RESEARCH ARTICLE OPEN ACCESS

User Authentication Based On Keystroke Dynamics

J.R Nisha[1], R.P. Anto Kumar[2]
[1]PG Student, Dept. of computer science and Engineering, St. Xavier's Catholic College of Engineering, Chunkankadai
[2]Assistant Professor, Dept. of Information Technology, St. Xavier's Catholic College of Engineering, Chunkankadai

Abstract— The most common way to enforce access control is user authentication based on username and password. This form of access control has many flaws which make it vulnerable to hacking. Biometric authentication such as the keystroke dynamics is used in which the keyboard is used in order to identify users. Then the classifier is tailored to each user to find out whether the given user is genuine or not. The contribution of this approach is twofold: first it reduces the possibility of over fitting second it allows scalability to a high volume of users. Here, measured mean, median values, and standard deviation of keystroke features such as latency, dwell time, digraph and their combination are used. The algorithms used for feature subset selection are Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO) and the proposed Renovated Artificial Bee Colony Optimization (RABCO) algorithm. Back Propagation Neural Network (BPNN) is used for classification.

Keywords— Keystroke Dynamics, Feature Extraction, Feature subset selection, Artificial Bee Colony Optimization, Back Propagation Neural Network.

I. INTRODUCTION

Now a day's many persons are trying to hack and misuse others identities like passwords, credit cards details, etc to prevent this type of actions a normal user should keep his user identities safe. A wide range of malicious activities are performed by stolen identities such as online purchases. The user is allowed to perform her intended activity after she entered her credentials. This form of access control, effective to a certain extent, but has many flaws which make it vulnerable to hacking. There are certain rules in order to make a password hard to hack, e.g., include at least eight characters, some of which must be capital letters and special characters (e.g. @, ?, !). The hard-to-hack passwords are also hard-to-remember. Many users choose passwords based on their private lives, e.g., digits from their social security number, pet's name, parent's or kids' name etc are easy to hack. Many users write their passwords on a note which may also easily stolen by hackers. If a hacker takes a user's password from a non-secure website without his knowledge there may be a chance to use his password unnecessary for hacking some of the user's website. This may incur devastating damage to the user. Because of these drawbacks, password-based user authentication methods provide only partial protection against hackers and hence additional authentication means e.g., physiological and behavioral biometrics is used. Behavioral biometrics such as keystroke dynamics can be used to identify the users based on their log-in or the time the user is logged-on. Authentication methods that \ employ this approach will uniquely identify each user.

Commonly, the keystroke dynamics of the user are extracted during login and compared with a reference model that was constructed based on the user's keystroke dynamics and/or similar features of other users. Physiological biometrics includes fingerprints, iris patterns, retina patterns, body heat, and keyboard typing pressure, palm lines, and haptic measurements. The Physiological biometrics based authentication systems that use hardware, and hence more expensive and time consuming to develop while keystroke dynamics does not need additional hardware and hence it is less expensive. The accuracy of biometric based systems may be affected by various factors such as if an injury is occurred in the fingerprint, the system may unable to identify that person; like that if any problem may occurred in the eye the system may unable to identify that person retina. Once a user fingerprints are stolen, then there is no way to change the fingerprints of that user to prevent future impersonation attempts a compromised password can be used. Keystroke dynamics extract and analyze the way an individual types. It also aims to identify the users based on the typing characteristics of the individuals. This may make the authentication process smoother and more user-friendly. The biometric features in addition to the password need to be stored in behavioral biometrics authentication systems

II. RELATED WORKS

Based on feature subset selection various algorithms have been used. They are Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization etc.

Feature subset selection is necessary for an optimization problem that chooses the most optimum or near optimal feature with respect to the performance measures. Since the aim is to obtain the maximum classification accuracy and minimize the classification error. Yu and Cho proposed a Genetic Algorithm – Support Vector Machine (GA-SVM) [4] based wrapper approach for feature subset selection. Here the GA is used for randomized search and SVM is used as a base learner. The main advantage is excellent performance and quick learning speed is desired. The disadvantage of using this method limited the range of parameter values to a small set, and thus it was not efficient enough to find an optimal solution. Second, the GA was still a time consuming searching method although the fast learning speed of the SVM showed its fitness as an induction algorithm. Third, in the FS-Ensemble, the diversity of classifiers was simply measured using hamming distance for feature subset difference and classifier distance for learner diversity.

