Int. J. Communications, Network and System Sciences, 2009, 2, 714-719
doi:10.4236/ijcns.2009.28082 Published Online November 2009 (http://www.SciRP.org/journal/ijcns/).

Application of a Dynamic Identity Authentication Model Based on an Improved Keystroke Rhythm Algorithm

Wenchuan YANG, Fang FANG
School of information and communication Engineering, Beijing University of Posts and Telecommunication, Beijing, China
*Email: yangwenchuan@bupt.edu.cn, fang.*16898@gmail.com
Received July 15, 2009; *revised August* 29, 2009; *accepted October* 2, 2009

Abstract

Keystroke rhythm identification, which extracts biometric characteristics through keyboards without additional expensive devices, is a kind of biometric identification technology. The paper proposes a dynamic identity authentication model based on the improved keystroke rhythm algorithm in Rick Joyce model and implement this model in a mobile phone system. The experimental results show that comparing with the original model, the false alarm rate (FAR) of the improved model decreases a lot in the mobile phone system, and its growth of imposter pass rate (IPR) is slower than the Rick Joyce model's. The improved model is more suitable for small memory systems, and it has better performance in security and dynamic adaptation. This improved model has good application value.

Keywords: Biometric Identification Technology, Keystroke Rhythm, Identity Authentication, Keystroke Latency Time, IPR, FAR

1. Introduction

Along with the technical development of computers and networks, unauthorized users or hackers do more and more invasions on the information system through the network. Therefore, protection of computer security has become a matter of urgency. User's identity authentication is an important means to carry out system security. The traditional login method only depends on a single password content, which has the drawback of password leak and security problems, so biometric identification technology is proposed to enhance system security by taking use of human biological characteristics.

Biological characteristics, the only difference from other people, can automatically identify and verify the physical characteristics or behavior patterns. The biometric identification technologies currently contain hand recognition, fingerprint identification, facial recognition, voice recognition, iris recognition and signature recognition, etc [1]. Although these methods can accomplish the identification, the processes of feature extractions need expensive hardware devices which limit the application range to a large extent.

Keystroke rhythm identification is a kind of biometric identification, and it has the advantages of biometric identification, in addition, the greatest advantage in terms of keystroke rhythm is low cost. It costs almost no hardware and can be applied in majority of systems which only require keyboards to accomplish the identification. Moreover, once it is applied, it can play a significant role on improving the security of keyboard input, so keystroke rhythm has a wide range of application.

20th century 80's, biometric identification using personal keystroke features was first proposed. This method can effectively prevent the illegal invasions. Keystroke rhythm identification extracts characteristics through keyboards without additional devices. So many scholars have paid much attention on this issue and have made some methods of pattern recognition applied to the identity authentication based on keystroke characteristics [2]. Rick Joyce and Gopal Gupta have also done some specific research on keystroke characteristics. In this paper, we will further improve the Rick Joyce algorithm to design a dynamic identification model based on keystroke rhythm which is suitable for small memory systems, then we realize it in a mobile phone system and do analysis of experimental data.

2. Keystroke Rhythm Algorithm of Rick Joyce Model and an Improved Model

Some studies have shown that users' keystroke rhythm characteristics like fingerprints can reflect persons' biological characteristics by the keystroke duration time and keystroke latency time [3,4]. For example, the definition of the n-th keystroke duration time means the time between the n-th button is pressed and released, the n-th keystroke latency time stands for the time between the n-th button is pressed and the (n+1)-th button is pressed. The following are two models of identity authentication based on keystroke latency time.

2.1. Rick Joyce Identity Authentication Model [5] and its Algorithm

Using keystroke rhythm as a means to study identity authentication at first is Rick Joyce and GopaIGupta. They design an identity authentication model based on keystroke rhythm – the keystroke latency time. In this system, a user is asked to type eight times of his username, password, firstname, lastname. Then the system processes the information extracted to establish a four-dimensional vector:

$$M = \left(M_{username}, M_{password}, M_{firstname}, M_{lastname}\right)$$

Each component of this four-dimensional vector is the average on eight vectors, and they use M as this user's behavioral characteristic profile. In the process of identity authentication, it is required to check the information typed by tested user (assuming he/she has been aware of the content typed), thus the input information is processed to form a test vector :

$$T = \left(T_{username}, T_{password}, T_{firstname}, T_{lastname}\right)$$

and the system just verify user's identity by comparing M with T.

