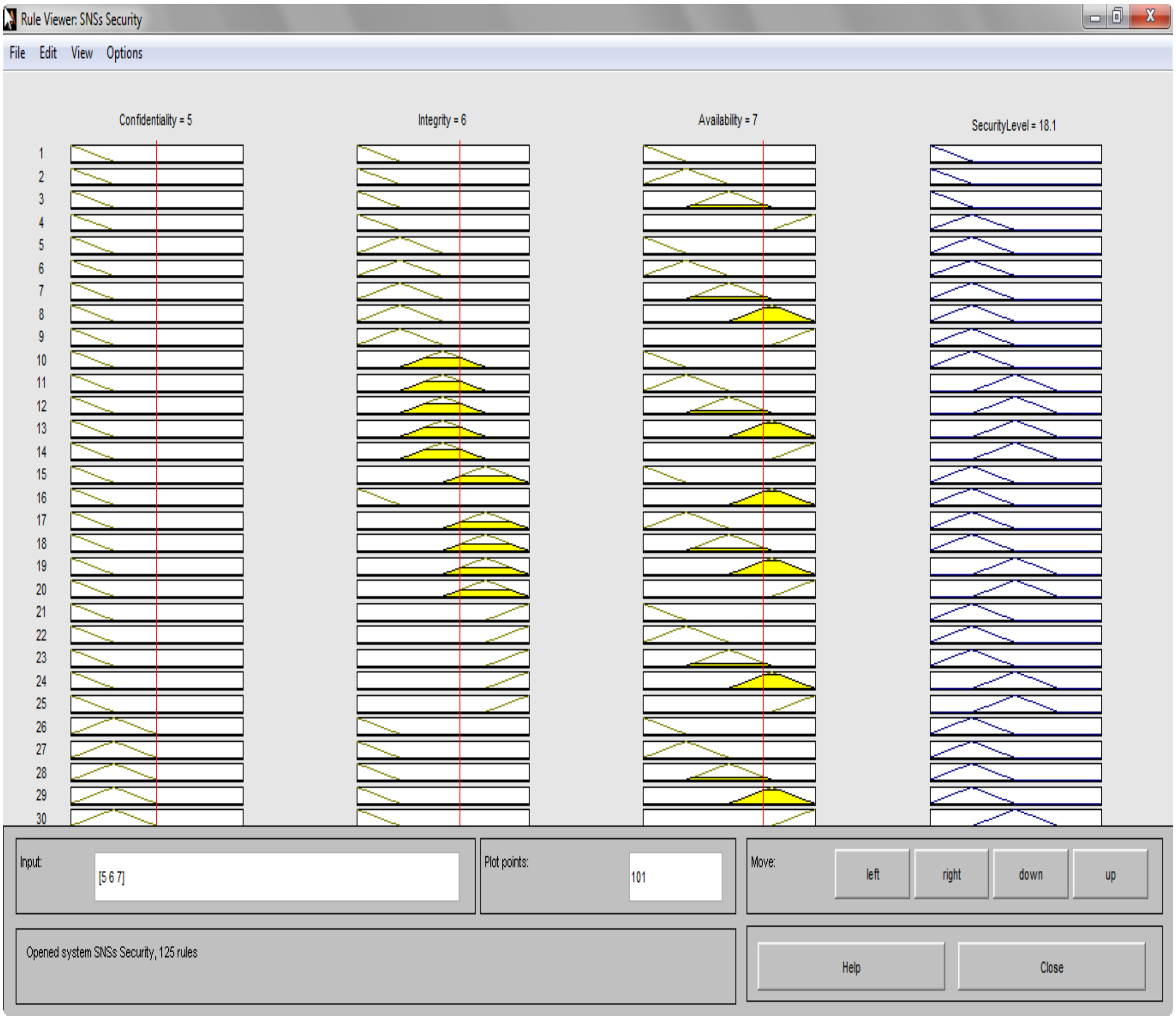


```
1. If (Confidentiality is NotConfidential) and (Integrity is VeryLow) and (Availability is NotOften) then (SecurityLevel is NotSecure) (1)
2. If (Confidentiality is NotConfidential) and (Integrity is VeryLow) and (Availability is RarelyOften) then (SecurityLevel is NotSecure) (1)
3. If (Confidentiality is NotConfidential) and (Integrity is VeryLow) and (Availability is Often) then (SecurityLevel is NotSecure) (1)
4. If (Confidentiality is NotConfidential) and (Integrity is VeryLow) and (Availability is AlwaysAvailable) then (SecurityLevel is SlightlySecure) (1)
5. If (Confidentiality is NotConfidential) and (Integrity is Low) and (Availability is NotOften) then (SecurityLevel is SlightlySecure) (1)
6. If (Confidentiality is NotConfidential) and (Integrity is Low) and (Availability is RarelyOften) then (SecurityLevel is SlightlySecure) (1)
7. If (Confidentiality is NotConfidential) and (Integrity is Low) and (Availability is Often) then (SecurityLevel is SlightlySecure) (1)
8. If (Confidentiality is NotConfidential) and (Integrity is Low) and (Availability is VeryOften) then (SecurityLevel is SlightlySecure) (1)
9. If (Confidentiality is NotConfidential) and (Integrity is Low) and (Availability is AlwaysAvailable) then (SecurityLevel is SlightlySecure) (1)
10. If (Confidentiality is NotConfidential) and (Integrity is High) and (Availability is NotOften) then (SecurityLevel is SlightlySecure) (1)
11. If (Confidentiality is NotConfidential) and (Integrity is High) and (Availability is RarelyOften) then (SecurityLevel is Secure) (1)
12. If (Confidentiality is NotConfidential) and (Integrity is High) and (Availability is Often) then (SecurityLevel is Secure) (1)
13. If (Confidentiality is NotConfidential) and (Integrity is High) and (Availability is VeryOften) then (SecurityLevel is Secure) (1)
14. If (Confidentiality is NotConfidential) and (Integrity is High) and (Availability is AlwaysAvailable) then (SecurityLevel is Secure) (1)
15. If (Confidentiality is NotConfidential) and (Integrity is VeryHigh) and (Availability is NotOften) then (SecurityLevel is SlightlySecure) (1)
16. If (Confidentiality is NotConfidential) and (Integrity is VeryLow) and (Availability is VeryOften) then (SecurityLevel is SlightlySecure) (1)
17. If (Confidentiality is NotConfidential) and (Integrity is VeryHigh) and (Availability is RarelyOften) then (SecurityLevel is SlightlySecure) (1)
```

Figure 4.8: MATLAB Rule editor

4.3.4 The MATLAB Rule Viewer

The three input variables (confidentiality, integrity and availability) needed to be fed into the system were supplied through the text box captioned *input*. The appropriate input corresponds to the weighted averages of the user responses in the questionnaire for each of the input variables followed appropriately by their center of gravity. For example, in the **figure 4.9**, the input values for the variables confidentiality, integrity and availability are respectively **[5, 6, 7]** and the corresponding output (security level) is 18.1, which is shown at the top of the corresponding graphs. The input for each of the input variables is specified at the top of the section corresponding to them, so also is the output variable.



Confidentiality = 5

Integrity = 6

Availability = 7

SecurityLevel = 18.1

Input: [5 6 7]

