Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105(Online)
Volume **3**, Number **3**, August **2018**

Complexity Analysis and Research based on the Chaotic System of Sample Entropy

Liu Chunyuan[1,2], Ding Qun[1]

[1]School of Electronic Engineering, Heilongjiang University, Harbin, 150080, China
[2]School of Computer and Information Engineering,
Heilongjiang University of Science and Technology,
Harbin 150022, China
Qunding@aliyun.com

Received March 2018; revised May 2018

Abstract. *This paper has analyzed the analysis method of Sample Entropy sequence linear complexity, and used this algorithm to calculate the measure entropy of Logistic chaotic system movement. Theoretical study and simulation realization show that, the method of short observation sequence can accurately reflect the chaotic system and the complexity size of chaos pseudo-random sequence, and can be used to judge the complexity of the pseudo-random sequence generated by chaotic system, as well as has the good consistency compared with approximate entropy (ApEn), meanwhile with high execution efficiency. The experimental results show the validity of this method and the correctness of the theoretical results.*
Keywords: Sample Entropy, Complexity, Logistic, Chaotic system

1. **Introduction.** Chaos, as a classic complex phenomenon of nonlinear dynamic systems, has been widely used in the research of secure communication in recent decades [1,2]. Because the chaotic sequence has the characteristics like wideband, noise like and being sensitive to the initial state, it can replace the traditional pseudo-random sequence, and apply it to the military and commercial spread spectrum communication system requiring high security degree.

The complexity of a finite length sequence refers to the similarity degree with a random sequence, and is a criterion of the difficulty degree in restoring the whole by using part of the sequence. The greater the complexity of the sequence, the greater the randomness, and the greater the difficulty of the sequence being restored. Therefore, complexity is an important index to measure the anti-interference and interception ability of spread spectrum sequence in secure communication system, and it is of great importance to further study the complexity of chaos.

At present, the chaotic sequence complexity measure algorithm mainly includes the approximate entropy (ApEn) algorithm [3] proposed by Pincus et al and the sample entropy (SampEn) [4] algorithm proposed by Richman et al. The similarity criterion adopts the Heaviside function, and is very sensitive to the threshold value r and the phase space dimension m, but SampEn does not calculate its matched statistical magnitude, because entropy is a criterion generated by new information, which can be regarded as the improvement to the ApEn, but there will be unmeaning $\ln 0$ in the absence of template matching.

In this paper, the SampEn algorithm is applied to the complexity analysis of chaotic system. First, it analyzes the characteristics of sample entropy and the advantages of sample entropy and approximate entropy in the aspect of complexity algorithm, and then applies this algorithm respectively to continuous chaotic system and uses different quantization accuracy of chaotic sequences for complexity analysis, and finally discusses that this algorithm occurs in digital chaotic sequence password, showing that this method can be applied in measuring analysis of complexity, with good stability, and can provide a theoretical basis for chaotic cryptography and chaotic secure communication.

2. **The Principle and Description of the Sample Entropy Algorithm.** Entropy is defined as a generation rate of information. $\mathrm{ApEn}(m, R, N)$ is the approximate value of the negative average natural logarithm CP, the conditional probability of mutually similar in $m+1$ point data segment mode, when the data length is N, the similarity tolerance is r, and the m point data segment mode is similar to each other. One thing to keep in mind is that the parameter r is usually expressed by the percentage of the SD (standard deviation) of the data, so the approximate entropy is a method of scale invariance. B is defined as the probability of sequence self-similarity when dimension is m, and A is the probability of sequence self-similarity when dimension is $m+1$, so $CP = A/B$.

From the approximate entropy algorithm, it can be seen that the approximate entropy takes $\log(CP)$ as the calculation model, and calculates the average value of all models. To avoid the occurrence of $\log(0)$, both A and B should not be 0. As a correction, CP should be defined as $(A+1)/(B+1)$.To ensure this point, there is a comparison between the data segments in the definition of approximate entropy, but this is obviously incompatible with the viewpoint of new information, and there must be some deviations. It contains two layers of meaning: on the one hand, the ApEn value is related to the data length, and the ApEn value of the short sequence is obviously smaller than the expected value. On the other hand, the consistency is poor. If a time sequence has a higher ApEn value than the other time sequence, then m value and z value should also have higher ApEn value, but Apen cannot meet. This deviation is also the cause of its insensitivity to small complexity changes. In fact, even so, when A and B are too small and the matching quantity is too little, which is close to $CP = 1$, APEn $= 0$, there will appear a large error. Therefore, the algorithm of approximate entropy needs improving to overcome the effect of this deviation [5].

