A Digital Identity Management System

by

Jackson Phiri

A thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae in the Department of Computer Science,

University of the Western Cape

Supervisor

Prof. Johnson Agbinya

February 2007

A Digital Identity Management System

Jackson Phiri

KEYWORDS

Artificial Intelligence
Authorization
Biometrics
Credentials
Digital Identity
Identity Fraud
Information Fusion
Multi-modal Authentication

ABSTRACT

The recent years have seen an increase in the number of users accessing online services using communication devices such as computers, mobile phones and cards based credentials such as credit cards. This has prompted most governments and business organizations to change the way they do business and manage their identity information. The coming of the online services has however made most Internet users vulnerable to identity fraud and theft. This has resulted in a subsequent increase in the number of reported cases of identity theft and fraud, which is on the increase and costing the global industry excessive amounts.

Today with more powerful and effective technologies such as artificial intelligence, wireless communication, mobile storage devices and biometrics, it should be possible to come up with a more effective multi-modal authentication system to help reduce the cases of identity fraud and theft.

A multi-modal digital identity management system is proposed as a solution for managing digital identity information in an effort to reduce the cases of identity fraud and theft seen on most online services today. The proposed system thus uses technologies such as artificial intelligence and biometrics on the current unsecured networks to maintain the security and privacy of users and service providers in a transparent, reliable and efficient way. In order to be authenticated in the proposed multi-modal authentication system, a user is required to submit more than one credential attribute. An artificial intelligent technology is used to implement a technique of information fusion to combine the user's credential attributes for optimum recognition. The information fusion engine is then used to implement the required multi-modal authentication system.

DECLARATION

I, the undersigned here declare that *Digital Identity Management System* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Name: Jackson Phiri

Date: February 2007

Signature: .....

DEDICATION

To our God and father in heaven through our Lord Jesus Christ who is the source of all knowledge and wisdom

ACKNOWLEDGEMENTS

I am extremely grateful to Prof. Johnson Agbinya who helped enormously and guided me with his knowledge and experience throughout my research to completion. I am also grateful for his continuous patience, support and advise during the course of my studies.

I would like to thank the Belgium government through the VLIR project for providing the scholarship, conference travel funds and valuable guidance. Special thanks go to Prof. Davison Theo, Prof. Shamitiba Kanyanga, Dr. Jameson Mbale, Sikaamba Mulavu and Joseph Sichangwa who facilitated the release of the scholarship funds.

It is very difficult to acknowledge everyone who has helped me in numerous ways. However, I would like to mention a few people who have helped me in many ways; Prof. Isabel Venter, Verna Connan, James Connan, Michael Norman, Daniel Leenderts, Paul Kogeda, Steven Mudenda, Wilson Wu, Sandro Da Silva, Chrispin Kabuya and all those who took part in answering the questionnaires.

Finally and most importantly, I would like to thank my lovely wife Janet M.C. Phiri for her patience, understanding and support while I was doing my research.

LIST OF PUBLICATIONS

[1] Phiri J. and Agbinya J., "Modelling and Information Fusion in Digital Identity Management Systems", *Proceedings of IEEE International Conference on Systems (ICONS 2006), Mauritius, 22nd* - 29th April 2006, pp. 181

[2] Phiri J. and Agbinya J., "Using Artificial Neural Networks to Implement Information Fusion in Digital Identity Management Systems", *Proceedings of the 10th Southern African Telecommunication Networks and Applications Conference on Network Management/OSS (SATNAC 2006), South Africa, 3rd* - 6th September 2006, pp. 1 - 6

[3] Agbinya J. and Phiri J., "Establishing the Significance of Credential Attributes in Digital Identity Management Systems", EUROCAST 2007, Las Palmas, Spain, 12th – 17th February 2007

[4] Phiri J. and Agbinya J., "Fusion of Multi-Modal Credentials for Authentication in Digital Identity Management Systems", Proceedings of the 2nd International Conference on Wireless Broadband and Ultra Wideband Communication (AusWireless 2007), Crowne Plaza Hotel, Sydney, Australia, 27th - 30th August 2007

TABLE OF CONTENTS

LIST OF FIGURES AND GRAPHS

LIST OF TABLES

Chapter 1

1 INTRODUCTION TO THE RESEARCH

1.1 Introduction

For thousands of years now, it has been a central responsibility of governments to manage the identities of its citizens, public property and public organizations. A good example is a historical reference story in the Bible of Joseph and Mary travelling to the town of Bethlehem to register for the census over 2,000 years ago [1]. This reference shows that there was already an established role for governments in the identification of its citizens during that time. Today with the invention of computers with improved storage and telecommunication system, identification of users and organisations has taken a slightly different approach. However this improved technology in access and storage mechanism of the credential attributes has also come with its own challenges. Due to the open nature of the Internet, without reliable identification and authentication, identity fraud and theft has been on the increase with security concerns emerging in the areas of immigration, border crossings, airline passengers, hazardous materials, driver's licenses, and pilot training [2].

Identity fraud has now become a global threat to the security of nations and global commerce as it facilitates a wide range of crimes and terrorism [3]. It has become a component of almost every major crime and its presence has been felt throughout the world. In today's corporate and government environments, information sharing is fundamental to cut cost and to manage increasing numbers of customer interactions [4]. Financial institutions, telecommunication organisations, governments, health care providers and other organisations are transforming their database records and their access methods into electronic forms [5]. The challenge is therefore to create a more effective method for validating, verifying and authenticating a user to reduce the cases of identity fraud seen on most online services today [2].

One of the effective ways to prevent identity fraud is to build defence against the use of false identities. This could be achieved by implementing a multi-modal identity authentication system during service delivery [3]. To balance convenience and security during multi-modal authentication, the strength of the authentication system needs to match the required level of trust. This is because if the implemented strength is lower than the required level of trust, it may introduce risk of fraudulent activities where as if the implemented strength is higher than the required level of trust, it may introduce inconvenience to the user hence preventing the usage of the system by the users [2]. In this thesis Digital Identity and Management System (DIMS) is introduced as a solution to identity fraud and theft seen on most online services today.

1.2 Scope

Digital Identity and Management System is a very broad topic and is comprised of three indispensable elements that include Policies, Processes and Technologies [4]. Policies are the standards and constrains that need to be followed in order to comply with regulations and business best practices. Processes are the sequences of steps that lead to the completion of business tasks or functions. Technologies refer to the automated tools that help accomplish business goals more efficiently and accurately while meeting the constraints and guidelines specified in the polices [4].

The technological segment is considered in this thesis. However the technology segment can further be broken down into three major areas. These are Identity Life Cycle, Directory Services and Access Management [4]. The Life Cycle process of digital identities involves the processes of their creation, utilization and termination where as the Directory Service refers to the processes of providing the infrastructure for secure data storage and organization [4]. The Access Management refers to the process of controlling and granting access to satisfy resource request [4].

Access Management is considered in this thesis. This process is usually completed through a sequence of Authentication, Authorization and Auditing actions. To be more specific this thesis will focus on multi-modal authentication system for remote online transaction over an unsecured network.

Multi-modal authentication system is considered as one of the solutions for the identity fraud and theft seen on most online services today. Identity fraud generally takes place in two stages [2]. The first stage involves creating a fictitious identity or stealing someone else's identity. In the second stage the fictitious or stolen identities are then used to access the restricted services. This thesis explores measures and creates mechanisms, which prevent users from using a fictitious or stolen identity to access restricted services. It does not however cover areas which prevent the users from creating fictitious identity or stealing others identity.

1.3 Problem Statement

The increase in the number of online services has seen a correspondent increase in the cases of identity theft and fraud which is becoming a major concern for both the public and private sectors especially as it relates to problems like terrorism, financial crime, drug trafficking and weapons smuggling. At the centre of these concerns is the need to authenticate individuals to determine if they are who they claim to be. A reliable multi-modal authentication system is required in a digital identity management system to ensure users are who they claim to be with reliable privacy and guaranteed security on both the service provider and the user. This thesis attempts to provide solutions to the following questions;

1) How can a digital identity representation be constructed from the real-space identity documents (credential tokens) to represent a person or a communication device used to access the online services from a remote area?
2) How can the credential attributes submitted during multi-modal identity authentication be modelled and strength of these credential attributes determined?
3) Is it possible to create a technique of information fusion to combine the credential attributes submitted by a user so as to create the overall combined strength for optimum recognition of the user or a non human communication device representing the user during multi-modal authentication?
4) Is it possible to build a prototype to create an online multi-modal authentication system that uses a technique of information fusion?

1.4 Aims

To answer the questions in the problem statement above, five aims are set to help in coming up with a multi-modal authentication system in a digital identity management system. These are:

1) Analyse the aspects of digital identities, identity management systems and then the aspects of information fusion techniques.
2) Formulate a scheme for digital identity representation of a person and a non-human communication device, which can be used for remote identity assertion.
3) Determine the grouping of the credential attributes and then formulate a scheme to estimate the weights of these credential attributes
4) Design and implement an information fusion technique using an artificial intelligence technology
5) Design and implement a system that uses a technique of information fusion to implement multi-modal authentication of a user accessing the services remotely

1.5 Motivation

Today's digital identity management products have fundamental design flaws. This is because they rely on a technology that was invented a quarter of a century ago, at the beginning of current cryptography in digital identity certificates [6]. Most of the digital identity management systems of today offer only little protection during authentication, do suffer a number of serious attacks and performance drawbacks because of their heavy reliance on trusted online repositories [6]. This has lead to a rapid increase in identity theft and fraud seen on most online services today. There is therefore a pressing need for digital identity management systems solutions that will address the entire range of needs of all system participants.

In this thesis, a Digital Identity Management System using multi-modal authentication is proposed for optimum recognition of a user or a device representing the user. The user will be required to submit more than one attributes in order to be authenticated. The attributes will include biometrics (features), attributes from the device being used to access the services, a secret code and the user' name or given national identity number. A

technique of information fusion will then be used to combine all the attributes submitted by the user for optimum recognition. Artificial intelligent technologies such as artificial neural network, fuzzy logic or Bayesian method can be used to implement an information fusion engine. The outcome of an information fusion will then be used to implement multi-modal authentication. The user will be successfully authenticated depending on the computed value of an information fusion engine. This value is dependant on the collect set of attributes submitted by the user. This most likely will make it difficult for a hacker or thief to either guess or forge the whole combination of the attributes, hence helping to reduce the cases of identity fraud and theft seen on most online services offered today.

1.6 Methodology

The goal in this thesis is to create a Digital Identity Management System using multi-modal authentication. To begin with, a Credential Attribute Mapping Modelling [2] scheme is employed to create a digital representation of a user. This scheme is extended to include non-human communication devices such as mobile phones and Internet terminals (e.g. a desktop computer) used to access online services. A questionnaire is used to estimate the initial average scores of the attributes using the desirable properties of digital identities [7]. Shannon's information theory is then used to compute the final weight of the attributes using the initial scores from the questionnaires.

Using the computed weights of the attributes an artificial neural network is used to implement an information fusion technique. MatLab software is used to train the neural network and simulate the network to obtain the weights and threshold values of the neurons in the network. Microsoft Visio 2000 is used to design the UML diagrams and an entity relational database diagram required to store the user attributes when implementing an experimental system. JavaServer Pages (JSP) language is used to code the classes in the UML diagram when implementing an online experimental system. The experimental system uses an information fusion engine implemented with an artificial neural network during multi-modal authentication. Apache Tomcat Web Server is used to test the online authentication system developed with JSP.

1.7 Outcomes

Section 1.4 above outlined five aims that need to be achieved in order to create a Digital Identity Management System using multi-modal authentication for optimum recognition of the user. To achieve these aims, the following are the expected outcomes of this thesis.

1.7.1 Background Theory

The background theory looks at the analysis of digital identity management systems and information fusion technologies. These include issues of privacy, identity fraud and how they affect digital identity management systems today. Related works in both digital identity management systems and information fusion techniques are also considered.

1.7.2 Selecting the Credentials

Using a questionnaire, the fourteen most commonly used credential tokens are selected. These are the tokens used in most countries to access services offered both in the Real-space and Cyber-space. It is these credential tokens that are used as the source of the attributes required to build an identity system.

1.7.3 Extracting the Attributes from the Credentials and Grouping them

The attributes are then extracted from the fourteen credentials and a set of biometric features. The attributes are then grouped into four groupings namely physical metrics, pseudo metrics, device metrics and biometrics for easy analysis.

1.7.4 Digital Identity Representation

In order to determine how much storage space is required to store each attributes in the database system or mobile storage device, each attribute is represented using the ISO standards encoding.

1.7.5 Determining the Weight of the Credential Attributes

Using a questionnaire with desirable properties of digital identities, the initial average score of each attribute is computed. Using the initial average scores, Shannon's information theory is then used to compute the final weight of each attribute.

1.7.6 Information Fusion Implementation Using Artificial Neural Networks

Using an artificial neural network with the attribute's weights and groupings, an information fusion engine is designed and implemented. MatLab software is used to train the neural network to obtain the desired input weights, layer weights and the threshold values of the neurons in the network.

1.7.7 System Design and Implementation

The experimental system is first designed and then implemented using UML diagrams, JavaServer Pages and HTML languages. The experimental system uses an information fusion engine to implement multi-modal authentication of a remote user accessing online services using an Internet terminal as a device used to access the online services.

1.8 Organisation of Thesis

To begin with, Chapter 2 looks at the literature review of the essential concepts surrounding digital identity management systems and technologies used to implement information fusion techniques. The chapter is concluded by the related works in digital identity management systems and technologies used to implement information fusion techniques. Chapter 3 focuses on the process of selecting the credential tokens (identity documents), which are then used as the sources of the credential attributes used to build an identity management system. In this chapter, a scheme is devised for grouping and representing the attributes. Using a questionnaire and desirable properties of digital identities, the initial average scores are first computed of the credential attributes. Shannon's information theory is then used to compute the final weight of these attributes.

Using the groupings and weights computed in Chapter 3, an information fusion engine is designed and implemented using artificial neural networks in Chapter 4. MatLab software is used to train the network in order to generate the input weights, layer weights and the threshold values of the neurons in the network. Using the system design and the information fusion engine developed in Chapter 4, an experimental system is implemented in Chapter 5 using JavaServer Pages. Finally the discussion, conclusion and future works are considered in Chapter 6.

Chapter 2

2 BACKGROUND THEORY

2.1 Introduction

This chapter looks at the background theory of digital identity management systems and information fusion technologies. The first half of the chapter gives an analysis of digital identities and management systems. This includes definitions, concepts and technologies surrounding digital identity systems. Cases in identity fraud and privacy are also highlighted. The second half of the chapter looks at the technologies used for implementing information fusion techniques. Finally related works done in the area of digital identities and management systems and information fusion techniques are highlighted.

2.2 Digital Identity and Management Systems

Digital identity management system is an integrated system of business processes, policies and technologies [13]. These enable organizations to facilitate and control the users' access to critical online applications and resources while protecting confidential information of the users and the service providers from unauthorized users [4]. It represents a category of unified solutions that are employed to administer user authentication, access rights, access restrictions, account profiles, passwords, and other attributes supportive of users' roles or profiles on one or more applications or systems [8].

2.2.1 Digital Identities

The definition of digital identity depends on the situation, purpose, use and many other factors. Eric and Adre define a digital identity as a, "virtual representation of a real identity that can be used in electronic interactions with other machines or people" [9]. In other words this is the electronic representation of a real-world entity. The term is usually taken to mean the online equivalence of an individual human being, which participates in

electronic transactions on behalf of the person in question [2]. However a broader definition also usually assigns digital identities to organizations, companies and even individual electronic devices [4].

The concept of digital identity usually depends on the usage, situation, purpose and several other factors [9]. In general digital identities can be considered and defined in terms of identity space, which can be categorised as *Real-space* and *Cyber-space* [2]. The Real-space identities are the physical identity tokens such as birth certificates, passports and driving licences while the digital identities include the credential attributes such as usernames, passwords and Internet Protocol (IP) addresses [2]. Table 1 gives a summary of the two types of identity spaces.

#	Real Space Identity	Cyber-Space Identity
1	Birth Certificate	Public Key / Private Key pairs
2	Driving Licence	Username / Password
3	Bank Cards	IP addresses
4	Passport	Mac Addresses

Table 1: Identity Space

An identity usually consists of traits, attributes and preferences upon which one may receive personalized services [10]. Such services may exist online, on mobile devices (e.g. mobile phone), at work and in many other places. A user acquires many forms of identifications usually stored in various forms and places as shown in Figure 1 [10];

Figure 1: The Various Forms of Identifications[1]

[1] http://www.projectliberty.org/

An attribute is a term used to refer to the properties of a given individual or entity that are of interest to and knowable by other entities [11]. Examples include an individual's height, age or eye colour. Figure 2 gives the illustration of a set of attributes [2].

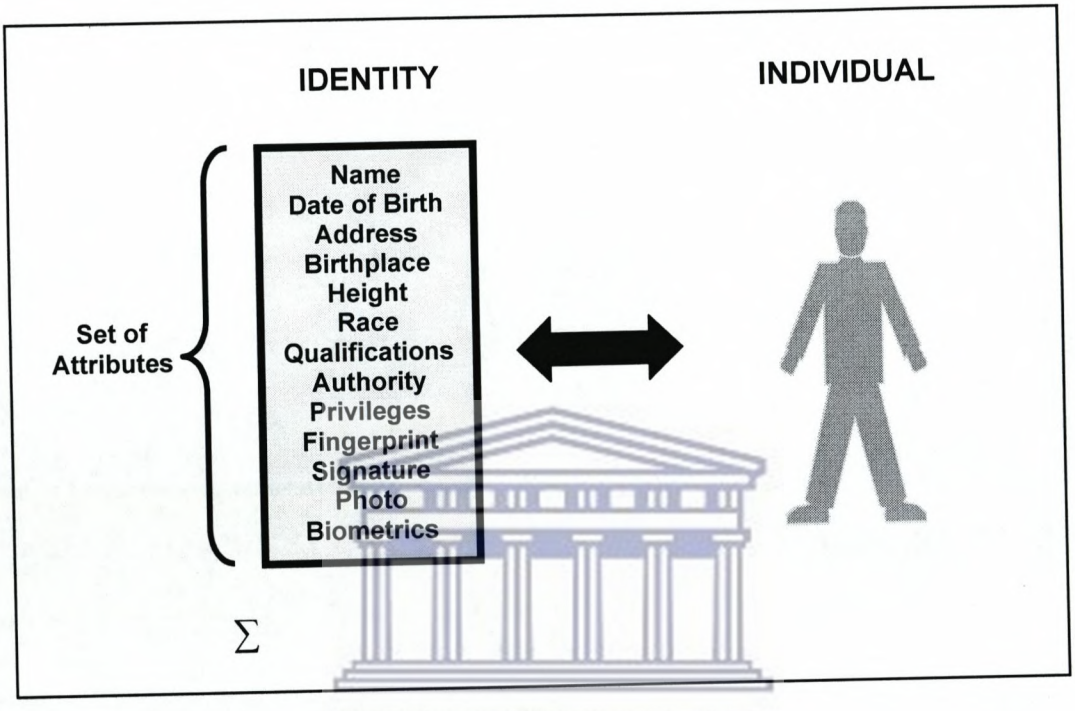

Figure 2: Concept of Personal Identity[2]

Credential is used to refer to the private or public data that could be used to prove authenticity of an identity claim [4]. Examples of credential tokens are a combination of a password and an email address in the cyber space or a passport in the real space [2].

2.2.2 Sources of the Attributes Used in Digital Identity Systems

Most of the time the choice of credentials tokens presented for identification of a given individual depends on the life history of that individual, his or her place of birth, citizenship, residence, qualifications or work. Today there are a number of different credential tokens used to access the services offered both in the cyber space and real space. These include the national passport, national identity book, birth certificates,

---

[2] Subenthiram Sittampalam, "Digital Identity Modelling and Management", MEng Thesis, 2005, UTS, Australia

citizenship certificates, driver's license, credit cards, a mobile phone (as used in m-commerce), an Internet terminal (as used in e-commerce) and biometrics [2]. These and many others are used to construct a digital representation of a personal identity or a device representing a user. Below is a brief description of the Internet terminal, mobile phone and biometric technologies as these will play an important role in this research.

The use of the Internet has changed in the last decade from using static web sites and email only. Today most organisations have dynamic web sites and a wide range of electronic commerce services. Good examples of such services include the electronic banking and electronic learning systems (e-learning)[12]. The challenge is to determine if the user is who he or she claims to be at the other end of the network. To do this the mode of connection to the Internet by the terminal used to access the online services need to be considered. The connection can therefore be done by either using a Network Interface Card (NIC) or a modem [13]. For home use and small offices usually a modem is used for the connection where as for bigger institutions the NIC is used to connect a computer mainly to a Local Area Network (LAN) which in turn connects to the Wide Area Network (WAN) which is then connected to the Internet. When using the modem, the user is identified by using the username and password provided by the service provider when registering for the service and the phone number of the home landline [13]. On the other hand when using a NIC, the computer will be identified using a Media Access Control (MAC) address, which consist of a 12 hexadecimal digits i.e. 48 bits [13] and the Internet Protocol (IP) address.

Telecommunications has been expanding and is still expanding beyond voice to new services that now allow people to do more than just speak to each other [14]. Today users can make use of sound, text, and video in their interactions with other users and are able to access online services from any place any time [15]. The mobile phone system has gone through a series of evolutions from first generation (1G), the second generation (2G) and now in the third generation (3G) and beyond [14]. There are a number of standards in the world today specifying the infrastructure for digital cellular services of which the most common is the Global System for Mobile Communication (GSM) [15]. This thesis focuses on GSM, which is a set of ETSI standards. The standard is used in approximately 85 countries in the world including such locations as [15] Europe, Japan

and Australia. The cellular network provider usually maintains its own subscriber database for the user's identity attributes. The database usually contains the following information about each customer; customer name and address, billing name and address, user name and address, billing account details, telephone number (MSISDN), International Mobile Subscriber Identity (IMSI), SIM serial number, PIN/PUK for the SIM and finally the type of services allowed [12]. It is these attributes that are used to authenticate a user when accessing the online services using a mobile phone.

Biometric features are increasingly being used as the best alternative way of preventing the identity fraud and theft seen on most online services today. There are seven biometric features that have found real life applications and have been deployed in a number of countries [2]. These include Palm Geometry Based Verification, Face Recognition, Voice Recognition, Iris or Retina Scan, Fingerprint Verification, Signature Verification and Infra Red Scan of face and body parts [16].

Hand geometry recognition system is based on a number of measurements taken from the human hand. These include its shape, size and length of the palm and finally the widths of the fingers [16]. The technique is very simple, relatively easy to use, and inexpensive with environmental factors such as dry weather not having any effect on verification accuracy. Also the individual anomalies such as dry skin do not appear to have any negative effects on the verification accuracy [2].

The facial images are probably the most common biometric characteristic used by humans to make a personal recognition [16] [17]. The most popular approaches to face recognition are based on both the location and shape of facial attributes such as the eyes, eyebrows, nose, lips and chin, and their spatial relationships or the overall analysis of the face image that represents a face as a weighted combination of a number of canonical faces [17]. While the verification performance of the face recognition systems that are commercially available is reasonable, they impose a number of restrictions on how the facial images are obtained, sometimes requiring a fixed and simple background or special illumination [18].

Voice is a combination of physiological and behavioural biometrics. The behavioural part of the speech of a person changes over time owing to age, medical conditions, emotional

state, and the speech text system being used [19]. Voice is not very distinctive and may not be appropriate for large-scale identification systems [16].

There are two main types of iris or retinal scans technology available today one of which is the infrared scanning and the second being the video camera encoding [20]. The eye's physiological response to light and natural pupillary oscillation prevents substitution of a photograph or some other imitation for living tissue [21]. The iris and the retina are considered to be more unique than any other biometrics feature [16].

A fingerprint is composed of a pattern of ridges and valleys. These patterns of ridges and valleys on the surface of a fingertip are determined during the first seven months of foetal development [16]. Because of the smaller size of the sensors used in fingerprint verification systems, they are becoming popular and also cheaper to deploy [22].

How an individual writes his signature is considered to be a unique characteristic of that individual. Signatures are considered to be behavioural biometrics that changes over a period of time and are influenced by the physical and emotional conditions of a signatory [16]. Most of the time, a person is required to sign a signature somewhere indicating an agreement with the terms and conditions of a given transaction. They have been accepted in governments, private institutions and commercial transactions as a method of verifying a transaction [16] [23]. Signature recognition systems have found a number of applications and have been deployed in a number of business centres [2].

