Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 1817 – 1825

Procedia Computer Science

19th International Conference in Knowledge Based and Intelligent Information and Engineering Systems

Personal ID System by Means of Random Input Passwords

Mieko Tanaka-Yamawaki[a]*, Yuki Tanaka[b], Katsutoshi Yoshii[a]

[a]*Department of Information.& Knowledge Eng., Graduate School of Engineering, Tottori University, 101-4, Koyamacho-Minami, Tottori,680-8662 Japan*

[b]*Being Co.,Ltd., Support Division, 312-1, Sakurabashi, Tsu, Mie, 514-0003, Japan*

Abstract

A new password system is proposed that utilizes unconscious habits in typing the keyboard. Such password, named the Random Input Password (RIP, in short), is a random sequence of a fixed length generated by individual each time. In other words, this systems uses the human-generated random number as a password. This ID system does not require any specific codes to memorize, thus has a potential to flee us from the flood of increasing number of passwords. Although the RIP is different every time, it reflects peculiarities of individuals and distinguish each person to a certain degree. The identifier is constructed by using 6 parameters extracted from the pre-registered data for each individual. The average accuracy evaluated by using 9 individuals was that the Type-I error rate that disapproves the right person is 23% and Type-II error rate that approves wrong persons is 19% on the average. A possible refinement is discussed by reconsidering the effectiveness of the parameters.

Keywords: Random Input Password (RIP), Human Random Generator (HuRG), Self Refusal Error(Type-I), Others Allowable Error(Type-II), Self Organizing Map (SOM)

1. Introduction

Nowadays we are surrounded by rapidly changing financial and social systems, and under such environment we have many occasions to use passwords and very often forget the right one. In fact, there are too many passwords to remember for most of us. Moreover, the number of passwords increases almost every day, when we make reservations for conferences, airlines, hotels, and restaurants. This situation must be simplified. As a possible way to escape from the flood of passwords, we propose a new kind of password, the Random Input Password (RIP).

The RIP is a random string generated by human. The use of human random generator (HuRG) [1,2,3] was originally practiced in the community of neurological doctors as a simple way of detecting the level of schizophrenia, based on the fact that the patients in the advanced stage of schizophrenia have difficulty to generated random numbers. The human random generation also caught the interests of computer scientists for the Turing Test to distinguish human

* Corresponding author. Tel.: +81-0857-31-5223; fax: +81-0857-31-0879.
E-mail address: mieko@eecs.tottori-u.ac.jp.

doi:10.1016/j.procs.2015.08.292

and a computer who impersonates man. It was also used in the study of developmental psychology in the context that the random generating ability reflects the stage of child developments. Our study of HuRG is not the same as any one of them. We have accumulated much data from normal young adults (namely, students in our Laboratory) to show that the random generation by human exhibits peculiarities of the person who generated the random strings. So far, we have investigated the relationship between the level of randomness of data strings and the character, age group, the level of dementia of the person who generate the data.

In this paper, we propose to apply the HuRG for a new type of ID system that is free from memorizing any specific strings for passwords.

2 Random Input Password (RIP)

In order to apply HuRG for the ID system, we set up a specific type of HuRG, that we call the "Random Input Password (RIP)", where the subject simply hits the selected keys on the PC keyboard *L* times by using one finger, within 2 seconds after hitting the previous key, as random as possible.

We choose the four contact keys "T, Y, G, H" which are coded by "0, 1, 2, 3", respectively, as shown in Fig.1. The conditions of RIP is summarized in Table 1. The input length *L* is chosen to be 50, 40, 30 and 20.

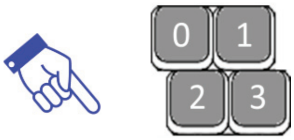


Fig.1 Hitting one of the 4 keys TYGH coded by 0123.

