Eng. & Tech. Journal, Vol. 29 No.7, 2011

The Frequency Theoretic Estimation of non-Linear Key Generator Sequences

Dr. Abdul Monem S. Rahma* Dr. Shatha A. Salman & Ayad G. Nasser****

Received on: 18/11/2009
Accepted on: 7/4/2011

Abstract

The Randomness is one of the basic criterions to measure Key Generator Efficiency. The key generator depends basically on Linear FeedBack Shift Register which is considered as one of the basic units of Stream Cipher Systems. In this paper, the frequency postulate of Randomness criteria is calculated theoretically for non-linear key generator before it be implemented or constructed (software or hardware), this procedure save time and costs. Two non-linear key generators are chosen to apply the theoretical studies; these key generators are the Product and Brüer.

Keywords: Key Generator, Randomness Criteria, Generator Efficiency, Stream Cipher Systems.

التخمين النظري لخاصية التردد للمتتابعات الناتجة من مولدات المفاتيح غير الخطي

الخلاصة

تعتبر العشوائية (Randomness) من أهم مقاييس الكفاءة الأساسية لمولدات المفاتيح (Key Generator Efficiency) المتمثل بمنظومات المسجلات الزاحفة الخطية (LFSR) كونه أحد نظم التشفير الانسيابي (Stream Cipher Systems). في هذا البحث، تم حساب خاصية التردد احد أسس العشوائية لثنائيات المتتابعة المولدة من نظام مولد مفاتيح غير خطي نظرياً قبل تنفيذ النظام عمليا (برمجيا أو مادياً)، وهذا الأسلوب سوف يوفر الوقت والجهد والكلفة لمصمم الشفرة. تم اختيار مولدي مفاتيح غير خطية لتطبيق الدراسة النظرية للبحث هما المنظومة الضربية ومولد بريور.

Introduction

Linear Feedback Shift Register (LFSR) and Combining Function (CF) are considered as basic units to construct key generator (KG) that used in stream cipher systems [1]. Any weakness in any one of these units means clear weakness in KG sequence, so there are some conditions must be available in KG before it is constructed; therefore the KG efficiency is concluded.

In this paper,some studies are applied on the KG sequences to determine the sequence frequency. The Basic efficiency for KG can be defined as the ability of KG and its sequence to withstand the mathematical analytical which the cryptanalyst applied on them, this ability measured by some basic criterions, the most important

* Computer Science Department, University of Technology/Baghdad
**Applied Science Department,University of Technology/Baghdad

https://doi.org/10.30684/ etj.29.7.2
University of Technology-Iraq, Baghdad, Iraq/2412-0758

criterion is the randomness, one of the randomness postulates is the frequency postulate.

As known, every system passes through collection of procedures. These procedures represent the system's life cycle. This cycle start with information collecting followed by designing, constructing, evaluation, maintaining and ending with developing. In this manner, since the KG is a system then it has a cycle life. The block diagram of the Classical KG Cycle Life is shown in figure (1).

As shown in figure (1) ,the repeating of the design, constructing and evaluating for many times will cause a large cost in time and resources because of constructing procedure every time.

In the next part of this paper, the frequency postulate of randomness criterion will be discussed in details and introduce the basic conditions to obtain efficient KG specially those related to frequency. It's important to mention that the zero input sequences must be avoided, this done when the non-all zeros initial values for LFSR's are chosen.

Let KG consist of n-LFSR's have lengths r1,r2,..,rn respectively with CF=Fn(x1,x2,…,xn), s.t. xi∈ {0,1} 1≤i≤n, represents the output of LFSRi, let S={s0,s1,…} be the sequence product from KG and sj, j=0,1,… represents elements of S. let Si be the sequence i product from LFSRi with aij elements 1≤i≤n, j=0,1,…,.

2. Conditions of the Theoretical Estimation

Definition (1) [3] : Let GCD2= gcd($\prod_{i=1}^{1} m_i$,m2.GCD1)=gcd(m1,m2), for convenient let GCD1=1 and so on the general form of the recursion equation will be:

GCDn=gcd($\prod_{i=1}^{n-1} m_i$,mn.GCDn-1) …(1)

where n≥2 s.t mi are positive integers, ∀1≤i≤n.

