International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Euler Movement Firefly Algorithm and Fuzzy Kernel Support Vector Machine Classifier for Keystroke Authentication

M. Rathi, A. V. Senthil Kumar

Abstract: ***User authentication can be successfully employed using keyboard typing patterns which is a form of behavioural biometrics. This modern method is highly analyzed for static authentication which refers to typing of fixed texts like 'password' and 'pin numbers'. Most of the methods with respect to keystroke dynamics are restricted to the study of user's activity involving fixed text. The formulated work concentrates on the investigation of the log of the user activity focused on the keyboard usage within the computer system through free text which refers to typing of texts throughout the login session. The Buffalo dataset is used in User Profiling Similarity Measurement (UPSM) stage and to recognize the time slice of the users, Euler Movement Firefly Algorithm (EMFA) is utilized. The typing behaviour is formulated in the form of time series in User Profiling Continuous Keystroke Authentication (UPCKA). Moreover the progression is made to user's Continuous Authentication so as to predict unauthorized users with the consideration of the classifier called Novel Fuzzy Kernel Support Vector Machine (NFKSVM). The experimental results provide the enhanced performance by utilizing the formulated UPCKA in correlation with the NFKSVM classifier when compared with SVM and Iterative Keystroke Continuous Authentication (IKCA) techniques.***

INDEX TERMS: ***Keystroke, Keystroke Time Series, Continuous Authentication, Buffalo dataset, User Profiling Similarity Measurement (UPSM) and User Profiling Continuous Keystroke Authentication (UPCKA).***

I. INTRODUCTION

The greater diffusion of the digital recognitions has led to the development of security issues due to data transmissions [1]. Nowadays, the perspective of the large diffusion involving in the various activities transmitted over the internet through events like online transactions in banking, transaction involving E-commerce, communication through e-mail tends to suffer security attacks [2]. Due to this, the theft regarding the identity of the person has become predominant and it has gained new momentum. The illegal use of personal information of someone else and pretending to be the actual person is generally termed as identity theft [3]. Under this situation, a variety of modern techniques have been developed for the purpose of user authentication. The process of confirming the users' identity is called Authentication. For instance, within workstations, initial authentication takes place, which is the system initialization. Moreover highly secured authentication techniques do not even render complete safety security mechanisms, As the computers may be subjected to unauthorized users whenever the user is left from the workstation without ending the session. Similarly the unauthorized users could handle the system pretending like a legitimate person, which leads to theft of identity [3]. Out of several methods one such technique to solve the issue caused by the intrusion is that the use of detection mechanisms that focus on workstation (host-based). Keystroke dynamics [4-5] (typing patterns) are considered to be the challenging tasks for the persistent authentication. To authenticate the typing patterns of an individual, the initial task was subjected to the text that remains static. For instance, static authentication related to that of typing pattern, the rhythm, recognize only when the users enter their credential data (username and password, or pin number) [6–7]. Keystroke dynamics is considered to be significantly precise to the choice of authentication due to its degree of transparency it produces. The most obvious way to take advantage of it is to gather timing information on data that users have already typed to login into the system—that is, username and password. Keystroke Static Authentication (KSA) has been subjected by considering the applications including username, password and pin number authentication [6–7]. KSA remains unsuitable for the applications that are in need of regular authentication like the context of the online assessments applied in eLearning environments. Hence, Keystroke Continuous Authentication (KCA) is required. When compared to KSA , KCA is considered to be more promising as the process focuses on discovering patterns from the text which is set free (not as to that of KSA planning for a fixed pattern which remains single). The working strategy of KCA till date has focused on feature vector based binary classification where the statistical features like the average hold time (duration of a key press) and digraph latency (duration between the start or end of pairs of common consecutive key presses) have been considered to be important [8–9]. These mechanisms function regularly by evaluating the similarity among the learnt user having a statistical profile and unseen data previously presented in the data stream. The authors are motivated behind the time series approach as it can be easily used to detect the suspicious behaviour during sequence of keystrokes. The idea presented in this work is to conceptualise the keystroke process in terms of time series from which the KCA have been identified without utilizing the feature vector based classification. More specifically the idea is to view keystrokes in terms of press-and-release temporal events such that a series of successive events can be recorded.

Manuscript published on 30 September 2019.
***Correspondence Author(s)
Rathi. M, Assistant Professor in Department of Computer Technology, Dr. NGP Arts and Science College, Coimbatore since 2016.
Dr. A. V. Senthil Kumar, Professor and Director of Department of Computer Applications, Hindusthan college of Arts and Science, Coimbatore.

In addition, Novel Fuzzy Kernel Support Vector Machine (NFKSVM) classifier need to be built for each user and this in turn improves the efficiency of the application of KCA in real environments.

II. LITERATURE REVIEW

Kang and Cho [10] formulated the relations among the topologies. The formulated method is correlated against other 13 novelty detectors depending on 21 benchmark sets of data from 2 different sources. Then the technique is applied with real time applications where incremental learning is required in keystroke dynamics-based user authentication. From the experimentation results, it has been concluded that the proposed method enhanced the performances in terms of both distance-based novelty detectors, and non-distance-based algorithms.