Table 1. The conditions used to data taking of RIP

keys	4 contact keys on the keyboard
input length *L*	30, 40, 50
conditions	as random as possible, and quick next input within 2 seconds only one finger used

There are four important elements to distinguish individuals from the random generations:

[a] sequential patterns of the input keys
[b] time intervals between key strokes
[c] directions of the finger motions
[d] intervals before returning to the same key

[a]Sequential patterns of the input keys

Sequential patterns of the key appearance carries information characteristic to individuals. In order to quantify the sequential pattern, we count the frequency of appearance of 4 keys, represented by 4 symbols, "0, 1, 2, 3" corresponding to "T, Y, G, H", the frequency of appearance of arrays of length 2, such as "00, 01, 02, 03, 10, 11, 12, 13,.., 33", and of length 3, such as "000,.., 333".

For example, the data sequence "32021" gives us the frequency of appearance of 1-digit to be "one" for the symbols "0", "1" and "3", and "two" for the symbol "2", and for 2-digit, "one" for the symbols "32", "20", "02", "21".

In particular, the 2-digit patterns are important to distinguish individuals. The patterns longer than 3 are not suitable, since the probability of appearance in the data sequence L=30-50 becomes too small.

Thus we have 3 indices:

P_1: frequency of appearances of 1-digit patterns 0, 1, 2, 3
P_2: frequency of appearances of 2-digit patterns 00, ..,33
P_3:frequency of appearances of 3-digit patterns 000,..,333

[b]*Time intervals between key strokes*

There are 3 types of intervals in key hitting as shown in Fig. 2. The accuracy of measurement is 0.1 second. The input data pass the identifier if two or more parameters out of three pass the test. Those three parameters are:

T_1: time between the two consecutive press-downs
T_2: time from the press-down to the release-up of a key
T_3: time from the release-up of the previous key to the press-down of the next key ($T_3 = T_1 - T_2$)

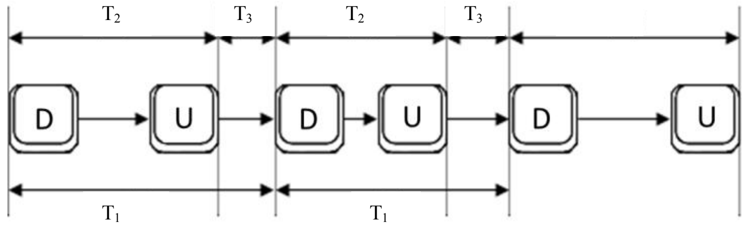


Fig. 2 Three time intervals, T_1, T_2, T_3 are illustrated

[c]Directions of the finger motions (D)

The frequency of appearance of the five directions, {still(・), horizontal(—), vertical(｜), left slant(＼), right slant (／)}, that the input finger moves between two consecutive inputs reflects characteristics of the individuals.

The frequencies of appearance of the 5 directions are registered as the reference vector for each subject. If the test vector coincides with the reference vector, the inner product between the test vector and the reference vector becomes one, and the test data proves to have been generated by the same subject. However, the reference vector involves variances over N=20 trials. Thus we set the evaluation criteria for each subject based on the average and the standard deviation of N=20 trials, in such a way that the 9 subjects separately recognized in this experiment.

[d]Intervals to return to the same key(R)

The intervals before returning to the same key also carries information of individuals. Usually, the distance larger than 10, or zero is hardly observed.

3 Learning the Pattern of Individuals

The pattern of individual subjects is constructed by using the 6 parameters of the 4 elements, P_2, P_3, T_1, T_2, D, and R. The average values of the 6 parameters with errors are used as identifiers.

3.1 Experimental Setup

A subject of the experiment is asked to input the ID and RIP, consecutively by N times. If all the 4 elements, P, T, D, R, are correct, then it passes the authentication. After collecting N RIPs from all the subjects, a set of identifiers are constructed. The process of authentication is one-to-one, as illustrated in Fig.3. When User B enters his/her ID and a RIP, the identifier of B is called from the system and used to check the similarity with the input.

Nine students in the age of 20's served as the subjects of this experiment. The job flow is shown in Fig.4. The number of input, N, is chosen to be 20 and 30. The display of the registration is shown in Fig. 5, and the display of the identification is shown in Fig. 6.