Particle Swarm Optimization [1] maximizes the classification performance and minimizes the number of features. The disadvantages of this method is multiobjective PSO-based feature selection approach is not used to better explore the Pareto front of nondominated solutions in feature selection problems. It does not know that whether using a given learning algorithm in a wrapper feature selection approach can select a good or near-optimal feature subset for other learning algorithms for classification tasks.

Gabriel L also proposed Particle Swarm Optimization [7] here each particle is represented by a vector of possibilities that indicate the possibility of selecting a particular feature and directly affects the original value of the feature. Support Vector Machine (SVM) is used for classification. The classification error was 1.57% with an FRR of 0.81% and an FAR of 0.76%. The feature reduction rate was far superior, achieving 77.04% and processing time was 1.13s. PSO exhibits a shorter processing time than GA. The disadvantages are auto-associative neural networks is not used here and not better studying variations in the parameters and their influence on results.

The Ant Colony Optimization [15] reduces the redundant feature values and minimizes the search space. Keystroke duration values gives optimum feature subset results when compared with other feature values. Better performance is achieved with keystroke duration feature.

III. PROPOSED SYSTEM

In the proposed work, the timing of each word that the user types are extracted and saved in the database. When the user enters into the application and types the word, timing is compared if matched user is considered as valid user otherwise not. The best timing for each word is found by feature subset selection using RABCO algorithm and the user is classified using back propagation neural network**.**

Figure1 shows the flow diagram of the proposed intelligent system for keystroke dynamics. In the proposed system, every user is characterized by a biometric profile, which is constructed in the following way: First the users are required to type their password for a given number of times. Next the features are extracted from the keystroke dynamics of every password entry and are represented as a vector-one for each password entry. The features extracted are duration, latency, digraph etc.

The features that are extracted from the password entries of a given user form her biometric profile and are stored in a profile database. Relevant, irrelevant and redundant features are usually introduced to the data set are not useful for classification and they may even reduce the classification performance due to the large search space. Hence only the relevant features are selected for classification which reduces the error rate and improves classification accuracy.

The mean and standard deviation for each feature are calculated. Second the feature subset selection is built. Here the Renovated Artificial Bee Colony Optimization (RABCO) algorithm is used as feature subset selection and Back Propagation Neural Network is used as classification. The most promising features in a given dataset are identified by feature subset selection. By using a subset of users instead of the entire set, it aims to achieve two goals: first, prevent over fitting and second facilitate scalability to handle a large number of users.