Ahmed and Traore [11] proposed an innovative method for the purpose of free text investigation of keystrokes that merges monograph and digraph analysis, and utilization of neural network to detect missing digraphs depending on the relation between the monitored keystrokes. The examination of experiments involving 53 users in various environments provides a ratio under a false acceptance rate (FAR) of 0.0152% and a false rejection rate (FRR) of 4.82% and at an equal error rate (EER) of 2.46%. Within the experiments conducted in the similar environment with 17 users, provides FAR of 0% and FRR of 5.01% and EER of 2.13%.

Hocquet et al [12] proposed a technique for the realization in classification of keystroke dynamics users earlier to the user authentication process. The aim of this technique is to fix dynamically the single attributes of the classification process for every users of each class. The user learning set is used to extract the features and the clusters are segmented into user set by utilizing the clustering algorithm. A group of parameters are evaluated for every cluster. The process of authentication is carried out in two steps. Initially, the users are related to the clusters and secondly the clusters containing the parameters are utilized during the authentication process. These 2 steps render better outputs when compared to that of a system with global settings.

Hu et al [13] presented a technique namely K-Nearest Neighbor which is used in the classification process of users' keystroke dynamics profiles. For the purpose of authentication, an input has been validated against the cluster profiles which have minimized the verification load effectively. Giot et al [14] presented an innovative technique which is combined with SVM learning, satisfying the operational conditions (less than 5 captures for the process of enrolment). In the formulated work, the users seem to be authenticated when the users satisfy the keystroke dynamics of a passphrase. The GREYC keystroke is used as a benchmark which comprises a huge range of users (100) for the purpose of validation. The experimental results show that the proposed method performs better in terms of context of operations.

Alshehri et al [15] formulated an innovative Keystroke Continuous Authentication (KCA) technique which differs from feature vector representations. It depends on the analysis of the timeline series, by considering the keystroke sequences. The authors have also suggested that KCA can be applied in the area of online examinations which are subjected to the environments like eLearning. Alshehri et al [16] proposed an innovative method in the identification of typing behaviour from subjective text in heterogeneous environments using time series analytics.

Patil and Renke [17] formulated an investigation which was carried out with the typing pattern of the humans. No extra hardware is necessary as the keystroke dynamics doesn't require any sort of hardware. For the purpose of password identification, software based technology becomes difficult task. The output renders the highest security for growth in the development of web based applications. Wangsuk and Anusas-amornkul [18] focused on the improvement in the credentials like username and scheme of the password authentication, which has retained certain weaknesses as the username is known publicly and a password can be entered by guessing. If the attacker is aware or guesses a password accurately, the system can be satisfied. Hence the work concentrates on the defects and formulates an extra token ensuring the security to this system by merging the dynamics of the keystrokes into the system.

Alshehri et al [19] proposed a new realistic Iterative Keystroke Continuous authentication (IKCA) by considering the typing behaviour of the user in time series format which restricts the demerits related to the feature vector method. The outputs of the experiments demonstrated a significant performance which is enhanced with the help of formulated technique correlated with the feature vector based approach. The application of keystroke dynamics authentication for Online Exams [20] and the importance of using free text instead of fixed text [21] were also studied in the recent works.

III. PROPOSED METHODOLOGY

The formulated technique considers the study of the users' activity log correlated with the usage of keyboard in a system. Depending on a Keystroke Dynamics dataset the User Profiling Similarity Measurement (UPSM) of text is carried out. UPSM is done by Euler Movement Firefly Algorithm (EMFA) in order to recognize the time slice of the same type of users. User Profiling Continuous Keystroke Authentication (UPCKA) is proposed by considering the typing behaviour as a form of time series. Moreover, a trial is made by the user in Continuous Authentication so as to identify the unauthorised persons with the help of classifier called Novel Fuzzy Kernel Support Vector Machine (NFKSVM). The representation of the overall formulated work is shown in fig. 1.

A. Dataset

The dataset consists of free text typing as well as the keystrokes corresponding to transcriptions. The datasets' segment is extracted with the help of variety of keyboards available at different sessions. Moreover the mouse is used to co-ordinate the information and its events provided for the keystroke dataset. The samples of the dataset are gathered from https://cubs.buffalo.edu/research/datasets.

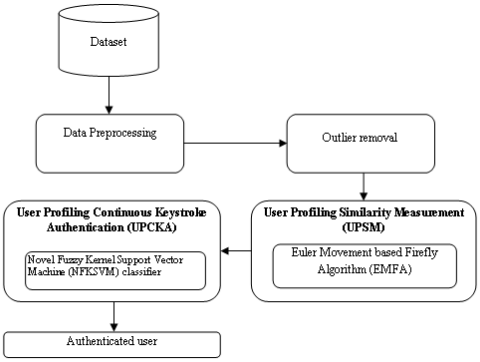

Fig. 1: Architecture of the proposed system

B. Keystroke Time Series Representation

Keystroke dynamics contains the events like key press-and-release temporal events. In addition, these events predefine its patterns of characteristic typing.