Sampen (Sample Entropy) is a new measure method of time sequence complexity proposed by Richman. It is CP's strict natural logarithm [6, 7], which can be represented by $Sampen(m, r, N)$. Among them, N stands for length, r stands for similar tolerance, and the dimension is m and $m+1$. The sample entropy is similar to the improvement on the approximate entropy algorithm in terms of algorithm: relative to entropy, the calculation of sample entropy is the sum logarithm, whose advantage lies in containing larger A, B value and more accurate CP estimation. Sample entropy aims at reducing the error of approximate entropy, which has closer consistency with the known random part. Sample entropy is a method similar to the current approximate entropy, but with better accuracy. The sample entropy algorithm consists of two parameters m and r. m is the embedding dimension, and r is the threshold value, also called the similarity coefficient. The calculation steps are as follows

(1) The sequence $\{x_i\}$ makes up of m dimensional vector in sequence:

$$X(i) = [x(i), x(i+1), \ldots, x(i+m-1)] \quad i = 1, 2, \ldots, N-m+1 \tag{1}$$

(2) Define $d[X(i), X(j)]$ the one with the maximum difference value in the corresponding element $X(i)$ and $X(j)$, that is,

$$d[X(i), X(j)] = \max_{k=0,2,\ldots,m-1} |x(i+k) - x(j+k)| \tag{2}$$

For each i value, calculate the distance $d[X(i), X(j)]$ from $X(i)$ to the rest vectors $X(j)$, $(j = 1, 2, \ldots, N - m + 1)$. The difference value between other corresponding elements in $X(i)$ and $X(j)$ at this time is naturally less than d.

(3) In accordance with the given threshold value $r\,(r > 0)$, count the number of $d[X(i), X(j)] < r$ for each i value and the ratio of this number to the number of the total vectors $N - m$, marked as $B_i^m(r)$.

$$B_i^m(r) = \frac{1}{N-m} num\{d[X(i), X(j)] < r\} \quad i = 1, 2, \ldots, N - m + 1, i \neq j \tag{3}$$

(4) Average all i values, marked as $B^m(r)$, that is:

$$B^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_i^m(r) \tag{4}$$

(5) The dimension plus 1 into $m + 1$, repeating the process of (1) – (4), and obtain $B^{m+1}(r)$ and $B_i^{m+1}(r)$:

$$B_i^{m+1}(r) = \frac{1}{N-m-1} num\{d[X(i), X(j) \leq r]\} \quad i = 1, 2, \ldots, N - m \tag{5}$$

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m+1} B_i^m(r) \tag{6}$$

(6) At this time, the theoretical value of the sample entropy is:

$$SampEn(m, r) = \lim_{N\to\infty} \left\{ -\ln[B^{m+1}(r)/B^m(r)] \right\} \tag{7}$$

But in actual work, N can't be ∞. When N takes a finite value, the estimated sample entropy is:

$$SampEn(m, r, N) = -\ln[B^{m+1}(r)/B^m(r)]$$

The $SampEn$ value is obviously related to the value of m and r. The sample entropy of different embedding dimensions m and similar tolerance R is also different. In general, the sample entropy obtained under $m = 1$ or 2, $r = 0.1 \sim 0.25$SD has reasonable statistical properties. Therefore, the parameters of the sample entropy in this paper are $m = 2$, and $r = 0.2$SD (SD is the criterion deviation of the original data). The study shows that, in addition to the very simple cases, the entropy of time sequence has a close relation with the variance SD.

The sample entropy is similar to the improvement on the approximate entropy algorithm in terms of algorithm, with the following good properties.

(1) The sample entropy does not contain the comparison of its own data segments, so it is the exact value of the negative natural logarithm of conditional probability, and the calculation of sample entropy does not depend on data length.

(2) The sample entropy has better consistency. If a time sequence has a higher ApEn value than the other time sequence, then it should also have higher $SampEn$ value for other m and r values.

(3) The sample entropy is not sensitive to the obliterated data. Even if the data is obliterated as much as 1/3, the effect on the SampEn calculation is still very poor.

