New Methodology of Block Cipher Analysis Using Chaos Game

Budi Sulistyo, Budi Rahardjo, Dimitri Mahayana & Carmadi Machbub

School of Electrical Engineering and Informatics, Institut Teknologi Bandung,
Bandung 40132, Indonesia
Email: budi241@yahoo.com

Abstract. Block cipher analysis covers randomness analysis and cryptanalysis. This paper proposes a new method potentially used for randomness analysis and cryptanalysis. The method uses true random sequence concept as a reference for measuring randomness level of a random sequence. By using this concept, this paper defines bias which represents violation of a random sequence from true random sequence. In this paper, block cipher is treated as a mapping function of a discrete time dynamical system. The dynamical system framework is used to make the application of various analysis techniques developed in dynamical system field becomes possible. There are three main parts of the methodology presented in this paper: the dynamical system framework for block cipher analysis, a new chaos game scheme and an extended measure concept related to chaos game and fractal analysis. This paper also presents the general procedures of the proposed method, which includes: symbolic dynamic analysis of discrete dynamical system whose block cipher as its mapping function, random sequence construction, the random sequence usage as input of a chaos game scheme, output measurement of chaos game scheme using extended measure concept, analysis the result of the measurement. The analysis process and of a specific real or sample block cipher and the analysis result are beyond the scope of this paper.

Keywords: *block cipher; chaos game; cryptanalysis; measure; random sequence.*

1 Introduction

Block cipher analysis covers randomness analysis and cryptanalysis. Randomness analysis is a very important thing in cryptographic field. Randomness analysis has become an approach for testing security or strength of an encryption algorithm. In stream cipher or pseudo random number generator (PRNG) context the randomness analysis objective is to test whether a random sequence produced by the stream cipher or PRNG can be distinguished from a sequence produced by a true random sequence generator.

In block cipher context, the randomness analysis is more specifically used to identify the existence of regularities in the block cipher algorithm. Katos is doing such analysis by systematically altering 1-bit of input of the block cipher

Received October 8th, 2009, Revised September 29th, 2011, Accepted for publication October 17th, 2011.
 ISSN: 1978-3086, DOI: 10.5614/itbj.ict.2011.5.2.3

[1]. The failure of a block cipher in passing this test indicates the existence of some relationship between its input and output. However, the method proposed by Katos cannot revealed the form of the relation explicitly. Hernandez suggested that a block cipher randomness testing cannot be performed by using a random input data (*high-entropy-feeding*) [2]. In his paper, he proposes a technique called low-entropy-feeding. In this technique, some bit of the input is fixed.

A block cipher randomness analysis technique can be potentially developed further to be a cryptanalysis method. In cryptanalysis, the randomness analysis technique will be used to identify the round subkeys. Because there are a set of subkeys consists of one right-subkey and several wrongsubkeys, the cryptanalysis method should distinguish the rightsubkeys from the other subkeys.

This paper proposes a new framework and methodology for block cipher analysis using chaos game. The proposed method consists of a dynamical system approach for analysis, a new chaos game scheme and extended measure concept related to the chaos game scheme. A discrete time dynamical system which created by using block cipher as a mapping function will produce a random sequence. The random sequence, then, will drive a chaos game scheme. The measure concept will be used to measure the output of the chaos game scheme and then the measurement result will be analyzed further to distinguish a block cipher from random permutation function or to distinguish the right subkey from the other wrong subkeys.

To explain some basic concepts related to the proposed method, this section will also discuss chaos game, dynamical system approach for randomness analysis, iteration of quadratic functions and iteration of an SPN block cipher.

1.1 Chaos Game

It is hard to explain chaos game without describe fractal in the first. Fractal is a geometric object which can be divided infinitely into smaller part that is identical to the larger part (Figure 1 is an example of fractal object). This characteristic is called self-similarity. The terminology of fractal originally is proposed by Benoit Mandelbrot. From early nineteen century until now, the research about fractal is a rich and fast moving subject. Mathematical fractal is a fractal constructed by a mathematical equation which inhere feedback and recursive process.

In mathematic field, the term chaos game, as explained by Michael Barnsley [3], is referred to a scheme for fractal image generation which uses a polygon to

create a dynamical system and random sequence as the input of the system. Fractal is generated by plotting a point proportionally between the former point and one corner of the polygon chosen by random sequence. This fractal generation process is an iterative process driven by a random sequence. As example, if the iterative process uses triangle (as polygon) and \frac ½ proportional constant, then the process will produce Sierpinski triangle.

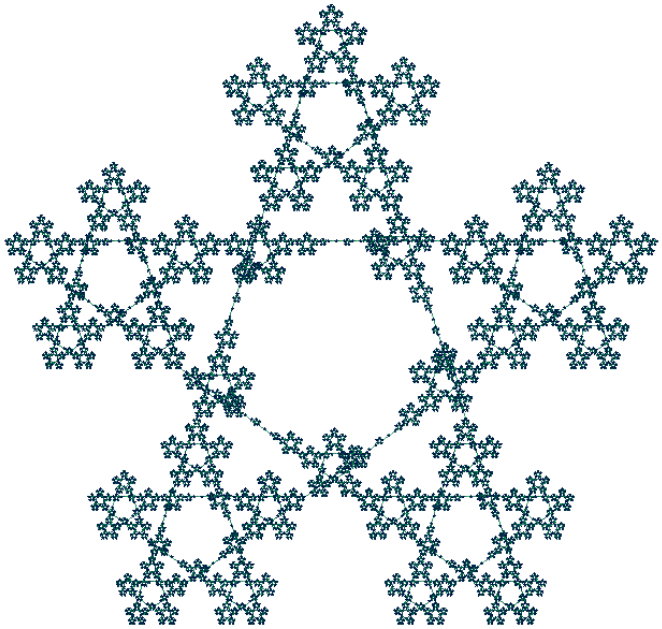

Figure 1 A fractal example: pentagram. (Source: http://en.wikipedia.org/wiki/Fractal).

