Security Enhancement of Knowledge-based User Authentication through Keystroke Dynamics

Soumen Roy[1], Utpal Roy[2], D. D. Sinha[1]

[1]*Department of Computer Science and Engineering, University of Calcutta, 92 APC Road, Calcutta -700 009, INDIA*
[2]*Department of Computer & System Sciences, Visva-Bharati, Santiniketan -731235, INDIA*

Abstract- Keystroke Dynamics is a behavioural biometrics characteristic in Biometric science, which solve the issues in user identification or verification. In Knowledge-based user authentication technique, we compromise with PIN or password which is unsafe due to different type of attacks. It is good to choose password with the combination of upper and lower case letter with some digits and symbols, but which is very hard to remember or generally we forget to distinguish those passwords for different access control systems. Our system not only takes the users' entered texts but their typing style is also account for. In our experiment, we have not taken hard password type texts, we have chosen some daily used words where users are habituated and comfortable at typing and we obtained the consisting typing pattern. Different distance-based and data mining algorithms we have applied on collected typing pattern and obtained impressive results. As per our experiment, if we use keystroke dynamics in existing knowledge based user authentication system with minimum of five daily used common texts then it increases the security level up to 97.6% to 98.2% (if we remove some of the irrelevant feature sets).

1 Introduction

Keystroke dynamics is a behavioral biometrics which is the method of analysing the way a user types on a keyboard and classify user based on regular typing rhythm. It is the study of whether people can be well-known by their typing rhythms, much like handwriting is used to recognize the author of a written text. A user's typing pattern may be unique because similar neuro-physiological factors that make written signatures unique. Here user can choose any text for password from his/her own dictionary. It is very simple and nothing to remember still it enhances the security level and can be used to identify an individual. Recognizing typing style promises a parameter like biometric characteristics that may facilitate non-intrusive, cost-effective and continuous monitoring.

System takes comprising of characters as well as the typing style of each subsequent character entered. It facilitates that no one can track the time or presses the character of password in same rhythm. It will prevent our system from off-line guessing attracts and also prevent to track by un-authorised people. Our objective is to minimize the probability of off-line guessing attacks, hide the password from public, minimize the hardware cost and minimize the software cost by making faster pattern recognition.

In our experiment we have consider five daily used strings ("kolkata123", "facebook", "yahoo.com", "gmail.com" and "123456") as password string laptop keypad as hardware configuration. We have collected the keystroke raw data samples from 12 users in four sessions, in each session 6 times using JAVA swing. Then we extracted all the features (key duration, up-down key latency, down-up key latency, up-up key latency, down-down key latency, tri-gap timing, four-gap timing and total timing) and we analysed features and combination of features using R Statistical tool and Microsoft Excel.

We have tested our raw data using eight different score-based, distance-based and features mining algorithms and we got 2.4% of Equal Error Rate for Z-score algorithm where we have consider all the five fixed strings and all the mentioned features.

2 Keystroke Dynamics

2.1 Basic Idea

Keystroke dynamics is set of some timing data or keystroke pressure data which is generated at typing on keyboard which is unique and can be used to classify the users.

2.2 Science and Features Selection

Placement of fingers on keyboard, hand weight, length of finger, neuro-physiological factors make typing style unique, where some timing factors easily

[a] Corresponding author: soumen.roy_2007@yahoo.co.in

calculate these pattern factors. We have calculated key press and release time P_i and R_i for all keys K_i which represent entered character set. Where $1 \le i \le$ length of the entered word. The features of the keystroke dynamics as follows:

Key Duration $(T_1)=R_i-P_i$ (1)
Up Up Key Latency $(T_2)=R_{i+1}-R_i$ (2)
Down Down Key Latency $(T_3)=P_{i+1}-P_i$ (3)
Up Down Key Latency $(T_4)=P_{i+1}-R_i$ (4)
Down Up Key Latency $(T_5)=R_{i+1}-P_i$ (5)
Total Time Key Latency $(T_6)=R_n-P_1$ (6)
Tri-graph Latency $(T_7)=R_{i+2}-P_i$ (7)
Four-graph Latency $(T_8)=R_{i+3}-P_i$ (8)

Some new features are keystroke pressure, finger tips size, finger placement on keyboard, keystroke sound, error correcting mechanism, sequence of left-right control keys.

