International Journal of Computational Intelligence Systems, Vol. 4, No. 4 (June, 2011), 420-430

Using Artificial Intelligence Techniques to Implement a Multifactor Authentication System

Jackson Phiri[1], Tie-Jun Zhao[2], Cong Hui Zhu[3]
School of Computer Science, Harbin Institute of Technology, No.11 Siling Street, Nangang District, Harbin, 150001, China
jackson.phiri@gmail.com[1], tjzhao@mtlab.hit.edu.cn[2], chzhu@mtlab.hit.edu.cn[3]
www.en.hit.edu.cn

Jameson Mbale
Department of Computer Science, University of Namibia, P/Bag 13301, Windhoek, Namibia,
jmbale@unam.na
www.unam.edu.na

Received 11 November 2010
Accepted 5 March 2011

Abstract

The recent years have seen a rise in the number of cases of cyber-crime committed through identity theft and fraud. To address this problem, this paper uses adaptive neural-fuzzy inference system, fuzzy logic and artificial neural network to implement a multifactor authentication system through a technique of information fusion. To begin with, the identity attributes are mined using the three corpora from three major sources namely the social networks, a set of questionnaires and application forms from the various services offered both in the real and cyberspace. The statistical information generated by the corpora is then used to compose an identity attribute metric model. The composed identity attributes metrics values classified as biometrics, device metrics and pseudo metrics are then fused at the score level through a technique of information fusion in a multifactor authentication system by using each of the above artificial intelligence technologies and the results compared.

Keywords: authentication, fuzzy logic, neural networks, identity attributes, metrics model, information fusion.

1. Introduction

The last two decades have seen a remarkable increase on the scale of network access and the number of businesses going online hence exceeding the original models for identity and access management for online services.[1] Most of the identity management systems developed today are too complex and come with a lot of flows which poses a lot of security and privacy challenges.[2] This has lead to the growth of cyber crime which is used to perpetuate terrorist activities.[3] Cyber-crime is now one of the biggest industries around the globe and is costing the global industry billions of dollars every year.[3,4] These crimes come in form of identity theft, phishing, company policies violations, fraud, extremism, child pornography and terrorism.[3-6]

This paper looks at a branch of digital identity management systems called authentication. We introduce statistical and artificial intelligence techniques to develop and implement a multifactor authentication system through information fusion. This is in an effort to curb the cases of cyber crime as seen on most online services today. [3,4] We therefore begin by mining the identity attributes from three major sources using the corpora. The first source is composed of identity

attributes mined from the services in the financial institutions, public services, healthy care systems and the education system.[7-11] The second source of identity attributes is a selected set of 25 social networks.[12-13] Finally we roll out questionnaires and 100 responses are used as the third source of the identity attributes. In section 3.1, the AntConc[14] corpus, ConcApp[15] corpus, and the TestSTAT[16] corpus are used to mine and generate the statistical information of the identity attributes from the above sources. The statistical information is then used to compose the identity attributes metrics model using term weight from text mining techniques and entropy from information theory.[17] In section 3.2, a selected set of three identity attributes is then used to develop an information fusion technique using an artificial neural network (ANN),[18] fuzzy inference system (FIS)[19] and a hybrid of the two called adaptive neural-fuzzy inference system (ANFIS).[20] The outcome of the information fusion technique is then used to implement a multifactor authentication system at the score level.[21] The Introduction of the identity attributes metrics values and artificial intelligent techniques in identity management authentication system will most likely help to improve the security features of most online services. [22, 23]

2. Background Information and Related Works

2.1. *Digital Identities Attributes*

In Ref. 23, an identity is defined as consisting of traits, attributes, and preferences upon which one may receive personalized services which could exist online, on mobile devices, at work, or in many other places. Ref. 21 goes further to classify the various forms of identity attributes into three general grouping called *biometrics*, *physical metrics* and *pseudo metrics*. Biometrics refers to the automated technology for measuring and analysing an individual's physical and behaviour characteristics.[22] Examples include the fingerprints, face recognition, iris scan, hand geometry and signature recognition.[23] In Ref. 6, the physical metrics refer to what we have and include all physical based credential token such personal computer, mobile phone and card based credential tokens such as smart cards. In this paper, the identity attributes from the physical tokens will be referred to as *device metrics.* Examples of the device metrics include the Internet Protocol (IP) address, International Mobile Equipment Identifier (IMEI), the Subscriber Identify Module (SIM) and a unique card identification number.[25] The pseudo metrics include all identity attributes in the category of something you know. Good examples are password and personal identity number (PIN).[21] The three groupings are used in this paper in the implementation of a multifactor authentication system through information fusion where the user is required to submit an identity attribute from at least each of the three groupings.

2.2. *Cyber-crime and Privacy*

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.[3] Cyber-crime has become one of the fastest growing crimes in the world.[2,5] Identity fraud has become a major concern for the public and private sectors, particularly as it relates to terrorism, money laundering, financial crime, drug trafficking, alien smuggling, and weapons smuggling as indicated by Ref. 26. To address these security challenges, in Ref. 27, virtual identities which are based on service sessions are introduced to protect the user's privacy towards service providers as well as towards access network providers. Another milestone is the frameworks such as the one introduced by Ref. 28 to trace the history of how a user's identity information is handled after it is transferred across domains of control. Finally Ref. 29 develops an approach that supports privacy controlled sharing of identity attributes in a federated environment. This enables uses to trace their personal information across the federations and verify if the information has been used according to the privacy preferences. This paper introduces statistical and artificial intelligence techniques in a multifactor authentication system to help address the cases of cyber crime.

