On Accuracy of Keystroke Authentications Based on Commonly Used English Words

Alaa Darabseh and Akbar Siami Namin

Department of Computer Science
Texas Tech University
Lubbock, TX, USA
alaa.darabseh@ttu.edu
akbar.namin@ttu.edu

Abstract: The aim of this research is to advance the user active authentication using keystroke dynamics. Through this research, we assess the performance and influence of various keystroke features on keystroke dynamics authentication systems. In particular, we investigate the performance of keystroke features on a subset of most frequently used English words. The performance of four features such as i) key duration, ii) flight time latency, iii) digraph time latency, and iv) word total time duration are analyzed. Experiments are performed to measure the performance of each feature individually as well as the results from the different subsets of these features. Four machine learning techniques are employed for assessing keystroke authentications. The selected classification methods are two-class support vector machine (TC) SVM, one-class support vector machine (OC) SVM, *k*-nearest neighbor classifier (K-NN), and Naive Bayes classifier (NB). The logged experimental data are captured for 28 users. The experimental results show that key duration time offers the best performance result among all four keystroke features, followed by word total time. Furthermore, our results show that TC SVM and KNN perform the best among the four classifiers.

1 Introduction

Biometric authentication technique concerns the use of human characteristics that make each individual unique. It involves any personal characteristics that can be used to uniquely verify a person's identity [MR00]. Biometrics are mainly classified as physiological biometrics features like fingerprint, face, iris, or behavioral biometrics features such as gait, handwritten signature, keystroke dynamics, etc.

Keystroke dynamics are defined as "*a behavioral biometric characteristic which involves analyzing a computer users' habitual typing pattern when interacting with a computer keyboard*"[MR00]. There are several benefits of using keystroke dynamics: First, keystroke dynamics are practical and feasible, since every computer user types on a keyboard; Second, it is inexpensive due to the fact that it does not require any additional or special tools nor components; Thirdly, typing rhythms can be still available even after the authentication phase has passed [GP05].

In building a keystroke-based authentication system, a number of features or measures pertinent to individuals typing styles are used. These features are one of the most important factors that may influence the performance and error rates of keystroke-dynamic detectors [KM10]. Hence, the proper selection of features plays an important role in enhancing the performance of such authentication system when adapting keystroke dynamic detectors.

This paper focuses on studying the influence of various keystroke features on the keystroke dynamics authentication system performance. The major contribution of this paper is the utilization of most frequently used English words in deciding about authenticating users when typing. Our keystroke authentication scheme captures necessary features such as latencies and duration times to determine which timing feature performs better in keystroke dynamics. The promising results demonstrate the performance accuracy of the proposed authentication approach.

The remainder of the paper is organized as follows: Section 2 highlights the motivation and the contributions of this paper. Section 3 describes our experiment procedure, having components for data capture, feature extraction, and classification. Section 4 describes the experiments and presents the experimental results. Section 5 presents the conclusions and future work.

2 Motivation and Contributions

When a person types on a keyboard there are two main timing events that occur: 1) the key down event when the person presses a key, and 2) the key up event when a person releases a key. Timestamps of each event are recorded to keep track of pressing and releasing a key. A variety of timing features can then be extracted from this timing information. Two of the most used features are 1) *duration of the key*, which is the time in which the key is being held down, and 2) *keystroke latency*, which is the time between two successive keystrokes.

Many different methods can be used to calculate latency. The most commonly used methods are: *press-to-press* (PP) latency, which is the time interval between consecutive key presses, PP is also called *digraph time*, *release-to-press* (RP) latency, which is the time interval between releasing the key and pressing the next one, RP is also called *flight time*, and *release-to-release* (RR) latency which is the time interval between releases of two consecutive keys.

It is also possible to capture other keystroke dynamic information such as the time it takes to write a word, two letters (digraph) and three letters (tri-graph).

Keystroke features are one of the most important factors that may influence the error rates of keystroke-dynamic detectors [KM10]. The process of feature selection plays a critical role in improving the performance when designing keystroke dynamic detectors. Revett et al [RGG$^+$07] states that the classification accuracy is substantially influenced by the feature selection process and to a lesser extent on the authentication algorithm employed. A recent survey of keystroke dynamics perceives that certain features have a tendency to be more helpful than others [BW12].