One of the most important applications of chaos game in the other field outside mathematics is as a DNA analytical tool which called chaos game representation (CGR) [4],[5],[6] Jeffrey uses CGR to reveal some important characteristics of Human Beta Globin. An example of CGR image can be seen in Figure 2. CGR method uses four symbol which each represented four kinds neucleotids, they are A, G, C and T.

This paper proposes chaos game as a tool for block cipher analysis. The main idea is to measure characteristic of the sequence produced by the block cipher using chaos game scheme and extended measure concept. The chaos game scheme proposed in this paper is especially designed for analyzing random sequence consists of n-bit binary number as its term. If the random sequence analyzed is very similar to true random sequence (defined later), then the

scheme will produce a uniformly distributed points along a rectangle. The rectangle is essentially the attractor of the proposed chaos game scheme. The measure concept is used to measure violation of the fractal image produced by a tested random sequence from the one produced by true random sequence. The method proposed in this paper may be potentially used as a distinguisher and then may be developed later as a cryptanalytic method.

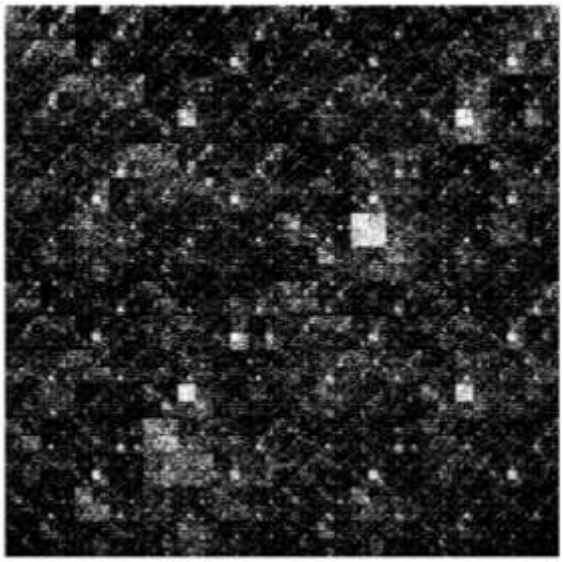

Figure 2 Chaos Game Representation (CGR) of genetic stream of E. coli K12 bacteria. (source:http://insilico.ehu.es/oligoweb/info/CGR.php?ZCGR).

1.2 Dynamical System Approach for Block Cipher Analysis

This section explains the dynamical system framework for block cipher analysis. The objective of system dynamic approach is in order to make the application of analytical tool developed in dynamical system field becomes possible. These are the proposed framework for block cipher analysis:

1. Block cipher is treated as mapping function of a discrete time dynamical system which is mathematically modeled as a difference equation. In this framework, the execution of the system is identical to block cipher iteration. This iteration will produce the orbit or trajectory of the system from an initial condition.
2. Next, the orbit of the system is analyzed by symbolic dynamic method [7],[8]In this method, domain of the system will be divided into several parts and each part is denoted by a distinct symbol. The analysis of the system dynamic is performed by analyze the sequence of symbols produced by the orbit or trajectory.

Definisi 1 [7],[9] *This is a mathematical model of a discrete time dynamical system*:

$$x\,(n+1) = f_k\,(x(n)) \tag{1}$$

with $x \in \mathbb{R}^m$ *and* $f_k : S \in \mathbb{R}^m \rightarrow \mathbb{R}^m$ *is a nonlinear function with* k *is the parameter of the mapping function. S denotes the domain of* f_k. *The sequence of* x *i.e.* $\{x(0), x(1), x(2), \ldots\} = \{x(0), f_k(x(0)), f_k(x(0))), \ldots\}$

1.3 Iteration of Quadratic Function

If iteration of a mapping function produces chaos dynamic, then the mapping function is a chaotic map. This subsection will explain iteration of a quadratic mapping function. This is a well known function in many literatures of dynamical system and chaos. The dynamic of the quadratic mapping function could be either asymptotic, periodic or chaotic depends on the value of its parameter. Generally, the quadratic function is represented as $Q_\mu(x) = \mu x - \mu x2$ with domain $S_Q = \{x/0 \leq x \leq 1, x \in \mathcal{R}\}$. The value of μ must satisfy $0 \leq \mu \leq 4$ to make Q_μ maps S_Q onto S_Q. As examples, the followings are iteration of two quadratic functions with $\mu = 3$ and $\mu = 4$.

1.3.1 Quadratic Function with μ = 3

Iteration of a quadratic function can be modeled as follows:

$$x\,(n+1) = \mu x_n - \mu x_n^2 \tag{2}$$

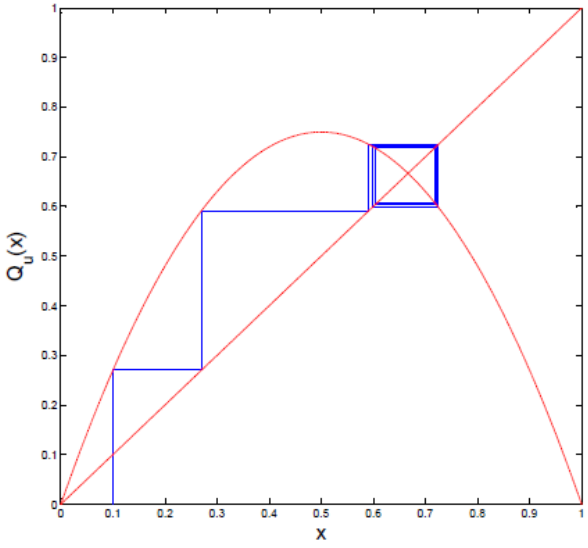

Figure 3 Cobweb plot of $Q_3(x)$ for x (0) = 0:1.

For $0 \leq \mu \leq 3$, Q_μ has only one attracting fixed point. For $3 < \mu \leq 1 + \sqrt{6}$, Q_μ has attracting two cycle periodic points [7]. Theoretically, for $\mu = 3$ the system orbit will converge to an attracting point. Suppose $x(0) = 0{:}1$ as initial point and set $\mu = 3$, by Eq. (2), we get the system orbit as follows:

0.1000, 0.2700, 0.5913, 0.7250, 0.5981, 0.7211, 0.6033, 0.7180, 0.6075, . . .