The user's keystroke rhythm can be expressed as a n-dimensional vector, in which n is the total number of input intervals, and the last number of the vector means the latency time between the last button is pressed and the enter button is pressed. Then M and T can be recorded as $M = (M_1, M_2 \cdots M_n)$, $T = \left(T_1, T_2 \cdots T_n\right)$. And the difference between M and T is expressed as $\left\|M - T\right\|_1$, namely $\sum_{i=1}^{i=n} \left|m_i - t_i\right|$, which is a norm for comparison.

Then let the first user enter eight times of information again to form eight training sequences $S_1, S_2 \cdots S_8$ as the reference vectors, and they can be used to calculate the value of $\left\|M - S_i\right\|_1, i = 1, 2 \cdots 8$, then the mean and standard deviation can be gained. Consider the false alarm rate (FAR) and imposter pass rate (IPR), the verification threshold for this system is the mean plus one-and-one-half standard deviation, namely:

$$\Phi = \frac{1}{8}\sum_{i-1}^{i-8}\left\|M - S_i\right\| + 1.5\sigma$$

If the tested user meets the inequality $\left\|M - T\right\|_1 > \Phi$, he/she is considered as an imposter. If the tested user meets the inequality $\left\|M - T\right\|_1 < \Phi$, he/she is verified as an authentic user stored in system.

2.2. A Dynamic Model Based on the Improved Algorithm

The shortcomings of the algorithm above lie in: first of all, once the characteristic profile is formed, it is difficult to modify. Secondly, the Rick Joyce algorithm ignores the consideration of signature curve shape [5], and it just compares M with T by $\left\|M - T\right\|_1 = \sum_{i-1}^{i-n}\left|m_i - t_i\right|$, in which they do not refer to the detailed information of signature curve shape. Therefore, in terms of the shortcomings above, we make a corresponding improvement.

Because the calculation amount of the algorithm above is large, we propose a dynamic model based on the improved algorithm with a small amount of calculation.

The user is required to type his password in our system, and then the system processes the information typed to establish a one-dimensional vector:

$$M = (M_{password})$$

$M_{password}$ is the mean vector of eight times of keystroke latency time typed. Because of the possibility of long waiting time, the time between the last button is pressed and enter button is pressed can be ignored in this algorithm.

In general, both the keystroke rhythm of user and the tested user can be expressed as n-dimensional vectors, in which n is the total number of input intervals, namely:

$$M = \left(M_1, M_2 \cdots M_n\right), \quad T = \left(T_1, T_2 \cdots T_n\right)$$

The difference between M and T is expressed as $\left\|M - T\right\|_2$, namely:

$$\vec{\Delta} = \left\|M - T\right\|_2 = \left(\left|m_1 - t_1\right|, \left|m_2 - t_2\right| \cdots \left|m_n - t_n\right|\right)$$

Again, the first user enter eight times to form training sequences $S_1, S_2 \cdots S_8$ as the reference vectors, then we get eight difference vectors:

$$\overrightarrow{\Delta_{s_i}} = \left\| M - S_i \right\|_2, i = 1,2 \cdots 8$$

And use these eight difference vectors to calculate the mean difference vector $\overrightarrow{\Delta_s}$ and the standard deviation vector $\vec{\sigma}$.

$$\overrightarrow{\Delta_s} = \frac{1}{8}(\sum_{i=1}^{i=8} \overrightarrow{\Delta_{s_i}}), i = 1,2 \cdots 8, \vec{\sigma} = \frac{1}{8}\sum_{i=1}^{i=8} (\overrightarrow{\Delta_{s_i}} - \overrightarrow{\Delta_s})^2, i = 1,2 \cdots 8$$

Then a threshold vector is pre-established in our model:

$$\vec{\Phi} = \overrightarrow{\Delta_s} + 3\vec{\sigma}$$

Here, we compare the difference vector $\left\| M - T \right\|_2$ with the threshold vector $\vec{\Phi}$. If $\vec{\Delta} < \vec{\Phi}$, namely:

$$(\Delta_1, \Delta_2 \cdots \Delta_n) < (\Phi_1, \Phi_2 \cdots \Phi_n)$$

It means that each component of difference vector Δ_i is less than the corresponding component of threshold vector Φ_i. At this time, the identity of the tested user is authentic. If $\vec{\Delta} > \vec{\Phi}$, the tested user is verified as an imposter. Consider the non-regularity of user's operations, the matching percentage of corresponding components comparison can be adjusted a little. However, in this paper we adopt the matching percentage of comparison is 100% when the tested user will be considered as authentic user.

