The Analyses of Key Stroke Timings using Self Organizing Maps and its Application to Authentication

Hiroshi Dozono+ Hisao Tokushima+ Masanori Nakakuni* Yoshio Noguchi+

hiro@dna.ec.saga-u.ac.jp tokusima@dna.ec.saga-u.ac.jp nakakuni@cc.miyazaki-u.ac.jp nogu@dna.ec.saga-u.ac.jp

+ Faculty of Science and Engineering, Saga University

1-Honjyo Saga 840-8502 JAPAN

* Information Processing Center, University of Miyazaki,

1-1 Gakuen Kibanadai Miyazaki 889-2192 JAPAN

Abstract

Recently, security of the computer systems becomes an important problem. Almost all computers use the password mechanism for the user authentication. But password mechanism has many issues. In this paper, we propose a kind of biometrics authentication method using the timings of key strokes of an identical simple phrase for all users. For this method, selection of the phrase is important. We analyzed the timings of key strokes using Self Organizing Map (SOM) and selected the suitable phrase for authentication. We examined the effectiveness of this method with the authentication experiments using the map organized by SOM.

Keywords: Biometrics, Authentication method, Neural network, Self Organizing Map

----

1. Introduction

Recently, almost all computers are connected to the networks and security of the computer systems becomes an important problem. Almost all computers use the password mechanism for the user authentication. But password mechanism has many issues. At first the password is the static plain text so it is easy to get the password from shoulder hacking, guessing from personal information or memo on which the list of the password are written not to forget the password. Secondly, some persons are using several systems recently. For the different systems, different passwords should be set, but almost users set the identical password for different systems because they can not remember many passwords. So, if a password is hacked, all system can be accessible. Thirdly, someone may feel that it is troublesome to enter password during login to the computers, especially for portable computers such that notebook computers and Personal Digital Assistant(PDA)s, because complex password phrase which includes symbols and combination of capital letters is recommended as strong password.

As the solution to the issue of password mechanism, biometric authentications are often used. The biometric authentication uses the biological characteristics, such that fingerprints, iris patters and blue pipe patterns, or the behavior characteristics, such that hand-writting patterns. The fingerprint authentication becomes the most popular recently and some computers are equipped with fingerprint readers and the readers which are connected to Universal Serial Bus (USB) are already sold. But, all of the biometric authentications using biological characteristics need additional hardware and they will cost up the price of computers. Recently, for the main stream or low end computers, the costs are the most important issue, so the makers do not want to add any additional hardware. And some persons feel bad to register their fingerprints in host systems of company. Furthermore, the fingerprint authentication can be hacked or after login, the irregular user may use the computer while the regular user is apart from the computer.

For these problems, we propose the authentication method using the integrated information of the biometrics of behavior characteristics taken from the instruments equipped with the standard computers such that keyboard, mouse and touch panel. And we propose the authentication method not only for login but also for the operating time of the computers. In this paper, we concentrate on the authentication using behavior characteristics of key stroke timings and analyze them using Self Organizing Map(SOM)[1]s. It is well known that key stroke timings are usable for user authentications. In this paper, we analyzed them using SOM and search for the phrase which is suitable to authenticate all users using a single simple phrase in login authentication and analyze the behavior characteristic which are common to Japanese persons for the use of login authentications and operating time authentications. SOM is a kind of neural network which was proposed by Kohonen. SOM can extract the feature of multi dimensional input vectors with unsupervised learning and can visualize the relations among the input vectors on 2 dimensional plane. We reported the application of SOM for the analysis of pen pressure patterns at BIOAU05[2]. In [3], an authentication method using key stroke timings and key stroke pressures measured by pressure sensors is reported, but only some specific phrases which are usually used as the password phrases are examined. In this paper, we analyzed the sequences of key stroke timings using SOM and analyzed the relations among the users and examined many phrases with changing the category of the phrases. From these analyses, we selected the suitable phrases for authentication and made the experiments of authentication using the map organized by SOM.