The pattern of heat radiated by human body is a characteristic of an individual and can be captured by an infrared camera in an unobtrusive way much like a regular photograph [20]. A thermogram-based system does not require contact with an individual and is non-invasive. Image acquisition may be challenging in an uncontrolled environment, where you have objects emitting heat from their surfaces in the vicinity of the body to be measured [16].

2.2.3 Desirable Properties of Digital Identities

Digital credentials used to build digital identity management systems can be analysed using a number of ways in an effort to determine their effectiveness. In this thesis, the initial average scores of the identity attributes are computed using the desirable properties

of digital identities. Uniqueness, Verifiability, Consistency, Persistency and Trust [7] (though the list is not exhaustive) are used to analyse all the identity attributes except for the biometrics. For a biometric feature to pass a test so that it may be used in a digital environment, it must pass a test of seven desirable properties of digital identities, which include Universality, Distinctiveness, Permanence, Collectability, Performance, Acceptability and Circumvention [16] [17]. Only five of these are applicable in this thesis and will be used to compute the initial average scores of the biometrics. The initial scores of the attributes are then used to compute the final weight of each attributes using Shannon's information theory [24]. The final weights of the attributes are then used to determine an information fusion engine used during multimode authentication. Below is a brief description of the above ten desirable properties of digital identities used in this thesis to analyse the identity attributes.

To be unique means that the identifier does not use the same value as someone else's and refers to only one particular identifier [7]. For example, date of birth is not a unique identifier but a student identity number has unique association with a person. When resolving identity, uniqueness prevents confusion of whom it is referring.

When a credential is presented to a relying party, the attributes association with the individual in the credential should be verified very easily [7]. For example, a name can not be verified because anyone can claim to be Jackson while on the other hand; the photo can easily be verified by comparing with the individual's face.

To be consistent means that all parties concerned interpret an identifier in the same way. [7]. For example, a particular person's height is not consistent because while he or she may be considered to be tall in South Africa, in other countries he or she would be considered medium or even short. On the other hand a passport number will be the same from whichever country a person is visiting thus a passport number is considered to be consistent as compared to height.

Persistency in terms of digital identities can be defined as an identity attribute that remains the same at all time without changing [7]. An example is an individual's date of birth, which cannot change hence very high persistence while the individual's address of

residence used as part of their identity changes and becomes invalid as the person moves into another house in the future.

Trust in terms of digital identities can be defined as the ability of a particular identity to withstand a challenge as to its validity hence reassuring other parties who would like to use this identity that they are not being deceived [7].

Distinctiveness in terms of biometrics means any two persons should be sufficiently different in terms of the characteristics in consideration [16]. The iris and retinal scans are considered to be very distinctive for every person [17].

A biometric system should not be easily fooled using fraudulent methods by an attacker [16]. Therefore the threshold for calculating scores based on circumvention is based on how a perpetrator could fool the system [17].

A biometric system should have a very high performance in terms of accuracy, producing consistent results within shortest possible processing time [17]. Some biometric systems' performance can be affected by environmental factors such as lighting, temperature, noise or humidity [2].

Permanence in terms of biometric refers to a biometric characteristic that remains unchanged over a period of time [2]. Because of the natural human aging process, the biometric characteristics can vary with time and the degree of variation depends on the individual biometric characteristics [16].

For a specific biometric system to be widely deployed every person should possess the characteristic. Therefore universality refers to the availability of a biometric feature to almost each and every user [16]. Although every user is expected to possess the biometric trait being acquired, in reality it is possible for some users not to possess a particular biometric due to birth defects or injury [17].

2.2.4 Classification of Digital Identities and their Representation

Credential attributes used to authenticate a user in a digital identity management system are many and come from diverse areas. As the individuals grow from the time they were born, they usually get involved in many activities and tend to acquire different forms of identities. In order to build the information fusion engine, the credential attributes need to

be grouped with a better digital identity representation for easy usage and analysis. The credentials acquired by the user are usually classified into two or more groups for easy analysis [3]. A brief discussion of the basic classification used by [2] and the scheme used to determine a digital identity representation of a user are described below.

In [2] all the credential attributes are classified into three groups only, namely the *Physicalmetrics, Pseudometrics* and *Biometrics.* The *Physicalmetric* group is used to refer to something you possess as a user [2]. In this grouping, authentication relies on an evidence of possession of a token-based credential, which could be in form of a passport, a driver license or a credit card. The owner of this token carries it to authorise access to an online service. *Pseudometric* group on the other hand is used to refer to something you know which is a secret shared between the service provider and the user in order to authenticate the user's identity at the time of service delivery [2]. Good examples include a Personal Identity Number (PIN) and a password. The third and last group referred to as *Biometrics* is usually used to refer to personal identity based on someone's biological, physiological and or behavioural characteristics [16]. Examples of biometrics in use today include iris scan, face recognition, finger print, hand geometry, voice recognition, digital signature, infrared thermogram of face vein and hand vein, keystrokes and retina recognition [17].

The representation and mapping of digital identities is another critical step towards building a digital identity management system. The Credential Attribute Mapping (CAMA) scheme [2] is used in this thesis. This process requires that a set of the user's identity attributes be encoded into digital information to represent their real-space identity. It is therefore used to construct a digital representation of personal identity from conventional identity documents such as birth certificates, citizenship certificates, passports, driving licences, bankcard etc. This method is then extended to obtain a digital representation of a non-human communication device used in accessing online services such as an Internet terminal (PCs) or mobile phones.

2.2.5 Digital Identity Management on the Internet

Since communication in the cyber-space is based on the Internet Protocol (IP), the analysis of digital identities on the Internet is usually based on the TCP/IP protocol suite.

Thus the issue here is how one can tell who is at the other end of the network session. Identities of the Internet users can be revealed to the other party in one of the two ways namely explicitly or implicitly identification [7].

In explicit identification a user may be asked to submit his or her username and a password, which is a shared secret so that the server can authenticate the user's identity. The server will then compare the username and the password with a database copy and the user is authenticated if the right identity is supplied. Once the identity is authenticated, there should be a mechanism to reliably track the session activity to maintain access control based on the user's privileges. Since the Hyper Text Transfer Protocol (HTTP) is stateless in the case of World Wide Web, it will not be possible for the HTTP protocol to track the subsequent interaction following the authentication session. Cookies, URL-rewriting or HTML hidden form may be employed as one of the solutions to this problem [7]. In any of these cases, following a logon session, the embedded token in the cookie, URL or HTML hidden form field will be considered as the digital identity of that person during the session. Once the session is terminated, association with that token is normally destroyed.

Most of the time, the user is not aware of the implicit way of the construction of a digital identity. To analyse this implicit way of composing digital identities in the Internet, an approach of moving from bottom to the top of the layered Internet Protocol architecture must be considered [7]. The TCP/IP model has the following four layers moving from top (layer 4) to bottom; Application layer, Transport layer, Internet layer and the Network Access Layer. Some of the most commonly used application layer protocols include the File Transfer Protocol (FTP) and the Hypertext Transfer Protocol (HTTP) while the common transport layer protocols include the Transport Control Protocol (TCP) and the User Datagram Protocol (UDP). The primary protocol of the Internet layer is the Internet Protocol (IP) while the network access layer refers to any particular technology used on a specific network [7]. The identity of the user in an implicit identification is thus constructed using these protocols from the four layers as the user communicates with the other user or machine on the other end. The user is usually not aware of this identification, which takes place automatically usually without his or her intervention.

2.2.6 Fraud and Privacy

Well-implemented digital identity management systems are considered to be the solution to reducing identity fraud, improve user's privacy, and improve organizational service delivery. If governments and business organisations implement identification management systems the right way then all the parties involved in the system should benefit [2]. It is therefore important to consider the social, economic and political consequences of identity management when doing the implementation. At present, there are a very large number of identity management projects and proposals that are being considered in a narrow range of circumstances without thinking about privacy of their clients and the dangers of identity fraud cases [25]. Below is the brief description of the identity privacy and identity fraud.

Like the definition of identity, the definition of privacy also differs according to context, culture, environment and usage. It is generally viewed as a social and cultural concept. Ferdinand Schoeman defined privacy as, "the right to determine what (personal) information is communicated to others" or "the control an individual has over information about himself or herself" [11]. To obtain fine identity privacy policies and technologies, the following need to be followed as the recommended practices [26].

1. A good identity management system should ensure confidence to the degree appropriate for the interaction that the organisation is dealing with the right person.
2. It should also make sure that it does not facilitate inappropriate, unnecessary data linkage with other organisation.
3. Authenticate the identity of a given user only when it is absolutely necessary.

A bad identity management system on the other hand, collects and handles as much identifying data as possible during enrolment and subsequent transactions to interconnect with other organisations without considering the person's identity privacy as the relevant issue [26]. With the large number of web services today, privacy has become an important issue that poses a set of challenges different from those faced before the Internet era [27]. It is therefore very important to consider identity privacy when constructing digital identity management systems. This will be adhered to in this thesis.

Usually Identity theft and identity fraud are terms used to refer to all sorts of crime in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud typically for economic gain [27]. Today with the rapid development of new technologies, telecommunications and Internet accompanied by the rapid spread of globalisation, identity fraud and theft have become some of the fastest growing crimes in the world with computer hacking, industrial espionage and cyber-terrorism as being some of the most common fraud cases [2]. Identity fraud has become a major concern for the public and private sectors particularly as it relates to terrorism, money laundering, drug trafficking and weapons smuggling [28]. Identity fraud is broader than identity theft in that identity fraud refers to the fraudulent use of any identity, real or fictitious, while identity theft is limited to the theft of a real person's identity [28]. Password sniffs and cracker programs have been extensively used to steal identity information for a long time, but recently Spyware [29] [30] and Phishing [31] [32] attacks have become wide spread and are becoming very difficult to defend against. At the centre of these concerns is the need to authenticate individuals to determine if they are who they claim to be [33]. A strong line of defence is essential to prevent skilled criminals and terrorists from gaining access to entry points that allow them to commit crimes of profit and terrorism [34]. Multi-modal authentication system using information fusion in digital identity management system is considered as one of the solutions to the problem of identity fraud and theft in this thesis.

2.2.7 Authentication, Authorization and Auditing Actions

Two of the most critical areas of identity management systems are identity authentication and authorization that fall in the area of access control. In the current information age of electronic information sharing, it is much harder to deal with fundamental security issues related to access control and authentication [35]. As the number of online services and its users increase, so are cases of identity theft and fraud which are now spreading almost instantaneously in uncontrollable ways thereby giving the traditional security techniques, such as simple passwords and firewalls insufficient during authentication and authorization [36].

From a security point of view the transaction that proves that the digital identity presented really represents who or what it says, is the process of authentication [36]. Without authentication, no other digital identity attribute can be meaningful and as the word implies, the purpose of authentication is to prove that a digital identity is authentic and may be trusted for a given user [37]. All discussions of the ability to forge or spoof a digital identity are really discussions of the authentication of a particular digital identity scheme [35]. Identity authentication can either use a single credential attribute (singer factor) or more than one credential attribute (multifactor). An example of multifactor authentication would be a logon system that requires you to have a hardware plug-in device (e.g. ATM card) along with a Personal Identity Number (PIN). This is much more secure than just the PIN number, since if someone steals the hardware key, they are unlikely to have the PIN number and vice versa. By including the biometrics and communication devices (e.g. mobile phones) when building identity systems, it is now possible to obtain a more secure multi-modal authentication system.

Once the user identity has been authenticated, the next operational level is Authorization. In order for a user to access certain items or systems he or she must be granted that permission through the process of authorization [38]. Authorization or access control can be implemented using any of the following methods. In the first method a credential token is passed that a digital identity can then carry to various systems and present to gain access [38]. The second method requires a user after being authenticated to directly open the circuits to allow the identity selective access to the services [38].

To avoid a user from memorising multiple passwords, Single Sign-On (SSO) authentication and authorisation system is usually used [39]. It is a mechanism whereby a single action of user authentication and authorization can permit a user to access all computers and systems where the user has access permission, without the need to enter multiple passwords. It reduces human error through the loss of multiple passwords being memorised, a major component of systems failure, and security threat [25]. Examples of such applications include .Net Passport [39], Kerberos [44] and Liberty Alliance [41].

Auditing in the context of digital identity management systems refer to the keeping of the records of exactly who did what and when it was done within the Information

Technology (IT) infrastructure [4]. The IT audit process typically involves the Audit Generation phases, Data Collection and Storage phase and lastly the Analysis and Feedback phase [4]. Audit trails can be generated by different infrastructure and application components for different purposes. For example, firewalls and VPN servers can generate events to help detect external intrusions while middleware components can generate instrumentation data to help detect performance anomalies and business applications [4]. The audit data need to be collected and stored somewhere after it has been produced and this can be done either by a centralised or distributed storage system. Finally after the data has been collected and stored, it is now processed and analysed manually or automatically. The audit analysis is designed to lead to conclusions on what corrective actions, if any are needed to improve the IT system and process [4].

2.2.8 Types of Storage Mechanisms of Digital Identities

Digital identity management systems may not be complete without considering the storage and access mechanisms [42]. The storage of digital identity information may be either remotely or locally. For remote storage, the digital identities have to be accessed through a Local Area Network (LAN) or Wide Area Network (WAN) using a number of protocols. For local storage these identities would be on the same machine being used to access these identities. Storage and protocols used to exchange this information form the backbone for identity management solutions [43]. Today's Identity repository solutions are based on X.500 and Light-weight Directory Access Protocol (LDAP) directory services [42]. Recently LDAP has become very popular due to its simplicity for the exchange of identity information [43]. Database systems or mobile memory units can be used for the storage of digital information. Below is the description of the advantages and disadvantages regarding the storage capacity and security aspects of the Barcode, Magnetic Stripe, Integrated Circuits and the database system.

Barcode is a method for data storage using printing techniques known from product labels with automated recognizable graphical structures consisting of black and white areas [34]. Two different kinds of barcodes are in use and these are one-dimensional and two-dimensional barcodes [42]. The most important type of barcodes is the two-dimensional barcodes known as the PDF417 as defined in ISO/IEC 15438 [34]. Different

approaches using two-dimensional barcodes for storing biometrics information have been presented and some countries have introduced biometrics for passports or Identity documents using two-dimensional barcodes [42]. Examples of these countries are Bosnia-Herzegovina, Nigeria and Guatemala [43].

Access cards, ATM cards and credit cards are good examples of storage devices that use a magnetic strip. A magnet strip is a band of magnetic metal-oxide similar to those found in radio cassette tapes and is read out by physical contact that is by swiping past a reading head [34]. The storage capacity of magnetic strip cards following ISO 7811 standard is 1288 bits, distributed over three data tracks [34]. The advantage of magnetic strips over data storage using integrated circuits is the low unit price. Because of its limited life cycle and its vulnerability to failures and modifications, magnetic strips should not be used for long-term ID documents [34].

Integrated Circuits are a type of data storage that uses integrated circuits (ICs) memory like ROM, EEPROM or non-volatile RAM [34]. In the two-dimensional documents, there are two different relevant types of ICs memory. These are contact based and contactless ICs [34]. The contactless ICs are often referred to as radio frequency identification (RFID) transponder [44]. Both types of these ICs can have memory only functionality or advanced processing capabilities [34]. Contact based ICs when equipped with adequate security logic are able to prevent unauthorized subjects from reading or changing memory contents with the processing power to perform strong cryptography [44].

Finally the database system provides solutions to most of these problems by providing a number of security features and unlimited storage space. It usually forms the back storage for Barcode storage, Magnetic Strip Card storage and RFID system where the credential attributes stored on these devices can be looked up in the central database system in order to authenticate a given user [34].

2.3 Information Fusion Technologies

The current management of digital identity authentication systems, which depends so much on a PIN number or a username and a password, has lead to an increase in online

fraud since these credentials are easy to guess by hackers [45]. Multi-modal authentication, which will involve combining a number of attributes in order to authenticate a user or a device, is considered as one of the solutions to this problem in this thesis. The process of combing these attributes is referred to as information or data fusion [46]. The application of information fusion technical systems requires mathematical and heuristic techniques from fields such as statistics, artificial intelligence, operations research, digital signal processing, pattern recognition, cognitive psychology, information theory and decision theory [47]. The aim of the information fusion engine is to compose the combined strength of the submitted attributes during multi-modal authentication. Information fusion technology has been applied most prominently in military applications such as battlefield surveillance and tactical situation assessment [46]. It has also emerged in commercial applications such as robotics, manufacturing, medical diagnosis, and remote sensing [47]. Artificial neural networks, fuzzy logic, Bayesian method, evolutional computation, hybrid intelligent systems and data mining technologies are considered as examples of the technologies that can be used to implement an information fusion technique [48].

2.3.1 Artificial Neural Networks

The segment of artificial intelligence called artificial neural network (ANN) aims at emulating the function of the Biological Nervous System that makes up the brains found in nearly all higher life forms found on Earth [48]. Neural networks are made up of neurons and a neuron is made up of a core cell and several long connectors, which are called synapses [47]. These synapses show how the neurons are connected amongst themselves. Both biological and artificial neural networks work by transferring signals from neuron to neuron across the synapses. An artificial neural network is therefore an "information-processing paradigm that is inspired by the way biological nervous systems like the brain processes information" [48]. The key element of this concept is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (called neurons) working in unison to solve specific problems.

An artificial neuron commonly known as a perceptron is a device with many inputs and one output as shown in Figure 3 [48]. The neuron has two modes of operation namely the training mode and the using mode [49].

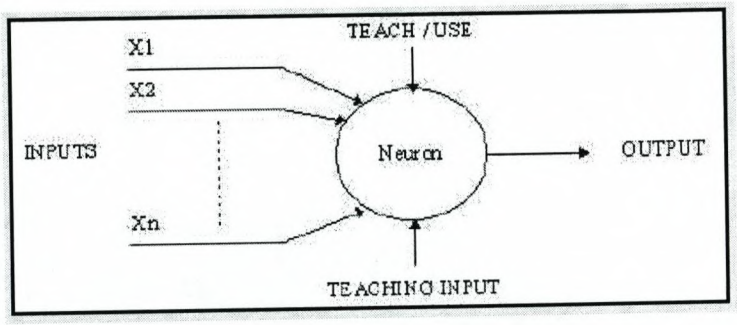

Figure 3: A Simple Neuron (Perceptron)

In the training mode, the neuron can be trained to fire or not to fire, for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not [49].

A trained artificial neural network is an expert in the category of information it has been given to analyse and can then be used to provide projections given new situations [48]. This makes it more useful when implementing information fusion. Once the network is trained and the network weights and threshold values are generated using a given range of training data to obtain a given set of targeted data, it can then work as an expert to compute the output of the information fusion based on the values given. The challenge when using artificial neural networks is how to compute the neuron weights and threshold values. Neural networks with their extraordinary ability to derive meaning from complicated or imprecise data can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques [49].

2.3.2 Fuzzy Logic

Most of the time human thinking and reasoning involve inexact information because much of human knowledge is vague and imprecise [48]. The sources and the nature of inexact information may be different for different problem domains. The following can be identified as the possible reasonable sources of inexactness of information. These are lack of adequate data, inconsistency of data, inherent human fuzzy concepts, matching of

similar rather than identical situations, differing opinions, ignorance, imprecision in measurements and lack of available theory to describe a particular situation [50].

Fuzzy inference is the process of mapping a given problem from a given input to an output using the theory of fuzzy sets. Two commonly used fuzzy inferences are Mamdani and Sugeno-style inference system [48]. Professor Ebrahim Mamdani of London University built one of the first fuzzy systems to control a steam engine and boiler combination [48]. He applied a set of fuzzy rules supplied by human experienced operators. In the Mamdani-style fuzzy inference process, there are four major steps, which include fuzzification of the input variables, rule evaluation, aggregation of the rule outputs, and finally defuzzification [48].

Sugeno-style fuzzy inference is very similar with that of Mamdani method but only differ where Sugeno changed the rule consequent [48]. Instead of a fuzzy set he used a mathematical function of input variable. Here a fuzzy singleton is a fuzzy set with a membership function that is unit at a single particular point on the universe of discourse and zero anywhere else. In this case the output of each fuzzy rule is a constant, which means that all consequent membership functions are represented by singleton spikes [48]. To get the crisp output (defuzzification), the weighted average of all the singleton points are computed.

Many practical applications, allows the natural use of vague and imprecise concepts of fuzzy logic for commonsense reasoning and explanation [50]. The disadvantages of fuzzy logic include membership functions being difficult to find. Also there are multiple ways of combining evidence in Fuzzy Logic and finally the problems with long inference chains are difficult to deal with [50].

2.3.3 The Bayesian Method

One of the most common characteristics of the information available to human experts is its imperfection [48]. Information can be incomplete, inconsistent, uncertain, or all the three. Therefore uncertainty can be described as "the lack of exact knowledge that would enable us to reach a perfect reliable conclusion" [51]. A number of numerical and non-numerical methods have been developed to deal with uncertainty in real based expert

systems [48]. One of the most popular uncertainty management paradigms is the Bayesian reasoning.

Bayesian reasoning or inference is "a statistical inference in which probabilities are interpreted as degrees of belief" [48]. The name comes from the frequent use of the Bayes' theorem named after the Reverend Thomas Bayes [52]. The equation below is used to represent Bayesian reasoning and provides a background for the application of probability theory to manage uncertainty in expert systems [48] [52].

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)} \tag{2.1}$$

In equation 2.1, A is an hypothesis. *Pr(A)* is called the prior probability of A. Pr(B | A) is called the conditional probability of seeing the evidence B given that the hypothesis A is true. . It is also called the likelihood function when it is expressed as a function of A given B. Pr(B) is called the marginal probability of B, the probability of witnessing the new evidence B under all mutually exclusive hypotheses.

As can be seen from the equation above, the outcome of the Bayesian formula is a single value. This equation can be used to implement a technique of information fusion to obtain the required single value. It is this outcome that can then be used in multi-modal authentication in a digital identity management system.

The Bayesians method has proved to be a useful technique and has a lot of applications in real life situations. Sound theoretical foundation and well-defined semantics for decision making are some of the advantages of the Bayesian method [52]. Some problems of the Bayesian method include requiring large amounts of probability data and sufficient sample sizes. The subjective evidence may not be reliable, also the relationship between the hypothesis and evidence is usually reduced to a number and lastly it involves high computational overhead [48].

2.3.4 Evolutionary Computation

This is another new area of artificial intelligence technologies. The evolutionary approach to machine learning is based on computational model of natural selection and genetics

[48]. They are called evolutionary computation. This is an umbrella term that combines genetics algorithms, evolutionary strategies and genetic programming [48].

Genetic algorithms (GA) are a class of stochastic search algorithms based on biological evolution [48]. By imitating the process of evolution using computer instructions and algorithms, scientists have been trying to mimic the intelligence associated with the problem solving capabilities of the evolution process through evolving the bit-string representation of some problem [53]. A GA represents an iterative process where each iteration is called a *generation.* A typical number of generations for a simple GA can range from 50 to over 500 [48]. The entire set of generations is called a *run.* At the end of a run it is expected to find one or more highly fit chromosomes [48].

One of the central challenges of computer science is to get a computer to do what needs to be done, without telling it how to do it. Genetic programming (GP) which is a recent development in the area of evolutionary computation offers a solution through the evolution of computer programs by methods of natural selection [48]. Genetic programming addresses this challenge by providing a method for automatically creating a working computer program from a high-level problem statement of the problem. It achieves this goal of automatic programming by "genetically breeding a population of computer programs using the principles of Darwinian natural selection and biologically inspired operations" [54]. GA is another technology that can be used to implement information fusion technique in addition to the above artificial intelligent technologies.

2.3.5 Hybrid Intelligent Systems

Of the intelligent technologies considered above, which include probabilistic reasoning, fuzzy logic and neural networks, each of these has a weak and strong point. Many real world applications would need a combination of these technologies. A hybrid intelligent system can therefore be defined as one that combines at least two of these technologies [48]. Examples would be a combination of probabilistic reasoning (Bayesian), fuzzy logic, neural networks and evolutionary computing to form *soft computing* (SC) where as a combination of neural networks with fuzzy logic results in a hybrid *neuro-fuzzy system* [48]. Finally a combination of neural networks and a rule-based expert system yields a *neural expert* system. Below is a brief description of two hybrid intelligence systems.

Neural networks and expert systems as intelligent technologies both attempt to imitate human intelligence and eventually create intelligent machines [48]. They however differ in the way they represent knowledge and do data processing techniques to achieve their goals. For example, while experts systems rely on logical inferences and decision trees and focus on modelling human reasoning, neural networks rely on parallel data processing and focus on modelling a human brain [48]. A hybrid system that combines a neural network with a rule-based expert system is called a *neural expert system* [48]. Learning, generalisation, robustness and parallel information processing makes neural networks a right component for building a new breed of expert system.