We first perform a closed test, using the same N=30 data set to test the performance of the system. Then, by using new data taken a few days to a few weeks after the days when the first set of data are taken, we perform open tests.

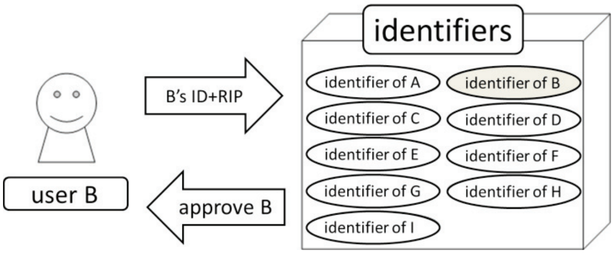


Fig. 3 One-to-one authentication method

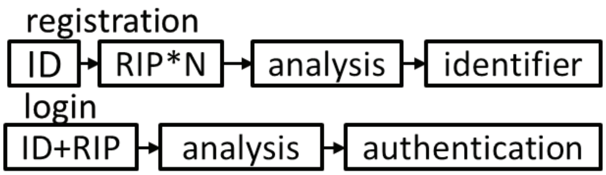


Fig. 4 System flowchart

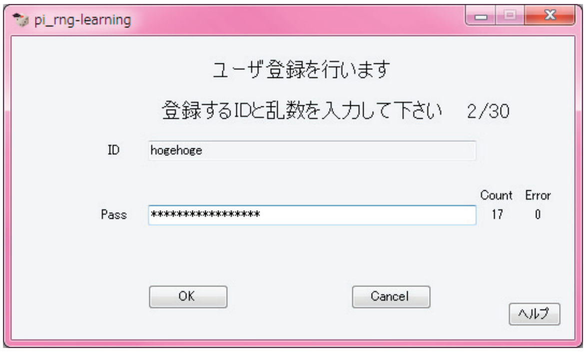


Fig. 5 Registration window is shown at the 2nd RIP of N=30

Fig. 6 Execution window is shown at the moment of login

3.2 Evaluation by the identifiers

The identifiers are constructed for each subject who offered data set of *N* strings of equal length *L*. Essentially they are eight dimensional real-valued vectors of the parameters,

$$V = (P, T, D, R) \tag{1}$$

Each element of this vector is obtained by averaging over the elements by *N* samples, with errors proportional to the standard deviations per element. The test data is also converted to a vector of the same dimension.

$$V_{test} = (P', T', D', R') \tag{2}$$

If the inner product between the test vector and the corresponding vector of the identifier is sufficiently large, the test string is admitted to be the RIP of the right person.

$$(V, V_{test}) > \text{threshold} \tag{3}$$

However, the threshold values of this evaluation criteria are currently chosen by experience.

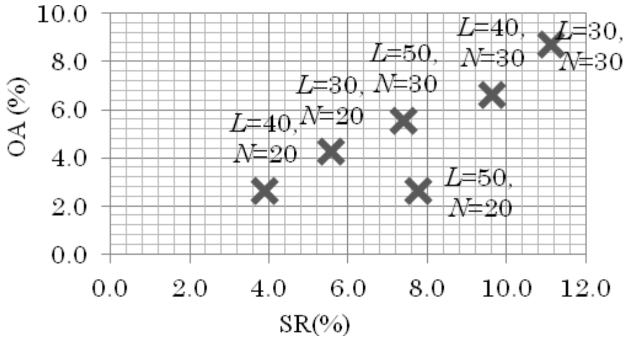
4 Experimental Results

The results are evaluated by two parameters, SR and OA. SR, the self refusal rate, is the Type-I error rate in which the right person is rejected by the ID system, and OA, the others allowable rate, is the Type-II error rate in which wrong persons are admitted by the ID systems[4]. Those two errors are reciprocal each other. It is impossible to eliminate both of them simultaneously. If the Type-I error rate is large, the right person has difficulty in entering the system, which is quite awkward but common to many biometrical IDs such as finger prints. On the other hand, if the Type-II error rate is large, the system is not secure. Those two errors depend on the choice of the threshold values. We have tested L = 30, 40, 50 and N = 20, 30.