2. Sampling of the key stroke timings

The methods to get the timing of the key stroke are different depending on the operating systems. The standard I/O functions can not be used for such applications, so the OS-dependent system call functions are used. For the following experiments, we used Windows XP and Visual C++ as the operating system and programming language respectively. The time is measured by reading the system clock directly as to get the finest resolutions of time in Windows XP. Figure.1 shows a sample of key stroke timings

Push A	Interval	Push B	Interval	Push C	phrase
|------------|		|-------|		|------------|	ABC
100ms	90ms	60ms	110ms	100ms	timings

Figure.1 Sample of the key stroke timings

For each key, the time pushing the key and the interval time between keys are sampled.
The shift key is also sampled independently, so inputted phrases can be determined from the sequence of key strokes.

3. Self Organizing Map(SOM)

SOM is an architecture of neural networks, which is classified as the network of feed forward type and of the unsupervised learning method. SOM can organize the feature of the input vectors on the 2-dimensional map on which the output neurons are arranged. After learning, the input vectors are mapped on the organized map, then the relations of the input vectors can be visualized on the map. Original SOM algorithm trains the map incrementally by updating the map for each presentation of input vector. The recent trend of SOM algorithm adopts Principal Component Analysis(PCA) and batch update to improve the performance. The PCA is used to reduce the dimension of input vectors and to initialize the map. The input vectors are analyzed by PCA before they are given to SOM and the vector of principal components with large contribution rates are used as the input vectors to SOM. The initial map is made as the plane of 1st and 2nd principal components. The batch update improves the effectiveness of training and the ambiguity of the resulting maps according to the order of the presentations of input vectors. In the learning phase, the inputs vectors are associated to the closest units on the map, and in the update phase, the units on the map are updated at once using the information of the associated vectors to the unit and those of the neighboring units. For this research, we used the SOM with batch update and PCA only for initialization of the map, because the dimension of input vector is not so large. The algorithm is shown as follows.

Batch SOM algorithm with PCA initialization

Step-1 PCA analysis
Calculate the 1st and 2nd Principal Components (PC) of input vectors.

Step-2 Initialization of the map
Initialize the map using the 1st and 2nd principal components as base vectors of 2-dimensional map.
Initialize the size of neighbors Sn=IS
and L=0, where IS is initial size of the neighbor.

Step-3 Batch learning phase

For each input vector $\mathbf{y}_k$, search for the unit U_{ij} which is associated to the closest vector $\mathbf{z}_{ij}$ and add $\mathbf{y}_k$ to the learning buffer of U_{ij}.

Step-4 Batch update phase

For each unit U_{ij}, update the associated vector $\mathbf{z}_{ij}$ using the weighted average of the vectors recorded in the buffers of U_{ij} and its neighboring units as follows.

(1) For all vectors $\mathbf{v}$ recorded in the buffer of U_{ij} and its neighboring units in distance d<=Sn, calculate weighted sum of the vectors $\mathbf{S}$ and the sum of weight values W.

$$\mathbf{S} = \mathbf{S} + fn(d) \times \mathbf{v}$$

$$W = W + fn(d)$$

where fn(d) is the neighborhood function which become 1 for d=0 and decrease with increase of d.

(2) Set the vector $\mathbf{z}_{ij}=\mathbf{S}/W$,

and set L=L+1.

Repeat Step-3 and Step-4 with decreasing the size of neighbors Sn until L<=M, where M is the number of learning and update loops.

4. The analysis of key stroke timings using SOM

In this section, we show the results of the analyses of key stroke timing using SOM. At first, we made experiments using the phrases based on the English words and mixture of capital characters, lowercase characters and symbols, in practice "sdfghjkl", "SdFgHjKl", "datafile", "DataFile", "sagauniv", "SagaUniv" and "Ht%n&MkA". "sdfghjkl" is the phrase of the keys in home position, datafile is simple English word, "sagauniv" is name of university and "Ht%n&MkA" is random phrase with symbols which are recommended for the password. The examinees are 11 Japanese university students and their typing skills are vary from person to person and they inputted the phrases sequentially in 6 times using same keyboard. Figure 2 shows the map sized 16x16 organized with all of the key stroke timings. Each small graph denotes the key stroke timing and the graphs are organized according to the similarities among them. In this paper, the cyclic maps are used, so upper side of the map is connected to lower side of the map and left side of the map is connected to right side of the map.