Fuzzy systems usually lack the ability to learn and cannot adjust themselves to a new environment, while on the other hand, although neural networks can learn, they are opaque to the user [48]. Therefore merging the two neural networks with fuzzy system offers a promising approach to building intelligent systems. A *neural-fuzzy* system can combine the parallel computation and learning capabilities of neural networks with the human like knowledge representation and explanation abilities of fuzzy systems [48]. In fact, the *neuro-fuzzy* system is a neural network that is functionally equivalent to the fuzzy inference model with a structure similar to a multi-layer neural network [48].

The hybrid intelligent systems offer a much better and robust technology for creating an information fusion technique as it integrates more than one artificial intelligent technologies. However it may be difficult to work with due to the complexity of the technologies.

2.3.6 Data Mining

The process of extracting knowledge from data is called data mining. It is defined as the "exploration and analysis of large quantities of data in order to discover meaningful patterns and rules" [48]. The ultimate goal of data mining is to discover knowledge. Data mining is based on intelligent technologies such as neural networks. However the most popular tool used in decision-making is a decision tree [48].

Data mining has been applied in a number of areas to solve problems. The types of problem areas include the classification problem, clustering and dependence modelling [48]. This makes it an ideal technology for building an information fusion technique.

2.4 Related Works

There are a number of standards and digital identity managements systems created by various organisations on the market today. Below is an example of some of these products in digital identity management systems and some related works in the area of information fusion techniques implementations using the technologies in section 2.3.

Microsoft created .NET Passport for digital identity management which is a centralised online user authentication service that allows web users to use their e-mail address and a single password to securely sign in and obtain services from any .NET Passport-participating websites [55]. Once someone has signed in at the .Net Passport website, he or she can access the services provided by the other web sites affiliated to Microsoft without signing in again at their websites. This is referred to as Single Sign On [56].

The Internet2 Consortium led by about 205 universities working in partnership with industry and government is working to develop and deploy advanced network applications and technologies, accelerating the creation of future Internet. Shibboleth is an identity management framework and specifies architectures, policy structures and technologies designed mainly for academic institutions using open source implementation [57].

Liberty Alliance project formed in September 2001 is made up of more than 160 companies, representing service providers, technology firms, financial institutions, educational institutions and government organisations [58]. Its main goal is to develop open standards for federated network identity management and identity based services. The Project does not deliver specific products or services but its goals are to create specifications creating a means for its members and other organisations to build products and services that will interoperate and promote secure federated identity management systems based on commonly available industry standards [59]. Currently there are more than 30 Liberty enabled products available in the market [56].

In the cyber-space, there have been efforts on creating a unique identity reference and development of frameworks to deliver services from converged service architectures [58]. Examples of such schemes are Electronic Number Mapping (E-NUM) [59], Universal Communications Identifier (UCI) [60] and i-Name [61]. These schemes require

a particular user to have a single identity different from other users around the world and which could be used to access any form of service from any point around the world.

Six technologies have been highlighted above that can be used to implement information fusion. Of the six examples, artificial neural networks have broad applicability to real world business problems. Since they are the best at identifying patterns or trends in data, they are well suited for prediction or forecasting needs including sales forecasting, industrial process control, customer research, data validation, risk management and target marketing. [49]. In almost all of the above mentioned applications information fusion is used to combine the data or information from different sources in order to obtain meaningful deduction from the input data [48].

A number of applications that require grouping and classifying of information without clear distinct boundaries have found Fuzzy Logic more applicable [48]. One of the main applications of fuzzy logic is fuzzy control. In most fuzzy control applications, Mamdani-type of approach is used. Some of these projects include fuzzy *fusion based landmine detection*, Fuzzy Systems to Evaluate Weather and Terrain Effects on Military Operations, Minimising Tremor in a Joystick Controller using Fuzzy Logic [62] all of which require combining more than one set of data to derive meaningful output [46].

Bayesian inference has been used in recent years to develop algorithms for identifying unsolicited bulk e-mail Spam [48]. Applications that make use of Bayesian inference for spam filtering include Bogofilter, SpamAssassin and Mozilla all of which implement some form of information fusion technique [52].

Data mining uses the technique of ANN and genetic algorithm to solve problems in classification, clustering, continuous classes and dependence modelling [48]. Data mining has found applications in Audio mining, Biometric mining, Image mining, Text mining etc. Applications currently in use that are using data mining are fraud detection, Stock market analysis, crime detection and homeland security all of which involve some form of data or information fusion technique [63].

2.5 Summary

This chapter looked at the background information of digital identities and management systems and information fusion technologies. The first part of Chapter 2 focused on digital identities and management systems. The basic terms used in digital identity management systems were first defined to set the stage. The common sources of the attributes used to build identity systems were then highlighted. These included the credentials used in daily life to access services offered both on the cyber space and real space. Desirable properties of digital identities used to analyse the credential attributes were discussed and this was followed by a way of classifying the identities for easy analysis. To obtain a digital representation of a person in an identity system, the scheme referred to as CAMA was discussed. Since the goal of this thesis is to reduce cases of identity fraud and theft, identity fraud was discussed together with issues surrounding identity privacy and how they affect the current digital identity systems. The authentication, authorisation and audit actions were then discussed. Lastly but not the least the four types of commonly used storage mechanism for digital identities in use today were discussed. The second part of the background information theory looked at the information fusion technologies. Six technologies were considered and these included artificial neural networks, fuzzy logic, Bayesian methods, evolutional computation, hybrid intelligent systems and data mining. Advantages and disadvantages were given showing the strength and weakness of each technology. The last part of the background theory focused on the related works in both digital identity management systems and information fusion technologies.

With the information fusion technologies and the digital identity management systems analysis given in this chapter, it should be possible to obtain an effective digital identity management system using the technique of information fusion to implement multi-modal authentication. The next chapter looks at the analysis of the credential attributes used to build a multi-modal authentication system in digital identity management system.

Chapter 3

3 CREDENTIAL ATTRIBUTES

3.1 Introduction

An identity management system needs a set of credential tokens whose attributes are used to identify a user when accessing the online services. This chapter looks at the process of selecting the credential tokens used in most countries to access diverse services. The attributes are then extracted from these credential tokens. This is then followed by the representation and then the grouping of these credential attributes. The chapter is concluded by the process of computing the weight of each attribute by using a questionnaire and Shannon's Information Theory.

3.2 Choosing the Credential Identity Tokens

Currently, individuals acquire many credentials and become involved in many activities as they move from one stage of life to another. This makes it extremely challenging as to which credentials must be selected as the source for the attributes to be used when coming up with a digital identity system. To maintain simplicity, this thesis considers the credentials in use today in most countries for service delivery.

A questionnaire was prepared and distributed to a wide range of responders in two countries namely Zambia and South Africa. This choice provides a wide range of individuals and racial mixtures. A number of people were targeted and they include accountants, students, IT specialists, lecturers, government officials and foreign students. In Zambia, the survey was carried out at the University of Zambia while in South Africa; it was carried out at the University of the Western Cape and University of Cape Town. In the questionnaire the most commonly used credentials [2] were selected, namely national passports, national identity cards, birth certificates, citizenship certificates, acceptable ID cards (such as employee ID cards), driving license, credit cards, bank cards (such as

ATM, MasterCard or VISA), insurance membership card, student cards, a mobile phone (as used in m-commerce), Internet terminal (as used in e-commerce) and school certificates (like diplomas and degrees earned from universities and colleges) [2].

Responders were requested to add any five credentials they have seen used either in their countries or any where else around the world. The respondents were then requested to grade each credential out of five. The grading was done based on their perceived level of importance of that token. A credential therefore received a grade of five if it was extremely important to the responder and a one if it was not import or useful.

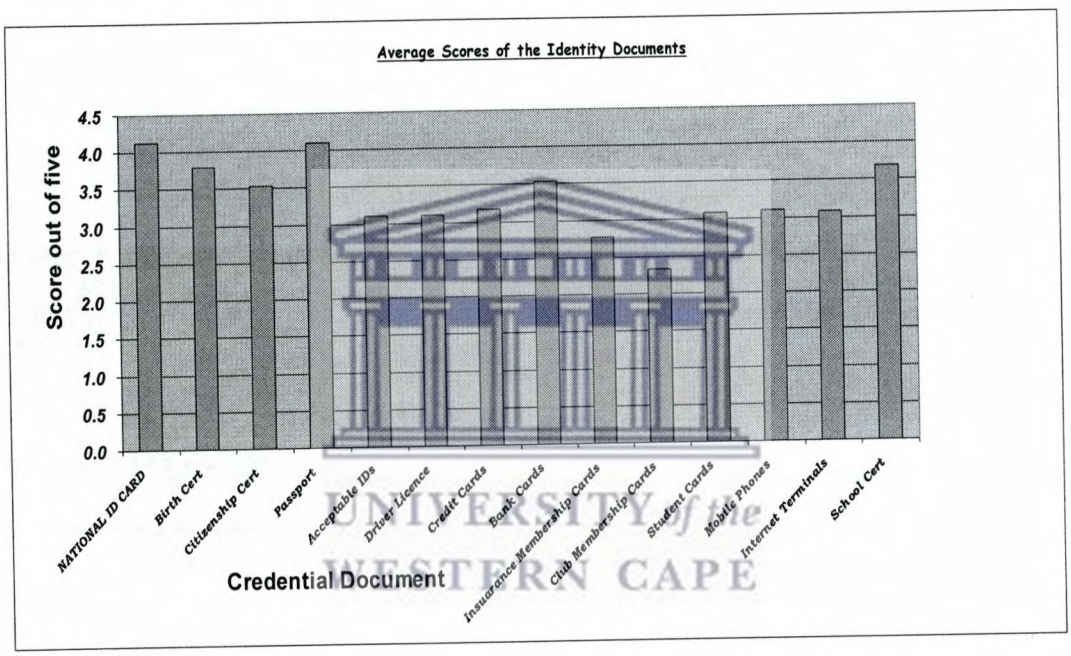

Figure 4: Average Scores of Identity Documents

Figure 4 shows the compiled results. As a result of the survey, fourteen credentials emerged as the most commonly used ones. As can be seen from the graph, the national ID cards and national passports have an average score of more than four meaning they are considered to be very important or useful documents. Bank cards, birth certificates and citizenship certificates all have average scores of more than 3.5 meaning that they are considered as second level of important documents.

Club membership card has the lowest score with less than 2.5 meaning that it is not considered to be a very important credential for digital identity. These credentials were then used as the source of the attributes required to build a multi-modal digital identity management system.

3.3 Extracting and Grouping the Attributes

After coming up with a set of credentials in section 3.2, the attributes were then extracted from those credentials. In practice this process needs to be automated. In this thesis the fourteen credentials above in addition to a set of biometric attributes (features) were used as the source of the attributes. Below is an outline of the process of extracting the attributes from the biometric features and the credential tokens above.

Almost every form of online service today requires the user to have a form of a secret code which is known by him or her and the service provider only. These secret codes come in different forms depending on the application. For example to access emails, the users need a password as a secret code in addition to their username (from the email address) and to use bank cards they need a Personal Identity Number (PIN) as a secret code in addition to their bank cards. Table 2 shows an example of two attributes used as secret codes.

As individuals move from one stage of life to another they acquire various forms of credentials. These credentials come in different forms depending on their use. Good examples are the national identity cards (ID books) given to all citizens in a country after the age of 16 (e.g. South Africa) and the national passports used to identify individuals crossing international boundaries. Nine of the fourteen credentials listed in Figure 4 were broken down into their respective identity attributes as shown in Table 2. The choice of which attribute is selected from these credentials is dependant on the security level of the application to be developed.

A number of biometric characteristics exist and are in use in various applications today [19]. Biometric technologies work by measuring and analyzing human physiological or behavioural characteristics [16]. Examples of physiological characteristics are finger print, face recognition and hand geometry while behavioural characteristics are based on

data derived from a person's actions and these include voice and signature recognition [16] as shown in section 2.2.2 in the background theory. Table 2 lists eight biometrics features in use today [19].

The recent years have seen a rise in a number of users using non-human communication devices such as mobile phones and Internet terminals to access the online services. With this trend of events, it is expected that the use of these devices in e-commerce and m-commerce will be on the increase in the coming years [2]. There is therefore a need to determine a more effective way to authenticate both the user and the device used to access the online services. This thesis will consider three types of communication devices namely a mobile phone, an Internet terminal and card based credentials used to access online services.

Apart from using a mobile phone to communicate one with another, currently it is also being used to access a wide range of online services [14]. A good example of such services is the telephone banking. The cellular network provider usually maintains its own subscriber database for the user's identity attributes. Section 2.2.2 in the background theory gives a set of attributes stored in the subscriber database. Only five of these attributes stored in the subscriber database by the service providers are considered in this thesis as shown in Table 2 .

The last two decades have seen a sharp rise in the number of users using the Internet terminals to access the online services [13]. This has also seen a rise in the number of cases in identity theft and online fraud. Therefore, there is a need to properly identify the Internet terminals used to access the online services in addition to the user using that particular terminal. Section 2.2.2 in the background theory highlighted the two methods used to connect to the Internet. Table 2 gives a summary of the attributes obtained from an Internet terminal using these two modes of connections.

The last group of credentials used to access the online services comes in form of a card. Good examples include credit cards, bank cards, insurance membership card and student cards. These cards use different data storage techniques and different access methods to access the data stored in them. Examples of the storage mechanisms are the barcode, magnetic strip and an Integrated Circuit (IC) memory like ROM, EEPROM or non-

volatile RAM (section 2.2.8) [34]. To maintain simplicity, only the magnetic strip storage mechanism is considered in this thesis and only five attributes extracted from the cards based credentials are considered as shown in Table 2 .

Grouping the credential attributes makes it easy to analyse and work with them [2]. In this thesis all the credential attributes are grouped into four classes. These are the pseudo metrics, physical metrics, device metrics and biometrics. Section 3.4 gives the definitions, advantages and disadvantages of using each group.

3.4 Description of the Multiple Metrics

The system being developed in this thesis depends on multiple metrics, which include pseudo, physical, biometrics, and device metrics. The *pseudo metrics* attributes refer to what you know. This is usually a secret code shared between the user and the service provider [19]. It comes in different forms depending on the type of service offered and the nature of the credential used to access the service. For example when using a card based credential like a smart card, a PIN number is used as a pseudo metric whereas when the Internet terminal is used to access emails, then the password is used as a pseudo metric. Since they are simple and easy to implement, pseudo metrics are highly in use today as they do not require specialised equipment like the device metrics. The disadvantage is that they are easy to guess and there is no guarantee that the user is the owner of the credentials when used online [2].

Physical metrics refers to what you have acquired by virtue of being born as an entity. These are the attributes which users acquire from the time they where born. The most important and commonly used physical metric is a name. Other examples include date of birth, address and the National Identity Number. The advantage of these attributes is that most users are familiar with them since they are used in daily life. The disadvantage is that it is difficult to prove online whether the person is whom he or she claims to be. Hence other metrics are usually combined with physical metrics to obtain more robust user digital identity.

Biometrics refers to what you are biologically and basically fall in one of the two classes [16]. The first class is referred to as active biometrics. These are physiological behaviour

of a human being and include features like voice recognition and digital signature. The second class is referred to as passive biometrics. These are the user's physical characteristics, which include fingerprint, retina and iris scan. Seven of these features have already been applied in real life applications [19]. The advantage of using biometrics over physical metrics, pseudo metrics or device metrics is that biometrics elements cannot be misplaced, stolen, forgotten or duplicated. The major disadvantage is that not all people may have the biometrics traits being sought. For example someone may not have a hand or may have poor and damaged finger print ridges [2].

Device metric refers to the attributes from telecommunication devices used to access services offered online. This thesis considers three devices namely a mobile phone [15], an Internet terminal [13] (computer connected to the Internet) and card based devices in the form of smart cards or magnetic strip cards [64]. Examples of device metrics include International Mobile Equipment Identifier (IMEI) from a mobile phone and the MAC address from an Internet terminal. The device metrics are not easy to forge and if the credential is stolen the owner can easily notice that and deactivate the token hence making them better in terms of security. The disadvantage of device metrics is that the token-based credentials are costly to deploy and operate. Users are also fond of losing the tokens more regularly and replacing them is very expensive [2]. Table 2 gives a summary of the extracted attributes, their sources and their groupings

Index	Attributes	Grouping and Source of Attribute
1	National Identity Number	**PHYSICAL METRICS** **From** National Identity Cards, Birth Certificates, Citizenship Certificates, National Passport, Acceptable ID Cards, Driver License, Student cards, Insurance Membership Card, School Certificates
2	Full name	
3	Residential Address	
4	Date of birth	
5	City of birth	
6	Country of birth	
7	Race	
8	Mother's Name	
9	Father' Name	
10	Eye colour	
11	Certificate number	
12	Citizenship	
13	Height	
14	Signature	
15	Personal Identity Number (PIN) (e.g. used on credit cards)	**PSEUDO METRICS**

16	Password (e.g. used on your emails)		From Secret Codes
17	International Mobile Equipment Identifier (Phone Identity Number)		DEVICE METRICS From Communication Devices (Mobile Phone)
18	International Mobile Subscriber Identifier (SIM Identity Number)		
19	SIM Serial Number (SSN)		
20	Mobile Station ISDN Number (Phone Number we dial)		
21	PUK/PIN Number		
22	Network Interface Card (NIC)	IP Address	DEVICE METRICS From Communication Devices (Internet Terminal)
		Media Access Control (MAC) Address	
23	Modem	Username	
		Password	
		Phone Number	
24	Unique Card ID Number		DEVICE METRICS From National Identity Cards, Credit Cards, Bank Cards, Insurance ID Cards, Student ID Cards, Acceptable IDs & Club Membership Cards
25	Name of User		
26	Expiration Date of Card		
27	Registration Date		
28	User's Unique Service Number		
29	Face Recognition		BIOMETRICS From Your body features and physiological behaviour
30	Facial Thermogram		
31	Fingerprint		
32	Hand Geometry		
33	Iris scan		
34	Retina Scan		
35	Signature Recognition		
36	Voice Recognition		

Table 2: Grouping of the Credential Attributes

3.5 Representation of Digital Identities

Before storing each of a user attributes, an effective method of representing them for storage is required. Only a general description on storage is given here and no attempt has been made to optimise or compress the codes given to the attributes. The alphanumerical characters are encoded using 2-octect UTF-8 Sequence Unicode [65]. This scheme complies with ISO10646 and has been designed for ease of use with existing ASCII-based systems [66]. Each 2-octect UTF-8 Unicode occupied 16-bit space. The dates on the other hand were represented by the "ISO 8601 standard format, which is YYYYMMDD where YYYY is the year in the Gregorian calendar, MM is the month of

the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31" [67]. For example, 20060901 represents the first day of September in 2006 [2]. Table 3 shows a summary of the representation of these attributes.

Representation of Digital Identities			
Index	**Attributes**	**Parameters**	**Bit Space**
PSEUDO METRICS			
1	Password	ASCII Representation (20 char)	160
2	Personal Identity Number (PIN)	ASCII Representation (10 char)	80
3	Private Keyword	ASCII Representation (20 char)	160
PHYSICAL METRICS			
1	National ID Number	ASCII Representation (18 char)	144
2	Full Name	ASCII Representation (40 char)	320
3	Residential Address	ASCII Representation (64 char)	512
4	Date of birth	Number of days since the year 1900	27
5	City of birth	Postcode (8 digits ASCII)	64
6	Country of birth	International dialling code (4 digits ASCIII)	16
7	Race	20 character ASCII	160
8	Mother's name	40 character ASCII	320
9	Father's name	40 character ASCII	320
10	Eye colour	4 bit representation of 16 colours	4
11	Certificate number	20 character ASCII	160
12	Citizenship	20 character ASCII	160
13	Height	12 bit representation of 3000 mm	12
BIOMETRICS			
1	Face recognition	84-byte digital template	672
2	Fingerprint	Fingerprint Template of 1000 bytes	8000
3	Hand and finger geometry	250 – 1000 bytes	8000
4	Iris	512 byte IrisCode(tm)	4096
5	Signature	1.5KBytes of signature verification data	12000
6	Voice	10000 - 20000	160000

DEVICE METRICS (MOBILE PHONE)				
1	International Mobile Equipment Identifier (IMEI)		15 digits ASCIII	120
2	International Mobile Subscriber Identity (IMSI)		15 digits ASCIII	120
3	SIM Serial Number		10 bytes	80
4	Mobile Station ISDN number		16 digits ASCIII	128
5	PUK/PIN number		4 bytes	32
DEVICE METRICS (INTERNET TERMINAL)				
1	Network Interface Card (NIC)	Media Access Control Address (MAC Address)	12 hexadecimal digits	48
		Internet Protocol (IP) Address	32 bit representation	32
2	Modem	Username	ASCII Representation (20 char)	160
		Password	ASCII Representation (20 char)	160
		Phone Number	ASCII Representation (12 char)	96
DEVICE METRICS (CARD BASED CREDENTIAL TOKEN)				
1	Unique Card ID number		Up to 19 characters	152
2	Format or File System used		One character (alpha only) ASCIII	8
3	Full Name of a User		Two to 26 characters	208
4	Expiration date of card		Four characters or one character ASCIII	32
5	Discretionary data		Enough characters to fill out maximum record length (79 characters total)	632

Table 3: Representation[3] of Credential Attributes[4]

This section looked at the representation of the attributes from the different groups. This information helped to determine how much space was actually required to store a given set of attributes required to authenticate a user when building a digital identity system. The next section looks at the methods used to compute the weights of the attributes that are used to build an information fusion engine.

[3] U.S. GAO, Using Biometrics for Border Security, (Washington, DC: GAO, November 2002), p. 46.
[4] http://money.howstuffworks.com/credit-card.htm

3.6 Computing the Weights of the Attributes

To effectively determine the weights of the attributes to be used in the information fusion engine, two methods were used. The initial scores of the attributes were computed using a questionnaire and then Shannon's information theory was used to compute the final weight of each attribute.

3.6.1 Using a Questionnaire

Apart from using a questionnaire to analyse the credentials, which acted as the source of the attributes, it was also used to obtain the initial scores of the attributes. In this mean opinion score method, respondents were requested to grade each attribute using the desirable properties of digital identities [7]. The physical metrics, pseudo metrics and device metrics were graded using the same set of desirable properties of digital identities namely Uniqueness, Verifiability, Consistency, Persistency and Trust [7]. For the biometric features, the standard method of testing a biometric feature is to see if it can pass [16] a test of seven standard desirable properties of digital identities [7]. These properties include Universality, Distinctiveness, Permanence, Collectability, Performance, Acceptability and Circumvention [2] but only five of these were directly applicable in analysing the biometrics.

Each attribute had five desirable properties of digital identities of which each was graded out of five. For example using uniqueness as a desirable property of digital identities, the grading was done based on the perceived level of uniqueness of that attribute to the respondent. A desirable property of digital identities therefore received a grade of five if it is extremely unique to the responder and a one if it is not unique at all. The average scores for each desirable property of the digital identities for every attribute were thereafter computed. Microsoft Excel was used to analyse the results in order to determine the average scores. With five desirable properties of digital identities for each attribute, Table 4 and Table 5 show the computed average scores from the questionnaires.

