ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

Volume 3, Special Issue 3, March 2014

2014 International Conference on Innovations in Engineering and Technology (ICIET'14)

On 21st&22ndMarch, Organized by

K.L.N. College of Engineering, Madurai, Tamil Nadu, India

Authentication of User Based On Mouse-Behavior Data Using Classification

G. Muthumari, R. Shenbagaraj, M. Blessa Binolin Pepsi

PG Scholar, Dept of Computer and Communication, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.

Assistant professor, Dept of Information technology, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.

Assistant professor, Dept. of Information technology, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.

Abstract- To authenticate a user, classification based on mouse operating behavior is proposed. The data includes co-ordinate axes, time Stamp value and mouse operations. The holistic and procedural features are extracted from the mouse behavior data. The feature distance vector is calculated using Manhattan and Dynamic Time Warping method based on feature vector for representing the original mouse feature space. Kernel PCA method is used to reduce the dimensionality of the feature distance vector. Then the one-class Support Vector Machine classifier is applied on the distance-based eigenspace feature to analyze whether the input sample is legitimate user or an imposter. The performance of the proposed method is measured by False Acceptance Rate and False Rejection Rate. The test result proves that, the proposed method KPCA with one-class Support Vector Machine provides low error rates with good accuracy than the existing method PCA with classifier.

Key words -Authentication, mouse dynamics, kernel Principal Component Analysis, Support Vector Machine, Dynamic Time Warping**.**

I.INTRODUCTION

Authentication is to prove the identity of a person. Biometric is defined as an automated use of a collection factor describing human behavioral or physiological characteristic to establish or verify a precise identity [5]. Biometric is used in computer science as form of identification and access control. Physiological biometric including finger prints, iris recognition, face recognition and palm print measured from the human body. Behavioral biometrics is related to pattern of behavior of a person including keystroke dynamics, gait and voice [8]. It requires specialized hardware. An behavior biometric, mouse dynamics does not need any specialized hardware to capture the data [16].

The authentication is done based on the users behavioral of mouse movements and clicks which is stated as mouse dynamics [6]. The types of decision made by biometric system is classified as genuine or imposter. Mouse dynamics assess the behavior of a user as a biometric. The extracted features of mouse movements are analyzed with legitimate user's profile. This authenticates the user and also monitored during the computer usage which can be later used for intrusion detection. If the match is not authenticated, then the access will be denied [11]. The performance of biometric system can be measured by two factors: False Acceptance Rate (FAR) and False Rejection Rate (FRR)**.** The mouse dynamics is less obtrusive and no need specialized hardware compared with other biometric such as finger print, Iris recognition and so on. The user log on to the system provides the user name and password by performing some mouse operation task, from that features are extracted and store it as legitimate user profile. The authentication is done during user's subsequent operation to enforce the full session identity [13][14]. The objective of the mouse based biometric, the user authentication task is done based on the mouse operating style of user. The paper includes the authentication based on the mouse behavior data consisting mouse operations, co-ordinates and time stamp. The features are extracted and distance vector is calculated. The dimensionality of input is reduced using the Eigen space transformation. This data is assessed by classification to perform the authentication task. Finally the performance is measured by the various parameters and the statistical analysis of the result states

that evaluation techniques are good compared to other techniques.

II.RELATED WORK

In this section, we describe the existing work related to the mouse based biometric authentication.

Fred and Gamboa [1], presents the web-based user authentication system based on mouse-dynamics. The mouse operation samples are analyzed and the spatial and temporal features are extracted. The mouse behavior data is classified using Sequential classifier. The classifier decides to accept or reject the identity based on the genuine and impostor distribution. The classifier performance is measured by greedy search algorithm. The authentication technique is inexpensive and reported with high error rates and low accuracy.

Brodley and Pusara [2], proposed decision tree as a classifier to analyze the given sample as legitimate user or imposter. Decision tree is a tree structure, which consists of internal node and branches in the tree. The speed, distance and angle of the mouse data are extracted. The internal node represents the test on the features. The branch in the tree indicates the outcomes of the test and leaf node represents the class label. The tree size can be resized to minimize the classification error. This researcher reported 12% FAR and 6% FRR.