The result is summarized in Fig.7 for the closed test, and in Fig.8 for the open test. The closed test is a test in which the test data are chosen from the set of the same data set used for learning, while the open test uses new data set independent of the data set used for learning.

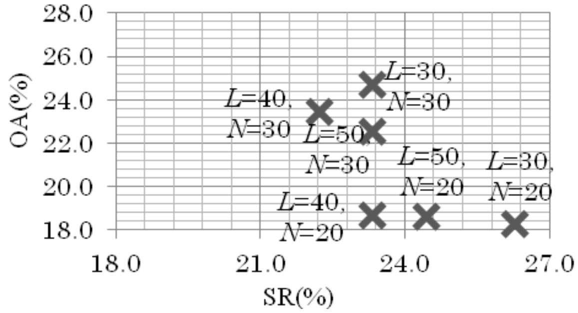
The best result, with the smallest errors, is obtained in the case of L=40, N=20 for both cases[5].

The result for individual subjects (A, B, …, H,I), together with the average value is summarized in Fig.8. The average values of the errors would be much smaller if we exclude the two subjects (E, G).



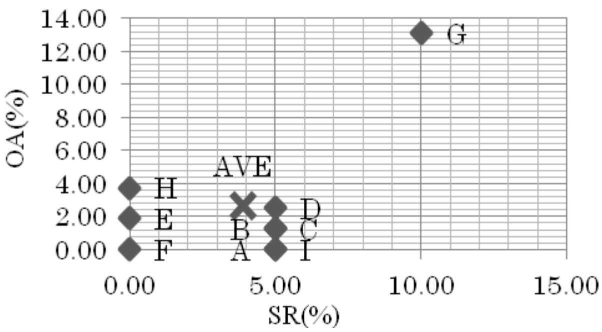
L,N	50,30	40,30	30,30	50,20	40,20	30,20
SR(%)	7.4	9.6	11.1	7.8	3.9	5.6
OA(%)	5.6	6.6	8.7	2.6	2.6	4.2

Fig. 7 Average error rates for six sets of L,N (close test)



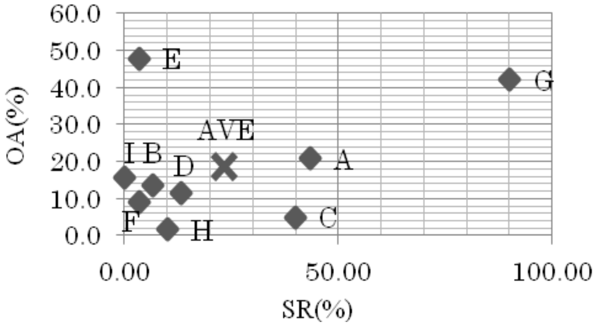
L,N	50,30	40,30	30,30	50,20	40,20	30,20
SR(%)	23.3	22.2	23.3	24.4	23.3	26.3
OA(%)	22.5	23.4	24.7	18.6	18.7	18.2

Fig. 8 Average error rates for six sets of L,N (open test)



	A	B	C	D	E	F	G	H	I	AVE
SR(%)	5.0	5.0	5.0	5.0	0	0	10	0	5.0	3.9
OA(%)	1.3	1.3	0	2.5	1.9	0	13.1	3.8	0	2.6

Fig. 9 Average error rates for individual subjects(close test)



	A	B	C	D	E	F	G	H	I	AVE
SR(%)	43.3	6.7	40.0	13.3	3.3	3.3	90.0	10.0	0	23.3
OA(%)	20.8	13.8	5.0	11.7	47.9	9.1	42.1	1.7	15.8	18.7

Fig. 10 Average error rates for individual subjects (open test)

5 Selection of Parameters

Encouraged by the partial success of the preliminary result, some attempts have been practiced in order to reduce the errors, by selecting the parameters as well as the subjects. By interviewing the nine subjects who offered the data, it is found that the subject G intentionally generated different types of data every time. On the other hand, I,B,D,F,H show quite encouraging results, as shown in Fig. 8 and Fig. 9. We therefore assume that the RIP is suitable to use among friends of a small circle who attempt to behave cooperatively. By this reason, we exclude G, E, A, C and concentrate to find good parameters to authenticate 5 subjects (I,B,D,F,H) who offer positive attitude in the experiment.