2.3. *Composition of the Metrics Values*

Metrics models help in the quantitative analysis of most systems thereby accelerating validation processes and improve the efficacy of model design.[30] For example, in Ref. 31, a quantitative metrics model is used to evaluate molecular level system biology model on cell metabolism. They explore the quantitative scoring metrics to compare systems biology modelling. In Ref. 32, they develop guidelines that are capable of helping researchers to select algorithms in systems biology

modelling. A set of quantitative metrics are proposed based on discrete observable units in terms of key bio-modelling considerations. In Ref. 30, they define metrics to automate the quantitative analysis of textual information within a web page. The metrics system is used to determine the difference between the text displayed by the web page and that hidden from users but obtained by the search engine robots. In Ref. 33 a fuzzy logical is applied in the development of a software metric model used to model the development effort estimation in software engineering. Finally in Refs. 34, term weight and entropy are used in the development an identity attributes metric model. They use application forms from the various services offered both in the real and cyberspace as the source of the identity attributes. The corpora are used to automate the mining of the identity attributes to generate the statistical information.[32] The statistical information is used to compute the empirical probability and the entropy is then computed using[34-36];

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

Where p_i is the probability of the identity attribute in the i^{th} sample space.

Using the generated statistical information by the corpora, term weight is used to compose the identity attributes metrics values using (2) where N is the number of the documents in the corpus, *t* the identity attribute being mined and *d* the number of documents where the mined attribute appears in the corpus[34,37];

$$w_{t,d} = (1 + \log tf_{t,d}) \times \log_{10} \frac{N}{df_t} . \quad (2)$$

In this paper we extend the methodology used by Ref. 34 to develop an identity attribute metric model.

2.4. *Information Fusion Techniques*

Information fusion has seen a lot of applications in the areas of robotics, geographical information systems and data mining technologies.[18-20] In Ref. 38 an artificial neural network is used to build an information fusion model in a coal mining monitoring system to ensure the accuracy of transmission of the multi-sensor information that comes from the coal monitoring systems. In Ref. 39, a combination of multiple artificial neural networks is used to improve on-line process fault diagnosis through information fusion. Different combination schemes such as averaging, majority voting and modified majority voting are applied. In Ref. 40 a combination of artificial neural networks, Dempster-Shef (D-S) evidence theory-based information fusion and Shannon entropy is used to form a weighted and selective information fusion technique to reduce the impact uncertainties on structure damage identification. Finally in Ref 41 they give a detailed implementation of an information fusion technique using an ANN and identity attributes metrics values composed using Shannon's information theory (entropy) values. In this paper FIS, ANN and ANFIS are used to implement information fusion technique in a multifactor authentication system.

3. Methodology

3.1. *General Implementation*

The goal is this paper is to implement a multifactor authentication system through a technique of information fusion. Information fusion is however realised by using artificial intelligence technologies and the identity attributes metrics values. The methodology therefore begins by looking at the development of an identity attribute metric model. In Ref. 34, a detailed methodology is given on the implementation of an identity attribute metric model using a total of 200 application forms from the various services offered by both the private and public sectors. In this paper, we extend this methodology by introducing two other sources of the identity attributes in addition to the 200 application forms used in Ref. 34. From the top 1000 website as of April 2010 generated by Google's double click ad planner, twenty-five most popular social networks in different regions of the world are selected and used in this study as the second source of the identity attributes (See Ref. 12 and 13). These include Linkedin, FaceBook, Friendster, reddit, FetLife, QQ, Twitter, MySpace, MSN, Zorpia.Com, Orkurt, Bebo, Netlog, digg, Youtube, Propeller, Stumble Upon, WAYN, Lifeknot, Hi5, Yahoo, Bando, Badoo, Okcupid and renren.com. Secondly we rolled out a set of questionnaires to obtain the user's opinion on identity attributes deemed as important in their respective countries. A total of 100 responses are then used as the

third source of the identity attribute. Using the AntConc, ConcApp and the TestSTAT corpora, the identity attributes are then mined from these sources to generate the statistical information (See Refs. 14-16). Shannon's information theory (entropy) represented by (1) and the text mining techniques (term weight) represented by (2), are then used to compute the two sets of identity attributes metrics values. The implementation details can be found in Ref. 34. Table 1 shows five examples of the computed identity attributes metrics values using term weight and entropy each of which is grouped into one of the three groupings.

Table 1: Term weight and entropy input values for the artificial neural network.

Inputs	Group	Entropy	Term Weight
Card Number	Device Metrics	0.3787	3.5030
Password	Pseudo Metrics	0.0532	3.8429
PIN Number		0.0533	1.5430
Fingerprint	Biometrics	0.5827	8.9740
Face Recognition		0.4004	1.9242

Using the computed metrics values, an information fusion technique is then developed and implemented by using ANN, FIS and ANFIS. In this paper we use three identity attributes shown in Table 2 one from each of the three groupings to implement an information fusion technique and then a multifactor authentication system.

Table 2: Term weight and entropy input values for the artificial neural network.

#	Grouping	Inputs	Term Weight	Entropy
1	Physical Metrics	Unique Card Number	3.5030	0.3787
2	Pseudo Metrics	PIN Number	1.5430	0.0533
3	Biometrics	Fingerprint	8.9740	0.5827

For each of these three implementations (ANN, FIS and ANFIS) first the term weight metrics values are used and then secondly the entropy metrics values are used to develop the information fusion technique. The results are then compared for the two classes of the metrics values. Finally the results from each of the three information fusion techniques implementation are also compared with the conclusion on the effectiveness of both the metrics values and information fusion techniques in the implementing a multifactor authentication system. Below is a brief description of the multifactor authentication system implementation through information fusion by using the three artificial intelligence techniques (ANN, FIS and ANFIS).

