I.J. Information Engineering and Electronic Business, 2012, 2, 29-35
Published Online April 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2012.02.05

Enhanced Password Based Security System Based on User Behavior using Neural Networks

Preet Inder Singh
Department of CSE/IT, Lovely Professional University (Punjab), Phagwara
Email: preetindermail@gmail.com

Gour Sundar Mitra Thakur
Department of CSE/IT, Lovely Professional University (Punjab), Phagwara
Email: cse.gsmt@gmail.com

***Abstract*—There are multiple numbers of security systems are available to protect your computer/resources. Among them, password based systems are the most commonly used system due to its simplicity, applicability and cost effectiveness But these types of systems have higher sensitivity to cyber-attack. Most of the advanced methods for authentication based on password security encrypt the contents of password before storing or transmitting in the physical domain. But all conventional encryption methods are having its own limitations, generally either in terms of complexity or in terms of efficiency.**

In this paper an enhanced password based security system has been proposed based on user typing behavior, which will attempt to identify authenticity of any user failing to login in first few attempts by analyzing the basic user behaviors/activities and finally training them through neural network and classifying them as genuine or intruder.

***Index Terms*—Artificial neural networks, Keystroke Dynamics, intrusion detection, Security & User Authentication.**

1. Introduction

It is often seen that to gain some personal benefit or attention or to harm someone some people always try to break cyber securities. The first step in preventing unauthorized access is to assure *user authentication.* User authentication is the process of verifying claimed identity. The authentication is accomplished by matching some short-form indicator of identity, such as a shared secret that has been pre-arranged during enrollment or registration for authorized users [1]. This is done for the purpose of performing trusted communications between parties for computing applications.

The well-known ID/password (static authentication) is far the most used authentication method. It is widely used despite its obvious lack of security. This fact is due to the ease of implementation of this solution, and to the instantaneous recognition of that system by the users that facilitates its deployment and acceptance. Increasing the password strength is a solution to avoid dictionary attacks or to make brute force attacks infeasible [2]. It is generally accepted that the length of the password determines the security it provides, however, it is not exactly true: the strength of the password is rather related to its entropy. For example, User that chooses a password of 7 characters is said to provide between 16 and 28 bits of entropy.

The conventional security system can be shown in figure - 1 given below.

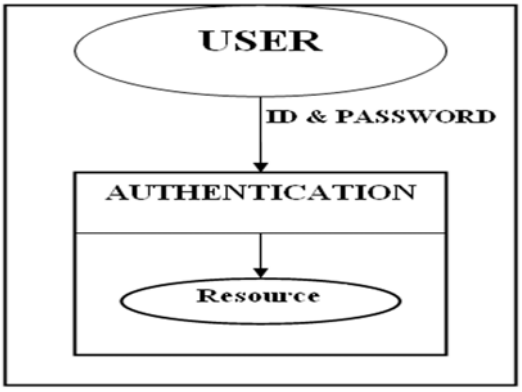


Figure1: Conventional Security System

Due to the deficiencies in traditional password-based access methods/Security systems, the new security system comes into existence which provides higher level of security is the Keystroke biometrics,

which seeks to identify individuals by their typing characteristics [3].

Conventionally, user authentication is categorized into three classes [4]:

- Knowledge - based,
- Object or Token - based,
- Biometric - based.

The knowledge-based authentication is based on something one knows and is characterized by secrecy. The examples of knowledge-based authenticators are commonly known passwords and PIN codes. The object-based authentication relies on something one has and is characterized by possession.

Behavioral characteristics are related to what a person does, or how the person uses the body. Voiceprint, traditional keys to the doors can be assigned to the object-based category. Usually the token-based approach is combined with the knowledge-based approach. An example of this combination is a bankcard with PIN code. In knowledge-based and object-based approaches, passwords and tokens can be forgotten, lost or stolen. There are also usability limitations associated with them. For instance, managing multiple passwords / PINs, and memorizing and recalling strong passwords are not easy tasks. Biometric-based person recognition overcomes the above mentioned difficulties of knowledge-based and object based approaches. The following figure - 2 shows the different classification of user authentication methods.