Let the sequence S has period P(S), the period of LFSRi denotes by P(Si), P(S) and P(Si) are least possible positive integers, so

P(S)=lcm(P(S1),P(S2),…,P(Sn))…(2)

$$P(S)=\frac{\prod_{i=1}^{n} P(S_i)}{GCD_n(P(S_i))} \quad \ldots(3)$$

s.t.GCDn(P(Si))=

$$\gcd\left[\prod_{i=1}^{n-1} P(S_i), P(S_n)\cdot GCD_{n-1}(P(S_i))\right]$$

If P(Si) are relatively prime with each other this mean GCDn(P(Si))=1 this implies [3]:

$$P(S)=\prod_{i=1}^{n} P(S_i) \quad \ldots(4)$$

It's known earlier that P(Si) $\leq 2^{r_i}-1$, and if the LFSRi has maximum period then P(Si)= $2^{r_i}-1$ [4].

Theorem (1) [3]

P(S)=$\prod_{i=1}^{n}(2^{r_i}-1)$ if and only if the following conditions are holds:

GCDn(P(Si))=1,.

the period of each LFSR has maximum period (P(Si)= $2^{r_i}-1$).

3. Randomness

The sequence that is satisfied the 3-randomness properties called Pseudo Random Sequence (PRS) [4]. The randomness criterion depends on LFSR's and CF units, therefore from the important conditions to get PRSR'S, the sequence must be the maximum and CF must be balanced [5].

To guarantee the KG to produce PRS, the sequence must pass randomness tests with complete period, these tests applied in two ways, on: [1]

1. Global sequence for complete period and that is the right way (but it's hard to applied for high periods).
2. Local sequence for many times for various lengths less than the original length.

In this part, the 1st way will be applied theoretically for any period.

If $GCD_n(P(S_i))=1$ then, P(S)=

$$2^{\sum_{i=1}^{n} r_i} + (-1)\cdot(2^{r_1+\cdots+r_{n-1}} + \cdots + 2^{r_2+\cdots+r_n} + \cdots + (-1)^{n-1}\cdot(2^{r_1} + \cdots + 2^{r_n}) + (-1)^n \quad \ldots(5)$$

Definition (2): Let R_m^t denotes the combination to sum m of numbers r_i from n of the numbers r_i, R_m denotes the set of all possibilities of R_m^t s.t.

$$R_m^t = \begin{pmatrix} r_1, r_2, \ldots, r_n \\ \sum_{j=1}^{m} r_{i_j} \end{pmatrix} \quad 0\le m\le n, \quad 1\le i\le n,$$

$t\in\{1,2,\ldots,C_m^n\}$

define $R_0=\{R_0^1\}$, $R_0^1=0$.

For instance let m=1 then

$$R_1=\{R_1^1, R_1^2, \ldots, R_1^{C_1^n}\}, R_1^1=r_1, \ldots, R_1^n=r_n$$

If m=n then $R_n=\{R_n^1\}$, $R_n^1=\sum_{i=1}^{n} r_i$

So equation (5) can be written in compact formula:

$$P(S)=\sum_{k=0}^{n}(-1)^k\cdot\sum_{t=1}^{C_k^n} 2^{R_{n-k}^t} \quad \ldots(6)$$

Golomb deduced three theorems about the maximal sequence generated from LFSR [4]. One of the three Golomb's theorems deduced from the frequency postulate. the next sections will introduce new theorems, as Golomb did on LFSR, to prove the good frequency distribution of Bruer KG and the weak randomness of the product KG by applying the frequency postulates.

1st Golomb's theorem says that if LFSR with length r has maximal sequence then $N_r(0)=2^{r-1}-1$ and $N_r(1)=2^{r-1}$, where $N_r(a)$ denotes the number of bit "a" in the maximal sequence s.t.:

$$P(r)=2^r-1=(2^{r-1}-1)+2^{r-1}=\sum_{a=0}^{1} N_r(a)$$

Let $N_S(a)$ be the frequency of bit "a" in S which generates from KG then:

P(S)=

$$\sum_{a=0}^{1} N_S(a) = N_{r_1}(0)\cdots N_{r_n}(0) + N_{r_1}(0)\cdots N_{r_n}(1) + \cdots + N_{r_1}(1)\cdots N_{r_n}(1) \quad \ldots(7)$$