Definition 1

'KT' is a keystroke time series, which is an ordered discrete sequential keystroke points 'KP', KT = {kp_1, kp_2,…, kp_n} where n is considered to be the length of the complete series, and in every point 'kp_i' is a keystroke event. For every press of a keystroke, four timing values can be obtained:

(1) Key-down time 'KDT'
(2) Key-up time 'KUT',
(3) Key-hold time 'KHT' and
(4) Flight-time 'FT'.

For any keystroke i, KHT_i can be evaluated by $KHT_i = KUT_i - KDT_i$. The value for FT_i can be extracted by $FT_i = KUT_{i-1} - KDT_i$. KHT and FT are the significant features of keystroke. The tuple having the form ⟨KHT, FT⟩ is used to describe the keystroke. In the proposed work, KT= {FT_1 , FT_2 ,…,FT_i} is used for implementation.

Definition 2

A keystroke time series subsequence 'ks', having a length l, is a subsequence of 'KT' that initiates at a point kp_i within KT and terminates at point kp_{i+l-1} , thus ks = { kp_i, kp_{i+1},…, kp_{i+l-1}}. A subsequence 'ks' of 'KT' is shown through the notation ks ≤ KT (∀kp_i ∈ ks, ∃kp_j ∈ KT such that kp_i ≡ kp_j).

Definition 3

A user template (profile) UP is considered to store the m keystroke timeseries subsequences of a user, say UP = {ks_1, ks_2,…, ks_m}. The template UP is recorded in a database called user enrolment.

C. Data Preprocessing

The initial phase of data investigation is the information pre-processing step which is used to collect the keystrokes which are dependent on time and extracted when the user is working with different software. Every events like the window change or any keyboard events (pressing or releasing the key) is recorded in the subsequent i-throw of the information which serves as input for the vector w_i,

$$w_i = [prefix, T_i, id] \quad (1)$$

where $prefix \in \{W, \text{KUT}, \text{KDT}, \text{KHT}, \text{FT}\}$ is an event's prefix-type, T_i is an event's timestamp from definition 1 and id is a key identifier. The vectors w_i are filtered in the pre-processing phase in such a way that keystrokes from the recognized software are identified and used for further analysis, where rest of the keystrokes are omitted. For example, the following labels are identified by the activity registration software: "Word", "Excel", "Chrome", "Firefox", "Internet Explorer", "MATLAB", "Notatnik", "Opera", "Outlook", "Thunderbird". The target of the information pre-processing step is experimented using MATLAB and the remaining keystrokes are removed. This principle is shown in Fig. 2.

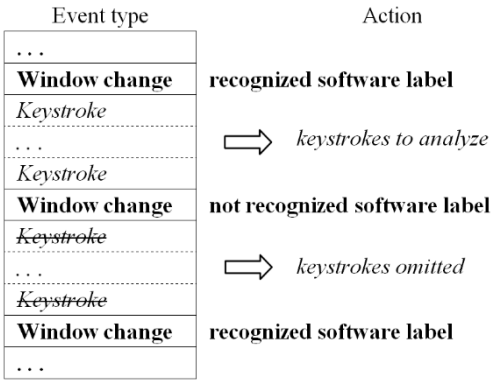

Fig. 2. Principle of data pre-processing

D. Outliers Elimination

The investigation of data is carried out with certain restrictions projected over the key events so as to remove the outliers. A user can make use of the keys of a keyboard freely. However in the data analysis procedure, it was assumed to be the keystrokes forming the event sequences. The procedure for outlier's removal is as follows:

1) A forthcoming event may not happen after the time T_{max} and
2) The consecutive events with the number of occurrences (met with the conditions at the initial level) could not be lesser than c_{min}.

The event is summed up along with the sequence only if the elapsed time fall as the previous event does not exceed the maximum allowed time T_{max} between two events. A sequence in which number of elements meeting the initial condition and do not attain the minimum count of elements c_{min}, is omitted in further analysis.

E. User Profiling Similarity Measurement (UPSM)

The formulated system is the process whereby same pairs of keystroke series of time sub-sequences are correlated from the pre-processed information, or with the keystroke series of time sub-sequences held in UP, or previously identified subsequence in the existing stream of data. Let us consider that there are two keystroke time series subsequences ks_1 and ks_2 of the similar length for different users, the easy way to find their similarity is by measuring the Euclidean Distances (ED) between each keysequence points in ks_1 and the related point in ks_2 for different users. The offsets are not considered by the ED measurement that lie in the time series pair.

