A WRAPPER BASED APPROACH FOR PERSONAL IDENTIFICATION THROUGH KEYSTROKE DYNAMICS USING SOFT COMPUTING TECHNIQUES

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO USER AUTHENTICATION
1.2 CLASSIFICATION OF USER AUTHENTICATION
   1.2.1 KNOWLEDGE BASED AUTHENTICATION
   1.2.2 OBJECT BASED AUTHENTICATION
   1.2.3 BIOMETRIC BASED AUTHENTICATION
1.3 BIOMETRICS OVERVIEW
   1.3.1 TYPES OF BIOMETRICS
   1.3.2 ANALYSIS OF BIOMETRICS TYPES
1.4 KEYSTROKE DYNAMICS AS BIOMETRICS
   1.4.1 MERITS OF USING KEYSTROKE DYNAMICS
   1.4.2 KEYSTROKE DYNAMIC APPROACHES
   1.4.3 APPLICATION AREAS OF KEYSTROKE DYNAMICS
1.5 PROBLEM JUSTIFICATION
1.6 OBJECTIVES OF THE THESIS
1.7 SIGNIFICANT CONTRIBUTIONS
1.8 ORGANIZATION OF THE THESIS
1.9 CHAPTER SUMMARY
CHAPTER 1

INTRODUCTION

1.1. INTRODUCTION TO USER AUTHENTICATION

The security of computer access is important today because of huge transactions being carried out every day via Internet. Username with password is the commonly used authentication mechanism. Most of the text based authentication methods are vulnerable to many attacks as they depend on text and can be strengthened more by combining password with key typing manner of the user.

Unauthorized access to highly confidential data threatens the security of the computer. The first and foremost step in preventing unauthorized access of information for providing information security is User Authentication.

1.2. CLASSIFICATION OF USER AUTHENTICATION

The confidential information can be secured from unauthorized users by providing authentication. *User authentication is defined as the process of verifying the identity claimed by an individual.* The general classification is shown in Figure 1.1.

User Authentication is normally classified into three categories [48] namely,

- Knowledge based Authentication - Something the user knows.
- Object based Authentication - Something the user has.
- Biometric based Authentication - Something you are.
1.2.1. Knowledge based Authentication

Knowledge based Authentication is based on “something the user knows”. User name and passwords come under this category. It is popular because of the following reasons:

- A password does not require any special software for incorporating.
- Passwords are portable as the users memorize them.
- Passwords authenticate the user directly because the user only knows the password.

Password systems suffer from the following problems:

- Passwords are vulnerable to password hackers because the users sometimes can't remember the strong passwords and they write them down.
- The recovery of forgotten passwords is either expensive or insecure.
- When the users share passwords, revenue is lost when multiple users share an account.
A password is a shared secret since the web site as well as the user knows it. An administrator can discover the password and use it to imitate the authenticated user.

The user must have a unique password for each site because a single password cannot be safely used with multiple web sites, since each site administrator can discover the password.

1.2.2. Object based Authentication

Something the user has or possesses is called Object based authentication. Tokens or Personal Identification Numbers (PINs) are the widely used examples of authentication through “something the user has”. Tokens have the following merits:

- Tokens prevent the stolen password from accessing the website by the unauthorized user as he/she must steal the physical token as well as the password.
- Tokens prevent accounts from being shared since the token must be duplicated.
- Tokens are portable.
- A token does not require any special software on the user's computer.

Object based Authentication has the following limitations:

- Tokens are expensive and must be replaced every few years.
- Tokens are easy to misplace or damage.
- Tokens are not convenient for the user since he/she must manually enter the token value as well as the password.
1.2.3. Biometric based Authentication

The final method, Biometric based authentication, provides a very reliable method of authenticating a user. The most common methods of authentication are fingerprint, iris, voice, retina pattern, typing behavior etc. They are difficult to duplicate. The hope for biometrics is based on the following:

- Biometrics cannot be stolen, shared, guessed, lost or forgotten unlike passwords and tokens.
- Biometric authentication is very strong because they are difficult to steal.
- Biometric systems authenticate a person without the necessity to carry or to memorize anything.

Biometric based authentication has the following limitations:

- Appropriate biometric sensor and software must be installed on the users' computers which are tending to be expensive.
- It leaks the personal information of the body.

1.3. BIOMETRICS OVERVIEW

Biometric is a measurement of a biological characteristic such as fingerprint, iris pattern, retina image, face or hand geometry; or a behavioral characteristic such as voice, gait or signature. The biometric technology uses these characteristics to identify individuals automatically. These characteristics are universally present, unique to an individual, stable over time and can be easily measured. Since the user has to present himself/herself during the authentication, it is not necessary to memorize or carry anything at the time of authentication. Biometrics also reduce password administration costs. Different biometrics will be suitable for different kinds of applications depending on whether the aim is to
identify someone with their cooperation or from a distance without their knowledge.

1.3.1. Types of Biometrics

Biometrics involve something a person is or does. Biometric Authentication is divided into physiological and behavioral types [28]. Physiological biometrics refer to what the person is or they measure physical parameters of certain parts of the body. Examples of Physiological biometrics are fingerprints, iris, retina, hand geometry, face etc. Behavioral biometrics show how the person is using the body for authentication. Some examples are keystroke dynamics, mouse dynamics, signature, gait, voice etc. Figure 1.2 shows an example of different types of biometrics.