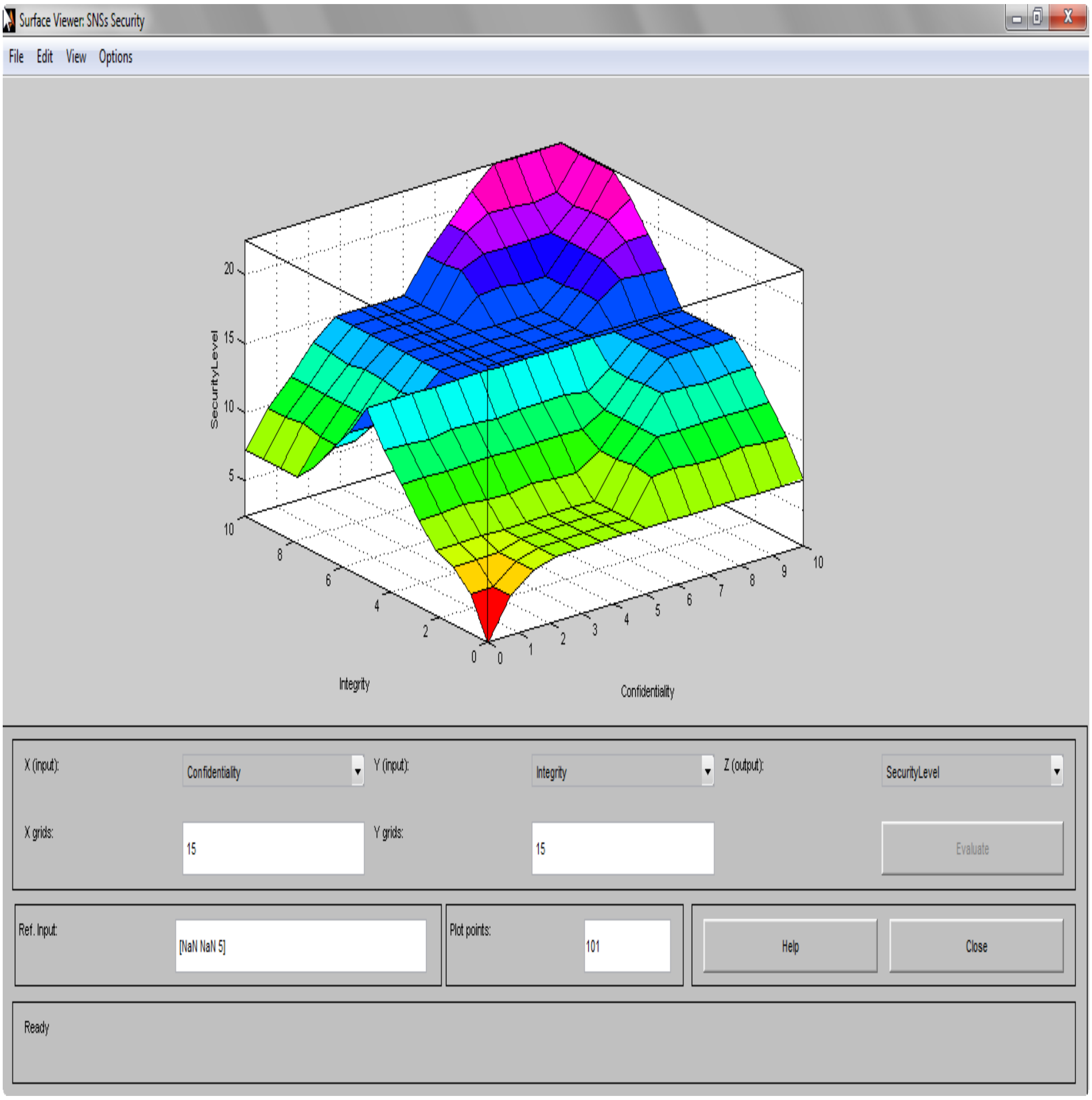
Plot points: 101

Opened system SNSs Security, 125 rules

Figure 4.9: MATLAB Rule Viewer

4.3.5 The Surface Viewer

The MATLAB surface viewer as shown in figure 4.10 is a 3-D graph that shows the relationship between the inputs and the output. The output (security Risk) is represented on the Z-axis while 2 of the inputs (Confidentiality and Integrity) are on the x and y axes and the other input (Availability) is held constant. The surface viewer shows a plot of the possible ranges of the input variables against the possible ranges of the output.



X (input): Confidentiality

Y (input): Integrity

Z (output): SecurityLevel

X grids: 15

Y grids: 15

Evaluate

Ref. Input: [NaN NaN 5]

Plot points: 101

Help

Close

Ready

Figure 4.10: MATLAB Surface Viewer

4.4 Evaluation

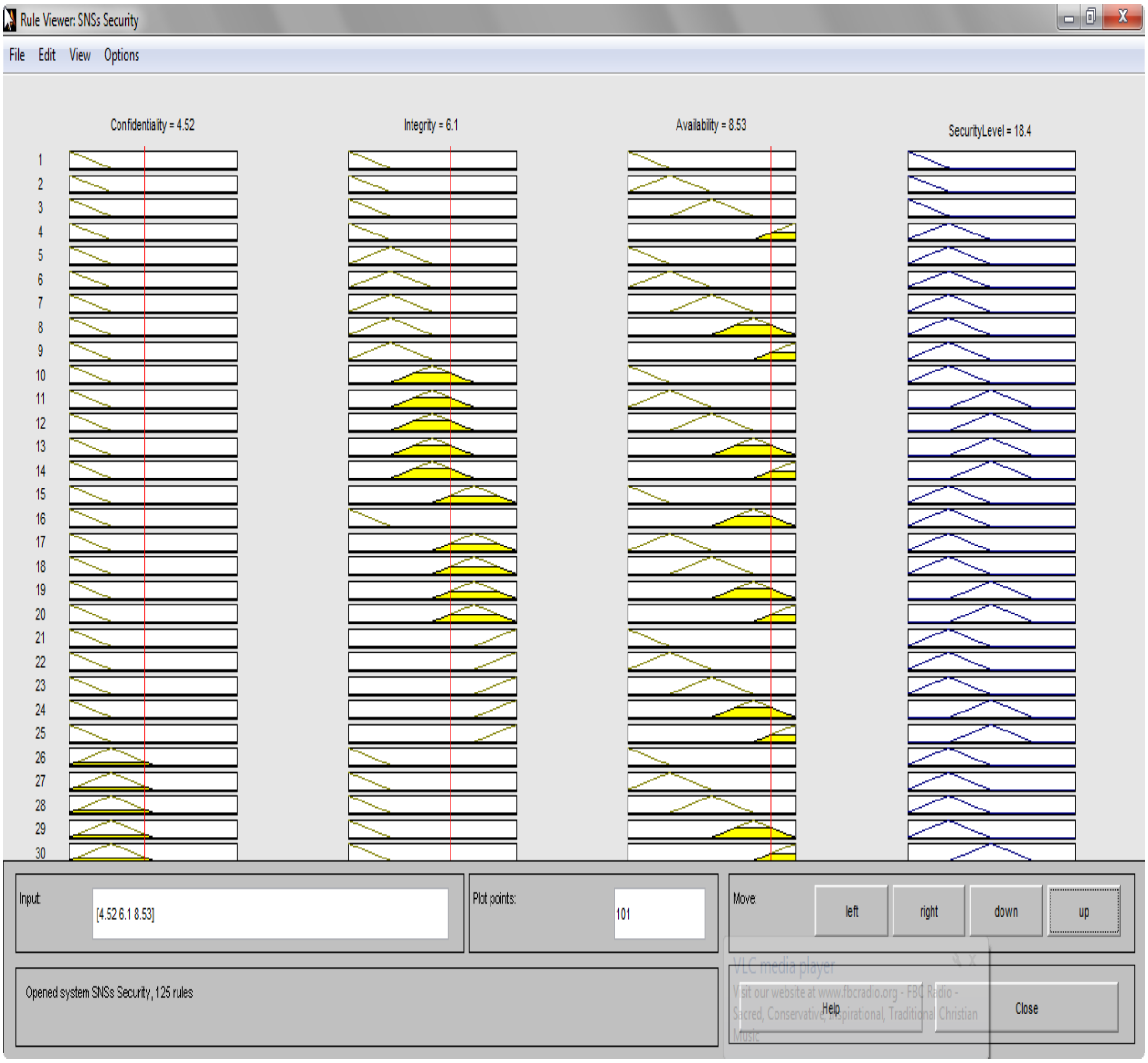
After aggregating the user responses for the three security criteria variables and respectively for the three selected online social media sites, Table 4.10 shows the security risk analysis system as evaluated for the SNSs. The output determines the security level as perceived by users per their responses for of each the social networking site (SNSs). The summary of the evaluation is given in Table 5. The input and the output for the three SNSs are shown in Table 4.11 below.

SNSs	Variable inputs	Crisp Output	Significance	Security Level
Facebook	[4.52, 6.10, 8.53]	18.4	40% slightly secure, 60% secure	61.33%
Twitter	[6.30, 7.42, 8.21]	22.4	25% secure, 75% very secure	74.67%
LinkedIn	[3.98, 2.70, 7.70]	14.9	45% slightly secure, 55% secure	49.67%

Table 4.11: Evaluation of Input Variables

4.5 User Security Perception of Facebook

Facebook was rated according to user responses with 4.52 as the score for confidentiality, 6.10 for integrity and 8.53 as the score for availability. This produced a crisp output of 18.4 representing the security level out of the set range of 30. On a scale of 30, this value of 18.4 shows Facebook is 61.33% secure. On the fuzzy sets defined for the security risk level, this value corresponds to around 40% slightly secure and 60% secure. The significance is that if the management of Facebook has set the site's security to a minimum of 70%, users feel that they are not yet secured and therefore more tightening must be made to get users of Facebook adequately secured.



Confidentiality = 4.52

Integrity = 6.1

Availability = 8.53

SecurityLevel = 18.4

Input: [4.52 6.1 8.53]