Dynamic Time Warping (DTW) has been described by considering two time series ks_1= {kp_1, kp_2,…, kp_i,…, kp_x} and ks_2 = {kq_1, kq_2,…, kq_j,…, kq_y} , where x and y are the lengths of the two series correspondingly. The time series are shown as keystroke time series and the data values depicted by each point $kp_i \in ks_1$ and each point $kq_j \in ks_2$ are flight time values

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

(FT) under the formulated CKA technique described in this work. A matrix M of size (x − 1) × (y − 1) is then created with the value on each cell $m_{i,j} \in M$ is computed between the distance from point $kp_i \in ks_1$to point $kq_j \in ks_2$:

$$m_{i,j} = \sqrt{(ks_i - ks_j)^2} \quad (2)$$

The distance among two time series of various users is computed with the help of the Euler Movement Firefly Algorithm (EMFA). Once the distance is determined the timestamps of keystrokes pair with the identifier 'id' is subjected to the cluster 'CL_{id}'. After the process of enrolment, the same user is combined to CL_{id} until it reaches the target samples. For instance, if an identifier 'L1' is subjected to the element 'id' of a vector (i.e, id=L1) then the vector v_{L1} will be added to the clusters and finally described as Cl_{keys}.

Euler Movement Firefly Algorithm (EMFA)

Firefly Algorithm (FA) is utilized for the evaluation of the similarity among variety of sequences of a user. This work FA is improved by an implicit backward Euler movement for similarity computation between users. Hence in every iterative process of the keystroke dataset, a linear system of operations must be dealt with, to identify the innovative cluster positions of the time sequences of the user. To validate the authentication process of the implicit movement is applied for the regular operations of optimization. The process of optimization is performed by Firefly Algorithm to predict the significance of the implicit movements.

Firefly Algorithm [22] was designed by the characteristic nature of the fireflies. The significant feature of these insects is the potential to glow; which is used for the attraction of the similar user profiles from the keystroke dataset. The FA spreads the instructions of the fireflies to turn towards the greater value of authentication in a Novel Fuzzy Kernel Support Vector Machine (NFKSVM) classifier. The '$m_{i,j}$' to be computed form an $\mathbb{R}^{m \times l}$ in which length 'l' represents the count of the users within the dataset containing keystroke between 'i' and 'j' users. As per the position of the vector of a user $UP_{i,j}$, a single i has a target operation whose value 'f' identifies the brightness attribute $In_{i,j}$ of the individual. According to the measurement of the similarity value, the brightness attributes value In_{ij} can basically be shown as:

$$In_{i,j(l)}(UP_{i,j(l)}) \propto f\left(UP_{i,j(l)}\right) \quad (3)$$

During the evaluation process of similarity, every user i and j with a single sequence number moves with a solution heading towards each bright user profile with similarity j, in order to enhance the highest value of similarity operation. The attractive $atr_{i,j(l)}$ for user i and user j focuses on the computation of distance $m_{i,j(l)}$with both users position and is detected by

$$atr_{i,j(l)} = atr_0 e^{-\gamma m^2_{i,j(l)}} \quad (4)$$

The parameter atr_0 is related to the attractiveness of the distance between the similar users and are shown by $m^2_{i,j(l)}$=0 and γ is assigned to be the light absorption coefficient. The distance is generally evaluated by the Euclidean Distances (ED)

$$m^2_{i,j(l)} = \left|\left|ks_{i,l} - ks_{j,l}\right|\right| \quad (5)$$

For single i, its direction of a brighter individual is measured by

$$UP^{t+1}_{il} = UP^t_{il} + atr_0 e^{-\gamma m^2_{i,j(l)}}\left(UP^t_{jl} - UP^t_{il}\right) + SZ_t \epsilon^t_{il} \quad (6)$$

In equation 6, the third term is correlated to movement which is random in space with the step size SZ_t and a vector ϵ^t_{il} is the random numbers derived from the Gaussian distribution. FA excluding the step in random compares between the processes of regressive diffusion having the form

$$\frac{\partial X}{\partial t} = atrl(UP) \quad (7)$$

with $E(UP_i) = \left(UP^t_{jl} - UP^t_{il}\right)$. The equation of diffusion (7) can be combined with the help of a scheme called Euler producing,

$$UP^{a+1} = UP^a + atrdtE(UP^a_l) \quad (8)$$

The following equation (9) is related to the equation of movement (6) in the basic FA without random movement

$$UP^{a+1}_i = UP^a_i + atr(UP^a_{jl} - UP^a_{il}) \quad (9)$$

The individual's position at an iteration level of a+1 will rely on the other individual position at an iterative level t. To enhance the behaviour of convergence of FA, the equation (7) is resolved by a process of implicit integration [23]. The individual's position at the level of iteration t+1, thus rely on the other individual's position at an iteration t+1. The integration of equation (7) by the implicit backward Euler method is

$$UP^{t+1} = UP^t + atrdtE(UP^{t+1}_l) \quad (10)$$

which results to a Linear System of Equation (LSE),

$$(I - atrdtE)(UP^{t+1}_l) = UP^t_l \quad (11)$$

$$SZUP^{t+1} = UP^t \quad (12)$$

where I is denoted as the identity matrix. The matrix size SZ=I− atrdtE equals $\mathbb{R}(m \cdot l) \times (m \cdot l)$, with the number of individuals l . Once after resolving the individual's random movement:

$$UP_i = ip_i + SZ\epsilon \quad (13)$$

The process is iterated for a specified number of times as the demanding features of time series sequences are processed.