2.3 Security Issues

Among various user authentication techniques knowledge-based, token-based and biometric-based authentication techniques, biometric authentication is most popular for their uniqueness characteristics and cannot be stolen or there is no chance to loss. Keystroke Dynamics is a behavioral characteristic which is unique and can be effectively implemented with the existing system with minimal alternation. It can be used as a safe guard of our password from different type of attacks.

2.4 Factors Affecting Performance

Some of the factors which affect the way of keystroke Dynamics as follows: Text length, sequences of character types, word choice, and number of training sample, statistical method to create template, mental state of the user, tiredness or level of comfort, keyboard type, keyboard position and height of the keyboard, hand injury, weakness of hand mussel, shoulder pain, education level, computer knowledge, and category of users.

2.5 Algorithms

Many classification methods have been applied in keystroke dynamics study over the last three decades. Following are the anomaly detector algorithm described in [8].

2.6.1 Canberra:

$$D_{car}=\sum_i^n \frac{|P_i-Q_i|}{P_i+Q_i} \quad (9)$$

2.6.2 Chebyshev:

$$D_{cheb}=\sum_i^n \max|P_i-Q_i| \quad (10)$$

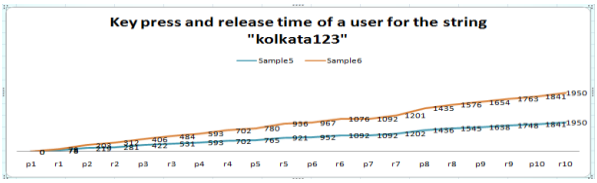

Figure 1. Two different samples from a user in line chart

2.6 Keystroke Dynamics as User Authentication

There are different ways in which a user can be authenticated. However all of these ways can be categorized into one of three classes: “Something we know” e.g. password, “Something we have” e.g. token, “Something we are” e.g. biometric property.

