Key-Stroke Pattern-based Behavioral Bio-metric Authentication system

Submitted by:
Mudit Rathore (ICM 2015 502)

Information Security

Information security (IS) is designed to protect the confidentiality, integrity and availability of computer system data from those with malicious intentions.

CONFIDENTIALITY : Hidden from unauthorised access

INTEGRITY : Protected from unauthorised change

AVAILABILITY : Available to an authorised entity

Security Services

- ACCESS CONTROL
- CONFIDENTIALITY
- INTEGRITY
- AUTHENTICATION
- NON REPUDIATION

**In this presentation we are specifically interested in a part of the authentication security service.

Authentication

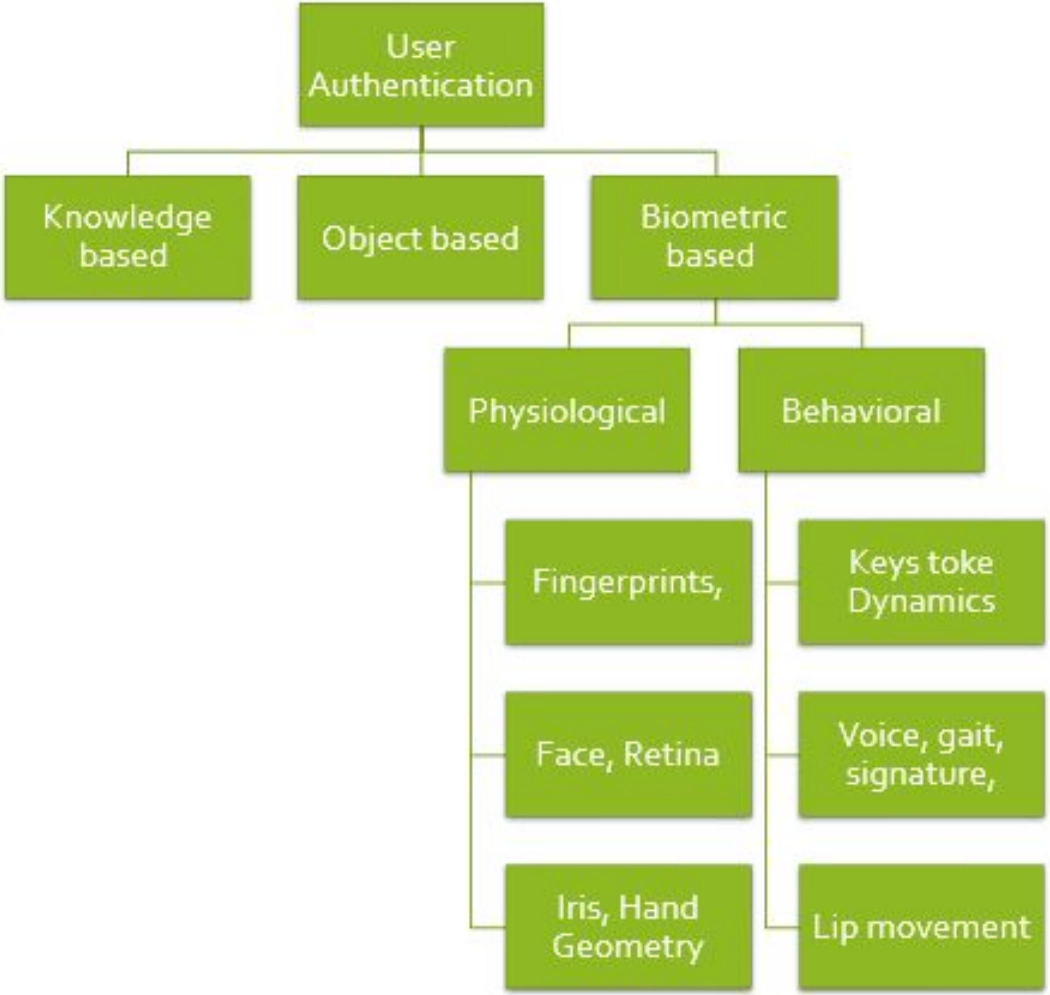
Authentication is the process of determining whether someone or something is, in fact, who or what it declares itself to be.

Authentication technology provides access control for systems by checking to see if a user's credentials match the credentials in a database of authorized users or in a data authentication server.

User Authentication

Users are usually identified with a user ID, and authentication is accomplished when the user provides a credential, for example a password, that matches with that user ID.

User authentication occurs within most human-to-computer interactions. With the increase of global resource and information, it has increase the possibility for new threats, and improved attacks. Thus it is necessary to enhance and improve the user authentication mechanism and make it more robust.