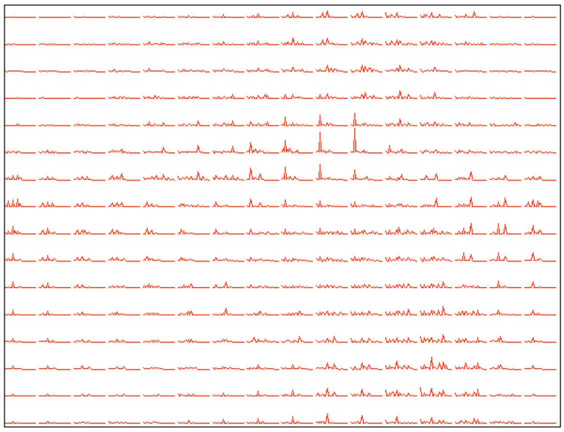

Figure 2 Magnified map organized with all key stroke timings

Figure 3 shows the map labeled with the user number. For each number, the clusters which correspond to the phrases are formed.

1	5	1	2	9	5	4	1	7	4	4	8	3	3	2	9
1	5	1	10	10	7	9	10	4	2	4	8	6	6	2	6
6	8	8	10	1	1	7	9	4	4	4	8	6	8	3	6
9	8	8	11	1	7	7	7	10	10	10	5	4	5	5	6
6	6	11	11	7	11	7	9	4	1	11	10	4	10	1	1
5	10	11	11	8	8	4	5	1	1	11	11	4	2	5	1
1	1	11	5	8	11	4	2	5	3	11	2	11	2	5	4
8	8	5	6	8	4	4	3	3	9	9	2	7	1	2	10
2	7	5	5	5	7	8	8	4	9	11	1	3	7	7	2
2	7	10	10	7	8	7	7	4	9	9	1	3	3	4	4
4	7	9	1	8	8	8	7	4	1	9	1	4	9	2	2
5	4	9	10	7	10	7	7	8	4	5	4	4	4	2	5
9	11	11	10	11	7	7	7	11	9	1	8	6	6	9	9
2	2	2	1	5	11	7	11	11	11	8	8	10	6	3	6
6	6	10	10	11	11	7	2	4	11	3	3	6	6	6	3
6	6	1	1	2	5	7	7	7	4	4	8	3	3	3	3

Figure 3 The map sized 16x16 labeled with user numbers

To examine more precisely, the map for each phrase was organized independently. The map size was 16x16. Figure 4 shows the map of phrases "sdfghjk", "datafile" and "Ht%n&MkA" respectively.

3	3	9	1	1	8	1	8	8	4	4	4	9	6	6	6
3	2	7	1	8	8	8	10	10	4	4	4	10	10	11	11
11	7	7	7	1	8	10	10	10	4	4	4	10	10	6	3
5	8	7	1	1	1	7	10	8	8	2	2	4	10	6	3
5	8	4	4	7	7	7	8	8	8	2	2	2	6	6	5
5	4	4	7	7	7	7	8	8	8	2	2	2	6	6	6
4	4	4	7	7	7	11	11	11	10	8	2	10	6	6	6
4	4	4	7	7	7	11	11	11	10	10	10	10	9	9	9
9	9	9	7	7	7	11	11	11	10	10	10	10	9	9	9
9	9	9	9	7	11	11	11	7	10	10	10	4	4	9	9
9	9	9	9	11	11	11	11	10	10	10	4	4	4	5	5
2	9	9	5	5	11	11	11	10	10	10	4	4	5	5	5
2	2	5	5	5	9	9	11	10	10	10	1	4	5	1	2
2	2	5	5	9	9	9	11	8	10	1	1	1	11	11	11
5	5	2	1	1	9	7	8	8	1	1	1	1	9	11	11
11	3	9	1	1	1	7	8	8	8	1	1	9	9	6	6

