

Using Artificial Neural Networks to Implement Information Fusion in Digital Identity Management Systems

Jackson Phiri and ++Johnson I Agbinya

Department of Computer Science

University of the Western Cape, South Africa, jphiri@uwc.ac.za

++University of Technology, Sydney, Australia, agbinya@eng.uts.edu.au

Abstract

The management of identities of the citizens, public institutions and other organisation has been the core function of the government for years now. The increase in the number of users accessing online services using communication devices such as computers, mobile phone and cards such credit cards has prompted most governments and business organizations to change the way they do business and manage their identity information. This technology and globalisation has also seen most Internet users being vulnerable to identity fraud, which is on the increase and costing the global industry excessive amounts. In this paper we propose Digital Identity Management System (DIMS) as a solution for managing digital identity information. The system proposed uses technologies like artificial intelligence and biometrics on the current open networks to maintain the security and privacy of users and service providers in a transparent, reliable and efficient way. DIMS in this paper will use Multimodal Authentication which will be implemented using a technique of information fusion with artificial neural networks to combine the user credentials for optimum recognition of a user.

Keywords Credentials, Attributes, Digital Identity, Authentication, Authorization, Information Fusion

1. Introduction

For the past millennia before the Information Technology age, governments have been managing the identities of their citizens, organizations and other public institutions using the traditional methods [1]. Today most countries are now implementing e-government and e-services [2] which require an effective way to manage digital identity information of the

users. Digital Identity and Management System is not just about technology but is comprised of three indispensable elements which include [3] Policies, Processes and Technology. However, in this paper we consider the technology segment which can further be subdivided into three major areas. These are Identity Life Cycle, Access Management and Directory Services [3]. The Life Cycle of digital identities involves the process of their creation, utilization and termination. Access Management refers to the process of controlling and granting access to satisfy resource request. This process is usually completed through a sequence of authentication, authorization and auditing actions. Finally Directory Services provides the infrastructure for secure data storage and organization [3].

The rest of this paper will look at the details of multimode authentication.

2. Background Information

2.1. Digital Identity

The definition of digital identity usually depends on the usage, situation, purpose and several other factors [4]. Eric and Adre [5] have defined digital identity as “A digital identity is a virtual representation of a real identity that can be used in electronic interactions with other machines or people”. An identity consists of traits, attributes, and preferences upon which one may receive personalized services. Such services could exist online, on mobile devices, at work, or in many other places.

2.2. Desirable Properties of Internet Identities

The desirable properties of Internet identities include [6] uniqueness, consistency, persistency and verifiability although not comprehensive. These properties will be used to

analyse the effectiveness of identity systems used in digital environment [4].

2.3. Fraud and Privacy in Digital Identity

Identity theft and identity fraud are terms used to refer to all types of crime in which someone criminally obtains and uses another person's personal data in some way that involves economic gain [4]. With the rapid development of new technologies in telecommunications, the Internet and the wide spread of globalisation, identity fraud has become one of the fastest growing crimes in the world today [4]. Identity fraud has become a major concern for the public and private sectors [7], particularly as it relates to terrorism, money laundering, financial crime, drug trafficking, alien smuggling, and weapons smuggling. Digital identity management system is considered to be a solution for preventing fraud, improve user's privacy and organizational services and increasing national security.

Ferdinand Schoeman [8] 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." A good identity management system should ensure confidence to the degree appropriate for the interaction that the organisation is dealing with the right person.

2.4. Standards and Specifications for Communication Devices

In this paper we consider the GSM [9] standard for mobile phones while IPv4 is considered for Internet Protocol in addressing the individual Internet terminal. Finally ISO 7816-4 which provides logical file structure for smart cards is considered as the standard for smart cards [10] to address security and storage standards.

