Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Effective User Recognition Using Mouse Gesticulation

Govardhini.S
PG Scholar Department of IT
SNS College of Engineering,
Coimbatore ,India

Kowshika.A
Assistant Professor Department of IT
SNS College of Engineering,
Coimbatore ,India

Abstract--- In this paper, we present a new mouse dynamics analysis framework that uses mouse gesture dynamics for static authentication. The gestures captured are analyzed using Ant Colony Optimization. In behavioral biometric technology, mouse dynamics biometrics is the one that extracts and analyzes the movement characteristics of the mouse input device when a computer user interacts with a graphical user interface for identification purpose. The biggest part of the presented study of mouse dynamics analysis has targeted primarily continuous authentication or user authentication for which promising results have been achieved. Static authentication (at login time) using mouse dynamics, however, there may be some challenges due to the limited quantity of data that can reasonably be captured during such an operation.

Index Terms—Behavioral biometrics, biometric authentication, computer security, identity verification, mouse dynamics

I. INTRODUCTION

WITH the rising number of hacking incidents and identity theft, the demand for strengthened security in networked environments is ever increasing. Simply having a password protected system is just not good enough these days. As a result, new methods are being developed to restrict user access as well as to protect the confidentiality and integrity of important data in various computer systems. One of these approaches is biometrics. Biometric recognition systems are widely used in various security applications, and are considered among the most accurate and efficient security systems in the market. In the Oxford dictionary, a generic definition of biometrics is given as "the application of statistical analysis to biological data" . In the particular field of computer security, biometrics is defined as "the automated use of a collection of factors describing human behavioral or physiological characteristics to establish or verify a precise identity". From finger-print scanning to voice recognition, biometrics is becoming a popular choice for enhancing the security of many computer systems. Unfortunately, a common limitation of most biometric systems is their reliance on special hardware devices for biometric data collection.

Although a few computer vendors have started integrating biometric readers in their products, an overwhelming number of machines are still not equipped with such special hardware devices. This restricts the scope of traditional biometrics systems to only networks or organizations where these devices are available. Thus, they are insufficient for securing computer systems for organizations that conduct business with a large and varied population over the Internet. This is where Mouse Dynamics, in other words, Mouse Actions, come into play. Behavioral biometrics based on mouse or keystroke dynamics represent some interesting alternatives, which address the above shortcomings since they can be implemented and collected using standard human-computer interaction devices readily available at any modern computing system.

The mouse dynamics biometric is a new biometric technology which has recently been developed in our lab for computer user recognition based on the way a user uses his/her mouse . Previous work on mouse dynamics has been limited mainly to improving the design of user interfaces . In our research, we target the biometric identification problem by focusing on extracting the behavioral features related to the mouse movements of a user and using these features for enhancing computer security. The mouse dynamics biometric involves a signature that is based on selected mouse movement characteristics. These characteristics are computed using statistical techniques as well as using neural networks. One of its key strength compared to traditional biometric technologies is that it allows dynamic and passive user monitoring.

Mouse dynamics biometrics could also be effective for dynamic authentication or identity confirmation in cases where the actions of an active user raise some suspicions. Besides these possible applications, we think that the most suitable use of mouse dynamics biometrics is for continuous monitoring applications such as detecting masqueraders in intrusion detection, or establishing the identity of perpetrators in digital forensics analysis. Mouse dynamics biometrics could also be used to generate and recognize electronic signatures in e-commerce transactions. So far the technology has been evaluated experimentally with 22 human participants achieving an equal error rate (False Rejection and False Acceptance) of about 2.46% recent years have seen an increasing interest in mouse dynamics as a new behavioral biometric Mouse dynamics refers to distinctive behavioral characteristics in a user's mouse actions when he interacts with the Graphical User Interface (GUI) of a computer system.

