Chapter 16
Intrusion Detection Using Keystroke Dynamics

Mahalaxmi Sridhar, Teby Abraham, Johnelle Rebello, Winchell D'souza and Alishia D'Souza

Abstract In an effort to confront the challenges brought forward by the increased need for access control, we present an improved technique for authorized access to computer system resources and data via keystroke dynamics. A database of keystrokes of login ids and passwords collected from 38 users is constructed. From the samples collected, signatures were formed using three membership functions of Fuzzy Logic. Users were authenticated by comparing the typing pattern to their respective signatures. We have included the usage of the SHIFT and the CAPS LOCK keys as part of the feature sets. We analyzed the performance of the three membership functions of Fuzzy Logic based on features like FAR and FRR to evaluate the efficiency of the detection algorithms. The paper presents the results of the analysis thereby providing an inexpensive method of intrusion detection as compared to other behavioral biometric methods.

Keywords Keystroke dynamics · Intrusion detection · Fuzzy logic · Computer security

16.1 Introduction

One of the primary means of authenticating users and providing security to computers are textual passwords. Passwords are convenient and require no specialized hardware. However, users frequently share password with others,

---

M. Sridhar (✉) · T. Abraham · J. Rebello · W. D'souza · A. D'Souza
Don Bosco Institute of Technology, Kurla, Mumbai, Maharashtra, India
e-mail: mahalaxmi90sridhar@gmail.com

T. Abraham
e-mail: projectkd2011@gmail.com

V. V. Das (ed.), *Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing*, Lecture Notes in Electrical Engineering 150,
DOI: 10.1007/978-1-4614-3363-7_16,

forget passwords, and select poor passwords that may be easily defeated. Compromised passwords and shared accounts are frequently exploited by both external attackers and insiders.

One idea to overcome this is to use keystroke dynamics. It is a novel approach in which a legitimate user's typing patterns such as durations of keystrokes, latencies between keystrokes etc. are combined with the user's password to generate a hardened password that is convincingly more secure than conventional passwords.

16.2 Literature Survey

Keystroke Dynamics is the manner and rhythm in which an individual types characters on a keyboard or keypad. It gives the detailed timing information that describes exactly when each key was pressed and released as a person is typing. Ever since Forsen et al. [1] investigated for the first time whether users could be distinguished by the way they type many different techniques for keystroke dynamics have been proposed.

In almost every technique the common feature sets used to form the signatures are:

- Enter: the Enter key is considered to be part of the password.
- KeyUp–KeyUp: Time between the key releases of consecutive keys is used as a feature.
- KeyUp–KeyDown: Time between the release of one key and the press of the next is used.
- KeyDown–KeyDown: Time between the key presses of consecutive keys is used as a feature.

16.2.1 Anomaly Detectors for Password Timing

Our main focus is on developing an intrusion detection system using the static login method. Various studies have been done on the use of anomaly detectors to analyze password-timing data.

Table 16.1 summarizes some of the anomaly detectors along with their results relevant to our study. False accept rate (FAR) denotes the rate that an imposter is allowed access. Similarly False reject rate (FRR) denotes the rate that the legitimate user is denied access. After thoroughly studying various anomaly detectors summarized in the Table 16.1 we concluded that fuzzy logic has a reasonable balance between FRR and FAR errors. Hence we planned to implement it using various membership functions.

Table 16.1 Comparison of various anomaly detectors and their error rates [4]

Algorithms	Feature sets			Results	
	Enter key	Keydown–keydown	Keyup–keydown	FRR	FAR
Euclidean		Y		2.8	8.1
Manhattans	Y	Y		0.25	16.4
Mahalanobis		Y		2.8	8.1
Neural-network		Y		0.2	0.22
Fuzzy-logic		Y		0.11	0.19
z-score		Y		0.02	0.13
K-means			Y	3.8	3.8