2.5. Information Fusion

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 [4]. Multimode credential authentication which will involve a combination of a number of credentials attributes (Physical metrics, Biometrics, Pseudo metrics, and Device metrics) in order to authenticate a user or a device representing a

user will be considered as a solution to this problem. The process of combing these credential attributes is refereed to as information fusion [11]. Information fusion technology has been applied most prominently to military applications such as battlefield surveillance and tactical situation assessment [11]. It has also emerged in commercial applications such as robotics, manufacturing, medical diagnosis, and remote sensing [12]. In this paper, the implementation of information fusion will involve the technique of artificial neural networks.

2.6. Related Works

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 [4]. Examples of such schemes are Electronic Number Mapping (E-NUM) [13] and Universal Communications Identifier (UCI) [14]. After the 2001 September 11 attack in the United States, most countries are now considering adopting national identity cards. These include the United States, United Kingdom, the Philippines, and the Netherlands [15]. There have been a number of organisations and companies working on digital identity management systems and identity related applications.

Microsoft Passport 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 [16].

Shibboleth is an identity management framework which specifies architectures, policy structures, technologies designed mainly for academic institutions using open source implementation [17]. It was developed by Internet2 Consortium being led by about 205 universities working in partnership with the industry and the government to develop and deploy advanced network applications and technologies, accelerating the creation of future Internet.

3. Methodology

We employ Credential Attributes Mapping [4] scheme to come up with the digital identity representation of a user or a device representing a user. To obtain optimum recognition of the user, artificial neural network, fuzzy logic or Bayesian methods will be used to combine the

credentials attributes presented by the user during information fusion. The result of information fusion will then be used to perform multimode authentication. In this paper we shall consider using artificial neural networks and MATLAB shall be used to train and simulate the network. Storage of attributes will relies on a database which is used as a distributed database system. Only the details of multimode authentication using a mobile phone are considered as an example.

4. Design and Implementation

4.1. Design

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 and then the credential 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 [4] as computed in table 3 or a zero is if it does not much a copy in the database system. With their assigned values, the credential attributes are then forwarded to the information fusion engine. The outcome value of the Information Fusion Engine is then used in multimode authentication as described in details in section 5 below.

4.2. Choosing Identity Credential

A questionnaire was passed to respondents from a wide range of countries (International Students) requesting them to list as many identity documents and tokens used in their countries to access the services or secure areas. We finally came up with 14 identity credentials. Another questionnaire was passed requesting another set of responders to grade these credentials according to their level of importance with 5 being an extremely important or useful credential and 1 being not important at all. Table 1 below gives us the average scores of 40 respondents. These credentials were used as the source for the identity attributes used to implement our system.

Table 1: Identity Credentials Used as the source of the identity attributes

Index	Identity Credential	Grade out of Five
-------	---------------------	-------------------

1	National ID Card	4.13
2	Birth Cert	3.80
3	Citizenship Cert	3.53
4	Passport	4.10
5	Acceptable IDs	3.10
6	Driver Licence	3.10
7	Credit Cards	3.18
8	Bank Cards	3.53
9	Insurance Membership Cards	2.78
10	Club Membership Cards	2.33
11	Student Cards	3.08
12	Mobile Phones	3.10
13	Internet Terminals	3.08
14	School Certificates	3.68

4.3. Choosing Credential Attributes

From table 1 above, we came up with 38 credential attribute and table 2 below shows 16 of the 38 credentials attributes.

Table 2: Identity Attributes from the 14 identity credentials in table 1

Index	Attributes	Source of Attribute
1	National Identity Number Or ID Number	From Commonly used identity documents like passport, driver's licence, birth certificate, national identity cards etc.
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	Cert. number	
12	Citizenship	
13	Height	
14	Signature	
15	Personal Identity Number (PIN) (e.g. used on ATM cards)	From Secrete Codes
16	Password (e.g. used on your emails)	

We then decided to group these attributes into four classes for easy analysis.

4.4. Classification of Digital Identity

A user [18] acquires many forms of identity attributes which are stored in various forms and places from the time he/she is born. In this paper, we divide our identity attributes into four major groups. These are physical metrics, pseudo metrics, device metrics and biometrics.

Physical metric authentication relies on evidence of what you have. These include attributes like your name, age, address and national ID number or social security number.