Besides, the system will merge the test vector into the user keystroke rhythm vector M to modify the user characteristic profile when the tested user is considered as authentic user [6]. For example, when the (k+1)-th result is legal, we can integrate this test vector with former user characteristic profile in system:

$$M_{k+1} = \frac{k \cdot M_k + T}{k+1}$$

M_k is the k-th characteristic profile of user. In this way the matching model for keystroke rhythm is variable along with the input of legal user. Accordingly, it is able to modify the keystroke rhythm and benefit the performance in authentication .

Suppose that an imposter enters a password and the certain parts of his/her keystroke latency time are longer (or shorter) but the others are normal, because Rick Joyce model considers the latency time as a whole, it may be verified as authentic user. It is required to enhance the security for some privacy systems [7]. The improved model in this paper take shape of the signature curve into account, and it contains each interval's threshold value, so the security performance is stronger.

3. System Implement

The improved dynamic identity authentication model is suitable for systems with small calculation amount, so in our study, a mobile phone system will be applied.

As the limited space and computing capability in the mobile phone system, it needs an algorithm of small calculation amount and the less storage space. For one thing, the improved algorithm does not require too much calculation, the calculation amount about comparison and modification of rhythm characteristics are small. For another thing, mobile phones are used by individuals, and it should not be required to identify more individuals and store more keystroke rhythm profiles for them, so the storage space requirements will be reduced. In addition, the amount of buttons on the mobile phone keyboard is small and the users' keystroke actions are relatively concentrated, so that it will help to form steady keystroke characteristics to enhance the security performance.

We have established a model of identity authentication using the characteristic profile of keystroke rhythm in a mobile phone system [8]. First of all, the tested users type their passwords, and then the system extracts information to form their keystrokes rhythm characteristics which are compared with the existed user characteristic profile in the mobile phone system. Only when users type the same content and their keystrokes rhythm characteristics are within the threshold value of model, they can successfully login. The system process is as follows (Figure 1):

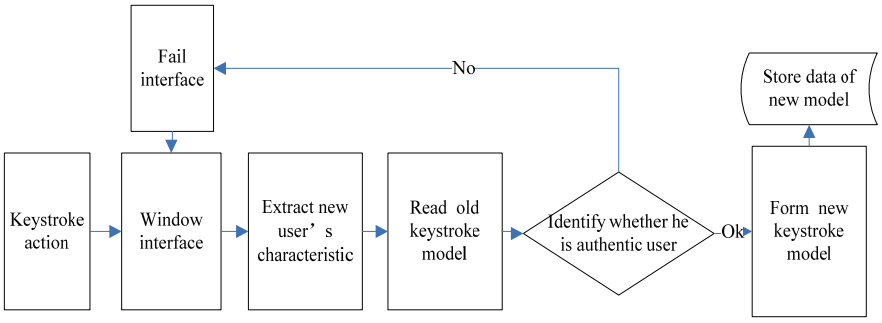

Figure 1. System process of identity authentication based on keystroke rhythm.

Table 1. Experimental data of latency time.

unit: Millseconds	latency time (rhythm features stored in system)				Latency time (authentic user)				latency time (imposters)			
1	19	33	19	32	16	38	19	37	23	32	27	30
2	17	37	20	35	17	37	19	32	21	32	26	29
3	18	37	17	36	22	40	18	35	23	28	24	27
4	16	42	22	37	17	41	19	31	25	37	29	29
5	18	41	19	32	17	39	17	37	22	32	26	28
6	18	35	19	39	18	37	19	36	23	36	26	31
7	17	37	18	34	17	34	18	34	24	29	26	32
8	16	36	20	36	18	37	17	36	21	32	27	32
9	15	38	19	32	17	34	19	35	22	29	26	30
10	19	39	14	37	18	35	20	34	19	30	25	25
11	17	40	18	30	17	35	15	31	25	30	21	26
12	17	38	14	32	17	39	17	32	20	34	21	27

4. Data Analysis and Results

The keystroke rhythms typed by users are our experimental data resources which are transformed from the mobile phone system to the computer through serial ports. The identification system requires users to login a password which contains five numbers [9]. Therefore, the vectors processed in system are four-dimensional vectors. We just choose three representive group of data in the Table 1 below. In these groups, the users need to type the password for twelve times. Then we do analysis based on these data.

Two groups of keystroke rhythm features can be gained by processing the data above, then we compare them with the rhythm features already stored in our system. The latency time curves of signature typed are shown in Figures 2-5.