Index	Attributes	Average Scores of Desirable Properties of Digital Identity				
		Uniqueness	Verifiability	Consistency	Persistency	Trust
PHYSICAL METRICS						
1	National Identity Number	4.63	4.04	4.15	3.74	3.85
2	Full name	3.11	2.85	3.31	3.42	2.92
3	Residential Address	2.44	2.59	2.59	2.00	2.41
4	Date of birth	3.19	3.48	3.73	3.58	3.15
5	City of birth	2.30	2.69	3.00	3.42	2.69
6	Country of birth	2.96	3.31	3.36	3.60	2.68
7	Race	2.56	3.15	3.33	3.45	3.04
8	Mother's Name	2.56	2.89	3.15	3.30	2.96
9	Father' Name	2.59	2.89	3.19	3.37	3.65
10	Eye colour	2.70	2.93	2.92	3.04	2.54
11	Cert. number	3.62	3.46	3.00	3.08	2.67
12	Citizenship	3.30	3.59	3.31	3.04	3.08
13	Height	2.04	2.93	2.65	2.52	2.31
14	Signature	3.67	3.22	3.15	3.00	3.15
PSEUDO METRICS						
1	Personal Identity Number	4.07	3.96	3.30	3.11	3.37
2	Password	3.96	3.63	3.22	2.56	3.22
DEVICE METRICS (Mobile Phone)						
1	International Mobile Equipment Identifier	3.73	3.62	3.54	2.82	2.88
2	International Mobile Subscriber Identifier	4.07	3.67	3.52	3.04	3.22
3	SIM Serial Number (SSN)	3.88	3.42	3.35	3.15	2.88
4	Mobile Station ISDN Number	3.50	3.27	3.15	2.88	2.92
5	PUK/PIN Number	3.58	3.27	3.12	2.62	2.77
DEVICE METRICS (Internet Terminal)						
1	Internet Protocol (IP) Address	4.06	3.24	3.12	2.82	2.76
2	Media Access Control (MAC) Address	3.91	3.55	3.27	3.23	3.05
DEVICE METRICS (Card Based Tokens)						
1	Unique card ID number	4.58	3.38	3.12	3.23	3.17
2	Format or File System used	3.22	3.04	3.19	2.85	2.65
3	Full Name of a User	2.84	2.92	2.75	3.08	2.91
4	Expiration date of card	2.42	2.91	2.70	2.61	2.51
5	Discretionary data	3.92	3.50	3.25	3.17	3.22

Table 4: Computed Average Scores of Physical, Pseudo and the Device Metrics

As seen from Table 4 the National Identity Number (NID) had the highest average score for all the five desirable properties of digital identities for the physical metrics group

meaning users had more faith in using the NID number than any other physical metric attribute. Height had the lowest score for uniqueness while residential address had the lowest score for verification, consistency, persistency and trust.

Only two most commonly used pseudo metrics namely password and the PIN number were considered. As shown in Table 4, the PIN number was rated as being more unique, verifiable, consistent, persistent and trustworthy than a password. However both the password and the PIN number had their lowest scores for persistent. Meaning they are changed more often which becomes an advantage in terms of security.

Only five attributes were analysed from a mobile phone as shown in Table 4. The International Mobile Subscriber Identity (IMSI) had the highest score for both uniqueness and verifiability while the Mobile Station ISDN number had the lowest score for uniqueness and verifiability. The IMSI is used by the service providers to identify users. Each IMSI may have more than one MSISDN numbers assigned to it but every IMSI is very unique as described in [15] hence confirming the above results.

Only the attributes from the terminal connected to the Internet using a Network Interface Card (NIC) were considered to maintain simplicity. The Internet Protocol was rated as being more unique than the MAC address. The MAC address on the other hand was rated as being more consistent, verifiable and trustworthy than the IP address which represents findings in [68].

Five attributes from magnetic strip cards were considered using the five desirable properties of digital identities. The card's unique number had the highest scores for both uniqueness and verifiability where as the card's expiry date had the lowest score for all the desirable properties of digital identities.

Unlike the other three groups, biometrics features were analysed using distinctiveness, universality, permanence, performance and circumvention as shown in Table 5. Fingerprint had the highest average scores in all the five desirable properties of digital identities. This means finger print was considered to be more distinctive, universal and could perform better as compared to the others biometrics. On the other hand voice recognition had the lowest score for distinctiveness while hand geometry had a lower score for universality.

Index	Attributes	Desirable Properties of Digital Identity				
		Distinctiveness	Universality	Performance	Permanence	Circumvention
1	Face Recognition	3.77	3.38	3.12	3.23	3.17
2	Facial Thermogram	3.50	3.00	3.18	2.91	3.05
3	Fingerprint	4.54	3.77	3.76	3.96	3.88
4	Hand Geometry	3.25	2.96	3.04	3.08	2.73
5	Iris scan	3.76	3.16	3.33	3.56	3.35
6	Retina Scan	3.83	3.46	3.52	3.58	3.27
7	Signature Recognition	3.42	3.08	2.80	2.81	3.00
8	Voice Recognition	3.19	3.00	2.88	2.62	2.67

Table 5: Computed Average Scores of Biometrics Attributes

These initial average scores obtained from the questionnaire were then used to compute the final weight of each attribute. The next section looks at how Shannon's Information theory was used to accomplish this.

3.6.2 Application of Shannon's Information Theory

Using the initial scores in the previous section, this section uses the principles of information theory to compute the weight of each attribute. In Shannon's information theory, the entropy of a set of related events is defined as the average information content [24]:

$$H(X)=\sum_{i=1}^{n} p(X_i)\log_2\left[\frac{1}{p(X_i)}\right] \qquad (3.1)$$

Where;

X is the message

X_i is the i^{th} symbol in the message

$p(X_i)$ is the probability of the occurrence of the i^{th} symbol

This formula was then applied to the attributes' initial scores in Table 4 and Table 5. The i^{th} symbol in the message was the first, second, third, fourth or fifth desirable property of

digital identity. The average information content of an attribute was therefore expressed as:

$$H = \sum_{i=1}^{n} p_i \log_2 (1/p_i) \qquad (3.2)$$

Where p_i is the contributing weight of the i^{th} desirable property of a digital identity. In this method, the probability p_i is given as the ratio of the computed average score of a desirable property of digital identity to the sum of the weight of all the average scores of desirable properties of that attribute:

$$p_i = \frac{a_i}{\sum_{i=1}^{n} a_i} \qquad 3.3$$

Where a_i is the weight of a desirable property of digital identity i. Below is an example of how to compute the information content of the National Identity Number (NID) using equations (3.2) and (3.3). Using the average scores of the desirable properties of NID in Table 4, the values for each desirable property were given as follows;

a_1 = Uniqueness = 4.63

a_2 = Verifiability = 4.04

a_3 = Consistency = 4.15

a_4 = Persistency = 3.74

a_5 = Trust = 3.85

Therefore the value of p_i of the NID from uniqueness using equation (3.3) is given by;

$$p_1 = {}^{4.63}/_{20.41} = 0.2268$$

The entropy contributed by uniqueness property of the NID is given by;

$$p_1 \log_2 (1/p_1) = 0.48550$$

Therefore performing similar calculations for the other four desirable properties of the NID and summing them gives the required average information content of the NID as:

$$H(National\_Identity\_Number) = \sum_{i=1}^{5} p_i \log_2(1/p_i) = 2.31786$$

Therefore 2.31786 is the equivalent entropy yield of the National Identity Number and is taken as its computed weight. Performing similar calculations on the other attributes using the initial scores from the questionnaire would give the required entropy yields as shown in Table 6 and Table 7. The variables W1, W2, W3, W4 and W5 in the tables represent the contributing entropies of the respective desirable properties of the digital identities shown in the third column.

Index	Attributes	Desirable Properties of Digital Identity					$P_i \log_2(1/P_i)$					$\sum P_i \log_2(1/P_i)$
		Uniqueness	Verifiability	Consistency	Persistency	Trust	W1	W2	W3	W4	W5	
PHYSICAL METRICS												
1	National Identity Number	4.63	4.04	4.15	3.74	3.85	0.48550	0.46256	0.46728	0.44861	0.45392	**2.31786**
2	Full name	3.11	2.85	3.31	3.42	2.92	0.46371	0.44794	0.47446	0.47990	0.45239	**2.31839**
3	Residential Address	2.44	2.59	2.59	2.00	2.41	0.46684	0.47701	0.47701	0.43035	0.46468	**2.31589**
4	Date of birth	3.19	3.48	3.73	3.58	3.15	0.45157	0.46712	0.47889	0.47200	0.44926	**2.31884**
5	City of birth	2.30	2.69	3.00	3.42	2.69	0.42672	0.45597	0.47503	0.49569	0.45597	**2.30938**
6	Country of birth	2.96	3.31	3.36	3.60	2.68	0.45140	0.47123	0.47378	0.48510	0.43285	**2.31435**
7	Race	2.56	3.15	3.33	3.45	3.04	0.42873	0.46685	0.47633	0.48215	0.46058	**2.31465**
8	Mother's Name	2.56	2.89	3.15	3.30	2.96	0.43710	0.45942	0.47441	0.48210	0.46367	**2.31670**
9	Father' Name	2.59	2.89	3.19	3.37	3.65	0.42900	0.44956	0.46726	0.47662	0.48943	**2.31186**
10	Eye colour	2.70	2.93	2.92	3.04	2.54	0.45625	0.47066	0.47008	0.47689	0.44506	**2.31895**
11	Certificate Number	3.62	3.46	3.00	3.08	2.67	0.48677	0.47951	0.45476	0.45950	0.43309	**2.31363**
12	Citizenship	3.30	3.59	3.31	3.04	3.08	0.46631	0.48056	0.46684	0.45162	0.45401	**2.31933**
13	Height	2.04	2.93	2.65	2.52	2.31	0.42758	0.49120	0.47510	0.46648	0.45090	**2.31127**
14	Signature	3.67	3.22	3.15	3.00	3.15	0.48539	0.46340	0.45950	0.45066	0.45950	**2.31845**
PSEUDO METRICS												
1	Personal Identity Number (PIN)	4.07	3.96	3.30	3.11	3.37	0.48666	0.48230	0.45065	0.43964	0.45448	**2.31373**
2	Password	3.96	3.63	3.22	2.56	3.22	0.49333	0.47968	0.45906	0.41603	0.45906	**2.30717**
DEVICE METRICS (Mobile Phone)												
1	International Mobile	3.73	3.62	3.54	2.82	2.88	0.48408	0.47923	0.47552	0.43457	0.43854	**2.31194**

	Equipment Identifier											
2	International Mobile Subscriber Identifier	4.07	3.67	3.52	3.04	3.22	0.48921	0.47240	0.46519	0.43845	0.44916	**2.31441**
3	SIM Serial Number	3.88	3.42	3.35	3.15	2.88	0.48942	0.46872	0.46512	0.45412	0.43752	**2.31491**
4	Mobile Station ISDN Number	3.50	3.27	3.15	2.88	2.92	0.48251	0.47120	0.46472	0.44857	0.45111	**2.31812**
5	PUK/PIN Number	3.58	3.27	3.12	2.62	2.77	0.48972	0.47513	0.46710	0.43522	0.44566	**2.31283**
DEVICE METRICS (Internet Terminal)												
1	IP Address	4.06	3.24	3.12	2.82	2.76	0.50205	0.46656	0.45990	0.44138	0.43734	**2.30724**
2	Media Access Control (MAC) Address	3.91	3.55	3.27	3.23	3.05	0.48758	0.47177	0.45734	0.45512	0.44459	**2.31640**
DEVICE METRICS (Card Based Tokens)												
1	Unique Card ID number	4.58	3.38	3.12	3.23	3.17	0.50629	0.45839	0.44374	0.45015	0.44669	**2.30526**
2	Format or File System used	3.22	3.04	3.19	2.85	2.65	0.47708	0.46729	0.47552	0.45583	0.44245	**2.31816**
3	Full Name of a User	2.84	2.92	2.75	3.08	2.91	0.46069	0.46559	0.45490	0.47476	0.46499	**2.32092**
4	Expiration date of card	2.42	2.91	2.70	2.61	2.51	0.44940	0.48153	0.46896	0.46304	0.45606	**2.31899**
5	Discretionary data	2.46	2.43	2.48	2.61	2.35	0.46396	0.46179	0.46538	0.47417	0.45579	**2.32108**

Table 6: Information Content of Physical, Pseudo and Device Metrics Attributes

The last columns of Table 6 and Table 7 have the computed values of the information content. As can be seen from Table 6, citizenship had the highest score while city of birth had the lowest score for the physical metrics group. The National Identity Number which was rated first using a questionnaire is now rated only sixth when using Shannon's information content.

The PIN number had a higher score as compared to the password in terms of the entropy yield for the pseudo metrics group meaning users had more faith and found it handy to use the PIN number as compared to a password. The Mobile Station ISDN Number (User's phone number) had the highest score while the International Mobile Equipment Identifier had the lowest score in Table 6. This means that the users had more confidence

in their phone numbers, which is increasingly becoming more useful to access online services. The MAC address had the highest score as compared to the IP address for the Internet terminal. This again reflects details in [7] where the MAC address is considered to be more unique, verifiable, consistent, persistent and trustworthy than the IP address.

Index	Attributes	Desirable Properties of Digital Identity					$P_i \log_2 (1/P_i)$					$\sum P_i \log_2 (1/P_i)$
		Distinctiveness	Universality	Performance	Permanence	Circumvention	W1	W2	W3	W4	W5	
1	Face Recognition	3.77	3.38	3.12	3.23	3.17	0.48502	0.46678	0.45249	0.45876	0.45538	**2.31843**
2	Facial Thermogram	3.50	3.00	3.18	2.91	3.05	0.48333	0.45694	0.46727	0.45141	0.45991	**2.31887**
3	Fingerprint	4.54	3.77	3.76	3.96	3.88	0.48632	0.45461	0.45413	0.46341	0.45979	**2.31825**
4	Hand Geometry	3.25	2.96	3.04	3.08	2.73	0.47740	0.46131	0.46601	0.46828	0.44662	**2.31962**
5	Iris scan	3.76	3.16	3.33	3.56	3.35	0.47991	0.44952	0.45903	0.47075	0.46010	**2.31931**
6	Retina Scan	3.83	3.46	3.52	3.58	3.27	0.47822	0.46074	0.46379	0.46675	0.45053	**2.32003**
7	Signature Recognition	3.42	3.08	2.80	2.81	3.00	0.48515	0.46771	0.45067	0.45132	0.46310	**2.31794**
8	Voice Recognition	3.19	3.00	2.88	2.62	2.67	0.48215	0.47194	0.46487	0.44781	0.45129	**2.31806**

Table 7: Information Content of Biometrics Attributes

Discretionary data had the highest score where as the *Unique Card ID* number had the lowest score for the card based credentials. The discretionary data area is where critical information such as account number for a bank card is usually stored. Having the highest score means this attributes can be used to store vital information.

As can be seen in Table 7, retina scan had the highest score while signature recognition had the lowest score. This reflects most of the scientific experiments carried out in [16] which show that retina scan is actually more distinctive, universal and permanent, with very high performance and can not easily be fooled as compared to other biometric features.

On average biometric group had the highest average score of 2.318814 followed by device metrics with 2.317237 and at the end of the group was the pseudo metrics group with the lowest average score of 2.310450. Biometric features are therefore considered to

be more trustworthy of the four groups. Of the three communication devices the card based credentials had the highest score of 2.316882 while the Internet terminal had the lowest average score of 2.311820 meaning users have more faith in the card based tokens to access online services than using the mobile phone or Internet terminal.

3.7 Summary

Chapter 3 analysed the credential tokens and identity attributes required to build a digital identity management system. The chapter began by coming up with the fourteen credential tokens which together with the biometric features were used as the sources of the identity attributes in section 3.2. Section 3.3 looked at how the attributes were selected and extracted from the fourteen credentials and biometric features. Then the extracted attributes were grouped into four groupings namely physical metrics, pseudo metrics, device metrics and biometrics. A detailed description of the multiple metrics, advantages and disadvantages were then given in section 3.4. In order to find out how much space is required to store a given set of attributes, section 3.5 looked at the representation of attributes. This chapter was concluded in section 3.6, which provided a process of computing the weights of attributes. This process was done in two stages. In the first stage questionnaires were used to determine the initial mean scores of the attributes using the desirable properties of the digital identities. Shannon's Information Theory was then applied to the mean scores of the desirable properties of digital identities so as to compute the information content for each attribute. The computed information content of the attributes was then taken as the required weights. The next chapter looks at the system implementation using the computed weights to implement the information fusion engine.

Chapter 4

4 SYSTEM DESIGN AND INFORMATION FUSION IMPLEMENTATION

4.1 Introduction

This chapter looks at the system design and the information fusion implementation using the attributes obtained in Chapter 3. The chapter therefore begins by looking at the system design of the digital identity management system using multi-modal authentication system. Multi-modal authentication system is however implemented by a technique of information fusion. Information fusion is used to combine all the attributes submitted by the user for optimum recognition. The second half of this chapter thus looks at the implementation of the information fusion engine using a multilayer artificial neural network. MatLab software is used to implement the multilayer artificial neural network to obtain the input weights, the layer weights and the threshold values required to implement an information fusion engine.

4.2 System Design

Figure 5 shows the system design of the digital identity management system using multi-modal authentication. Users in this design may access the services using any of the three devices namely a mobile phone, Internet terminal (PC) or a smart card (e.g. credit card). The type of device used to access the services is first identified by the system. In the next stage the attributes forwarded by the device are verified against the copies stored in the database system. Each attribute that has been verified successfully is then assigned its credential strength [3] as computed in Table 8 or a zero if it does not match a copy in the database system. If a new device is presented it is recorded in anticipation of the user registering it as an identity device. With their assigned values, the attributes are forwarded to the information fusion engine. The information fusion engine then computes a single value to be used in a multi-modal authentication system. The threshold value is then set depending on the level of security for a given application. The multi-modal

authentication system therefore gives three possible outcomes using the computed value of information fusion and the threshold value set for the system. If the computed value of the information fusion is less than the set threshold value, the user will be denied access to the system hence unsuccessful multi-modal authentication.

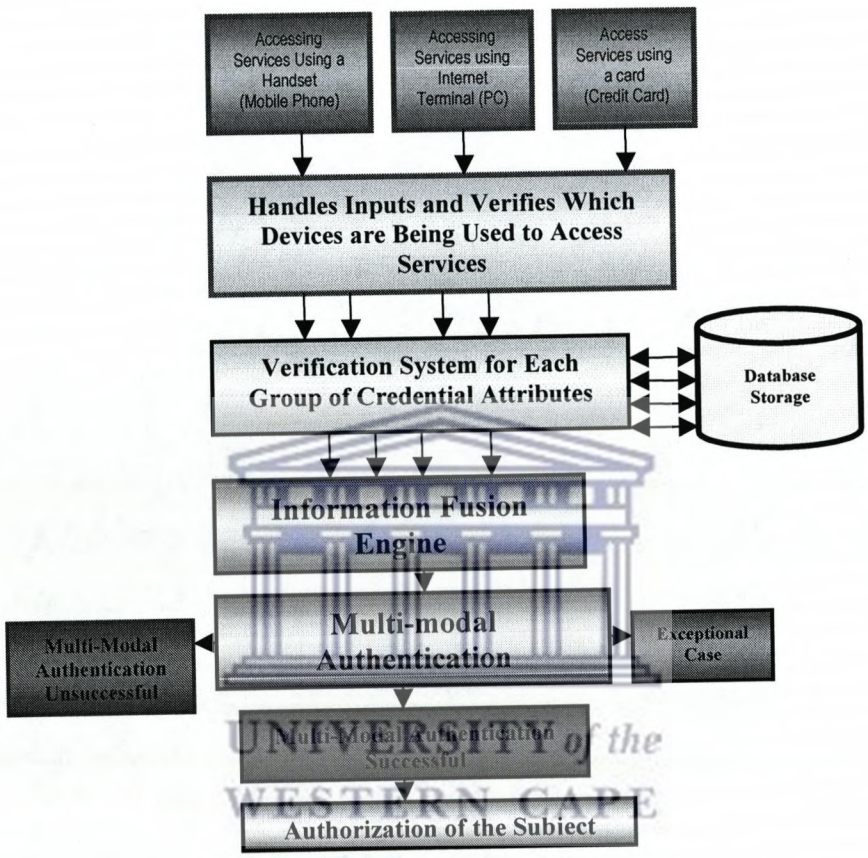

Figure 5: Multi-modal Authentication Model

If however the computed value of the information fusion is outside the required range of values then that would be an exceptional case and the user is denied access to the system. Finally if the computed value of the information fusion is greater than or equal to the set threshold value, the user is allowed access to the services thereby having successful multi-modal authentication. This is followed by the process of authorization. In this thesis, an Internet terminal is used to develop an experimental demonstration project.

4.3 Implementation of Information Fusion Engine

Information fusion is the technique of combining different kinds of data from different sources in order to obtain meaningful joint information [46] [47].

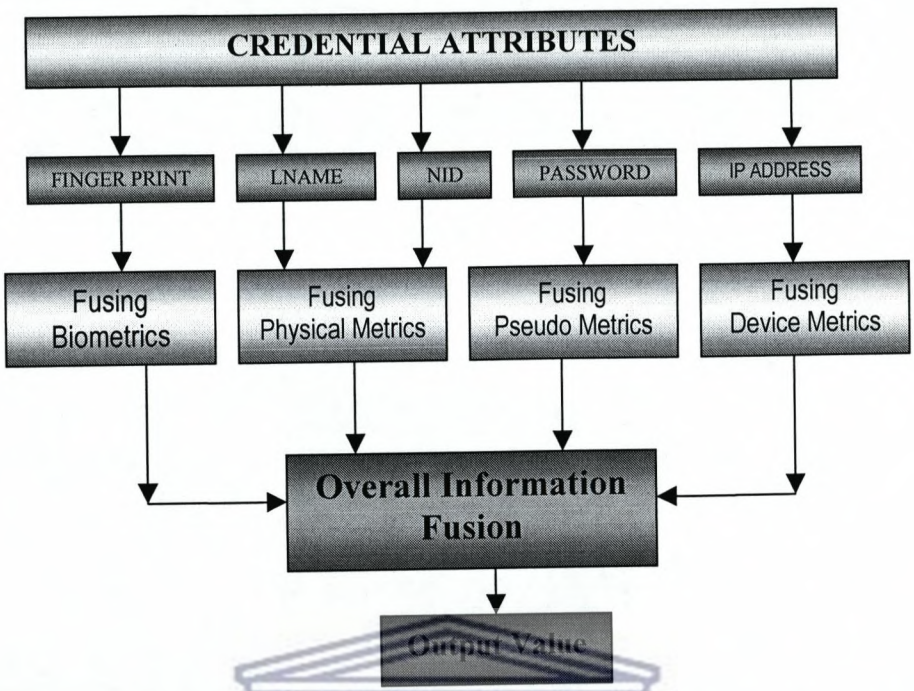

Figure 6: Information Fusion Engine in Multi-modal Authentication System

Information fusion technology has been applied most prominently to military applications such as battlefield surveillance and tactical situation assessment [46]. It has also emerged in commercial applications such as robotics, manufacturing, medical diagnosis and remote sensing [47]. Artificial neural networks, fuzzy logic, Bayesian method, evolutional computation, hybrid intelligent systems or data mining technologies can be used to implement an information fusion engine [46]. The aim of the information fusion engine is to compose the combined strength of the submitted attributes.

The first step of the information fusion engine is to compute the combined weight of the attributes in a given group. The second and final stage then uses the outputs from the groupings and computes the overall weight of all the attributes submitted by combining the weights from the groupings. This gives a single value representing the combined strength of the submitted attributes as shown in Figure 6. This strength is representative of the identity of the person. The computed value will depend on the collection of attributes submitted by the user during multi-modal authentication. It is the computed

value that will be used to authenticate a user in a multi-modal digital identity management system as shown in Figure 5.

4.3.1 Information Fusion Implementation Using Artificial Neural Networks

In Section 2.3, the advantages and disadvantages of using each of the six artificial intelligent technologies was given. Of the six, an artificial neural network is better suited for the implementation of an information fusion engine in this thesis owing to its advantages. The key element of this concept is the novel structure of the information processing system. Neural networks have broad applicability to real world business problems. Since they are the best at identifying patterns or trends in data, they are well suited for prediction or forecasting needs. These include sales forecasting, industrial process control, customer research, data validation, risk management and target marketing. Most of these applications involve combining data or information from different sources to obtain meaningful joint information [49].

The choice of the number and type of the attributes used as the input vectors help to determine the architecture design of the multilayer artificial neural network. However the choice of the attributes depends on the level of security required by the system. For example, the security level required for an online student record system is higher as compared to an email system. This implementation looks at an online application that requires a higher level of security.

The neural network in this implementation uses five attributes as the input vector from the four classes. The National Identity (NID) number and the full name are used as the physical metrics. The password is used as the pseudo metrics while the IP address as the device metric and the fingerprint as the biometric feature. Using too many attributes slows down the system and discourages users who have to submit a lot of credentials to access the services while too few attributes may compromise the required security of the system. The IP address of the terminal used to access the service is captured automatically by the system. Since a high level of security is required five attributes are considered adequate. These attributes form the input vector for the network as shown in Table 8. They are used in a multi-modal authentication system using artificial neural networks to implement information fusion.

Index	Attribute	Information Content ($\sum Pi \log_2 (1/Pi)$)	Input Variable
Physical Metrics			
1	National Identity Number (NID)	2.31786	x1
2	Last Name	2.31839	x2
Pseudo Metrics			
3	Password	2.30717	x3
Device Metrics			
4	Internet Protocol (IP) Address	2.30724	x4
Biometrics			
5	Finger Print	2.31825	x5

Table 8: Weights of the Input Vectors

These weights are then assigned to the input variables x1 to x5 as shown in the fourth column of Table 8.