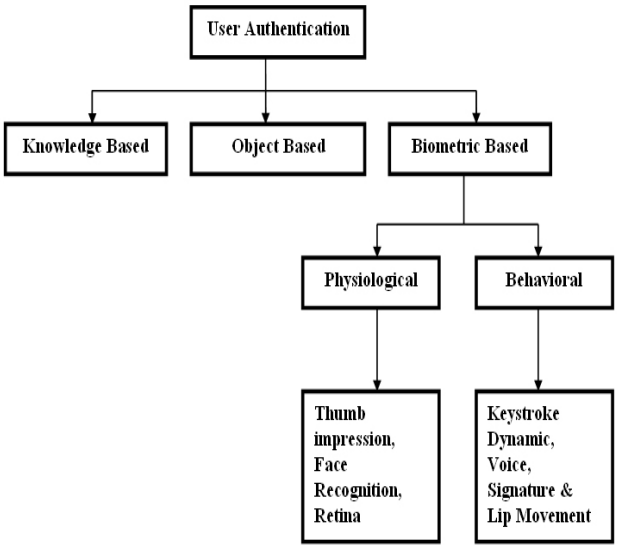


Figure2: Classification of User Authentication approaches

Biometrics technologies are gaining popularity due to the reason that when used in conjunction with traditional methods of authentication they provide an extra level of security. Biometrics involves something a person is or does. These types of characteristics can be approximately divided into physiological and behavioral types [4]. Biometric technologies are defined as automated methods of verifying or recognizing the identity of a living person based on physiological or behavioral characteristics [5]. Physiological characteristics refer to what the person is, or, in other words, they measure physical parameters of a certain part of the body. Some examples are Fingerprints, Hand Geometry, Vein Checking, Iris Scanning, Retinal Scanning, Facial Recognition, recognition, Signature Recognition, Mouse Dynamics and keystroke dynamics, are good examples of this group.

Keystroke dynamics is considered as a strong behavioral Biometric based Authentication system [6]. It is a process of analyzing the way a user types at a terminal by monitoring the keyboard in order to identify the users based on habitual typing rhythm patterns. Moreover, unlike other biometric systems, which may be expensive to implement, keystroke dynamics is almost free as the only hardware required is the keyboard.

A person's identity is checked in the verification case. If the user behavior is matched with the existing parameters then the login is successful otherwise the login is unsuccessful. If a person tries number of times then after three unsuccessful attempts the system is automatically locked [7].

The ultimate level of security can be achieved with the help of keystroke dynamic only if when combine all the parameters/features like key code, two keystroke latencies, three keystroke latencies and key duration. This approach can be used to improve the usual login-password authentication when the password is no more a secret [9].

The remainder of this paper is organized as follows: Section 2 gives an enhanced method of password based security system. Section 3 describes the experimental results. Section 4 presents the advantages of the proposed system. Conclusion is given in the final section.

2. Proposed security system

In this proposed model the different behaviors in entering passwords are recoded as per the above mention criteria. A sample dataset will be prepared by recoding some attempts by authentic as well as non-authentic user in any simple password based user authentication system because for known regularly-typed strings (e.g., username and password), such features are quite consistent. Then those data are to be trained using Feed Forward Neural Network to classify authentic and non-authentic user. After that the trained network will be used to identify authentic or non-authentic user based on his runtime behavior in time of entering password.

2.1 Password based behavioral characteristics

In the behavior keystroke dynamics, behavior of the user password is checked by the system at the real time i.e while entering the password. It includes the multiple parameters to check the behavior of the user like total time to enter password, average time in all the attempts, latency between two characters, latency between three characters, deviation from the original/last attempt password, number of times Shift/Caps used to enter the uppercase letter, time between the two passwords attempts, if in first time behavior of user/person does not match with the existing behavior/rhythm etc. It is noted that the behavior of the users are different from one another while entering the password in real time.

2.2 Neural network

Artificial Neural Network is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems [7].

MLP neural network and RBF networks have become the most widely used network architectures in pattern classification problems. The general difference between the two neural networks is that MLP is a more distributed approach compared to RBF, which only responds to a limited section input space [8].

2.3 Neural network's features suitable for security systems

Neural network has the tremendous property of learning from the environment which is suitable for security systems. Learning can be fall into two categories i.e Supervised learning and Unsupervised learning. Neural network can use a set of observations to solve the task in an optimal sense. It is the branch of artificial intelligence (AI).