Figure 3 shows the cobweb plot of $Q_3(x)$:

1.3.2 Quadratic Function with $\mu = 4$

Suppose $x(0) = 0{:}1$ is an initial point and $\mu = 4$, then we get the orbit of $Q\mu$ as follows:

0:1000, 0:3600, 0:9216, 0:2890, 0:8219, 0:5854, 0:9708, 0:1133, 0:4020, . . .

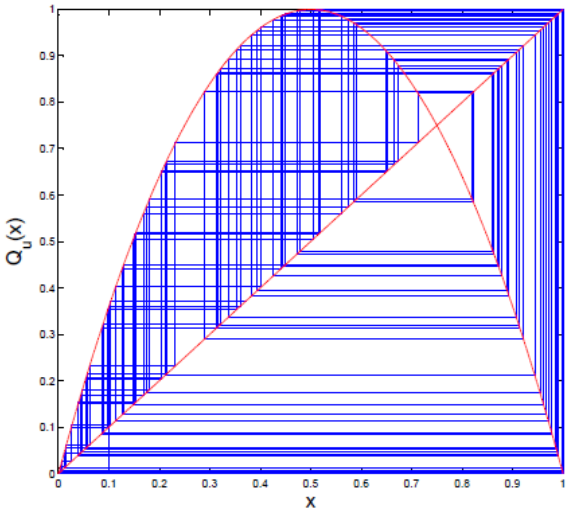

Figure 4 Graphical analysis of orbit of x for $\mu = 4$.

For $\mu = 4$, the orbit of Q_μ is neither periodic nor convergent to a point. The orbit seems to fill all of its domain. Figure shows the cobweb plot of Q_4.

1.4 Iteration of an SPN Block Cipher

Now, we will iterate a block cipher as shown in Figure 5. It is an SPN block cipher which uses substitution and permutation as its building blocks.

The SPN processes data block in four rounds. Each round implements three basic operations, they are (1) substitution, that is a mapping from a set of binary number to its self usually using a lookup table, (2) permutation, that is a interchange of bit position, and (3) key mixing that is using a key to XORed

with the data in a round. The block cipher has 16-bit data block and 32-bit key length. All parameters of the block cipher are identical to the one described in [10],[11]. The followings will explain iteration on E_k, i.e. mapping function formed by a block cipher and how to analyze the output of the iteration by using *symbolic dynamic*.

1.4.1 Iteration of E_k

Due to its 16-bit block size, the cipher domain can be presented as a set as follows:

$$SE = \{0, 1, 2, \ldots, 2^{16} - 1\}$$

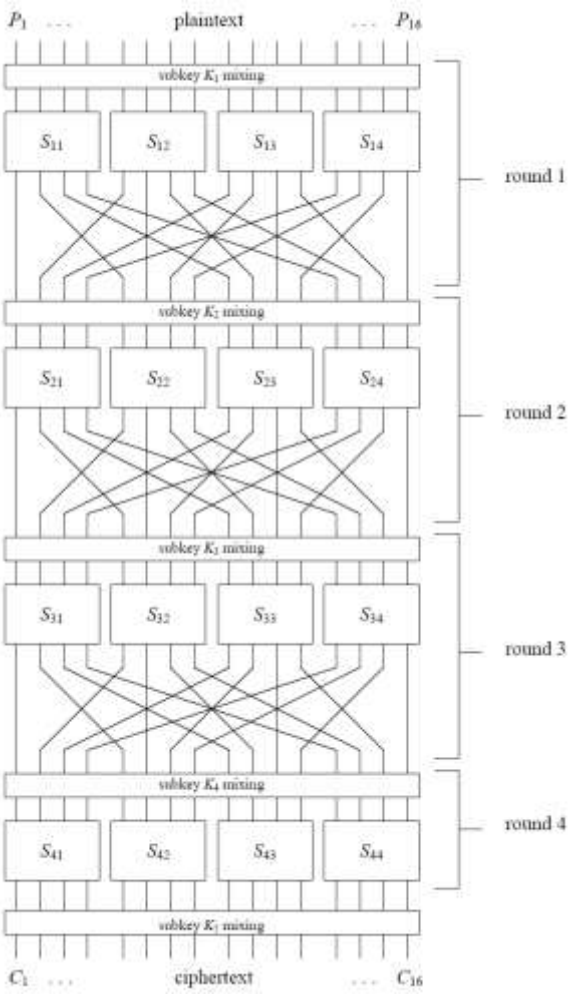

Figure 5 Substitution and permutation network.

Encryption function with key k is denoted by E_k. The key length is 32-bit and due to the key length, the set of all keys can be presented as follows:

$$\mathcal{K} = \{ 0, 1, 2, \ldots, 2^{32} - 1\}$$

For each value of k, E_k is a mapping function which maps the set S_E to itself. The mapping function is essentially a permutation function. Based on characteristics of block cipher, the picking of a key k is identical to a random selection process of one permutation function from a set of permutation functions.

In the followings, we choose a key k in random as a sample. key k in random as a sample. The plot of $E_k(p)$ for

$$k = 11010111001110101000110101101110,$$

and for an initial point which is chosen randomly from a set $S_E = \{ 0, 1, 2, \ldots, 2^{16} - 1\}$ is shown in Figure 6.

By looking at the Figure 6 it is clear that for any two close point *p1* and *p2*, the encryption mapping function will reveal E_k (*p1*) and E_k (*p2*) which is not always close from each other.