Plot points: 101

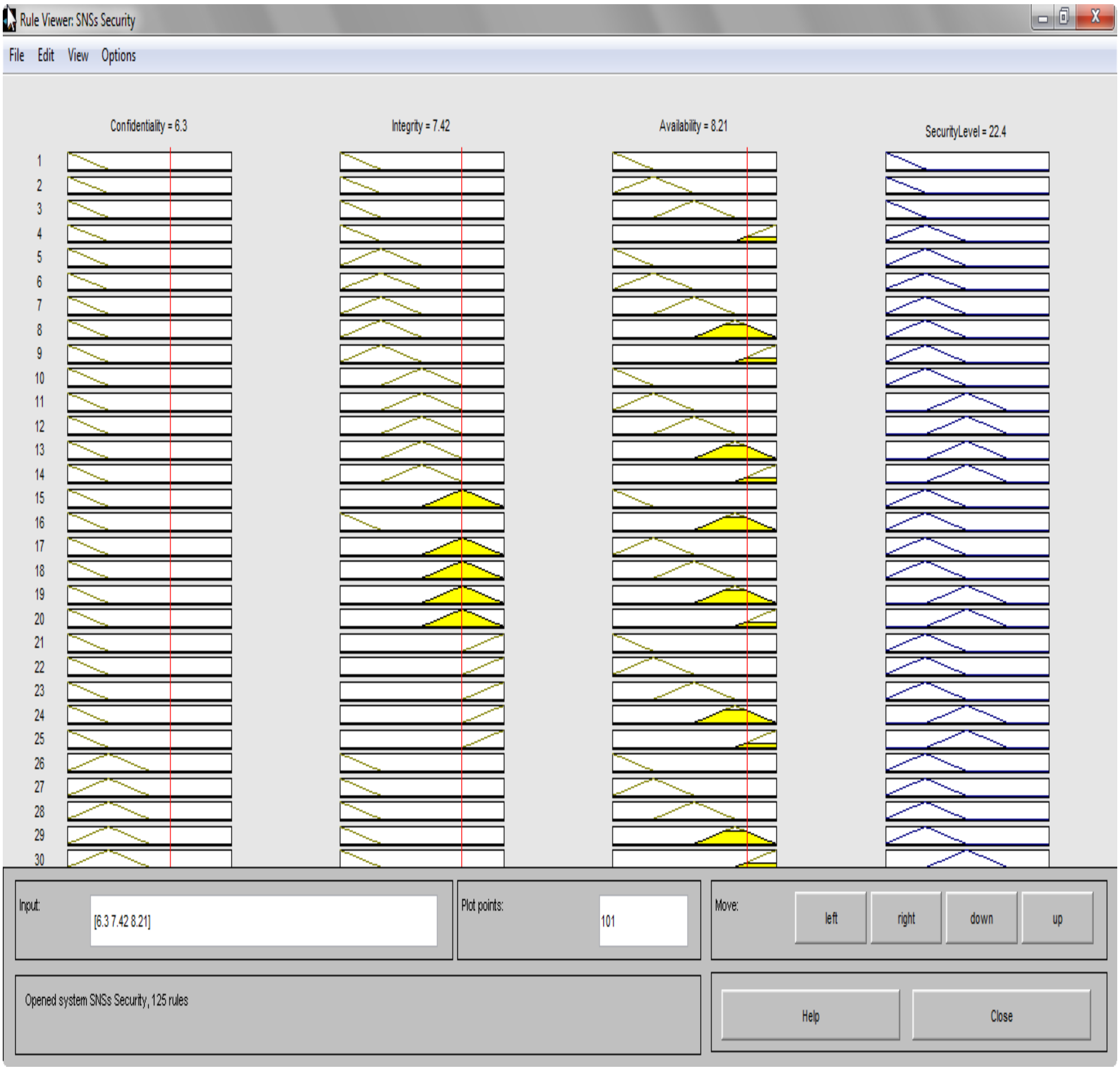
Opened system SNSs Security, 125 rules

Rule Viewer: SNSs Security
File
Edit
View
Options
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
Move:
left
right
down
up
Help
Close
VLC media player
Visit our website at www.fbcradio.org - FBC Radio - Sacred, Conservative, Inspirational, Traditional Christian Music

Figure 4.11: Rule Viewer Security Level for Facebook

4.6 User Security Perception of Twitter

Twitter scored inputs of 6.3, 7.42 and 8.21 for confidentiality, integrity and availability respectively. The crisp output was 22.4. This value corresponded to 25% secure and 75% very secure on the fuzzy set scale for the output. This implies that user view Twitter to be 74.67% secure. Assuming that Twitter as an organization has set 70% as the minimum standard for the site’s security, it can be deduced that from their users’ point of view, security is relatively stable or enough for their numerous users.



Input: [6.3 7.42 8.21]

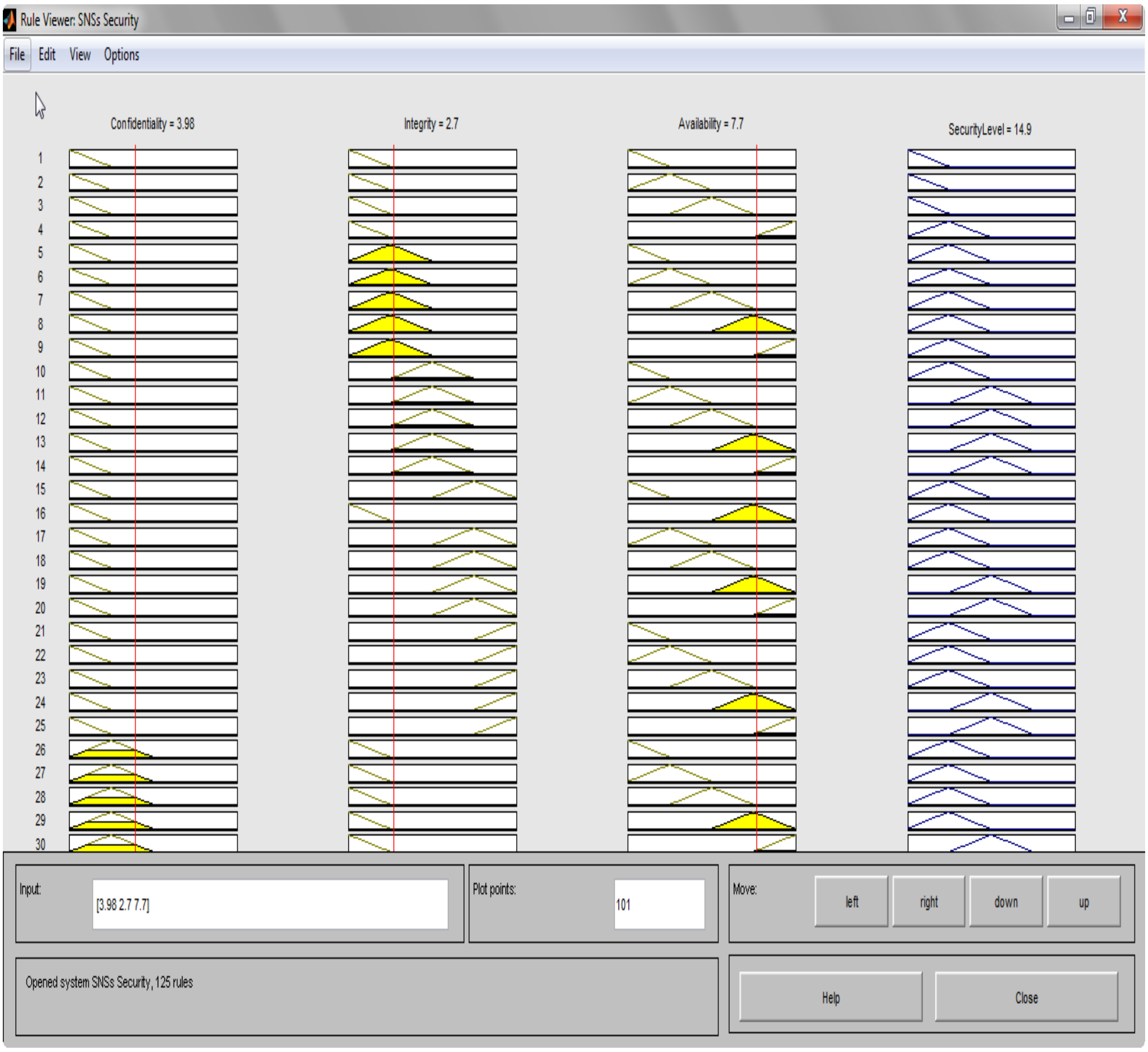
Plot points: 101

Opened system SNSs Security, 125 rules

Figure 4.12: Rule Viewer Security Level for Twitter

4.7 User Security Perception of LinkedIn

LinkedIn was judged with scores of inputs of 3.98, 2.70 and 7.70 for confidentiality, integrity and availability respectively. The crisp output was 14.9. This value corresponded to 45% slightly secure and 55% secure on the scale of fuzzy set output. The significance is that users perceive LinkedIn to be only 49.67% secure. Again if LinkedIn's minimum security standard is 70%, it would have fallen short of its own standard judging by how users rate the security of the website and its content. The organization can rely on this rating of their website to improve upon security.



Input: [3.98 2.7 7.7]

Plot points: 101

Opened system SNSs Security, 125 rules

Figure 4.13: Rule Viewer Security Level for Twitter

CHAPTER 5
Conclusion and Recommendations

5.1 Introduction

Users are by far the main building block of any online social networking site and therefore their security and privacy should be of utmost concern to mangers of these social media sites. One user complaint, user perception, is an important element when considering the concepts of social networking security. Conventional methods for analyzing social networking site user perception of security have limitations and do not fully explain the user perception phenomena. The perception processes of humans cannot be analyzed and assessed by a binary approach or in a simple quantitative way. The human thought process is subjective, imprecise and complicated, and human perception usually uses a linguistic approach, as opposed to a numerical approach, to classify, describe, or "value" a system. In addition, user perception of security risk on social networking site is solely affected by an individual evaluator's needs and requirements of what would make him or her secured on a social networking site.

In this study, a fuzzy system was implemented using fuzzy logic theory to evaluate user perception of security risk on social networking sites. Facebook, Twitter and LinkedIn were used as case studies for this research. Employing MATLAB and its associated fuzzy logic toolbox to design the Fuzzy Inference System, an overall user perception of security risk on SNSs were realized.

5.2 Implications of study

The findings of this study allow for the research to conclude, with certainty, the assertion that most users of social networking sites feel insecure giving the many issues of identity breaches. Through this research there has been an attempt to design a system that can be used to evaluate the security risk associated with the use of social media from the point of view of those that matter, the users. This will definitely help management of these social networking sites to beef up security on their websites from time to time when a thorough study such as this is employed to adequately get users views and perceptions of security. The result of this study points to the fact that if social networking sites would incorporate security risk analysis into the design and maintenance of their systems, the issue of insecurity by users would be minimized if not eliminated. Finally, security risk analysis is a useful way towards designing secured information systems and must therefore be considered a significant activity by software companies especially those that have direct effects on users such as social media sites applications.