It is expected that mouse dynamics can provide a convenient source of information for personal identification which is applicable in most current computer systems. Unlike other biometrics such as fingerprints or voice, collection of mouse dynamics does not require a specialized hardware and can proceed non-intrusively as a user operates the computer using a mouse, so there may be less resistance in the acceptance of the biometric by ordinary computer users. Research on mouse dynamics began to appear in literature from around 2003. Reference reported feasibility studies of authentication methods based on mouse dynamics. In,[2] S. Yitzhaki. Gini' s proposed a remote authentication system based on signatures written by mouse. They reported a false acceptance rate (FAR) of around 7%, a false rejection rate (FRR) of 38.6% for single attempt and a FRR of 0.2% for multiple attempts. In [4], Gamboa and Fred studied the possibility of authentication based on mouse dynamics in graphical interactions, which reported an equal error rate of 1/50 based on 30 seconds of mouse interaction data. The results of [4] are less encouraging, Hocquet etc. studied mouse dynamics in a computer game and an equal error rate of 37.5% was reported. Reference reported researches on identity monitoring and tracing by using information of mouse dynamics. In[5], Ahmed and Traore collected about 13 hours of data of mouse activities from 22 users to test whether the users could be separated according to a set of features defined to characterize mouse dynamics. They reported an equal error rate of 2.46%. In [1], Pusara and Brodley focused on mouse dynamics in specific applications. They collected 2 hours of mouse data in Internet Explorer browsing activities from 18 users. An average FAR of 1.75% and average FRR of 0.43% are reported. Although there are some variations among reported accuracies, all previous findings support the basic assumption that mouse dynamics can be used as a source of information for personal identification.

II. LITERATURE SURVEY

A. Fuzzy Classification

These techniques are based on looking at the mean and covariance of the features to determine if they belong in a given class. In this section of the course, we will examine the notions of crisp and fuzzy classes. We will also examine some relatively simple techniques for determining the non-linear transfer functions that map features into classes. These transfer functions enable the HCI designer to develop systems that can find patterns in the data obtained from input sensors and map these patterns to specific actions which can control the operation of the computer. Fuzzy sets, on the other hand, allow elements to be *partially* in a set. Each element is given a degree of membership in a set. This membership value can range from 0 (not an element of the set) to 1 (a member of the set). It is clear that if one only allowed the extreme membership values of 0 and 1, that this would actually be equivalent to crisp sets. A membership function is the relationship between the values of an element and its degree of membership in a set.

The European standard requires, for commercial biometric systems, a false acceptance rate (FAR) of less than 0.001%. and a false rejection rate (FRR) of less than 1%. This correspondence paper takes as its point of departure the feature set and data set provided in and investigates, for the first time, the use of fuzzy logic for mouse dynamics biometric analysis. Fuzzy logic provides more adequate answers (than classical logic) when dealing with the uncertainty underlying human actions. A fundamental and challenging aspect of fuzzy logic is the extraction of a fuzzy membership function. We use the Learning Algorithm for Multivariate Data Analysis, an unsupervised learning technique, for such a purpose. Supervised learning may be used to discriminate between users in a closed setting in which mouse data can be collected for all the users.

B. Support Vector Machines

Support Vector Machines 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. A schematic example is shown in the illustration below. In this example, the objects belong either to class GRAY or BLACK. The separating line defines a boundary on the right side of which all objects are GRAY and to the left of which all objects are BLACK. Any new object (white circle) falling to the right is labeled, i.e., classified, as GRAY(or classified as BLACK should it fall to the left of the separating line).

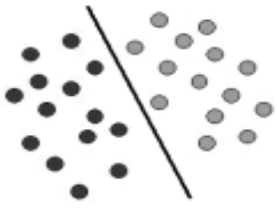

Fig 1. Support Vector Machine

The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GRAY and BLACK in this case) with a line. Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks. Unfortunately, they only provide one-time verification, and the verified users are still vulnerable to both session hijacking and the divulging of the secret information. To achieve a timely response to an account breach, more frequent user verification is needed. However, frequent verification must be passive and transparent to users, as continually requiring a user's involvement for re-authentication is too obtrusive and inconvenient to be acceptable

C. Dimensionality Reduction

Dimension reduction1 is the mapping of data to a lower dimensional space such that uninformative variance in the data is discarded, or such that a subspace in which the data lives is detected. Dimension reduction has a long history as a method for data visualization, and for extracting key low dimensional features. In some cases the desired low dimensional features depend on the task at hand. Apart from teaching us about the data, dimension reduction can lead us to better models for inference. The need for dimension reduction also arises for other pressing reasons. In[6], Chao Shen, Zhongmin Cai ; Xiaohong Guan ; Huilan Sha ; Jingzi Du, proposed a remote authentication system based on signatures written by mouse. They reported a false acceptance rate (FAR) of around 7%, a false rejection rate (FRR) of 38.6% for single attempt and a FRR of 0.2% for multiple attempts. In Gamboa and Fred studied the possibility of authentication based on mouse dynamics in graphical interactions, which reported an equal error rate of 1/50 based on 30 seconds of mouse interaction data. The results of are less encouraging, Hocquet etc.