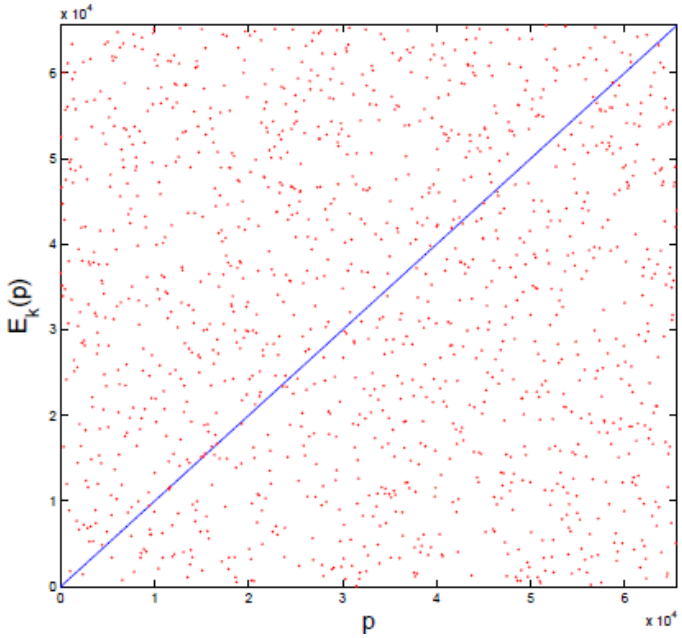

Figure 6 The plot of E_k (*p*) on *p*.

1.4.2 Analysis of the Output by Using Symbolic Dynamic

The sequence generated by the block cipher will be analyzed further. First, divide the mapping domain into two parts and denotes each part with a distinct symbol. This analysis will use two symbol alternatives $m = 2$. The sequence $s_E = k_1k_2 \ldots$ is constructed by using these following equations:

$$k_n = \begin{cases} L & \text{if } 0 \le p_{dec}(n) \le 2^{15} - 1 \\ R & \text{if } 2^{15} \le p_{dec}(n) \le 2^{16} \end{cases} \tag{3}$$

This domain division into two parts as mathematically presented by Eq. (3) is identical to domain division by referring to one of the bit of the block data. As example, we can use the most significant bit of the block data to divide the domain into two parts.

This can be presented mathematically as follows:

$$k_n = \begin{cases} L & \text{if the msb of } p(n) = 0 \\ R & \text{if the msb of } p(n) = 1 \end{cases} \tag{4}$$

Table 1 Random test of three symbols for E_k with episodes amount $Nr = 4$

Symbol	fr	Ideal Probability	bias
LLL	0.1375	0.125	0.0125
LLR	0.1267	0.125	0.0017
LRL	0.1214	0.125	-0.0036
LRR	0.1251	0.125	0.0001
RLL	0.1265	0.125	0.0015
RLR	0.1198	0.125	-0.0052
RRL	0.1248	0.125	-0.0002
RRR	0.1182	0.125	-0.0008

The iteration is performed by choosing one initial condition randomly, that is

p (0) = 0110111001010001

and the encryption key

K = 11010111001110101000110101101110

The sequence generated by s_{Ek} is as follows

s_{Ek} = LLLRLRLRLLRLLLRRLLRLR …

Here, we will perform a 3-term testing to the sequence generated by the block cipher. It means that we use all possible alternatives 3-terms sequence of symbols, Δ^3. The result can be seen in Table 1.

2 The Proposed Chaos Game Method for Randomness Analysis and Criptanalysis

This section explains the proposed method for randomness analysis and cryptanalysis. The first part explains a theorem (Theorem 1) which reveals the probability of occurrence of a certain finite length sequence along a true random sequence. This probability value of occurrence is an ideal value, as we supposed true random sequence in the theorem. The value then will be used as a center by which we can calculate bias of actual relative frequency of a generated sequence from a true random sequence. This theorem gives basic for the theorem discussed in the third part.

The second part explains the proposed chaos game algorithm (Eq. (9)) used in the analysis method. To generate an attractor, this algorithm needs a random sequence as input. Suppose we use a true random sequence as input then the chaos game algorithm will produce points that uniformly distributed along a chaos game attractor.

The third part explains metric or measure used to measure the output of the chaos game scheme. Theorem 2 reveals distribution measure of points generated by a chaos game algorithm if a true random sequence is used as input of the chaos game algorithm. Finally, Eq. (14) can be used to calculate distribution measure bias of the actual random sequence from the true random sequence.

2.1 True Random Sequence

Let s = l_0, l_1, l_2 , ... l_N denotes a random sequence whose $N + 1$ number of terms. Each term represents a certain symbol picked from M number of symbols alternatives or $l_n \in \{\alpha_1, \alpha_2, ... \alpha_M\}$ for $n = 0, 1, 2, ..., N$. Probability value for each symbol is denoted by $\{p_{\alpha 1}, p_{\alpha 2}, ... , p_{\alpha M}\}$. Let $\Delta^K = \delta_1, \delta_2, \ldots, \delta_K$ denotes a sequence with $K < N$ number of terms whose the same alternatives symbols as previously described, $\delta_K \in \{\alpha_1, \alpha_2, ... \alpha_M\}$ for $k = 1, \ldots, K$. Pick an integer I and point a subsequence of s whose K terms of length start at index I, then we points at this subsequence:

$$s = l_0 l_1 \ldots \underbrace{l_i l_{(i+1)} \ldots l_{(i+K-2)} l_{(i+K-1)}}_{s_i^{(i+K-1)}} \ldots l_{(N-1)} l_N$$

For some index *I*, an event $s_i^{(i+K-1)} = \Delta^K$ has some probabilistic value to occur.

True random sequence will be defined as follows:

Definition 2 *Let $s = l_0, l_1, l_2, \ldots l_N$ denotes a random sequence with l_n is the n-th term of s and $l_n \in \{\alpha_1, \alpha_2, \ldots \alpha_M\}$ for n = 0, 1, 2, …, N. The random sequence s is a true random sequence (denoted by s_{true}) if for each index n, each symbol$\{\alpha_1, \alpha_2, \ldots \alpha_M\}$is equally probable to occur in l_n and the occurrence of an alternative symbol in l_{n1} is independent from the occurrence of an alternative symbol in l_{n2} for any n1 ≠ n2. Because the occurrence of symbols in l_n for any n is equally probable, then $p_{\alpha 1} = p_{\alpha 2} = \ldots = p_{\alpha M} = \frac{1}{M}$ for each l_n.*

We formulate the following problem statement for s_{true}:

> For each index *i* which randomly chosen, what is the probability of event $s_{ideal}^{(i+K-1)} = \Delta^K$ or in other words what is the probability of occurrence of certain sequence of symbols with *K*-terms length in s_{true} starting at *i*-th index which is chosen randomly?

