International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

Active Authentication using Keystroke Sound

Princy Ann Thomas[1], Ms. Nimmy K.[2], Prof. P. Jayakumar[3]

[1]MTech Cyber Security, Dept. of CSE, SNGCE, Kadayiruppu, Kerala, India.
[2]Asst. Prof. MTech Cyber Security, Dept. of CSE, SNGCE, Kadayiruppu, Kerala, India.
[3]Prof. & Head, Dept. of CSE, SNGCE, Kadayiruppu, Kerala, India.

Abstract - *User authentication is usually implemented at login. However, one-time validation of the user's identity is increasingly becoming insufficient. There is a requirement for continuous non-intrusive verification of the user's identity. In this paper the discriminative power of keystroke sound is used to continuously monitor the authenticity of a user after initial authentication at login. Digraphs are used in modeling keystroke dynamics. Digraph latency within the pairs of virtual letters and other statistical features are used to generate match scores. Active authentication using keystroke sound adds to the measure of confidence that the user has not changed during a session. Mouse gesture dynamics is added to improve the accuracy of the system from 0.86 to 0.92 on detection of an imposter. The log generated can be useful as corroborative evidence on forensic investigations. The system is to be implemented in areas where security is critical and environmental noise would be low.*

Key Words: Continuous Authentication, Keystroke Sound, Mel Frequency Cepstral Coefficient (MFCC), K-Means Clustering, Code Book, Digraph Latency.

1.INTRODUCTION

Authentication allows a user to claim his identity. This allows privilege assignment and authorization according to user requirements. In the authentication process the credentials provided by the user are compared to stored values. If a match occurs the process is completed and user is granted authorization for access. Authentication is categorized into type 1 based on something the user knows like passwords, type 2 based on something the user has and type 3 based on something the user is. Biometrics is of type 3 category. Biometrics is once again categorized into physical biometrics and behavioural biometrics. Physical biometrics uses the physical characteristics of an individual for authentication like fingerprint or iris and so on. Behavioural biometrics uses the behavioural pattern of an individual when using specific device. In this work we use the keyboard usage behavioural pattern. This system requires no additional hardware as opposed to physical biometric authentication. All authentication methods have their own advantages and disadvantages and the environment determines which authentication method is best suited. The disadvantage that all static authentication methods is that the user is authenticated once and any subsequent change of user after that will be unnoticeable to the system. Active authentication on the other hand is used after an authenticated user has logged into a system to monitor any change in user. Such a system must work in the background in order to be non-intrusive. A natural way for active authentication is the monitoring of the keyboard and mouse behaviour of the user. Keystroke dynamics measures how a user uses a keyboard. The least hassle method is to collect timing information of a key when it is pressed down and when it is released, collected using software that checks if a key is pressed down or released and in the event of these two occurrences collects information on what key was used and what the time was. This method has the disadvantage that the actual keys being pressed are logged which in itself presents a security hazard. The keystroke sound technique on the other hand avoids keystroke logging and uses virtual alphabet concept to represent the acoustic stream of keys being pressed [1]. Duration, latency and digraph information is exploited to correctly identify an imposter from a genuine user. The keystroke sound method has the disadvantage of being less accurate in the presence of environment noise. To remedy this situation mouse gesture dynamics is used once the system detects an imposter. Though mouse gesture is intrusive, it is only used once there is a doubt in the identity of the user.

2. MULTIMEDIA DATA MINING

Vijayarani et al [2] provides the basic concepts of multimedia mining and its essential characteristics. Multimedia data mining is used to extract information from audio, video, images, graphics, speech, text and combination of data sets all converted from different formats into digital media. Multimedia data are classified into text data, image data, audio data which may be speech or music and video data. This work is based on the concepts of audio data mining where the contents of an audio signal can be automatically searched, analyzed and transformed. Band energy, frequency centroid, zero crossing rate, pitch period and band-width are often used for audio processing. Fig - 1 shows the basic architecture of multimedia data mining. Input stage captures and decides which multimedia database is used for the data mining

process. Multimedia Content requires the user to select the databases, subset of features or data to be used for data mining. Spatio-temporal segmentation identifies the spatial and temporal coherences in the multimedia content and derive appropriate representation for the data sequence [3]. Feature extraction is part of pre-processing involving data integration and characterizing or coding relevant data fields. The hidden patterns and trends in the data are discovered in this stage. Association, classification, clustering, regression, time-series analysis and visualization are used to find similar patterns. Evaluations of results determine whether prior stage must be repeated and consists of reporting and using the extracted knowledge to produce decision making.