3.2. *Multifactor Information Fusion Technique*

3.2.1 Using Artificial Neural Networks

This implementation uses a two layer feedforward network and *sigmoid transfer function* given by the following equations[36,38];

$$X = \sum_{i=1}^{n} x_i w_i - \theta \tag{3}$$

Where θ is the threshold value applied to the neuron, x_i is the identity attribute metric value of the input variable and w_i is the input or layer weight of the neuron. Using the output X, the neuron output is then computed by[36,38];

$$Y^{sigmoid} = \frac{1}{1+e^{-X}} \tag{4}$$

The network is trained using *trainlm* which uses *Levenberg-Marquardt optimization*. Using three identity attributes as an inputs, one from each of the three categories a minimum of eight possible combinations is shown in Table 3 for the term weight and Table 4 for entropy metrics values. The eight combinations are used to train, validate and test the network with the data in the last column used as the target data. The *training state*, *performance* and *regression* are used to validate the network performance.

Table 3: Eight combinations of term weight identity attributes metrics input values and the corresponding target values.

Input #	PIN	Fingerprint	Card-ID	Target
1	0.0000	0.0000	0.0000	0.0000
2	1.5430	0.0000	0.0000	0.1000
3	0.0000	8.9740	0.0000	0.4000
4	1.5430	8.9740	0.0000	0.6000
5	0.0000	0.0000	3.5030	0.3000
6	1.5430	0.0000	3.5030	0.5000
7	0.0000	8.9740	3.5030	0.7000
8	1.5430	8.9740	3.5030	0.9000

Table 4: Eight combinations of entropy identity attributes metrics input values and the corresponding target values.

Input #	PIN	Fingerprint	Card-ID	Target
1	0.0000	0.0000	0.0000	0.0000
2	0.0533	0.0000	0.0000	0.1000
3	0.0000	0.5827	0.0000	0.4000
4	0.0533	0.5827	0.0000	0.6000
5	0.0000	0.0000	0.3787	0.3000
6	0.0533	0.0000	0.3787	0.5000
7	0.0000	0.5827	0.3787	0.7000
8	0.0533	0.5827	0.3787	0.9000

3.2.2 Using Fuzzy Inference System

The fuzzy inference system implementation uses three inputs, seven rules and one output. Sugeno-Style fuzzy inference is used in this paper since it is computationally effective and works well with optimisation and adaptive techniques.[19,42] The *Sigmoidally function* (*sigmf*) with two valuables a and c is used in the fuzzification of biometrics inputs[42];

$$f(x,a,c) = \frac{1}{1+e^{-a(x-c)}}. \quad (5)$$

Triangular membership function (*trimf*) which depends on three scalar parameters a, b and c, where the parameters a and c locate the feet of the triangle while the parameter b locate the peak is used in the fuzzification of the pseudo and device metrics[42];

$$f(x:a,b) = \begin{Bmatrix} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0 & c \le x \end{Bmatrix}. \quad (6)$$

Fingerprint is used to name the *sigmf* while PIN and Card-Number are used to name the *trimf* for the pseudo metrics and device metrics respectively. Table 5 shows the corresponding values of the variables *a, b* and *c* for (5) and (6) for the term weight identity attributes metric values. Table 6 on the other hand shows the corresponding values of the variables *a, b* and *c* for (5) and (6) for the entropy identity attributes metric values. The ouput variables consist of seven constants. These are *VeryLow, Low, Medium, Average, High, VeryHigh* and *All.* The seven rules are then composed from the input variables, output constant variables and the membership functions as follows;

1. *If (PseudoMetrics is PIN) then (Output is VeryLow)(1)*
2. *If (DeviceMetrics is Card-ID) then (Output is Low)(1)*
3. *If (Biometrics is Fingerprint) then (Output is Medium) (1)*
4. *If (PseudoMetrics is PIN) and (DeviceMetrics is Card-ID) then (Output is Average)(1)*
5. *If (PseudoMetrics is PIN) and (Biometrics is Fingerprint) then (Output is High)(1)*
6. *If (Biometrics is Fingerprint) and (DeviceMetrics is Card-ID) then (Output is VeryHigh) (1)*
7. *If (PseudoMetrics is PIN) and (Biometrics is Fingerprint) and (DeviceMetrics is Card-ID) then (Output is All)(1)*

Table 5: Fuzzification functions and variable values using term weight identity attributes metrics values

Name	Type of Function	Function Variable Values
PIN	Trimf	A=-0.9755, b=1.066, c=3.486
Fingerprint	Sigmf	a=1.56, b=3.687
Card-ID	Trimf	A=-1.12, b=2.478, c=5.024

Table 6: Fuzzification functions and variable values using entropy identity attributes metrics values

Name	Type of Function	Function Variable Values
PIN	Trimf	a=-0.06052, b=1.126,c=3.494
Fingerprint	Sigmf	a=0.6894, b=3.074
Card-ID	Trimf	a=-0.9058, b=2.472, c=4.999

Using a Sugeno Adaptive Neural - Fuzzy Inference System (ANFIS)

In this implementation, the ANFIS has three inputs, eleven neurons, three layers, a total of seven rules and one output neuron. Sigmoid shaped membership function is used in the fuzzification of the biometrics while the triangular shaped membership function is used in the fuzzification of pseudo metrics and device metrics. Table 4 and Table 5 show the eight possible inputs and targeted values using term weight and entropy identity attributes metrics values respectively. A combination of the least squares type of method and a backpropagation algorithm are used to train the network in order to generate the output linguistic variable values, the fuzzification membership function variable values and the required output based on the training data. This implementation uses a three layer feedforward network with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer.