The first attempt is to compare the relative contribution of P_1, P_2, ..., P_6. Fig. 11[6] shows the degree of matching between the identifier and the test data for those parameters. The better performance is obtained as the pattern length grows up to P_4, then get saturated and falls after P_5.

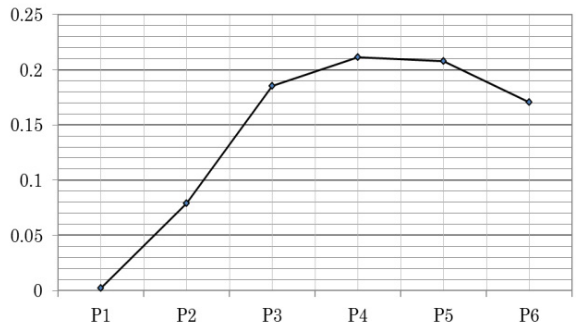


Fig. 11 Performance of P_1, P_2,..., P_6

Next, the relative performance of 8 parameters in Eq. (1), such as P_1, P_2, P_3, T_1, T_2, T_3, D, R, have been examined by mans of self-organizing maps (SOM)[7]. In the practice, the spherical SOM[8] is used in order to eliminate the boundary effect. SOM is a leaning machine based on teacher-less neural network, and used to represent multi-dimensional data in lower dimensional spaces, such as one, two , or three. The SOM is a useful to classify multidimensional data into groups. By using this tool,, various combinations of parameters are examined.

In order to examine the effectiveness of parameter P_2, the frequency of appearance of TT, TH, YT, YY, YG, YH, GT, GY, GG, GH, HT, HY, HG, HH for the selected 5 subjects are classified.

If P_2 is a good parameter to distinguish the 5 subjects, the 30 data files of each subject aggregate separately. The selected 5 subjects are coded as {A,B,C,D,E} for convenience. The 30 independent data files per subject (150 data files in total) classified into 5 distinct areas, as shown in Fig.12. For example, the 30 files of subject A are named as A-1, A-2,..., A-30, and 100% of them are gathered into the region A in the picture (a), which is viewed from the side of A. Also 100% of the 30 files of subject D, named as D-1, D-2,.., D-30, are gathered into the region D in the picture D. However, the rate of aggregation of the region B is 26/30, the region C is 29/30, the region E is 25/30, as shown in the picture (b), (c), (e), respectively.

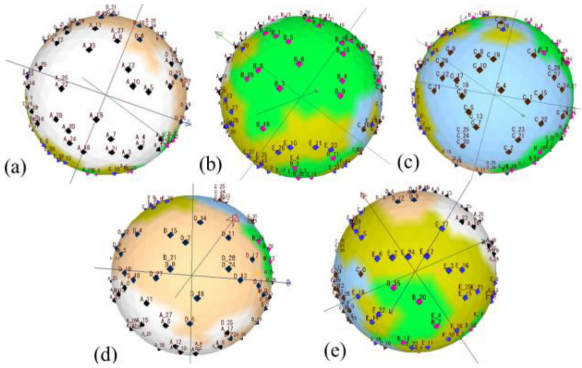
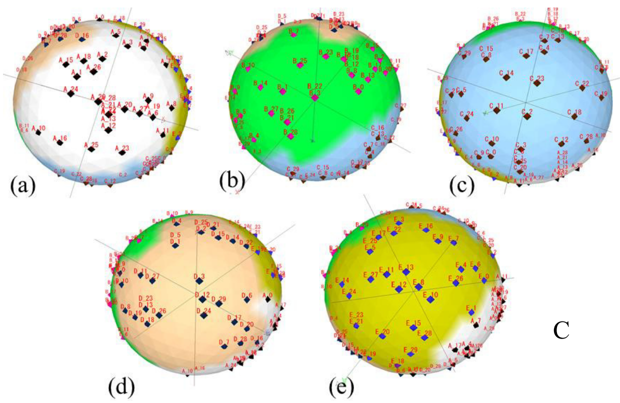


Fig.12 Result of SOM classification by using SOM based on P_2. The regions A,B,C,D,E are shown in the pictures (a), (b), (c), (d), (e), respectively.