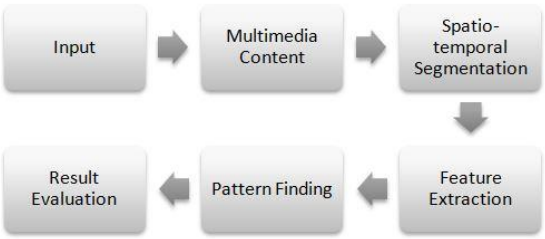

Fig -1: Multimedia Mining Architecture

3. ACTIVE AUTHENTICATION SYSTEM

Biometric authentication verifies the identity of a user by their physical trait. In the system proposed in this paper we use the behavioural traits of a user when using the keyboard and mouse. Since these devices are used on a continuous basis by most users they are the logical choice for a continuous authentication system. Keystroke dynamics although non-intrusive employs key logging to record typed texts which could lead to a privacy risk. Roth et al [1] consider keystroke sound as a biometric modality for active authentication. They were able to achieve an EER of 11% in ideal conditions. In this work we propose to apply the concepts of authentication using keystroke sound for continuous user authentication and to improve the accuracy of the system the concept of mouse gesture dynamics is incorporated. Sayed et al [4] in their work has achieved an EER of 5.11% using mouse gesture dynamics. Using the same database used by Roth et al we were able to increase the accuracy of the system from 86% to 92% with a decreased EER of 9%.

Fig - 2 gives the basic overview of the system. The keystroke sounds as a user types is captured and recorded by a microphone, the system then performs feature extraction and matching continuously at designated time intervals to verify the identity of the user. If the user is authenticated then he will be allowed to continue using the session without interruption. In case the system detects a change in user during an ongoing session, then the user will be required to re-authenticate with a mouse gesture. If he fails to reclaim his identity, the intruder will be blocked and relevant log entries will be made.

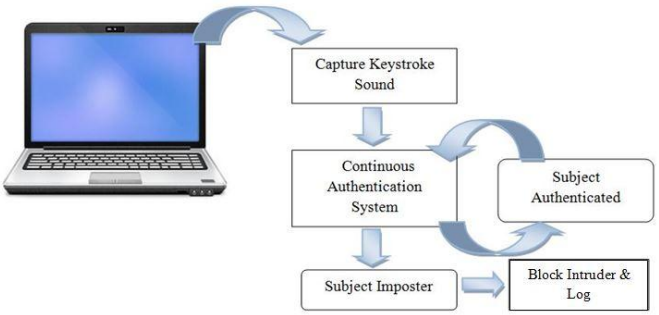

Fig -2: Overview of Authentication System

3.1 Technical Approach

To match two keystroke sound streams, digraph latency which is one of the most often used features in keystroke dynamics is used [5]. It is the time difference between pressing the keys of two successive letters. The system is trained using acoustic signals of training samples. Temporal segmentation will be used to detect segments of keystrokes. The MFCC of these keystrokes will then be input to a clustering routine. Each cluster centroid will be considered a virtual letter or code which allows the computation of the most frequent digraphs i.e. a pair of cluster centroids. The statistical attributes of each subject is also calculated. Histograms of virtual letters and digraphs in the sound streams are then calculated. A number of similarity scores are generated and fused to determine whether probe streams match the gallery streams. On mismatch, mouse gesture re-authentication will be required. A mouse gesture results from the combination of mouse movements and clicks in a way that the system recognizes as specific commands [4]. If an intruder is identified then the event will be entered in the log.