Limitations

At this stage, it may seem that fuzzy logic is the right answer to the analysis of subjective and imprecise concepts such as user security because most evaluation studies produce simple statistically aggregated results without considering the variety of user opinions. Through statistically aggregated method of evaluation, subjective data and user opinion cannot be adequately analyzed.

In spite of the above, this particular study had some limitations in the design of the evaluation of user perception of security on social networking sites as follows:

- The input variables (confidentiality, integrity and availability) were all given the same weight since in the evaluation of information systems, standardized bodies such ITSEC (the European standard), CTCPEC (the Canadian standard) and TCSEC (The Department of Defense, US standard) all agree that the three criteria of confidentiality, integrity and availability are all equally important. However, the research could have allowed users to rank these criteria; the ones they feel are important to them and then assign appropriate weights to them.
- Whiles the 221 respondents are adequately representative, the research could have covered more users than was done in this research.
- The questionnaire that was used to create the fuzzy membership functions and their ranges using the interval estimation method was answered only by about one-third of the total respondents for the entire survey**.** Though their results were used, it cannot be conclusive enough giving the number of users who answered it.
- Experts' opinions are very important when it comes to the design of most fuzzy systems but for this research, no experts' opinions were used. In the future if there are experts in the area of social networking sites use, it would be interesting to compare their opinions to user opinions.

Future Developments

It might be necessary in the near future to redesign this system to still be efficient without the use of MATLAB fuzzy toolbox. It might be necessary to program everything in C or C++ language to allow for more customization. In the future research, experts' opinions if any, can be incorporated into the design or at least compare their opinions with user opinions.

BIBLIOGRAPHY

[1]. Nielsen Online**,** "Social Networks/Blogs Now Account for One in Every Four and a Half Minutes Online" [internet]. 2010 June 5 [cited on 2011-09-13]; Available from: http://blog.nielsen.com/nielsenwire/online_mobile/social-media-accounts-for-22-percent-of-time-online/

[2]. Nielson Online, News Release: "Time Spent on Facebook up 700 Percent, but MySpace.com Still Tops for Video" [internet] 2009 June 2 [cit. 2011-09-13] Available from: http://www.nielsen-online.com/pr/pr_090602.pdf

[3]. Facebook, "Statistics-Facebook", [internet] 2011 July 8 [cit. 2011-09-15] Available from http://www.facebook.com/press/info.php?statistics (2011)

[4]. Alexa- Web Information Company, "Facebook.com Site Info", [internet] 2011 August 8 Available from: http://www.alexa.com/siteinfo/facebook.com (2011) [cit. 2011-07-09]

[5]. Accuracy in media -US, "threat-of-cyber-crime-continues-to-increase" [internet] 2011 August 8 [cit. 2011-09-16] Available from: http://www.aim.org/guest-column/threat-of-cyber-crime-continues-to-increase/

[6]. S. Shah, "Measuring Operational Risks using Fuzzy Logic Modeling, "Article, Towers Perrin, JULY 2003.

[7]. Dahal K. P, Hussain K P and Hossain M A (2005): "Loan Risk Analyzer based on Fuzzy Logic ", *Proceedings of IEEE International Conference on E-Technology,E-Commerce and E-Service, IEEE Computer Society Press, pp 363-366*, Hong Kong, Hong Kong.

[8]. Maxwell Dondo, A Fuzzy Risk Calculations Approach for a Network Vulnerability Ranking System, Technical Memorandum 2007-090, Defence R&D Canada – Ottawa, May 2007, http://www.ottawa.drdc-rddc.gc.ca/docs/e/TEO-TM-2007-090.pdf [cit. 2011.11.05]

[9]. Samir Shah, Measuring Operational Risk Using Fuzzy Logic Modeling, International Risk Management Institute, Inc. (IRMI), September 2003, [cit. 2011-10-11] Available from: http://www.irmi.com/Expert/Articles/2003/Shah09.aspx

[10]. Sodiya, A. S., Onashoga, S. A., and Oladunjoye, O. B. Towards building secure software products. *Journal of Issues in Informing Science and Information Technology, U. S. A*, 13:635.646, 2006.

[11]. Sodiya, A. S., Longe H. O. D, .Fasan O. M. Software security risk analysis using fuzzy expert system "Article, JULY 2008.

[12]. William McGill, Bilal M. Ayyub, Multicriteria Security System Performance Assessment Using Fuzzy Logic, The Journal of Defense Modeling and Simulation

(JDMS): Applications, Methodology, Technology, Special Issue: Homeland Security, October 2007, vol. 4, no. 4, http://www.scs.org/pubs/jdms/vol4num4/McGill.pdf

[13]. Smith, E. and Eloff, J. Transaction based risk analysis using cognitive fuzzy techniques. *Advances in Information Security Management&Small Systems Security*, 2001.

[14].Smith, E. and Eloff, J. Cognitive fuzzy modeling for enhanced risk assessment in a health care institution. *IEEE Intelligent systems & their applications*, 15(2), 2000.

[15]Donath, J., 2007. Signals in social supernets. Journal of Computer-Mediated Communication 13 (1), 231–251.

[16]Boyd, D., 2008a. Facebook’s Privacy Trainwreck: exposure, invasion, and social convergence. Convergence 14 (1), 13–20.

[17] Boyd, D., 2008b. Facebook’s Privacy Trainwreck. Convergence 14 (1), 13–20.

[18] M. Srivatsa, D. Agrawal, and S. Reidt. A metadata calculus for secure information sharing. In *CCS’09*, IL, USA, 2009.

[19] C. Dwyer, S.R Hiltz, K. Passerini “Trust and Privacy: A Comparison of Facebook and MySpace” [internet] 2007-04-12 [cit. 2011-1023]http://csis.pace.edu/~dwyer/research/DwyerAMCIS2007.pdf

[20]. Mayer, R. C., J. H. Davis, and F. D. Schoorman (1995) “An Integrative Model of Organizational Trust,” The Academy of Management Review (20) 3, pp. 709-734.

[21]. Fukuyama, F. (1995) Trust: The Social Virtues and the Creation of Prosperity. New York, NY: Simon & Schuster, Inc.

[22]. Lewis, J. D. and A. Weigert (1985) “Trust as a Social Reality,” Social Forces (63) 4, pp. 967-985.

[23]. Chiaramonte, P. and E. Martinez (2006) “Jerks In Space,” in The New York Post, pp. 6. New York

[24]. Hass, N. (2006) “In Your Facebook.com,” in The New York Times, pp. 30-31. New York. http://www.nytimes.com/2006/01/08/education/edlife/facebooks.html?pagewanted=all

[25] 22. S. Bosworth and M.E. Kabay, “Computer Security Handbook”, John Wiley & Sons, Fourth Edition, 2002

[26]. Bhaiji, Y (2008) “Network Security Technologies and Solutions”, A comprehensive, all-in-one reference for Cisco network security. Pg 8-9. Cisco Press.

[27]. NIST (2010): "Loan Risk Analyzer based on Fuzzy Logic ", *Proceedings of IEEE International Conference on E-Technology,E-Commerce and E-Service, IEEE Computer Society Press, pp 363-366*, Hong Kong, Hong Kong.

[28] ITSEC "Information Technology Security Evaluation Criteria": Department of Trade and Industry, London, [internet] 1991 June 12. [cit. 2011-10-28] Available at: http://www.iwar.org.uk/comsec/resources/standards/itsec.htm.

[29]. Wikipedia (2006), "Data confidentiality" [cit. 2011-10-14] Available from: http://en.wikipedia.org/wiki/Social_networking_service

[30]. Cutillo, L.A., Molva, R., Strufe, T. (2009) "Privacy preserving social networking through decentralization, in Proceedings of the Sixth International Conference on Wireless On-Demand Network Systems and Services" (WONS), pp.145-152.

[31] Microsoft Corporation "Data Security and Data Availability in the Administrative Authority". [cit.2011-10-28] Available from: http://technet.microsoft.com/en-us/library/cc722918.aspx

[32] Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230.

[33] Palfrey & Gasser, (2008), "*Born Digital: Understanding the First Generation of Digital Natives*. Basic Books, 288 pp

[34] Bull, G., Thompson, A., Searson, M., Garofalo, J., Park, J., Young, C., & Lee, J (2008). Connecting informal and formal learning: Experiences in the age of participatory media. *Contemporary Issues in Technology and Teacher Education, 8*(2). http://www.citejournal.org/vol8/iss2/editorial/article1.cfm (accessed on 18 August 2011)

[35] Barnes S. (2006), *A privacy paradox: social networking in the United States*. [Cit.2011-12-13] Available from: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1394/1312

[36] L. Greiner, "Hacking Social Networks," networker, Mar.2009, pp9-11

[37] C. Patsakis, A. Asthenidis, and Chatzidimitriou, "Social Networks as an Attack Platform: Facebook Case Study," Proc. 2009 8th Int'l Conf. Networks (ICN 09), IEEE CS Press, 2009, pp. 245-247.

[38] E. Athanasopoulos et al., "Antisocial Networks: Turning a Social Network into a Botnet," Proc. 11th Int'l Conf. Information Security (ISC 08), LNCS 5222, Springer, 2008, pp. 146-160.

[39] A. Besmer et al., "Social Applications: Exploring a More Secure Framework," Proc. 5th Symp. Usable Privacy and Security (SOUPS 09), ACM Press, 2009, article no. 2.