2.6.3 Czekanowski:

$$D_{cze}=\frac{\sum_i^n |P_i-Q_i|}{\sum_i^n (P_i+Q_i)} \quad (11)$$

2.6.4 Gower:

$$D_{gow}=\frac{1}{n}\sum_i^n |P_i-Q_i| \quad (12)$$

2.6.5 Intersection:

$$D_{ins}=\frac{1}{2}\sum_i^n |P_i-Q_i| \quad (13)$$

2.6.6 Kulczynski:

$$D_{kuld}=\frac{\sum_i^n |P_i-Q_i|}{\sum_i^n \min(P_i,Q_i)} \quad (14)$$

2.6.7 Lorentzian:

$$D_{lor}=\sum_i^n \ln(1+|P_i-Q_i|) \quad (15)$$

2.6.8 Minkowski:

$$D_{mink}=\sqrt[p]{\sum_i^n |P_i-Q_i|^p} \quad (16)$$

2.6.9 Motyka:

$$D_{mot}=\frac{\sum_i^n \max(P_i)}{\sum_i^n (P_i+Q_i)} \quad (17)$$

2.6.10 Ruzicka:

$$D_{ruz}=1-\frac{\sum_i^n \min(P_i,Q_i)}{\sum_i^n \max(P_i,Q_i)} \quad (18)$$

2.6.11 Soergel:

$$D_{soe}=\frac{\sum_i^n |P_i-Q_i|}{\sum_i^n \max(P_i,Q_i)} \quad (19)$$

2.6.12 Sorensen:

$$D_{sor}=\frac{\sum_i^n |P_i-Q_i|}{\sum_i^n (P_i+Q_i)} \quad (20)$$

2.6.13 Wavehedges:

$$D_{wv}=\frac{\sum_i^n |P_i-Q_i|}{\sum_i^n \max(P_i,Q_i)} \quad (21)$$

2.6.14 Manhattan Distance:

$$M=\sum_{i=1}^n (|P_i-Q_i|) \quad (22)$$

2.6.15 Euclidean Distance:

$$E=\sqrt[2]{\sum_i^n (|P_i-Q_i|)^2} \quad (23)$$

2.6.16 Mahanobolis Distance:

$$Eh = \sqrt[2]{\sum_i^n((|P_i - Q_i|)/\alpha_i)^2} \quad (24)$$

2.6.17 Z Score:

$$Z=\sum_{i=1}^{n}(|P_i| - \mu(|Q_i|))/\alpha_i \quad (25)$$

3 Background details

In 30+ years of experience, many researchers have proposed their algorithms, taking various features, various length of pattern string.

Table 1. Background of keystroke dynamics

Authors	Classifiers	Length of the pattern	Features	EER (%)
Joyce & Gupta [2]	Manhattan	33	UD	0.25-16.36
Bleha et al. [1]	Euclidian	11-17	UD	2.8-8.1
Haider et al. [6]	Nural Network	7	UD	16.1
Yu & Cho [5]	SVM	6-7	UD	10.2
Killourly S. [4]	Manhattan (Scaled)	10+	UD	9.6
Giot et al. [7]	SVM	100	KD, UD	15.28

4 Experimental results

We have implemented a program in JAVA for experimental purpose, which has the capability of capturing all key pressing and releasing events, which are used to create the database of different sample of passwords and timing templates. Here we have calculated average equal error rate for all eight algorithms considering some single feature and combination of features for all five strings.

Table 2. Average equal error rate for fixed-text "kolkata123"

FEATURES & COMBINATION OF FEATURES	Euclidean	Manhattan	Scaled Manhattan	Outlier Count	Mahalanobis	KMeans	AutoAssocNNet	SVM
UU	0.163	0.163	0.140	0.152	0.180	0.128	0.168	0.130
UD+DU	0.151	0.139	0.106	0.102	0.147	0.112	0.150	0.108
UD + Tri	0.160	0.151	0.123	0.129	0.144	0.134	0.149	0.120
UD	0.175	0.174	0.159	0.176	0.222	0.167	0.167	0.160
Tri Gap	0.166	0.159	0.135	0.143	0.175	0.140	0.157	0.124
Tri + 4gap	0.178	0.177	0.157	0.137	0.153	0.124	0.159	0.136
KD+DD+UU+DU+UD+tri+4gap	0.166	0.147	0.113	0.089	0.141	0.152	0.155	0.150
KD+DD+UU+DU+UD+tri	0.155	0.132	0.114	0.092	0.144	0.151	0.154	0.149
KD+DD+UU+DU+UD	0.151	0.128	0.117	0.094	0.141	0.151	0.157	0.149
KD + UD + Tri + 4gap	0.172	0.161	0.110	0.092	0.137	0.155	0.149	0.147
KD + UD + Tri	0.157	0.137	0.127	0.095	0.136	0.149	0.144	0.144
KD + UD	0.164	0.126	0.134	0.101	0.169	0.154	0.150	0.160
KD + Tri	0.163	0.149	0.131	0.105	0.159	0.159	0.150	0.156
KD + DU + UD + Tri + 4gap	0.169	0.152	0.115	0.088	0.140	0.150	0.154	0.150
KD + DU + UD + tri	0.157	0.130	0.116	0.093	0.138	0.150	0.147	0.148
KD + DU + Tri + 4gap	0.172	0.156	0.112	0.092	0.135	0.150	0.157	0.146
KD + DU + Tri	0.157	0.129	0.126	0.096	0.141	0.144	0.145	0.145