F. User Profiling Continuous Keystroke Authentication (UPCKA)

UPCKA is performed using a Novel Fuzzy Kernel Support Vector Machine (NFKSVM). It is formulated with optimal kernel matrix which is created with the combination of output of two kernels with the help of SVM model. An appropriate NFKSVM model proves that the kernel matrix gained is positive and seems to be optimal regarding the keystroke dataset. Let $X = (x_i, y_i)$, i = 1, ...,m be a keystroke dataset with m sample numbers having the instance–label pairs, where $x_i \in \mathbb{R}^n$ states the input vector and y_i shows the relative class label if the user is authorized or not. Let $X \subseteq \mathbb{R}^n$ be the space of input, Y the space of output and h: X → Y the classifier. The output space in the binary classification issue is Y={±1}. The algorithm called SVM learning identifies the authentication process along with the parameters.

Let us consider that the gradient vector and γ^*(scalar bias) with the help of keystroke dataset is used to segregate the data points of keystrokes accurately. Linear inseparability issue is used to track the data points of keystrokes into a greater dimensional space F which is termed as feature space to the keystroke data points X [24].

Keystroke data points are embedded to Kernel methods ([24-25]) into $\mathbb{R}$ in expecting for linear relation. The technique which states this process of tracking is termed as kernel function, K: X×X → $\mathbb{R}$. With the actual feature space of sufficient dimensionality, any data set with consistency are split. Describing the following reproducing kernel map as $\phi : x \rightarrow K(\cdot, x)$, a function $K(\cdot, x)$ is related to the data points of keystroke $x \in X$ in the input keystroke space.

$$\min \frac{1}{2}\left|\left|We\right|\right|^2 + C\sum_{I=1}^{m}\xi_i \; s.t.\, y_i(we^T\phi(x_i) + \gamma) \geq 1 - \xi_i, i = 1, \ldots m, \xi_i \geq 0, i = 1, \ldots m \quad (14)$$

Where the trade-off parameter C maintains penalty over patterns that are misclassified or near to the SVM decision boundary, restricting the difficulty of the decision operation vs. the reduction of errors during the training process ξ_i. The use of equation (14) is the formulation of duality ([24])

$$\max W(\alpha) = \sum_{I=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\, y_i y_j K_{ij} \;\; s.t \sum_{I=1}^{m}\alpha_i\, y_i = 0, 0 \leq \alpha_i \leq C \quad (15)$$

Where α states the Lagrangian multiplier. Kernel method is extensively utilized in SVM classifier and it depends on the information which is shown only through the dot products among the arguments of the user. The plan behind the kernel operation is the benefaction of a link from the linear information to the non-linear keystroke of separable information. The two advantages of the kernel operations are: (i) it has the capacity of creating non-linear limits of the decisions are taken by the classifier and (ii) the kernel operation is subjected to allow the user apply the keystroke information into the classifier leaving out the vector of a fixed-dimensional space notation. Exponential and tangential are the two kernel operations which are used by the keystroke data vector, where the classifier effectively verifies the keystroke data of the similar kind of user. Thus, the innovatively constructed kernel is shown by

$$K_{ij} = a_1 \exp\left(-(x_i * x_j)\right) + a_2 \tanh\left(-(x_i * x_j)\right) \quad (16)$$

Where $a_1$and a_2 are the weights that are evaluated by using fuzzy triangular membership operation. The exponential kernel operation is stated to be similar to that of Gaussian kernel operation without the procedure of square norm function. It is best suitable for radial-basis kernel operation. The tangential kernel is called as sigmoid or multilayer perceptron kernel operation. This kernel operation is well known for SVM classifier. The triangular membership function is used to determine the weights of classifier with user defined by three limits, like lower value, middle value and upper value. The triangular membership function consumes very minimal duration and provides a comparatively better efficiency than the other membership functions. The membership function is represented like:

$$\phi = x_i * x_j \quad (17)$$

Where ϕ is the product of 2 $x_i * x_j$ vectors. a_1 and a_2 values are evaluated through the triangular membership operation. The triangular membership operation is represented by

$$TF(a_1, p, q, r) = \begin{cases} 0, \phi \leq p \\ \dfrac{\phi - p}{q - p}, p \leq \phi \leq q \\ \dfrac{r - \phi}{q - p}, q \leq \phi \leq r \\ 0, q \leq \phi \end{cases} \quad (18)$$

Where p shows the lesser value of ϕ, r is the highest value of ϕ and q is the middle value which is identified through p and r. Thus, the value a_1 is evaluated by the membership function. Then, the a_2 value is derived as $a_2 = 1 - a_1$. The new kernel operation is applied innovatively through the values of a and b, which are combined into spherical SVM classifier for the improved performance of authentication.