3.2 Architecture

The authentication system operates in three different stages: training, enrolment and authentication. During the training phase the keystroke sound streams of various subjects are used to learn the parameters that will be suitable for the algorithm in the given environment Fig - 3. Given a raw keystroke sound stream, the keystroke portions are estimated and separated from the silent periods in the temporal domain. The MFCC features are then extracted the segments are clustered using K-means clustering. Each centroid is considered as virtual letters in the virtual alphabet. The system is then able to calculate the top N digraphs, frequency of digraphs and virtual letters and the scores for digraph statistics, histogram of digraphs virtual letters and intra letter distance. These scores are fused to give the final score which helps with decision making.

The enrollment phase records the keystroke sound streams of a subject typing a static text feed. Temporal segmentation and feature extraction is used to create and store a user template for the subject. Fig - 4 shows the steps in the enrolment stage.

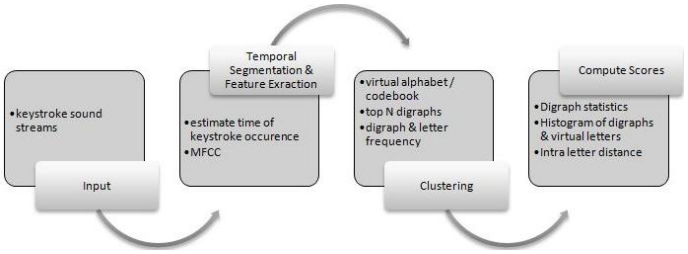

Fig -3: Training Phase

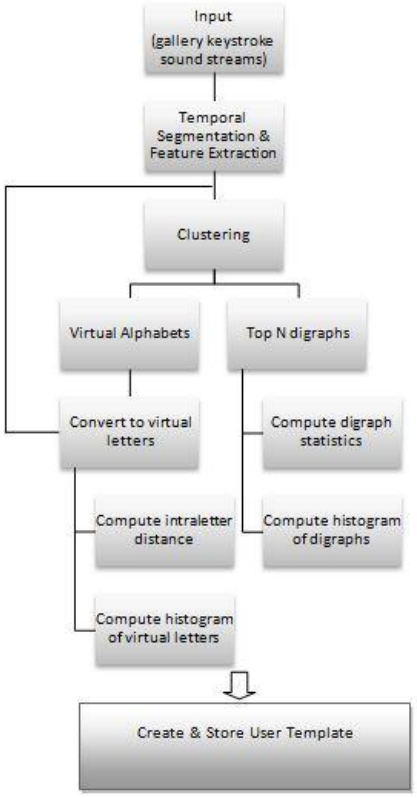

Fig -4: Enrolment Phase

In the authentication phase, a user claims his identity with an ordinary login screen using user id and password. The system then captures the keystroke sound of the user continuously during the session. The temporal extraction and feature extraction takes place in real time after which the system compares the current user template with the stored user templates. Similarity scores are then computed and fused. If the fused similarity score is high enough then the user is accepted and may continue to use the session. If the fused similarity score is not high enough then the user is designated an intruder and a mouse gesture re-authentication is required. Fig - 5 shows the score fusion and authentication steps.

The combined score of four similarity scores are taken to make the authentication decision in this architecture. In the case of keystroke acoustics the digraph information is not directly available, we use digraph latency from the timing information. The similarity score based on digraphs should be high for an authentic user and low for an intruder. Since we have clustered the keystroke segments using k means clustering algorithm we will have k virtual letters. In this case it would be possible to generate k^2 digraphs. In the training stage we generate a list of top N digraphs and the frequency of digraphs. To calculate the similarity of two acoustic signals of digraphs, the mean (m) and deviation (σ) of the digraph latency is calculated using S_{1d}. It calculates the overlapping region between two Gaussian distributions of the same digraph using the Bhattacharyya coefficient. The coefficient can be used to determine the relative closeness of the two samples being considered [6].

$$S_{1d}(K, K') = \sum_{\triangle t} \sqrt{N(\triangle t; m_n) N(\triangle t; m_{n'})}$$