From this equation the act of CF will starts to distribute the ratio of "0" and "1" in S. If the terms of equation (7) rearranged s.t. $0=F(a_{i1},a_{i2},..,a_{in})$, $1\le i\le m_0$ for the 1st m_0 terms, and $1=F(a_{i1},a_{i2},..,a_{in})$, $1\le i\le m_1$ for 2nd m_1 terms $2^n=m_0+m_1$ then,

$$N_S(a)=\sum_{i=1}^{m_a}\prod_{j=1}^{n} N_{r_j}(a_{ij}) \quad \ldots(8)$$

subject to $a=F(a_{i1},a_{i2},..,a_{in})$ s.t. $1\le i\le m_a$, $a=0,1$

m_a denotes the number of states which are subject to above condition.

4. Product System (n-PKG)

This system consists of n-LFSR's with different lengths. The product KG using the non-linear product CF s.t:

$$F_n(x_1,x_2,..,x_n)=\prod_{i=1}^{n} x_i \quad \ldots(9)$$

The product function is not balanced (which expects that the n-PKG will not produces pseudo random sequences).

In the next theorem the number of ones ($N_S(1)$) of the sequence generated from n-LFSR's can be calculated.

Theorem (2):

Let $N_S(a)$ be the number of a-bit in the sequence S generated from n-PKG, a∈{0,1}, which satisfies theorem (1), then:

$$N_S(1)= 2^{\sum_{i=1}^{n} r_i - n} \quad \ldots(10)$$

Proof:

Recall equations (7) and (8).

$$P(S)= N_{r_1}(0).N_{r_2}(0).N_{r_3}(0)+\ldots+N_{r_1}(1).N_{r_2}(1).N_{r_3}(1)$$

$$N_S(1)= \prod_{i=1}^{n} N_{r_i}(1) =$$

$$N_{r_1}(1).N_{r_2}(1)\cdots N_{r_n}(1) =$$

$$2^{r_1-1}\cdot 2^{r_2-1}\cdots 2^{r_n-1} = 2^{\sum_{i=1}^{n} r_i - n} = 2^{R_n^1 - n}$$

From the result of the above theorem:

$$N_S(0)=P(S)- 2^{R_n^1 - n} \quad \ldots(11)$$

The proof of non-balance frequency of 0's and 1's in S generated from n-PKG is given in the next lemma.

Lemma (1): In the n-PKG, $\lim_{r_i \to \infty} (N_S(1)/P(S)) = \frac{1}{2^n}$, 1≤i≤n.

Proof:

$$\frac{N_S(1)}{P(S)} = \frac{2^{\sum_{i=1}^{n} r_i - n}}{\prod_{i=1}^{n} (2^{r_i} - 1)}$$

As $r_i \to \infty$, then $2^{r_i} - 1 \to 2^{r_i}$ (ignore 1), then $P(S) \approx \prod_{i=1}^{n} 2^{r_i}$.

therefore $\frac{N_S(1)}{P(S)} \approx \frac{2^{\sum_{i=1}^{n} r_i - n}}{2^{\sum_{i=1}^{n} r_i}} = \frac{1}{2^n}$

Example (1):

Table (1) shows the proportion of $N_S(1)$ to P(S) for various n-PKG.

Table (1) the proportion of $N_S(1)$ to P(S) for various n-PKG.

As shown in table in table (1) ,if the length of combined LFSR's be as high as possible then the observed proportion of NS(1) to P(S) approximate 1/2n for different n.

5. Brüer System (3-BKG)

The Brüer system consists of odd number of LFSR's with different lengths, in this paper the proposed system will consist of three LFSR's. The Brüer KG [2] using the non-linear CF called Majority CF s.t:

$$F_n(x_1,x_2,x_3)=x_1x_2 \oplus x_1x_3 \oplus x_2x_3 \quad \ldots(12)$$

The majority function is balance and symmetric (which expect that the 3-BKG will produces PRS).

In the next theorem the number of 1's in the sequence S generated from 3-BKG is calculated.

Theorem (3): Let $N_S(a)$ be the number of a-bit in the sequence S generated from 3-BKG, a∈{0,1}, then:

$$N_S(1)= 2^{r_1+r_2+r_3-1} - (2^{r_1+r_2-2} + 2^{r_1+r_3-2} + 2^{r_2+r_3-2})$$

.