3. Complexity Analysis of Sample Entropy in Chaotic System.

3.1. Complexity Analysis of Logistics System.

One-dimensional Logistic mapping is a very simple chaotic mapping from the mathematical form, but this system has complex dynamic behavior and is widely applied in the field of secure communication. Its mathematical expression is:

$$x_{n+1} = \mu x_n(1 - x_n) \quad 0 \le x \le 1, 0 < \mu \le 4 \tag{8}$$

Among them, x_n is the state value of the Logistic mapping, and μ is the parameter coefficient of the iterative equation. The value of the adjustment parameter μ can generate the chaotic phenomenon. When Logistic mapping is in the parameter $\mu \in (3.57, 4]$, iteration 1000 enters the chaotic state. Figure 1 shows the time sequence mapping after 10000 iterations of one dimensional Logistic chaotic mapping when the initial value $x_0 = 0.3$, $\mu = 4$, and the length $N = 1000$.

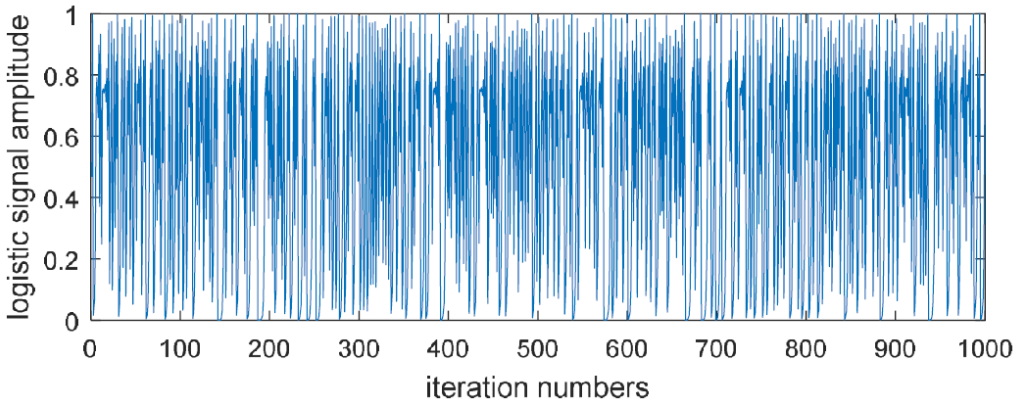

FIGURE 1. Tent signal in time domain

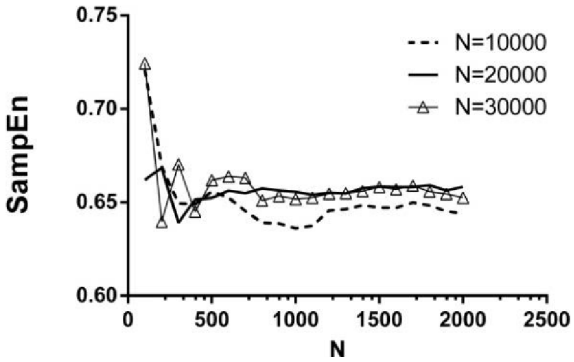

FIGURE 2. The complexity of sequence indifferent iteration times

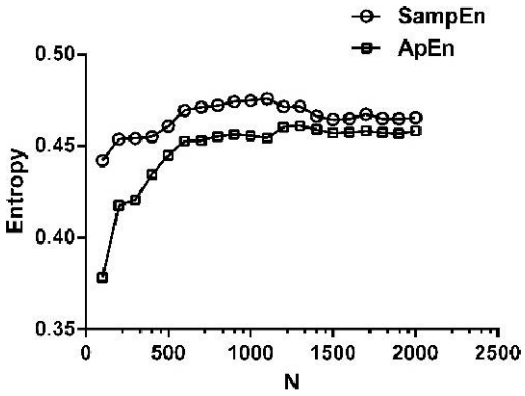

FIGURE 3. The measure entropy ofthesame sequence intwomethods

Figure 2 gives the analysis results of sample entropy of chaotic random sequence when the Logstic sequence whose length is respectively from100 to 2000 goesthrough10000, 20000 and 30000 iterations. Among them, the initial value $x_0 = 0.3$, $\mu = 4$, $r = 0.15$SD and $m = 2$. The results show that when the iteration times reaches 20000, the value of sample entropy of random sequence basically tends to be uniform. When 10000 iteration

times, for the value of sample entropy, its randomness is significantly lower than the sequence with high iteration times. In addition, it can be concluded that the SampEn value of each iteration time is basically consistent after the sequence length reaches 1000, which means that the complexity analysis method is not very dependent on the sequence length.

