Indian Institute of Technology, Kharagpur
Department of Electronics and Electrical Communication Engineering

Biometric Authentication using Mouse Dynamics

Machine Intelligence and Expert Systems

Group 6 Team Members:

Avinab Saha (15EC10071)
Atul (15EC10067)
Kanakvi Aggarwal (15EC10068)
Joyjit Paul (15EE10059)
Sohom Chakraborty (15EE10060)

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

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 three 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. User with the same or higher clearance can conceal his identity by performing malicious actions under the guise of a coworker.
3. A person who is not affiliated with the company in anyway can gain access to the internal network.

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

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Biometric Authentication

What is it?

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

The biometrics can be broadly classified into two types:

1. Physiological biometrics
2. Behavioral biometrics

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. They are not affected by any mental conditions such as stress or illness.

1. **Iris pattern:** Iris recognition technology are used primarily in high security environments, where low error rates are essential. But the problem with Iris patterns is that its accuracy is affected by changes in lighting, its scanners are more expensive. Moreover, the recognition is difficult to perform at a distance longer than few meters.
2. **Fingerprint:** comparison of several features of the print pattern. These patterns don't change much over the time period and differ from human to human, hence if properly recognized the biometric can be used to a great level of accuracy. The sensor is cheap and its size is also small. any intruder may access a legitimate users fingerprints and can use it to login to the system. Hence Replicable

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.

1. **Mouse Dynamics**: This behavioral biometric is characterized by the way an individual moves the mouse or clicks on the screen of the desktop/laptop. Mouse actions like mouse movements, clicks, drag and drop etc. can be used as useful features. This behavioral biometric also has issue with variability of features over time.

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Components of Biometric Identification system

1. Feature extraction which captures the data generated by standard input devices such as a mouse or a keyboard.
2. Feature extraction and classifier module that constructs the users signature based on his behavioral biometrics.
3. A signature database consisting of behavioral signatures of registered users.

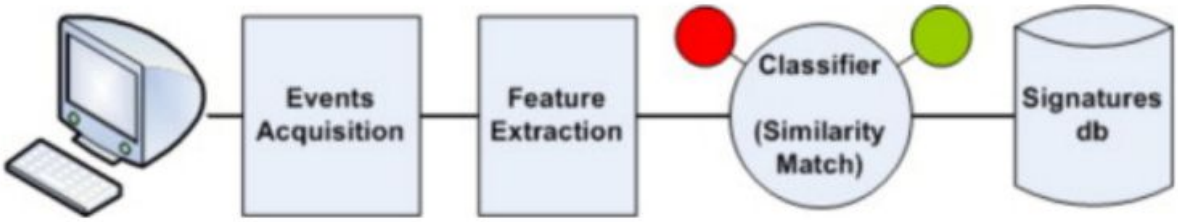

Fig 1: Typical framework of a behavioral biometric identification system

Requirements of a Biometric

1. **Universality:** Every person should have the characteristics.
2. **Uniqueness:** No two person should have the same biometric characteristics.
3. **Permanence:** The biometric should not be variant with time. Behavioral biometrics can be highly variable with time.
4. **Collect-ability:** The characteristics must be measurable quantitatively and obtaining the characteristics should be easy.
5. **Performance:** Acceptable accuracy should be achieved after identification/ verification.
6. **Circumvention:** This property indicates to how difficult it is to fool the system by fraudulent techniques.

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Problem Definition

The objective of the project is to create a continuous user authentication system for PCs/laptops to prevent threat against intruder, using biometrics involving mouse dynamics.

During the course of the semester, the program was implemented to authenticate the user by training and testing the assigned classifier by the neutral, happy and sad mood data.

Mouse Dynamics

For getting an idea about mouse dynamics we need to look into different types of mouse actions which can occur from user interaction with the PC through a mouse. The main strength of mouse dynamics biometric technology is in its ability to constantly monitor the legitimate and illegitimate users based on their session based usage of a computer system.

The different types of possible mouse actions are listed below:

1. **Mouse Move:** Mouse move is a simple movement involving no clicks. Mouse move can be 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, it is used to move/copy a file to a particular location.
3. **Point and Click:** It is a movement of the mouse ending in a click.
4. **Silence:** This action suggests no mouse movement.

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

1. Mouse move
2. Mouse pointer location
3. Mouse wheel movement
4. Mouse Pressed
5. Mouse released

From these events mouse actions given below can be extracted:

1. Left Click
2. Left Double Click
3. Right Click
4. Drag and Drop

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Feature Extraction

The mined frequent-behavior patterns cannot be used directly by a detector or classifier. Instead, dynamic characteristics are extracted from these patterns. Some characteristics that can be extracted are:

1. **Click Time**: It is the time required for the user to click a button.I
2. **Pause Time:** It is the amount of time spent pausing between pointing to an object and actually clicking on it.
3. **Horizontal Velocity:** Horizontal Velocity is change in X coordinate value for the given change in time.
4. **Vertical Velocity**: Vertical Velocity is change in Y coordinate value for the given change in time.
5. **Straightness:** The straightness feature characterizes the nature of movement of the
6. **Mouse move**: It was seen in the previous chapter that the feature was not discriminating.

The information vectors used are:

SPATIAL INFORMATION	TEMPORAL INFORMATION
1. Horizontal coordinates 2. Vertical coordinates 3. Path distance from the origin 4. Angle of the path with respect to X axis 5. Curvature of the path 6. Derivative of the curvature of the path	1. Input x values 2. Input y values 3. Input t values 4. Horizontal velocity 5. Vertical velocity 6. Tangential velocity 7. Tangential acceleration 8. Tangential jerk 9. Angular velocity

For each information vector, we extracted the following features:

1. Statistical features:
    a. Mean
    b. Standard deviation
    c. Maximum
    d. Minimum
    e. Range
2. Straightness of the path
3. Jitter
4. High curvature points (also called critical points, can be multiple for the same vector)
5. Number of pauses, paused time and paused time ratio

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Classification

Random Forest

Decision Tree Ensembles, also referred to as random forests, are useful for feature selection in addition to being effective classifiers. One approach to dimensionality reduction is to generate a large and carefully constructed set of trees against a target attribute and then use each attribute's usage statistics to find the most informative subset of features.