[40] Hoglund, G. and McGraw, G. Exploiting software, how to break the code. Addison-Wesley publisher, 2004.

[41] Myagmar, A. J., Lee, S. and Yurcik, W. Threat modeling as basis for security requirements. In Symposium on Requirements Engineeing for Information Security (SREIS), 2005.

[42] Leveson, N. G. System safety and computers. *Addison-Wesley publisher,* 1995.

[43] Sheyner, O. and Wing, J. Tools for generating and analyzing attack graphs. In Proceedings of formal methods for component and Objects, 2005.

[44] Sodiya, A. S., Onashoga, S. A., and Oladunjoye, O. B. Towards building secure software products. Journal of Issues in Informing Science and Information Technology, U. S. A, 13:635.646, 2006.

[45] Smith, E. and Eloff, J. Cognitive fuzzy modeling for enhanced risk assessment in a health care institution. IEEE Intelligent systems & their applications, 15(2), 2000.

[46] Smith, E. and Eloff, J. Transaction based risk analysis using cognitive fuzzy techniques. Advances in Information Security Management & Small Systems Security, 2001.

[47] J.C. Bennett, G.A. Bohoris, E.M. Aspinwall, R.C. Hall, Risk analysis techniques and their application to software development, European Journal of Operational Research 95 (1996) 467– 475.

[48]] D. White, Application of systems thinking to risk management: a review of the literature, Management Decision 3 (10) (1995) 35– 45.

[49] R.K.J.R. Rainer, C.A. Snyder, H.H. Carr, Risk analysis for information technology, Journal of Management Information Systems 8 (1) (1991) 129– 147.

[50] B.W. Boehm, Software Risk Management, IEEE Computer, Society Press, Washington, DC, 1989.

[51] Fuzzy Sets. Zadeh, L. A. Berkeley : University of California, 1965, Information and Control, Vol. 8, pp. 338 - 353.

[52] W.G. de Ru, J.H.P. Eloff, Risk analysis modeling with the use of fuzzy logic, Computer Security 15 (3) (1996) 239– 248.

[53]] D.R. Moscato, Database gateway processor risk analysis using fuzzy logic, Information Management and Computer Security 6 (3) (1998) 138–144.

[54] H.M. Lee, Applying fuzzy set theory to evaluate the rate of aggregative risk in software development, Fuzzy Sets and Systems 79 (3) (1996) 323–336.

[55] J.B. Bowles, C. Pelaez, Enrique fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis, Reliability Engineering and Systems Safety 50 (2) (1995) 203–213.

[56] J.H.M. Tah, V. Carr, A proposal for construction project risk assessment using fuzzy logic, Construction Management & Economics 18 (2000) 491–500.

[57] E.N. Wirba, J.H.M. Tah, R. Howes, Risk interdependencies and natural language computations, Engineering Construction and Architectural Management 3 (4) (1996) 251–269.

[58] P.C. Pandey, S.V. Barai, Sensitivity-based weighted-average in structural damage assessment, Journal of Performance of Constructed Facilities 8 (4) (1994) 243–263.

[59] T.J. Ross, H.C. Sorensen, S.J. Savage, J.M. Carson, DAPS: expert system for structural damage assessment, Journal of Computing in Civil Engineering 4 (4) (1990) 327– 348.

[60] J.H.M. Tah, V. Carr, A proposal for construction project risk assessment using fuzzy logic, Construction Management & Economics 18 (2000) 491–500.

[61] C. Huang, Fuzzy risk assessment of urban natural hazards, Fuzzy Sets and Systems 83 (2) (1996) 271–282.

[62] S. Bonvicini, P. Leonelli, G. Spadoni, Risk analysis of hazardous materials transportation: evaluating uncertainty by means of fuzzy logic, Journal of Hazardous Materials 62 (1) (1998) 59– 74.

[63] Y.W. Lee, M.F. Dahab, I. Bogardi, Fuzzy decision making in ground water nitrate risk management, Water Resources Bulletin 30 (1) (1994) 135– 148.

[64] Tsuen, H. H., Feng-Chuan, P., and Kuo-Chien, C. (2004)” Using fuzzy logic in the evaluation of customer perceived value on healthcare services” MCDM 2004,

[65] Moraes, O.B., and Abiko A.K.(2007), Dweller perception using fuzzy logic for slum upgrading. Institution of Civil Engineers (ICE) Proceedings of the Institution of Civil Engineers Municipal Engineer 161(1) (2007) 151-161.

[66] Dongmin L., A generalized approach for analyzing Transportation User perception using fuzzy sets [PhD thesis]. Pennsylvania: Pennsylvania State University; 2007.

[67] F. Gardin, R. Power, E. Martinelli, Liquidity management with fuzzy qualitative constraints, Decision Support Systems 15 (1995) 147– 156.

[68] S.H. Tsaur, G.H. Tzeng, K.C. Wang, Evaluating tourist risks from fuzzy perspectives, Annals of Tourism Research 24 (4) (1997) 796– 812.

[69] Zadeh, L. A. (1965_b) “Fuzzy Sets”. Information and Control, 8(3), pp. 338-353.

[70] Aburrous, Maher Ragheb, Hossain, M.A., Thabatah, Fadi and Dahal, Keshav (2008) *Intelligent Quality Performance Assessment for E-Banking Security using Fuzzy Logic.* In: Fifth International Conference on Information Technology: New Generations, 2008. ITNG 2008. , 7-9 April 2008, Las Vegas, Nevada.

[71] K. Haslum, A. Abraham, and S. Knapskog (2008). *Fuzzy online risk assessment for distributed intrusion prediction and prevention systems*. In Tenth International Conference on Modeling and Simulation, IEEE Computer Society press, volume I, pp 1-12 2008.

[72] A. Abraham, C. Grosan, H. Liu andY. Chen (2009) Hierarchical Takagi-Sugeno Models for Online Security Evaluation Systems, In Fifth International Conference on Information Assurance and Security, IEEE Computer Society press, volume I, pp 1-6, 2009.

[73] A. Abraham, C. Grosan, H. Liu andY. Chen (2009) Hierarchical Takagi-Sugeno Models for Online Security Evaluation Systems, In Fifth International Conference on Information Assurance and Security, IEEE Computer Society press, volume I, pp 1-6, 2009.

[74] Chennakesava R. Alavala (2008) "Fuzzy Logic and Neural Networks: Basic Concepts and Applications "New Age International Publishers Ltd pp 6-9, 2008.

[75] Lucero, Yvonne C. and Nava, Patricia A. El Paso. A fuzzy method for automatic generation of membership function using fuzzy relations from training examples: University of Texas at El Paso, 2002. 0-7803-7461-4.

[76] Zadeh, L. A. (1975$_a$) "Fuzzy logic and approximate reasoning". Synthese, 30(1), pp. 407-428.

[77] John, R. Fuzzy Logic and Knowledge Based Systems. 2006.

[78] George.J., Klir andYuan. Bo, "Fuzzy sets and Fuzzy Logic", 1995 Edition 574pg. ISBN 0-13-101171-5

[79] Jantzen, J**.** *Design of Fuzzy Controllers.* Department of Automation, Technical University of Denmark. s.l. : Technical University of Denmark, 1998. 98-E 864.

[80] Jang, J.-S. R. and C.-T. Sun, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* Prentice Hall, 1997.

[81] Mamdani, E.H. and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[82] Sugeno, M., *Industrial applications of fuzzy control,* Elsevier Science Pub. Co., 1985.

[83] Mamdani, E.H. and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[84] MATLAB, 2010 “Fuzzy Toolbox” [internet] 2010 April 5 [cit. 2012-01-03] Available from: http://www.mathworks.com/products/fuzzy-logic/

[85] Loizos, A., A Simplified Application of Fuzzy Set Theory for the Evaluation of Pavement Roughness. *Road and Transport Research*, Vol. 10, No. 4 2001, pp. 21-32.

[86] Juang, C. H., X. H. Huang, and S. D. Shiff. Determination of weight of criteria for decision making by the fuzzy eigenvector method. *Civil Engineering System,* Vol. 9, 1992, pp. 1-16.

[87] Lee, D., M. T. Pietrucha, and S. K. Sinha, Application of Fuzzy Logic to Evaluate Driver Perception of Variable Message Signs. *Transportation Research Record 1937* TRB, National Research Council, Washington, D.C., 2005, pp.96-104.