IV. RESULTS AND DISCUSSION

A. Dataset Description

Buffalo dataset [27] contains keystrokes based on transcription as well as free text typing. The dataset segment is extracted with the help of various types of keyboards in all sessions. Moreover the coordinated information and its associated events data are obtained along with the keystroke dataset. The dataset was collected from 148 subjects in 3 separate laboratory sessions (session 0, 1 and 2). Each session was about 50 minutes, and contains 5.7k keystrokes. The average time interval between the sessions was 28 days. Four different types of keyboards were used across sessions. There were two subsets of users divided based on the keyboards they use. i) Baseline subset with 75 users using the same type of keyboard across 3 sessions. ii) Rotation subset with 73 users using 3 different types of keyboard across 3 sessions. Each subject was asked to perform Task 0 and Task 1 in each session. Task 0 was Transcription of

Steve Jobs' commencement speech split in three pieces. Task 1 was free text questions with one being 2 survey style questions plus one scene description and the other was to mimic the realistic daily working scenario, e.g., checking email, sending email and web surfing. In this work, we have taken the samples of session 1 in baseline subset.

B. Evaluation

This section reports on the results obtained when the proposed method is compared with the existing methods of the form frequently referenced in the literature. The evaluations of the Buffalo dataset with authentication methods are presented in this section. The metrics used for the evaluation are: (1) False Match Rate (FMR), (2) False Non-Match Rate (FNMR), (3) Authentication Accuracy (Acc.), (4) Error and (5) F-Measure. FMR and FNMR are the standard metrics used to measure the performance of Biometric systems [28], although some researchers, in the literature, have used the terms False Acceptance Rate (FAR) and False Rejection Rate (FRR) instead. The experimental results are obtained, with respect to the above evaluation objectives.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

To evaluate the effectiveness of user authentication using the proposed approach, for each dataset and each subject, the continuous typing process was simulated by presenting the keystroke dynamics in the form of a data stream. Comparison of every subsequence with a previously stored subsequence recorded whether this was a True Positive (TP), False Positive (FP), False Negative (FN) or True Negative (TN). In this manner, a confusion matrix was built up from which FMR, FNMR, Acc. , Error, F-Measure could be calculated (using Eqs. (19), (20), (21), (22) and (23) below).

$$FMR = FP/ FP + TN \quad (19)$$

$$FNMR = FN/ FN + TP \quad (20)$$

$$Acc. = TP + TN/ TP + FP + FN + TN \quad (21)$$

$$Error = 1-Acc \quad (22)$$

$$F\text{-}Measure = 2.P.R/ (P+R) \quad (23)$$

The results are presented in Table 1 by comparing the proposed UPCKA approach with the other approaches. From the table, it can be observed that the proposed UPCKA obtained a much better performance than the IKCA (the baseline approach) and SVM approach, with respect to dataset.

Table-I: Metrics comparison between proposed and existing methods

Methods	FMR	FNMR (%)	Acc. (%)	Error (%)	F-Measure (%)
SVM	0.1621	24	76	24	75.6743
IKCA	0.1163	17.333	82.667	17.33	82.4997
UPCKA	**0.0626**	**9.33**	**90.667**	**9.33**	**90.3545**

Fig. 3 shows the performance comparison results of the SVM, IKCA and UPCKA authentication methods in terms of FMR. The proposed UPCKA method produced lesser FMR results of 0.0626, where the other methods such as SVM and IKCA produced higher FMR results of 0.1621 and 0.1163 respectively.

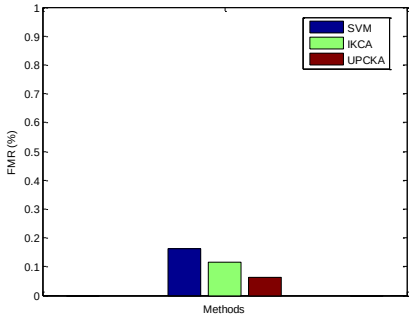

Fig. 3. FMR results of the authentication methods

Fig. 4 shows the performance comparison results of the SVM, IKCA and UPCKA authentication methods in terms of FNMR. The proposed UPCKA method produced lesser FNMR results of 9.33%, whereas the other methods such as SVM and IKCA produced higher FNMR results of 24% and 17.33% respectively.

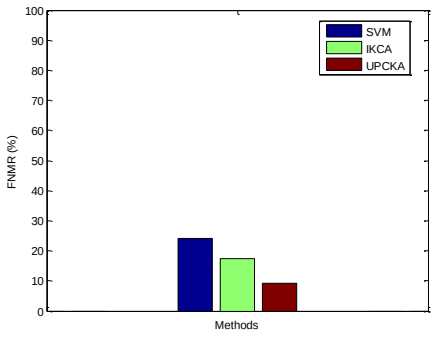

Fig. 4. FNMR results of the authentication methods

Fig. 5 shows the performance comparison results of the SVM, IKCA and UPCKA authentication methods in terms of accuracy. The proposed UPCKA method produced higher accuracy of 90.667%, whereas the other methods such as SVM and IKCA produced lesser accuracy of 76% and 82.667% respectively.