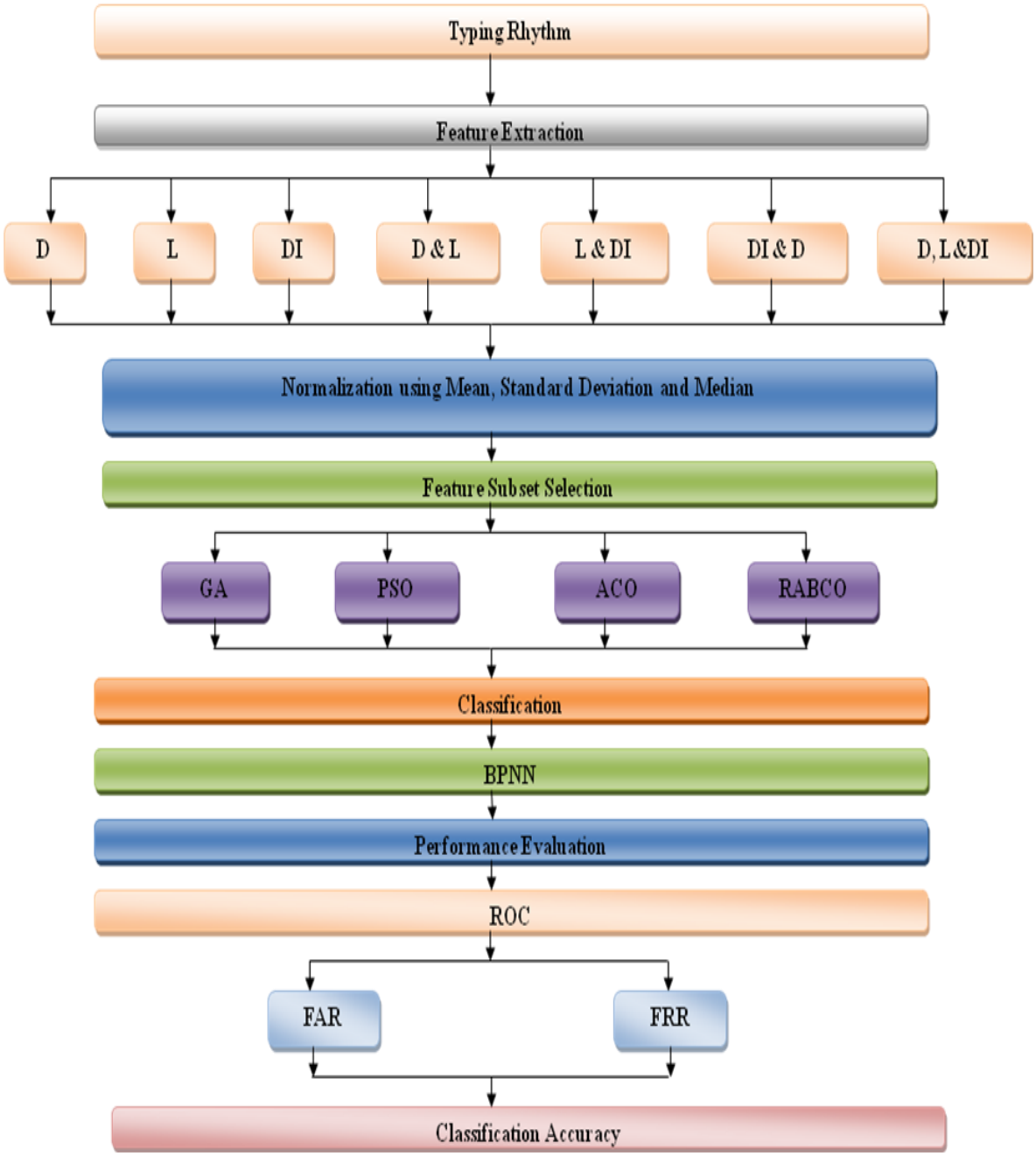

Figure1. Flow Diagram of the proposed Intelligent System for Keystroke Dynamics

A) Renovated Artificial Bee Colony Optimization (RABCO)

Using Renovated Artificial Bee colony Optimization (RABCO) algorithm the best timing for each word is found. The RABCO model consists of three categories of bees: employed bees, onlooker bees and scout bees. Assume that only one artificial employed bee is present in each food source. Hence the number of food sources is equal to the number of employed bees. Employed bees go to their food source and evaluating their nectar amounts. The employed bees memorize the higher fitness value and forget the lower fitness value when it finds the new food source. The employed bees share the nectar information of food source with the onlooker bees. The onlooker bees select the food source based on the information given by employed bees and calculate the nectar amount of the food source. Then the scout bees are sending randomly to find the new food sources. This process is repeated until the requirements are met. The main steps of the algorithm are given below:

- Initial the food sources.
- Evaluate the population
- Cycle=1
- REPEAT
- Each employed bee goes to a food source in her memory and determines a neighbor source, then evaluates its nectar amount and dances in the hive.
- Each onlooker bees select the food sources based on the information of the employed bees and then go to that source. Then she evaluates its nectar amount.
- Abandoned food sources are determined and are replaced with the new food.
- Sources discovered by scouts.
- The best food source found so far is registered.
- UNTIL (requirements are met)

The first step in Renovated Artificial Bee Colony Optimization algorithm is initializes the population and then evaluate the population. Each food sources contains one employed bee and the employed bees go the food sources in her memory and determine a neighbor source. The employed bees memorize the food sources with highest fitness value and forget the older one. Each onlooker bees select the food sources based on the information of the employed bees and then go to that source. Then she evaluates its nectar amount. Then the scout bees are sending to the food source to find any new food sources are available. These steps are repeated until the requirements are met.

B) Back Propagation Neural Network (BPNN)

Back Propagation Neural Network (BPNN) algorithm is used for classification. The various combinations of the dominant features from the output of feature subset algorithm RABCO are used in BPNN. The reference subset samples from the feature subset selection algorithms were used to train the neural network. The training result was stored in the Training database. The back propagation neural network uses a training set composed of input vectors and a desired output (here the desired output is usually a vector instead of a single value). These elements or nodes are arranged into layers: input, middle and output. The main step of this algorithm is

- Initialize weights (typically random!)
- Keep doing epochs
  - For each example e in training set do
    - forward pass to compute
      - ✓ O=neural-net-output(network,e)
      - ✓ miss = (T-O) at each output unit
    - backward pass to calculate deltas to weights
    - update all weights
  - end

- until tuning set error stops improving

Next the performance evaluation is calculated by drawing a curve that is related to the ROC curve plots the FAR versus the FRR.

