ON THE CRYPTOGRAPHIC MEASURES AND CHAOTIC DYNAMICAL COMPLEXITY MEASURES

[1]Hanping Hu, [2]LingFeng Liu*

[1, 2]Institute of Pattern Recognition & Artificial Intelligence, Huazhong University of Science &Technology, Wuhan, PRC
[1]hphu@mail.hust.edu.cn; [2]vatanoilcy@163.com;

ABSTRACT

The relationship between cryptographic measures and chaotic dynamical complexity measures is studied in this paper, including linear complexity and measure entropy, nonlinear complexity and source entropy. Moreover, a method is presented to guarantee the complexity of chaos-based pseudorandom sequence due to this relationship. This is important to the development of chaos-based cryptography.

KEYWORDS

Chaos-based cryptography, Cryptographic measures, Linear complexity, Nonlinear complexity, Chaotic dynamical complexity measures, Measure entropy, Source entropy

1. INTRODUCTION

Pseudorandom binary sequences are widely applied to the various engineering domains such as biology system control, spread spectrum communications and cryptography [1-3]. Particularly, the security of cryptographic systems is strongly contingent on the pseudorandomness of the key stream [3]. Depending on the cryptography application, a sequence is required to present many properties so as to be considered as pseudorandom. The linear complexity and nonlinear complexity of a sequence s are both important cryptographic measures. The linear complexity is defined as the length of the shortest linear feedback shift register (LFSR) that generates s, which has already been widely studied for several decades [4-9]. The nonlinear complexity is defined as the length of the shortest feedback shift register (FSR) that generates s. On the contrary, it has not been studied to the same extent [10-15].

Chaotic system is regarded as an important pseudorandom source in the design of pseudorandom number generators. The main reason is in accordance with its complex dynamic behaviors, such as sensitive to initial condition, no periodic and long-term unpredictability etc. Lyapunov exponent and measure entropy are the most widely used complexity measures to chaotic system. Lyapunov exponent, which provide a qualitative and quantitative characterization of dynamical behavior, are related to the exponentially fast divergence or convergence of nearby orbits in phase space. A system with one or more positive Lyapunov exponents is defined to be chaotic. Measure entropy describes the unpredictability of chaotic system. In practice, it is hard to calculate and may be inaccurate, so approximate entropy was proposed as a complexity measure of chaotic system for its facilitation to calculate [16]. Approximate entropy appears to have potential

application to a wide variety of relatively short (greater than 100 points) and noisy time series data [17].

However, by using chaotic system as a pseudorandom source, the complex dynamic behaviors of chaos may not lead to a high complexity in cryptography. The relationship between chaotic dynamical complexity measures and cryptographic measures still remains an open problem, which is a serious impediment to the development of chaos-based cryptography, such as the design or selection of chaotic system. Without this relationship, we can only select a chaotic system arbitrarily, then, analyze the security of this chaos-based cryptography. If it doesn't satisfy the security requirements, we should select again. Obviously, it is rather blind and time-consuming. Till now, no such researches show the relationship between these two kinds of complexities. Thus, this is what we will do in this paper. The establishment of relationship between these two kinds of complexity measures will link the chaos-based cryptography into the mainstream cryptographic research thinking, and guide us to design pseudorandom binary sequences which satisfy the requirement of cryptographic complexity.

The paper is organized as follows. In Section 2 the basic terminology and definitions are introduced. The connections of linear complexity with measure entropy are presented in Section 3. In Section 4, we explore the relationship between nonlinear complexity and source entropy. Finally, concluding remarks are given in Section 5.

1. PRELIMINARIES

Let F_2 denote the binary field, and F_N denote the finite field with N elements. Let $y=y_0y_1y_2\ldots$ be a sequence and denote by y_i^j, with $i \le j$, the tuple $y_i...y_j$. If y has finite length N, then $y^N:=y_0^{N-1}$ denotes the whole sequence. For any $0\le j<N$, the tuple y_0^j is a prefix of y^N; for the special case that $j<N-1$, such a prefix is called proper prefix. A suffix of y^N is any tuple y_j^{N-1}, $0\le j\le N-1$; a proper suffix is defined similarly. If there exist $t_0\ge 0$ and $T>0$ such that $y_i=y_{i+T}$ for all $i\ge t_0$ then the sequence is called ultimately periodic. If $t_0=0$, the sequence is simply periodic. The least t_0, T with this property are called preperiod and period respectively.