2.4 Training and testing with feed-forward neural network

A multilayer perceptron in Weka is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. Feed Forward architecture is used in this paper as shown in figure - 3. This neural network is formed in three layers, called the input layer, hidden layer, and output layer. Each layer consists of one or more nodes, represented in this diagram by the small circles. The lines between the nodes indicate the flow of information from one node to the next. In this particular type of neural network, the information flows only from the input to the output. The nodes of the input layer are passive, meaning they do not modify the data. They receive a single value on their input, and duplicate the value to their multiple outputs. In comparison, the nodes of the hidden and output layer are active. The values entering a hidden node are multiplied by weights; a set of predetermined number is stored in the program. The weighted inputs are then added to produce a single number (output).

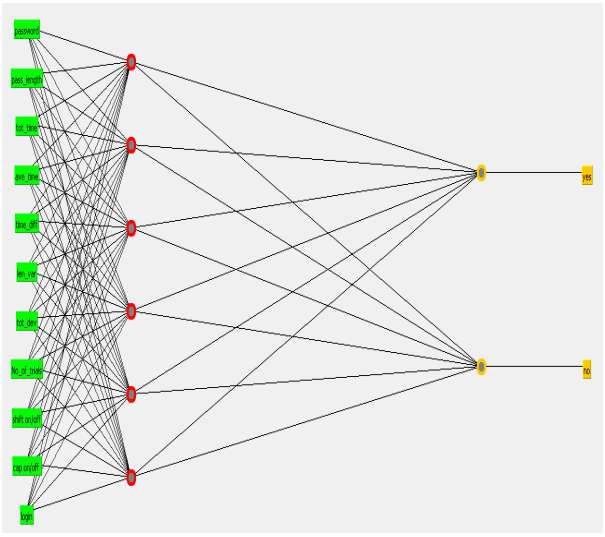


Figure3: Feed forward Architecture with two outputs

2.5 Transfer function: Sigmoid function

Sigmoid function, mathematically described by the equation $S(x) = 1/(1+e^{-\lambda x})$, having several advantages like, (1) Soft limiter, i.e. having sensitivity w.r.t variation in input. (2) Mathematical model of biological neuron, firing phenomenon the characteristic is appearing like sigma function. It derivative is easily available which is required in learning process, $S'(x) = s(x)[1 - s(x)]$ [7].

2.6 Authentication process

At the real-time user enter his/her ID and Password to enter into their account and to access the resources as shown in figure 5. Login information may have various types of information like strings, characters (Block/Small letters) or any alpha numeric data. This information is then matched with the saved login behavior of the user. If this information is matched with the saved login behavior then the login is successful otherwise the login is unsuccessful. This includes the various logical parameters like Number of trails, Length of password, Time taken to insert the password, Time taken to reenter the password if in first attempt password is wrong, Deviations from the first attempt password, If the password is in the capital letter; whether Shift is used or not, If the password is in the capital letter; whether Caps Lock is used or not.

These parameters are forming a hierarchical structure to enhance the speed of computation and saving the unnecessary involvement of other inner layer modules associated with this security system. These layers logically define the types of activity with knowing the contents of password & behavior of entering the password. Neural network based authentication is the innermost layer, which is taking care of contents available with the password as shown in Figure - 4.

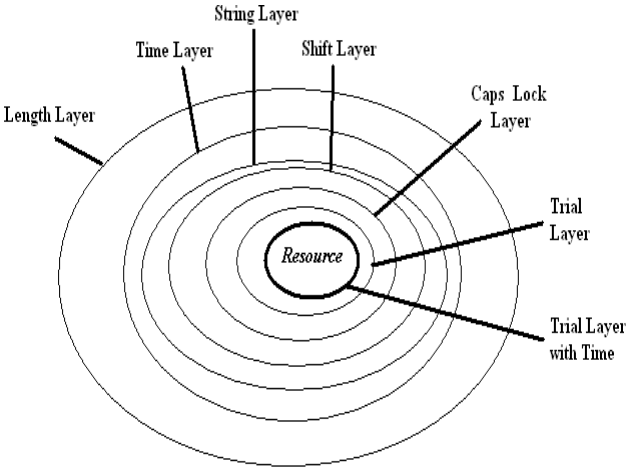


Figure 4: Number of Layers is used to protect the resources from unauthorized users.