Figure 3 has given the Logistic sequences with different lengths after 20000 iterations when $x_0 = 0.6$, $\mu = 4$, $r = 0.15$SD, $m = 2$. The results of two test methods for the same random sequence are observed. First, for the approximate entropy, when the sequence length N reaches a certain length, the entropy value of the sequence basically tends to be uniform. But for the shorter sequence, the difference of approximate entropy value is quite big. Compared with the sample entropy, it is more dependent on the sequence length N. For sample entropy, the sample entropy of time sequence is almost around 0.45, and the sample entropy has better consistency than the approximate entropy. In addition, the SampEn method has a higher execution efficiency, and the execution time of the program is far less than that of the ApEn method. Thirdly, it should be noted that the value of μ and the initial value x_0 have certain influence on the entropy of the random sequence when studying Logistic random sequence.

3.2. **Complexity Analysis of Sample Entropy of the Binarization Logistics Time Sequence.** The chaotic random sequence after quantization is called as the binary sequence. In fact, in the random number generation process of any chaotic sequence, however good the chaotic characteristic of the original sequence is, and however good the randomness the chaotic sequence shows, the data compression and information loss in the quantization process will have different degrees of effects on the randomness of digital chaotic sequences. Since the emergence of chaotic encryption, the quantization process has always been a necessary link for the discrete chaotic system transforming to the digital chaotic system. Due to the limitation of computation accuracy, the energy loss of dynamic degradation and other reasons, the quantization accuracy in the binarization process also has certain influence on the complexity of sequence [12–16].

In order to check the test results of complexity of the binary sequence after quantization by the sample entropy, the simulation experiment generates the sequence to be tested by the following method: by using the Logistic equation, the initial value x_0 of $= 0.7$, $\mu = 4$, and M fixed-point precision to start iteration, the chaotic sequence after quantization generated by function with the length of 10^4 is called the sequence to be tested, and the quantization precision were given, which are respectively $M = 8, 16, 24, 32, 64$. Different sequences to be tested have been checked practically, respectively by SampEn method and ApEn method.

Table 1. The entropy of Logistic Mapping

M	8	16	24	32	64
SampEn	0.0161	0.6823	0.6930	0.6930	0.6931
ApEn	0.0453	0.6859	0.6920	0.6911	0.6911

From Table 1, it can be seen that for chaotic pseudo-random sequences, both SampEn method and ApEn method can reflect the complexity of phase space structure of Logistic chaotic binary sequence to a certain extent. However, from the numerical display, the complexity value measured by the sample entropy method is more consistent, which is in accordance with the previous analysis result. The phenomena of different complexity under the same condition and the same sequence length when accuracy $M = 8$ and

$M = 24$, shows that, for Logistic chaotic system, the selection of quantization accuracy and quantification method in sequence binarization have relatively large effect on the complexity of sequence. Meanwhile, for Logistic mapping, the stability value of SampEn is about 0.69 [8], while the Lyapunov index of Logistic mapping is 0.69, which proves from another aspect that the SampEn can be applied to the measure entropy of digital chaotic sequence in Logistic system.

4. **Complexity Analysis of Digital Chaotic Pseudorandom Sequence.** In order to test the actual digital sequence of sample entropy, the experimental device first applies DSP Builder and Quartus II to design and implement a chaotic sequence generator [9,10]. The design schematic diagram of a chaotic sequence generator is shown in Figure 4.