Pseudo metric authentication uses a shared secret between the service provider and the user. A good example is a password or a personal identification number (PIN).

The term biometrics refers to the ‘automatic’ recognition of personal identity based on someone’s biological, physiological and/or behavioural characteristics. Examples of biometrics include iris scan, finger print and face recognition.

Device metrics refers to attributes of non-human communication device. Examples of these devices include mobile phones, Internet terminals (PCs) and smart cards. Examples of Device metrics are International Mobile subscriber Identity for a mobile phone and an IP address for a PC.

4.5. Determining Credential Attribute Strength Using a Questionnaire and Shannon’s Information theory

A questionnaire was distributed and respondents were requested to grade these attributes out of five using the desirable properties of Internet Identities [6] such as uniqueness, consistency, verifiability and persistence to the physical metrics, pseudo metrics and device metrics [4]. Permanence, universality, distinctiveness, performance and circumvention were applied to biometrics. For example a 5 means extremely unique while 1 not unique at all for a National ID number. Table 3 shows the average scores of identity attributes from questionnaires.

Table 3: Average Scores from a questionnaire showing only 8 of the 38 Identity Attributes

Index	Attributes	Average scores of Desirable Properties of Internet Identities				
		Uniqueness	Verifiability	Consistency	Persistence	Trust
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

Table 4: Assigning weights to Inputs variables for mobile phone Multimode Authentication

Index	Attributes	$P_i \text{Log}_2 (1/P_i)$					$\sum P_i \text{Log}_2 (1/P_i)$
		Uniqueness	Verifiability	Consistency	Persistence	Trust	
1	National Identity Number	0.4855	0.4626	0.4673	0.4486	0.4539	2.318
2	Full name	0.4637	0.4479	0.4745	0.4799	0.4524	2.318
3	Residential Address	0.4668	0.4770	0.4770	0.4304	0.4647	2.316
4	Date of birth	0.4516	0.4671	0.4789	0.4720	0.4493	2.319
5	City of birth	0.4267	0.4560	0.4750	0.4957	0.4560	2.309
6	Country of birth	0.4514	0.4712	0.4738	0.4851	0.4328	2.314
7	Race	0.4287	0.4668	0.4763	0.4822	0.4606	2.315
8	Mother’s Name	0.4371	0.4594	0.4744	0.4821	0.4637	2.317

With the average scores shown in table 3, we then employed Shannon’s information theory to compute the weights of the attributes given by the following equation;

$$H(X) = \sum_{i=1}^n p(X_i) \log_2 [p(X_i)] \dots (1)$$

Where; X is the message

- Xi is the ith symbol in the message
- p(Xi) is the probability of the occurrence of the ith symbol

$$H = \sum_{i=1}^n p_i \log_2 (1/p_i) \dots (2)$$

Here;

- X is the desirable property of Internet identities
- Pi is the ratio of the weight of the ith desirable property of Internet identities to the sum weights of all desirable property of Internet l identities considered

The information content computed in table 4 is the input weight of the network described below.

4.6. Implementation using Multilayer Neural Network on Mobile Phone

4.6.1 Input Variables

We decided to choose 5 identity attributes for our system as shown in table 5 one at least from each of the four classes.

Table 5: Inputs Variables for Mobile Phone Multimode Authentication

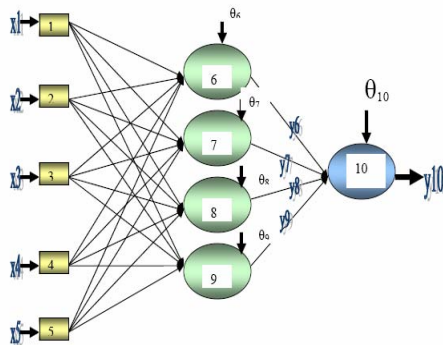
Index	Attribute	Information Content ($\sum \text{PI} \text{Log}_2(1/\text{PI})$)	Input Variable
Physical Metrics			
1	National Identity Number (NID)	2.31786	x1
2	Last Name	2.31839	x2
Pseudo Metrics			
3	Personal Identity Number (PIN)	2.30717	X3
Device Metrics			
4	International Mobile Equipment Identifier (IMEI)	2.30724	X4
Biometrics			
5	Voice Recognition	2.31825	X5