User Authentication

Knowledge based

Object based

Biometric based

Physiological

Behavioral

Fingerprints,

Face, Retina

Iris, Hand Geometry

Keys toke Dynamics

Voice, gait, signature,

Lip movement

Biometric Authentication

Biometric technologies are defined as “automated methods of verifying or recognizing the identity of a living person based on a physiological or behavioral characteristic”. Thus
Biometrics can be a good choice for user authentication.

Using the biometrics for authentication purpose is termed as Biometric Authentication.

Biometrics provides an extra level of security as these cannot be lost, stolen or imitated and these can also be used with the traditional method for verification.

Behavioural Characteristics v/s Physiological Traits

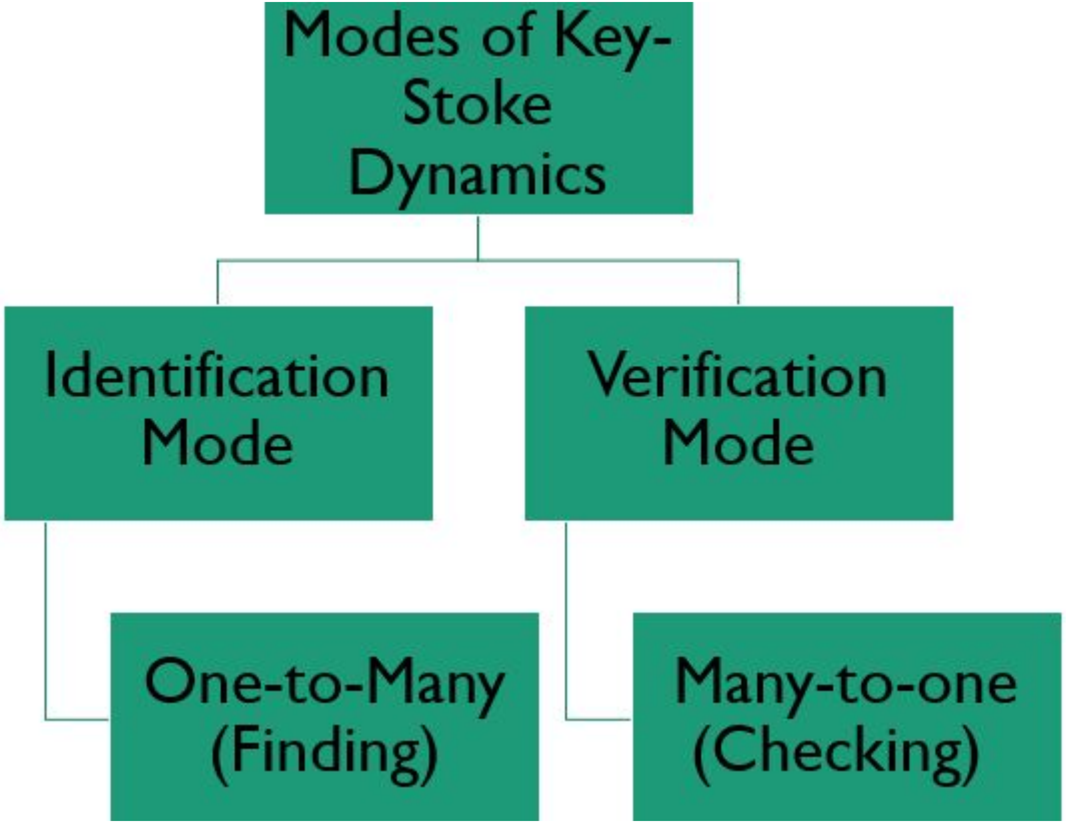
Behavioural characteristics are linked with the neurophysiological characteristics of an individual thus cannot be lost, stolen or imitated. These are easy to deploy when particular technique is considered.

Physiological traits like iris, retina and fingerprints though provide a very robust environment but are easy to fool and are very difficult and expensive to deploy.

We have considered KeyStroke Dynamics a behavioral biometric authentication technique in our study.

Keystroke Dynamic : Definition

Keystroke dynamics is the process of analyzing the way a user types at a terminal by monitoring the keyboard inputs thousands of times per second and attempts to identify them based on habitual rhythm patterns in the way they type.



Modes of Key-Stoke Dynamics

Identification Mode

Verification Mode

One-to-Many (Finding)

Many-to-one (Checking)

Assumptions

- We have considered for the scope of this project that the typing pattern patterns and rhythm of an individual is a characteristic of an individual that does not change with time and only show a slight changes when typed in different mood, on different hardwares and over number of times.
- We have considered that every character out of the extended ascii table list of 256 characters are typed at the beginning of the process.
- We have considered time as a floating point number expressed in seconds since the epoch, in UTC.

