User Authentication using Mouse Dynamics

Bachelor of Technology
in
Electronics and Electrical Communication

Bandaru Sai Manoj 17EC35007
D Rajesh 17EC35037
Daiveek Sai P 17EC10011
M Vamsi Krishna 17EC10029
K Sudheer Kumar 17EC35038

Electronics and Electrical Communication Engineering
Indian Institute of Technology Kharagpur
India

November 8, 2020

Contents

1 Introduction

Today most computer systems identify users by means of secret phrases known as passwords. However, this authentication system does nothing to protect the computer from unauthorized access once the user has started an active session. Unattended computers with an active session present a much larger security threat. Users who are not tech-savvy frequently leave their computers unlocked with an active session.

This allows for two types of attacks:

1. A user of lower clearance can gain access to a terminal with higher clearance and access files or functions of the network to which he is not supposed to have access to.
2. Users with the same or higher clearance can conceal his identity by performing malicious actions under the guise of a coworker.

These limitations of password-based authentication lead to the introduction of authentication techniques based on biometrics.

1.1 Biometric Authentication

Human recognition can be done by using his physiological or behavioral characteristics. Biometrics offer automated methods of identity verification or identification on the principle of measurable physiological or behavioral characteristics. The characteristics are measurable and unique. Thus, biometrics play an important role in recognizing a human being.

Types of Biometrics:

1. Physiological Biometrics
2. Behavioral Biometrics

1.1.1 Physiological Biometrics

Physiological biometrics involve physiological characteristics of a human being used as biometric such as voice, DNA, fingerprint, IRIS pattern or hand geometry. These biometrics are more reliable and accurate.

1.1.2 Behavioral Biometrics

Behavioral biometrics involve the behavioral characteristics of a human being. These biometric characteristics are acquired over time by an individual, and are at least partly based on acquired behavior. Thus, it is something known to an individual and can be exploited for authentication purposes.

Mouse dynamics comes under behavioral biometrics. In this we consider the features like individual moves the mouse or clicks on the screen of the desktop/laptop. Mouse actions like mouse movements, clicks, drag and drop etc.

2 Problem Definition

The main aim of our project is to create a User Authentication system using continuous mouse dynamics data for personalised Laptops/PCs to prevent threats from any intruder/imposter accessing the authenticated system.

This is done as a part of the course project during the semester to authenticate the user by training and testing the Gaussian Mixture Model by the continuous mouse dynamics data.

3 Mouse Dynamics

Mouse dynamics, a means of biometric identification based on users characteristic interactions with the mouse, is a largely unexplored area of research. As a result, data sets containing unrestricted mouse usage data is still remarkably scarce.

So the data set preparation has been done by every student taking the course and the corresponding data has been used for the purpose.

The real ability of the mouse dynamics biometric technology is in its ability to constantly monitor the legitimate and illegitimate users based on their session based usage and some typical mouse related actions.

To understand about mouse dynamics further, we can look into different types of mouse actions which can occur from user interaction with the PC through a mouse

1. MOUSE MOVE: Mouse move is a simple movement in the 2D region of the screen involving no clicks. This can happen between two click events or non-click events.
2. DRAG AND DROP: It is the action which starts by a mouse button held down followed by a movement and finally the button released. Generally ,we use this to move/copy a file into a particular location.
3. POINT AND CLICK: It is the act of moving to point at a location ending with a click on it.
4. SILENCE: This action suggests no mouse movement at all for a fixed duration of time.

To capture these kind of mouse actions we would require the data collection software to capture events like:

1. Mouse Move
2. Mouse Wheel Movement
3. Mouse Pressed
4. Mouse Released
5. Mouse Pointer 2D Location

These above events can be used to extract the some mouse actions like:

1. Left Click
2. Right Click
3. Left Double Click
4. Right Double Click

4 Feature Extraction

The frequent-behaviour patterns cannot be used directly by a detector or a classifier. Instead, dynamic features are extracted from these patterns. Characteristics that can be extracted are:

1. **Click Time:** The time required for the user to click a button.
2. **Pause Time:** The amount of time spent pausing between pointing to an object and actually clicking on it.
3. **Horizontal Velocity:** The change in X-coordinate value for the given change in time.
4. **Vertical Velocity:** The change in Y-coordinate value for the given change in time.
5. **Straightness:** This characterizes the nature of movement of the mouse by the user.

The information used is :

1. **Temporal Information:**
   (a) Horizontal coordinates
   (b) Vertical Coordinates
   (c) Path distance from origin
   (d) Angle of the path with respect to X-axis
   (e) Curvature of the path
   (f) Derivative of the curvature of the path
2. **Spacial Information:**
   (a) X values
   (b) Y values
   (c) Horizontal Velocity
   (d) Vertical Velocity
   (e) Tangential Velocity
   (f) Angular Velocity

For each data sample, we extracted the following:

1. Statistical features:
   (a) Mean
   (b) Standard Deviation
   (c) Minimum
   (d) Maximum
   (e) Range
2. Straightness of the path
3. Jitters
4. Critical Points or High Curvature points
5. Number of pauses, Paused time

5 Feature Selection

For selecting features from a candidate data set we have used a common feature selection technique called **Sequential Feature Selection**.

This Sequential Feature Selector adds (forward selection) or removes (backward selection) features to form a feature subset in a greedy fashion. At each stage, this estimator chooses the best feature to add or remove based on the cross-validation score of an estimator.

Here, we used **knn** as estimator and reduced to 5 feautures with scoring as accuracy.

6 Classifier

The classifier used here is **Gaussian Mixture Model**.

Gaussian Mixture Models are a powerful clustering algorithm working on the principle of unsupervised learning. Gaussian Mixture Models (GMM's) assume that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Hence, a Gaussian Mixture Model tends to group the data points belonging to a single distribution together. Gaussian Mixture Models are probabilistic models and use the soft clustering approach for distributing the points in different cluster.

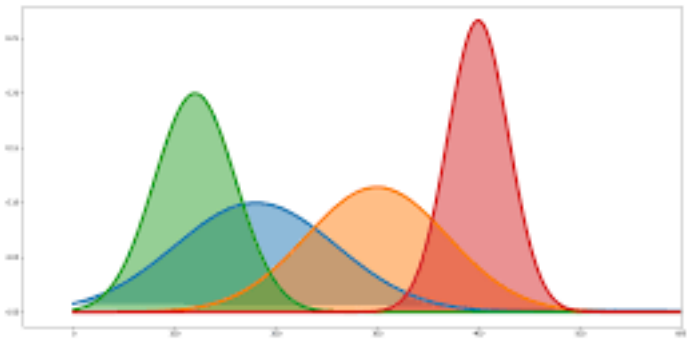

Figure 1: Different Gaussian's representing different clusters

The idea behind clustering is grouping data points together such that each individual cluster holds the most similar points. This algorithm is different from k-means clustering in a way that this takes into account the mean as well as the variance of the data.

Here we have used **GaussianMixture** function from **scikit-learn**.

Guassian Mixture Model from scikit-learn: GaussianMixture.

7 Results

K-Fold	Training Accuracy(%)	Testing Accuracy(%)
1st Fold	96.63016802107306	93.62322442801779
2nd Fold	95.1273505842548	97.23906717821176
3rd Fold	95.25622583977102	96.20463681564797
4th Fold	95.69929020874315	95.0859320304987
5th Fold	97.45703716656706	94.07101048526768
Overall Accuracy	96.0340143640818	95.24477418752878