KD + DU	0.150	0.109	0.134	0.097	0.143	0.162	0.138	0.165
KD + 4gap	0.177	0.180	0.130	0.109	0.145	0.166	0.172	0.160
KD	0.153	0.143	0.179	0.170	0.276	0.204	0.153	0.208
DU + Tri	0.159	0.144	0.115	0.114	0.168	0.118	0.156	0.111
DU	0.158	0.142	0.112	0.125	0.158	0.119	0.157	0.112
DD+UU	0.156	0.150	0.123	0.123	0.157	0.127	0.160	0.112
DD	0.151	0.147	0.122	0.142	0.157	0.132	0.155	0.120
All	0.175	0.156	0.113	0.089	0.140	0.152	0.167	0.150
4gap	0.181	0.201	0.197	0.201	0.159	0.153	0.169	0.148

Table 3. Average equal error rate for fixed-text “yahoo.com”

FEATURES & COMBINATION OF FEATURES	**Euclidean**	**Manhattan**	**Scaled Manhattan**	**Outlier Count**	**Mahalanobis**	**KMeans**	**AutoAssocNNet**	**SVM**
UU	0.224	0.201	0.196	0.199	0.269	0.199	0.248	0.197
UD+DU	0.226	0.200	0.187	0.172	0.218	0.182	0.239	0.192
UD + Tri	0.223	0.216	0.196	0.204	0.202	0.203	0.215	0.176
UD	0.229	0.215	0.200	0.209	0.244	0.196	0.234	0.189
Tri Gap	0.228	0.228	0.215	0.239	0.243	0.216	0.207	0.216
Tri + 4gap	0.231	0.234	0.212	0.230	0.259	0.220	0.211	0.221
KD+DD+UU+DU+UD+tri+4gap	0.224	0.209	0.145	0.127	0.249	0.170	0.229	0.158
KD+DD+UU+DU+UD+tri	0.218	0.200	0.142	0.118	0.246	0.169	0.236	0.165
KD+DD+UU+DU+UD	0.226	0.195	0.139	0.109	0.253	0.169	0.235	0.179
KD + UD + Tri + 4gap	0.227	0.212	0.146	0.124	0.233	0.160	0.218	0.173
KD + UD + Tri	0.220	0.193	0.141	0.109	0.218	0.168	0.224	0.173
KD + UD	0.221	0.184	0.138	0.095	0.246	0.176	0.229	0.170
KD + Tri	0.226	0.213	0.131	0.135	0.224	0.149	0.222	0.154
KD + DU + UD + Tri + 4gap	0.226	0.210	0.148	0.122	0.242	0.161	0.223	0.168
KD + DU + UD + tri	0.220	0.196	0.140	0.110	0.233	0.162	0.222	0.169
KD + DU + Tri + 4gap	0.230	0.219	0.139	0.127	0.257	0.162	0.226	0.173
KD + DU + Tri	0.221	0.203	0.136	0.115	0.260	0.169	0.222	0.177
KD + DU	0.228	0.176	0.141	0.101	0.261	0.171	0.240	0.174
KD + 4gap	0.229	0.217	0.141	0.145	0.231	0.161	0.211	0.164
KD	0.158	0.155	0.187	0.163	0.321	0.205	0.158	0.217
DU + Tri	0.222	0.225	0.190	0.193	0.252	0.197	0.221	0.181
DU	0.236	0.205	0.175	0.188	0.242	0.206	0.243	0.189
DD+UU	0.235	0.212	0.196	0.196	0.264	0.196	0.256	0.189
DD	0.250	0.223	0.198	0.213	0.236	0.203	0.251	0.194
All	0.211	0.203	0.145	0.128	0.249	0.170	0.215	0.161
4gap	0.234	0.245	0.241	0.263	0.262	0.228	0.247	0.232

Table 4. Average equal error rate for fixed-text "gmail.com"