Proof:

Recall equations (7) and (8), and when n=3:

$$P(S)= N_{r_1}(0).N_{r_2}(0).N_{r_3}(0)+\ldots+N_{r_1}(1).N_{r_2}(1).N_{r_3}(1)$$

$$N_S(1)= N_{r_1}(0).N_{r_2}(1).N_{r_3}(1)+N_{r_1}(1).N_{r_2}(0).N_{r_3}(1)+$$

$$N_{r_1}(1).N_{r_2}(1).N_{r_3}(0)+N_{r_1}(1).N_{r_2}(1).N_{r_3}(1)$$

=

$$(2^{r_1-1}-1)2^{r_2-1}2^{r_3-1}+2^{r_1-1}(2^{r_2-1}-1)2^{r_3-1}+$$

$$2^{r_1-1}2^{r_2-1}(2^{r_3-1}-1)+2^{r_1-1}2^{r_2-1}2^{r_3-1}$$

=

$$2^{\sum_{i=1}^{3} r_i - 3} - 2^{r_2+r_3-2} + 2^{\sum_{i=1}^{3} r_i - 3} - 2^{r_1+r_3-2} + 2^{\sum_{i=1}^{3} r_i - 3} - 2^{r_1+r_2-2} + 2^{\sum_{i=1}^{3} r_i - 3}$$

=

$$2^{\sum_{i=1}^{3} r_i - 1} - (2^{r_1+r_2-2} + 2^{r_1+r_3-2} + 2^{r_2+r_3-2})$$

$$= 2^{R_3^1-1} - (2^{R_2^1-2} + 2^{R_2^2-2} + 2^{R_2^3-2})$$

The last formula can be viewed as the following formula:

$$N_S(1)=\sum_{k=0}^{1}(-1)^k \sum_{t=1}^{C_k^3} 2^{R_{3-k}^t-(k+1)} \quad \ldots(13)$$

From equation (13):

$$N_S(0)=P(S)-\sum_{k=0}^{1}(-1)^k \sum_{t=1}^{C_k^3} 2^{R_{3-k}^t-(k+1)} \quad \ldots(14)$$

The proof of the balance frequency of 0's and 1's in S generated from 3-BKG is given in the next lemma.

Lemma (2): In the 3-BKG, $\lim_{r_i \to \infty} (N_S(1)/P(S)) = 0.5, 1 \le i \le 3.$

Proof:

$$\frac{N_S(1)}{P(S)} = \frac{2^{\sum_{i=1}^{3} r_i - 1} - (2^{r_1+r_2-2} + 2^{r_1+r_3-2} + 2^{r_2+r_3-2})}{\prod_{i=1}^{n}(2^{r_i}-1)}$$

As $r_i \to \infty$, for $1 \le i \le 3$, then $2^{r_i} - 1 \to 2^{r_i}$ (ignore 1), then

$$P(S) \approx \prod_{i=1}^{n} 2^{r_i} .$$

$$\frac{N_S(1)}{P(S)} = \frac{2^{\sum_{i=1}^{3} r_i - 1} - (2^{r_1+r_2-2} + 2^{r_1+r_3-2} + 2^{r_2+r_3-2})}{2^{\sum_{i=1}^{3} r_i}}$$

=

$$\frac{2^{\sum_{i=1}^{3} r_i - 1}}{2^{\sum_{i=1}^{3} r_i}} - \frac{2^{r_1+r_2-2} + 2^{r_1+r_3-2} + 2^{r_2+r_3-2}}{2^{\sum_{i=1}^{3} r_i}}$$

$$= \frac{1}{2} - \left(\frac{1}{2^{r_3+2}} + \frac{1}{2^{r_2+2}} + \frac{1}{2^{r_1+2}}\right)$$

As $r_i \to \infty$, then $2^{r_i+2} \to \infty$, for $1 \le i \le 3$, then:

$$\therefore \frac{N_S(1)}{P(S)} \approx \frac{1}{2} - 0 = 0.5$$

Example (2):

Table (2) shows the proportion of $N_S(1)$ to P(S) for various 3-BKG.

Table (2) The proportion of $N_S(1)$ to P(S) for various 3-BKG.

As shown in table (2) if the length of combined LFSR's be as high as possible then the observed proportion of $N_S(1)$ to P(S) approximate 0.5 for different n.