4.6.2 Artificial Neural Network

A multimode neural network with ten neurons and three layers as shown in figure 1 will be used in this implementation. Neuron 1 to 5 will be the input neurons while neuron 6, 7, 8 and 9 represent the groupings physical metrics, biometrics, pseudo metrics and device metrics respectively. Neuron 10 is the output neuron with y10 used for multimode authentication.

Attributes strengths from the metrics as shown in table 5 are used as input neurons for the network. Input variables x1 to x2 form the vector for the physical metrics while x3 form the vector for biometrics. Neurons x4 form the pseudo metrics vector and finally neurons x5 form the device metrics vector.

Figure 1: Three Layer Artificial Neural Network for Information Fusion



4.6.3 Activation Function

Many activation functions have been tested but only a few like the step, sign, sigmoid and

linear functions have found practical application [19]. In this paper we shall consider the use of sigmoid function. Sigmoid function transforms the input, which can have any value between plus and minus infinity into a reasonable value in the range between 0 and 1. 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 neuron output is bound between 0 and 1 [19]. The activation function will be represented by the following equations [19];

$$X = \sum_{i=1}^n x_i w_i - \theta \quad \dots\dots\dots (3)$$

Here X is the net weighted input to the neuron, n is the number of neuron inputs, and θ is the threshold to the neuron. x_i is the value or strength of the input variable and w_i is its respective weight [19].

$$Y = \frac{1}{1 + e^{-X}} \quad \dots\dots\dots (4)$$

Using the output X of equation (1), equation (2) computes the output value Y for each neuron in the layers two and three. Y6, y7, y8 and y9 are the respective output strengths of neurons 6, 7, 8 and 9 representing the output strength of physical metrics, biometrics, pseudo metrics and device metrics respectively. y10 is the final output resulting from combining the strengths of the four output y6, y7, y8 and y9. It represents the overall strength of the information fusion engine and is used for multimode authentication decision.

4.6.4 Computing the Neuron Output

Here is an example on how to compute the output of neuron 6 using 5 inputs at least one from each grouping. Using table 5, figure 1 and table 6 with w61, w62, w63, w64, and w65 as the initial input weights to neuron 6 from the 5 input vectors and θ_6 as its threshold value, we computed the output of neuron 6 using equation 1 and 2 as follows;

$$y_6 = 1 / (1 + e^{-((2.3177x1) + (2.3184x0) + (2.3072x0) + (2.3072x0) + (2.3183x0) - 1)})$$

$$\therefore y_6 = 1 / (1 + e^{-1.31786}) = 0.788815 \approx 0.79$$

to 2 decimal places. By adjusting θ_6 and initial weights (w_{ij}), we can come up with the targeted value set for y6. Using the same method, we computed the values of y7, y8, y9 and y10.

Table 4: Neural 6's Initial Inputs Values

#	w61	w62	w63	w64	w65	θ_6
1	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000

5. Results and Discussion

We trained our network using MatLab software to get input weights, layer weights and threshold values of our network which would give us the targeted value of y_{10} [19] as the final output. The final result of the network (y_{10}) will be determined by the number of correct attributes submitted during multimode authentication which would give any of the following outcome. Unsuccessful multimode authentication is when the value of y_{10} is less than the threshold value set for system and occurs when the wrong attributes are submitted. An exceptional case is one which occurs when the computed value of y_{10} is not in between the range 0 and 1 (sigmoid function). In both cases above and the user will be denied access to the system. Successful multimode authentication occurs when the submitted credential attributes give the value of y_{10} equal to or greater than the threshold value. This system will require a user to submit at least one credential attribute from each grouping which must much a copy in the database in order to be authentication. Only a mobile phone was considered here but this could also be extended to Internet terminals and card based device.