FEATURES & COMBINATION OF FEATURES	Euclidean	Manhattan	Scaled Manhattan	Outlier Count	Mahalanobis	KMeans	AutoAssocNNet	SVM
UU	0.175	0.167	0.149	0.164	0.191	0.163	0.170	0.170
UD+DU	0.172	0.164	0.139	0.139	0.158	0.149	0.177	0.153
UD + Tri	0.186	0.168	0.141	0.158	0.172	0.144	0.180	0.150
UD	0.182	0.176	0.144	0.153	0.191	0.159	0.180	0.165
Tri Gap	0.184	0.191	0.169	0.206	0.203	0.173	0.188	0.173
Tri + 4gap	0.199	0.210	0.188	0.203	0.212	0.176	0.202	0.171
KD+DD+UU+DU+UD+tri+4gap	0.193	0.162	0.108	0.118	0.162	0.139	0.178	0.143
KD+DD+UU+DU+UD+tri	0.183	0.166	0.106	0.116	0.164	0.136	0.194	0.140
KD+DD+UU+DU+UD	0.135	0.151	0.153	0.161	0.194	0.144	0.134	0.142
KD + UD + Tri + 4gap	0.193	0.174	0.095	0.099	0.153	0.129	0.176	0.120
KD + UD + Tri	0.178	0.150	0.086	0.088	0.158	0.129	0.174	0.117
KD + UD	0.171	0.145	0.092	0.083	0.160	0.130	0.167	0.125
KD + Tri	0.196	0.169	0.092	0.099	0.151	0.105	0.173	0.099
KD + DU + UD + Tri + 4gap	0.193	0.170	0.099	0.104	0.158	0.128	0.192	0.122
KD + DU + UD + tri	0.181	0.157	0.093	0.098	0.162	0.127	0.180	0.128
KD + DU + Tri + 4gap	0.193	0.178	0.098	0.103	0.157	0.125	0.194	0.116
KD + DU + Tri	0.182	0.151	0.089	0.093	0.155	0.124	0.182	0.114
KD + DU	0.168	0.138	0.105	0.087	0.151	0.122	0.167	0.114
KD + 4gap	0.200	0.188	0.093	0.103	0.155	0.097	0.190	0.096
KD	0.135	0.151	0.153	0.161	0.194	0.143	0.133	0.140
DU + Tri	0.187	0.172	0.148	0.168	0.192	0.157	0.183	0.154
DU	0.175	0.164	0.148	0.151	0.172	0.159	0.184	0.161
DD+UU	0.184	0.173	0.163	0.160	0.188	0.170	0.182	0.164
DD	0.190	0.180	0.178	0.167	0.184	0.180	0.197	0.176
All	0.195	0.166	0.108	0.117	0.163	0.139	0.199	0.142
4gap	0.212	0.228	0.204	0.231	0.191	0.194	0.216	0.190

Table 5. Average equal error rate for fixed-text "facebook"

FEATURES & COMBINATION OF FEATURES	Euclidean	Manhattan	Scaled Manhattan	Outlier Count	Mahalanobis	KMeans	AutoAssocNNet	SVM