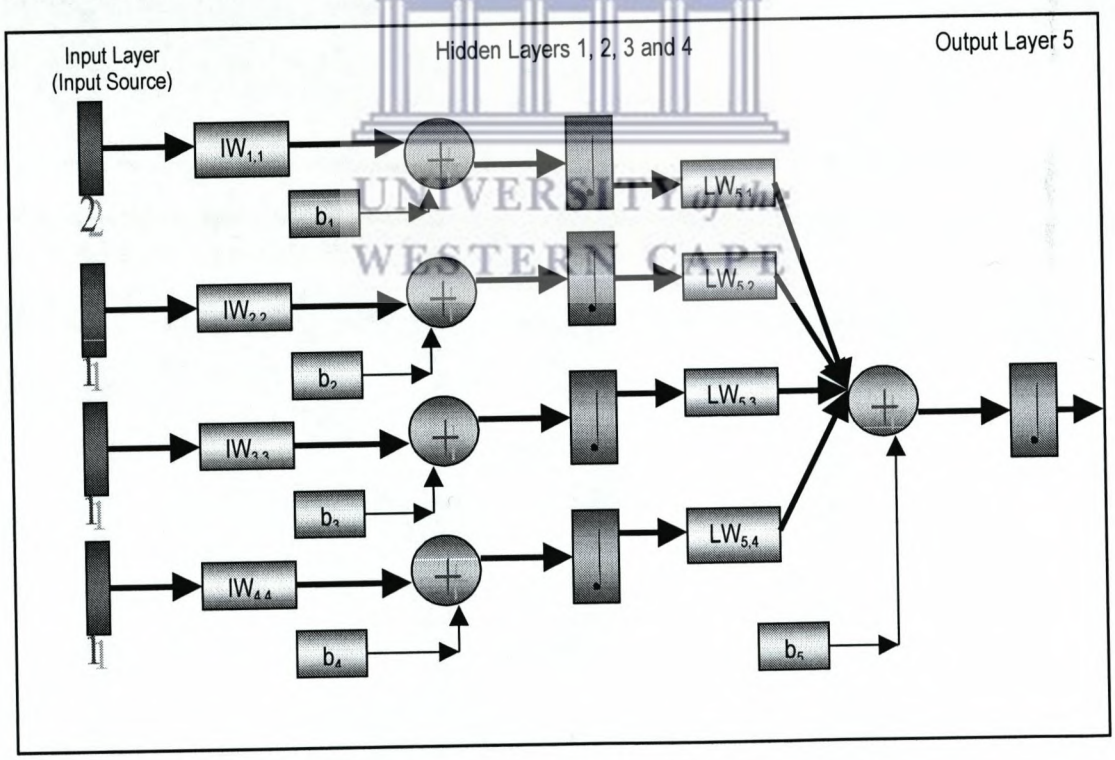

Figure 7: MatLab Design of the Artificial Neural Network

The information fusion engine will use a set of five input vectors from four input sources. The input sources are selected from each group namely physical metrics, pseudo metrics, device metrics and biometrics. The first input source will have two input vectors (last name and NID) represented by 2 as shown in Figure 7 while the rest of the input sources each has a single input vector (password, IP address and fingerprint). Figure 7 shows the designed artificial neural network with four input sources, four hidden layers, one output layer, input weights (IW), layer weights (LW) and the threshold values or biases (bi). A more simplified diagram representing the design in Figure 7 is shown in Figure 8.

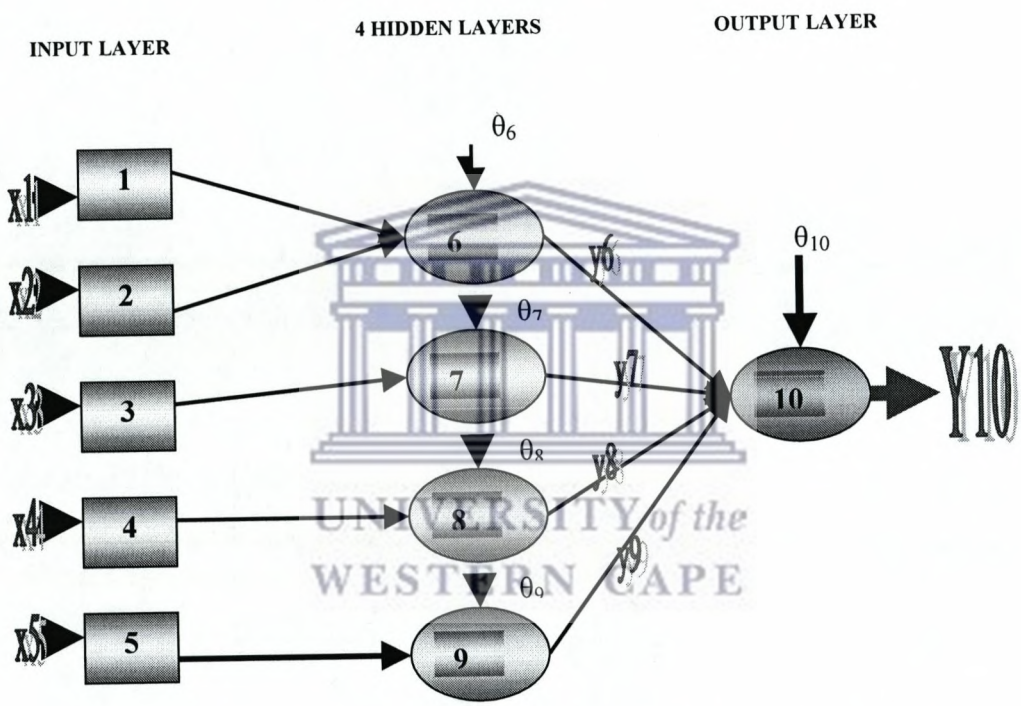

Figure 8: A Multilayer Artificial Neural Network Used in Information Fusion

It is a multilayer artificial neural network with ten neurons and six layers. Neuron 1 to 5 are the input neurons while neuron 6, 7, 8 and 9 represent the physical metrics, pseudo metrics, device metrics and biometrics respectively. These four neurons also represent the four hidden layers of the network where each neuron belongs to a single layer as depicted both in Figure 7 and Figure 8. Neuron 10 is the output neuron representing the output layer, which receives the inputs from the four hidden layers to compute the network output.

The attributes' weights from the metrics as shown in Table 8 are used as input vectors for the network. Input variables x1 and x2 form the vector for the physical metrics. These are then forwarded to the first hidden layer representing the physical metrics grouping. The input variable x3 forms the vector for pseudo metrics. This input is fed into the second hidden layer representing the pseudo metric grouping. The input variable x4 form the device metrics vector and is fed into the third hidden layer representing the device metrics grouping. Finally input variable x5 form the biometrics vector and is fed into the fourth hidden layer representing the biometrics groupings. These groupings will then compute their individual group weights, which are then fed into the output layer. The output layer then computes the output Y10 representing the overall combined weight of the submitted attributes. It is this value (Y10) obtained from the information fusion engine that is used in a multi-modal authentication system.

Many activation functions have been tested but only a few like the step, sign, sigmoid and linear functions have found practical application [48]. This thesis uses sigmoid function. Sigmoid function transforms the input, which can assume any value between plus and minus infinity into a reasonable value in the range between zero and one. Neurons with this function are used in the back-propagation network. The derivative of this function is easy to compute and also guarantees that the output of the neuron is bound between zero and one [48]. Back propagation algorithm is used for learning with initial weights w assigned at random between the values of 0 and 1. The activation function will be represented by the following equations [48];

$$X = \sum_{i=1}^{n} x_i w_i - \theta \tag{4.1}$$

Where X is the net weighted input, n is the number of inputs and θ is the threshold applied to the neuron. x_i is the value of the input variable and w_i is the input or layer weight [48] of the neuron.

$$Y^{sigmoid} = \frac{1}{1+e^{-X}} \tag{4.2}$$

From equation (4.1), the output value X is then used as the input to equation (4.2). Equation (4.2) computes the final output Y of a given neuron. These two equations will be used to compute the output of the neurons in the hidden and output layers.

Using neuron 6 as an example, this section demonstrates how to compute its output. Neuron 6 computes the strength of the physical metrics submitted by the user. It has two inputs vectors, two initial input weights and one threshold value as shown in Table 9.

#	x1	x2	*w61*	*w62*	θ_6
1	2.31786	2.31839	1.00000	1.00000	1.00000

Table 9: Neural 6's Input Values, Input Weights and Threshold Value

Using the table, equation (4.1) and (4.2), the output of y6 will be as follows;

$$y_6 = 1/(1+e^{-3.63625}) = 0.97432557$$

This gives y6 = 0.97432 to five decimal places as the final output. By adjusting the threshold value θ_6, the initial input weights *w61* and *w62*; it is possible to obtain the targeted value set for y6. Using the above method, it is possible to compute the targeted values of y7, y8, y9 and y10 [3]. In order to obtain the five initial input weights, four layer weights and five threshold values for the network (Figure 7), MatLab software was used as illustrated in the next section.

4.3.2 Using Matlab Software to Program and Train the Network

This section looks at how MatLab was used to program and train the designed multilayer artificial neural networks shown in Figure 7 in order to obtain the threshold values, input weights and layer weights. Below is the description of the code used to program the network in MatLab. The full code of this network is included in Appendix I.

To begin with an empty network called net was defined. This network held all the other components of the designed multilayer artificial neural network in Figure 7. It had four input sources as shown by this code;

```
net = network     %%Defines the Network
net.numInputs=4; %%Defines the Number of Input Sources
```

The network had five layers and these were four hidden layers and one output layer. In MatLab code the input layer is not considered as a layer. In the code below, layer 1 therefore stands for the first hidden layer represented by neuron 6 in Figure 8 . All the five layers had a bias connection as shown by this code:

```
net.numLayers=5; %%Defines the Number of Layers

%%Bias Connections to the Layers
net.biasConnect(1) = 1;
net.biasConnect(2) = 1;
net.biasConnect(3) = 1;
net.biasConnect(4) = 1;
net.biasConnect(5) = 1;
```

The four input sources specified above were each connected to the four hidden layers. The first input source was connected to the first hidden layer, the second input source to the second hidden layer, the third input source to the third hidden layer and the fourth input source to the fourth hidden layer as shown in Figure 8. This was represented by the following four lines of code.

```
%%Connecting the Input Sources to the Layers
net.inputConnect(1,1) = 1;
net.inputConnect(2,2) = 1;
net.inputConnect(3,3) = 1;
net.inputConnect(4,4) = 1;
```


The next step was to connect the hidden layers to the output layer. The four hidden layers (physical metrics, pseudo metrics, biometrics and device metrics) representing the four groupings were each connected to the output layer as shown in this code:

```
%%Connecting the Layers to the Output Layer
net.layerConnect(5,1) = 1;
net.layerConnect(5,2) = 1;
net.layerConnect(5,3) = 1;
net.layerConnect(5,4) = 1;
```

Layer 6 represented by a number 5 in the code below was specified as the target and output layer. This layer is responsible for computing the required output of the network.

```
%%Specifying the Target and Output Layers
net.outputConnect(5) = 1;
```

net.targetConnect(5) = 1;

In this network, the first input source connecting to the first layer had two input vector elements while each of the remaining input sources each had a single element as the input vector. The computed information content of the attributes as shown in Table 8 is in the range between zero and ten. Therefore each of the input vector elements was set to a value in between zero and ten as shown below.

%%Define the Input Vector for Every Input Source
net.inputs{1}.range = [0 10; 0 10];
net.inputs{2}.range = [0 10];
net.inputs{3}.range = [0 10];
net.inputs{4}.range = [0 10];

The next stage was to specify the number of neurons, the transfer function and the initialisation or activation functions for each layer. The first hidden layer has a single neuron and uses sigmoid transfer function represented by '*transig*'. The initialisation function used was the Nguyen-Widrow function as shown below:

%%Specify Number of Neurons, Transfer Functions and Initialisation Function for Layer One
net.layers{1}.size = 1;
net.layers{1}.transferFcn = 'logsig';
net.layers{1}.initFcn = 'initnw';

The second hidden layer representing the grouping pseudo metrics has one input vector element and was therefore represented by a single neuron. Sigmoid transfer function was also used as the transfer function and Nguyen-Widrow function was used to initialise the network as shown by the MatLab code below.

%%Specify Number of Neurons, Transfer Functions and Initialisation Function for Layer Two
net.layers{2}.size = 1;
net.layers{2}.transferFcn = 'logsig';
net.layers{2}.initFcn = 'initnw';

The third hidden layer representing the grouping device metrics with only one input vector element was represented by a single neuron. Sigmoid transfer function and Nguyen-Widrow function were used as transfer function and to initialise the network respectively.

%%Specify Number of Neurons, Transfer Functions and Initialisation Function for Layer Three
net.layers{3}.size = 1;

net.layers{3}.transferFcn = 'logsig';
net.layers{3}.initFcn = 'initnw';

The fourth and last hidden layer representing the grouping biometrics also has only a single neuron. Sigmoid transfer function was also used as the transfer function where as Nguyen-Widrow function was used to initialise the network as shown below.

%%Specify Number of Neurons, Transfer Functions and Initialisation Function for Layer Four

net.layers{4}.size = 1;
net.layers{4}.transferFcn = 'logsig';
net.layers{4}.initFcn = 'initnw';

The last layer is the output layer. The above four hidden layers represented by neurons 6, 7, 8 and 9 in Figure 8 connect to this layer and their outputs become its inputs. This layer was represented by a single neuron that computes the final output Y10. Sigmoid function was used as the transfer function while Nguyen-Widrow function was used to initialise the network.

%%Specify Number of Neurons, Transfer Functions and Initialisation Function for Layer Five

net.layers{5}.size = 1;
net.layers{5}.transferFcn = 'logsig';
net.layers{5}.initFcn = 'initnw';

To set up the bias between the four input sources and the four hidden layers to which the input sources are connected the following four lines of code were used.

%%Specify the Bias between Input Source and Four Hidden Layers

net.inputWeights{1,1}.delays = 1;
net.inputWeights{2,2}.delays = 1;
net.inputWeights{3,3}.delays = 1;
net.inputWeights{4,4}.delays = 1;

The next four lines of code below show the bias connection between the four hidden layers and the output layer. The first line of code for example shows that there is a connection between the first hidden layer and the output layer. These two layers are represented by neuron 6 and neuron 10 in Figure 8.

%%Specify the Bias between Four Hidden Layers and Output Layer

net.layerWeights{5,1}.delays = 1;
net.layerWeights{5,2}.delays = 1;
net.layerWeights{5,3}.delays = 1;
net.layerWeights{5,4}.delays = 1;

The final stage in building this network was to set the universal functions. Nguyen-Widrow initialisation function was set as the global initialisation function in order to initialise the network according to the layer initialisation function set earlier on. The training function was set to Levenberg-Marquardt backpropagation (*trainlm)* as the training algorithm. Finally the performance function was set to mean squared error (*mse)* function as shown below.

```
%%Defining the Globe Functions and Variables
net.initFcn = 'initlay';
net.trainFcn = 'trainlm';
net.performFcn = 'mse';
```

With the above network specification of the multilayer artificial neural network program, the network was then trained to generate the required threshold values, input weights and layer weights. Before training this network, there was a need to obtain the training and target data which were however dependant on the input vectors for each input source.

The total number of input vectors for this network was five elements. Each input vector element had only two possible values assigned to it which were either a zero when a wrong copy of an attribute was submitted such that it did not match a copy stored in the database or the computed weight shown in Table 8 when the correct attribute was submitted such that it matched a copy in the database. This means there were 2^5 or 32 different combinations of the input vectors. The letter P was used to represent the training data as follows:


```
P = {
[2.31786; 2.31839] [0.00000; 2.31839] [2.31786; 0.00000] [2.31786; 2.31839]
[2.31786; 2.31839] [2.31786; 2.31839] [0.00000; 0.00000] [2.31786; 0.00000]
[2.31786; 2.31839] [2.31786; 2.31839] [2.31786; 0.00000] [2.31786; 2.31839]
[0.00000; 2.31839] [0.00000; 2.31839] [2.31786; 0.00000] [0.00000; 2.31839]
[2.31786; 2.31839] [0.00000; 2.31839] [0.00000; 0.00000] [0.00000; 0.00000]
[0.00000; 2.31839] [0.00000; 0.00000] [2.31786; 0.00000] [2.31786; 0.00000]
[0.00000; 2.31839] [2.31786; 0.00000] [2.31786; 0.00000] [0.00000; 2.31839]
[0.00000; 0.00000] [0.00000; 0.00000] [0.00000; 0.00000] [0.00000; 0.00000];

[2.30717] [2.30717] [2.30717] [0.00000] [2.30717] [2.30717] [2.30717] [0.00000]
[0.00000] [2.30717] [2.30717] [0.00000] [0.00000] [2.30717] [2.30717] [2.30717]
[0.00000] [2.30717] [0.00000] [2.30717] [0.00000] [2.30717] [2.30717] [0.00000]
[0.00000] [0.00000] [0.00000] [0.00000] [2.30717] [0.00000] [0.00000] [0.00000];

[2.30724] [2.30724] [2.30724] [2.30724] [0.00000] [2.30724] [2.30724] [2.30724]
```

[0.00000] [0.00000] [0.00000] [2.30724] [2.30724] [2.30724] [2.30724] [0.00000]
[0.00000] [0.00000] [2.30724] [2.30724] [2.30724] [0.00000] [0.00000] [0.00000]
[0.00000] [2.30724] [0.00000] [0.00000] [0.00000] [2.30724] [0.00000] [0.00000];

[2.31825] [2.31825] [2.31825] [2.31825] [2.31825] [0.00000] [2.31825] [2.31825]
[2.31825] [0.00000] [2.31825] [0.00000] [2.31825] [0.00000] [0.00000] [2.31825]
[0.00000] [0.00000] [2.31825] [0.00000] [0.00000] [2.31825] [0.00000] [2.31825]
[2.31825] [0.00000] [0.00000] [0.00000] [0.00000] [0.00000] [2.31825] [0.00000]
}

With the above training data and using sigmoid transfer function; the network computes 32 different output values ranging between zero and one. The letter T was used to represent the targeted data as follows:

T = {
0.99999 0.80000 0.80000 0.80000 0.80000 0.80000 0.60000 0.60000 0.60000 0.60000
0.60000 0.60000 0.60000 0.60000 0.60000 0.60000 0.40000 0.40000 0.40000 0.40000
0.40000 0.40000 0.40000 0.40000 0.40000 0.40000 0.20000 0.20000 0.20000 0.20000
0.20000 0.0000
}

The network was therefore trained with training and targeted data. To begin with the network was initialised with the first line of code shown below and then the performance goal was set to 1e-10 by using the second line of code. The training and targeted data were then fed to the network as shown by the third line of code below.

%%Training the Network, where P is the Training Data and T is the Targeted Data
net = init(net)
net.trainParam.goal = 1e-10
net = train(net,P,T);

4.3.3 Results of Training and Simulating the Network

The performance goal was continually changed during training until the desired results were obtained. Figure 9 is an example of a graph and the summary generated with different performance gaols.

TRAINLM, Epoch 0/100, MSE 0.0288781/0.001, Gradient 2.97205e-010/1e-010
TRAINLM, Epoch 1/100, MSE 0.0288781/0.001, Gradient 2.97205e-010/1e-010
TRAINLM, Maximum MU reached, performance goal was not met.

Figure 9 shows the results generated when the performance goal was set to 0.001. The performance goal was not reached in this case.

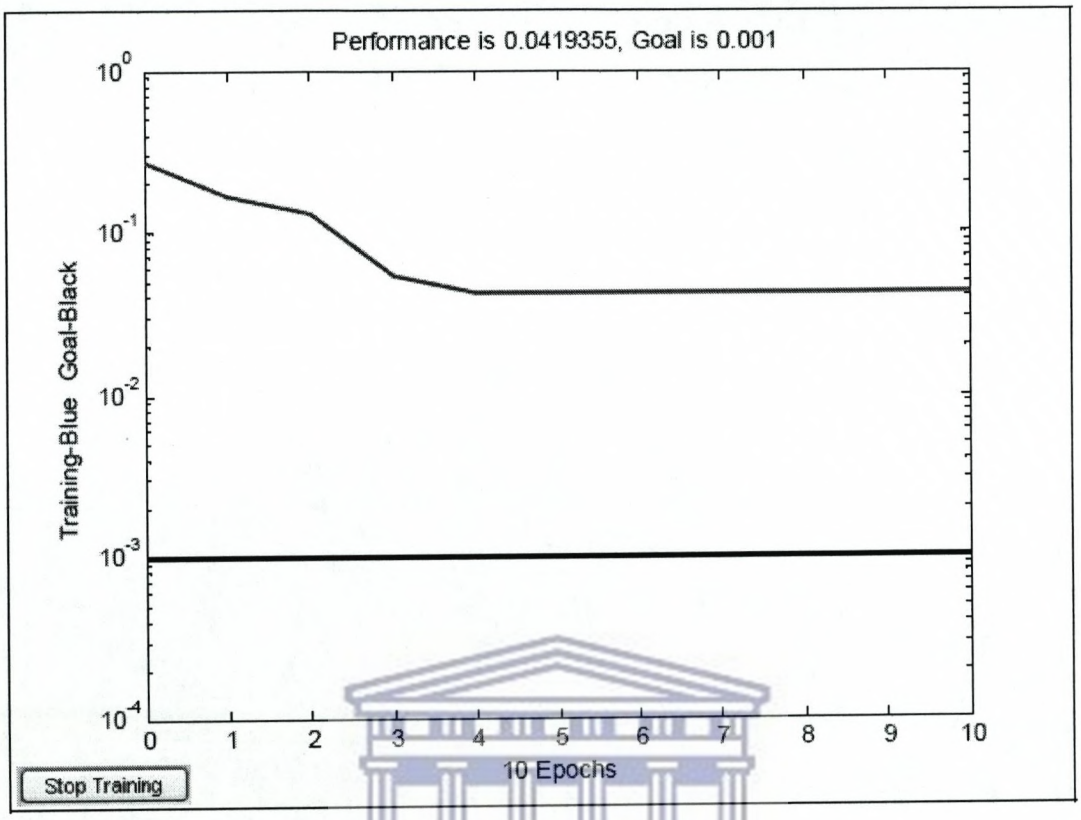

Figure 9: The Graph Showing the Output of Training the Network

Every time the performance goal was met, the network was simulated to see if the generated results were the required results. This process was performed several times until the required results were obtained.

TRAINLM, Epoch 0/100, MSE 0.137045/1e-008, Gradient 0.219577/1e-010
TRAINLM, Epoch 8/100, MSE 2.86996e-009/1e-008, Gradient 1.37465e-008/1e-010
TRAINLM, Performance goal met.

Table 10 shows the values of the input weights, layer weights and the threshold values of the network generated in relation to Figure 7 and Figure 8 . The second column shows five input weights for five input neurons. The third column shows the layer weights for four hidden layers of which each is represented by a single neuron which stands for one of the four groupings of physical metrics, pseudo metrics, device metrics or biometrics. The four threshold values belong to four hidden layers with the last threshold value that belongs to the output neuron 10.

Index	Inputs Weights	Layer Weight	Threshold Values to the Hidden Layer Neurons	Threshold Value to the Output Layer Neuron
1	0.5490	7.0970	3.2971	2.5270
2	0.1104	5.0617	2.8000	
3	0.5600	4.4144	2.8000	
4	0.5600	5.7191	2.8000	
5	0.5600			

Table 10: Input Weights, Layer Weights and the Threshold Values

The weights and threshold values in Table 10 were then programmed into the multilayer neural network in order to perform information fusion. Table 11 gives a sample of the outputs from a set of six inputs out of the thirty-two possible inputs as indicated by the letter P in section 4.3.2.

No.	Input to the Network					Output of the Network
	x1	x2	x3	x4	x5	Y10
1	2.31786	2.31839	2.30717	2.30724	2.31825	0.7815606326653765
2	2.31786	2.31839	2.30717	2.30724	0.00000	0.6366732167598268
3	2.31786	2.31839	2.30717	0.00000	0.00000	0.5024924327884972
4	2.31786	2.31839	0.00000	0.00000	0.00000	0.3493705690641579
5	2.31786	0.00000	0.00000	0.00000	0.00000	0.3040545860240985
6	0.00000	0.00000	0.00000	0.00000	0.00000	0.1973734463611742

Table 11: Inputs and Respective Output of the Neural Network

If all submitted attributes in Table 8 matched the copies in the database and were assigned weights as shown in the first row of Table 11, the computed weight of the neural network (information fusion) will then be 0.7815606326653765 as shown in Table 11. If however none of the submitted credential attributes matched the copies in the database and all the variables were therefore assigned zeros then the network would compute 0.1973734463611742 (last row). The other 30 combinations of the attributes will be spread out in between this range. With this output range, it is possible to build a multi-modal authentication system that would authenticate a user by setting the threshold value of the developed system between 0.20000 and 0.78156 depending on the level of security of the system.