The virtual alphabet represents the collection of virtual letters which in this work is simply represented using numbers from 1 to k. It is possible that different users may produce different virtual letter for the same key. This gives us the benefit of being able to use frequency of the top N digraphs for authentication. *h* represents the histogram of top N digraphs and is calculated using h_n. Here, δ is the indicator function, numerator is the number of digraph dn = k_{n1}; k_{n2} , |K| is the length of acoustic signal within which the digraphs are considered. S_2 computes the similarity between two histograms of digraphs.

$$h = [h_1, h_2, ..., h_N]^T$$

$$h_n = \frac{\sum_i \delta(l_i = k_{n2})\delta(l_{i-1} = k_{n1})}{|K| - 1}$$

$$S_2(K, K') = h^T h'$$

MFCC and clustering is done to decrease the distance between elements within a cluster and increase the distance between elements in different clusters. There is a distinct intra letter separation due to the fact that different users pressing different keys may be represented using a single virtual letter due to similarity in keystroke acoustics and the individuality of the typing behaviour of each user. f_k calculates the mean of K MFCC features for each virtual letter and S_3 calculates the intra letter distance. K and K' are two corresponding acoustic signals, w_3^n is the overall frequency of each virtual letter among all keystroke segments in the training set and -1 ensures that a genuine probe will be larger than an imposter probe.

$$\overline{f_k} = \frac{1}{|l_i = k|} \sum_{l_i = k} f_i$$

$$S_3(K, K') = -\sum_{k=1}^{K} w_3^n ||\overline{f_k} - \overline{f'}_k||_2$$

Different users may produce different sounds pressing the same key. This allows users to be identified based on these acoustic distributions. The histogram of virtual letters is represented by η and computed by η_n. k_n is the number of keystrokes assigned to a virtual letter. S_4 calculates the similarity between two histograms of virtual letters.

$$\eta = [\eta_1, \eta_2, ..., \eta_K]^T$$

$$\eta_n = \frac{\sum_i \delta(l_i = k_n)}{|K|}$$

$$S_4(K, K') = \eta^T \eta'$$

The fused similarity score is calculated by S. Here, c_v is the optimal LDA projection vector $[c_1,c_2,c_3,c_4]^T$ learned on the training data set, m_{sv} is the mean of the imposter score distributions learned from the training data set, σ_{sv} standard deviation of the imposter score distributions learned from the training data set and S_v are the computed individual similarity scores.

$$S = \sum_{v=1}^{4} c_v \frac{S_v - m_{sv}}{\sigma_{sv}}$$

This would cause imposter scores to fall within a standard normal distribution and genuine score to be outliers on the positive side. The authentication decision is based on a threshold τ, if S>τ then user is genuine else user is imposter.

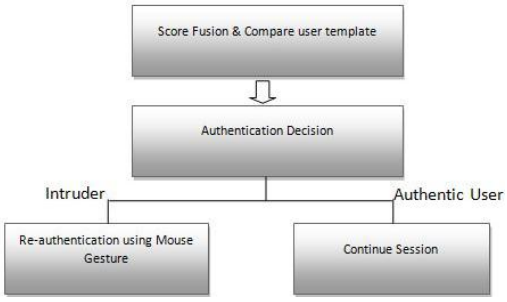

Fig - 5: Authentication Phase

The mouse gesture detection method is very similar to that used by Sayed et al. Here once the enrolment is done with 10 repetitions of the same gesture, a freehand drawing area is presented to the user on detection of a imposter probe by the keystroke sound authentication system. If the replication is within the given threshold then the user is deemed authentic and the user is allowed to continue using the system. The raw data will consist of the horizontal coordinate which is the x-axis, vertical coordinate which is the y-axis and the elapsed time in milliseconds at each pixel. A gesture is a sequence of triples represented as (x,y,t). The number of data points that are recorded for a given gesture would be the size of the gesture. The gesture creation is attained by the free hand drawing area where users can freely draw a set of gestures. The gestures are restricted to uni-stroke gestures without closed loops. The gesture creation module both normalizes the input to the center of the drawing area and normalizes the gesture to get a size of 16 data points. Table -1 shows the extracted features. The euclidean distance between the features is calculated and within a reasonable threshold the user is considered to be authentic.