Any ultimately periodic sequence can be generated by a feedback shift register, satisfying a recurring relation

$$y_{i+n} = h(y_i,...y_{i+n-1}), \quad i \ge 0$$

Where $n>0$ equals the length of the FSR. The function h is called the feedback function of FSR. If h is a linear function, then the sequence is generated by a linear feedback shift register with the length n.

Definition 2.1: The length of the shortest LFSR generating a sequence y^N is referred to as linear complexity of y^N, and is denote by $L(y^N)$.

Definition 2.2: The length of the shortest FSR generating a sequence y^N is referred to as nonlinear complexity of y^N, and is denote by $c(y^N)$.

Definition 2.3: Let $y^N \in F_2$ be a binary sequence. Let M be a given Turing machine generating y^N. Let $p(M)$ be the program, which makes this Universal Turing Machine (*UTM*) simulate M. Then

the number $\mu(M):=|p(M)|$ is called the size of M, where $|p(M)|$ means the length of $p(M)$. The number $K(y^N):=\mu_M(y^N)=\min\{\mu(M)|M$ generates $y^N\}$ is called the kolmogorov complexity of the sequence y^N.

Practically, all sequences of moderately large length have a kolmogorov complexity close to the length of the sequence [18].

Let (X,T) denote a chaotic system. For any initial conditions $x_0 \in X$, let $T^*(x_0)=(x_0, T(x_0),\ldots,T^m(x_0),\ldots)$ denote the orbit of the chaotic system (X,T). Let $\alpha=\{C_0,\ldots,C_{N-1}\}$ denote a finite measurable partition of X. For any $x \in X$, let us define

$$\psi_\alpha(x)=\{\omega\in\{0,1,...,N-1\}^N : \forall k\in N, T^k(x)\in C_{\omega(k)}\}$$

The set $\psi_\alpha(x)$ is the set of all possible codings of the orbit of x relative to the partition α. Let $T^*(x_0|\alpha)=((x_0|\alpha), T(x_0|\alpha),\ldots,T^m(x_0|\alpha),\ldots)$ denote the coding of the orbit $T^*(x_0)$ relative to the partition α.

2. CONNECTIONS OF LINEAR COMPLEXITY WITH MEASURE ENTROPY

In this section, we take kolmogorov complexity as a bridge, reveal the connections of linear complexity with measure entropy.

Lemma 3.1 [18]: For any real number ε, $0<\varepsilon<1$, arbitrary integer N, and an arbitrary finite sequence $y^N \in F_2$, we have:

$$\text{Prob}\left\{y^N\in F_2 \,|\, (1-\varepsilon)\cdot 2L\left(y^N\right)\le K\left(y^N\right)\le(1+\varepsilon)\cdot 2L\left(y^N\right)\right\}\to 1$$

for moderate large N, where $L(y^N)$ means the linear complexity of the sequence y^N.

***Lemma 3.2 [18]*:** For any arbitrary infinite sequence y^N, y^n is the prefix of y^N, with length n, m is a Lebesgue measure, then we have:

$$\lim_{n\to\infty}\frac{K\left(y^n\right)}{2L\left(y^n\right)}=1.$$

for m-almost everywhere.

From these two lemmas, we can see that the kolmogorov complexity and the linear complexity are the same for practically all 0-1-sequences of length N, already for moderately large N. So we have

$$K\left(y^N\right)=2L\left(y^N\right)$$

for moderately large *N*.

Let (*X*,*T*) be a chaotic system with a partition α. The complexity of the orbit $T^*(x_0|\alpha)$ with the initial condition x_0 is defined as:

$$\sup K(x_0, T \mid \alpha) = \limsup_{n\to\infty} \frac{K(T^*(x_0 \mid \alpha))}{n}$$

$$\inf K(x_0, T \mid \alpha) = \liminf_{n\to\infty} \frac{K(T^*(x_0 \mid \alpha))}{n}$$

The next theorem will reveal the connection of kolmogorov complexity with measure entropy.