Hashia and Pollett [3], deals with results on mouse behavior based authentication. The speed, deviation and angle are extracted as features. The authentication task was performed by Distance based classifiers, which is used for comparing the verification data with enrollment data. The classifier can be built based on the scaled Euclidean distance using both legitimate users and impostors. These researchers reported the result with 15% FAR and 20% FRR.

Jorgenson and Yu [4], proposed the Machine learning algorithm for classification. The Velocity mean, Velocity standard deviation and angle are extracted as features. The mouse operation samples are divided in to training and testing group and assigned the label. The Machine learning algorithm like neural network used to analyze whether given input sample is legitimate user or imposter. The neural network consists of input layer, hidden layer and output layer. This machine learning algorithm consists of training and testing phase. These results achieved the error rates are somewhat high.

The results of the related work illustrates with error rates and moderate accuracy. A technique related to the work is proposed which defines procedural features, holistic features and Kernel Principal Component Analysis for Eigen space computation with One-Class Support Vector Machine for classification.

III.SYSTEM DESCRIPTION

The mouse dynamics based user authentication is done by following methodologies on mouse behavior dataset.

1. Feature extraction
2. Eignspace Computation
3. Classification

The feature extraction consists of holistic and procedural features. In this phase the mouse features are extracted from mouse-behavior data and the distance matrices are calculated as distance based feature vector. Kernel PCA is applied on the distance based feature vectors to determine the predominant features for reducing the dimensionality and user's profile is built using one class Support Vector Machine classifier. The classification consists of training and testing phase. In training phase the legitimate user's profile was built based on the mouse operations. In testing phase the new mouse features are compared with legitimate user's profile. The proposed system provides the low error rates and high accuracy than the existing approaches.

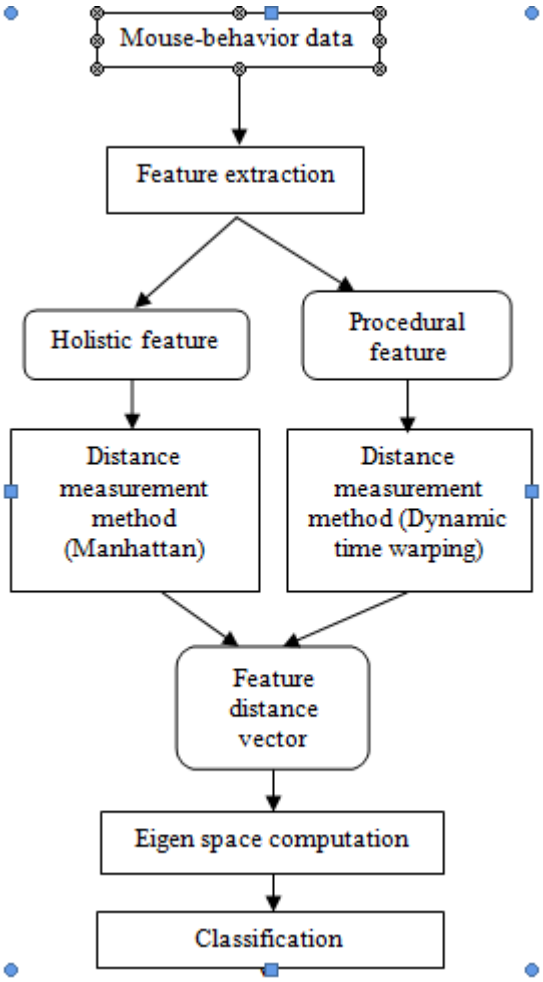

Fig1.System Description

The figure1 shows that the overview of the proposed system. The mouse behavior data with co-ordinate axes, time stamp and clicks are given as input to the proposed system. The holistic and procedural features are extracted from the mouse behavior data.

The distance measurement method and Kernel PCA method are used to solve the behavioral variability and dimensionality problem respectively. The output of Kernel PCA method is given input to the one-class SVM

classifier which is used to perform the authentication task. This classifier classifies the given input sample is legitimate users or imposters. The classifier performance is measured by False Acceptance Rate and False Rejection Rate.

IV.IMPLEMENTATION

This section describes the implementation of proposed work. The proposed system implemented with the following modules:

- Feature Extraction
- Distance Measurement
- Eigen space Computation
- Classification

A. *Feature Extraction*

The mouse behavior data consist of four attributes: type of mouse operation, x-coordinate, y-coordinate and time stamp of the mouse operation. The types of mouse operations are single click or double click and mouse movements. These are converted into set of features called feature vector [16]. The feature vector is used to characterize the user's unique mouse behavior.