Table -1: Features extracted from mouse gesture

Mouse Features Extracted	
X co-ordinate	x
Y co-ordinate	y
Time	t
Horizontal velocity	$\frac{\Delta x}{\Delta t}$
Vertical velocity	$\frac{\Delta y}{\Delta t}$
Slope angle of tangent	$\arctan\left(\frac{y_i}{x_i}\right)$

4. RESULT AND ANALYSIS

The pilot system was designed to analyse the effectiveness of keystroke sound as an authentication biometric. Chart - 1 shows the ROC curve for the digraph statistics.

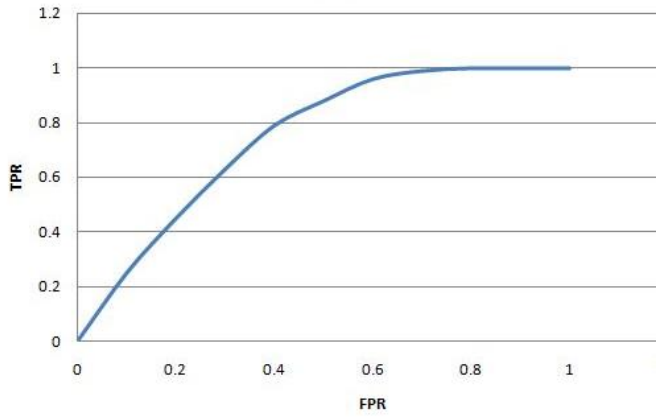

Chart -1: ROC for digraph statistics

The probability distribution score plots for duration and digraph are shown in Chart - 2 and Chart - 3.

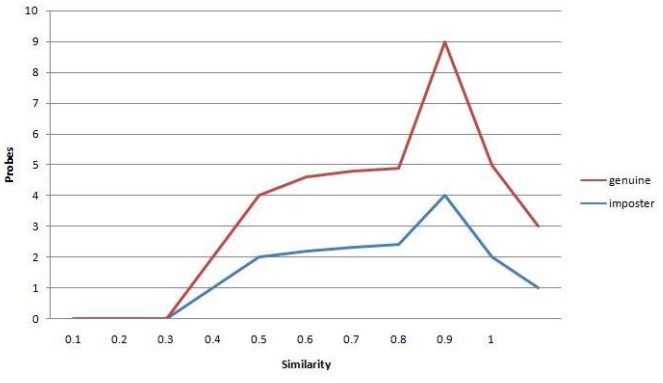

Chart -2: Duration based similarity scores.

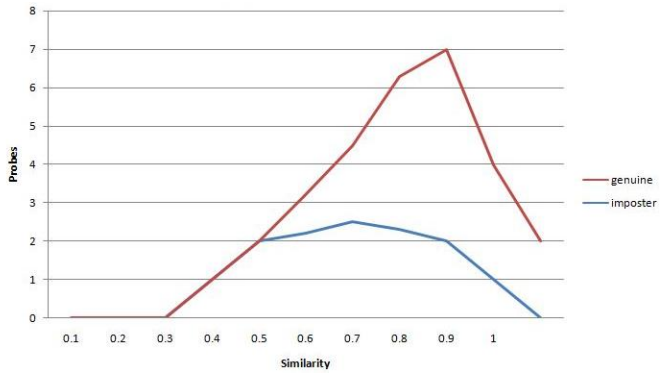

Chart -3: Digraph based similarity scores

Table - 2 shows the analysis of the pilot system giving a TPR (true positive rate) of 0.89, FPR (false positive rate) of 0.11 and accuracy 86%.