Existing works in the literature of keystroke dynamics demonstrate conflicting results regarding which feature is the most effective timing feature in keystroke dynamics domain. The existing works indicate that duration times are more important than latencies times in reducing false positive error rates [TYT12, RLCM98]. It is also observed that using trigraph time offers better classification results than using digraphs or higher order n-graphs [BGP02]. Revett et al [RGG$^+$07] also reported that the digraph and tri-graph times were more effective compared to duration time and flight time.

Another observation we can conclude from the existing literature of keystroke dynamics is the lack of adequate studies on less common features such as speed and frequency of typing errors which occur during typing. These less common features may embrace effective timing information that might improve the performance of keystroke dynamics systems.

This paper focuses on studying the influence of various keystroke features on the keystroke dynamics authentication system performance. In particular, we investigate the performance of keystroke features on a subset of most 20 frequently used English alphabet letters , most 20 frequent appearing pairs of English alphabet letters , and most 20 frequent appearing of English words. Four features including key duration, flight time latency, digraph time latency, and word total time duration were analyzed using four machine learning techniques namely two different classes of support vector machines (SVM), *k*-nearest, and Naive Bayes. This work poses the following research questions and addresses them throughout the paper:

1. *What feature items contribute more to the accuracy of the feature*? The feature items are defined as the exact instances of letters in each feature, e.g. “a”, “Th”, etc.
2. *What timing feature (e.g. flight time) performs better in keystroke dynamics*?
3. *How the accuracy of the features is impacted from one prediction technique (e.g. SVM) to another*?
4. *How comparable is the total time to flight, digraph, and duration time when typing an entire word*?

3 Experimental Setup

This section describes the experimental procedure, data collection, and the keystroke features used in this experimental study.

3.1 Data Collection

A VB.NET form application was developed to capture raw keystroke data samples. For each keystroke, the time (in milliseconds) when a key was pressed and the time when a key was released, were captured. The participants had the choice of using all keys on the keyboard including special characters such as `Shift Key` and `Caps Lock`, as well as combination of keys. More importantly, they were also able to use the `Backspace` key to correct or modify their typing.

Our experiment of keystroke timing dynamic involved 28 participants. The participants were graduate students majoring in Computer Science. All participants were asked to type the same prepared English passage (5000 characters) one time. Participants had the choice to either use our laptop computer or download the application for collecting data on their own devices. The layout of the application was split into two sections. The top section displayed text that was required to be typed, i.e., fixed text. The bottom section had a space to allow the participant type the text seeing on the top. The texts used in our experiment were a collection of English sentences and passages drawn from a randomly selected pool of English passages.

3.2 Extracted Keystroke Features

The collected raw data from both data sets were used to extract the following features from the experimental data:

1. *Duration* (F1) of the key presses for the most 20 frequent appearing English alphabet letters (e, a, r, i, o, t, n, s, h, d, l, c, u, m, w, f, g, y, p, b) [Gai89].
2. *Flight Time Latency* (F2) for the most 20 frequent appearing pairs of English alphabet letters (in, th, ti, on, an, he, at, er, re, nd, ha, en, to, it, ou, ea, hi, is, or, te) [Gai89].
3. *Digraph Time Latency* (F3) for the most 20 frequent appearing pairs of English alphabet letters (in, th, ti, on, an, he, at, er, re, nd, ha, en, to, it, ou, ea, hi, is, or, te) [Gai89].
4. *Word Total Duration* (F4) for the most 20 frequent appearing English words (for, and, the, is, it, you, have, of, be, to, that, he, she, this, they, will, i, all, a, him) [FKF00].

Every feature (F1, F2, F3, and F4) consisted of 20 items data where feature items are the possible instance values that the underlying feature may have. For instance, F1 contained 20 English alphabet letters where each letter represented an item. Figure 1 illustrates an example of four keystroke features extracted for the word “THE.”

3.3 Classification Algorithms

Four machine learning techniques are adapted to classify the data and address the posed research questions. The selected classification methods are two-class support vector ma-

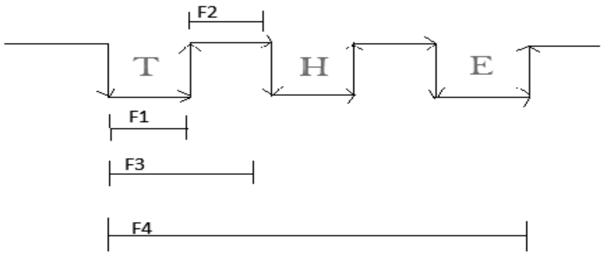

Figure 1: An example of four keystroke features extracted for the word “THE.”

chine (TC) SVM, one-class support vector machine (OC) SVM, the *k*-nearest neighbors classifier (K-NN), and Naive Bayes (NB). Due to the space limitations, we do not provide any discussion about mathematical and fundamental background of these algorithms.