1) Holistic Feature

This feature characterizes the properties of the mouse operations such as single-click and double clicks statistics which includes,

- *Mean and standard Deviation:*
  Mean and standard deviation of overall time of single click operation
- *Movement offset*:
  The distance between the starting x & y coordinates and ending x & y coordinates of the mouse movement
  Movement offset=$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
- *Movement Elapsed Time*:
  The time difference between the starting and ending point of mouse movement.
  MET = Timestamp (end)-Timestamp (start)

2) Procedural Feature

- *Speed curve*: The ratio of the movement offset to the movement elapsed time, for each movement.
- *Acceleration curve*: The ratio of the speed value to the movement elapsed time.

B. *Reference Feature Distance Vector Generation*

The feature vector consists of both holistic and procedural features; in this vector the distance measurement is applied to address the problem of behavioral variability. Consider x_i (i=1, 2…n) is the feature vector extracted from i^{th} training sample, n is the number of training sample for one subject.

Distance Measurement Method

The distance measurement method used to address the behavioral variability i.e., the human can't produce the same behavior twice under the same condition. There are two types of distance measurement methods are used to eliminate this behavioral variability in the mouse data points produced in such a way using,

- Manhattan distance method
- Dynamic Time Warping method

Manhattan Distance Method

This method used to measure the distance between two data points [17]. The distance between two vectors is the sum of the difference between all points in the corresponding vector. This method used to measure the distance between holistic features and also it preserve the physical interpretation of this features.

Dynamic Time Warping method

This method used to measure the similarity between the two sequences; it may vary in time or speed. The procedural features such as movement speed curve and acceleration curve of two data samples are not likely to consist of the exactly same number of points. So use the DTW [10] method to find whether these samples are generated by the same or by different user.

The following steps are used to calculate the feature vector,

Step1:

The feature distance vector of holistic feature between all pairs of training sample using Manhattan distance measurement method is found by using the equation(1),

$$x^H_{i,j}=|f_i^H-f_j^H| \quad (1)$$

The feature distance vector of procedural feature between all pairs of training sample using dynamic time warping method is calculated using the equation(2),

$$x^p_{i,j}=DTW(f_i^p, f_j^p) \quad (2)$$

Step2:

The procedural and holistic features obtained are concatenated to calculate feature distance vector.

$$x_{i,j}=[x_{i,j}^P, x_{i,j}^H] \quad (3)$$

Step3:

The arithmetic mean difference for all distance vectors and the reference vector with minimum mean distance is found.

Step4:

Final feature distance vector X_i is calculated between the reference vector and feature vector. Now the feature distance vector Xi ($X_i \in R^d$, i=1, 2…n) is feature distance vector of i^{th} training sample.

C. *Eigenspace Computation*

The Kernel Principal Component Analysis (KPCA) used to reduce the dimensionality and retain the variation in the data set as much as possible .The feature distance vector can't be directly given as input to the classifier because of high dimensionality. So the Kernel PCA is addressed to reduce the dimensionality [7]. This method used to convert the number of correlated variable into number of uncorrelated variable in the feature space. It is a mathematical technique used to solve the Eigen values and Eigen vectors of a square symmetric matrix with sums of squares and cross products. In the Eigen space computation, calculated the largest eigenspace of feature distance vector .This largest eigenspace is called the principal component.

Kernel PCA

The PCA is traditional method based on the linear principle. But the kernel PCA method is used to represent

the non linear structure. The kernel PCA [18] is done in two steps:

- Kernel PCA training
- Kernel PCA projection

Kernel PCA training

In Kernel PCA training calculate the Eigen vector of feature distance vector for each subject. Let consider, Xi ($X_i \in R^d$, i=1, 2…n) be the i^{th} feature distance vector in the training sample and n is the number of such vector.