Appendix A: Online Questionnaire for Survey

Online Social Networking Site Security Risk Questionnaire

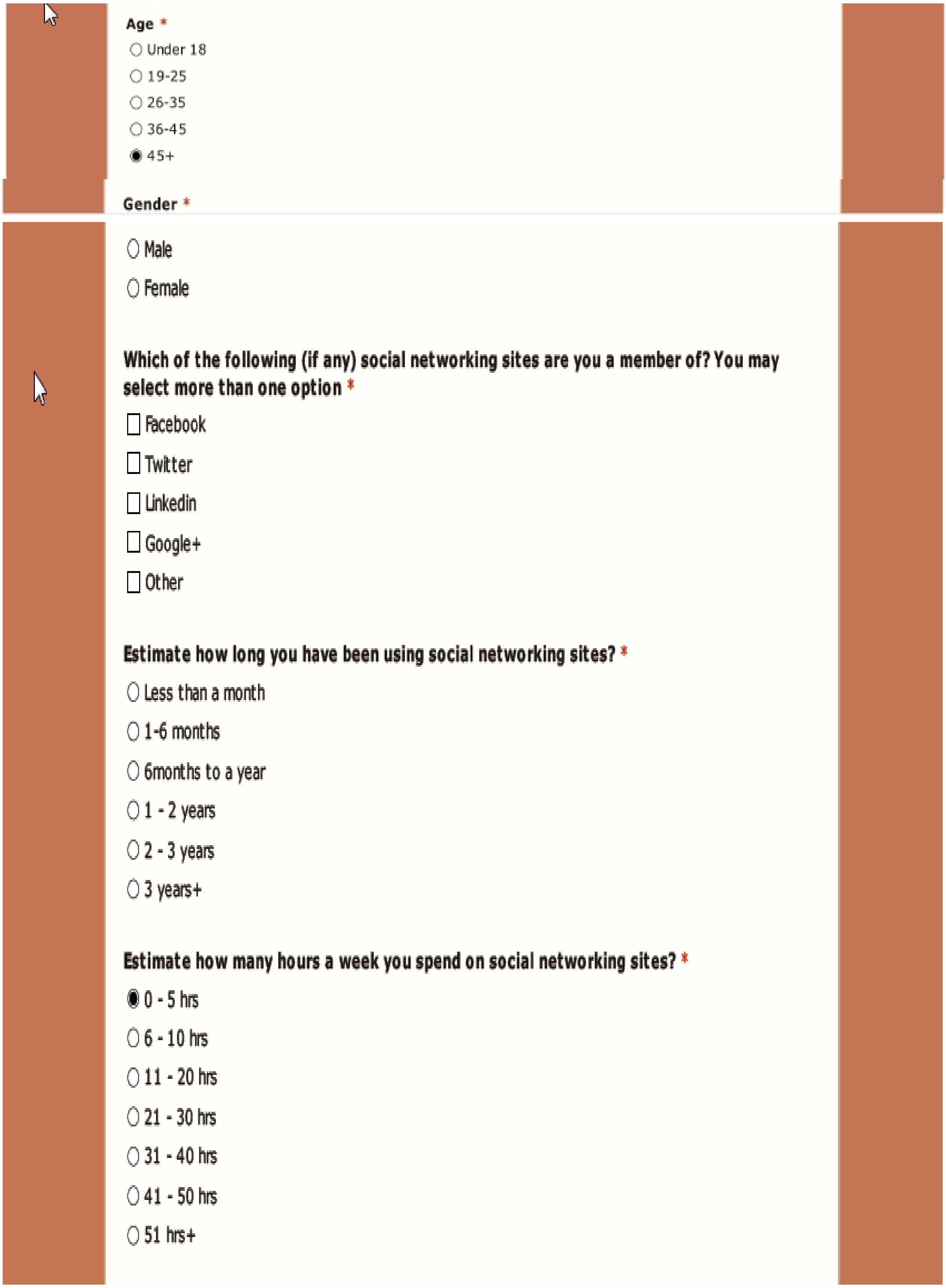
Czech University of Life Sciences-Prague
Department of Information Engineering
Academic Year 2012/2013

TOPIC: Evaluation of User Perception of Security Risk on Social Networking Sites using Fuzzy Logic

Introduction

You are welcome to take part in this survey which is aimed at identifying user perception of security risk on online social networking sites. To participate you must be a member of a Social Networking Site (For the purposes of this research, we focus only on Facebook, Twitter and LinkedIn). Participation in this survey is entirely voluntary and anonymous so you will not be associated with the answers you provide. You can withdraw at any time simply by leaving this page and are free to ignore any questions you prefer not to answer. The results to be obtained from this questionnaire will be used in Czech University of Life Sciences Dissertation for an MSc Degree. Please note that you will not be referred to by any personally identifiable information (i.e. name, email address, etc) in any of the reports. If you have any questions do not hesitate to email me at paadadzie2002@gmail.com and I will get back to you as soon as possible. This survey has been approved by my supervisor, Doc. Ing. Arnost Vesely. Information that will help you answer the following questions

- *Social Network Sites (SNSs) refer to web sites such as Facebook, Twitter, LinkedIn, etc*



Age *

○ Under 18

○ 19-25

○ 26-35

○ 36-45

● 45+

Gender *

○ Male

○ Female

Which of the following (if any) social networking sites are you a member of? You may select more than one option *

□ Facebook

□ Twitter

□ Linkedin

□ Google+

□ Other

Estimate how long you have been using social networking sites? *

○ Less than a month

○ 1-6 months

○ 6months to a year

○ 1 - 2 years

○ 2 - 3 years

○ 3 years+

Estimate how many hours a week you spend on social networking sites? *

● 0 - 5 hrs

○ 6 - 10 hrs

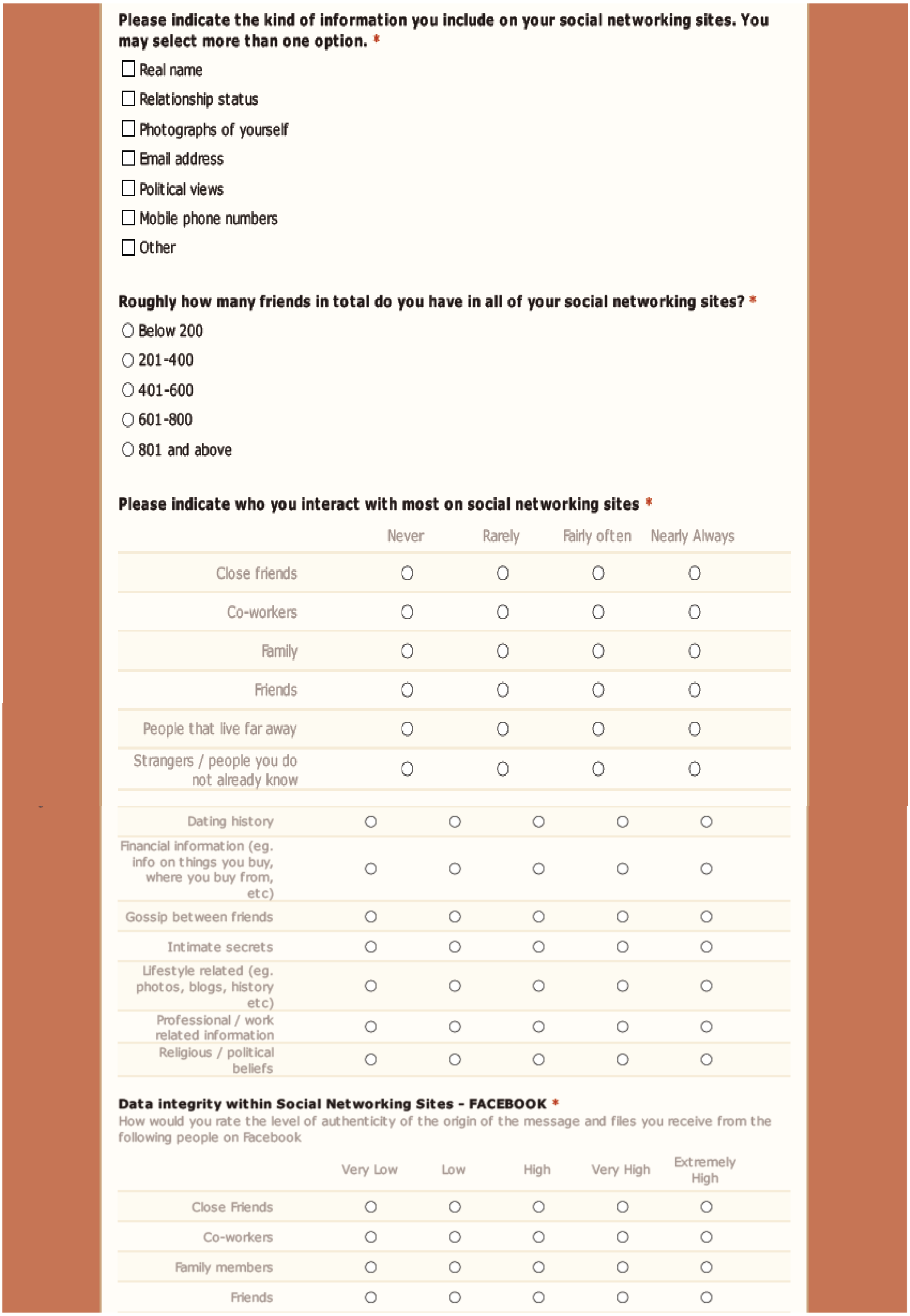
○ 11 - 20 hrs

○ 21 - 30 hrs

○ 31 - 40 hrs

○ 41 - 50 hrs

○ 51 hrs+



Please indicate the kind of information you include on your social networking sites. You may select more than one option. *

□ Real name

□ Relationship status

□ Photographs of yourself

□ Email address

□ Political views

□ Mobile phone numbers

□ Other

Roughly how many friends in total do you have in all of your social networking sites? *