UU	0.204	0.173	0.117	0.152	0.205	0.158	0.221	0.148
UD+DU	0.190	0.155	0.105	0.129	0.168	0.124	0.199	0.123
UD + Tri	0.199	0.177	0.124	0.144	0.154	0.148	0.180	0.127
UD	0.220	0.185	0.140	0.169	0.197	0.164	0.205	0.165
Tri Gap	0.206	0.196	0.169	0.188	0.176	0.199	0.222	0.180
Tri + 4gap	0.209	0.210	0.173	0.188	0.255	0.191	0.193	0.181
KD+DD+UU+DU+ UD+tri+4gap	0.197	0.161	0.082	0.083	0.169	0.132	0.203	0.122
KD+DD+UU+DU+ UD+tri	0.190	0.146	0.081	0.080	0.169	0.121	0.186	0.116
KD+DD+UU+DU+U D	0.190	0.147	0.090	0.082	0.166	0.126	0.206	0.119
KD + UD + Tri + 4gap	0.200	0.178	0.099	0.099	0.162	0.133	0.185	0.129
KD + UD + Tri	0.188	0.148	0.113	0.096	0.146	0.131	0.176	0.128
KD + UD	0.205	0.136	0.116	0.099	0.153	0.133	0.198	0.137
KD + Tri	0.197	0.170	0.121	0.103	0.172	0.134	0.197	0.129
KD + DU + UD + Tri + 4gap	0.197	0.163	0.093	0.086	0.161	0.131	0.187	0.120
KD + DU + UD + tri	0.186	0.146	0.098	0.081	0.155	0.127	0.176	0.126
KD + DU + Tri + 4gap	0.202	0.170	0.096	0.097	0.174	0.126	0.198	0.130
KD + DU + Tri	0.195	0.150	0.110	0.092	0.181	0.129	0.183	0.128
KD + DU	0.176	0.116	0.106	0.099	0.176	0.130	0.197	0.129
KD + 4gap	0.209	0.187	0.125	0.116	0.205	0.128	0.197	0.127
KD	0.165	0.166	0.172	0.163	0.182	0.157	0.166	0.158
DU + Tri	0.204	0.171	0.122	0.145	0.169	0.148	0.189	0.142
DU	0.190	0.153	0.107	0.148	0.190	0.145	0.201	0.137
DD+UU	0.203	0.172	0.113	0.132	0.181	0.139	0.206	0.129
DD	0.210	0.166	0.117	0.145	0.175	0.144	0.214	0.139
All	0.195	0.158	0.080	0.084	0.169	0.132	0.189	0.113
4gap	0.219	0.231	0.209	0.237	0.200	0.198	0.229	0.213

Table 6. Average equal error rate for fixed-text “123456”

FEATURES & COMBINATION OF FEATURES	Euclidean	Manhattan	Scaled Manhattan	Outlier Count	Mahalanobis	KMeans	AutoAssocNNet	SVM
UU	0.217	0.207	0.202	0.231	0.228	0.229	0.212	0.211
UD+DU	0.159	0.163	0.152	0.170	0.285	0.173	0.163	0.163
UD + Tri	0.175	0.186	0.173	0.195	0.247	0.179	0.172	0.174
UD	0.232	0.232	0.232	0.263	0.258	0.236	0.221	0.224
Tri Gap	0.204	0.208	0.203	0.233	0.215	0.241	0.198	0.207
Tri + 4gap	0.210	0.214	0.205	0.223	0.243	0.227	0.197	0.214
KD+DD+UU+DU+ UD+tri+4gap	0.179	0.175	0.142	0.154	0.307	0.181	0.156	0.162
KD+DD+UU+DU+ UD+tri	0.176	0.172	0.153	0.154	0.323	0.188	0.173	0.170
KD+DD+UU+DU+UD	0.187	0.173	0.144	0.153	0.305	0.181	0.170	0.173
KD + UD + Tri + 4gap	0.159	0.155	0.146	0.156	0.323	0.182	0.160	0.162

KD + UD + Tri	0.179	0.170	0.180	0.172	0.323	0.193	0.181	0.173
KD + UD	0.183	0.173	0.166	0.166	0.313	0.193	0.194	0.179
KD + Tri	0.179	0.170	0.145	0.146	0.306	0.182	0.167	0.162
KD + DU + UD + Tri + 4gap	0.153	0.154	0.145	0.146	0.323	0.181	0.155	0.158
KD + DU + UD + tri	0.194	0.178	0.151	0.145	0.314	0.183	0.181	0.170
KD + DU + Tri + 4gap	0.176	0.170	0.157	0.150	0.323	0.184	0.172	0.174
KD + DU + Tri	0.162	0.146	0.163	0.169	0.323	0.189	0.159	0.169
KD + DU	0.197	0.170	0.184	0.188	0.303	0.215	0.189	0.188
KD + 4gap	0.259	0.261	0.272	0.288	0.353	0.294	0.259	0.286
KD	0.154	0.156	0.138	0.151	0.323	0.181	0.149	0.161
DU + Tri	0.196	0.199	0.192	0.197	0.317	0.217	0.199	0.192
DU	0.203	0.196	0.187	0.221	0.262	0.236	0.209	0.214
DD+UU	0.193	0.195	0.187	0.203	0.370	0.218	0.189	0.209
DD	0.198	0.204	0.196	0.228	0.253	0.229	0.194	0.221
All	0.189	0.179	0.146	0.155	0.306	0.180	0.160	0.165
4gap	0.218	0.222	0.216	0.262	0.229	0.251	0.233	0.240