6. Applying of Chi-Square Tests on Study Cases

In this section we will apply chi-square test on the results gotten from calculations of frequency postulates on two study cases.

Let M be the number of categories in the sequence S, c_i be the category i, $N(c_i)$ be the observed frequency of the category c_i, Pr_i the probability of occurs of the category c_i, then the expected frequency E_i of the category c_i is $E_i = P(S) \cdot Pr_i$, the T (chi-square value) can be calculated as follows:

$$T=\sum_{i=1}^{K}\frac{(N(c_i)-E_i)^2}{E_i} \quad \ldots(15)$$

Assuming that T distributed according to chi-square distribution by υ=M-1 freedom degree by α as significance level (as usual

α=0.05%), which it has T_0 as a pass mark. If $T \le T_0$ then the hypothesis accepted and the sequence pass the test, else we reject the hypothesis and the sequence fails to pass the test, this mean that T not distributed according to chi-square distribution (for more information about chi-square see [6] or any book in statistics and probability).

In order to test our results we have to suggest an example suitable to our three studied cases. Let n=3, r_1=7, r_2=9 and r_3=11. P(S)= 132844159, E_i=66422079.5.

In **Frequency** test υ= 1, with α=0.05%, then T_0=3.84 (see chi-square table).

Before we discuss this postulate, Since E_i=P(S)/2, then we can conclude from equation (15) the following result:

$$T=\frac{(N(0)-P(S)/2)^2}{P(S)/2}+\frac{(N(1)-P(S)/2)^2}{P(S)/2}$$

$$=\frac{(N(0)-N(1))^2}{P(S)} \quad \ldots(16)$$

1. **3-PKG**: from equation (9) we get $N_S(1)$=16777216, then: $N_S(0)=P(S)-N_S(1)$= 116066943. By using equation (15), we get: T=74210638.706>>T_0=3.85, then S generated from n-PKG **fails** to pass this test.
2. **3-BKG**: from equation (11), $N_S(1)$=66424800, then: $N_S(0)=P(S)-N_S(1)$= 66419359. By using equation (16), we get:

T=0.223<T_0=3.85, then S generated from 3-BKG passes this test. Notice that proportion of $N_S(1)$=0.50002%.

In this section, we focus in three important procedures which are considered as the main procedures of cycle life. These three procedures are designing, constructing and evaluating. Figure (2) shows the proposed modern key generator's life cycle.

As shown in figure (2) ,this diagram is more saver in time and the resources from classical method (showed in figure (1)) since we repeat just the design and theoretical evaluation procedures without any constructing for the key generator when the key generator fail in pass the frequency criterion. After applying the practical evaluating then repeat the previous step unless the results of evaluation are success.

7. Conclusions

1. the proposed system proves that the product cryptosystem has weak statistical frequency properties, this done in deterministically by using theorem (1) which is found to calculate $N_S(1)$ of sequence generated from n-PKG.
2. The proof of non-balance frequency of 0's and 1's in the sequence generated from n-PKG given in the lemma (1).
3. The Brüer has good statistical frequency properties, this done in deterministically by using theorem (2) which is found to calculate $N_S(1)$ of sequence generated from 3-BKG.
4. The proof of the balance frequency of 0's and 1's in the sequence generated from 3-BKG given in lemma (2).
5. In this paper we introduce theoretical and practical evaluation for non-linear key generator before the practical constriction of the key generator, while before this paper all this work done statistically and after the practical constriction of the key generator.

8. Future Work

1. As future work the proposed system suggests applying another properties of randomness criterion such as, run and autocorrelation on linear or non-linear KG.
2. These theoretical studies can be applied on other kind of KG,s to calculate the frequency of these KG,s which are use combining functions with some combinations of variables.

References

[1]. Stallings, W., "***Cryptography and Net-work Security: Principles and Practices***", Pearson Prentice-Hall, 4th Edition, 2006.

[2]. Brüer, J. O., "***On Nonlinear Combinations of Linear Shift Register Sequences***" Internal Report LITH-ISY-1-572,1983.

[3]. Dr. Abdul Monem S. Rahma, Dr. Nadia M. G. Al-Saidi, and Ayad G. Nasser, "***The Theoretic Estimation of the Basic Criterions to Evaluate the Key Generator Efficiency before the Practical Construction***", The 1st Conference of Iraqi Association of Information Technology-Iraq, 17, Jen., 2009.