Lemma 3.3 (Brudno, White) [19]: Let (*X*,*T*) be a chaotic system. If *X* is compact, μ is an ergodic probability measure on (*X*,*T*), then

$$\inf\ K(x_0, T \mid \alpha) = \sup\ K(x_0, T \mid \alpha) = h_\mu(T \mid \alpha)$$

for μ-almost each $x_0 \in X$. $h_\mu(T|\alpha)$ is the measure entropy of the dynamical system (*X*,*T*) relative to the partition α.

From this theorem and the definition above, we have

$$\lim_{n\to\infty} \frac{K(T^*(x_0 \mid \alpha))}{n} = h_\mu(T \mid \alpha).$$

Till now, we can get the following main theorem.

Theorem 3.1: Let (*X*,*T*) be a chaotic system with a partition α. *X* is compact, μ is an ergodic probability measure on (*X*,*T*), $T^*(x_0|\alpha)$ is a coding sequence of the orbit $T^*(x_0)$ relative to α, then the linear complexity of $T^*(x_0|\alpha)$ increases while the entropy $h_\mu(T|\alpha)$ increases for moderately large length *N*.

Proof: This theorem can be easily acquired from the Lemmas above. According to Lemma 3.3, we have $\lim_{n\to\infty} \frac{K(T^*(x_0 \mid \alpha))}{n} = h_\mu(T \mid \alpha)$. From this equation we can see that when $h_\mu(T|\alpha)$ increases, $K(T^*(x_0|\alpha))$ will also increase for moderately large length *N*. Combining with the equation $K(T^*(x_0|\alpha)) = 2L(T^*(x_0|\alpha))$, concludes the proof.

According to this theorem, we can evaluate the security of chaos-based cryptography by using $h_\mu(T|\alpha)$. More importantly, we can use $h_\mu(T|\alpha)$ to select or construct a suitable chaotic system and encoding to guarantee the chaos-based pseudorandom sequence has a high linear complexity. The method for making $h_\mu(T|\alpha)$ large can be shown as follow.

First, choose or construct a chaotic system which has a large Kolmogorov-shannon entropy $h_\mu(T)$. The Kolmogorov-shannon entropy $h_\mu(T)$ defines as the biggest entropy in all partitions.

Second, choose a proper partition α (we always call this generating partition) so as to make the entropy $h_\mu(T|\alpha)$ close to $h_\mu(T)$.

Accordingly, we choose the following logistic map as its Kolmogorov-shannon entropy $h_\mu(T)$ is rather large in the most frequently-used one-dimension chaotic system, for $h_\mu(T)$=0.6929.

$$x_{k+1} = 4x_k(1-x_k)$$

Now we get the generating partition. For a given chaotic system, it is known that the generating partition can be specified by using the set of an infinite number of unstable periodic orbits (UPO's) embedded in the underlying dynamical invariant set. Here we used the method proposed in [20] to get our generating partition.

The topological entropy of this attractor is $h_T \approx 0.69 < \ln 2$, so the symbolic dynamics is likely to be encoded with two symbols {0,1}. We begin construction of the partition by assigning symbols to the fixed point. The position of the generating partition for the attractor given by the logistic map is indicated in Fig. 1, where orbit points with periods up to 11 are colored according to their symbolic representation. It is obviously that the critical point 0.5 in this case is the generating partition for this logistic map.

By choosing this logistic map and the generating partition, we can guarantee that the chaos-based pseudorandom sequence has a high linear complexity. Set the initial condition x_0=0.1, the P-value [21] of the linear complexity is as high as 0.995.