Table 7. Average equal error rate for all combination of fixed texts

STRINGS & COMBINATION OF STRING	Euclidean	Manhattan	Scaled Manhattan	Outlier Count	Mahalanobis	KMeans	AutoAssocNNet	SVM
Kolkata123	0.217	0.207	0.202	**0.231**	0.228	0.229	0.212	0.211
Yahoo.com	0.159	0.163	0.152	**0.170**	0.285	0.173	0.163	0.163
Gmail.com	0.175	0.186	0.173	**0.195**	0.247	0.179	0.172	0.174
Facebook	0.232	0.232	0.232	**0.263**	0.258	0.236	0.221	0.224
123456	0.204	0.208	0.203	**0.233**	0.215	0.241	0.198	0.207
Kolkata123 + facebook	0.152	0.134	0.087	**0.060**	0.184	0.140	0.181	0.137
Kolkata123+facebook+ gmail.com	0.178	0.134	0.084	**0.042**	0.205	0.150	0.250	0.155
Kolkata123+facebook+ gmail.com+yahoo.com	0.204	0.152	0.097	**0.038**	0.236	0.171	0.291	0.171
Kolkata123+facebook+ gmail.com+yahoo.com+ 123456	0.205	0.144	0.088	**0.024**	0.260	0.184	0.204	0.160

5 Evaluation and Analysis

Here we see that no combination of features and algorithms give bellow 0.08 average equal error rate for all five type of fixed string.

We have tested combining these five strings and we got the following result. Here minimum average equal error rate is 0.024 where all five strings and all features are considered.

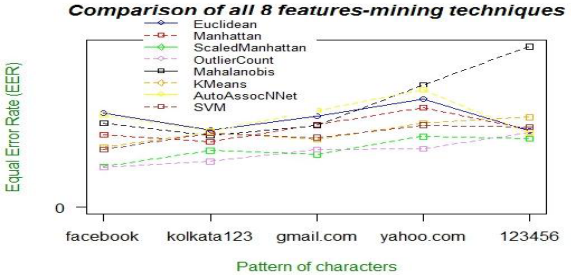

Figure 2. Line chart of all 8 classifiers for each dataset

In the above figure, we see that for all the strings outlier Count (z-score) is achieved best result after scaled Manhattan.

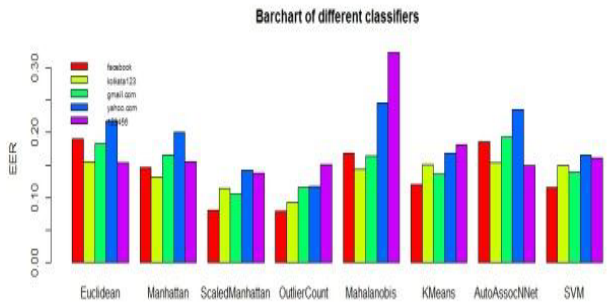

Figure 3. Bar chart of all 8 classifiers for each dataset

Here we see that no combination of features and algorithms give bellow 0.08 average equal error rates for all five type of fixed string.

We have tested combining these five strings and we got the following result. Here minimum average equal error rate is 0.024 where all five strings and all features are considered.