Table -2: Analysis of pilot system

Total Population (140)	Predicted Genuine	Predicted Imposter	Row Total
Genuine condition	62	8	70
Imposter condition	8	62	70
Column Total	70	70	140

The mouse gesture authentication system shows an accuracy of 95% at the optimal threshold of 0.8. The accuracy of this system is dependent on the kind of gesture that is used by the user for authentication. Table - 3 shows the result of analysis on various mouse gestures in this system.

Table -3: Gesture based analysis

Mouse Gesture	TPR	FRR	FPR	Accuracy
L	0.8	0.2	0.2	0.8
v	0.8	0.2	0.6	0.6
٦	0.9	0.1	0.7	0.6
ρ	1	0	0.1	0.95
Ɔ	0.9	0.1	0.2	0.85
C	0.8	0.2	0.3	0.75
Z	0.9	0.2	0.1	0.9

Analysis based on different threshold values for the mouse gesture authentication is shown in Table - 4. We see that at 0.8 we have optimal threshold. Table – 5 shows the comparison between different biometric systems and how the proposed system increases the accuracy of keystroke sound biometric when combined with mouse gesture.

5. CONCLUSIONS

When using keystroke sound as a biometric for authentication the major drawback is the presence of environmental noise which caused inefficient performance.

Table -4: Gesture threshold optimization

Threshold	TP	FP	FN	TN	TPR	FRR	FPR	Accuracy
0.2	0	0	70	70	0	1	0	0.5
0.3	5	1	65	69	0.07	0.93	0.01	0.53
0.4	13	1	57	69	0.19	0.81	0.01	0.59
0.5	31	1	39	69	0.44	0.56	0.01	0.71
0.6	42	5	28	65	0.6	0.4	0.07	0.76
0.7	49	12	21	58	0.7	0.3	0.17	0.76
0.8	61	22	9	48	0.87	0.13	0.31	0.78
0.9	65	30	5	40	0.93	0.07	0.43	0.75
1	67	37	3	33	0.96	0.04	0.53	0.71
2	70	66	0	4	1	0	0.94	0.53

Table -5: Gesture threshold optimization

	Keystroke dynamics	Keystroke sound	Mouse dynamics	Mouse gesture	Proposed system
Static/Dynamic	Both	Both	Both	Static	Both
EER	0.09	0.11	0.03	0	0.9
Accuracy	0.91	0.86	0.97	0.95	0.92

When mouse gesture is used as an added biometric the system still works as a non-intrusive active authentication system till an imposter is identified. Then the mouse gesture re-authentication blocks the user till his identity is positively verified. Log entries are based on the result of the mouse gesture authentication making it more accurate. These logs may be used as corroborative evidence if forensic evidence collection is required. In the experimental system, an existing database from the work of Roth et al is used to design an intruder detecting system. Mouse gesture dynamics is used to increase the accuracy of the identification decreasing the EER from 11% to 9% and increase accuracy from 86% to 92%.

REFERENCES

[1] A. J.Roth, X.Liu and D.Metaxas, “Investigating the discriminative power of keystroke sound”, IEEE Trans. on Information Forensics and Security, Vol. 10, No. 2, pp. 333-345, 2015.

[2] Dr.S.Vijayarani and Ms.A.Sakila, “Multimedia mining research – an overview”, International Journal of Computer Graphics Animation (IJCGA) Vol.5, No.1, pp. 69-77, 2015.

[3] J. Y. A. Wang and E. H. Adelson, “Spatio-temporal segmentation of video data”, M.I.T. Media Laboratory Vision and Modeling Group, Technical Report No. 262, Appears in Proceedings of the SPIE: Image and Video Processing II, vol. 2182, San Jose, pp. 1-12, 1994.

[4] I. W. B. Sayed, Issa Traore and M. Obaidat, “Biometric authentication using mouse gesture dynamics,” IEEE Systems Journal, vol. 7, pp. 262-273, 2013.

[5] F. Z. Li Zhuang and J. D. Tygar, “Keyboard acoustic emanations revisited,” Proceedings of the 12th ACM Conference on Computer and Communications Security, November 2005, pp. 373-382, ACM New York, 2005.