Figures 2-5 show that only when the keystroke rhythm vector is nearly the same with the user characteristic profile, the tested user is considered as authentic user. However, the keystroke rhythm vector of imposters is far different from the user characteristic profile which means the difference is beyond the threshold value pre-established in system. Furthermore, the result demonstrates the improved model is able to accomplish the identity authentication based on keystroke rhythm in the mobile phone system.

In Table 2, M is the characteristic profile vector of user in system, and we choose four representive data to illustrate the performance of the improved model. Through the table we can clearly see that by the increase of threshold value, it has a good security performance in this system. Though we do not show the whole data, those listed above certainly have representive meanings in our study to a large extent. And the Rick Joyce model is also implemented in our system to make comparison with the improved model. The performance comparisons are showed in Figure 6.

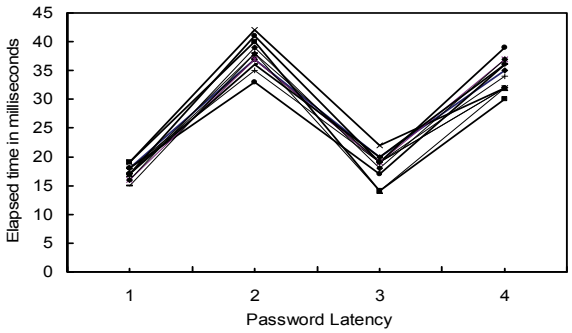

Figure 2. Keystroke rhythm features stored in system.

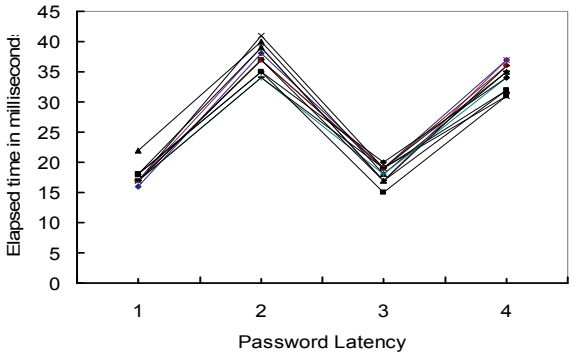

Figure 3. Twelve login attempts of authentic user.

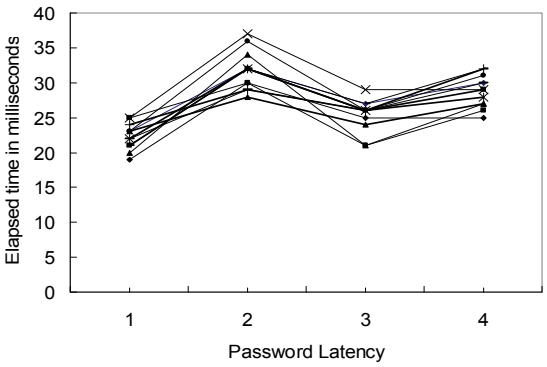

Figure 4. Twelve login attempts of imposters.

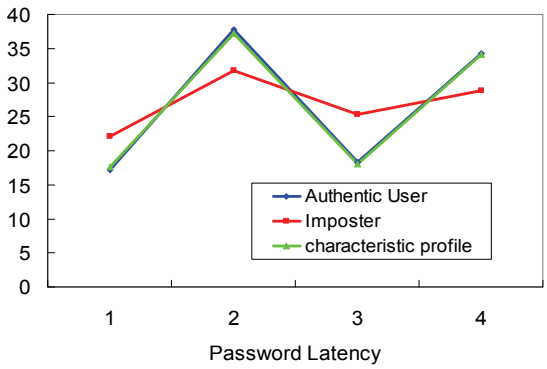

Figure 5. Comparison with the rhythm characteristic profile.

Table 2. Results comparison under different thresholds.

	latency	latency	latency	latency	Identity result
M	17.25	37.75	18.33	34.3	
$T_{authentic1}$	16	38	19	37	
$\sigma:1$	Ok	Ok	Ok	Ok	pass
$\sigma:3$	Ok	Ok	Ok	Ok	pass
$T_{authentic2}$	22	40	18	35	
$\sigma:1$	No	Ok	Ok	Ok	imposter
$\sigma:3$	Ok	Ok	Ok	Ok	pass
$T_{imposter}$	22.08	31.75	25.33	28.83	
$\sigma:1$	No	No	No	No	imposter
$\sigma:3$	Ok	No	No	No	imposter
$\sigma:4$	Ok	Ok	No	No	imposter