For s_{true} case, because the *i*-th index is randomly chosen, the probability of occurrence of δ_1 at l_i is $p_{\delta 1}$ or $Pr\{l_i = \delta_1\} = p_{\delta 1}$. Also in this case $p_{\alpha 1} = p_{\alpha 2} = \ldots = p_{\alpha M} = \frac{1}{M'}$. Next, based on the independence of occurrence of symbols at di erent terms of s_{true}, the probability of occurrence of δ_2 at $l_i + 1$ is $p_{2=}\frac{1}{M}$, and still because the independence, we get $\Pr\{l_i l_{i+1} = \delta_1\delta_2\} = (\frac{1}{M})^2$. We can repeat the same argument until the (*i + K - 1*)-th index. Now, we can conclude that for each *i*-th index which is chosen randomly, the probability of the occurrence of symbol Δ^K in $s_{truei}^{\ i+K-1}$ is $(1/M)^K$ or $\Pr\{s_{true_i}^{(i+K-1)} = \Delta^K\} = \left(\frac{1}{M}\right)^K$.

The above explanation will be restated by the following lemma:

Lemma 1 Let $s_{true} = l_0 \ldots l_N$ denotes a true random sequence as defined in definition 2 and $\Delta^K = \delta_1 \ldots \delta^K$ denotes a sequence with $K \leq (N + 1)$. The probability of occurrence of in Δ^K *in* $s_{true\ i}^{(i+K-1)}$, with index *i* is chosen randomly, is

$$\Pr\{s_{ideal_i}{}^{(i+K-1)} = \Delta^K\} = \left(\frac{1}{M}\right)^K \quad (5)$$

Proof. The proof is obvious by the above explanation.

Based on Lemma 1, if we choose i randomly, then the probability of occurrence of Δ^K in s_{true} starting from the *i*-th is $\left(\frac{1}{M}\right)^K$.

Definition 3 *Let* $s = l_0 l_1 \ldots l_N$ *denotes a random sequence with* $l_n \in \{\alpha_1, \alpha_2, \ldots \alpha_M\}$ *for n = 0, 1, 2, ..., N. Relative frequency of occurrence a sequence* Δ^K *in a random sequence s is the expected frequency of occurrence of* Δ^K *along random sequence s divided by number of subsequence with K-terms length along the random sequence s.*

The following is a theorem on relative frequency of occurrence of a certain sequence of symbol along a true random sequence s_{true}:

Theorem 1 *The* relative frequency *of occurrence of a certain sequence of symbol* Δ^K *along a true random sequence* s_{true} *defined in Definition 2 is* $\left(\frac{1}{M}\right)^K$.

Proof. For true random sequence $s_{true} = l_0 l_1 \ldots l_N$, there are (N - K + 2)-alternatives of position for occurrence of a sequence Δ^K. These alternatives of position is also the alternatives of index *i* i.e *i* = 0, 1, …, (*N* - *K* + 1) where any occurrence of Δ^K starts. Because there are (*N* - *K* + 2) alternatives of position and because, based on Lemma 1, for any alternatives of position which is chosen randomly, the probability of occurrence of Δ^K is $\left(\frac{1}{M}\right)^K$, then the expected frequency of occurence of Δ^K along (*N* - *K* + 2)-alternatives of position is (*N* - *K* + 2) x $\left(\frac{1}{M}\right)^K$. Then, the relative frequency of occurrence of a sequence Δ^K in a true random sequence s_{true} is

$$\frac{(N-K+2)\times(1/M)^K}{(N-K+2)} = \left(\frac{1}{M}\right)^K \quad (6).$$

We can calculate directly the actual relative frequency of occurrence of sequence Δ^K along a sequence Θ. The following is the definition of actual relative frequency of occurrence of a sequence Δ^K along a sequence Θ.

Definition 4 *Let $\Theta = \theta_0, \theta_{1,\ldots,} \theta_N$ denotes a sequence which is a realization of a random sequence s as previously explained and $\Delta^K = \delta_1\delta_2\ldots \delta_K$ denotes a sequence with K-terms length which satisfies $\frac{M^K}{N+1} \ll 1$. The number of occurrence of Δ^K along Θ is denoted by $[\Delta^K]_{[\Theta]}$. The actual relative frequency of occurrence of a sequence Δk in a sequence Θ is*

$$Fr_{actual}\left\{occurrence\ of\ \Delta^K\ along\ \Theta\right\} = \frac{\left[\Delta^K\right]_{[\Theta]}}{N-K+2} \tag{6}$$

The following is a definition of bias of actual relative frequency of occurrence of a sequence Δ^K along a sequence Θ from the relative frequency of occurrence of Δ^K along a true random sequence s_{true}.

Definition 5 *The bias of actual relative frequency of occurrence of a sequence along a sequence $\Delta^K = \delta_1\delta_2\ldots \delta_K$* along a sequence $\Theta = \theta_0, \theta_{1,\ldots,} \theta_N$ *which is a realization of a random sequence s is defined by this following equation*:

$$\begin{aligned} &bias\left\{actual\ relative\ frequency\ of\ occurrence\ \Delta^K\ along\ \Theta\right\} \\ &= Fr_{actual}\left\{occurrence\ of\ \Delta^K\ along\ \Theta\right\} \\ &\quad - Fr\left\{occurrence\ of\ \Delta^K\ along\ s_{true}\right\} \end{aligned} \tag{7}$$

with the length of Δ^K is K-terms length which satisfies $\frac{M^K}{N+1} \ll 1$.