1.3.2. Analysis of Biometrics Types

Each type of biometrics has its own advantages and limitations. Physiological biometrics are considered to be extremely robust and hence more secure. However, these types of biometrics are not portable and cannot be used for web based applications. Behavioral biometrics such as voice, gait etc., face the problem of high implementation cost as they require special hardware or sensors for capturing the template. An analysis of different types of biometrics is given in the following Table 1.1.
Figure 1.2 Different Types of Biometrics
### Table 1.1 Analysis of Different Types of Biometrics

<table>
<thead>
<tr>
<th>Physiological Biometrics</th>
<th>Behavioral Biometrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>What the person is or measure physical parameters of a certain part of the body</td>
<td>What a person does or how the person uses the body</td>
</tr>
<tr>
<td>Examples: Fingerprint, Iris, Face Recognition, etc.,</td>
<td>Examples: Voice, Handwriting, walking gait, Typing rhythm or Keystroke Dynamics</td>
</tr>
<tr>
<td>Limitations:</td>
<td>Limitations:</td>
</tr>
<tr>
<td>• Can be spoofed</td>
<td>• Voice Recognition, the most promising technology but it is difficult to distinguish a live user from a recorded one</td>
</tr>
<tr>
<td>• Difficulties in implementing</td>
<td>• Handwriting, gait requires a special device, which can be cost prohibitive and signatures can be forged</td>
</tr>
<tr>
<td>• Expensive</td>
<td></td>
</tr>
<tr>
<td>• Requires special Hardware</td>
<td></td>
</tr>
<tr>
<td>• More Intrusive</td>
<td></td>
</tr>
</tbody>
</table>

### 1.4. KEYSTROKE DYNAMICS AS BIOMETRICS

Keystroke dynamics is a strong behavioral biometric [1] and it is the process of analyzing the way a user types at a terminal by monitoring the keyboard in order to identify the user based on habitual typing rhythm patterns. Unlike other biometric systems, which may be expensive to implement, keystroke dynamics is almost free and it does not require any sophisticated hardware as the only hardware required is the keyboard.

#### 1.4.1. Merits of using Keystroke Dynamics

The merits of using Keystroke dynamics are as follows:
- It does not require any special equipment.
- It is user-friendly and noninvasive.
- Flexible enrollment is possible.
- The typing rhythm of the person cannot be lost or forgotten.
- If the template is stolen or guessed, the new one can be easily generated. So it is the only resettable biometric.
- It can be used for remote applications over the Internet.
- Keystroke dynamics can be combined with other authentication technologies.

1.4.2. Keystroke Dynamics Approaches

Keystroke verification contains two approaches Static and Continuous [37]. In Static approach, the system checks the user only at the authentication time. It provides additional security than the username/password and provides more robust user verification than simple passwords. In this approach, the analysis is performed on typing samples produced using the same predetermined text for all the individuals under observation [37]. The static analysis is done at login time in conjunction with other authentication methods such as passwords.

In the Continuous approach [69], the system checks the user continuously throughout the session and the user’s typing behavior is monitored throughout his/her typing time using the keyboard. It means that even after a successful login, the user typing patterns are constantly analyzed. The continuous approach is not considered in the thesis because the fatigue factor will have a play in the entire process. The approach used in the thesis is Static since the authentication is done during login time.
1.4.3. Application areas of Keystroke Dynamics

Keystroke Dynamics finds its applications in the following vital areas namely,

- Internet Banking
- Online Examination
- ATM
- Mobile Phones
- Email Security
- Detecting Synthetic Forgeries
- Simply wherever the keystroke (Typing) is involved

1.5. PROBLEM JUSTIFICATION

For the remote access applications such as Internet banking, online examination, payment systems etc., the authentication methods are based on knowledge or objects and are not secured as the authentication is done either with the passwords or Personal Identification Number (PIN). The usage of simple passwords is often easy to guess using either dictionary [56] or brute force attacks and PIN may be lost or stolen. These methods may not be sufficient and additional security measures should be taken.

Utilizing Biometric authentication for secure and comfortable access control is becoming convenient and considerably more accurate than current methods like passwords or PINs. The significant limitations of most biometric authorization or verification techniques are the need for specialized hardware like camera, microphone, finger scanner, hand scanner, drawing tablet, signature tablet, and others. These hardware may not be available at the remote locations since they are not portable. Therefore using keystroke dynamics authorization techniques is useful because of lack of any sophisticated hardware requirements.
Moreover, keystroke can be easily integrated with other authentication systems passwords or PINs. Keystroke when combined with traditional password authentication can be a powerful multifactor authentication technology. The accuracy of such technology should be high for high security applications. Hence an attempt is made in this thesis to improve the accuracy using wrapper based approach.

1.6. **OBJECTIVES OF THE THESIS**

The major objectives formulated in the thesis are as follows:

- To extract the new keystroke features for better performance.
- To reduce the dimensionality of the extracted features.
- To improve the classification accuracy and
- To reduce the Training and Testing time.

1.7. **SIGNIFICANT CONTRIBUTIONS**

The following are the significant contribution of this thesis work:

i. A new feature, ‘Virtual Key force’ (VKF) has been introduced.

ii. Wrapper based feature subset selection is proposed using the following methods:

e. Particle Swarm Optimization-Extreme Learning Machine (PSO-ELM).


iii. Wrapper based classification using the following classifiers:


1.8. ORGANIZATION OF THE THESIS

The rest of the thesis is organized as follows:

- Chapter 2 provides the review of existing work in the area of static keystroke dynamics.
- Chapter 3 explains the proposed methodology and discusses the different phases of proposed methodology elaborately.
- Chapter 4 discusses the experiments conducted and results.
- Chapter 5 concludes the thesis with future scope.

1.9. CHAPTER SUMMARY

This chapter discussed the overview of the user authentication and its types. The need for keystroke dynamics and their applications are elaborated. The objectives of the thesis and the significant contribution are given in this chapter. The next chapter discusses the related works.