○ Below 200

○ 201-400

○ 401-600

○ 601-800

○ 801 and above

Please indicate who you interact with most on social networking sites *

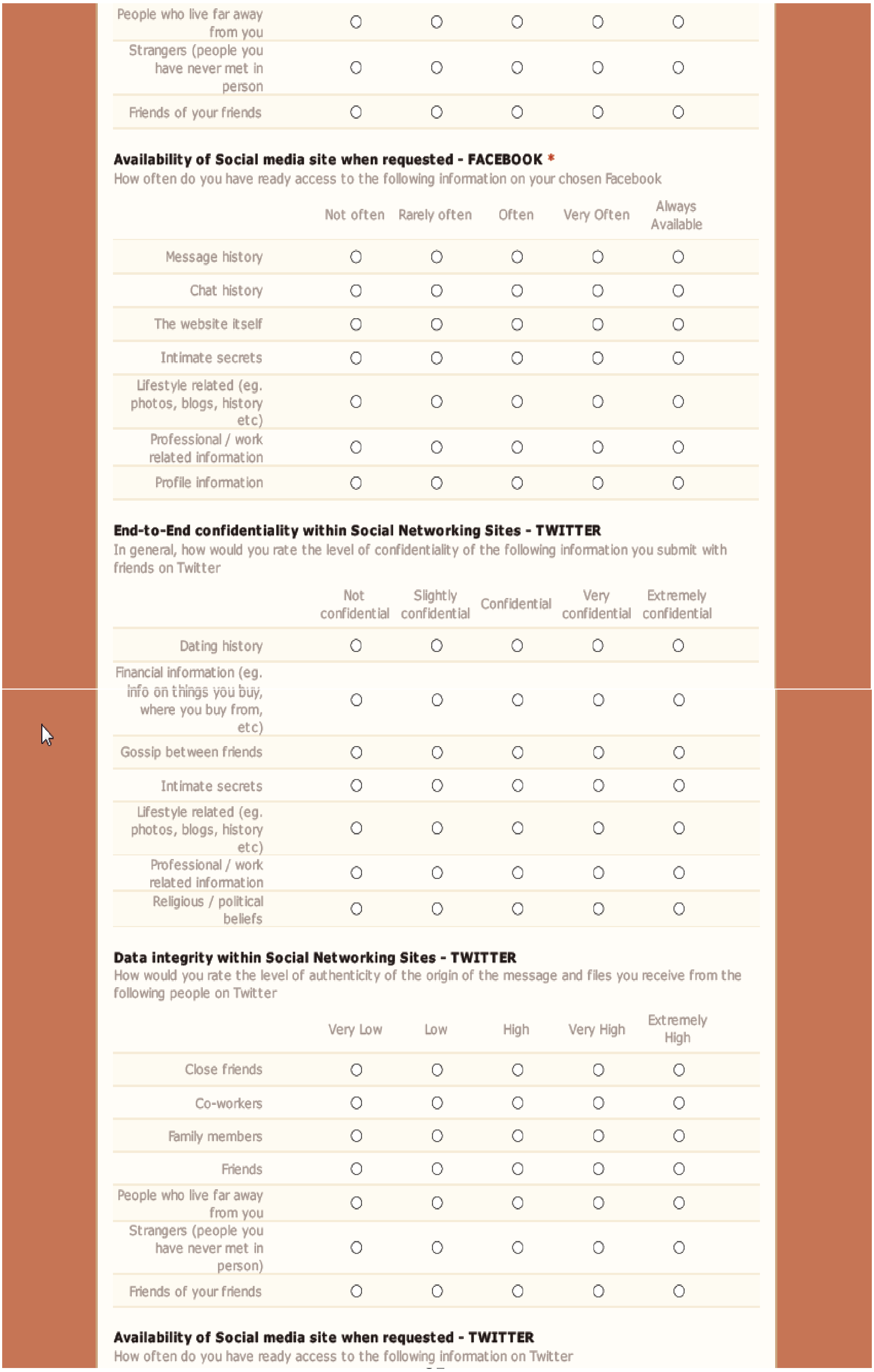
	Never	Rarely	Fairly often	Nearly Always
Close friends	○	○	○	○
Co-workers	○	○	○	○
Family	○	○	○	○
Friends	○	○	○	○
People that live far away	○	○	○	○
Strangers / people you do not already know	○	○	○	○

Dating history	○	○	○	○	○
Financial information (eg. info on things you buy, where you buy from, etc)	○	○	○	○	○
Gossip between friends	○	○	○	○	○
Intimate secrets	○	○	○	○	○
Lifestyle related (eg. photos, blogs, history etc)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Religious / political beliefs	○	○	○	○	○

Data integrity within Social Networking Sites - FACEBOOK *

How would you rate the level of authenticity of the origin of the message and files you receive from the following people on Facebook

	Very Low	Low	High	Very High	Extremely High
Close Friends	○	○	○	○	○
Co-workers	○	○	○	○	○
Family members	○	○	○	○	○
Friends	○	○	○	○	○



People who live far away from you	○	○	○	○	○
Strangers (people you have never met in person	○	○	○	○	○
Friends of your friends	○	○	○	○	○

Availability of Social media site when requested - FACEBOOK *

How often do you have ready access to the following information on your chosen Facebook

	Not often	Rarely often	Often	Very Often	Always Available
Message history	○	○	○	○	○
Chat history	○	○	○	○	○
The website itself	○	○	○	○	○
Intimate secrets	○	○	○	○	○
Lifestyle related (eg. photos, blogs, history etc)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Profile information	○	○	○	○	○

End-to-End confidentiality within Social Networking Sites - TWITTER

In general, how would you rate the level of confidentiality of the following information you submit with friends on Twitter

	Not confidential	Slightly confidential	Confidential	Very confidential	Extremely confidential
Dating history	○	○	○	○	○
Financial information (eg. info on things you buy, where you buy from, etc)	○	○	○	○	○
Gossip between friends	○	○	○	○	○
Intimate secrets	○	○	○	○	○
Lifestyle related (eg. photos, blogs, history etc)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Religious / political beliefs	○	○	○	○	○

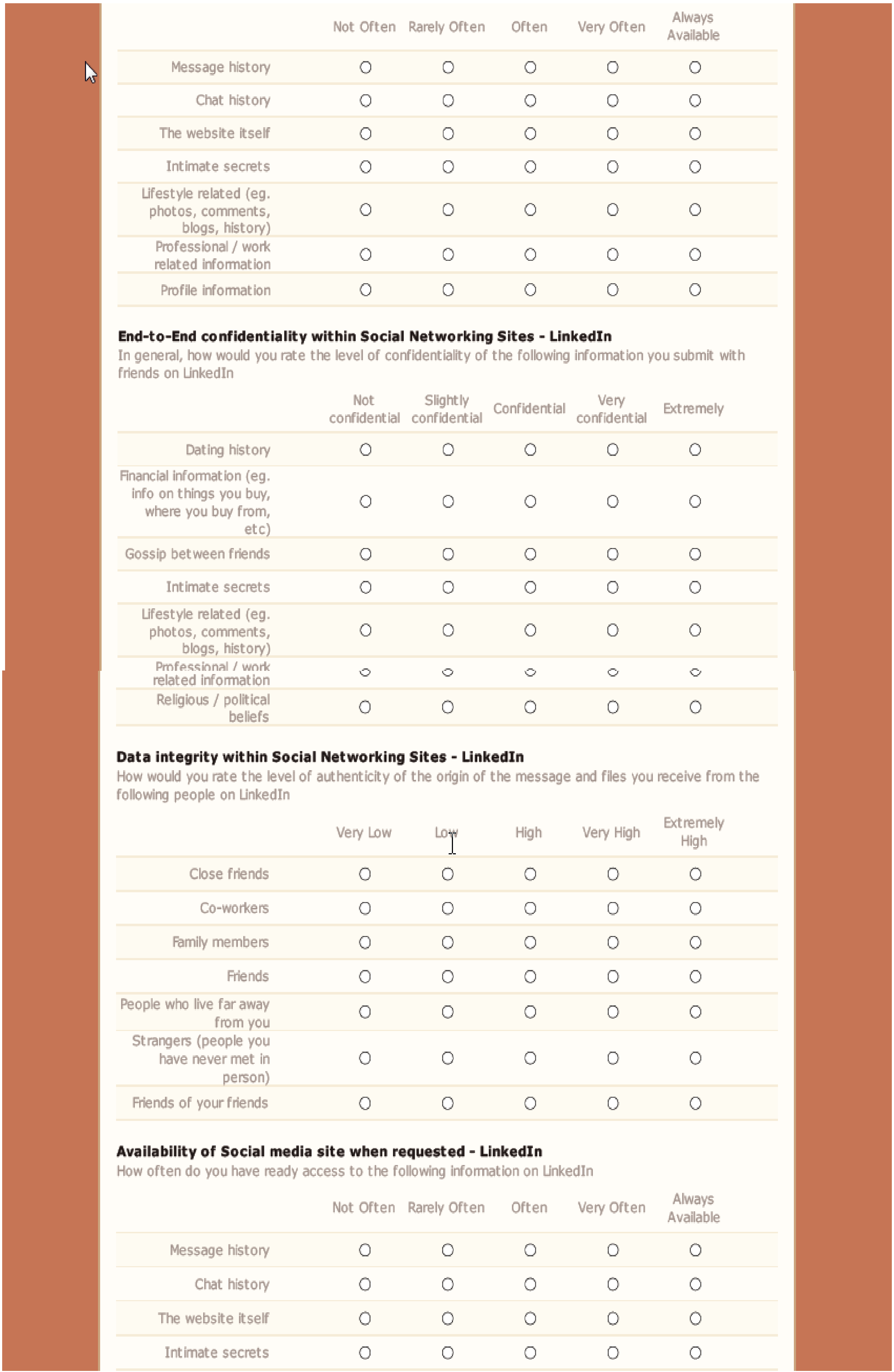
Data integrity within Social Networking Sites - TWITTER