If an attribute is often selected as best split, it is most likely an informative feature to retain. A score calculated on the attribute usage statistics in the random forest tells us − relative to the other attributes − which are the most predictive attributes.

Random Forest was used to select the most discriminating features among the initially 63 extracted features. On multiple iterations among various users, the number of discriminating features was around 4-10. On creating the feature-set based on the extracted features, we use support vector machine for further classification

One Class Support Vector Machine

In general, Support Vector Machines (SVMs) are able to achieve comparable or even higher accuracy with a simpler and thus faster scheme than neural networks. In the two-class formulation, the basic idea of SVMs is to map feature vectors to a high dimensional space and compute a hyperplane, which separates the training vectors from different classes and further maximizes this separation by making the margin as large as possible.

We aim to find a hyperplane that not only separates the data points but also maximizes the separation.Thus, the procedure to resolve a classification problem using SVMs is:

1. Choosing a kernel function,
2. Setting the penalty parameter C and kernel parameters as well, if any,
3. Constructing the discriminant function from the support vectors.

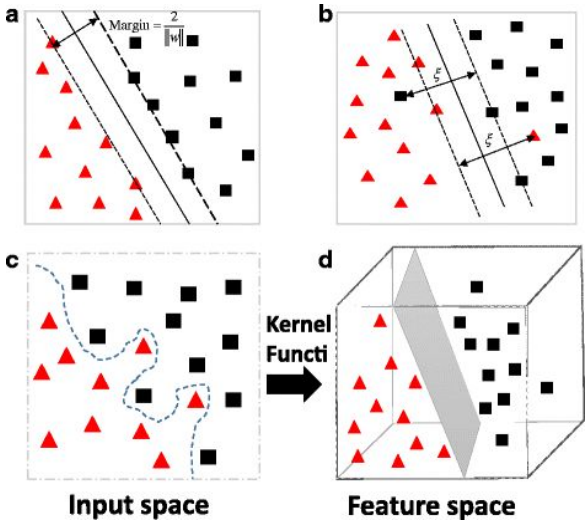

Fig: Working of One Class Support Vector Machine

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Details on tuning of parameters and corresponding results:

1. The kernel chosen was 'rbf' with ν = 0.01 to obtain the maximum training accuracy of 94.642% and a test set accuracy of 86.172% which was obtained by averaging out the metrics for all the 5 users.
2. ν was varied between 0.01 and 0.22 to obtain a variety of results with a tradeoff between training and test accuracy as a low value of ν leads to overfitting and a value close to 1.0 would lead to underfitting.

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Results

User-13	User-1
1. Training accuracy: 94.9475065617 Test accuracy: 86.9121706986 Precision: 0.1098546042 Recall: 0.890052356021 F1 Score: 0.195570894449	1. Training accuracy: 78.1124497992 Test accuracy: 41.4629479022 Precision: 0.0310827007591 Recall: 0.779559118236 F1 Score: 0.0597817734747
2. Training accuracy: 94.8196721311 Test accuracy: 87.6930276088 Precision: 0.112495845796 Recall: 0.887287024902 F1 Score: 0.199675564076	2. Training accuracy: 78.0230807827 Test accuracy: 41.8763754665 Precision: 0.0309642226857 Recall: 0.774322968907 F1 Score: 0.0595472251147
3. Training accuracy: 94.3606557377 Test accuracy: 85.2316331306 Precision: 0.107771802172 Recall: 0.893356643357 F1 Score: 0.192340265362	3. Training accuracy: 77.8223783241 Test accuracy: 40.9673715434 Precision: 0.0312175648703 Recall: 0.784615384615 F1 Score: 0.0600460711543
4. Training accuracy: 94.4262295082 Test accuracy: 85.8072063641 Precision: 0.106895208414 Recall: 0.899672131148 F1 Score: 0.191086350975	4. Training accuracy: 78.273958856 Test accuracy: 42.7279686154 Precision: 0.031311468829 Recall: 0.78273958856 F1 Score: 0.0602142236804
5. Training accuracy: 94.6885245902 Test accuracy: 85.2363125877 Precision: 0.104859020824 Recall: 0.895592864638 F1 Score: 0.187737145999	5. Training accuracy: 78.4746613146 Test accuracy: 42.632283992 Precision: 0.0312525171164 Recall: 0.778803693296 F1 Score: 0.0600935476876

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)

Remarks

Factors Affecting Performance

1. Environmental conditions
    a. Height of chair
    b. Distance between mouse and body
    c. Touchpad vs conventional mouse
2. User conditions
    a. Mood
    b. Knowledge & practice of application
    c. Typing errors
3. GUI/mouse setting
    a. Screen resolution
    b. Pointer speed
4. Noise
    a. Hardware error
    b. Software error

Future Improvements

1. Mood analysis
    a. A separate label for mood could constitute a useful feature
2. Data collection
    a. More amount of data
    b. Standard environment
    c. Standard computer settings
3. Data Preprocessing
    a. Noise removal, smoothening and error nullification
4. Training
    a. State-of-the-art classification algorithms

Reference: Gamboa, H. and Fred, A.L., 2003, April. An Identity Authentication System Based On Human Computer Interaction Behaviour. In PRIS (pp. 46-55)