4. Results

4.1. *Identity Attributes Metrics*

Table 1 shows an example of five identity attributes with the corresponding computed metrics values from each of the three groupings namely pseudo metrics, biometrics and device metrics. They are composed from a set of 200 application forms, for the services offered both by the private and public sectors, 100 questionnaires and 25 social networks. Term weight and entropy represented by (1) and (2) are used to compose the metric values using the statistical information generated by the three corpora. Three of these identity attributes as shown in Table 2 are used in the implementation of an information fusion technique in this paper using ANN, FIS and ANFIS.

4.2. *Multifactor Information Fusion Technique*

4.2.1 Using Artificial Neural Networks

Table 7 shows the eight possible term weight identity attributes metrics values as an input and the corresponding target and output values. Also shown is the error generated as the difference between the output values and the targeted values. Table 8 on the other hand shows the results when using the entropy identity attributes metrics values. Fig. 1 shows the graph for the training, validation and testing values. The target is met after seven epochs. The performance, training state and regression graphs were used in the training and testing of the network until target values were met.

Table 7: Input values, target values and the corresponding output and error values generated by an ANN information fusion technique using entropy identity attributes metrics values.

P	B	D	Target	Output	Error
0.0000	0.0000	0.0000	0.0000	0.0000	-0.0000
1.5434	0.0000	0.0000	0.1000	0.1000	0.0000
0.0000	8.9744	0.0000	0.4000	0.4000	-0.0000
1.5434	8.9744	0.0000	0.6000	0.6000	0.0000
0.0000	0.0000	3.5035	0.3000	0.3000	0.0000
1.5434	0.0000	3.5035	0.5000	0.5000	-0.0000
0.0000	8.9744	3.5035	0.7000	0.8372	-0.1372
1.5434	8.9744	3.5035	1.0000	0.9228	0.0772

Table 8: Input values, target values and the corresponding output and error values generated by an artificial neural network information fusion technique using term weight identity attributes metrics values.

P	B	D	Target	Output	Error
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0533	0.0000	0.0000	0.1000	0.1000	0.0000
0.0000	0.5827	0.0000	0.4000	0.4829	-0.0829
0.0533	0.5827	0.0000	0.6000	0.6000	0.0000
0.0000	0.0000	0.3787	0.3000	0.3000	0.0000
0.0533	0.0000	0.3787	0.5000	0.5000	0.0000
0.0000	0.5827	0.3787	0.7000	0.7880	-0.0880
0.0533	0.5827	0.3787	1.0000	1.0000	0.0000

Using the term weight identity attributes metrics values, as shown in Table 7, the system gives -0.0000 correct to four decimal places when none of the submitted identity attribute is correct. On the other hand when all the submitted identity attributes are correct the ANN information fusion system gives 0.9228 to four decimal places. The rest of the combinations are spread out in between this range. When using entropy identity attributes metrics values, the system gives 0.0000 when none of the submitted identity attributes is correct and 1.0000 correct to four decimal places when all the submitted identity attributed are correct. Table 8 shows the rest of combinations. In each of the two cases above, the generated results for the eight combinations are all between zero and one. Using the results in Figure 2, a multifactor authentication system is then implemented by setting the threshold value which the user must meet in order to be authenticated.

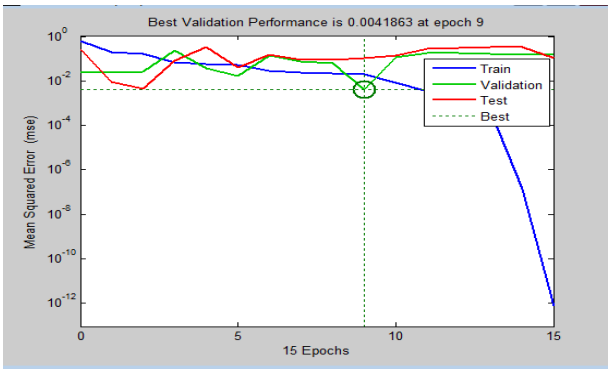

Fig. 1. Performance graph showing the training, validation, test and best performance after nine epochs for the artificial neural network when using term weight metrics values to implement an information fusion technique.

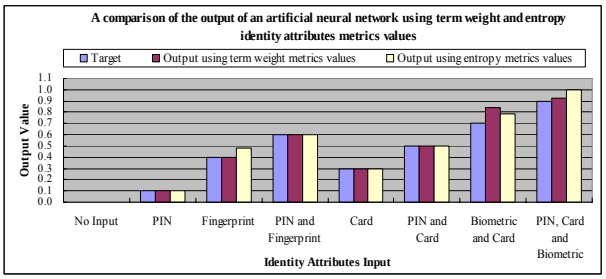

Fig. 2. A comparison of the output values generated by an artificial neural network information fusion technique using term weight and entropy identity attributes metrics values.

4.2.2 Using Fuzzy Inference System

In this implementation, the information fusion technique has three inputs, seven rules and one output. When all the three inputs using term weight identity attributes metrics values are correct, the FIS information fusion technique system gives the maximum output of 3.45. However, when none of the submitted inputs is collect, the system gives an output of 0.0455. The rest of the combination is spread out in between this range as shown in Table 9. Table 10 shows the eight possible inputs and the corresponding output values using entropy identity attributes metrics values. The system gives an output of 0.046 when none of the inputs is correct and the output of 2.670 when all the entropy identity attributes input metrics values are correct. Fig. 3 shows the surface viewer graph of the pseudo metrics, the device metrics inputs and the corresponding output values. When using their term weight metrics, the output values are in between 0.045 and 3.450 while when using the entropy identity attributes metrics values, the output values are in between 0.045 and 2.67 as shown in Table 9 and Table 10 respectively.