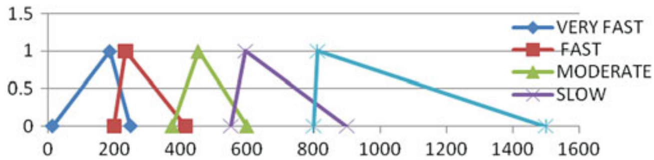


Fig. 16.1 Fuzzy sets for triangle membership function

16.3 Design

Fuzzy logic is a form of many-valued logic or probabilistic logic; it deals with reasoning that is approximate rather than fixed and exact. In contrast with traditional logic theory, where binary sets have two-valued logic: true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Every element requires a degree of membership to determine how strongly it belongs to a certain class. Degree of membership is calculated by a Membership Function. Of the many membership functions for fuzzy logic we have selected three membership functions for our study:

- Triangle
- Trapezoidal
- Gaussian

16.3.1 Formation of Intervals

A dedicated software module was designed to collect the features of 35 volunteers. Data collected from these volunteers were stored in a database and used to form the intervals of various classes; where each class represents different typing speeds. Different classes of typing speed that we decided for our project are: Very Fast, Fast, Moderate, Slow and Very Fast.

Based on the sample collected, the intervals for the three membership function mentioned before were designed as follows. (Figs. 16.1, 16.2, 16.3, 16.4).

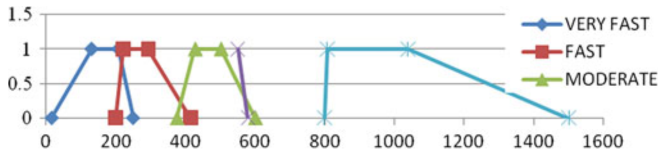


Fig. 16.2 Fuzzy sets for trapezoidal membership function

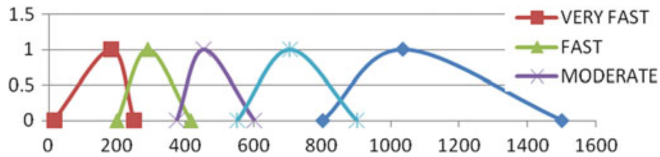
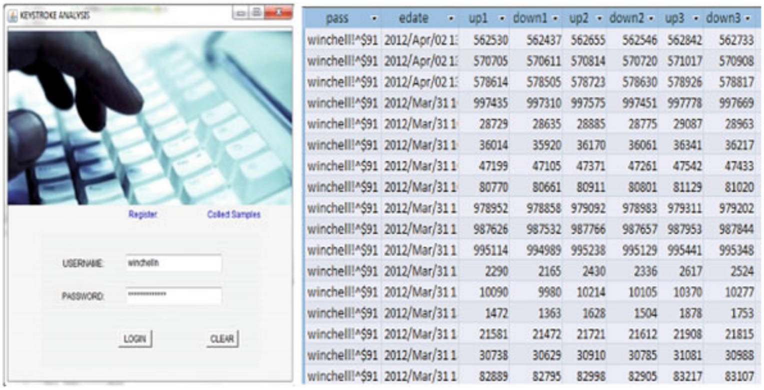


Fig. 16.3 Fuzzy sets for Gaussian membership function



pass	edate	up1	down1	up2	down2	up3	down3
winchell!^$91	2012/Apr/02 1	562530	562437	562655	562546	562842	562733
winchell!^$91	2012/Apr/02 1	570705	570611	570814	570720	571017	570908
winchell!^$91	2012/Apr/02 1	578614	578505	578723	578630	578926	578817
winchell!^$91	2012/Mar/31 1	997435	997310	997575	997451	997778	997669
winchell!^$91	2012/Mar/31 1	28729	28635	28885	28775	29087	28963
winchell!^$91	2012/Mar/31 1	36014	35920	36170	36061	36341	36217
winchell!^$91	2012/Mar/31 1	47199	47105	47371	47261	47542	47433
winchell!^$91	2012/Mar/31 1	80770	80661	80911	80801	81129	81020
winchell!^$91	2012/Mar/31 1	978952	978858	979092	978983	979311	979202
winchell!^$91	2012/Mar/31 1	987626	987532	987766	987657	987953	987844
winchell!^$91	2012/Mar/31 1	995114	994989	995238	995129	995441	995348
winchell!^$91	2012/Mar/31 1	2290	2165	2430	2336	2617	2524
winchell!^$91	2012/Mar/31 1	10090	9980	10214	10105	10370	10277
winchell!^$91	2012/Mar/31 1	1472	1363	1628	1504	1878	1753
winchell!^$91	2012/Mar/31 1	21581	21472	21721	21612	21908	21815
winchell!^$91	2012/Mar/31 1	30738	30629	30910	30785	31081	30988
winchell!^$91	2012/Mar/31 1	82889	82795	82998	82905	83217	83107