D. Principle Component Analysis

Principal component analysis (PCA) is a mathematical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

In [5], Ahmed and Traore collected about 13 hours of data of mouse activities from 22 users to test whether the users could be separated according to a set of features defined to characterize mouse dynamics. They reported an equal error rate of 2.46%. In , Pusara and Brodley focused on mouse dynamics in specific applications. They collected 2 hours of mouse data in Internet Explorer browsing activities from 18 users. An average FAR of 1.75% and average FRR of 0.43% are reported. Although there are some variations among reported accuracies, all previous findings support the basic assumption that mouse dynamics can be used as a source of information for personal identification.

III. PROPOSED SYSTEM

A. Ant Colony Optimization

The algorithm used is Ant Colony Optimization, The complex social behaviors of ants have been much studied by science, and computer scientists are now finding that these behavior patterns can provide models for solving difficult combinatorial optimization problems. The attempt to develop algorithms inspired by one aspect of ant behavior, the ability to find what computer scientists would call shortest paths, has become the field of ant colony optimization (ACO), the most successful and widely recognized algorithmic technique based on ant behavior. This book presents an overview of this rapidly growing field, from its theoretical inception to practical applications, including descriptions of many available ACO algorithms and their uses. In all Ant Colony Optimization algorithms, each ant gets a start city. Beginning from this city, the ant chooses the next city according to algorithm rules. After visiting all customer cities exactly once, the ant returns to the start city. The ants might travel concurrently or in sequence. Each ant deposits some amount of pheromone on his path. The amount of pheromone depends on the quality of the ant's path: a shorter path usually results in a greater amount of pheromone. The deposited pheromones suffer from evaporation. The algorithms use different rules for selection of the next city to move to, for evaporation, and for deposition of pheromone.

B. Euclidean distance algorithm

Euclidean algorithm is a method for computing the greatest common divisor (GCD) of two (usually positive) integers, also known as the greatest common factor (GCF) or highest common factor (HCF). It is named after the Greek mathematician Euclid, who described it in Books VII and X of his *Elements*. The Euclidean algorithm proceeds in a series of steps such that the output of each step is used as an input for the next one. Let k be an integer that counts the steps of the algorithm, starting with zero. Thus, the initial step corresponds to $k = 0$, the next step corresponds to $k = 1$, and so on.

Each step begins with two nonnegative remainders r_{k-1} and r_{k-2}. Since the algorithm ensures that the remainders decrease steadily with every step, r_{k-1} is less than its predecessor r_{k-2}. The goal of the kth step is to find a quotient q_k and remainder r_k such that the equation is satisfied

$$r_{k-2} = q_k r_{k-1} + r_k$$

where $r_k < r_{k-1}$. In other words, multiples of the smaller number r_{k-1} are subtracted from the larger number r_{k-2} until the remainder is smaller than the r_{k-1}. In the initial step ($k = 0$), the remainders r_{-2} and r_{-1} equal a and b, the numbers for which the GCD is sought. In the next step ($k = 1$), the remainders equal b and the remainder r_0 of the initial step, and so on.

C. Modules

a. Training User

This Module is used to capture the mouse movements recorded *xi* and *yi* position values of the mouse pointer at a sampling rate of 100 samples per second. Each run was stored in a separate data file, identifying both the user and the number of the run he or she was performing. Given the specific task for the participants, the resulting data gave a clear view of the performed task.

b. Data Preprocessing

The data preprocessing consisted of transforming the collected positioning data into velocity data and then reducing the noise in the transformed data by applying a noise reduction filter. Transforming positioning data into velocity data is done by simply taking the derivative, transformed velocity data in the horizontal (upper part of figure) and vertical (lower part of figure) direction, plotted against time. Various different noise reduction filters can be applied. We decided to use the Moving Average (MA) filter with a window size of $N = 5$. For the MA filter, a filtered value at time *ti* is the normal average of that value plus $(N - 1)/2$ neighbouring values to the left and to the right.