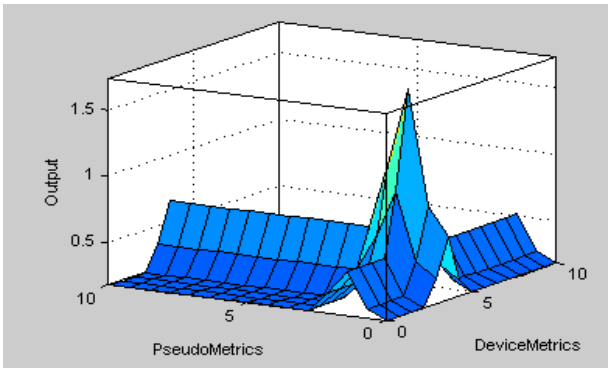

Fig. 3. Surface viewer for the pseudo metrics input, the device metrics input and the corresponding output values.

Table 9: Input values and the corresponding output values generated by the FIS information fusion technique using term weight identity attributes metrics values.

#	Pseudo	Biometrics	Device Metrics	Output
1	0.000	0.000	0.000	0.046
2	1.500	5.000	0.000	0.291
3	0.000	8.970	0.000	0.528
4	0.000	0.000	3.500	0.491
5	1.500	0.000	3.500	1.210
6	1.500	8.970	0.000	1.260
7	0.000	8.970	3.500	1.460
8	1.500	8.970	3.500	3.450

Table 10: Input values and the corresponding output values generated by the FIS information fusion technique using entropy identity attributes metrics values.

#	Pseudo	Biometrics	Device Metrics	Output
1	0.0000	0.0000	0.0000	0.046
2	0.0533	0.0000	0.0000	0.291
3	0.0000	0.5827	0.0000	0.359
4	0.0533	0.5827	0.0000	0.491
5	0.0000	0.0000	0.3787	1.210
6	0.0533	0.0000	0.3787	0.918
7	0.0000	0.5827	0.3787	1.120
8	0.0533	0.5827	0.3787	2.670

Fig. 4 shows the comparison of the results generated by a fuzzy inference system using the two sets of the metrics values. In general, the output values from the FIS information fusion technique gives a wider range as compared to those generated by ANN. Using this range, a multifactor authentication system is implemented by

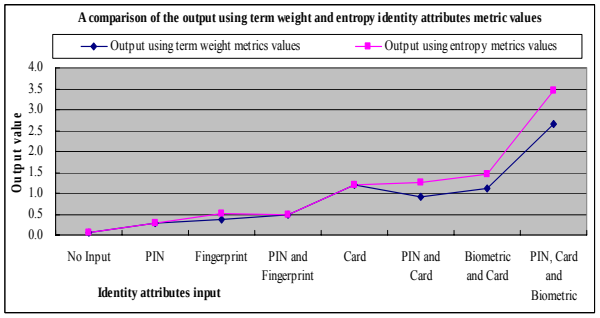

Fig. 4. A comparison of the output values generated by the FIS information fusion technique using term weight and entropy identity attributes metrics values.

setting a threshold value which the user must meet in order to be authenticated.

3.2.3 *Using Sugeno Artificial Neural - Fuzzy Inference System (ANFIS)*

Using the hybrid implementation of an ANN and the FIS called the ANFIS, Table 11 shows the eight possible inputs, target values and the corresponding output values using term weight metrics values. When none of the submitted identity attributes is correct, the system gives an output of 0.0009 while when all the submitted identity attributes are correct the system gives an output of 1.0000 correct to four decimal places. The rest of the combination is spread out in between this range as shown in Table 11.

Table 11: Input values, target values and the corresponding output values generated by an adaptive neural-fuzzy inference system information fusion technique using term weight identity attributes metrics values.

#	PIN	Fingerprint	Card-ID	Target	Output
1	0.0000	0.0000	0.0000	0.0000	0.0009
2	1.5430	0.0000	0.0000	0.1000	0.0995
3	0.0000	8.9740	0.0000	0.4000	0.4000
4	1.5430	8.9740	0.0000	0.6000	0.6010
5	0.0000	0.0000	3.5030	0.3000	0.2990
6	1.5430	0.0000	3.5030	0.5000	0.5010
7	0.0000	8.9740	3.5030	0.7000	0.7000
8	1.5430	8.9740	3.5030	0.9000	1.0000

However, when using the entropy metrics values, the information fusion technique gives an output of -0.0006 when none of the submitted identity attributes is correct and an output of 0.9890 when all the submitted identity attributes are correct. The rest of the combination is spread in between this range as shown in Table 12. Fig. 5 shows an ANFIS generated surface viewer of the biometrics inputs, device metrics inputs and their corresponding outputs. The ANFIS also generated the values of the fuzzification function variable varies represented by (5) and (6), input range of the function and display range of the function for each of the three inputs. Table 13 shows these values when term weight metrics values are used while Table 14 show the generated values when entropy metrics values are used to implement an information fusion technique.

Table 12: Input values, target values and the corresponding output and error values generated by an adaptive neural-fuzzy inference system information fusion technique using entropy identity attributes metrics values.