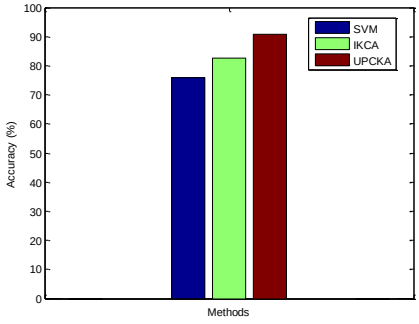

Fig. 5. Accuracy results of the authentication methods

Fig. 6 shows the performance comparison results of the SVM, IKCA and UPCKA authentication methods in terms of error. The proposed UPCKA method produced lesser error rate of 9.33%, where as the other methods such as SVM and IKCA produced higher error rate of 24% and 17.33% respectively.

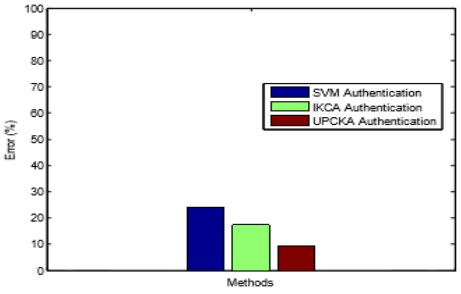

Fig. 6. Error results of the authentication methods

Fig. 7 shows the performance comparison results of the SVM, IKCA and UPCKA authentication methods in terms of F-Measure. The average results of TPR and TNR is measured using this metric. The proposed UPCKA method produced higher F-Measure results of 90.3545%, where as the other methods such as SVM and IKCA produced lesser F-Measure results of 75.6743% and 82.4997% respectively.

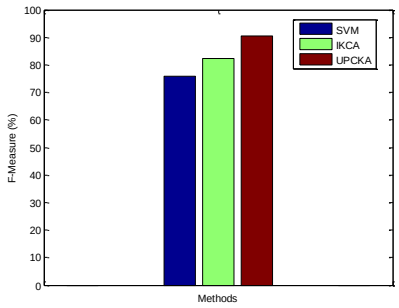

Fig. 7. F-Measure results of the authentication methods

V. CONCLUSION AND FUTURE WORK

In this paper, a novel User Profiling Continuous Keystroke Authentication (UPCKA) is proposed for authentication. Firstly, User Profiling Similarity Measurement (UPSM) is proposed in order to find the similar pairs of keystroke time series subsequence from pre-processed data, either with the keystroke time series subsequence held in user profiles or previously identified subsequence in the current data stream. In UPSM, similarity between the user profiles is measured based on the Euler Movement Firefly Algorithm (EMFA). Then, Novel Fuzzy Kernel Support Vector Machine (NFKSVM) classifier is proposed to use sequences of keystroke dynamics in the form of time series. This classifier is used to monitor such time series as a continuous data stream, by periodically extracting subsequence from these time series and authenticating the subsequence. NFKSVM classifier optimal kernel matrix is formed by combining the results of the two kernels, generated using a SVM model. A suitable model NFKSVM assures that the obtained kernel matrix is positive and is optimal with respect to the keystroke Buffalo dataset under consideration. In terms of accuracy, the best overall accuracy of 90.38% achieved when compared to other methods. In future, finding the number of clusters should be extended by applying other algorithms in order to improve its accuracy. The authors have also intended to focus on time series samples which are made up of hold time and flight time.

ACKNOWLEDGMENT

Authors wish to thank the Management and Principal for their kind support and providing the infrastructure facilities required for this research work. (DrNGPASC2019-20 CS003)