This curve is useful for the evaluation of authentication system since FAR corresponds to malicious users who are logged into the system, while FRR corresponds to legitimate users being blocked from accessing the system. We aim to minimize both but usually the FRR increases with the decrease in the FAR and thus, ERR describes the point both achieve the best measure with respect to one another. The performance analysis comparing the various algorithms is given below.

Performance analysis of a classifier

Algorithm	Average Accuracy (%)	Average Error Rate
Genetic algorithm	87.54	0.067
Particle Swarm Optimization	89.23	0.059
Ant colony optimization	92.8	0.050
Artificial Bee Colony Optimization	93.5	0.045

IV. CONCLUSION

Comparison with various algorithms shows that RABCO shows the better classification results. After feature subset selection is calculated, a classifier is built using the timing vector patterns. In the proposed work, Back Propagation Neural Network (BPNN) is used for classification.

REFERENCES

[1] Bing Xue, Mengjie Zhang and Will N.Browne, "Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach," *IEEE Trans.Cybern*., oct. 2012

[2] S.Bleha, C. Slivinsky, and B.Hussein, "Computer-access security systems using keystroke dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 12, pp. 1217–1222, Dec. 1990.

[[3] A. El-Saddik, M. Orozco, Y. Asfaw, S. Shirmohammadi, and A. Adler, "A novel biometric system for identification and verification of haptic users," *IEEE Trans.*

Instrum. Meas., vol. 56, no. 3, pp. 895–906, Jun. 2007.

[[4] Enzhe Yu and Sungzoon Cho, "GA-SVM Wrapper Approach for Feature Subset Selection in Keystroke Dynamics Identity Verification" *IEEE Trans*, 2003.

[[5] Enzhe Yu and Sungzoon Cho, "Keystroke dynamics identity Verification- its problems and practical solutions," *computer and security*, vol. 23, pp. 428-440, Feb. 2004.

[[6] Fabian Monrose and Aviel D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation Com. Syst.*, pp. 351–359, Mar. 1999.

[[7] Gabriel L, F.B.G. Azevedo, George D.C.Cavalcanti and E.C.B Carvalho Filho, "An approach to feature selection for keystroke dynamics systems based on PSO and feature weighting," *IEEE Trans.*, 2007

[[8] Glaucya C. Boechat, Jeneffer C. Ferreira, and Edson C. B. Carvalho, Filhoin, "Using the Keystrokes Dynamic for Systems of Personal Security," *World acad. of sci., Eng., and Tech.*, pp. *887-892*, 2008.

[9] N. J. Grabham and N. M. White, "Validation of keypad user identity using a novel biometric technique," *J. Phys.: Conf. Ser.*, vol. 76, no. 1, pp. 012023-1–012023-6, 2007.

[[10] N. Hidetosshi and M. Kurihara, "Sensing pressure for authentication system," in Proc. Int. Conf. Comput. Intell., Istanbul, Turkey, Dec. 17–19, 2004, pp. 19–22.

[[11] D. Hosseinzadeh and S. Krishnan, "Gaussian mixture modeling of keystroke patterns for biometric applications," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 6, pp. 816–826, Nov. 2008.

[[12] John A. Robinson , Vicky M. Liang, J. A. Michael Chambers, and Christine L. MacKenzie , "Computer User Verification Using Login String Keystroke Dynamics," *IEEE Trans. on sys., man, and cybern.*, Vol. 28, No. 2, March 1998.

[13] K. S. Killourhy and R. A. Maxion, "Comparing anomaly detectors for keystroke dynamics," in *Proc. 39th Annu. Int. Conf. Dependable Syst. Netw.*, Lisbon, Portugal, Jun. 29–Jul. 2, 2009, pp. 125–134.

[14] Lívia C. F. Araújo, Luiz H. R. Sucupira Jr., Miguel G. Lizárraga, Lee L. Ling, and João B. T. Yabu-Uti , "User Authentication Through Typing Biometrics Features," *IEEE Trans on sign. proc*., vol. 53, No. 2, pp. 851-855, Feb. 2005.

[15] Marcus Karnan et.al(2009) in "Feature subset selection in keystroke dynamics using ant colony optimization," *journel of Engin.and tech. research*, vol.1(5), pp. 072-080, Aug. 2009