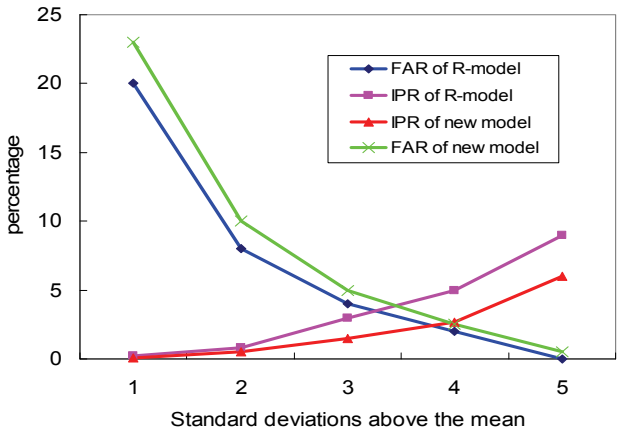

Figure 6. IPR and FAR versus threshold.

Figure 6 shows that though the IPR is not high in the Rick Joyce model (R-model) when the threshold value using one-and-one-half standard deviation [5], the FAR is not as ideal as IPR, nearly 14%, so it is not suitable for daily operations in the small memory system due to the troubles which will take to the phone users. In the improved new model, the growth rate of IPR is less than the R-model's, and it is only 1.5% in the improved model when the threshold value using triple standard deviation, which means the threshold work well. In terms of FAR, though the value is higher than the R-model's under the same threshold, the FAR is only about 5% in new model, and it is much lower than the R-model's FAR. With the increase of threshold, the FAR decreases a lot and the IPR increases slowly in the new model. Consider roundly the application in a small memory system, the new model is better to meet the actual operation needs, and this algorithm is dynamic, so along with the increase of authentication times, the identity authentication system can be improved automatically [10]. In the actual system design, through changing the multiple of the standard

deviation so as to influence the threshold value, we can easily adjust the performance in security and dynamic adaptation.

5. Conclusions

In this paper, a dynamic model based on the improved keystroke rhythm algorithm has been proposed and a mobile phone authentication system has been implemented. In this mobile phone system, the users could login the system only by typing a password which contains five known numbers. The system verifies the identity by using the dynamic model based on the improved algorithm. The experimental results show that it is able to accomplish the identification. Then, we compare the improved model with the original model in this system. It has illustrated that the FAR is 14% in the original model and 5% in the improved one, in which the FAR decreases by 9%, and the IPR increases slowly in the new model comparing with the original model. Therefore, the improved model has a high capability of authentication in the mobile phone system. Moreover, the performance in security and dynamic adaptation can be easily adjusted by changing the threshold value in the actual system design. And the improved model can be applied to meet different requirements and protect individual privacy. This model has good application value.

Although the dynamic identity authentication model based on the improved algorithm has been demonstrated in the small memory system, more experiments will be required to investigate how the keystroke rhythm model could be applied to other aspects. For the further study, other parameters, such as the length of password vector, the number of reference vectors, should be considered to enhance our ability of devising new and better models against unauthorized users or hackers.

6. References

[1] M. Zhu, J. Zhou, and J. K. Wang, "A new approach for user authentication based on biometrics," Computer Engineering, Vol. 28, No. 10, October 2002.

[2] S. Bleha, C. Slivinsky, and B.Hussien, "Computer-access security systems using keystroke dynamics[J]," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 12, December 1990.

[3] F. Monrose and A. D. Rubin, "Keystroke dynamics as a biometric for authentication. future generation computing systems (FGCS)," Journal: Security on The Web (Special issue), March 2000.

[4] F. Monrose, M. K. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamic," Proceedings of the 6 ACM Conference on Computer and Communication Security, 1999.

[5] R. Joyce and G. Gupta, "Identity authentication based on keystroke latencies," Communications of the ACM, February 1990.

[6] F. Bergadano, D. Gunetti, and C. Picardi, University of Torino, "User authentication through keystroke dynamics [J]," ACM Transactions on Information and System Security, Vol. 5, No. 4, November 2002.

[7] W. G. de Ru and J. Eloff, "Enhanced password authentication through fuzzy logic[J]," IEEE Expert, Vol. 12, No. 6, November/December 1997.

[8] Z. H. Deng, S. Q. Zhuo, etc., "The development of applications in mobile phone systems," Science Press, March 2004.

[9] R. Gaines, W. Lisowski, and S. Press, "Authentication by keystroke timing: Some preliminary results[R]," Rand Corporation:Rand Report R-2560-NSF, 1980.

[10] J. Sleggett, G. Williams, and J. Usnick, "Dynamic identity verification via keystroke characteristics," International Journal of Man-Machine Studies, 1991.