Table 8. Comparison by average equal error rates of different distance-based algorithm

Classifiers	EER	Sd	Classifiers	EER	Sd
Chebyshev	**0.083**	0.289	Ruzicka	0.871	0.109
Canberra	**0.071**	0.104	Soergel	0.129	0.109
Czekanowski	0.129	0.109	Sorensen	0.129	0.109
Gower	0.515	0.264	Wavehedges	0.129	0.109
Intersection	0.579	0.255	Euclidean	0.205	0.123
Kulczynski	0.129	0.109	Manhattan	0.144	0.127
Kulczynskis	0.129	0.109	**ScaledManhattan**	**0.088**	0.097
Lorentzian	**0.044**	0.076	**OutlierCount**	**0.024**	**0.072**
Minkowski	0.219	0.119	Mahalanobis	0.260	0.181
Motyka	0.129	0.109	KMeans	0.184	0.095

The Table 8 represents that Outlier Count, Lorentzian, Canberra, Chebyshev and Scaled Manhattan are the suitable model in identification of user through typing pattern.

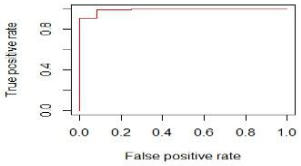

Figure 4. ROC Curve of the Outlier Count

6 Discussion and Further Area of Research

Keyboard is essential for a computer device, which can recognize our typing style and very much unique as per our experiment and cannot be copied or stolen. It can be used as safe guard of our password in any access control system. This technique can be used in online criminal investigation, back door account identification, online typing examination, emotion recognition, lab attendant system many more.

Sometimes, score of different algorithms varies due to mental state of user or aging problems. It can be solve by updating mechanism.

7 Conclusion

Different classification algorithms we have applied on 5 similar keystroke databases taking in our consideration all 8 features and combination of features, so we can compare the classifiers on an equal basis. In our evaluation process, we have identified the best classifier (z-score). It achieved 91.2% of accuracy for the string “kolkata123” (considering KD, DU, UD, Trigap and 4gap timing features), 90.5% of accuracy for the string “yahoo.com” (considering KD, UD), 91.7% of accuracy for the string “gmail.com” (considering KD, UD), 92.0% of accuracy for the string “facebook” (considering KD, DD, UU, DU, UD and Trigap), 85.5% of accuracy for the string “123456” (considering KD, DU, Trigap and 4gap timing features). Z-score classification algorithm gives the highest accuracy for all the string patterns. We also have tested this algorithm on the entire

strings database and we got 97.6 % of accuracy. If we remove some of the feature sets like four graph times for the string "kolkata123" (e.g. four graph times of "ata1", "ta12", "a123", three graph times between "lka" etc), then we can get up to 98.2% of accuracy in this technique. So it has been established that this technique can be used as a safe guard of password or PIN in knowledge-based user authentication. But in practical there are many affecting factors may affect way of this process. Need much more experiment on it like key-pressure, finger placement on keyboard, finger tips size etc. can be calculated.

References

1. S. Bleha, C. Slivinsky, and B. Hussien, "Computer-access security systems using keystroke dynamics,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 12, pp. 1217–1222, 1990.
2. R. Joyce & G. Gupta, Identity authorization based on keystroke latencies. Communication of ACM 33 (2) 168–176, 1990.
3. F. Monrose & A. D. Rubin, Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems, Vol. 16, No. 4, pp. 351–359.
4. K. S. Killourhy, A Scientific Understanding of Keystroke Dynamics. PhD thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, US, 2012
5. E. Yu and S. Cho, "Novelty detection approach for keystroke dynamics identity verification," in Intelligent Data Engineering and Automated Learning, vol. 2690, pp. 1016–1023, Springer, Berlin, Germany, 2003.
6. S. Haider, A. Abbas, and A. K. Zaidi, "A multi-technique approach for user identification through keystroke dynamics," in Proceedings of the 2000 IEEE Interantional Conference on Systems, Man and Cybernetics, vol. 2, pp. 1336–1341, October 2000.
7. R.Giot, M. El-Abed, andC. Rosenberger, "GREYCkeystroke: a benchmark for keystroke dynamics biometric systems," in Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems (BTAS '09), pp. 1–6, September 2009.
8. S. Roy, U. Roy, D.D. Sinha, "Enhanced Knowledge-Based User Authentication Technique Via Keystroke Dynamics", International Journal of Engineering and Science Invention (IJESI), Vol 3, Issue 9, Sep, 2013, 41-48.