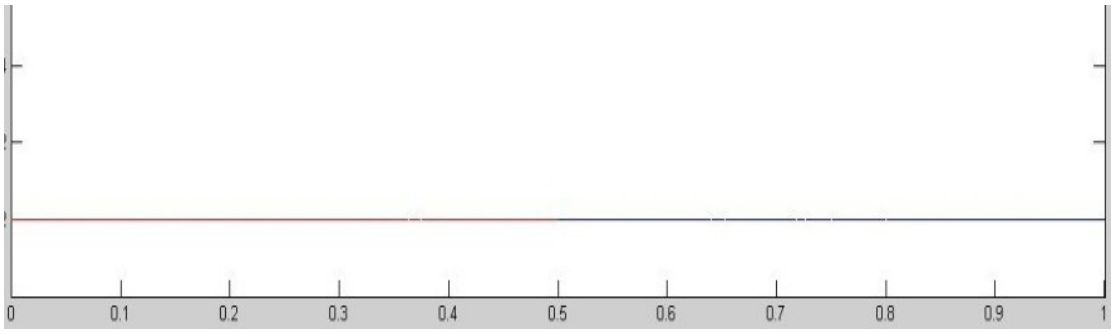

Fig.1. (Color) Orbit points up to period-11 for the logistic attractor colored according to their symbolic representation: red and blue dots represent orbit points encoded with symbols **0** and **1**, respectively.

3. CONNECTIONS OF NONLINEAR COMPLEXITY WITH SOURCE ENTROPY

In this section, we take Lempel-Ziv complexity as a bridge, reveal the connections of nonlinear complexity with source entropy.

In [22] a complexity measure for finite length sequences is proposed, called Lempel-Ziv complexity. We first recall the basic definitions from [22].

A sequence y^N is called reproducible by its proper prefix $y^j=y_0\ y_1...y_{j-1}$, $j<N$, if there exists an integer $p<j$ such that $y_{j+i}=y_{p+i}$ for all $0\le i<N-j$. It is denoted by $y^j \to y^N$. Furthermore, a sequence y^N is called producible by its proper prefix y^j if it is reproducible by y^j possibly with the exception of its last term and is denoted by $y^j \Rightarrow y^N$. A production process $S(y^N)$, or history, of sequence y^N is any partitioning

$$S(y^N) = y_0^{h_0} y_{h_0+1}^{h_1} ... y_{h_{s-1}+1}^{h_s}$$

Where $h_0=0$, $h_s=N-1$, $h_{i-1}< h_i$ and $y_0^{h_{i-1}} \Rightarrow y_0^{h_i}$, $1\le i\le s$. A tuple $y_{h_{i-1}+1}^{h_i}$ is called word of the history. A particular word $y_{h_{i-1}+1}^{h_i}$ is called exhaustive if $y_0^{h_i}$ cannot be reproducible by its prefix $y_0^{h_{i-1}}$. If all the words of a history are exhaustive (with a possible exception of the last one), the history is called exhaustive. Every sequence y^N has a unique exhaustive history $S_e(y^N)$. [22] proved that among all histories of a sequence its exhaustive history has the least number of words $LZ(y^N)$, referred to as the Lempel-Ziv complexity of y^N.

Theorem 4.1 [22]: For every $y^N \in F_\delta$ and positive ε,

$$\lim_{n\to\infty} \Pr\left[LZ(y^N) < \frac{hN(1-\varepsilon)}{\log_\delta N} \right] = 0$$

where N is the length of this sequence and h is the source entropy, $0\le h\le 1$.

Now we consider a chaotic system (X,T) with a partition α. For almost every initial conditions $x_0 \in X$, without a set of zero Lebesgue measure, the Lempel-Ziv complexity of the orbit sequence $T^*(x_0|\alpha)$

$$LZ(T^*(x_0 \mid \alpha)) \sim \frac{hN}{\log_\delta N}$$

for moderate large N, "~" means positive correlation [23]. Thus, we have $h \sim LZ/(N/\log_\delta N)$. Denote $b(N)=N/\log_\delta N$, normalize Lempel-Ziv complexity with $b(N)$ as

$D(N)=LZ/b(N)$

Then, the complexity of sequence can be measured with the normalized $D(N)$. $D(N)$ approaches 1 for true random sequence. $D(N)$ approaches 0 for periodic sequence. The relative complexity $D(N)$ describes the degree of a given sequence approaches to random one. If the $D(N)$ of a given sequence approaches to 1, the sequence approaches to random one.