Based on the Theorem 1, then we get

$$\begin{aligned} &bias\left\{actual\ relative\ frequency\ of\ occurrence\ \Delta^K\ along\ \Theta\right\} \\ &= Fr_{actual}\left\{occurrence\ of\ \Delta^K\ along\ \Theta\right\} - \left(\frac{1}{M}\right)^K \end{aligned} \tag{8}$$

All analysis of random sequence in the proposed method will always refer to true random sequence concept that has been explained in this section.

2.2 New Chaos Game Scheme for Bock Cipher Analysis

The scheme uses a rectangle *F* as a set where all points generated by the scheme lay down. This *F* is essentially the attractor of the scheme. This scheme is different from sierpinski triangle scheme and the scheme used to generate CGR due to its number of symbols which are independent from number of corner point of the rectangle *F* This scheme is very flexible in the sense that we could use different number of symbols without changing the form of attractor *F* The dynamic of the chaos game scheme can be represented mathematically as follows:

$$\begin{bmatrix} x_1(i+1) \\ x_2(i+1) \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} . \begin{bmatrix} x_1(i) \\ x_2(i) \end{bmatrix} + length. \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} . \begin{bmatrix} \left(l_{i+1} \bmod 2^{p_1}\right). \\ \left\lfloor \frac{l_{i+1}}{2^{p_1}} \right\rfloor \end{bmatrix} \quad (9)$$

with

- $i = 0; 1; 2; : : :$ is an index which denotes the iteration step,
- $\vec{x}(i) = \begin{bmatrix} x_1(i) & x_2(i) \end{bmatrix}^T$ is a variable state which denotes points coordinate,
- $\vec{x}(0) = \begin{bmatrix} x_1(0) & x_2(0) \end{bmatrix}^T$ is the initial condition of the system,
- the input of the system is a sequence $s = l_1, l_2, \ldots,$
- n is number of bits analyzed by the scheme,
- $m = 2n$ is number of symbols used in the scheme,
- l_i is a term of a sequence s with $l_i \in \{0, 1, 2, \ldots, m\text{-}1\}$ for $i = 1, 2, \ldots$
- $\begin{bmatrix} c_1 & c_2 \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{2}\right)^{p_1} & \left(\frac{1}{2}\right)^{p_2} \end{bmatrix}$,
- $p_1 = \left\lfloor \frac{n}{2} \right\rfloor$
- $p_2 = n - p_1$,
- *length* is the length of side of F
- the area of F can be expressed as follows $F = \{(x,y) \in \mathbb{R}^2\}$ with $k_1 \le x_1 \le (k_1 + length)$ and $k_2 \le x_2 \le \{k_2 + length\}$

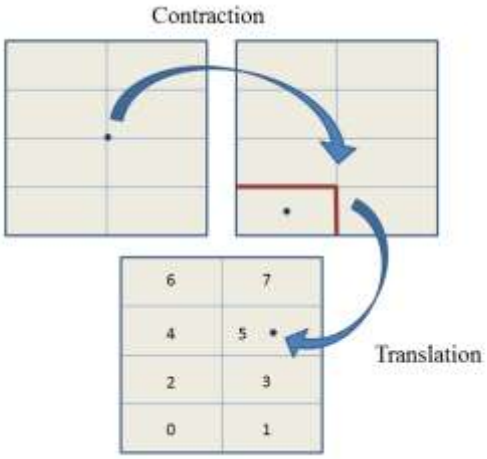

Figure 7 The graphical illustration of chaos game scheme to analyze n = 3-bits data by using $m = 2^3$ symbols.

Next, we denote the set of all points generated by the chaos game scheme until N-th iteration by $X_N = \{\vec{x}(0), \vec{x}(1), \ldots, \vec{x}(N)\} \in F$. For each iteration step, this scheme uses a contraction mapping and m translation mappings. The scheme is illustrated in Figure 7. The set X_N will be analyzed further by the extended measure concept described in the next section.

2.3 The Measure on X_N and Its Bias

Let B_i with $i = 1, 2, \ldots, n_{rect}$ denote small rectangle areas in rectangle F. $B_i \cap B_j = \emptyset$- for any $i \neq j$ and $B_1 \cup B_2 \cup \ldots B_{n_{rect}} = F$. All B_i-s have the same size. Here is a definition of measure on X_N with Bi, $i = 1, \ldots, n_{rect}$ are the measured areas.

Definition 6 *Let $X_N = \{\vec{x}(0), \vec{x}(1), \ldots, \vec{x}(N)\}$ denotes points generated by the chaos game scheme and $\mathfrak{N}(B_i, X_N)$ denotes number of X_N -points which lie within the area Bi. Measure on $X_N = \{\vec{x}(0), \vec{x}(1), \ldots, \vec{x}(N)\}$ by using $B_i = 1, 2, \ldots, n_{rect}$ is defined as follows*

$$\mu(B_{i=1}, \ldots, n_{rect}, X_N) = \begin{bmatrix} \frac{\mathfrak{N}(B_1, X_N)}{N+1} \\ \frac{\mathfrak{N}(B_2, X_N)}{N+1} \\ \vdots \\ \frac{\mathfrak{N}(B_{n_{rect}}, X_N)}{N+1} \end{bmatrix} \tag{10}$$

The following theorem explains bias of measure of X_N generated by a chaos game scheme driven by a random sequence s from X_N^{true} generated by a chaos game scheme driven by s_{true}.