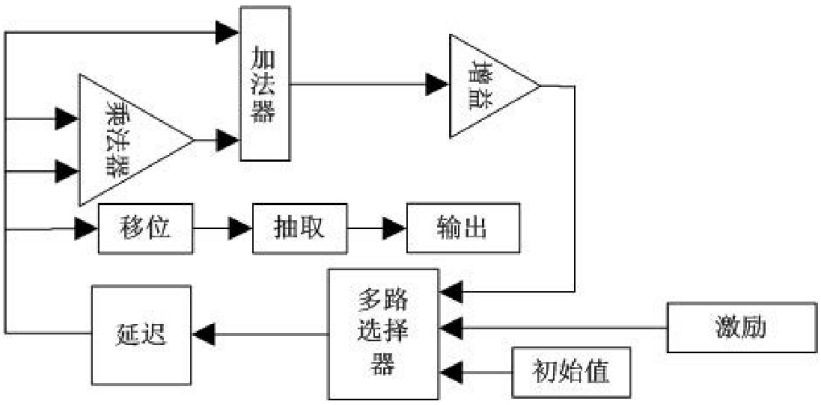

FIGURE 4. The principle diagram of the chaotic sequence generator

Using the above chaotic sequence generator and the quantization method used in this paperis: use irreversible transformation function $T[x(n)]$ to transform the real value sequence generated by the chaotic equation into 0-1 sequence. The definition of the transformation function $T[x(n)]$ is as follows:

$$T[x(n)] = \begin{cases} 0 & x(n) \in \bigcup_{k=0}^{2^m-1} I_{2k}^m \\ 1 & x(n) \in \bigcup_{k=0}^{2^m-1} I_{2k+1}^m \end{cases} \tag{9}$$

The value is 1 or 0 according to the difference of original value $x(n)$ lying in the continuous equal interval of the conversion function $T[x(n)]$. Among them, $m > 0$ is an arbitrary integer, $I_0^m, I_1^m, I_2^m, \ldots$ is 2^m continuous equal interval in the interval $[0, 1]$. In order to facilitate the realization of the function, the formula (9) is transformed as follows:

$$T[x(n)] = \begin{cases} 0 & 2^m x(n) \in [2k,(2k+1)) \\ 1 & 2^m x(n) \in [(2k+1),(2k+2)] \end{cases} \tag{10}$$

In order to furthest simplify the design of circuit, the formula (10) can be used to judge the parity of the single digit of the integer of the $2^m x(n)$ product term to determine whether the output of the function is 1 or 0.In the design of DSP Builder, the unit extractor is used to realize the parity judgment of the single digit of the integer, and the left shifting function of the register shift is used to complete $2^m x(n)$, to furthest simplify the design of circuit [11]. Next, approximate entropy and sample entropy are applied to test and analyze the theoretical values generated by Logistic chaotic equation iteration, the actual value generated by hardware and the digitalized values after quantification by the above

transformation function. For the parameters in the algorithm, N are respectively 300, 500, 1000, 5000, 10000, parameter $m = 2$, $r = 0.2$SD. The test results are shown in table 2.

TABLE 2. The results of digitalized chaotic sequence in different test methods

	ApEn $= (m, r, N)$		SampEn $= (m, r, N)$	
N	Actual Value	Quantized Value	Actual Value	Quantized Value
300	0.2470	0.2265	0.6483	0.6554
500	0.4559	0.4211	0.6734	0.6764
1000	0.6314	0.6200	0.6835	0.6809
5000	0.6936	0.6760	0.6930	0.6806
10000	0.6929	0.6931	0.6815	0.6739

From table 2, it can be seen that the complexity of the digitalized chaotic random sequence declines compared to the complexity of the chaotic sequence before quantization. The decrease of complexity is likely to lose some useful information of the original chaotic dynamical system. This information needs to be further studied. The digitalized chaotic sequence basically maintains the characteristics of the original sequence, and its complexity can reflect the complexity of the original system. The data show that the sample entropy method is more consistent, more independent on the sequence length N. When length < 1000, the complexity value of the SampEn method is more stable. For the ApEn method, the value when $N = 300$ and $N = 500$ is markedly less than 0.69, showing the characteristic more restricted by the sequence length. For $N = 10^4$, the complexity of the sample entropy method is lower than that of the approximate entropy. The reason is that when the sequence generator generates a long sequence, after quantization, the periodicity generated by the sequence increases, so the measurement entropy decreases. From another side, it is shown that the sample entropy can be well used in the complexity measure of digitalized chaotic sequence. The research results provide theoretical and experimental basis for the application of chaotic sequence ininformation security. Studying the complexity of digitalized chaotic sequence is of great significance for studying the complexity of chaotic sequence.