Map of the phrase “sdfghjkl”

4	4	6	6	6	6	6	5	1	2	2	9	9	7	10	10
1	8	9	6	1	6	6	5	2	2	9	9	9	9	4	1
1	8	9	9	1	1	1	4	10	10	10	2	9	9	4	1
3	8	3	9	6	1	4	4	4	10	2	2	2	9	4	1
3	3	3	3	6	6	4	4	4	1	10	2	7	7	7	3
3	3	3	8	6	11	11	4	1	1	1	7	7	7	7	1
3	3	8	8	9	11	11	11	11	1	1	7	7	7	7	3
8	8	8	8	11	11	11	11	11	11	11	11	11	7	2	2
8	8	8	10	10	11	11	11	11	11	11	11	11	7	2	2
2	2	10	10	10	10	1	1	11	11	2	11	11	2	4	4
4	4	10	10	10	6	1	1	1	2	2	2	11	10	10	10
4	9	9	10	6	6	5	1	1	2	2	11	7	11	10	10
9	9	9	5	6	6	6	5	5	2	11	7	7	7	1	9
9	9	10	5	5	6	5	5	5	4	4	11	7	7	1	9
10	4	5	5	5	5	5	5	5	4	4	4	7	7	10	10
4	4	4	5	6	6	5	5	5	2	4	4	7	7	10	10

Map of the phrase “datafile”

6	6	1	1	2	3	6	3	8	8	8	1	1	8	11	11
6	6	5	5	2	6	6	6	6	6	6	1	1	9	11	4
4	6	6	5	9	9	6	5	5	6	6	6	9	9	9	4
4	4	6	6	9	9	7	7	7	6	2	2	2	9	4	4
4	4	1	9	9	9	8	7	7	8	2	2	2	8	8	4
4	4	1	1	2	8	8	8	8	8	8	2	2	2	8	4
4	4	1	5	2	2	8	10	8	8	7	7	2	2	2	4
10	4	5	5	5	10	10	10	10	7	7	7	7	2	2	10
10	4	4	5	10	10	10	10	7	7	7	7	11	11	10	10
4	4	4	7	7	10	11	11	7	7	7	7	11	11	11	10
11	11	7	7	7	7	11	11	11	7	3	3	5	11	11	11
11	11	9	7	7	7	11	11	11	10	3	3	3	5	11	11
11	3	3	5	5	3	9	9	10	10	3	3	3	9	9	11
3	3	3	5	5	3	9	9	10	10	10	3	9	9	9	9
3	3	1	1	5	3	9	3	8	10	10	3	1	9	9	9
6	1	1	1	1	3	3	3	8	8	8	1	1	1	11	11

Map of the phrase “Ht%n&MkA”

Figure 4 The maps sized 16x16 labeled with user numbers of the phrases

For the phrase "sdfghjk", the cluster is not identical for each user even though it can be easily typed from home position of the keyboard. During the experiments, the examinees typed each key one by one with confirming the key, so the timing is vary in each iteration. For the phrase "datafile", better clusters are formed compared with "sdfghjkl" because it based on the English word. For the phrase "Ht%n&MkA", the result became the worst because the symbols are typed one by one and typed slower than English characters. The results for phases which are mixture of capital characters and lowercases are worse than those for the lowercase phrases. In all cases, the results of the clustering are not good enough to apply to authentications.

Next, we made experiments using the phrases Romaji(Roman alphabet). Romaji is the ideographic writing used to input Japanese from English keyboard. The examinees (Japanese university students) do not usually input English words and they are rather used to input Japanese in Romaji. The phrases, "kirakira", "sagadai", "arigato", "denatsu" and "kousatsu", are selected and 10 examinees inputted each word in 8 times.

Figure 5 shows the map of the "denatsu" which is the worst case and the map of the phrase "kirakira" which is the best case.