First the feature distance vector nonlinearly mapped in the hyper dimensional feature space.

$$\Phi: X_i \in R^d \quad (4)$$

The mean and covariance matrix of each feature distance vector in the feature space calculated as,

Mean

$$m_\Phi = \frac{1}{n}\sum_{i=1}^{n} \Phi(X_i) \quad (5)$$

Covariance matrix

$$\sum \Phi = \frac{1}{n}\sum_{i=1}^{n} \bar{\Phi}(X_i)\bar{\Phi}(X_i)^T \quad (6)$$

Where,

$$\bar{\Phi}(X_i) = \Phi(X_i) - m_\Phi$$

The Eigen value problem for eqn (6) is solved as,

$$\lambda V = \sum_\Phi V = \frac{1}{n}\sum_{i=1}^{n} \bar{\Phi}(X_i)^T V \bar{\Phi}(X_i) \quad (7)$$

Where,

$\lambda > 0$ and $V \neq 0$.

Then kernel matrix calculated by the inner product between the feature subspaces as follows,

$$K_{ij} := \left(\bar{\Phi}(X_i).\bar{\Phi}(X_j)\right) = k(X_i, X_j) \quad (8)$$

Then the Eigen value problem for the kernel matrix is solved to get the Eigen vectors and Eigen values as,

$$\lambda\alpha = K\,\alpha \quad (\alpha = (\alpha_1, \ldots, \alpha_n)^T) \quad (9)$$

The smallest Eigen values and the corresponding Eigen vectors are eliminated.

Kernel PCA projection

In kernel PCA projection, Compute the projection on those Eigen vector i.e., project original feature in to the point p as follows,

$$P = V^k.\bar{\Phi}(X) = \sum_{i=1}^{n} \alpha_i^k K(X_i, X) \quad (10)$$

The value of V^k is calculated by taking the largest eigenvalues and the corresponding eigenvectors. This matrix is called transform matrix.

Now the point P contain the principal components i.e. contain the predominant feature. The principal component used to represent the mouse behavior in the K dimensional Eigen space. The value of K is usually much smaller than the dimensionality of the original feature space .The output of this method is given to input of the classifier.

Implementation of Kernel PCA

Step1

A kernel mapping K (x_m, x_n) is chosen for defining the kernel matrix. It is calculated by inner product between the feature spaces.

Step2

K valued based on training data $\{x_n, (n=1,2\ldots N)\}$ is obtained.

Step3

Eigen value problem of K to get λ_i and α_i is solved, i.e., to get Eigen value and corresponding Eigen vector.

Step4

For each given data point X, obtain its principal components in the feature space. The smallest Eigen value and Eigen vectors are eliminated using the eqn (10).

Step5

The projection on the low dimensional Eigen space is performed to obtain the principal components. Now the mouse data point represents low dimensional Eigen space than the input space.

D. Classification

The output of the Eigen space transformation is given as the input for the process of classification. The classifier assesses whether the given input is legitimate or imposter sample**.** The classifier technique, one class SVM is used for classification.

One class Support Vector Machine

The one class classifications build the model based on the legitimate sample and try to find the imposter sample using that model [15]. This type of classification tries to find the object that belongs to one class. Support Vector Machine [12] is a discriminative classifier formally defined by a separating hyper plane. SVM are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. The aim of SVM is find the hyper plane that separate the data points between two classes by the largest margin by solving the dual quadratic programming problem [9]. This classifier used to discriminate between the legitimate user and impostor. The dual quadratic programming problem solved as follows,

$$w(\alpha) = \sum_{i=1}^{l} \alpha_i\, \alpha_j K\left(X_i, X_j\right) \quad (11)$$

Where,

α-vector of Lagrangian multipliers

$K(X_i, X_j)$-Kernel function

The decision function

$$f(x) = sign\left(\sum_{i=1}^{l} \beta_i\ K\left(X_{i\,,} X_j\right) - \rho\right) \quad (12)$$

The eqn (12) produces positive if authorized sample is input otherwise false rejection cases occurs. It produces negative if unauthorized sample is input otherwise false acceptance cases will occurs [15]. This decision function used to discriminates between two classes.

Classification Algorithm

The mouse behavior data is analyzed and the holistic, procedural features are extracted and the Eigen space transformation is performed. The extracted principal component analysis is given as input to the classifier. The input samples are divided into testing and training samples, the classifier perform the following training and testing procedure on those samples as follows,

Step1:

The training samples were trained to build the model based on the legitimate user sample and the classifier use that model to recognize the given sample.

Step2:

The testing is done based on the ability of classifier to recognize the legitimate user by calculating the genuine score and imposter score.

V.PERFORMANCE

The classifier performance can be calculated by the False Acceptance Rate (FAR) and False Rejection Rate (FRR).