Now we analyze the relationship between Lempel-Ziv complexity and nonlinear complexity. For a random binary sequence with length N, the expectation of the nonlinear complexity is $2\log_2 N$. For a pseudorandom sequence, in order to satisfy the security requirement of cryptography, we

should make its nonlinear complexity close to this value. Also, we can denote $g(N)=2\log_2 N$, normalize nonlinear complexity with $g(N)$ as
$G(N)=c/g(N)$

Then, the nonlinear complexity of sequence can be measured with the normalized $G(N)$. If $G(N)$ approaches to 1, then the nonlinear complexity of the given sequence approaches to $2\log_2 N$, which satisfies the security requirement of cryptography.

Without loss of generality, we take the following logistic map as an example.

$$x_{k+1} = ax_k(1-x_k)$$

We get our chaotic binary sequence by using the following algorithm.

$$x_k = \begin{cases} 0 & x_k \le 0.5 \\ 1 & x_k > 0.5 \end{cases}$$

Before analyzing the relationship between $D(N)$ and $G(N)$, we should first analyze the stability of $D(N)$ and $G(N)$, for both of them may vary with the increase of length N. Let a=4 and 3.6, initial value x_0=0.1, the results are shown in fig 2 and 3.

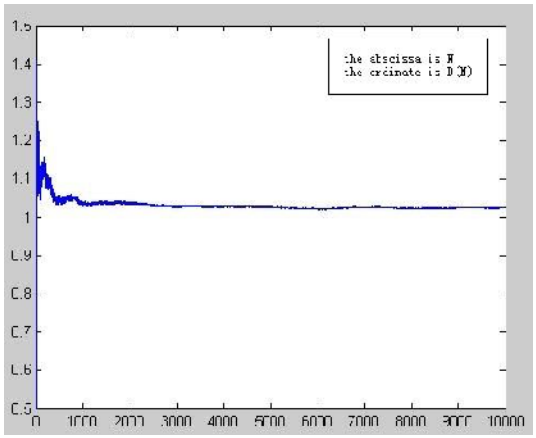

Fig.2. The stability of $D(N)$ when a=4

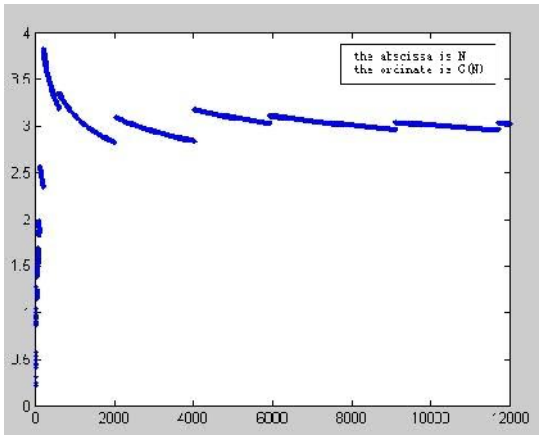

Fig.3. The stability of $G(N)$ when a=3.6

The selections of parameter are due to the extent consideration. For a=4, the Lempel-Ziv complexity of the binary sequence grows fastest in all parameters. Thus, the changes of $D(N)$ may be more apparent. For a=3.6, the nonlinear complexity of the binary sequence grows fast in most parameters, so the changes of $G(N)$ is also apparent. From these two figs we know, $D(N)$ and $G(N)$ may reach stable with the growth of the length of sequence.

Now we can analyze the relationship between $D(N)$ and $G(N)$. Set the initial value x_0=0.1, and the length N=10000. We choose 100 parameters a from the chaotic area [3.6, 4] uniformly. For each a, $D(N)$ and $G(N)$ are calculated separately. By using a tunable kernel function to fitting the data,

the relationship between $D(N)$ and $G(N)$ is shown in fig 4. In fig 4, we find that with the growth of $D(N)$, $G(N)$ gradually approaches to 1.