Fig. 16.4 Snapshots of software used to collect keystroke samples and the associated database

16.3.1.1 Triangle Membership Function

$$
\begin{array}{llll}
\Lambda(\mathrm{u}{:}\alpha,\beta,\gamma) & = 0 & \mathrm{u} < \alpha & \text{Where} \quad \alpha - \text{ Lower Bound Value} \\
 & = (\mathrm{u}-\alpha)/(\beta-\alpha) & \alpha <= \mathrm{u} <= \beta & \beta - \text{ Modal Value} \\
 & = (\alpha - \mathrm{u})/(\beta-\alpha) & \beta <= \mathrm{u} <= \gamma & \Gamma - \text{ Higher Bound Value} \\
 & = 0 & \mathrm{u} > \gamma &
\end{array}
\qquad (16.1)
$$

16.3.1.2 Trapezoidal Membership Function

$$\begin{aligned} F(x, a, b, c, d) &= 0 && \text{when } x < a \text{ and } x > d \\ &= (x - a)/(b - a) && \text{when } a <= x <= b \\ &= 1 && \text{when } b <= x <= c \\ &= (d - x)/(d - c) && \text{when } c <= x <= d \end{aligned} \quad (16.2)$$

16.3.1.3 Gaussian Membership Function

$$G\ (u:\ m, \sigma) = \exp[-\{(u - m)/\sqrt{2}\sigma\}2] \quad \text{Where m---Mean Value} \quad (16.3)$$

16.4 Implementation

16.4.1 Sample Collection and Signature Formation

Inter-key delays were collected using dedicated software and the samples were stored in a simple database.

Once sufficient samples were collected, a minimum of 15 samples for each user were collected, intervals generated and algorithms for each of the three membership functions are generated. An example of a simple algorithm [2] implementing the triangle membership function as an anomaly detector is shown below. Similar algorithms were developed by us for the other two membership functions.

If (Input < LowerBound OR Input > UpperBound)

Then 0

Else If (Input < Midvalue)

Then (Input − LowerBound)/(Midvalue − LowerBound)

Else If (Input = Midvalue)

Then 1

Else (UpperBound − Input)/(UpperBound − Midvalue)

The Feature Sets used for our study are listed below in Table 16.2. We included the SHIFT and the CAPS LOCK key in the feature sets of our fuzzy logic. It is often observed that the tendency to use the RIGHT_SHIFT or the LEFT_SHIFT or CAPS LOCK to type special characters and upper-case letters differ from user to

Table 16.2 Feature sets

Algorithm	Feature sets			
	Keydown-keyup	Keyup–keyup	Shift key	Caps lock
Fuzzy logic	No	Yes	Yes	Yes

user [3]. This variation can thus be used as an additional parameter to validate legitimate users from imposters.

A membership function calculates the degree of membership to each class for each inter-key delay (KeyUp–KeyUp) given as input.

Based on the input a signature for a particular user is determined. One such signature formed is shown in Table 16.3.

16.4.2 Signature Comparison

In the working phase the real time signature of a user is compared with the stored signature. If both signatures match up to a certain limit (in this case it is up to 70 %) then the user is verified as the genuine user and granted access; else they are not granted access.

16.5 Testing

To increase the confidence in the correctness (accuracy) of specified membership function of Triangular, Trapezoidal and Gaussians, we conducted testing by supplying typical test inputs (request) and subsequently checking test output (responses) against expected ones to enhance the correctness of specified algorithm (Fig. 16.5)

As we can see from Table 16.4, comparison of the FAR and FRR of all the three membership function shows that Gaussian function yields the best results as compared to the other two membership functions.

16.6 Conclusion and Future Scope

We believe keystroke dynamics can be used effectively to safeguard against unauthorized access of computer as well as mobile resources [2]. When implemented in conjunction with traditional schemes, it allows for the design of more robust authentication systems than traditional password based alternatives alone.