2.7 Hierarchical structure of security

In this method a multilayer, multi-parametric security system has developed. In most of the public oriented services there are service providers along with service users. Each party wanted to allocated a security means to protect their resource. Hence for each case a separate password provision can be allocated. To provide the intruder detection several parameters can taken as measuring consideration likes,

i. Number of trails.
ii. Length of password.
iii. Time taken to insert the password.
iv. Time taken to reenter the password if in first attempt password is wrong.
v. Deviations from the first attempt password.
vi. If the password is in the capital letter, whether Shift is used or not.
vii. If the password is in the capital letter, whether Caps Lock is used or not.

These parameters can surely help to define the activity as normal or intrusion users. When intrusion declared, to protect the resource, security environment not allow entering the password further. The other benefit of this facility is, when the right person will try to access the resource, system will not permit to enter the password hence an auto- information mechanism about intrusion available to right user**.**

3. Experimental results

3.1 Tools of data collection

Data is collected in real time using *Visual Basic 6.0* as a front end and *Microsoft Access 2003.* This includes the various logical parameters as received from the intruders intending to breach our security set-up.

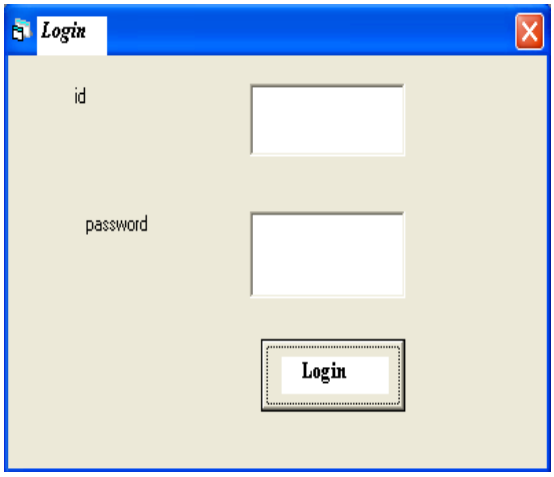
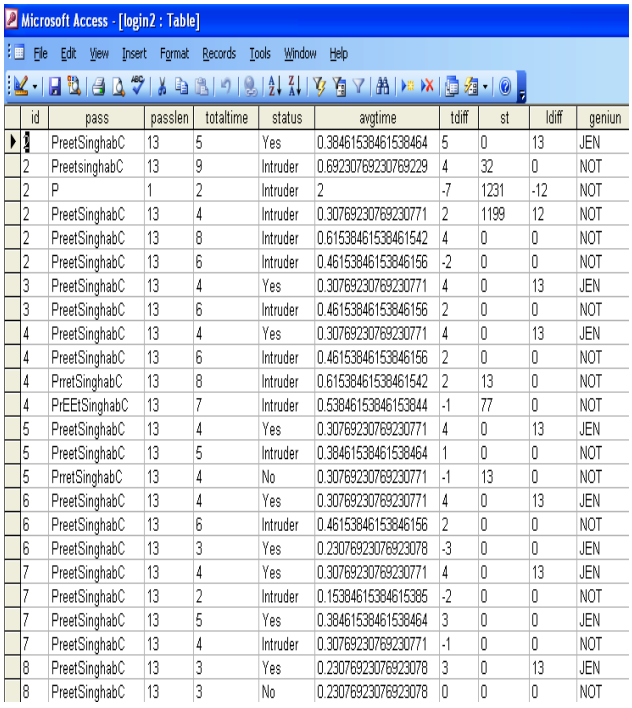


Figure 5: Sample Login Window


Microsoft Access - [login2 : Table]