6. Conclusion

Multimode authentication is likely to be the solution for most problems of fraud and identity theft seen on the cyber space today. As shown in section 4, artificial neural networks combined with biometrics and communication devices forms a very effective and intelligent multimode authentication system in a digital identity management system. It is hoped that most countries will embrace multimode authentication in managing the digital credential attributes of their citizens when implementing e-governments.

7. References

- [1] The National Electronic Commerce Coordinating Council (NECCC), "Identity Management", Presented at the NECCC Annual Conference, Dec. 4-6, 2002, New York, NY
- [2] Economic Commission for Africa, "E-Strategies", National, Sectoral and Regional ICT Policies, Plans and Strategies, <http://www.uneca.org/aisi/nici/Documents/E-Strategies.pdf>, 2003

- [3] Fredrick Chong, "Identity and Access Management", Microsoft Architect Journey, <http://msdn.microsoft.com/library/en-us/dnmaj/html>, July 2004
- [4] Subenthiram Sittampalam, "Digital Identity Modelling and Management", MEng. Thesis, 2005, UTS, Australia
- [5] Abdelmounaam Rezgui, Athman Bouguettaya, and Mohamed Y. Eltoweissy, "Privacy on the Web: Facts, Challenges, and Solutions", IEEE Security & Privacy, Vol.1, No.6, Nov-Dec 2003, pp.40 - 49
- [6] P. Faltstrom and G. Huston, "A Survey of Internet Identities", *Internet Engineering Task Force - Work in Progress - draft-iab-identities-01.txt*, www.ietf.org, 28 April 2004
- [7] Lorrie Cranor, Marc Langheinrich, and Massimo Marchiori, "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
- [8] Abdelmounaam Rezgui, Athman Bouguettaya, and Mohamed Y. Eltoweissy, "Privacy on the Web: Facts, Challenges, and Solutions", IEEE Security & Privacy, Vol.1, No.6, Nov-Dec 2003, pp.40 - 49
- [9] Performance Technologies, "Introduction to GSM", Performance Technologies, Inc, <http://www.pt.com/products/gsmintro.html>, 2005
- [10] ISO/IEC 7816-4, "ISO/IEC 7816 Part 4: Interindustry command for interchange, first edition 1995-09-01", http://www.tffn.net/techno/smartcards/iso7816_4.html
- [11] Hall D., "Mathematical Techniques in Multisensor Data Fusion", Artech House, Boston, MA, 1992.
- [12] SilkRoad publications, "IDS Data Fusion", SilkRoad Inc, <http://www.silkroad.com/papers/html/ids/node3.html>, 2005
- [13] RSA Security, "An Enterprise Perspective on Identity Theft", www.rsasecurity.com, December 2003
- [14] Specialist Task Force 180, "Universal Communications Identifier (UCI): System framework EG 202 067", ver. 1.1.1, ETSI, http://portal.etsi.org/docbox/EC_Files/EC_Files/eg_202_067v010101p.pdf, September 2002
- [15] Lorrie Cranor, Brooks Dobbs, *et al.*, "The Platform for Privacy Preferences 1.1 (P3P1.1) Specification - W3C Working Draft", <http://www.w3.org/TR/2005/WD-P3P11-20050104/>, 4 January 2005
- [16] Microsoft, ".NET Passport: Balanced Authentication Solutions", <http://www.microsoft.com/net/services/passport/balanced.asp>, April 2003
- [17] Interent2, "Shibboleth Project", <http://shibboleth.internet2.edu/>, March 2004
- [18] Liberty Alliance Project, "Introduction to the Liberty Alliance Identity Architecture", <http://www.projectliberty.org/>, 2003
- [19] Michael Negnevitsky, "Artificial Intelligence, first edition", Addison Wesley Publishing, ISBN Number 0-201-71159-1

8. Biography

Jackson Phiri obtained a BSc Computer Science at the University of Zambia in 2004. Currently an MSc Computer Science student at UWC, South Africa. His interests include Biometrics, Artificial Intelligent Technologies and Online Security Systems.