[4]. Golomb, S. W., "***Shift Register Sequences***" San Francisco: Holden Day 1967.

[5]. Brüer, J. O., "***On Nonlinear Combinations of Linear Shift Register Sequences***" Internal Report LITH-ISY-1- 572,1983.

[6]. Martinez, W. L. and Martinez, A. R., "***Computational Statistics Handbook with MATLAB***", Chapman & Hall/CRC, Library of Congress Cataloging-in-Publication Data, 2002.

Table (1) the proportion of N_S (1) to P(S) for various n-PKG.

n	r_i	$N_S(1)$	P(S)	Proportion	
				Expected	Observed
2	2,3	8	21	0.25	0.38
	2,5	32	93		0.34
	5,7	1024	3937		0.26
	7,11	65536	25996		0.25
3	2,3,5	128	651	0.125	0.197
	3,4,5	512	3255		0.157
	4,5,7	8192	59055		0.139
	4,5,11	131855	951855		0.131
4	2,3,5,7	8192	82766	0.0625	0.099
	3,4,5,7	32768	413385		0.079

Table (2) the proportion of N_S (1) to P(S) for various 3-BKG.

r_1	r_2	r_3	$N_S(1)$	P(S)	Proportion
2	3	5	408	651	0.62
3	4	5	824	3255	0.56
3	5	7	15040	27559	0.54