c. Feature Extraction

The dataset can be split in different parts, each part representing a single track of the maze. So the task during the feature extraction will be to split the data in horizontal and vertical tracks. We know that on a horizontal track the velocity in the *x* direction will be high (in the absolute sense), while the velocity in the *y* direction will be around 0 (but not always equal to 0 because of some small trembling in the vertical direction). For a vertical track the opposite holds. The basic rule used to determine at a specific point in time if we are on a horizontal or a vertical track was to look at the absolute values of the velocities in both directions. If the absolute velocity in the horizontal direction is highest, we declare to be on a horizontal track, otherwise on a vertical track. This method is however too simple to work perfect under all circumstances, so we added some extra checks. One extra check looked at the occurrence of short tracks. It could be that when going from a horizontal to a vertical track (or the other way around) the basic rule could switch between horizontal and vertical tracks various times. This is obviously due to the fact that horizontal and vertical velocity on the corners can be almost the same, i.e. sometimes the horizontal velocity is the highest, and sometimes the vertical velocity. In order to not detect all these switches, we have a threshold value *t* for the minimal length of a track. Furthermore we also checked that the total number of detected tracks at the end.During the feature extraction, we will create a so-called transition vector indicating where the (horizontal and vertical) tracks start and end. Note that the end of a track indicates directly the beginning of the next track.

d. Distance Metrics Calculation

In this module various distance metrics, being the Euclidean Distance, the Manhattan Distance and the Edit Distance. The first two distance metrics are well known and can be applied to sequences of equal length. Both these distance metrics consider the difference in the values of the input sequences at a specific point in time. In other words they regard the cost of a substitution of one value in another. The edit distance (also called Levenshtein distance or referred to as Dynamic Time Warping) on the other hand, does not only consider the costs of substituting one value by another, but also the costs of inserting or deleting values in the input sequences.

e. User Validation

A solution for this is the use of biometrics as a way to authenticate a user. Various options are available for getting access to a computer. A technique already implemented in a number of laptops is the use of fingerprints. If however no fingerprint scanner is present in a computer, extra costs need to be made to acquire the hardware. Two other biometric features, especially suitable for access to computer systems, are keystroke dynamics and mouse dynamics. The first one measures how the user uses the keyboard, the second one measures how he uses the mouse. Although both methods have their own specific disadvantages, the major advantage of these methods is that they do not require any extra equipment. Many computer systems are nowadays equipped with a mouse and a keyboard that are needed to measure the features. Laptop systems however are generally equipped with touchpads that exhibit completely different user behavior compared to an external mouse

IV. CONCLUSION

In this paper, we highlighted the challenges faced by mouse dynamics biometric technology when it is applied to static authentication and proposed a new framework based on mouse gesture dynamics that achieves encouraging results toward addressing those challenges. The proposed framework uses a modular design of the Ant Colony Optimization which is efficient than the above mentioned. But there are still some challenges in authenticating the users by mouse gestures.

REFERENCES

[1] M. Pusara, C. E. Brodley. User re-authentication via mouse movements. In Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security. Washington DC, USA, 2004: 1-8.

[2] S. Yitzhaki. Gini' s Mean difference: a superior measure of variability for non-normal distributions. Metron-International Journal of Statistics, 2003. LXI, 285-316.

[3] Y. Aksari and H. Artuner. Active authentication by mouse movements. In Proceedings of 24th International Symposium on Computer and Information Science, 2009: 571-574.

[4] H. Gamboa and A. Fred. A behavioral biometric system based on human computer interaction. In Proceedings of SPIE, 2004, 54: 4-26.

[5] A. A. E. Ahmed and I. Traore. A new biometric technology based on mouse dynamics. IEEE Transactions on Dependable and Secure Computing, 2007, 4(3): 165-179.

[6] Chao Shen, Zhongmin Cai ; Xiaohong Guan ; Huilan Sha ; Jingzi Du, Feature Analysis of Mouse Dynamics in Identity Authentication and Monitoring. Communications, 2009. ICC '09. IEEE International Conference 1938-1883.

[7] Nakkabi, Y. Traore, I. ; Ahmed, A.A.E. Improving Mouse Dynamics Biometric Performance Using Variance Reduction via Extractors With Separate Features. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on (Volume:40 , Issue: 6) Biometrics Compendium, IEEE. 1083-4427.

[8] Nan Zheng, Aaron Paloski, and Haining Wang Department of Computer Science The College of William and Mary {nzheng,appalo,hnw}@cs.wm.edu Williamsburg, VA 23185, USA.