False Acceptance Rate

False acceptance is the error in biometric based authentication that causes an unauthorized person to be authenticated i.e., the system accept the user even though that user is an impostor. False acceptance rate is ratio of the number of false acceptance to the number of test sample of imposter.

$$FAR = \frac{Number\ of\ False\ Acceptance}{Number\ of\ imposter\ sample} \quad (13)$$

False Rejection Rate

The system fails to recognize an authorized person and reject that person as an imposter is called false rejection. False rejection rate is ratio of the number of false rejection to the number of test sample of the legitimate user.

$$FRR = \frac{Number\ of\ False\ Rejection}{Number\ of\ legitimate\ \ sample} \quad (14)$$

Statistical Analysis of the Result

The statistical analysis of the result can be calculated by the Half Total Error Rate (HTER).The HTER calculated by the average of false acceptance rate and false rejection rate.

$$HTER = \frac{FAR + FRR}{2} \quad (15)$$

Confidence Interval are calculated around the HTER with different accuracy level as follows

$$HTER \pm \sigma . Z_{\alpha/2} \quad (16)$$

Where,

$$\sigma = \sqrt{\frac{FAR(1-FRR)}{4.NI} + \frac{FRR(1-FAR)}{4.NG}}$$

$$Z_{\alpha/2} = \begin{Bmatrix} 1.645\ for\ 90\%\ CI \\ 1.960\ for\ 95\%\ CI \\ 2.576\ for\ 99\%\ CI \end{Bmatrix}$$

Where,

NI -the total number of imposter score and
NG- the total number of genuine score.

VI.RESULTS AND ANALYSIS

This section describes the results and analysis of the proposed work. The proposed method compares one class SVM with the PCA space and original features space by calculating FAR and FRR. The input is divided into testing and training samples, the results and analysis performed on the twenty training samples for each user and testing done with the ten testing samples of unique users. The False acceptance rate and false rejection rate are calculated from the output of the classifier as given in eqn (13) and eqn (14). The statistical analysis is performed by using the eqn (16). The following results illustrate the performance of the proposed method.

TABLE I
FAR AND FRR FOR THREE DIFFERENT FEATURE SPACE

Algorithm	**FAR (%)**	**FRR (%)**
One-class SVM with KPCA space	8.98	8.25
One-class SVM with PCA space	10.69	9.53
One-class SVM with original space	11.84	10.77

The Table I shows that the comparison of FAR and FRR for three different feature space such as KPCA feature space, PCA feature space and original feature Space. This results shows that, the proposed system produces the 8.98% FAR and 8.25% FRR. It provides low error rates than the previous method

.

TABLE II
HTER PERFORMANCE AND CONFIDENCE INTERVAL

Algorithm	HTER	Confidence interval(%) around HTER for		
		90%	95%	99%
1-Class SVM (KPCA)	8.6	±1.32	±3.14	±4.07
1-Class SVM(PCA)	10.11	±2.60	±3.47	±4.15
1-Class SVM (original space)	11.3	±2.93	±4.18	±5.33

The Table II shows that the Confidence interval around HTER for one-class SVM with KPCA, PCA and original space at different confidence interval. This result shows that the statistical analysis and provides the better performance at different confidence interval than the PCA and original spaces.

The KPCA with one class-SVM compared with the PCA with one class SVM by ROC curve. ROC curve is the graphical plot which illustrates the performance of the classifier.

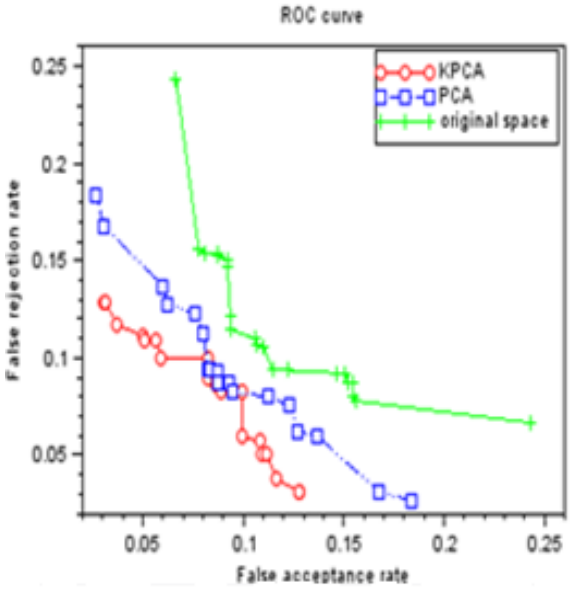

Fig 2 ROC curve for three different feature space

The Figure 2 shows that the KPCA with one-class SVM provides the better performance than the PCA and Original space with one-class SVM classifier.