3	5	9	9	9	8	8	8	8	2	2	5	5	5	6	3
3	5	9	7	7	7	8	8	8	2	2	2	5	2	2	2
2	4	7	7	7	7	7	9	9	10	2	2	5	2	2	2
4	4	4	7	7	7	7	9	9	9	10	10	7	7	2	2
10	4	4	7	7	7	7	9	9	10	10	10	7	7	7	2
10	10	6	6	7	7	7	9	9	10	10	10	7	7	6	10
10	10	6	6	6	7	4	7	7	7	7	10	8	6	6	6
10	10	10	6	6	4	4	4	7	7	7	8	8	6	6	6
5	10	10	6	4	4	4	4	7	7	7	8	8	8	9	9
5	5	8	10	10	4	4	6	8	8	3	3	8	9	9	5
5	8	8	10	10	2	2	8	8	8	3	3	3	1	1	5
8	8	8	10	2	2	2	8	8	8	3	3	1	1	1	1
5	6	8	10	10	2	10	4	4	3	3	1	1	1	5	5
9	3	3	6	6	6	4	4	4	3	3	3	1	1	5	9
9	3	3	6	6	6	4	4	4	3	3	6	6	6	4	9
3	5	9	9	6	6	8	8	8	5	5	5	5	4	4	4

Map of the phrase “denatsu”

8	8	6	6	6	6	3	5	5	10	10	10	10	8	8	8
10	10	3	6	6	6	3	3	7	7	7	8	8	10	10	10
6	3	3	3	6	3	3	3	7	7	7	8	8	8	10	10
1	1	3	4	4	3	7	7	7	7	7	7	8	1	1	1
1	1	4	4	4	3	7	7	7	7	7	7	1	1	1	1
1	4	4	4	4	3	7	7	7	7	7	7	1	1	1	1
4	4	4	4	4	3	3	3	7	7	7	2	2	1	1	1
4	4	3	3	3	3	3	3	7	7	7	2	2	2	9	4
4	4	3	3	3	3	3	3	7	7	7	2	2	9	9	4
4	4	3	3	3	6	6	6	7	7	5	9	9	9	9	9
5	5	5	5	3	6	6	6	6	5	9	9	9	9	9	9
3	5	5	5	5	6	6	6	5	10	9	9	2	9	9	9
2	5	5	5	3	5	5	5	5	10	10	2	2	2	2	2
2	5	5	6	6	6	5	5	5	10	10	10	2	2	2	2
5	5	5	6	6	6	5	5	5	10	10	10	10	2	8	2
8	5	6	6	6	6	5	5	5	10	10	10	8	8	8	8

Map of the phrase “kirakira”

Figure 5 The map sized 16x16 labeled with user number for romaji phrases

The difference among these phrases is that "kirakira" is simple and has one way to type in Romaji and "denatsu has some way to type in, so the examinees felt embarrassed during typing. The phrase "kirakira" is considered to be the best in these experiments and most of romaji phrases showed better results than those of previous experiments as expected. We made the authentication experiments based on the results of these analyses.

5. Authentication experiments using self organizing maps

In this section, the results of authentication experiments are shown. We made these experiments using the same key stroke data mentioned in the previous section and the half of the data were used to organize the map and remained half data were used as the test data for authentication experiments. We used the map labeled with user numbers for these experiments and a test data is classified to the user numbered to the unit which is the closest to it. The map size was set to 16x16.

Table 1, 2, 3 shows the results. In these table, success, FRR and FAR means the rate of successful authentication, false reject rate and false accept rate respectively.