REFERENCES

1. Windley PJ (2005) Digital identity. O'Reilly Media, Sebastopool
2. Hosseinzadeh D, Krishnan S (2008) Gaussian mixture modelling of keystroke patterns for biometric applications. IEEE Trans SystMan Cybernetics Part C: Appl Rev 38(6):816–826
3. Moskovitch R, Feher C, Messerman A, Kirschnick N, Mustafic T,Camtepe A, Lohlein B, Heister U, Moller S, Rokach L, EloviciY (2009) Identity theft, computers and behavioral biometrics. IEEE International conference on intelligence and security informatics, 2009. ISI '09. pp 155–160.
4. Zhong, Y., Deng, Y. and Jain, A.K., 2012, Keystroke dynamics for user authentication. IEEE computer society conference on computer vision and pattern recognition workshops,pp. 117-123.
5. Pisani, P.H. and Lorena, A.C., 2013. A systematic review on keystroke dynamics. Journal of the Brazilian Computer Society, 19(4), p.573.
6. Ogihara A, Matsumuar H, Shiozaki A (2006) Biometric verification using keystroke motion and key press timing for ATM user authentication. Intelligent Signal Processing and Communications. ISPACS'06. International Symposium on. IEEE, pp 223–226.
7. Syed Z, Banerjee S, Cukic B (2014) Normalizing variations in feature vector structure in keystroke dynamics authentication systems. Softw Qual J:1–21.
8. Messerman, A., Mustafic, T., Camtepe, S.A., Albayrak, S.: Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. In: 2011International Joint Conference on Biometrics (IJCB), pp. 1–8. IEEE (2011)
9. Bours,P.:Continuous keystroke dynamics: A different perspective towards biometric evaluation. Inf. Secur. Tech. Rep.17(1), 36–43 (2012).
10. Kang, P. and Cho, S., 2009. A hybrid novelty score and its use in keystroke dynamics-based user authentication. Pattern recognition, 42(11), pp.3115-3127.
11. Ahmed, A.A. and Traore, I., 2014. Biometric recognition based on free-text keystroke dynamics. IEEE transactions on cybernetics, 44(4), pp.458-472.
12. Hocquet, S., Ramel, J.Y. and Cardot, H., 2007, User classification for keystroke dynamics authentication. In International Conference on Biometrics (pp. 531-539). Springer, Berlin, Heidelberg.
13. Hu J., Gingrichd., Sentosaa. A K-Nearest Neighbor approach for user authentication through biometric keystroke dynamics. IEEE International Conference on Communications, 2008. pp. 1556–1560.
14. Giot R, El-Abed, M, Rosenberger C (2009) Keystroke dynamics with low constraints SVM based passphrase enrollment. In: IEEE 3rd International Conference on biometrics: theory, applications, and systems, 2009. BTAS 2009, pp 1–6.
15. Alshehri A, Coenen F, Bollegala D (2016a) Towards keystroke continuous authentication using time series analytics. In: Proc. AI 2016, Research and Development in Intelligent Systems XXXIII. Springer, New York, pp 325–338
16. Alshehri A, Coenen F, Bollegala D (2016b) Keyboard usage authentication using time series analysis. In: International Conference on Big Data Analytics and Knowledge Discovery. Springer, New York, pp 239–252.
17. Patil, R.A. and Renke, A.L., 2016. Keystroke dynamics for user authentication and identification by using typing rhythm. International Journal of Computer Applications, 144(9), pp.27-33.
18. Wangsuk, K. and Anusas-amornkul, T., 2013. Trajectory mining for keystroke dynamics authentication. Procedia Computer Science, 24, pp.175-183.
19. Alshehri, A., Coenen, F. and Bollegala, D., 2018. Iterative Keystroke Continuous Authentication: A Time Series Based Approach. KI-Künstliche Intelligenz, pp.1-13.
20. A.V. Senthil Kumar and M. Rathi, "Keystroke Dynamics – A Behavioral Biometric Model For User Authentication In Online Exams," in Biometric Authentication in Online Learning Environments, IGI Global, 2019, pp. 183–207, DOI: 10.4018/978-1-5225-7724-9.ch008.
21. M. Rathi, Dr. A. V. Senthil Kumar, "Exploration of Keystroke Dynamics Based Authentication on Fixed-Text and on Free-Text", International Journal of Computer Sciences and Engineering, Vol.7 , Issue.1 , pp.807-812, 2019 .
22. X.-S. Yang, Nature-Inspired Metaheuristic Algorithms, Luniver Press, 2008.
23. Bartz, R., Fiebig, S., Franke, T., Falkenberg, P. and Axmann, J., 2017, Enhanced firefly algorithm with implicit movement. In World Congress of Structural and Multidisciplinary Optimisation (pp. 700-709). Springer, Cham.
24. Shawe-Taylor J,Cristianini N. Kernel methods for pattern analysis. Cambridge: Cambridge UniversityPress;2004.
25. Conforti, D. and Guido, R., 2010. Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. Computers & Operations Research, 37(8), pp.1389-1394.
26. Kevin S. Killourhy and Roy A. Maxion. Free vs. transcribed text for keystroke-dynamics evaluations. In Learning from Authoritative Security Experiment Results (LASER-2012), July 18–19, 2012, Arlington, VA, 2012. ACM Press.

27. Sun, Y., Ceker, H. and Upadhyaya, S., 2016, Shared keystroke dataset for continuous authentication. IEEE International Workshop on Information Forensics and Security (WIFS) ,pp. 1-6.
28. Unar JA, Seng WC, Abbasi A (2014) A review of biometric technology along with trends and prospects. Pattern Recognition 47(8):2673–2688

AUTHORS PROFILE

Rathi. M received her Master of Computer Applications (MCA) degree from Bharathiar University in 2003 and M.Phil Computer Science from Bharathiar University in 2005. She is currently pursuing part-time Ph.D in Hindusthan College of Arts and Science and working as Assistant Professor in Department of Computer Technology, Dr. NGP Arts and Science College, Coimbatore since 2016. She has 12+ years of teaching experience and has published papers in reputed international journals. Her research interests include Data Mining, Information Security.

Dr. A. V. Senthil Kumar is a Professor and Director of Department of Computer Applications, Hindusthan college of Arts and Science, Coimbatore. He obtained his Ph.D in Computer Science in 2009. He has to his credit 9 Book Chapters, 165 papers in International Journals, 2 papers in National Journals, 30 papers in International Conferences, 5 papers in National Conferences, and edited 7 books published by IGI Global, USA. He is an Editor-in-Chief for 4 International Journals and Key Member for India, Machine Intelligence Research Lab (MIR Labs). He is an Editorial Board Member and Reviewer for various International Journals. He is also a Committee member for various International Conferences.