VII.CONCLUSION

Mouse behavior based biometrics represents an emerging field of research which may used in the general areas of continuous user authentication, network forensics, and intrusion detection. Mouse behavior based user authentication is a simple technique that can perform the authentication task within a short period of time and provides the good accuracy. The newly defined procedural features and holistic procedural features are extracted from the mouse-behavior sample to characterize the user's unique data. The distance-based feature vector and Eigen space transformation used to reduce the dimensionality. Finally a one class classification algorithm is used for authentication task. . This algorithm used to perform the authentication task, which is more appropriate for mouse dynamics based user authentication in real applications. The performance of the proposed system provides 8.98% FAR 8.25% FRR. This test result proves that, the proposed method KPCA with one-class Support Vector Machine provides low error rates with good accuracy than the existing method PCA with classifier.

REFERENCES

[1] H. Gamboa and A. Fred, "A behavioral biometric system based on human computer interaction," in *Proc. SPIE*, Orlando, FL, 2004, pp.381–392.

[2] M. Pusara and C. E. Brodley, "User re-authentication via mouse movements," in *Proc. 2004 ACM Workshop on Visualization and Data Mining for Computer Security*, Washington, DC, 2004, pp. 1–8.

[3] S. Hashia, C. Pollett, and M. Stamp, "On using mouse movements as a biometric," in *Proc. Int. Conf. Computer Science and Its Applications*, Singapore, 2005, pp. 143–147.

[4] Z. Jorgensen and T. Yu, "On mouse dynamics as a behavioral biometric for authentication," in *Proc*. 6th *ACM Symp. Information, Computer and Communication Security*, Hong Kong, 2011, pp. 476–482.

[5] Y. Aksari and H. Artuner, "Active authentication by mouse movements," in *Proc. 24th Int. Symp. Computer and Information Science*, Guzelyurt, 2009, pp. 571–574.

[6] A.A.E.Ahmed and I. Traore, "A new biometric technology based on mouse dynamics," *IEEE Trans. Depend. Secure Comput.*, vol. 4, no. 3, pp. 165–179, Jul./Sep. 2007.

[7] Schölkopf, A. J. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel Eigen-value problem," *Neural Computat.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[8] P. Bours and C. J. Fullu, "A login system using mouse dynamics," in *Proc. 5th Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, Japan, 2009, pp. 1072–1077.

[9] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

[10] J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. Advance in Knowledge Discovery in Database: Papers from the 1994 AAAI Workshop*, Jul. 1994, pp. 359–37.

[11] Chao Shen,Zhongmin Cai , Xiaohong Guan and Roy A. Maxion, " User authentication through mouse dynamics," *IEEE Transaction on Information Forensics and security*,Vol. 8,no.1,JAN 2013.

[12] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Proc. Advances in Neural Information Processing Systems*, Denver, CO, 1999, pp. 526–532.

[13] N. Zheng, A. Paloski, and H. M. Wang, "An efficient user verification system via mouse movements," in *Proc. ACM Conf. Computer and Communications Security*, Chicago, IL, 2011, pp. 139–150.

[14] R. Biddle, S. Chiasson, and P. C. van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Computing Surveys*, vol. 44, no. 4, 2012.

[15] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

[16] C. Shen, Z. M. Cai, X. H. Guan, H. L. Sha, and J. Z. Du, "Feature analysis of mouse dynamics in identity authentication and monitoring,"in *Proc. IEEE Int. Conf. Communication (ICC)*, Dresden, Germany, 2009, pp. 1–5.

[17] S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *Proc. Int. Conf. Dependable Systems& Networks*, Estoril, Portugal, 2009, pp. 125–134.

[18] S. Mika, B. Schölkopf, A. J. Smola, K.-R. Müller, M. Scholz, G. Rätsch, M. S. Kearns, S. A. Solla, and D. A. Cohn, "Kernel PCA and de-noising in feature spaces," in *Proc. Advances in Neural Information Processing Systems*, Denver, CO, 1999, pp. 536–542