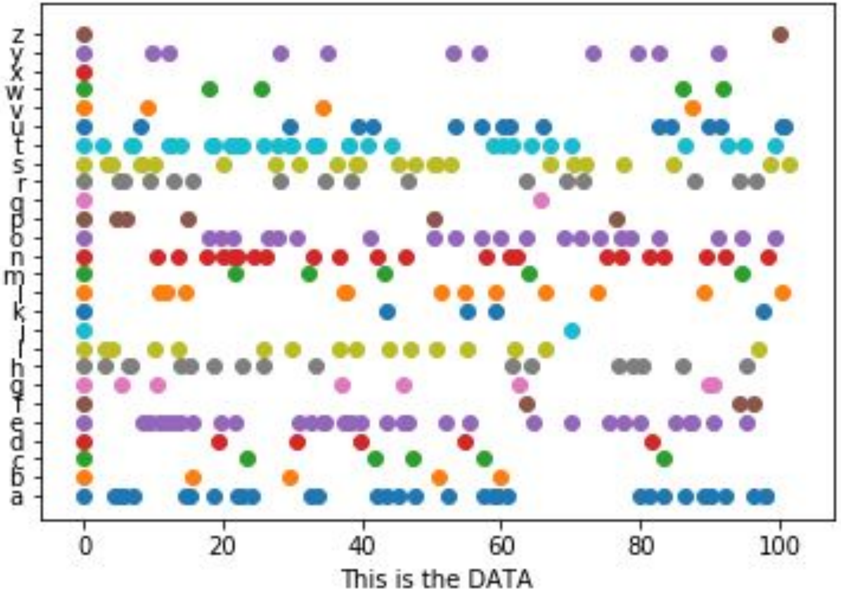
Data Set

- The database is created using a paragraph that is input by the user.
- We have stored the pressing time of the key in floating point and the releasing time of the key in floating point. We have tried to cover as many characters as possible in the extended ascii list of 256 characters.
- The user have to input his/her name in the beginning and it is appended by the last 5 time digits to form the csv format file name. For every user we have 7 different file in the dataset. This will help us increase the accuracy of our model.

Data Set

The scatter plot is created by user entry of the paragraph below.

“this is a paragraph that uses every single letter in the alphabet. now, that doesn’t mean this can be a paragraph with no story, but it does mean that every single letter is used. you can make it as generic or fanciful as you’d like. you can talk about anything from quilts to jets to xylophones. oh yeah, and you can use whatever language you want, from afrikaans to zulu.”



Features

We have considered the following features.

- The key pressing time
- The key release time
- The key seek time
- How many keys are simultaneous pressed
- The key holding time
- The error rate
- Method of the correction of error

Model Selection

We have considered many model for this classification like

- **Time series autocorrelation function**
- **KNN clustering algorithm**
- **Logistic Regression used for classification.**
- **Artificial Neural Networks.**

We have selected **Artificial Neural Networks (ANN).**

Why ANN ?

- ANNs have the ability to learn and model non-linear and complex relationships
- ANNs can generalize after learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well.
- It assigns weights to features as per their importance in the prediction process. This model also learns from error on every data point to improve overall accuracy. This is achieved by “Backpropagation”

Flow of Work

- Creation of the data set
- Training of ANN
    - Feature extraction
    - Pattern recognition
- Testing through ANN

How to Incorporate this into the security applications

- This can be incorporated into the security applications very easily for authentication of users.
- While creation of a new user they must be asked to type a very big paragraph so that our model can be trained via this Fixed text mode and our algorithm must be active while the user create the account and types different fields of the algorithm, free text mode. This will help us get all the necessary detail for the verification purposes.
- Then as the password is entered user must be given a pass phrase which the user enters into the system the user is authenticated if the patterns match upto a certain threshold.

Disadvantages of Keystroke Dynamics

The disadvantages of using the keystroke dynamics are

- **Drastic change in the typing pattern renders the user locked out of the account.**
- **Injuries might affect the typing patterns.**
- **Typing patterns might change with the hardware i.e. keyboard.**
- **Requires adaptive learning.**
- **Access is granted if typing pattern of a user matches within a particular threshold of the other user but falsely claimed.**

Advantages of Keystroke Dynamics

The advantages of using the keystroke dynamics are

- **Based on neuro-physiological characteristics, thus cannot be stolen, lost or imitated.**
- **Only Software requirement, no costly application specific hardware required.**
- **Simple to Deploy**
- **Cost Effective**
- **No end user training**
- **Simple and natural way to increase computer security**

Future scope

We can inculcate following features

- **Words-per-minute count**
- **Key latencies**
- **Error rate per specific key**
- **Pressure on a particular keyboard key**
- **Adaptive techniques during a particular session.**
- **Removal of outliers.**

Thank you !!