File Edit View Insert Format Records Tools Window Help

id	pass	passlen	totaltime	status	avgtime	tdiff	st	ldiff	geniun
1	PreetSinghabC	13	5	Yes	0.38461538461538464	5	0	13	JEN
2	PreetsinghabC	13	9	Intruder	0.69230769230769229	4	32	0	NOT
2	P	1	2	Intruder	2	-7	1231	-12	NOT
2	PreetSinghabC	13	4	Intruder	0.30769230769230771	2	1199	12	NOT
2	PreetSinghabC	13	8	Intruder	0.61538461538461542	4	0	0	NOT
2	PreetSinghabC	13	6	Intruder	0.46153846153846156	-2	0	0	NOT
3	PreetSinghabC	13	4	Yes	0.30769230769230771	4	0	13	JEN
3	PreetSinghabC	13	6	Intruder	0.46153846153846156	2	0	0	NOT
4	PreetSinghabC	13	4	Yes	0.30769230769230771	4	0	13	JEN
4	PreetSinghabC	13	6	Intruder	0.46153846153846156	2	0	0	NOT
4	PrretSinghabC	13	8	Intruder	0.61538461538461542	2	13	0	NOT
4	PrEEtSinghabC	13	7	Intruder	0.53846153846153844	-1	77	0	NOT
5	PreetSinghabC	13	4	Yes	0.30769230769230771	4	0	13	JEN
5	PreetSinghabC	13	5	Intruder	0.38461538461538464	1	0	0	NOT
5	PrretSinghabC	13	4	No	0.30769230769230771	-1	13	0	NOT
6	PreetSinghabC	13	4	Yes	0.30769230769230771	4	0	13	JEN
6	PreetSinghabC	13	6	Intruder	0.46153846153846156	2	0	0	NOT
6	PreetSinghabC	13	3	Yes	0.23076923076923078	-3	0	0	JEN
7	PreetSinghabC	13	4	Yes	0.30769230769230771	4	0	13	JEN
7	PreetSinghabC	13	2	Intruder	0.15384615384615385	-2	0	0	NOT
7	PreetSinghabC	13	5	Yes	0.38461538461538464	3	0	0	JEN
7	PreetSinghabC	13	4	Intruder	0.30769230769230771	-1	0	0	NOT
8	PreetSinghabC	13	3	Yes	0.23076923076923078	3	0	13	JEN
8	PreetSinghabC	13	3	No	0.23076923076923078	0	0	0	NOT

Figure 6: Sample of Data Stored regarding Passwords

The various parameters are shown above in Microsoft Access figure-6 are

1. Login ID *(id.)*
2. Password *(pass.)*
3. Password Length *(passlen).*
4. Total time to enter the password *(totaltime).*
5. Status means login is done or not *(status.)*
6. Average time to enter the whole password *(avgtime).*
7. Time difference between the two passwords if in the first attempt login is not done/successful *(tdiff).*
8. Deviation from the current/actual password *(st).*
9. Length difference from the current/old attempt password *(ldiff).*
10. Whether a user is genuine or not *(geniun).*

login : Table

id	pass	passlen	totaltime	shift
1	Preet	5	1	100000
2	PreetSinghabC	13	4	0000010000000010
3	PreetSinghabC	13	4	0000010000000010
4	PreetSinghabC	13	3	0000010000000010
5	PreetSinghabC	13	3	0000010010000000010
6	PreetSinghabC	13	3	0000010000000010
7	PreetSinghabC	13	5	0000010000000010
8	PreetSinghabC	13	3	0000010000000010
9	PreetSinghabC	13	6	0000010000000010

Figure 7: Recognize the Behavior of user in first attempt

In the figure 7, in column shift, it recognizes/saves the typing behavior (*Shift* and/or *Caps Lock* used to enter the *Block letters* in *Password*) of the user when the first time authorized user login to his/her account and matches it with when the user attempts to login again to his/her account. If the typing behavior is matched with the authorized user as well as other logical parameters as discussed above are also matched then login is successful otherwise the login is unsuccessful.

3.2 Training using neural networks

With the help of *Weka* Simulator (Version 3.7) using the above logical parameters as received from the intruder intending to breach our security system, we train the neural network with *Multilayer Perceptron* to find whether the user is genuine or not based upon user behavior at real-time and found that the accuracy is 100% as shown in table-1.

Table1:Defining the Network Parameters

Parameter	**Values**
Number of Training Data	120
Number of Testing Data	20
Number of Hidden Layers	2
Learning Rate	0.3
Momentum	0.2
Validation Threshold	20
Total no of Epochs	500
Error Per Epoch	0.0000375
Accuracy	100 %

The above table -1 shows maximum accuracy obtained during training of multilayer perceptron with 10 cross validation. For the cross validation purpose we divide 70% data for training, 15% data for validation and 15% data for testing of networks.