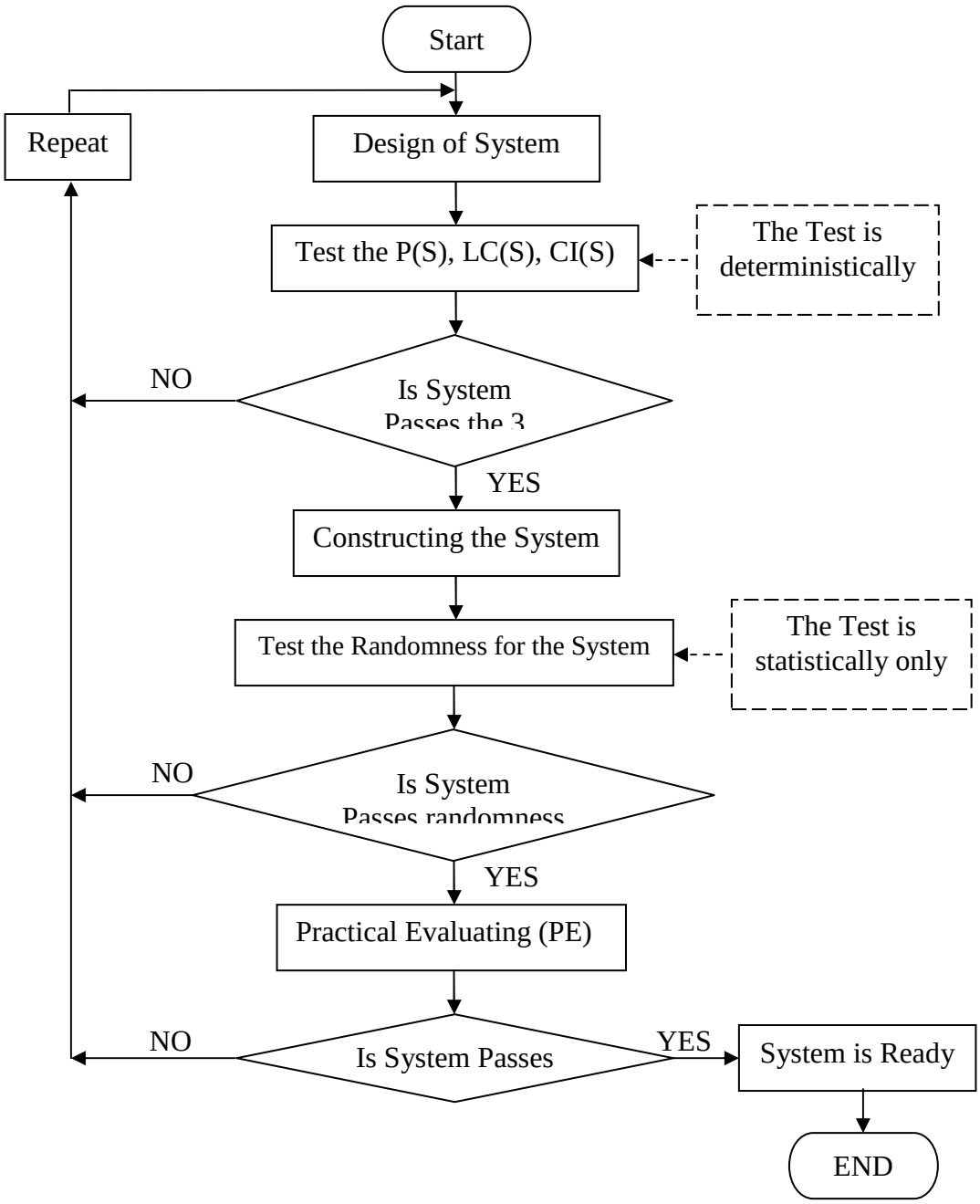

Figure (1) the classical key generator's life cycle

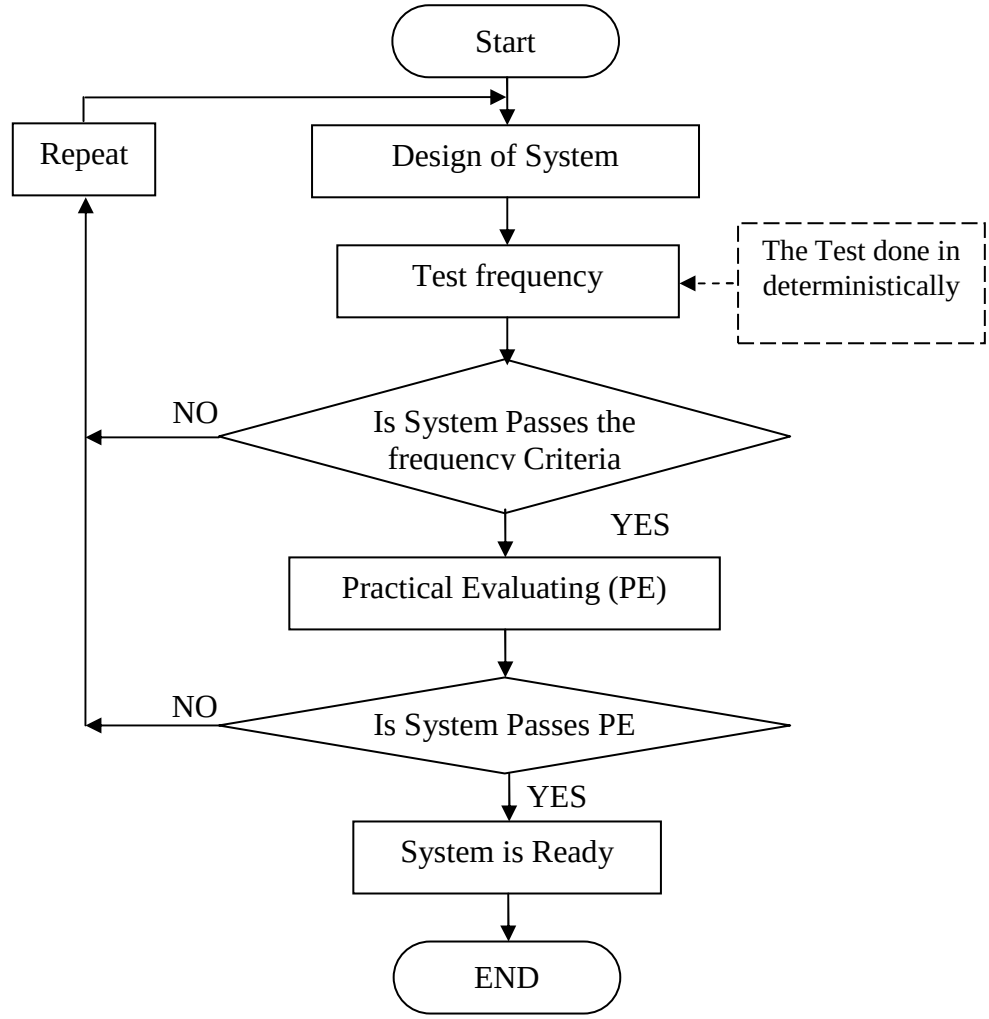

Figure (2) the proposed modern key generator's life cycle