4 Results

4.1 Experiments

Authentication system performance was evaluated by assigning each one of the 28 users in the training data set the role of a genuine user. Then, each of the remaining users played the role of the impostor. Classifiers were trained and tested to measure the accuracy of each feature's item individually in terms of their ability to classify users. On each repetition, we kept track the number of positive and negative testing cases were correctly classified by the classifier between two selected users.

The accuracy of each feature item was then measured as the percentage of positive and negative testing cases correctly classified by the classifier of all users. In other words, accuracy of the features was measured as the percentage of patterns that were correctly mapped to their true users, and was evaluated by the following formula:

$$Accuracy = \frac{T}{N} \tag{1}$$

where T is the number of cases correctly classified and N is the total number of sample cases.

For each user, the first 25 typed timing repetitions of all feature's items were selected to build a user's profile. The first 20 repetitions of each feature's item (of the 25 samples) for each user were chosen for the training phase, and the remaining 5 repetitions of each feature's item were used for the testing phase. The results of this experiment had two major benefits: Firstly, it measured the accuracy of each feature's item individually, so we could determine which items contributed more for the precision of the features. Secondly, we were able to compare the features' performances by measuring the accuracy of each feature individually.

4.2 Results

This section reports the accuracy measures of each feature's item individually. As stated earlier, the purpose is to determine which items are more salient for the accuracy of each feature. Tables 1 - 4 illustrate the accuracy values for all feature's items in each experiment based on using TC SCM, OC SVM, KNN, and NB, respectively. We can observe that some items are performing better than the others. For instance, the accuracy value measured using TC SVM for letter "B" was 90% while the accuracy value measured for letter "R" was 73%. The high accuracy values for the feature's items suggest that these items might be good indicators for distinguishing users.

Moreover, as we can observe from Tables 1 - 4, some feature's items appear more than once in the topmost of each classifier results. For instance: for F1 feature, B, H, T, I, Y items appear in the 10 topmost-appearing items of F1 feature item. For F2 feature, IN, OU, ED, EA, TH items appear in the 10 topmost-appearing items of F2 feature items. For F3 feature, IN, OU, ED, HI, EA items appear in the 10 topmost-appearing items of F3 feature items. For F4 feature, WILL, ALL, A, AND, HAVE items appear in the 10 topmost-appearing items of F4 feature items. Similarly, These feature items that appear in the topmost suggest that these items are good indicators for distinguishing users.

Another interesting finding we can point out, by observing Tables 1 - 4, is that F4 items are performing better than F2 and F3 items, which shows that the total time of typing the most frequent used English words might contain valuable timing information that could be utilized in order to differentiate users.

Apart from the accuracy values of feature's items, Tables 1 - 4 report the overall accuracy of each feature individually by calculating the average of its items accuracy values. We use these averages to compare between the four different types of keystroke features. By observing Tables 1 - 4, we can note that by using F1, we are able to obtain a better result compared to the other keystroke features (F2, F3, and F4). Three classifiers (OC SVM, TC SCM, KNN) out of the four used classifiers agree that F1 is performing better than other features followed by F4. For instance, the average accuracy value measured using TC SVM for F1 was 84% followed by F4 with the accuracy value was 81%. Thus, we can conclude that duration times are more effective timing feature than latencies times. This result is consistent with existing works [RLCM98, TYT12].

Finally, by observing Tables 1 - 4 we can note that TC SVM and KNN perform the best among the four classifiers. Both classifiers achieved good recognition accuracy. The other two classifiers, OC SVM and NB, did not perform well.

5 CONCLUSION AND FUTURE WORKS

This paper reports the results of a series of experiments to study the influence of four keystroke features when using four major classifiers algorithms. The experimental results show some feature's items are contributing more to the accuracy of active authentication. Also, our experimental results show that duration time (F1) feature offers the best per-

Table 1: Items performances using TC SVM (ACC: Accuracy).