4.4 Summary

This chapter looked at the system design of a multi-modal digital identity management system and the implementation of the information fusion engine. Users in this system would use either a mobile phone, Internet terminal or a credential in form of a card (e.g. credit card) to access the online services. The designed system used a technique of information fusion to combine the attributes submitted by the user during multi-modal authentication for optimum recognition. Section 4.3 looked at the implementation of an information fusion engine using a multilayer artificial neural network. The sub section 4.3.1 discussed the design of the multilayer artificial neural network with 10 neurons, one input layer, four hidden layers and one output layer. Matlab software was used to program and train the network in Figure 7 in section 4.3.2. Finally section 4.3.3 discussed the results of training and simulating the network in order to obtain the input weights, layer weight and the threshold values. Table 10 shows the generated input weights, layer weights and threshold values of the network while Table 11 shows a sample of the output values (Y10) of the information fusion engine using five input vectors.

Using these generated values of the information fusion engine and the system design above, the next chapter will consider the system implementation of the digital identity management system. The experimental system is implemented using an Internet terminal as a device used to access the online services.

Chapter 5

5 SYSTEM IMPLEMENTATION

5.1 Introduction

This chapter looks at the system implementation of a digital identity management system. It uses the system design and the information fusion engine developed in Chapter 4. To begin with the UML diagram is modelled using the system design in Chapter 4 and from this UML diagram, a UML Static Structure diagram is then developed using Microsoft Visio 2000. The classes in the UML Static Structure diagram are then programmed using Java language with Netbeans IDE as the programming environment. The implemented system uses Tomcat Web Server [69] running on Microsoft Windows XP as the operating system.

5.2 System Modelling Using a UML Diagram

The UML diagram as shown in Figure 10 was modelled based on the system design in section 4.2. It shows the six major steps required in order to authenticate a user as described below.

The Actor named users represents the individual accessing the online services. The person uses either a mobile phone, Internet terminal or the credential in form of a card to access the online services. In this implementation, an Internet terminal is used as a non-human communication device to access the online service.

The User Input is responsible for collecting the user's credential attributes. The user submits the required attributes by filling in a form as shown in Figure 13. These attributes are then forwarded to the Database Verification system.

The Database Verification is responsible for verifying the submitted attributes against the copies stored in the database system. After each attribute is compared with a copy stored in the database, it is then assigned a weight as shown in Table 8 if it matches a copy in the database or a zero if it does not match a copy in the database system.

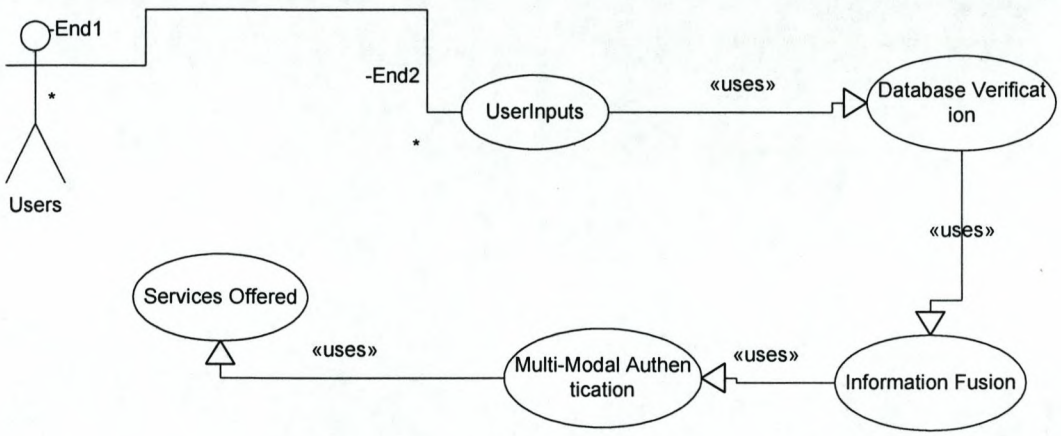

Figure 10: A UML Diagram

These weights are then forwarded to the Information Fusion. This then uses artificial neural networks to combine the attributes weights from the Database Verification in order to obtain a single value representing the combined weight of all the submitted attributes. The computed value is then forwarded to the Multi-Modal Authentication.

The Multi-Modal Authentication uses the computed value from the Information Fusion to authenticate a user. Depending on the value received and the threshold value set for the system, the user will either be allowed access to the services or denied access during multi-modal authentication.

The Services Offered is responsible for delivering the services offered by the system. A user gets access to this only after being successfully authenticated by the Multi-Modal Authentication.

5.3 System Modelling Using UML Static Structure Diagram

The UML Static Structure diagram in Figure 11 shows the classes derived from Figure 10. These classes are the building blocks of the multi-modal authentication system. The sub-sections below give a brief description of the classes comprising the digital identity management experimental system.

The first class is referred to as WelcomePage class. The information about the system and what is required in order to authenticate a user are specified and displayed by this class. The user is then directed to the UserInput class.

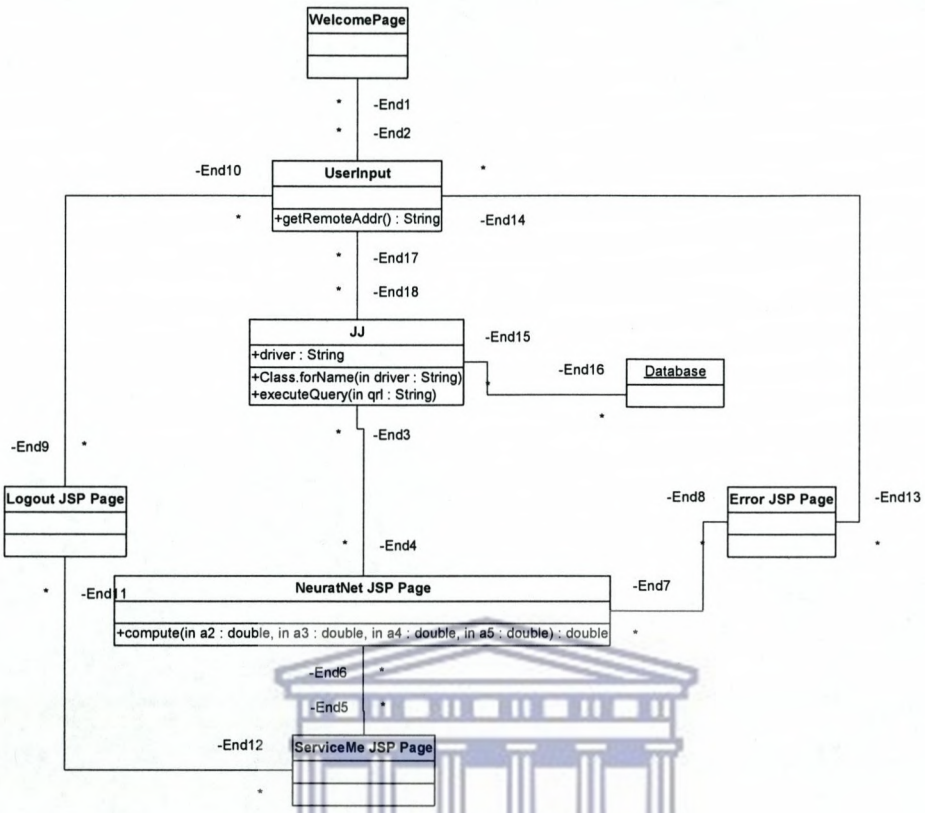

Figure 11: UML Static Structure Diagram

The UserInput class is responsible for collecting the user's credential attributes as shown in Figure 13. The collected attributes are then submitted to the JJ class.

The JJ class compares the attributes received from the UserInput class against the copies stored in the database system. After the verification, each attribute is assigned a weight as computed in Table 8 or a zero if the attribute did not match a copy in the database system. The weights are then used as the input vectors for the information fusion engine implemented by the NeuralNet class.

The NeuralNet class is responsible for performing information fusion using a multilayer artificial neural network. The artificial neural network uses the inputs vectors from the JJ class. The input weights, layer weights and the threshold values generated using MatLab software in Chapter 4 are used to construct an information fusion engine. The output of the information fusion engine, which is a single value, is then used to perform multi-modal authentication of a user. Depending on the value computed during information

fusion and the threshold value set for the system, a user is either authenticated or denied access to the services by the system.

If the computed value of the information fusion is greater than or equal to the threshold value set for the system, then the user is successfully authenticated and is directed to the ServiceMe class. This class as shown in Figure 15 handles all the services offered by the system. The user may then logout of the system to leave his or her privileged profile after accessing the required services. The user is then directed to the Logout class where he or she is assured that the logout process is successful and is then given an option either to close the page or login again.

However, if the computed value of the information fusion engine is less than the threshold value set for the system, then the user is denied access to the services and is directed to the Errorpage class as shown in Figure 14. Here the user is given an option of either to login again or close the page and leave the system.

5.4 Database Design and Implementation

This implementation uses four tables for the storage of the credential attributes. The user's National Identity Number forms the primary key in all the four tables and is used to link the four tables as shown in the entity relationship diagram (Figure 12).

Four tables are used in this implementation for the storage of the attributes with each group of the identity attributes as classified in Chapter 4 having a single table. Microsoft Access is used as a database system.

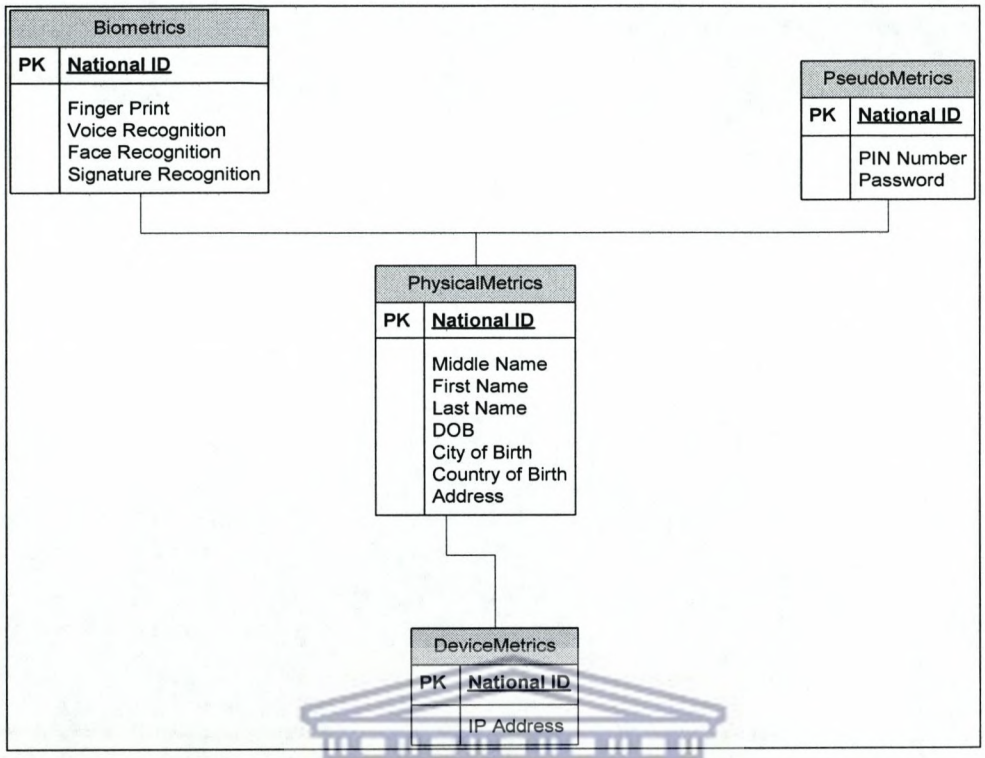

Figure 12: Entity Relationship Diagram Showing the Relationship of the Four Tables

5.5 System Programming

The implemented system is run on Apache Tomcat Web Server using Windows XP as the operating system. The Server uses 172.16.38.105 as the IP address and port 8084 to communicate with clients. It is linked to the author's homepage from the UWC Computer Science Department so that the system could be accessed from outside the University domain since a reserved IP address is used. The sections below give an illustration of the code used to program the classes in Figure 11 (see Appendix I for full code).

The WelcomePage is the front-end page. This is the first contact made by the user when interacting with the system. It contains some basic instructions on what is expected of the user. The user is then directed to the UserInput class by clicking on the login button.

The UserInput class is responsible for collecting the attributes of the user. In this implementation system the user is requested to submit the National Identity Number, the last name, the given password, the Internet Terminal's IP address and the fingerprint biometric features as shown in Figure 13.

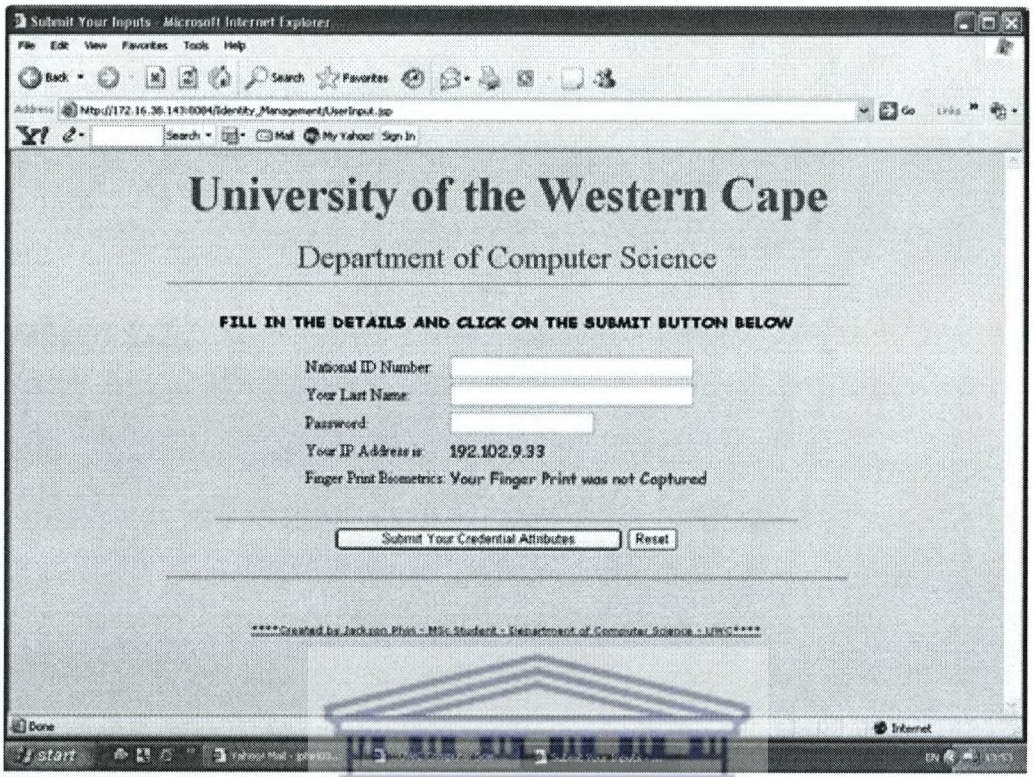

Figure 13: The Page Showing UserInput Class

It is these attributes that are used to authenticate a user in a multi-modal authentication system. The fingerprint is supposed to be captured first and separately. This is followed by the IP address captured automatically from the terminal being used to access the services. The Java method used to capture the IP address is called *getRemoteAddr()* and is called by using the request object from the HttpServletRequest class. Two buttons are included one for submitting the attributes on the form while the other for resetting the form as shown in Figure 13. The user attributes are submitted to the JJ class by clicking the submit button.

The JJ class operates in the background and performs two major functions. The first one is to compare the attributes submitted by the user against the copies stored in the database system. The second function is to assign the weights to these attributes using Table 8 if the attribute matched a copy in the database or a zero if it did not match a copy in the database system. Since the database is implemented using Microsoft Access database system, the drivers used to connect to the database are the *sun.jdbc.odbc.JdbcOdbcDriver* drivers. The URL used is *jdbc:odbc:DIMS* where DIMS is the name of the database

system used to store the user's attributes. To load the drivers above the *Class.forName(driver)* method is used with one parameter called *driver*. The following code shows how this is achieved using JavaServer Pages code.

```
<%  try {
        Class.forName(driver); %>        //Load the drivers
        <%} catch (ClassNotFoundException e) {%>    //Catch the exceptional errors
        <p>ClassNotFoundException: <%= e.getMessage() %>    //Print exceptional errors
        <%}%>
```

The connection to the database is made using the *DriverManager.getConnection(url, "", "")* method with three parameters. The url is represented by *jdbc:odbc:DIMS*, while the two double quotes represent the username and the password used to make the connection to the database system. The *createStatement()* method is used to create the statement used to query the database for the results. Using the *con* object of the Connection class and *stmt* object of the Statement class, the code below shows how the connection to the database is made and statement creation is achieved.

```
con = DriverManager.getConnection(url, "", "");    //Connect to the database
stmt= con.createStatement();    //Create the statement used to query the database
```


The *executeQuery(qrl1)* method is used to query the database system for the results. The qrl1 parameter is the SQL query statement used to compare the NID against the copy stored in the database system as shown below.

```
String qrl1 = "SELECT NID FROM PhysicalMetrics WHERE NID     = '"+ myNid+"'";
```

This statement is used to select the NID from the PhysialMetrics table by comparing if the NID in the table is the same as the submitted value found in the variable *myNid*. The results of this query are obtained using the combination of the *next()* and the *getString("NID")* methods. The retrieved results are then loaded into the variable Nid as shown below. If the right copy is submitted the copy in the database system is then loaded into the Nid variable but if it didn't match a copy in the database system then a null value is loaded into the Nid variable.

```
rs = stmt.executeQuery(qrl1);   //Query the database using an sql statement
while (rs.next()){              //Retrieve the results from the database
Nid = rs.getString("NID");      //Load the results into the Nid valuable
```

This process is done for all the submitted attributes of the user and then each attribute is assigned a weight. By comparing the submitted copy with the copy retrieved from the database system, each attribute is assigned a weight. If the submitted copy is the same as the copy stored in the database system, then the submitted attribute is assigned the weight as shown in Table 8, but if the submitted copy does not much the copy in the database then it is assigned a zero. The code below shows an example of how the Nid is assigned a weight;

```
if(Nid.equals(myNid)){ //Compare if the copy from the database is the same as that submitted
    x1 = 2.31786;      //If they are the same assign this weight to the input vector
} else{ x1 = 0.0000; } // If they are not the same then assign a zero to this input vector
```

These attributes with their assigned weights are then forwarded to the NeuralNet class. The following JSP code is used to complete this process and transfer control to the NeuralNet class.

```
<jsp:forward page="NeuralNet.jsp"> // Transfer control to the NeuralNet.jsp class
//Forward control together with the weights of the attributes from the JJ class
<jsp:param name="z1" value="<%= x1 %>" />
<jsp:param name="z2" value="<%= x2 %>" />
<jsp:param name="z3" value="<%= x3 %>" />
<jsp:param name="z4" value="<%= x4 %>" />
<jsp:param name="z5" value="<%= x5 %>" />
</jsp:forward>
```

The weights assigned to the attributes are then stored in the variables z1, z2, z3, z4 and z5. The NeuralNet class uses these attributes weights as the input vectors for the multilayer artificial neural network used to implement information fusion. This class performs two basic tasks. In the first task, it uses the attributes weights from the JJ class to perform information fusion. Then the computed value of the information fusion is used

in the second task to perform multi-modal authentication of the user. The following JSP code is used to access the attributes weights from the JJ class by the NeuralNet class.

```
<%
String v1 = request.getParameter("z1"); //Retrieve the value stored in z1 and load it into v1
String v2 = request.getParameter("z2"); //Retrieve the value stored in z2 and load it into v2
String v3 = request.getParameter("z3"); //Retrieve the value stored in z3 and load it into v3
String v4 = request.getParameter("z4"); //Retrieve the value stored in z4 and load it into v4
String v5 = request.getParameter("z5"); //Retrieve the value stored in z5 and load it into v5
%>
```

These attributes weights from the JJ class are then loaded into the variables v1, v2, v3, v4 and v5 in the NeuralNet class. These variables are first converted from the String format to the Double format and then loaded into the *compute* method which requires five parameters in Double format to implement information fusion as shown below.

```
compute(Double.parseDouble("v1"),Double.parseDouble("v2"),Double.parseDouble("v3
"),Double.parseDouble("v4"),Double.parseDouble("v5"));
```

This method computes and returns a single double value using sigmoid transfer function of the multilayer artificial neural network. In this method, the attributes weights from a1 to a5 are first assigned to the network input vectors x1 to x5 respectively. The input weights and layer weights as computed in Chapter 4 are represented by the variable *w*ij where i and j are integers representing the destination and source neurons respectively. The threshold values are represented by the variables q6 to q10. Sigmoid function represented by equations (4.1) and (4.2) in Chapter 4 is used in the neural network as the transfer function. The Java code used to implement the information fusion engine via the *compute* method is as shown below.


```
//The compute method used to implements the information fusion engine
public double compute(double a1,double a2,double a3,double a4,double a5){

//Assign the input variables from the JJ class a1,a2,a3,a4 and a5 to the input vectors x1,x2,x3,x4 and x5
    x1 = a1;
    x2 = a2;
    x3 = a3;
    x4 = a4;
    x5 = a5;

//Define the weights of the input layer neurons
```

```
w11 = 0.5490;
w12 = 0.1104;
w22 = 0.5600;
w33 = 0.5600;
w44 = 0.5600;

//Define the weights of the hidden layer neurons
w106 = 7.0970;
w107 = 5.0617;
w108 = 4.4144;
w109 = 5.7191;

//Define the threshold values of the hidden layers and output layer neurons
q6 = 3.2971;
q7 = 2.8000;
q8 = 2.8000;
q9 = 2.8000;
q10 = 2.5270;

X6 = (((x1*w11)+(x2*w12)) - q6); //Equation 4.1 used to compute overall input of Physical metrics
X7 = ( (x3*w22)-q7);             //Equation 4.1 used to compute overall input of Pseudo metrics
X8 = ( (x4*w33)-q8);             //Equation 4.1 used to compute overall input of Device metrics
X9 = ( (x5*w44)-q9);             //Equation 4.1 used to compute overall input of Biometrics

y6 = Math.exp(-X6);   //Equation 4.2 used to compute overall output of Physical metrics
Y6 = 1 / (1 + y6) ;

y7 = Math.exp(-X7);   //Equation 4.2 used to compute overall output of Pseudo metrics
Y7 = 1 / (1 + y7) ;

y8 = Math.exp(-X8);  //Equation 4.2 used to compute overall output of Device metrics
Y8 = 1 / (1 + y8) ;

y9 = Math.exp(-X9); //Equation 4.2 used to compute overall output of Biometrics
Y9 = 1 / (1 + y9) ;

//Equation 4.1 used to compute overall input from the four groupings
X10 = (((Y6*w106)+(Y7*w107)+(Y8*w108)+(Y9*w109))-q10);

y10  = Math.exp(-X10); //Equation 4.2 used to compute overall output from the four groupings
Y10 = 1 / (1 + y10);

return Y10;// Final Output of the information fusion engine
```

The compute method returns a double value Y10. It is this value that is used in multi-modal authentication of a user. Therefore to successfully authenticate the user, the computed value Y10 of the submitted attributes is supposed to be equal to or greater than

the threshold value set for the system. In this implementation the security level or threshold value for the system is set to 0.6000. If the computed value Y10 is greater than or equal to this value, then the user is authenticated and given access to the services. The user is directed to the ServiceMe class as shown in Figure 15 . However, if the computed value is less than the threshold value, access to the services is denied and the user is directed to the Errorpage class as shown in Figure 14 . This is achieved by the following JSP code.

```
<% if(Y10 < 0.6000){%>          //Check  if Y10 is less than 0.6000
//If Y10 is less than 0.6000 deny the user access and forward control to the Errorpage.jsp class
    <jsp:forward page="Errorpage.jsp" >
//Load the attributes weights to be printed in the Errorpage
    <jsp:param name="p1" value="<%= v1 %>" />
    <jsp:param name="p2" value="<%= v2 %>" />
    <jsp:param name="p3" value="<%= v3 %>" />
    <jsp:param name="p4" value="<%= v4 %>" />
    <jsp:param name="p5" value="<%= v5 %>" />
    <jsp:param name="p6" value="<%= Y10 %>" />
    </jsp:forward>
  <%}else if(Y10>=0.6000){%> //Otherwise Check if Y10 is equal to or greater than 0.6000
//If Y10 is equal to or greater than 0.6000 allow the user access and forward control to the ServiceMe.jsp
 <jsp:forward page="ServiceMe.jsp" />
 <%}%>
```

Figure 14 shows the output of the Errorpage class. Included in the page are the submitted attributes weights and the final computed value of the information fusion. In the example below only three attributes are successfully authenticated and these are the national ID, the user's last name and the captured IP address of the terminal used to access the services. They are assigned a value from Table 8 by the JJ class because they matched their respective copies in the database system. The password and finger print were either not submitted or wrong copies were submitted hence each is assigned a zero. The computed value by the information fusion engine is 0.48130679483223904. This is less than the threshold value of 0.6000 set as the security level of the system and thus the user

is denied access to the system. The user then has an option to login again, return to the home page or close the window and leave the system.