The following are the snapshots of data training (figure -8) & data testing (figure -9) with the help of W*eka* simulator.

```
Time taken to build model: 1.84 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         120               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.005
Root mean squared error                  0.0061
Relative absolute error                  1.1903 %
Root relative squared error              1.3356 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      50      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0          1         1         1          1        JEN
                 1         0          1         1         1          1        NOT
Weighted Avg.    1         0          1         1         1          1

=== Confusion Matrix ===

  a  b   <-- classified as
 36  0 |  a = JEN
  0 84 |  b = NOT
```

Figure 8: Data training in Weka simulator

In the above figure 8, *Weka* build a model which has taken 1.84 seconds, total number of instances are taken '120' out of which '36' users are *genuine* and '84' users are *not-genuine* as shown in *Confusion Matrix.*

```
Time taken to build model: 0.66 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances          20               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0063
Root mean squared error                  0.007
Relative absolute error                  1.3028 %
Root relative squared error              1.3559 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      50      %
Total Number of Instances               20

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0          1         1         1          1        JEN
                 1         0          1         1         1          1        NOT
Weighted Avg.    1         0          1         1         1          1

=== Confusion Matrix ===

  a  b   <-- classified as
  9  0 |  a = JEN
  0 11 |  b = NOT
```

Figure 9: Data testing in Weka simulator

In the above figure 9, *Weka* build a model which has taken 0.66 seconds, total number of instances are taken '20' out of which '9' users are *genuine* and '11' users are *not-genuine* as shown in *Confusion Matrix.*

4. Advantages of the proposed system

In summary the security system given in this paper having advantages like: -

1. Simple & similar design like conventional system.
2. Easy to implement.
3. No extra hardware required.
4. Free from service provider faith ness circumference.
5. It provides intrusion detection facility.
6. Hierarchical protection gives optimum use of security model with high processing speed.
7. It provides multi-user facility from same security environment.
8. Can be used /implemented in wide range of applications i.e. for standalone computers, Network systems and/or online systems etc.
9. Account can be protected by allocating the maximum number of attempts/trials to the user by locking the account, when login to account is un-successful. Hence, provides better security.
10. It is impossible to break the password with brute force attack because it is depending upon the user behavior and other logical parameters like total time, average time etc.

5. Conclusion

This new method based on user behavior using neural network is simple in designing which provides high level of security & at the same time is also cost effective because it does not need any extra hardware. Keyboard Dynamics, being one of the cheapest forms of biometric, has great scope. It is easy to implement on the password based system or systems. This system also discriminate the users on the basis of their typing behavior as a genuine user and non-genuine user. This method have the number of application numerous irrespective of their nature. With this method two ways security is used which provides more security to password based systems & gives new direction of development to password based security system.

References

[1] D. Shanmugapriya and G. Padmavathi, "A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 5, No. 1, 2009.

[2]R. Giot, M. El-Abed, C. Rosenberger, "Keystroke Dynamics Authentication for Collaborative Systems" 2009.

[3] A. Peacock, X. Ke and M. Wilkerson, "Typing patterns: A key to user Identification", IEEE Security and Privacy 2(5) (2004).

[4] L. O'Gorman, "Comparing Passwords, Tokens, and Biometrics for User Authentication", Proceedings of the IEEE, Vol. 91, No. 12, Dec, pp. 2019-2040, 2003.

[5] A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, January 2004.

[6] A. Awad, E. Ahmed, and I. Traore, "Anomaly Intrusion Detection based on Biometrics", Proceedings of the IEEE, 2005.

[7] M. K. Singh, "Password Based a Generalize Robust Security System design using Neural Network", IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2, 2009

[8] N. Harun, W. L. Woo and S. S. Dlay, "Performance of Keystroke Biometrics Authentication System Using Artificial Neural Network (ANN) and Distance Classifier Method", International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[9] Araujo, L.C.F., Sucupira, L.H.R., Lizarraga, M.G., Ling, L.L., Yabu-Uti, J.B.T., "User authentication through typing biometrics features", IEEE Trans. on Signal Processing, 53 (2), 851–855, (2005).

Preet Inder Singh: M.Sc computer Science from D.A.V College, Amritsar in 2010. Currently pursuing M.Tech (CSE/IT) from Lovely Professional University, Phagwara, interested in Network Security, Multi-media and Artificial Intelligent Systems.

Gour Sundar Mitra Thakur: B.Tech(C.S.E), M.Tech (C.S) Currently Pursuing Ph.D from National Institute of Technology, Durgapur in Mathematics. Areas of Interests are Fuzzy Logic and Fuzzy Mathematics, Soft Computing, Intelligent Systems and Neural Networks.