F1		F2		F3		F4	
Item	**Acc**	**Item**	**Acc**	**Item**	**Acc**	**Item**	**Acc**
B	0.90	IN	0.85	IN	0.86	WILL	0.87
H	0.88	OU	0.80	HI	0.82	ALL	0.86
P	0.88	ED	0.79	ED	0.81	A	0.85
T	0.88	HI	0.79	OU	0.81	AND	0.84
Y	0.88	EA	0.78	EA	0.80	HAVE	0.84
I	0.86	TH	0.77	HE	0.80	THE	0.84
N	0.86	ER	0.76	TH	0.80	TO	0.84
U	0.86	RE	0.73	TO	0.80	IS	0.83
W	0.86	TO	0.73	EN	0.79	OF	0.83
G	0.85	AT	0.70	IS	0.79	BE	0.83
L	0.85	HE	0.70	ER	0.78	FOR	0.81
M	0.84	AN	0.68	HA	0.78	IT	0.81
O	0.82	EN	0.68	OR	0.76	SHE	0.81
S	0.82	IS	0.67	RE	0.76	THIS	0.80
C	0.81	IT	0.67	AN	0.75	HE	0.79
D	0.81	ND	0.67	IT	0.75	I	0.79
F	0.80	HA	0.66	AT	0.74	YOU	0.78
A	0.79	OR	0.66	ND	0.72	THEY	0.77
E	0.76	ON	0.64	ON	0.70	THAT	0.74
R	0.73	TE	0.62	TE	0.70	HIM	0.73
AVG	0.84	AVG	0.72	AVG	0.78	AVG	0.81

Table 2: Items performances using OC SVM (ACC: Accuracy).

F1		F2		F3		F4	
Item	**Acc**	**Item**	**Acc**	**Item**	**Acc**	**Item**	**Acc**
Y	0.80	OU	0.73	ED	0.72	TO	0.79
L	0.79	EA	0.71	IN	0.71	HAVE	0.77
M	0.79	IN	0.70	EA	0.69	OF	0.77
T	0.79	HI	0.68	OU	0.69	THE	0.76
G	0.78	TH	0.68	HA	0.65	A	0.76
H	0.77	RE	0.67	OR	0.65	ALL	0.75
I	0.77	ER	0.65	TH	0.65	AND	0.74
N	0.77	ED	0.63	TO	0.65	YOU	0.73
O	0.77	AN	0.60	HE	0.64	WILL	0.73
U	0.76	HA	0.60	HI	0.64	FOR	0.72
W	0.75	ND	0.60	IS	0.64	IS	0.72
S	0.74	TO	0.60	IT	0.64	IT	0.72
B	0.73	IS	0.59	ND	0.64	BE	0.72
C	0.73	IT	0.59	TE	0.62	HE	0.72
F	0.73	AT	0.58	EN	0.61	THIS	0.72
D	0.72	EN	0.57	AN	0.60	I	0.72
A	0.71	TE	0.57	RE	0.59	SHE	0.71
R	0.71	HE	0.54	AT	0.58	THEY	0.69
P	0.70	ON	0.53	ER	0.58	HIM	0.69
E	0.69	OR	0.53	ON	0.54	THAT	0.64
AVG	0.75	AVG	0.62	AVG	0.64	AVG	0.73

Table 3: Items performances using KNN (ACC: Accuracy).

F1		F2		F3		F4	
Item	**Acc**	**Item**	**Acc**	**Item**	**Acc**	**Item**	**Acc**
B	0.86	IN	0.86	IN	0.86	A	0.91
M	0.86	OU	0.85	IS	0.84	WILL	0.90
H	0.85	EA	0.81	TH	0.83	ALL	0.88
L	0.85	ED	0.81	ED	0.82	BE	0.87
O	0.85	TH	0.81	HE	0.82	TO	0.87
A	0.84	HI	0.80	HI	0.82	I	0.87
I	0.84	RE	0.80	OU	0.82	AND	0.86
S	0.84	ER	0.79	EN	0.81	OF	0.86
T	0.84	TO	0.74	TO	0.81	THE	0.86
Y	0.84	AT	0.73	EA	0.80	HAVE	0.84
C	0.83	AN	0.72	AN	0.78	IS	0.83
D	0.83	ND	0.71	ER	0.78	IT	0.83
F	0.83	EN	0.70	RE	0.78	FOR	0.82
G	0.83	HE	0.70	AT	0.77	YOU	0.82
N	0.83	IT	0.70	HA	0.77	HE	0.82
P	0.83	HA	0.69	IT	0.77	SHE	0.82
R	0.83	IS	0.69	OR	0.76	THIS	0.82
E	0.82	OR	0.69	ND	0.74	THEY	0.82
U	0.82	ON	0.65	TE	0.71	THAT	0.74
W	0.82	TE	0.65	ON	0.70	HIM	0.74
AVG	0.84	AVG	0.75	AVG	0.79	AVG	0.84

Table 4: Items performances using NB (ACC: Accuracy).