Table.1 The result of authentication experiments using the phrase “datafile”

	Success	FRR	FAR
User1	0.000	1.000	0.000
User2	0.000	1.000	0.120
User3	0.500	0.500	0.120
User4	0.000	1.000	0.000
User5	0.000	1.000	0.040
User6	0.000	1.000	0.040
User7	0.500	0.500	0.080
User8	1.000	0.000	0.000
User9	0.000	1.000	0.160
User10	0.000	1.000	0.080
User11	1.000	0.000	0.000

Table.2 The result of authentication experiments using the phrase “denatsu”

	Success	FRR	FAR
User1	1.000	0.000	0.000
User2	0.250	0.750	0.190
User3	0.000	1.000	0.143
User4	0.250	0.750	0.095
User5	0.500	0.500	0.048
User6	0.000	1.000	0.429
User7	0.500	0.500	0.095
User8	0.500	0.500	0.095
User9	0.250	0.750	0.095
User10	0.000	1.000	0.095

Table.3 The result of authentication experiments using the phrase “kirakira”

	Success	FRR	FAR
User1	0.500	0.500	0.000
User2	1.000	0.000	0.048
User3	0.250	0.750	0.238
User4	1.000	0.000	0.000
User5	0.000	1.000	0.000
User6	0.750	0.250	0.095
User7	0.750	0.250	0.095
User8	0.750	0.250	0.095
User9	1.000	0.000	0.048
User10	0.500	0.500	0.000

Table 1 shows the results for the phrase “datafile” which organized the best map in no-romaji experiments. But, the rate of successful authentication is not good. For 2 examinees, the authntication rates were 1.0, because these users are used to type in English. Table 2 shows the results for the phrase "denatsu" and the results is the almost same as those of "datafile". Table 3 shows the results for the phrase "kirakira". For 8 of 10 examines, the authentication rates were greater than 0.5 and for 6 examines, the rates were greater than 0.75. From these results, simple phrases in Romaji are suitable for authentication of Japanese users, but the effectiveness depends on the skills of typing. Next, we made the experiment of authentication using login name of each user. Table 4 shows the results.

Table4 The result of authentication experiments using the login name of each user

	Success	FRR	FAR
User1	1.000	0.000	0.000
User2	1.000	0.000	0.083
User3	1.000	0.000	0.111
User4	1.000	0.000	0.083
User5	0.500	0.500	0.000
User6	0.500	0.500	0.083
User7	1.000	0.000	0.028
User8	1.000	0.000	0.083
User9	0.250	0.750	0.083
User10	0.500	0.500	0.111

The best results are obtained among all experiments. The login names are typed at every login time, so the authentication with key stroke timings will be effective. The major reasons for the failures come from the method of the experiments and the skill of typing of the examinees. As for the method, the examinees felt troublesome to type in same sequences of the phrases repeatedly, so their typing became loose after some repetition, and during the experiments, the examinees who are not authenticated well were complaining about it. Additionally, the last half of the data were used in the authentication experiments, so successful authentication rates became worse and worse. As for the skill of typing, some users are not using computers in their study frequently. They typed the keys one by one with confirming the keys even in Romaji, so the key stroke timings were vary and the successful authentication rates went down.

6. Conclusion

We proposed the analysis method of key stroke timings using Self Organizing Map(SOM) and made the authentication experiments based on the results of analysis.

In consequence, the authentication using the key stroke timings of typing login name were the best and simple Romaji phrases will be available to authenticate Japanese users who have a certain level of typing skills. The combination of the login name and Ronaji phrase, or login name and hand-written symbols will be more effective for login authentication. And this authentication method will be available for operating time authentication by observing the frequent Romaji phrases or frequent patterns of the sequences of keys. Furthermore, this method of the analysis using SOM will be available not only for this application but also for other biometric data and the data concerning security.

References

[1] T. Kohonen: Self Organizing Maps, Springer, ISBN 3-540-67921-9

[2] Hiroshi Dozono, et.al: The Analysis of Pen Pressures of Handwritten Symbols on PDA Touch Panel using Self Organizing Maps, Proceedings of the International Conference on Security and Management 2005, pp.440-445(2005)

[3] Chen Change LOY, Chee Peng LIM and Weng Kin LAI, Pressure-Based Typing Biometrics User Authentication Using the Fuzzy ARTMAP Neural Network, Proceeding of the 12th International Conference on Neural Information Processing 2005