How would you rate the level of authenticity of the origin of the message and files you receive from the following people on Twitter

	Very Low	Low	High	Very High	Extremely High
Close friends	○	○	○	○	○
Co-workers	○	○	○	○	○
Family members	○	○	○	○	○
Friends	○	○	○	○	○
People who live far away from you	○	○	○	○	○
Strangers (people you have never met in person)	○	○	○	○	○
Friends of your friends	○	○	○	○	○

Availability of Social media site when requested - TWITTER

How often do you have ready access to the following information on Twitter



	Not Often	Rarely Often	Often	Very Often	Always Available
Message history	○	○	○	○	○
Chat history	○	○	○	○	○
The website itself	○	○	○	○	○
Intimate secrets	○	○	○	○	○
Lifestyle related (eg. photos, comments, blogs, history)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Profile information	○	○	○	○	○

End-to-End confidentiality within Social Networking Sites - LinkedIn

In general, how would you rate the level of confidentiality of the following information you submit with friends on LinkedIn

	Not confidential	Slightly confidential	Confidential	Very confidential	Extremely
Dating history	○	○	○	○	○
Financial information (eg. info on things you buy, where you buy from, etc)	○	○	○	○	○
Gossip between friends	○	○	○	○	○
Intimate secrets	○	○	○	○	○
Lifestyle related (eg. photos, comments, blogs, history)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Religious / political beliefs	○	○	○	○	○

Data integrity within Social Networking Sites - LinkedIn

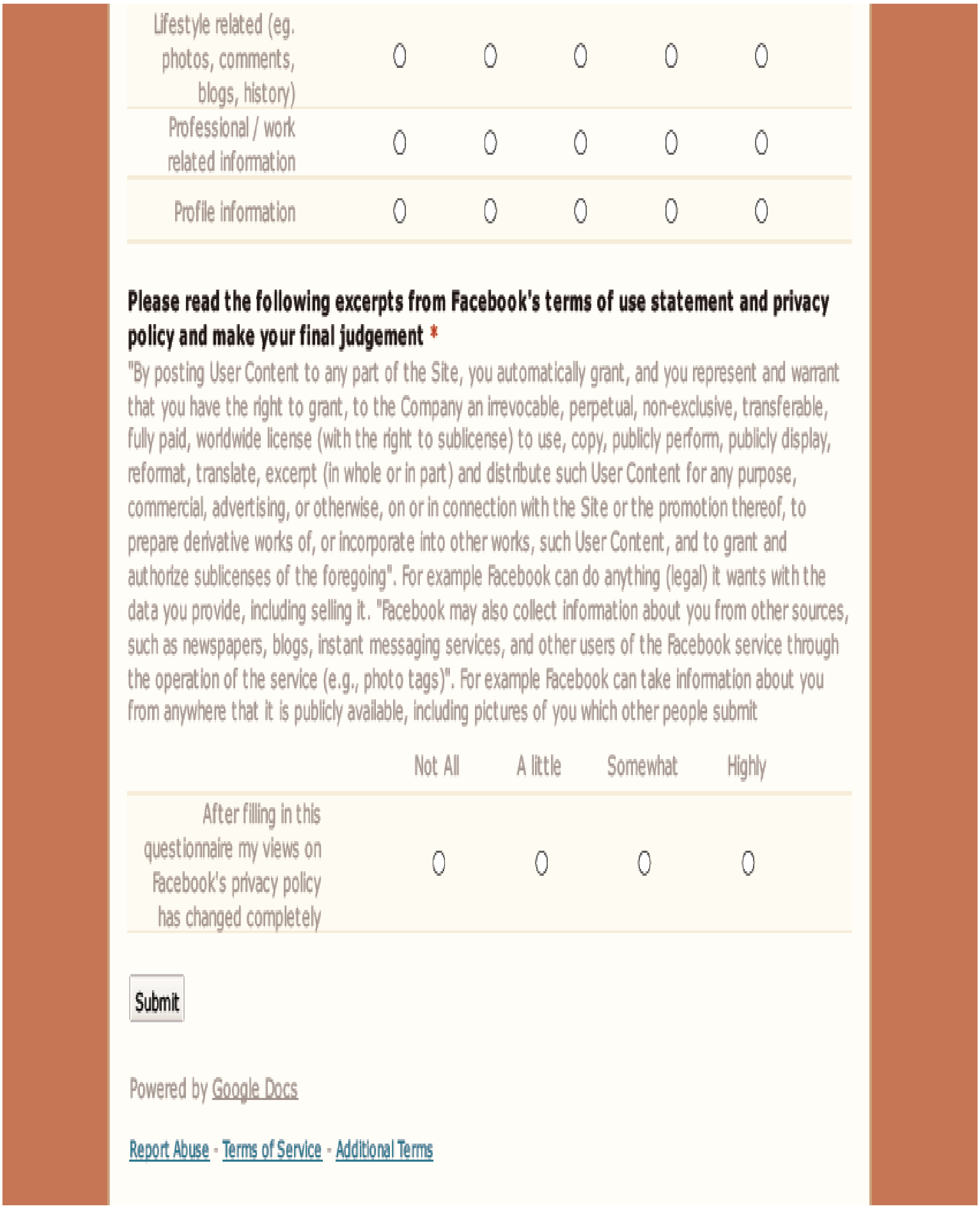
How would you rate the level of authenticity of the origin of the message and files you receive from the following people on LinkedIn

	Very Low	Low	High	Very High	Extremely High
Close friends	○	○	○	○	○
Co-workers	○	○	○	○	○
Family members	○	○	○	○	○
Friends	○	○	○	○	○
People who live far away from you	○	○	○	○	○
Strangers (people you have never met in person)	○	○	○	○	○
Friends of your friends	○	○	○	○	○

Availability of Social media site when requested - LinkedIn

How often do you have ready access to the following information on LinkedIn

	Not Often	Rarely Often	Often	Very Often	Always Available
Message history	○	○	○	○	○
Chat history	○	○	○	○	○
The website itself	○	○	○	○	○
Intimate secrets	○	○	○	○	○



Lifestyle related (eg. photos, comments, blogs, history)	○	○	○	○	○
Professional / work related information	○	○	○	○	○
Profile information	○	○	○	○	○

Please read the following excerpts from Facebook's terms of use statement and privacy policy and make your final judgement *

"By posting User Content to any part of the Site, you automatically grant, and you represent and warrant that you have the right to grant, to the Company an irrevocable, perpetual, non-exclusive, transferable, fully paid, worldwide license (with the right to sublicense) to use, copy, publicly perform, publicly display, reformat, translate, excerpt (in whole or in part) and distribute such User Content for any purpose, commercial, advertising, or otherwise, on or in connection with the Site or the promotion thereof, to prepare derivative works of, or incorporate into other works, such User Content, and to grant and authorize sublicenses of the foregoing". For example Facebook can do anything (legal) it wants with the data you provide, including selling it. "Facebook may also collect information about you from other sources, such as newspapers, blogs, instant messaging services, and other users of the Facebook service through the operation of the service (e.g., photo tags)". For example Facebook can take information about you from anywhere that it is publicly available, including pictures of you which other people submit

	Not All	A little	Somewhat	Highly
After filling in this questionnaire my views on Facebook's privacy policy has changed completely	○	○	○	○

Powered by Google Docs

Report Abuse - Terms of Service - Additional Terms

Appendix B

Questions used to create the fuzzy membership functions and their ranges for the variables

Based on answers you might have given to the questions above, please answer the following questions to classify the five "Not confidential/Extremely confidential" levels you would have used in answering the survey using a scale from 0 to 10.

For Example: My range is between 0-2 or 2-3 or 7-9 etc for the different levels all between 0-10

a. I will answer in 'Not confidential' when my agreement ranges from ______ to ______.

b. I will answer in 'Slightly confidential' when my agreement ranges from ______ to ______.

c. I will answer in 'Confidential' when my agreement ranges from ______ to ______.

d. I will answer in 'Very confidential' when my agreement ranges from ______ to ______.

e. I will answer in 'Etremely confidential' when my agreement ranges from ______ to ______.

Based on answers you have given to the questions above, please answer the following questions to classify the five "VeryLow/ExtremelyHigh" levels you would have used in answering the survey using a scale from 0 to 10.

For Example: My range is between 0-2 or 2-3 or 7-9 etc for the different levels all between 0-10

a. I will answer in 'VeryLow' when my agreement ranges from ______ to ______.

b. I will answer in 'Low' when my agreement ranges from ______ to ______.

c. I will answer in 'High' when my agreement ranges from ______ to ______.

d. I will answer in 'VeryHigh' when my agreement ranges from ______ to ______.

e. I will answer in 'EtremelyHigh' when my agreement ranges from ______ to ______.

Based on answers you have given to the questions above, please answer the following questions to classify the five "Not Often/Always Available" levels you would have used in answering the survey using a scale from 0 to 10. For Example: My range is between 0-2 or 2-3 or 7-9 etc for the different levels all between 0-10

a. I will answer in 'Not Often' when my agreement ranges from ______ to ______.

b. I will answer in 'Slightly Often' when my agreement ranges from ______ to ______.

c. I will answer in 'Often' when my agreement ranges from ______ to ______.

d. I will answer in 'Very Often' when my agreement ranges from ______ to ______.

e. I will answer in 'Always Available' when my agreement ranges from ______ to ______.

Appendix C

Implementation in MATLAB on CD-ROM