5. **Conclusion.** In this paper, the SampEn algorithm and the ApEn algorithm are used to systematically calculate and analyze the complexity of the Logistic discrete chaotic system and the digitalized chaotic sequence. It is analyzed that the sample entropy method can more effectively measure the complexity of chaotic sequence and chaotic pseudorandom sequence, and this method shows higher execution efficiency in simulation experiment. Compared with the ApEn method, the two both have low requirements of the sequence length N, but SampEn requires less for the sequence length N, which shows better consistency. In the aspect of quantization accuracy, for the random sequence generated by chaotic system in the binarization process, the quantization precision has some influence on the randomness of binary sequence. Too low quantization accuracy will cause the decline of sequence randomness and complexity. This characteristic can be proved through the SampEn method. In terms of the complexity test of digital random sequence generated by the experimental device, the sample entropy method is consistent with the Lyapunov index 0.69 on the complexity test, indicating that the SampEn method is effective in the chaotic random sequence and the digitalized chaotic sequence.

Acknowledgment. This work is partially supported by Natural Science Foundation of China (No.61471158).

REFERENCES

[1] J. M. Amigó, L. Kocarev, J. Szczepanski, Theory and practice of chaotic cryptography, *Physics Letters A*, vol. 366, no. 3, pp. 211–216, 2007.

[2] L. L. Huang, K. T. Yin, Chaotic synchronization secure communication system based on output control, *Journal of Electronics and Information*, vol. 31, no. 10, pp. 2402–2405, 2009.

[3] S. Pincus, S. M. Pincus, Approximate entropy (ApEn) as a complexity measure, *Chaos*, vol. 5, no. 1, pp. 110–117, 1995.

[4] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology Heart & Circulatory Physiology*, vol. 278, no. 6, pp. H2039, 2000.

[5] D. M. Bai, Characteristics analysis and feature extraction of EEG, of Dalian University of Technology, 2006.

[6] J. S. Richman, J. S. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology Heart & Circulatory Physiology*, vol. 278, no. 6, pp. H2039, 2000.

[7] D. E. Lake, J. S. Richman, M. P. Griffin, Sample entropy analysis of neonatal heart rate variability, *Am J Physiol Regul Integr Comp Physiol*, vol. 283, no. 3, pp. R789, 2002.

[8] J. P. Cai, Z. Li, W. T. Song, A complexity analysis methodof chaotic pseudorandom sequence, *Physical Journal*, vol. 52, no. 8, pp. 1871–1876, 2003.

[9] Q. Ding, L. Wang, Periodic extension method and implementation of new digital chaotic key sequence generator, *Journal of Scientific Instrument*, vol. 32, no. 10, pp. 2316–2323, 2011.

[10] H. J. Wang, B. B. Song, Q. Liu, FPGA Design and Applicable Analysis of Discrete Chaotic Maps, *International Journal of Bifurcation & Chaos*, vol. 24, no. 04, pp. 917–921, 2014.

[11] Y. B. Zheng, Y. Song, B. X. Du, A new method testingthe periodic characteristics of chaotic cipher sequences, *Physical Journal*, vol. 61, no. 23, pp. 230501–230501, 2012.

[12] C. M. Chen, K. H. Wang, T. Y. Wu, and Eric Ke Wang, On the Security of a Three-party Authenticated Key Agreement Protocol based on Chaotic Maps, *Data Science and Pattern Recognition*, vol. 1, no. 2, pp. 1–10, 2017.

[13] C. M. Chen, W. C. Fang, K. H. Wang, T. Y. Wu, Comments on An improved secure and efficient password and chaos-based two-party key agreement protocol , *Nonlinear Dynamics*, vol. 87, no. 3, pp. 2073–2075, 2017.

[14] C. M. Chen, L. L. Xu, K. H. Wang, S. Liu, T. Y. Wu, Cryptanalysis and improvements on three-party-authenticated key agreement protocols based on chaotic maps, *Journal of Internet Technology*, 2017, DOI: 10.6138/JIT.2018.19.5.20160710.

[15] N. Lin, H. F. Zhu, Enhancing The Security of Chaotic Maps-based Password-Authenticated Key agreement Using Smart Card, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 6, pp. 1273–1282, 2017.

[16] H. F Zhu, R. Wang, Multi-party Password-Authenticated Key Exchange Scheme with Privacy Preserving using Chaotic Maps in Random Oracle Model, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 1, pp. 42–53, 2017.