Theorem 2 *Let X_N^{true} denotes the output of chaos game scheme driven by s_{true} with m symbols, F is a rectangle area related to the chaos game scheme, $B_i = 1, 2, \ldots, n_{rect}$ is a set of small rectangle with the same size, with $B_i \cap B_j = \emptyset$ for any $i \neq j$ and $B_1 \cup B_2 \cup \ldots B_{n_{rect}} = F$. All these Bi-s are used to measure an output of chaos game scheme. n_{rect} is a square number which denotes the number of those small rectangles. Then we have*

$$\mu(B_i, X_N^{ideal}) = \frac{size\ B_i}{size\ F} = \frac{1}{n_{rect}} \tag{11}$$

Proof. Based on Theorem 1 the relative frequency of occurrence of K-terms length for a true random sequence with m-symbols is $(1/m)k$ (for each K-terms length of sequence of symbols denoted by Δ^K). Based on the explanation in the

[3] and [8], each combination of *K*-terms length sequence of symbol expresses an address of an area within the rectangle *F*. Because the attractor of the proposed chaos game scheme is a *non-touching* fractal and identical to *F*, then each *K*-terms length sequence of symbol will express a unique area, given K is a finite integer. For s_{true} , because each alternatives of sequence of symbols will have the same relative frequency of occurrence on s_{true}, we have the same relative frequency for X_N to occur in all parts of *F* In other words, X_N will distribute uniformly within *F* Then for each $B_i, \mu\left(B_i, X_N{}^{ideal}\right)$ will close to ratio (size of B_i)/(size of *F*).

Bias of the set of points X_N from $X_N{}^{true}$, by using $B_i = 1, \ldots, n_{rect}$, can be expressed as follows:

$$\begin{aligned} bias\left(X_N, B_{i=1}\ldots, n_{rect}\right) &= \mu\left(B_{i=1},\ldots, n_{rect}, X_N\right) \\ &\quad - \mu\left(B_{i=1},\ldots, n_{rect}, X_N{}^{ideal}\right) \\ &= \begin{bmatrix} \frac{\mathfrak{N}(B_1,N)}{N+1} \\ \frac{\mathfrak{N}(B_2,N)}{N+1} \\ \vdots \\ \frac{\mathfrak{N}\left(B_{n_{rect}},N\right)}{N+1} \end{bmatrix} - \begin{bmatrix} \frac{1}{n_{rect}} \\ \frac{1}{n_{rect}} \\ \vdots \\ \frac{1}{n_{rect}} \end{bmatrix} \end{aligned} \tag{12}$$

We also define the normalization of the previous bias as follows:

$$\text{bias}_{norm}\left(X_N, B_{i=1},\ldots, n_{rect}\right) = \frac{\text{bias}\left(X_N, B_{i=1},\ldots, n_{rect}\right)}{1/n_{rect}} \tag{13}$$

The objective of defining the normalized bias is to avoid the miss interpretation of small bias value which caused by small $1 = n_{rect}$. Next, to calculate the total normalized bias, we use RMS as follows:

$$\text{RMS}\left(\text{bias}_{norm}\left(X_N, B_{i=1},\ldots, n_{rect}\right)\right) = \sqrt{\frac{\text{bias}_{norm}\left(X_N, B_{i=1},\ldots, n_{rect}\right)^T .\text{bias}_{norm}\left(X_N, B_{i=1},\ldots, n_{rect}\right)}{n_{rect}}} \tag{14}$$

3 General Procedure for Randomness Analysis and Cryptanalysis

The proposed method can be used to perform randomness analysis and cryptanalysis to a block cipher. Applicability of a method for randomness analysis is a necessary condition for the method to be used further for cryptanalysis. By randomness analysis method proposed in this paper, we need to iterate the cipher several times and, choose some bit position based on some consideration and pick the stream of output to construct a random sequence. Figure 8 explains how to generate data being analyzed by this method.

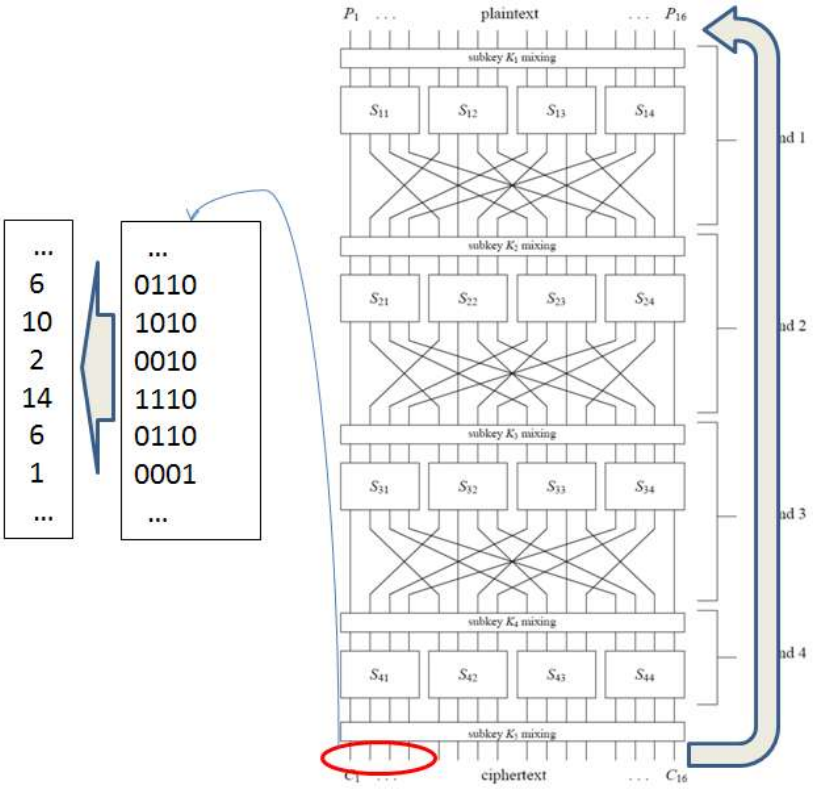

Figure 8 Illustration of how to generate data in randomness analysis by using only 4-bit msb of an SPN Block Cipher.

The strategy used to attack or cryptanalyze an SPN block cipher is to recover its subkeys layer by layer. This strategy is similar to the one used in linear and differential cryptanalysis [10]. In the proposed method we also need to generate one or several sequences, by choosing some bit output of the cipher, to be

analyzed in order to break the corresponding subkeys. Figure 9 explains how to generate data being used by the proposed cryptanalysis method.