#	PIN	Fingerprint	Card-ID	Target	Output
1	0.0000	0.0000	0.0000	0.0000	-0.0006
2	0.0533	0.0000	0.0000	0.1000	0.1030
3	0.0000	0.5827	0.0000	0.4000	0.4010
4	0.0533	0.5827	0.0000	0.6000	0.5980
5	0.0000	0.0000	0.3787	0.3000	0.2920
6	0.0533	0.0000	0.3787	0.5000	0.5080
7	0.0000	0.5827	0.3787	0.7000	0.7100
8	0.0533	0.5827	0.3787	0.9000	0.9890

The ANFIS information fusion technique also generated the values of the linguistic output variable values during the training process. Table 15 shows the linguistic variables and the corresponding values by using term weight and entropy identity attributes metrics values.

Table 13: Adaptive neural-fuzzy inference system fuzzification functions and the generated variable values using term weight identity attributes metrics values.

	Pseudo Metrics	Biometrics	Device Metrics
Input Range	0.000 - 1.543	0.000 - 8.974	0.000 - 3.503
Display Range	0.000 - 1.543	0.000 - 8.974	0.000 - 3.503
Function	trimf	sigmf	trimf
Variable Values of Function	a=-0.9755, b=1.066,c=3.486	a=1.56, b=3.687	a=-1.12, b=2.478, c=5.024

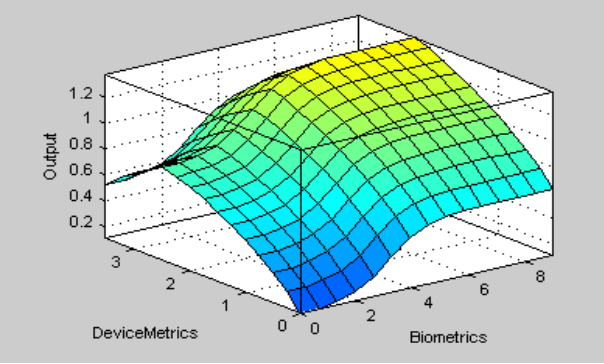

Fig. 5. Adaptive neural-fuzzy inference system surface viewer for the biometrics input, device metric inputs and the corresponding output values.

Table 14: Adaptive neural-fuzzy inference system fuzzification functions and the generated variable values using entropy identity attributes metrics values

	Pseudo Metrics	Biometrics	Device Metrics
Input Range	0.0000 - 0.05328	0.0000 - 0.5827	0.0000 - 0.3787
Display Range	0.0000 - 0.05329	0.0000 - 0.5827	0.0000 - 0.3787
Function	trimf	sigmf	Trimf
Variable Values of Function	a=-0.06052, b=1.126,c=3.494	a=0.6894, b=3.074	a=-0.9058, b=2.472, c=4.999

Table 15: Adaptive neural-fuzzy inference system information fusion technique output linguistic variables and the corresponding generated values using term weight and entropy identity attributes metrics values.

Output Variable	Term Weight metrics	Entropy Metrics
VeryLow	-0.9287	-8.867
Low	-0.6331	-1.256
Medium	-0.3327	-1.755
Average	4.2910	20.140
High	0.7150	39.600
Very High	0.2563	16.030
All	7.1540	16.080

The membership function variable values in Table 13 and Table 14, the output constant values for the output variables in Table 15 enable the system to generate the values shown in Table 11 and Table 12. Fig. 6 shows the comparison of the ANFIS output results when using either term weight or entropy metrics values as the input. In both cases the system was able to give the output values in between zero and one. The values are very close to those generated by the ANN information fusion technique. The ANFIS results shown in Figure 19 are then used to implement a multifactor authentication system by setting a threshold value which the user must met to be authenticated.

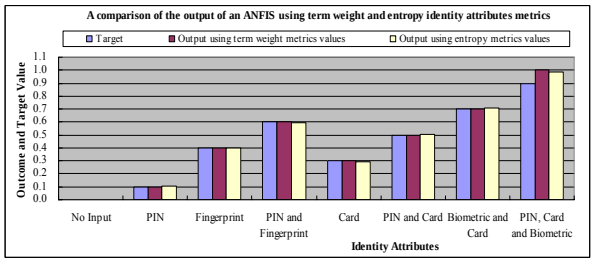

Fig. 6. A comparison of the output values generated by an adaptive neural-fuzzy inference system information fusion technique using term weight and entropy identity attributes metrics values.

5. Discussion and Conclusion

In this paper, a multifactor information fusion technique was successfully implemented by using the identity attributes metrics values composed by term weight from text mining techniques and entropy from Shannon's information theory. Fig. 7 shows the output results of an information fusion technique for an ANN, FIS and ANFIS when using term weight metrics values while Fig. 8 shows similar results when using entropy metrics values. In all the three cases, when none of the identity attributes submitted was correct, the system gave an output of zero correct to two decimal places. For both the ANN and ANFIS, the maximum output was one correct to two decimal places while the FIS gave a maximum output of 2.67 and 3.45 for the entropy and term weight metrics respectively. With this range, a multifactor authentication system is then implemented using the three groupings of the identity attributes by setting a threshold value which the user must meet in order to be authenticated. For example using the results in Fig. 7 and setting a threshold value of the system at 0.5, it means that for anyone to be successfully authenticated, they need a combination of any two of the three identity attributes. If a user has a correct biometrics but using the wrong card or PIN number, they will not be allowed access to the system. If however the threshold was set at 0.75, and the implementation is done using an ANN or an ANFIS, then the user will need a combination of the correct device metrics (unique card number) and a biometrics

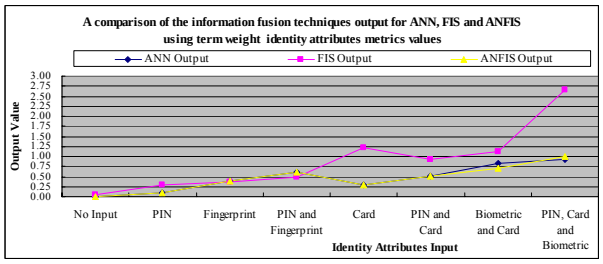

Fig. 7. A comparison of the information fusion techniques output for an artificial neural network, fuzzy inference system and adaptive neural-fuzzy inference system using entropy identity attributes metrics values.