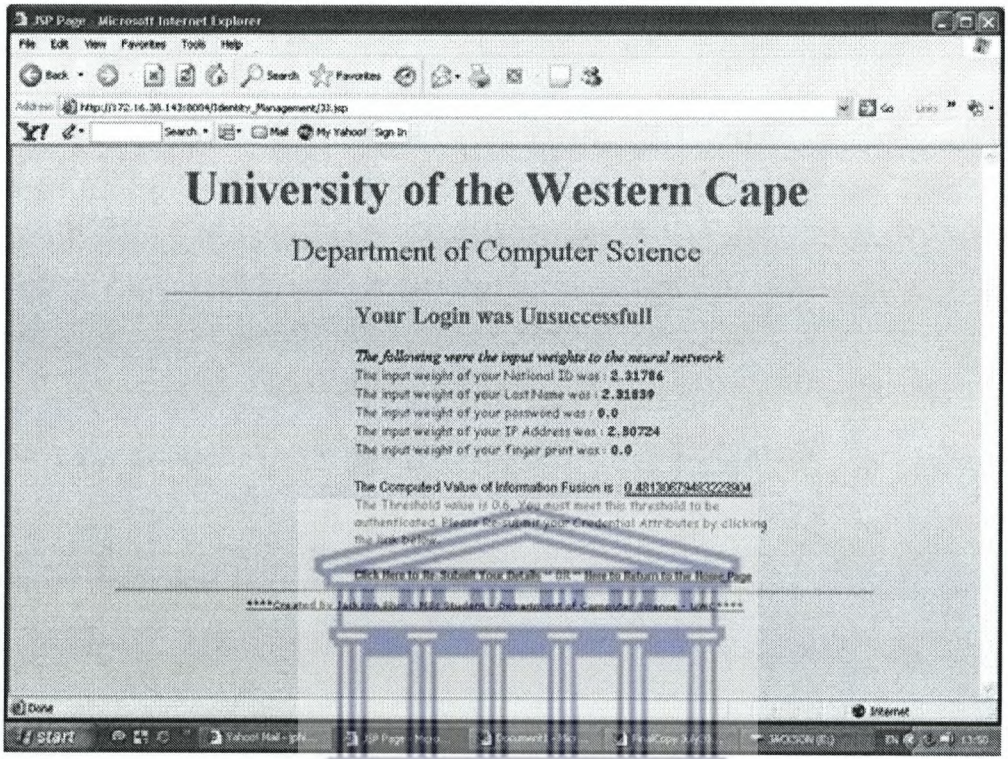

Figure 14: The Page Showing Errorpage Class

Figure 15 shows the ServiceMe class. With the security level of the system set to 0.6000, the user is required to submit at least four collect attributes that need to match the copies stored in the database system in order to be authenticated. With any four collect attributes the user is authenticated and directed to the ServiceMe class. The user is then availed with a number of services as shown in Figure 15. The user would then logout of the privileged profile by clicking the logout link.

The user is then directed to Logout class where he or she is assured that the logout process is successful. Here the user is given an option of either closing the page or login again. This page marks the end of the whole circle required to authenticate a user in this digital identity management system.

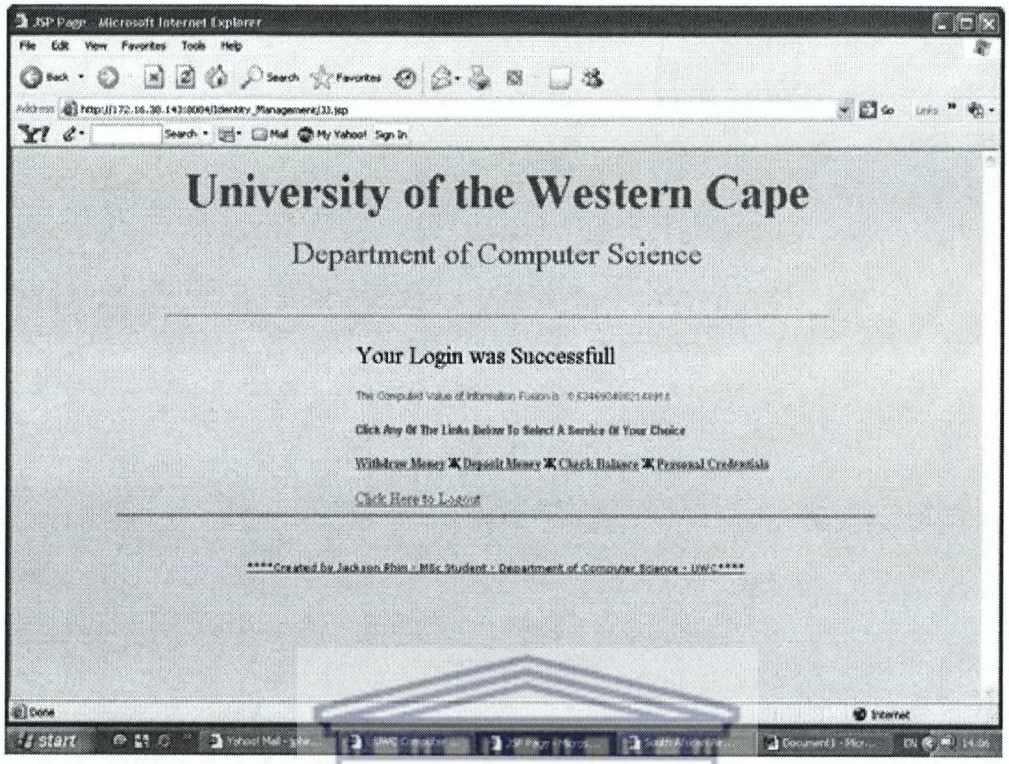

Figure 15: The Page Showing ServiceMe Class

5.6 Summary

Chapter 5 focused on the system implementation of the digital identity management system. The system design in Chapter 4 is first modelled in section 5.2 using UML diagram. In section 5.3, the UML diagram is used to obtain the UML Static Structure diagram. Microsoft Access database is used to implement the entity relationship diagram in Figure 12 of section 5.4. Section 5.5 then concluded this chapter by looking at the system programming of the classes in Figure 11 of section 5.3. The developed experimental system used the information fusion engine from Chapter 4 to implement multi-modal authentication of the user. Microsoft Window XP is used as the operating system and Apache Tomcat Web Server as the server. The system is able to authenticate a remote user with at least four collect credential attributes using a technique of information fusion to implement multi-modal authentication. The next chapter looks at the discussion and conclusion of the work done in this thesis.

Chapter 6

6 DISCUSSION AND CONCLUSION

6.1 Introduction

This thesis addressed the problem of identity fraud and theft seen on most online services today. This was achieved by developing a digital identity management system using multi-modal authentication system. This chapter therefore looks at the discussion and then the conclusion of the work done so far. Section 6.2 will discuss the work done and then section 6.3 will look at the conclusion. The chapter closes with the proposed future work in section 6.4.

6.2 Discussion

To help in developing the solution of the problem of identity theft and fraud seen on most online services today, section 1.3 highlighted four questions in the problem statement. To successfully determine the answers to these questions, five aims were set in section 1.4. The goal was to create a *Digital Identity Management System* using multi-modal authentication. It is this system that is finally considered as one of the solutions to the problem of identity theft and fraud. The discussion in this section will examine how the four questions in the problem statement were answered and how the five aims were met.

Chapter 2 focused on the background theory. In this chapter, the topics on *digital identities and management systems* and thereafter the technologies used to implement *information fusion* were explored in great detail. These formed the platform on which the implementation of this thesis was based. This was the first aim.

The first challenge in the system implementation was to come up with a set of credential tokens which would be used as the source of attributes used to build the identity system. These credentials were supposed to represent both humans and telecommunication devices used to access services offered both in the Real-space and Cyber-space in several countries. This was the focus of Chapter 3. Using the questionnaires fourteen credentials were chosen as the most commonly used tokens to access the services in most countries as shown in Figure 4. These tokens together with a set of the most commonly used

biometrics were then used as sources of the attributes for developing a digital identity management system. This was the second aim and it answered the first question in the problem statement.

Once the choice of credentials was made, their attributes were then extracted from these credentials. In practice this process needs to be automated. However in this thesis the fourteen credentials in addition to the set of biometrics features were used as the source of the attributes. Section 3.3 of Chapter 3 outlined the process used to extract the attributes from the above credentials. Table 2 gives a summary of the 36 attributes extracted from these credentials though the list is not conclusive. In section 3.3 of Chapter 3, all the 36 attributes were then classified into four groups (physical metrics, pseudo metrics, device metrics and biometrics). The description of the multiple metrics was given in section 3.4, where the definitions, advantages and disadvantages of using each group were given. To compute the weights of the attributes, two methods were used. The initial scores of the attributes were computed using a questionnaire and then Shannon's information theory was used to compute the final weight of each attribute. Section 3.6 of Chapter 3 has the detailed description how this was achieved. This answered the second question in the problem statement and was the third aim.

Using the groupings and the weights assigned to the attributes computed in Chapter 3, Chapter 4 used one of the six artificial intelligent technologies highlighted in the literature review (section 2.3) to implement an information fusion engine. A multilayer artificial neural network was used owing to its advantages over the other five technologies as highlighted in section 4.3.1. The neural network used five attributes as the input vector from the four classes as shown in Table 8. Using too many attributes would have slowed down the system and discouraged users who would be required to submit a lot of attributes to access the services while too few attributes would have compromised the required security of the system. Since a high level of security was required, five attributes were considered adequate. These attributes formed the input vector for the network. They were used to design a multilayer artificial neural network thereby implementing information fusion. The neural network design is shown in Figure 7 and 8 and section 4.3.1 has the detailed description. Section 4.3.2 gives an illustration of the network source code. The network was then trained and simulated using MatLab to

obtain the desired input weights, layer weights and the threshold values of the neurons required for the information fusion engine. Table 10 shows the input weights, layer weights and the threshold values of the neurons in Figure 8. Table 11 shows how the information fusion engine responded using the generated weights and thresholds. If all the submitted attributes in Table 8 matched the copies in the database and were assigned weights as shown in Table 11, the computed weight from the information fusion is 0.78156 to five decimal places. If however none of the submitted credential attributes matched the copies in the database such that all the variables were assigned zeros then the computed weight is 0.19737 (Table 11) to five decimal places. The other 30 combinations of the attributes are spread out in between this range. With this output range, it was possible to build a multi-modal authentication system to authenticate a user by setting the threshold value of the developed system between 0.20000 and 0.78156 depending on the required level of security for the system. This was the fourth aim and answered the third question in the problem statement.

The final challenge was to use the information fusion engine to develop an experimental system required to authenticate a user accessing online services remotely. Figure 5 shows a diagram of the proposed system design of the digital identity management system using multi-modal authentication. Section 4.2 of Chapter 4 has the detailed explanation on how the system works. The implemented experimental system used an Internet terminal as the device. A UML diagram was used to model the system design in Figure 5 and a UML Static Structure diagram as shown in Figure 11 was then created using Microsoft Visio 2000. The classes in the UML Static Structure diagram were then programmed using JavaServer Pages for the required access experiment. The project had seven classes as shown in Figure 11.

In the implemented system, four tables were used to store the user's attributes. The user's national identity number formed the primary key in all the four tables and was used to link the four tables as shown in the entity relationship diagram in Figure 12. Each group of the identity attributes had its own table and Microsoft Access was used to implement the database system on the Internet terminal running as the server. Section 3.5 of Chapter 3 focused on the representation of the digital identities to determine the amount of storage space required to store each of the user's credential attributes in the database system. This

was used to determine how much space was required to store each of the five attributes in Table 8.

In the implemented system, the user first comes in contact with the system through the WelcomePage. This page contains some basic instructions on the usage of the system and what is expected of the user. The user is then directed to the UserInput class by clicking on the login button. As shown in Figure 13, the user's credential attributes are submitted through this class. These are then forwarded to the JJ.jsp class, which compares these credential attributes against the copies stored in the database, and assigns a zero to an attribute if it didn't match a copy in the database or a value in Table 8 if it matched a copy in the database. These weights are then fed to the NeuralNet.jsp class that implement the information fusion using artificial neural networks. The neural network uses sigmoid function as the transfer function meaning the output is always between the value 0 and 1. The NeuralNet.jsp class then computes a single value (Y10 in Figure 8), which is the output of the information fusion engine. This value is then used to authenticate a user remotely. This is achieved by comparing Y10 with the value set as the threshold value or security level of the system, which is 0.6000 in this experimental system. If the computed value (Y10) is greater than or equal to this value, then the user is authenticated and is directed to the ServiceMe class as shown in Figure 15. However, if the computed value is less than the threshold value, access to the services is denied and the user is directed to the Errorpage class as shown in Figure 14. The implemented system is run on Apache Tomcat Web Server using Windows XP as the operating system. The server uses 172.16.38.105 as the IP address and port 8084 to communicate with clients accessing the services on this server. The system is able to authenticate a remote user with at least four collect credential attributes using a technique of information fusion to implement multi-modal authentication. This was the fifth aim and answered the fourth and last question in the problem statement.

Hence a multi-modal authentication system was successfully implemented using artificial neural network to implement information fusion. However a number of difficulties were encountered in the process of achieving the five aims in section 1.3. Not all those required to answer the questionnaires answered them correctly. Some respondents lost the questionnaires by the time they were being collected for the final compilation after being

answered. The questionnaires were re-distributed four times so as to get the required sample space of 100 respondents.

Identity fraud and identity theft have become a major concern for the public and private sectors especially as they relate to problems like terrorism, financial crime, drug trafficking, alien and weapons smuggling [2]. With the emerging security concerns in the areas of immigration, border crossings, airline passengers and driver's licenses [33], systems like the one developed would play a very big role to help reduce cases of identity theft and fraud seen on most online services. Today most online services like Internet banking, online stores, student records, e-governments depend so much on PIN numbers and passwords. This has exposed most users to identity theft and fraud since these credential attributes are easy to guess or forge [2]. Therefore to improve the security features of the current systems, this thesis proposed a digital identity management system using multi-modal authentication. The system proposed an addition of biometrics and attributes from the telecommunication devices used to access online services (e.g. mobile phone) when authenticating a remote user. These credential attributes are then combined using an artificial intelligent technology to derive the overall combined weight of all the submitted credentials. With this combination of credential attributes from different groupings having different strengths, it should be more difficult for hackers or thieves to guess or forge the whole set of the submitted credential attributes hence helping to reduce cases of identity fraud and theft seen on most online services today.

6.3 Conclusion

As seen from the previous section, the four questions in the problem statements and the five aims set in section 1.3 and section 1.4 respectively were all met. In the experiments the user was required to submit five attributes (Table 8). A multilayer artificial neural network was then used to combine these attributes for optimum recognition of the user during multi-modal authentication. Figure 15 shows how the user was successfully authenticated with at least four collected credential attributes when the security level for the system was set to 0.6000. Hence a multi-modal authentication system using an artificial neural network to implement information fusion was successfully implemented.

The developed system can be applied in a number of areas. Today most doors leading to secure areas usually use an electronic door where the user can either swipe a card or enter a secret code to gain access. Therefore if someone gains access to the card or PIN number used as the secret code then the secure area can easily be infiltrated. Combining for example your biometrics (e.g. face recognition or iris scan) with your PIN number and/or your card and then using information fusion improves the security level of the restricted area. The other application systems where the work in this thesis has direct application includes online banking system, student record systems, e-governments, border security programs and critical applications that include travel security systems with passport, ticket, and baggage verification systems.

6.4 Future Work

A multilayer artificial neural network was used to implement the information fusion engine in the experimental system. In the future, technologies such as evolutional computation (e.g. Genetic Algorithm) or Bayesian method should be used to develop the information fusion engine. Also five attributes which included fingerprinting and an Internet terminal were used to develop an experimental system. In the future, other application areas with different combinations of the attributes should be used. Good examples includes systems used to open the secure doors, system for accessing student records and for accessing the government services offered online. In all the above cases, a user would be required to submit more than one credential attribute, which would then be combined using a technique of information fusion for optimum recognition.

7 REFERENCES

[1] The National Electronic Commerce Coordinating Council (NECCC), "Identity Management", Presented at the NECCC Annual Conference, December 4-6, 2002, New York, NY

[2] Sittampalam S., "Digital Identity Modelling and Management", MEng Thesis, 2005, UTS, Australia

[3] Phiri J. and Agbinya J., "Modelling and Information Fusion in Digital Identity Management Systems", Proceedings of IEEE International Conference on Systems (ICONS 2006), Mauritius, 22nd - 29th April 2006, pp. 181

[4] Chong F., "Identity and Access Management", Microsoft Architect Journey, http://msdn.microsoft.com/library/en-us/dnmaj/html, July 2004

[5] Editor TB1, "Electronic Identity White Paper V 0.5", E-Europe Electronic Identity, November 2002

[6] Brands S. and Legare F., "Digital Identity Management based on Digital Credentials", Credentica Inc., http://ls6-www.informatik.uni-dortmund.de/issi/cred_ws/papers/brands.pdf, May 2005

[7] Faltstrom P. and Huston G., "A Survey of Internet Identities", *Internet Engineering Task Force - Work in Progress - draft-iab-identities-01.txt*, www.ietf.org, 28 April 2004

[8] Enterprise Identity Management, "Strategy White Paper", Microsoft Windows 2000 Server,

[9] Norlin E. and Durand A., "Federated Identity Management", PingID Network, Inc. White paper, 2002

[10] Liberty Alliance Project, "Introduction to the Liberty Alliance Identity Architecture", http://www.projectliberty.org/, 2003

[11] Office of the Federal Privacy Commissioner, "We Must Get Identification Management Right to Avoid Losing Our Privacy", http://www.privacy.gov.au/news/media/04_04.html, 31 March 2004

[12] Performance Technologies, Inc, "Introduction to GSM", Performance Technologies, Inc, http://www.pt.com/products/gsmintro.html, 2005

[13] Lerner M., “Connecting to the Internet”, Michael Lerner Productions http://www.learnthenet.com/english/html/04connec.htm, 2005

[14] Scourias J., “Overview of GSM”, University of Waterloo, http://www.shoshin.uwaterloo.ca/publications/pdfs/TR-96-01.pdf, 1996

[15] Ericsson White Paper, “Mobile Multimedia, The Next Step in Richer Communication”, Ericsson White Paper, http://www.ericsson.com/products/white_papers_pdf/mobile_multimedia.pdf, 2004

[16] Jain A. K., Ross A., and Prabhakar S., "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.14, No.1, January 2004, pp.4 - 20

[17] Prabhakar S., Pankanti S., and Jain A. K., "Biometric Recognition: Security and Privacy Concerns", *IEEE Security and Privacy Magazine*, Vol.1, No.2, March/April 2003, pp.33-42

[18] BiometriTech, "Facial-Recognition Solutions Roundup", http://www.biometritech.com/features/roundup051502.htm, 15 May 2002

[19] O'Gorman L., "Comparing passwords, tokens, and biometrics for user authentication", *Proceedings of the IEEE*, Vol.91, No.12, Dec. 2003, pp.2019 - 2020

[20] International Biometric Group, "Iris Recognition: How it Works", *Technology Report*, www.biometricgroup.com, October 2004

[21] Tracy K., Koerper V., “Biometrics: A Brief Introduction”, CSC 490 - Security Seminar, http://csc.noctrl.edu/f/kwt/590/projects/VICKIK~1.HTM, March 10, 1998

[22] International Biometric Group, "Biometric vs. Non-Biometric Fingerprinting", *Technology Report*, www.biometricgroup.com, October 2004

[23] Dingley D., "Signature Verification Technology", *CyberSIGN*, http://www.cybersign.com/news_press2.htm, 15 November 1999

[24] Wikipedia, “Information theory”, Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Information_theory , 2006

[25] The Open Group, "Single Sign-On", The Open Group, http://www.opengroup.org/security/sso/, 2005

[26] Rezgui A., Bouguettaya A., and Eltoweissy M. Y., "Privacy on the Web: Facts, Challenges, and Solutions", IEEE Security & Privacy, Vol.1, No.6, Nov-Dec 2003, pp.40 – 49

[27] Crompton M., "Proof of ID Required? Getting Identity Management Right," Federal Privacy Commissioner, http://www.privacy.gov.au/news/speeches/sp1_04p.pdf, 30 March 2004.

[28] Cranor L., Langheinrich M., and Marchiori M., "A P3P Preference Exchange Language 1.0 (APPEL1.0) - W3C Working Draft", *World Wide Web Consortium*, http://www.w3.org/TR/2002/WD-P3P-preferences-20020415/, 15 April 2002

[29] RSA Security, "An Enterprise Perspective on Identity Theft", www.rsasecurity.com, December 2003

[30] Ames W., "Understanding Spyware: Risk and Response", *IEEE IT Professional*, Vol.6, No.5, September - October 2004, pp.5 – 8

[31] HSBC Bank USA, N.A, "Terms and Definitions", Security Site HSBC Bank USA, N.A., http://www.us.hsbc.com/securitysite/termsanddefinitions.html, 2005

[32] APWG, "Anti-Phishing Working Group", http://www.antiphishing.org/, 2004

[33] Gordon G. R, Willox N. A., "Identity Fraud: A Critical National and Global Threat", Economic Crime Institute, http://www.lexisnexis.com/presscenter/hottopics/ECIReportFINAL.pdf, October 2003

[34] Lowry J. S., "The Identification Process Deconstructed", *NIST Workshop on Storage and Processor Card-based Technologies*, http://csrc.nist.gov/card-technology/privacy.html, July 2003

[35] Digital Identity World, "What is Digital Identity?", http://www.digitalidworld.com/, December 2003

[36] Liberty Alliance, "Identity Systems and Liberty Specification Version 1.1 Interoperability", *A Liberty Alliance Technical Whitepaper*, http://www.projectliberty.org/resources/whitepapers/, 14th February 2003

7 REFERENCES

[1] The National Electronic Commerce Coordinating Council (NECCC), "Identity Management," Presented at the NECCC Annual Conference, New York, NY, 4th – 6th December 2002.

[2] S. Sittampalam, "Digital Identity Modelling and Management," MEng Thesis, UTS, Australia, 2005.

[3] J. Phiri and J. Agbinya, "Modelling and Information Fusion in Digital Identity Management Systems," *Proceedings of IEEE International Conference on Systems (ICONS 2006)*, Mauritius, 22nd – 29th April 2006, pp.181 – 186.

[4] F. Chong, Microsoft Architect Journal, "Identity and Access Management," July 2004, http://msdn.microsoft.com/library/en-us/dnmaj/html.

[5] Editor TB1, "Electronic Identity White Paper V 0.5," E-Europe Electronic Identity, November 2002.

[6] S. Brands and F. Legare, Credentica Inc., "Digital Identity Management Based on Digital Credentials," May 2006, http://ls6-www.informatik.uni-dortmund.de/issi/cred_ws/papers/brands.pdf.

[7] P. Faltstrom and G. Huston, Internet Engineering Task Force, "A Survey of Internet Identities," April 2004, http://www.ietf.org.

[8] Enterprise Identity Management, "Strategy White Paper," Microsoft Windows 2000 Server, June 2005.

[9] E. Norlin and A. Durand, "Federated Identity Management," PingID Network Inc. White paper, 2002.

[10] Liberty Alliance Project, "Introduction to the Liberty Alliance Identity Architecture," 2003, http://www.projectliberty.org.