C

Fig.13 Result of SOM classification by using SOM based on T_1 , T_2, T_3. The regions A,B,C,D,E are shown in the pictures (a), (b), (c), (d), (e), respectively.

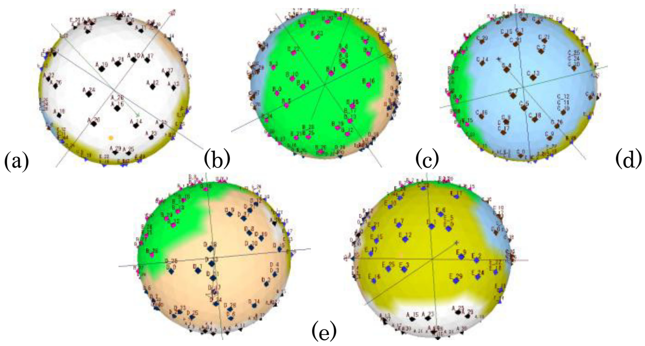


Fig.12 Result of SOM classification by using SOM based on T_1, T_2, T_3, and P_2. The regions A,B,C,D,E are shown in the pictures (a), (b), (c), (d), (e), respectively.

Table 2 Occupation rates of 30 files in each region A-E

Index \ Subjects	A	B	C	D	E
P_2	30/30	26/30	29/30	30/30	25/30
T_1 , T_2, T_3	30/30	30/30	30/30	30/30	30/30
P_2+T_1, T_2,T_3	30/30	30/30	30/30	29/30	30/30

6. Conclusion

We have constructed a personal authentication system based on the Random Input Password (RIP) and tested the system by using 9 subjects, for the parameters L=30, 40, 50, and N=20, 30. The relative frequencies of appearance of the 4 keys, the time intervals of consecutive inputs, and the patterns of RIPs are used to identify individuals. We have tested this system on 9 individuals as subjects, by closed tests as well as open tests. Using the parameters L=40 and N=20 which are supposed to be the best in our experiments, the result shows 23 percent of Type-I error, and 19 percent of Type-II error. While the level of Type-I error may be tolerable, Type-II should be significantly reduced for the sake of security. We expect to achieve improvement by reconsidering the indicators in more detail. Still, this kind of random password may not be safe by itself, if used alone. However, we can imagine various situations under which the RIP may add the level of security, if used together with other means.

References

1. W.A.Wagenaar, Generation of Random Sequences by Human Subjects: A Critical Survey of Literature, Psychological Bulletin, Vol.77, pp.65-72,1972.
2. J.N.Towse, D.Nell(1998), Analyzing Human Random Generation Behavior: A Review of Methods Used and a Computer Program for Describing Performance, Instruments & Computers, Vol.30, pp.583-591.
3. M.Tanaka-Yamawaki and M.Mishima, Effective indices to characterize short sequences of human random generations, Artificial Life and Robotics,vol.12, pp.184-187, 2008.
4. ITSCJ biometrics, http://itscj.ipsj.or.jp/topics/sc37.html
5. M.Tanaka-Yamawaki, Y. Tanaka, Proposing a System for Individual Authentication Using the Random Input Password, talk at COST Action TD 0120 held at The University of Seville, September 16-17, Seville Spain, 2014.
6. Y. Mitani, Effectiveness of input patterns in RIP, Bachelors Thesis submitted to Tottori University (2015)
7. T. Kohonen, The Self-Organizing Map, Proceedings of the IEEE, Vol.78, Issue 9, pp.1464-1480, 1990
8. The authors are grateful to SOM-Japan led by Prof. H. Tokutaka for providing us to use their product.