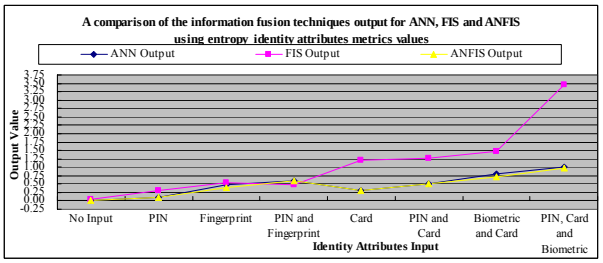

Fig. 8. A comparison of the information fusion techniques output for artificial neural network, fuzzy inference system and adaptive neural-fuzzy inference system using entropy identity attributes metrics values.

(fingerprint) or a combination of all the three attributes. Therefore depending on the security level of the system, the three groups of the identity attributes can be combined to achieve information fusion in various ways. Each category can also take different types of identity attributes. For example, in our example we used the fingerprint as an example of the biometrics, but face recognition, iris scan or any other biometrics can be used as well. Also instead of the unique card number and the PIN number combination, the password, username and IP address combination can be used in addition to the biometrics.

In this paper, we demonstrated that it is possible to combine the pseudo metrics, biometrics and physical metrics (device metrics) as the *score level* for effective authentication of a user through a technique of information fusion. An artificial neural network, fuzzy inference system and an adaptive neural-fuzzy inference system were successfully used to implement an information fusion technique using the identity attributes metrics values composed by term weight and entropy. This is the major contribution in this paper.

In the last decade, there has been a rise in the number of services going online each year, however this has also seen a corresponding increase in the number of cases related to cyber-crime (See Refs. 3-6). A multifactor authentication system such one proposed in this paper will most likely help to curb the cases of cyber-crime which is on the increase and costing the global industry and governments billions of dollars each year.

Future work will look into other methods of implementing information fusion such as game programming and evolution computing.

References

1. S. Mike, Unify and Simplify: Re-thinking Identity Management, *Network Security*. **2006**(7) (2006), 11-14. doi: dx.doi.org/10.1016/S1353-4858(06)70411-1
2. R. Dhamija, and L. Dusseault, The Seven Flaws of Identity Management: Usability and Security Challenges, *Security & Privacy, IEEE*. **6**(2) (2008) 24-29. doi: dx.doi.org/10.1109/MSP.2008.49.
3. S. Clare, Digital identity - The Legal Person? *Computer Law & Security Review*, *Elsevier*. **25**(3) (2009) 227-236. doi: dx.doi.org/10.1016/j.clsr.2009.03.009.
4. P. Geraint, The benefits and drawbacks of using electronic identities, *Information Security Technical Report, Elsevier*. **13(**2) (2008) 95-103. doi: dx.doi.org/10.1016/j.istr.2008.07.002.
5. G. Goth, Identity management, access specs are rolling along, *Internet Computing, IEEE*. **l9**(1) (2005), 9- 11. doi; dx.doi.org/10.1109/MIC.2005.16.
6. B. Geoff, The use of hardware tokens for identity management, *Information Security Technical Report* **9**(1) (2004) 22-25. doi: dx.doi.org/10.1016/S1363-4127(04)00012-3.
7. H. Marit, P. Andreas and S. Sandra, Identity management throughout one's whole life, *Information Security Technical Report, Elsevier*. **13**(2) (2008) 83-94. doi: dx.doi.org/10.1016/j.istr.2008.06.003.
8. EconomyWatch, List of Commercial Banks; Available (January 2011): http://www.economywatch.com/banks/commercial-banks/
9. Wikipedia, The free encyclopedia, Civil service; Available (January 2011): http://en.wikipedia.org/wiki/Civil_service
10. Top University, QS World University Rankings; Available (January 2011): http://www.topuniversities.com/university-rankings/world-university-rankings, 2011
11. Thomson Reuters, 100 Top Hospitals; Available (January 2011): http://www.100tophospitals.com/
12. Google Double Click Ad Planner; Available (April 2010): http://www.google.com/adplanner/static/top1000/#
13. Wikipedia, The free encyclopedia, List of social networking websites; Available (January 2011): http://en.wikipedia.org/wiki/List_of_social_networking_websites
14. L. Anthony. AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom, *in Professional Communication Conference Proceedings*, (2005), pp. 729. doi:10.1109/IPCC.2005.1494244
15. C. Greaves, ConcApp Version 4 Concordancer, Edict Virtual Language Centre, Available (November 2010): http://www.edict.com.hk/PUB/concapp/.
16. TextSTAT Corpus, Available (November 2010) on: http://www.edict.com.hk/PUB/concapp/