F1		F2		F3		F4	
Item	**Acc**	**Item**	**Acc**	**Item**	**Acc**	**Item**	**Acc**
F	0.71	IN	0.82	IN	0.83	BE	0.79
B	0.70	TH	0.74	ED	0.78	TO	0.79
C	0.70	ED	0.73	TH	0.78	WILL	0.79
P	0.70	ER	0.73	HE	0.77	ALL	0.79
H	0.69	HI	0.73	HI	0.77	HAVE	0.78
O	0.69	EA	0.70	TO	0.76	THE	0.78
T	0.69	HE	0.70	EA	0.75	OF	0.77
I	0.68	OU	0.70	OU	0.75	A	0.77
G	0.67	RE	0.68	HA	0.74	AND	0.76
R	0.67	TO	0.67	EN	0.73	IS	0.76
S	0.67	HA	0.65	ER	0.73	THIS	0.75
U	0.66	IS	0.65	IS	0.73	HIM	0.74
L	0.65	AN	0.63	AN	0.71	FOR	0.73
W	0.65	IT	0.63	RE	0.71	IT	0.73
Y	0.65	OR	0.63	IT	0.71	YOU	0.73
D	0.64	AT	0.62	OR	0.69	SHE	0.73
M	0.64	EN	0.61	RE	0.69	I	0.71
A	0.63	ND	0.61	ND	0.67	THAT	0.70
E	0.63	ON	0.59	ON	0.66	HE	0.70
N	0.61	TE	0.59	TE	0.65	THEY	0.66
AVG	0.66	AVG	0.67	AVG	0.73	AVG	0.74

formance result among all four keystroke features, followed by word duration time (F4). Lastly, the paper introduced new features' items that could be effectively used in keystroke dynamics domain. Future work will involve more classifiers and trying to train the classifiers using less samples sizes and test the impact of that on the performance accuracy.

References

[BGP02] Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. User Authentication Through Keystroke Dynamics. *ACM Trans. Inf. Syst. Secur.*, 5(4):367–397, November 2002.

[BW12] Salil Banerjee and Damon Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139, 2012.

[FKF00] E.B. Fry, J.E. Kress, and D.L. Fountoukidis. *The Reading Teachers Book of Lists*. Jossey-Bass, 3rd edition, 2000.

[Gai89] H.F. Gaines. *Cryptanalysis: A Study of Ciphers and Their Solution*. Dover Publications, Dover, New York, 1989.

[GP05] D. Gunetti and C. Picardi. Keystroke analysis of free text. *ACM Trans. Inf. Syst. Secur.*, 8(3):312–347, 2005.

[KM10] Kevin Killourhy and Roy Maxion. Why Did My Detector Do That?!: Predicting Keystroke-dynamics Error Rates. In *Proceedings of the 13th International Conference on Recent Advances in Intrusion Detection (RAID'10)*, pages 256–276. Springer-Verlag, 2010.

[MR00] Fabian Monrose and Aviel D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation Comp. Syst.*, 16(4):351–359, 2000.

[RGG$^+$07] K. Revett, F. Gorunescu, M. Gorunescu, M. Ene, S. Magahaes, and H. Santos. A machine learning approach to keystroke dynamics based user authentication. *TInternational Journal of Electronic Security and Digital Forensics*, 1(1):55–70, 2007.

[RLCM98] J. A. Robinson, V. M. Liang, J. A. M. Chambers, and C. L. MacKenzie. Computer User Verification Using Login String Keystroke Dynamics. *IEEE Transactions on Systems, Man,and Cybernetics - Part A: Systems and Humans*, 28(2):236–241, March 1998.

[TYT12] Pin S. Tech, Shigang Yue, and Andrew B.J. Teoh. Feature Fusion Approach on Keystroke Dynamics Efficiency Enhancement. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 1(1), 2012.