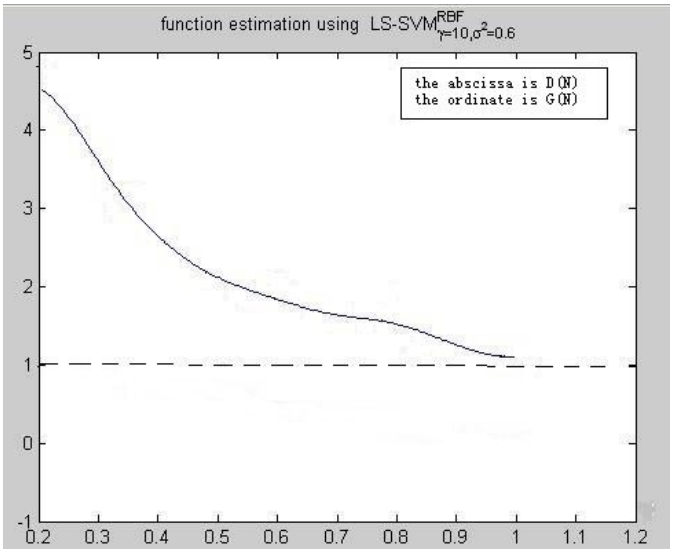

Fig.4. The relationship between normalized $D(N)$ and normalized $G(N)$

From the analysis above we know, in order to make the nonlinear complexity of the chaos-based pseudorandom sequence satisfy the security requirement of cryptography, we should make source entropy h as large as possible.

Let $\delta=m$, we can calculate the source entropy h as the following formula

$$h = \sum_{i=1}^{m} -p_i \log_m p_i$$

where p_i means the probability of the symbol exit in the sequence. From the formula, if $p_1=p_2=\ldots=p_m$, entropy h reaches the biggest value 1. Next, we take the following logistic map as an example to show how to make the distribution of the symbols in the sequence $T^*(x_0|\alpha)$ reach uniform.

$$x_{k+1} = -1 + 8x_k^2 - 8x_k^4$$

Theorem 4.2: Consider the chaotic map above, $\Omega=[-1,1]$ is its chaotic invariant set. The partition α divide the invariant set into m subsets: $[t_0, t_1)$, $[t_1, t_2),\ldots,$ $[t_{m-1}, t_m]$, where

$$t_i = -\cos(\frac{i}{m}\pi), \qquad i = 0,1,...,m$$

For almost every initial condition x_0, the distribution of the symbols in the orbit sequence $T^*(x_0|\alpha)$ reaches uniform.

Proof: This logistic map has the following probability density:

$$\rho(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad -1 \le x \le 1$$

The probability of the element fall into the set $[t_i, t_{i+1})$ is

$$\Pr(i) = \int_{t_i}^{t_{i+1}} \rho(x)dx = \frac{1}{m}$$

thus concluding our proof.

Therefore, for this logistic map, with the selection of a proper partition α analyzed above, can guarantee the nonlinear complexity of the chaos-based pseudorandom sequences satisfy the security requirement of cryptography.

4. CONCLUSION

This paper explores the relationships between linear complexity and measure entropy, nonlinear complexity and source entropy. We use these relations to select or construct a suitable chaotic system and partition to guarantee the linear complexity and the nonlinear complexity of chaos-based pseudorandom sequences satisfy the security requirement of cryptography. It is great helpful to make chaos-based cryptography consolidate as a real alternative. Connections of other cryptographic criteria with chaotic dynamical complexity measures remain an interesting open problem.

REFERENCES

[1] S. Rakesh, A. K. Ajitkumar, B. C. Shadakshari and B. Annappa, "Image Encryption using Block Based Uniform Scrambling and Chaotic Logistic Mapping", International Journal on Cryptography and Information Security (IJCIS), Vol. 2, No. 1, pp. 49-57, 2012.

[2] M. Ahmad, B. Alam and O. Farooq, "Chaos Based Mixed Keystream Generation For Voice Data Encryption", International Journal on Cryptography and Information Security (IJCIS), Vol. 2, No. 1, pp. 39-48, 2012.