17. R. Togneri, and C. J. S. DeSilva, *Fundamentals of Information Theory and Coding Design*, (Chapman & Hall/ CRC Press, FL, 2005).
18. M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd edn. (China Machine Press, 2005)
19. W. Chih-Hung, Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks, *Expert Systems with Applications*. **36**(3) (2009) 4321-4330. doi:10.1016/j.eswa.2008.03.002.
20. M. Fazle Azeem, M. Hanmandlu, N. Ahmad, Structure identification of generalized adaptive neuro-fuzzy inference systems, *Fuzzy Systems, IEEE Transactions*. 11(5) (3003) 666–681. doi:10.1109/TFUZZ.2003.817857.
21. J. I. Agbinya, R. Islam and C. Kwok, Development of Digital Environment Identity (DEITY) System for Online Access, in *Broadband Communications, Information Technology & Biomedical Applications*, *Third Int. Conf.,* (Australia 2008), pp. 23-26, doi: 10.1109/BROADCOM.2008.52.
22. M. He, et al, Performance Evaluation of Score Level Fusion in Multimodal Biometric Systems, *Pattern Recognition*. **43**(5) (2010) 1789-1800. doi: 10.1016/j.patcog.2009.11.018.
23. L. Nanni, A. Lumini, S. Brahnam, Likelihood Ratio Based Features for a Trained Biometric Score Fusion, *Expert Systems with Applications*. **38**(1), (2011) 58-63. doi: 10.1016/j.eswa.2010.06.006.
24. F. Chong, Identity and Access Management, *Microsoft Architect Journey*. (2004).
25. European Technology Assessment Group, RFID and identity management in everyday life, Available online (October 2010) at: http://www.europarl.europa.eu/stoa/publications/studies/stoa182_en.pdf
26. The National Electronic Commerce Coordinating Council (NECCC), Identity Management, *Presented at the NECCC Annual Conference*, (New York, 2002).
27. M. Hansen, A. Schwartz and A. Cooper, Privacy and Identity Management, *Security & Privacy, IEEE.* **6** (2008). doi: 10.1109/MSP.2008.41.
28. M. Barisch, Modelling the Impact of Virtual Identities on Communication Infrastructures, Conference on Computer and Communications Security, *in Proc. of the 5th ACM workshop, Digital identity management*, (Chicago, Illinois, 2009) pp. 45–52. doi: http://doi.acm.org/10.1145/1655028.1655040.
29. G. Hidehito, User-Centric Identity Governance Across Domain Boundaries, Conference on Computer and Communications Security, *in Proc. of the 5th ACM workshop, Digital identity management*, (Chicago, Illinois, 2009) pp. 35–44. doi: http://doi.acm.org/10.1145/1655028.1655038.
30. V. Avram, Defining metrics to automate the quantitative analysis of textual information within a web page, in Int. Conf. of Application of Information and Communication Technologies, (AICT 2009), pp.1-5. doi: http://dx.doi.org/10.1109/ICAICT.2009.5372575.
31. C. Kaddi, E. D. Oden, C. F. Quo, and M. D. Wang, Exploration of Quantitative Scoring Metrics to Compare Systems Biology Modeling Approaches, *in Int. Conf. of Engineering in Medicine and Biology Society (*EMBS 2007) pp. 1121-1124. doi: http://dx.doi.org/10.1109/IEMBS.2007.4352493.
32. S. Liao, H. Ho, F. Yang, Ontology-based data mining approach implemented on exploring product and brand spectrum, *Expert Systems with Applications*, **36**(9) (2009) 11730-11744.
33. L. Mazlack and S. Coppock, Granulating data on non-scalar attribute values, *in Proc. of the 2002 IEEE Int. Conf. on Fuzzy Systems.* (2002), pp. 944-949. doi: http://dx.doi.org/10.1109/FUZZ.2002.1006631.
34. J. Phiri and T. J. Zhao, Identity attributes quantitative analysis and the development of a metrics model using text mining techniques and information theory, i*n Int. conf. on Information Theory and information security (ICITIS2010)* (Beijing, 2010), pp. 390-393. doi: 10.1109/ICITIS.2010.5689588.
35. D. Applebaum, *Probability and Information: An Integrated Approach* (Cambridge University Press, Cambridge, 1996).
36. X. Zhang, Y. Wang, H. Yu, Neural Network Based Algorithm and Simulation of Information Fusion in the Coal Mine, *Journal of China University of Mining and Technology*, **17**(4) (2007) 595-598. doi: http://dx.doi.org/10.1016/S1006-1266(07)60153-9.
37. B. Croft, D. Metzler and D. Strohman, *Search Engines: Information Retrieval in Practise* (Addison Wesley, 2010)
38. J. Zhang, Improved on-line process fault diagnosis through information fusion in multiple neural networks, *Computers & Chemical Engineering*, **30**(3), (2006), 558-571. doi: http://dx.doi.org/10.1016/j.compchemeng.2005.11.002.
39. L. Hui, B. Yuequan, and O. Jinping, Structural damage identification based on integration of information fusion and shannon entropy, *Mechanical Systems and Signal Processing*, **22** (6) (2008) 1427-1440.
40. B. Kitchenham, What's up with software metrics? - A preliminary mapping study, *Elsevier, Journal of Systems and Software* **83**(1) (2010) 37-51. doi: http://dx.doi.org/10.1016/j.jss.2009.06.041.
41. J. Phiri, J. I. Agbinya, Modelling and Information Fusion in Digital Identity Management Systems, *in Int. Conf. on Systems and Int. Conf. on Mobile Communications and Learning Technologies*, (Mauritius, 2006), pp. 181-186. doi: 10.1109/ICNICONSMCL.2006.152
42. Fuzzy logic toolbox manual; Available (November 2010): http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/rn.pdf