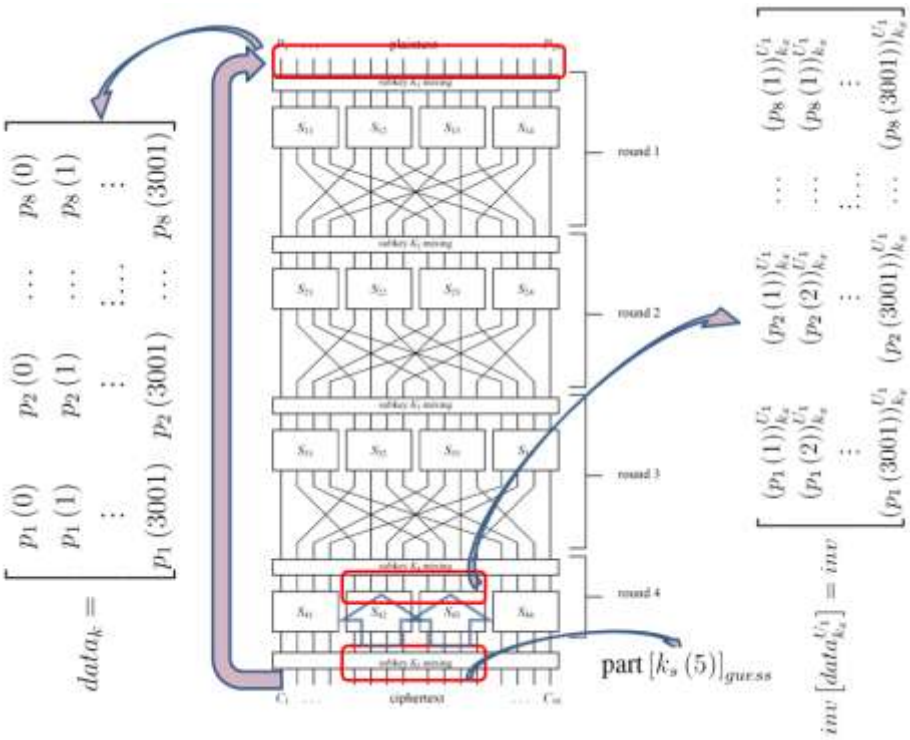

Figure 9 Illustration of how to generate data being used for block cipher cryptanalysis. The matrix which is used to arrange plaintext-plaintext data pair generated by iterating the cipher block with key *k* is data *k*. The matrix which is used to arrange last round inverse data is $\text{inv}\left[data_{k_x}^{U_1}\right]$.

To summarize all the process, here is a procedure for randomness analysis and cryptanalysis using concepts and algorithms explained above:

1. Iterate the block cipher to generate two streams or sequence, i.e. plain-text and cipher-text pair. The iteration must be initialized by randomly chosen initial conditions.
2. Perform symbolic analysis to the output of the iteration. To perform symbolic analysis, we must divide the mapping domain into several definite parts. Each parts will be denoted by a unique symbol. This step will produce a sequence s which each term express a certain symbol. We may use integer number as a symbol.
3. Use the sequence s as the input of the chaos game scheme.

4. Quantify the output of the chaos game scheme by using the proposed metric or measure. The objective of using this metric is to measure the density of points and to compute the bias from ideal density.
5. In case ciphers block randomness analysis, use the metric result to distinguish randomness of a cipher block from another cipher block. In case ciphers block cryptanalysis, use the metric to distinguish a true key from a set of all possible key.

4 Conclusions

The paper describes a proposed method used for randomness analysis and cryptanalysis. The analysis of a block cipher can be performed by treating the block cipher as a mapping function in a discrete time dynamical system. Next, the symbolic dynamic method is used to generate a random sequence whose m number of alternatives symbols for each term. The basic process used in the proposed method is to measure the violation of the sequence generated by the cipher from a true random sequence.

The proposed method is developed mainly to attack an SPN block cipher. Principally, the method also can be applied to attack any block cipher composed by several simple functions or rounds arrange in a layered structure.

In the next research, a real attacking experiment is very important to verify the applicability of the method against a simplified block cipher as used as example in this paper or against a real standard block cipher, for example DES, AES Serpent.

References

[1] Katos, V., *A Randomness Test for Block Ciphers*, Science Direct: Applied Mathematics and Computation, **162**, pp.29-35, 2005.

[2] Hernandez, J.C., Isasi, P., Sierra, J.M. & Tablas, A.G., *How to Distinguish Between A Block Cipher and A Random Permutation By Lowering the Input Entropy*, in IEEE 35th International Carnahan Conference on Security Technology, 2001.

[3] Barnsley, M, *Fractals Everywhere*, Academic Press Professional, Inc., San Diego, CA, USA, 1988.

[4] Jeffrey, H.J., *Chaos Game Visualization of Sequences*, Computer & Graphics, **16**(1), pp. 25-33, 1992.

[5] Jeffrey, H.J., *Chaos Games Representation of Genetic Sequences*, Nucleic Acids Research, **18**(8), pp. 2163-2170, 1990.

[6] Yang, J-Y., Yu, Z-G. & Anh, V., *Protein Structure Classification Based on Chaos Game Representation and Multifractal Analysis*, International Conference on Natural Computation, **4**, pp. 665-669, 2008.
[7] Gulick, D., *Encounter with Chaos*, McGraw Hill, 1992.
[8] Peitgen, H-O, Jurgens, H. & Saupe, D., *Fractals for The Class-Room. Part 1: Introduction to Fractals and Chaos*, Springer-Verlag New York, Inc., New York, NY, USA, 1992.
[9] Kocarev, L., Jakimoski, G., Stojanovski, T. & Parlitz, U., *From Chaotic Maps to Encryption Schemes*, In Proceedings of the 1998 IEEE International Symposium on Circuits and Systems. ISCAS '98., **4**, pp. 514-517, 1998.
[10] Heys, H.M., *A Tutorial on Linear and Differential Cryptanalysis*, Technical report, 2001.
[11] Stinson, D.R., *Cryptography: Theory and Practice*, Second Edition. Chapman & Hall/CRC, Februari 2002.