[3] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, Handbook of Applied Cryptography. Boca Raton, FL: CRC, 1996.

[4] E. R. Berlekamp, Algebraic Coding Theory. New York: McGraw-Hill, 1968.

[5] J. L. Massey, "Shift register synthesis and BCH decoding," IEEE Trans. Inf. Theory, vol. 15, No. 1, pp. 122-127, 1969.

[6] J. L. Massey and S. Serconek, "A Fourier transform approach to the linear complexity of nonlinearly filtered sequences," in Adv. Cryprology-CRYPTO '94, Berlin, Germany, vol. 839, Lecture Notes in Computer Science, pp. 332-340, 1994.

[7] N. Kolokotronis and N. Kalouptsidis, "On the linear complexity of nonlinearly filtered PN-sequences," IEEE Trans. Inf. Theory, vol. 49, No. 11, pp. 3047-3059, 2003.

[8] K. Limniotis, N. Kolokotronis, and N. Kalouptsidis, "New results on the linear complexity of binary sequences," in Proc. IEEE Int. Symp. Inf. Theory, pp. 2003-2007, 2006.

[9] N. Kolokotronis, K. Limniotis, and N. Kalouptsidis, "Lower bounds on sequence complexity via generalized Vandermode dissertation, in Sequence and Their Application, Berlin Germany, vol. 4086, pp. 271-284, 2006.
[10] C. J. A. Jansen, "The maximum order complexity of sequence ensembles," in Adv. Cryptology-Eurocrypt '91, Berlin, Germany, vol. 547, Lectures Notes in Computer Science, pp. 153-159, 1991
[11] D. Erdmann and S. Murphy, "An approximate distribution for the maximum order complexity," Des. Codes. Cryptogr., Vol. 10, No. 3, pp. 325-339, 1997.
[12] P. Rizomiliotis and N. Kalouptsidis, "Results on the nonlinear span of binary ssequences," IEEE Trans. Inf. Theory, Vol. 51, No. 4, pp. 1555-1563, 2005.
[13] P. Rizomiliotis, N. Kolokotronis, and N. Kalouptsidis, "On the quadratic span of binary sequences," IEEE Trans. Inf. Theory, Vol. 51, No. 5, pp. 1840-1848, 2005.
[14] P. Rizomiliotis, "Constructing periodic binary sequences of maximum nonlinear span," IEEE Trans. Inf. Theory, Vol. 52, No. 9, pp. 4257-4261, 2006.
[15] K. Limniotis, N. Kolokotronis, and N. Kalouptsidis, "On the nonlinear complexity and Lempel-Ziv complexity of finite length sequences," IEEE Trans. Inf. Theory, Vol. 53, No. 11, pp. 4293-4302, 2007.
[16] S. M. Pincus, "Approximate entropy as a measure of system complexity," National. Acad. Science, Vol. 88, No. 6, pp. 2297-2301, 1991.
[17] S. M. Pincus, "Approximate entropy (ApEn) as a complexity measure," Chaos, Vol. 5, No. 1, 1995.
[18] T. Beth, Zong-Duo Dai, "On the complexity of pseudo-random sequences-or: If you can describe a sequence it can't be random," Advances in Cryptology—EUROCRYPT'89, 1990 - Springer
[19] A. A. Brudno, "Entropy and the complexity of the trajectories of a dynamical system," Trans. Moscow. Math, Soc. 2, 1983.
[20] R. L. Davidchack, Y. C. Lai, "Estimating generating partition of chaotic system by unstable periodic orbits," Phys Review E, 2000.
[21] A. Rukin, J. Soto, and J. Nechvatal, et al., "A Statistical Test Suite for Random and Pseudorandom Number generators for Cryptographic Applications," NIST Special Publication 800-22, 15.Mar.2001.
[22] A. Lempel and J. Ziv, "On the complexity of finite sequences," IEEE Trans. Inf. Theory, Vol. 22, No. 1, pp. 75-81, 1976.
[23] F. Kaspar, H. G. Schuster, "Easily calculable measure for the complexity of spatiotemporal patterns," Phys. Rev. A, Vol. 36, 1987.