[11] Office of the Federal Privacy Commissioner, "We Must Get Identification Management Right to Avoid Losing Our Privacy," March 2004, http://www.privacy.gov.au/news/media/04_04.html.

[12] Performance Technologies, Inc., "Introduction to GSM," 2005, http://www.pt.com/products/gsmintro.html.

[13] M. Lerner, Michael Lerner Productions, "Connecting to the Internet," 2005 http://www.learnthenet.com/english/html/04connec.htm.

[14] J. Scourias, University of Waterloo, "Overview of GSM," 1996, http://www.shoshin. uwaterloo.ca/publications/pdfs/TR-96-01.pdf.

[15] Ericsson White Paper, "Mobile Multimedia, The Next Step in Richer Communication," 2004, http://www.ericsson.com/products/white _papers_pdf/mobile_multimedia.pdf.

[16] K. A. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.14, No.1, pp 4 – 20, January 2004.

[17] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric Recognition: Security and Privacy Concerns," *IEEE Security and Privacy Magazine*, Vol.1, No.2, pp.33 – 42 March – April 2003.

[18] BiometriTech, "Facial-Recognition Solutions Roundup," May 2002, http://www.biometritech.com/features/roundup051502.htm.

[19] L. O'Gorman, "Comparing Passwords, Tokens, and Biometrics for User Authentication," *Proceedings of the IEEE*, vol.91, No.12, pp. 2019 – 2020, December 2003.

[20] International Biometric Group, "Iris Recognition: How it Works," Technology Report, www.biometricgroup.com, October 2004.

[21] K. Tracy, V. Koerper, "Biometrics: A Brief Introduction," CSC 490 – Security Seminar, http://csc.noctrl.edu/f/kwt/590/projects/VICKIK~1.HTM, March 1998.

[22] International Biometric Group, Technology Report, "Biometric vs. Non-Biometric Fingerprinting," October 2004, www.biometricgroup.com.

[23] D. Dingley, CyberSIGN, "Signature Verification Technology," November 1999, http://www.cybersign.com/news_press2.htm.

[24] Wikipedia, The Free Encyclopaedia, "Information theory," 2006, http://en.wikipedia.org/wiki/Information_theory.

[25] The Open Group, "Single Sign-On," 2005, http://www.opengroup.org/ security/sso.

[26] A. Rezgui, A. Bouguettaya, and M. Y. Eltoweissy, "Privacy on the Web: Facts, Challenges, and Solutions," *IEEE Security & Privacy*, vol.1, No.6, pp. 40 – 49, Nov-Dec 2003.

[27] M. Crompton, Federal Privacy Commissioner, "Proof of ID Required? Getting Identity Management Right," March 2004, http://www.privacy.gov.au/news/speeches/sp1_04p.pdf.

[28] L. Cranor, M. Langheinrich and M. Marchiori, World Wide Web Consortium "A P3P Preference Exchange Language 1.0 (APPEL1.0) - W3C Working Draft," April 2002, http://www.w3.org/TR/2002/WD-P3P-preferences-20020415.

[29] RSA Security, "An Enterprise Perspective on Identity Theft," December 2003, www.rsasecurity.com.

[30] W. Ames, "Understanding Spyware: Risk and Response," *IEEE IT Professional*, vol.6, No.5, pp. 5 – 8, September – October 2004.

[31] HSBC Bank USA, N.A, Security Site HSBC Bank USA, N.A., "Terms and Definitions," 2005, http://www.us.hsbc.com/securitysite/termsanddefinitions.html.

[32] APWG, "Anti-Phishing Working Group," 2004, http://www.antiphishing.org.

[33] R.G. Gordon and A. N. Willox, Economic Crime Institute, "Identity Fraud: A Critical National and Global Threat," October 2003, http://www.lexisnexis.com/presscenter/hottopics/ECIReportFINAL.pdf.

[34] S. J. Lowry, NIST Workshop on Storage and Processor Card-based Technologies, "The Identification Process Deconstructed," July 2003, http://csrc.nist.gov/card-technology/privacy.html.

[35] Digital Identity World, "What is Digital Identity?" December 2003, http://www.digitalidworld.com.

[36] Liberty Alliance, A Liberty Alliance Technical Whitepaper, "Identity Systems and Liberty Specification Version 1.1 Interoperability," February 2003, http://www.projectliberty.org/resources/whitepapers.

[37] A. Reed, "Definitive Guide to Identity Management," Rainbow Technologies, 2002.

[38] S. Schimke, S. Kiltz, C. Vielhauer, T. Kalker, "Security Analysis for Biometric Data in ID Documents," SPIE-IS&T/ Vol. 5681, http://wwwiti.cs.uni-magdeburg.de/~sschimke/5681_48.pdf, 2005.

[39] Y. Chen, "Untrustworthy Passport," 2003, http://www.securityfocus.com/guest/20225.

[40] Microsoft, "Microsoft .NET Passport Review Guide," 2003, http://www.microsoft.com/net/downloads/passport_reviewguide.doc.

[41] Liberty Alliance Project, "Identity Systems and Liberty Specification Version 1.1 Interoperability," 2003, http://www.projectliberty.org/resources/whitepapers/Liberty%20and%203rd%20Party%20Identity%20Systems%20White%20Paper.pdf.

[42] J. Picard, C. Vielhauer, and N. Thorwirth, "Towards Fraud-Proof ID Documents Using Multiple Data Hiding Technologies and Biometrics," in *SPIE Proceedings – Electronic Imaging, Security and Watermarking of Multimedia Contents VI*, 2004, pp. 123–234.

[43] T. Petermann, C. Scherz, and A. Sauter, " Biometrie und Ausweisdokumente (Biometrics and Identity Documents)," TAB Working Report No. 93, http://www.tab.fzk.de/de/projekt/zusammenfassung/ab93.pdf (in German), 2003.

[44] G. M. Kuhn, and R. J. Anderson, "Soft Tempest: Hidden Data Transmission Using Electromagnetic Emanations," in *Information Hiding*, 1998, pp. 124–142.

[45] N. Poh and S. Bengio, IDIAP Research Institute, Rue du Simplon 4, CH-1920 Martigny, Switzerland, "A Score-Level Fusion Benchmark Database For Biometric Authentication," 2005, http://eprints.pascal-network.org/archive/00000864/01/norman_bmark.pdf.

[46] D. Hall, "Mathematical Techniques in Multisensor Data Fusion," Artech House, Boston, MA, 1992.

[47] SilkRoad Publications, SilkRoad Inc., "IDS Data Fusion," 2005 http://www.silkroad.com/papers/html/ids/node3.html.

[48] N. Negnevitsky, *Artificial Intelligence a Guide to Intelligent Systems*, Addison Wesley, 2002.

[49] C. Stergiou and D. Siganos, Tech. rep., Imperial College, London, "Neural Networks," 2005, http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.

[50] KBCS, KBCS-98 Secretariat, "Basics of Fuzzy Logic and Fuzzy Set Theory," May 1999, http://www.ncst.ernet.in/kbcs/vivek/issues/11.1/sam/node2.html.

[51] H. E. Stephanou and A. P. Sage "Perspective on imperfect information processing," *IEEE Transactions on Systems, Man, and Cybernetics, SMC-75(5),* p780 – 798, 1987.

[52] Wikipedia, Answers.com, "Bayesian Inference," May 2005, http://www.answers.com/topic/ bayesian-inference.

[53] S. Mohaghegh, Intelligent Solutions, Inc, "Virtual Intelligence and its Applications in Petroleum Engineering," October 2000, http://www.Intelligentsolutionsinc.Com/ Part2.htm.

[54] R. J. Koza, Genetic Programming Inc., "What is Genetic Programming?" August 2003, http://www.genetic-programming.com/gpanimatedtutorial.html.

[55] Microsoft, ".NET Passport: Balanced Authentication Solutions," April 2003, http://www.microsoft.com/net/services/passport/balanced.asp.

[56] Liberty Alliance, Version 1.1.1, ETSI "Identity Specialist Task Force 180, Universal Communications Identifier (UCI): System framework EG 202 067," September 2002, http://portal.etsi.org/docbox/EC_Files/EC_Files/ eg_202067v010101p.pdf.

[57] Internet2, "Shibboleth Project," March 2004, http://shibboleth.internet2.edu.

[58] L. Cranor, M. Langheinrich, and M. Marchiori, World Wide Web Consortium, "A P3P Preference Exchange Language 1.0 (APPEL1.0) - W3C Working Draft," April 2002, http://www.w3.org/TR/2002/WD-P3P-preferences-20020415/.

[59] RSA Security, "An Enterprise Perspective on Identity Theft," December 2003, www.rsasecurity.com.

[60] Specialist Task Force 180, Version 1.1.1, *ETSI*, "Universal Communications Identifier (UCI): System Framework EG 202 067," September 2002, http://portal.etsi.org/docbox/ EC_Files/EC_Files/eg_202067v010101p.pdf.

[61] G. Wachob, D. Reed *et al.*, *OASIS*, "Extensible Resource Identifier (XRI) Generic Syntax and Resolution Specification (Committee Draft)," January 2004, http://www.oasis-open.org/committees/xri/xri-syntax-resolution-1.0-cd.

[62] G. Dounias, University of the Aegean Dept. of Business Administration, "Hybrid Computational Intelligence in Medicine," 2005, http://cyber.felk.cvut.cz/EUNITE03-BIO/pdf/Dounias.pdf.

[63] StatSoft, Electronic Text Book StatSoft, StatSoft Inc., "Data mining Techniques," 2003, http://statsoft.com/textbok/stdatmix.html.

[64] Howstuffworks, "How Credit Cards Work," 2005, http://money.howstuffworks.com/credit-card.htm.

[65] F. Yergeau., The Internet Engineering Task Force, "RFC 3629 UTF-8, A Transformation Format of ISO 10646," November 2003, www.ietf.org.

[66] The Unicode Consortium, "The Unicode Standard, Version 4.0," August 2003, www.unicode.org.

[67] ISO, International Organisation for Standardisation, "ISO 8601:2004 Data Elements and Interchange Formats – Information Interchange – Representation of Dates and Times," 2004, http://www.iso.org/iso/en/prodsservices/popstds/datesandtime.html.

[68] Cisco Networks Networking Academy, "The TCP/IP Model, CCNA1: Networking Basics v3.0," 2005, http://www.pku.edu.cn/academic/research/computer-center/tc/html/TC0102.html.

[69] M. Hall, *Servlets and JavaServer Pages*, Second Edition, Sun Microsytems Press Publisher, 2002.

8 APPENDIX I

8.1 UserInput.jsp class

```
<%@page contentType="text/html"%>
<%@page pageEncoding="UTF-8"%>
<html><head>
<title>Submit Your Inputs</title></head>

//Setting the title, background colours, foreground colour and default font of the form
<BODY BGCOLOR="#CADFDF" TEXT="#000000" LINK="#666600" ALINK="#FF0000"
VLINK="#800080" BACKGROUND="bckgrnd15.jpg">
<H1><CENTER><FONT FACE="Times New Roman" SIZE="10" COLOR="#3F66D6">University of
the Western Cape </FONT></CENTER></H1>
<CENTER><FONT FACE="Times New Roman" SIZE="6"
 COLOR="#3F66D6">Department of Computer Science
</FONT>
</CENTER>
<HR WIDTH="70%" SIZE="2" NOSHADE>
<H4 ALIGN="CENTER"> FILL IN THE DETAILS AND CLICK ON THE LOGIN BUTTON BELOW
</H4>

//Setting and defining the form input variables
<FORM action=JJ.jsp method=post>
<TABLE ALIGN="CENTER">
 <TBODY>
 <TR>
  <TD>National ID Number:
  <TD><INPUT size=35 name=NID>
 <TR>
  <TD>Your Last Name:
  <TD><INPUT size=35 name=lname>
 <TR>
  <TD>Password:
  <TD><INPUT type=password name=password>
 <TR>
  <TD>Your IP Address is:
  <% String ipAddr = request.getRemoteAddr(); %> //Get the IP address of User
   <INPUT type=hidden name=IPaddr value="<%= ipAddr %>"> //Load IP address
  <TD>  <%= ipAddr %>  //Print IP address of User on the form
  </TBODY>
</TABLE>
 <BR>

<HR WIDTH="60%" SIZE="2" NOSHADE>
 <CENTER>
 <INPUT type=submit value="Please Login ...">
 <INPUT type=reset value="Reset">
 </CENTER>
 <BR>
 <HR WIDTH="70%" SIZE="2" NOSHADE>
</FORM>
<BR>
<CENTER><FONT SIZE="1" FACE="VERDANA, ARIAL">
```

```
</A><a href="http://www.uwc.ac.za" target=>
****Created by Jackson Phiri - MSc Student - Department of Computer Science - UWC****</a>
</FONT>
</CENTER>
</body></html>
```

8.2 JJ.jsp class

```
<%@page contentType="text/html"%>
<%@page pageEncoding="UTF-8"%>
<%@page import="java.awt.*,java.awt.event.*,javax.swing.*,java.sql.*"%>

<html><head><title>JSP Page</title></head><body>
  <%!
//Define the variables
String driver = "sun.jdbc.odbc.JdbcOdbcDriver";
String url = "jdbc:odbc:DIMS";
String record, Nid,NID,fname,lname,dob,ip,mac,pwd, Lname;
Connection con;
Statement stmt;
ResultSet rs;
%>
 <%!
double x1  = 0.0000;
double x2  = 0.0000;
double x3  = 0.0000;
double x4  = 0.0000;
double x5  = 0.0000;
double x6  = 0.0000;
double x7  = 0.0000;
double x8  = 0.0000;
double x9  = 0.0000;
double x10 = 0.0000;
double x11 = 0.0000;
double finaloutput = 0.0000;
%>

<BR>
<HR>
<H2 ALIGN="CENTER"> THE FOLLOWING ARE YOUR DETAILS</H2>
<BR>
<HR>
<CENTER>

<%
//Load the Database drivers
try {
Class.forName(driver);
%>
<%} catch (ClassNotFoundException e) {%>
<p>ClassNotFoundException: <%= e.getMessage() %>
<%}%>

//---------Retrieving the submitted user input attributes from the UserInput class ----------
      <%
```

```
//Physical Metrics
String myNid = request.getParameter("NID");
String myLname = request.getParameter("lname");
//Pseudo Metrics
String mypwd = request.getParameter("password");
//Device Metrics
String myip = request.getParameter("IPaddr");
%>

//---------Authenticate the submitted copies against those in the database system----------
 <%
try {
String qrl1 = "SELECT NID FROM PhysicalMetrics WHERE NID    = '"+ myNid+"'";

String qrl2 = "SELECT LName FROM PhysicalMetrics ph WHERE  ph.NID  = '"+ myNid+"'  AND
ph.LName ='"+ myLname+"'";

String qrl3 = "SELECT Password FROM PseudoMetrics ps,PhysicalMetrics ph  WHERE ps.Password =
'"+ mypwd+"' AND ph.Nid = '"+ myNid+"' AND ps.Nid = '"+ myNid+"'AND ps.Nid = ph.NID";

String qrl4 = "SELECT IP FROM DeviceMetrics de,PhysicalMetrics ph   WHERE de.IP    = '"+ myip+"'
AND ph.NID = '"+ myNid+"' AND de.NID = '"+ myNid+"'AND de.NID = ph.NID";

con = DriverManager.getConnection(url, "", "");
stmt= con.createStatement();

//----- Retrieving the results from the Database ---
        rs = stmt.executeQuery(qrl1);
            while (rs.next()){
            Nid = rs.getString("NID");
            }
        rs = stmt.executeQuery(qrl2);
                while (rs.next()){
                    Lname   = rs.getString("LName");
            }
        rs = stmt.executeQuery(qrl3);
                    while (rs.next()){
                    pwd   = rs.getString("Password");
            }
        rs = stmt.executeQuery(qrl4);
                    while (rs.next()){
                    ip   = rs.getString("IP");
            }
//----Assigning the weights to the attributes to be used in the information fusion engine ----
        try{
           if(Nid.equals(myNid)){
           x1 = 2.31786;
           }
           else{ x1 = 0.0000; }

           if(Lname.equals(myLname)){
           x2 = 2.31839;
           }
           else{ x2 = 0.0000; }

           if(pwd.equals(mypwd)){
```

```
        x3 = 2.30717;
        }
        else{ x3 = 0.0000; }

        if(ip.equals(myip)){
        x4 = 2.30724;
        }
        else{ x4 = 0.0000; }
```

//Close the Database and Catch all the errors

```
        } catch(Exception ex1){
        System.err.println("Exception: " + ex1.getMessage());
        }
        stmt.close();
        con.close();
        } catch(SQLException ex) {
        System.err.println("SQLException: " + ex.getMessage());
        System.out.println("SQLException: " + ex.getMessage());
        }
 %>
```

//--Forward the weights to the information fusion Engine class called NeuralNet.jsp --

```
<jsp:forward page="NeuralNet.jsp">
<jsp:param name="z1" value="<%= x1 %>" />
<jsp:param name="z2" value="<%= x2 %>" />
<jsp:param name="z3" value="<%= x3 %>" />
<jsp:param name="z4" value="<%= x4 %>" />
<jsp:param name="z5" value="<%= x5 %>" />
</jsp:forward>

</CENTER>
<BR><HR>
</body></html>
```

8.3 NeuralNet.jsp class

```
<%@page contentType="text/html"%>
<%@page pageEncoding="UTF-8"%>
<%@page import="java.awt.*,java.awt.event.*,javax.swing.*,java.sql.*"%>

<html><head><title> Information Fusion Computation </title></head><body>
<%--Defining the valuables for use to compute information fusion-- %>
  <%!
   public double x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11;
   public double w11,w12,w22,w33,w44,w106,w107,w108,w109;
   public double q6,q7,q8,q9,q10;
   public double y6,y7,y8,y9,y10;
   public double X6,X7,X8,X9,X10,Y6,Y7,Y8,Y9,Y10;
   %>
```

//Get the input weights of the credential attributes from the JJ.jsp class

```
     <%
      String v1 = request.getParameter("z1");
```

```
String v2 = request.getParameter("z2");
String v3 = request.getParameter("z3");
String v4 = request.getParameter("z4");
String v5 = request.getParameter("z5");
%>

<%!
//----compute method used to implement Information fusion engine using artificial neural networks --
public double compute(double a1,double a2,double a3,double a4,double a5){

//Assigning input variables from JJ class a1,a2,a3,a4 and a5 to the input vectors x1,x2,x3,x4 and x5
    x1 = a1;
    x2 = a2;
    x3 = a3;
    x4 = a4;
    x5 = a5;

//----Define the weights of the input layer neurons computed with the help of MatLab------
    w11 = 0.5490;
    w12 = 0.1104;
    w22 = 0.5600;
    w33 = 0.5600;
    w44 = 0.5600;

//---Define the weights of the hidden layer neurons computed with the help of MatLab-----
    w106 = 7.0970;
    w107 = 5.0617;
    w108 = 4.4144;
    w109 = 5.7191;

//----Define the threshold values of the hidden and output layer neurons computed with MatLab-----
    q6 = 3.2971;
    q7 = 2.8000;
    q8 = 2.8000;
    q9 = 2.8000;
    q10 = 2.5270;

X6 = (((x1*w11)+(x2*w12)) - q6); //Equation 4.1 used to compute overall input of Physical metrics
  X7 = ( (x3*w22)-q7);          //Equation 4.1 used to compute overall input of Pseudo metrics
  X8 = ( (x4*w33)-q8);          //Equation 4.1 used to compute overall input of Device metrics
  X9 = ( (x5*w44)-q9);          //Equation 4.1 used to compute overall input of Biometrics

   y6 = Math.exp(-X6);          //Equation 4.2 used to compute overall output of Physical metrics
  Y6 = 1 / (1 + y6) ;
  y7 = Math.exp(-X7);  //Equation 4.2 used to compute overall output of Pseudo metrics
  Y7 = 1 / (1 + y7) ;
  y8 = Math.exp(-X8);  //Equation 4.2 used to compute overall output of Device metrics
  Y8 = 1 / (1 + y8) ;
 y9 = Math.exp(-X9); //Equation 4.2 used to compute overall output of Biometrics
 Y9 = 1 / (1 + y9) ;
//Equation 4.1 used to compute overall input from the four groupings
X10 = (((Y6*w106)+(Y7*w107)+(Y8*w108)+(Y9*w109))-q10);

y10 = Math.exp(-X10); //Equation 4.2 used to compute overall output from the four groupings
```

```
Y10 = 1 / (1 + y10);
return Y10;// Final Output of the information fusion engine
} %>

<HR>
//--- Load the user attributes into the formula and print its value of the information fusion Y10 -------
The Final Value of Information Fusion is:
<%=
compute(Double.parseDouble("v1"),Double.parseDouble("v2"),Double.parseDouble("v3"),
Double.parseDouble("v4"),Double.parseDouble("v5"));
%>
<BR><HR>
//--Check the Computed Value Y10 and see if it is greater than the threshold value in this case 0.6000,
//--if it is not, then forward control to the ErrorPage together with the values assigned to attributes
//-- Otherwise if the computed value Y10 is greater than or equal to threshold value,
//--then grant access to the user and forward control to the ServiceMe class
<% if(Y10 < 0.6000){%>
   <jsp:forward page="Errorpage.jsp" >
   <jsp:param name="p1" value="<%= v1 %>" />
   <jsp:param name="p2" value="<%= v2 %>" />
   <jsp:param name="p3" value="<%= v3 %>" />
   <jsp:param name="p4" value="<%= v4 %>" />
   <jsp:param name="p5" value="<%= v5 %>" />
   <jsp:param name="p6" value="<%= Y10 %>" />
   </jsp:forward>

   <%}else if(Y10>=0.6000){%>
   <jsp:forward page="ServiceMe.jsp" />
   <%}%>

<HR>
</body>
</html>
```

8.4 MatLab Code for the Multilayer Artificial Neural Network

```
net = network     %%Defines the network
net.numInputs=4; %%Defines the number of input sources
net.numLayers=5; %%Defines the number of layers

%%Bias connections to the layers
net.biasConnect(1) = 1;
net.biasConnect(2) = 1;
net.biasConnect(3) = 1;
net.biasConnect(4) = 1;
net.biasConnect(5) = 1;

%%Connecting the layers one to another
net.inputConnect(1,1) = 1;
net.inputConnect(2,2) = 1;
net.inputConnect(3,3) = 1;
net.inputConnect(4,4) = 1;
net.layerConnect(5,1) = 1;
net.layerConnect(5,2) = 1;
net.layerConnect(5,3) = 1;
```

```
net.layerConnect(5,4) = 1;
net.outputConnect(5) = 1;
net.targetConnect(5) = 1;

%%Define the input vector for every input source
net.inputs{1}.range = [0 10; 0 10];
net.inputs{2}.range = [0 10];
net.inputs{3}.range = [0 10];
net.inputs{4}.range = [0 10];

%%Specify number of neurons, transfer functions and initialisation weight for layer one
net.layers{1}.size = 1;
net.layers{1}.transferFcn = 'logsig';
net.layers{1}.initFcn = 'initnw';

%%Specify number of neurons, transfer functions and initialisation weight for layer two
net.layers{2}.size = 1;
net.layers{2}.transferFcn = 'logsig';
net.layers{2}.initFcn = 'initnw';

%%Specify number of neurons, transfer functions and initialisation weight for layer three
net.layers{3}.size = 1;
net.layers{3}.transferFcn = 'logsig';
net.layers{3}.initFcn = 'initnw';

%%Specify number of neurons, transfer functions and initialisation weight for layer four
net.layers{4}.size = 1;
net.layers{4}.transferFcn = 'logsig';
net.layers{4}.initFcn = 'initnw';

%%Specify number of neurons, transfer functions and initialisation weight for layer five
net.layers{5}.size = 1;
net.layers{5}.transferFcn = 'logsig';
net.layers{5}.initFcn = 'initnw';

%%Specify the input and layer weight
net.inputWeights{1,1}.delays = 1;
net.inputWeights{2,2}.delays = 1;
net.inputWeights{3,3}.delays = 1;
net.inputWeights{4,4}.delays = 1;
net.layerWeights{5,1}.delays = 1;
net.layerWeights{5,2}.delays = 1;
net.layerWeights{5,3}.delays = 1;
net.layerWeights{5,4}.delays = 1;

%%Define the globe variables
net.initFcn = 'initlay';
net.trainFcn = 'trainlm';
net.performFcn = 'mse';
net = init(net)

%%Training the network, where P is the Training data and T is the Targeted data
net.trainParam.goal = 1e-10;
net = train(net,P,T);
Y=sim(net,P)
```