Table 16.3 Signature formation and comparison

Signature	1st	2nd	3rd	4th	5th	6th	7th	8th	Shift left	Shift right	Caps lock
Stored	Very fast	Fast	Moderate	Very fast	Fast	Fast	Very fast	Very fast	1	0	0
Detected	Fast	Fast	Moderate	Very fast	Slow	Fast	Very fast	Very fast	0	1	0

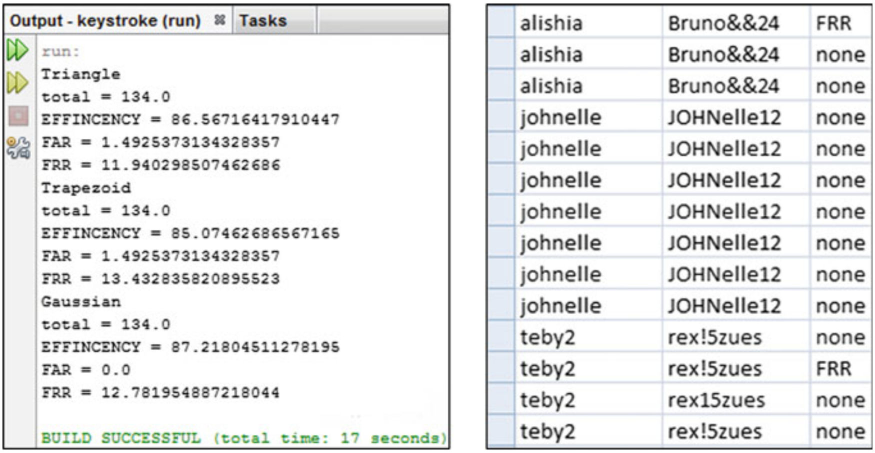


Fig. 16.5 Snapshots showing the experimental results generated, stored and evaluated

Table 16.4 Performance measure of membership functions

	FAR	FRR	Accuracy (%)
Triangle	1.4	11.19	87.41
Trapezoidal	1.4	12.59	86.01
Gaussian	0.38	11.97	88.03

In this project we compared triangular, trapezoidal and Gaussian membership functions of fuzzy logic to authenticate users based on their typing speed and proved that among the three, Gaussian membership function is the most effective means of implementing an intrusion detection system using Fuzzy logic. To implement such a system, the code developed by us in Java could be used as a plug-in for intrusion detection, once the database has been created for the authentic users.

The approach of using keystroke dynamics in our project was limited only to passwords. This can be extended to include all the text typed by a user during his work. This way, not only will there be monitoring at the login stage but also during the entire active session for a particular user.

References

1. Maxion RA, Killourhy KS (2010) Keystroke biometrics with number-pad input, Computer Science Department, Carnegie Mellon University
2. Haider S, Abbas A, Zaidi AK (2000) A multi-technique approach for user identification through keystroke dynamics. In: IEEE international conference on systems, man and cybernetics
3. Killourhy KS, RA Maxion Comparing anomaly-detection algorithms for keystroke dynamics
4. Forsen G, Nelson M, Staron R Jr (1977) Personal attributes authentication techniques. Technical Report RADC-TR-77-333, Rome Air Development Center
5. Ahmed Awad EA, Traore I Detecting computer intrusions using behavioural biometrics
6. Monrose F, Rubin AD (1999) Authentication via keystroke dynamics
7. Killourhy KS (2012) A scientific understanding of keystroke dynamics
8. Joyce and G. Gupta (1990) Identity authentication based on keystroke latencies. Commun ACM
9. Ahmed AAE, Traore I Department of Electrical and Computer Engineering, University of Victoria. Detecting computer intrusions using behavioural biometrics
10. Lane Department of Computer Science and Electrical Engineering (2005) Morgantown, West Virginia, Username and password verification through keystroke dynamics
11. Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley
12. Mandal SN, Choudhury JP, De D, Chaudhuri SRB (2008) Roll of membership functions in fuzzy logic for prediction of shoot length of mustard plant based on residual analysis