Statistical analysis of the LFSR generators in the NIST STS test suite

Rafał Stępień, Janusz Walczak
Silesian University of Technology
44-100 Gliwice, ul. Akademicka 10, email: rafal.stepien@polsl.pl,
janusz.walczak@polsl.pl

In this article the statistical tests results of the pseudo random sequences generated by the Linear Feedback Shift Registers (LFSR) generators were described. LFSR generators' structures were shown and used statistical tests were described. The generator output sequence was analyzed in the NIST Statistical Test Suite STS 2.1.1. Interpretation of data obtained from the NIST STS 2.1.1 and the analysis' results of the pseudo random test sequences were discussed.

KEYWORDS: pseudo random signal generators, LFSR, pseudo random signal testing, statistical test suit, NIST STS 2.1.1 test suite.

1. Introduction

Generators of the pseudo random sequences are used in many cipher systems, e.g. in stream ciphers [1], in the scrambling and descrambling circuits of the TETRA system [2] or in the A5 system, which is used in the GSM mobile phones.

The pseudo random signals are generated by means of hardware circuits, by the digital electronics components or by means of software by the microprocessors circuits. The pseudo random signal generating circuits and algorithms are fully deterministic, as a result these signals cannot be considered as random signals.

In order to assure the security of the transmission or cipher with the use of the pseudo random signals, it is crucial to use the pseudo random sequence that is possibly the most difficult to reconstruct [1]. When the pseudo random sequence, in a given long time, has the properties of the fully random signal, the sequence and generator's structure's reconstruction will be very difficult.

All systems that use the pseudo random signal generators are as safe (difficult to reconstruct) as the generator's output sequence. To determine the generator's and sequence's safety is not easy [4], 5] and one of the common methods of the pseudo random sequence's and generator's security are the statistical tests [3], [4], [5], [6].

2. LFSR generator's structure

One of the pseudo random signal generators class is the class of the generators built on a shift register. This class includes two basic types of generators: Linear

Feedback Shift Registers - LFSR [1], [8] and Non Linear Feedback Shift Register – NLFSR [1]. The feedback loop parameters are invariable in time. The generator feedback is described the most often by polynomial [8].

An exemplary scheme of the LFSR generator is shown in Figure 1. This generator is described with a polynomial (1) where the power of variable x are numbers of particular bits of the shift register.

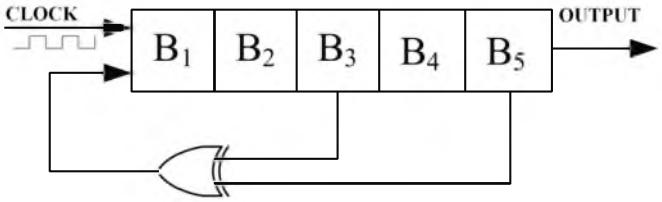

Fig. 1. Exemplary LFSR generator

$$L(x) = x^5 + x^3 + 1 \tag{1}$$

LFSR generators circuits are used in many applications, i.a in telecommunication and cryptography [1], [6], [7]. Additional details relating to structures and description of the LFSR generators may be found in [8], [9].

3. Statistical tests

The pseudo random sequences that are used in science and technology are obtained as a result of the fully deterministic algorithms, therefore this sequences cannot fully assure (with probability that equals 1) secured ciphering. Probability of the pseudo random sequence reconstruction that was used in a cipher system depends from the generator's construction. The more statistical defects and regularities are in generator's pseudo random sequence, the easier is to break the cryptography system that is used this kind of generator.

One of the methods of the pseudo random sequence's quality evaluation is the measurement of the similarity between the pseudo random sequence and the fully random sequence. Consequently, under this measurement it is possible to determine if the pseudo random signal generator assures the set level of security [3], [4]. In order to measure the above, a number of statistical test are used, for instance: DIEHARD, ENT, NIST STS and others, e.g. DIEHARDER or selected non-parametrical tests [4]. These tests find with certain probability the pseudo random sequence randomness – this is called as testing the H_0 hypothesis. The goal of each of these tests is to find whether the tested hypothesis H_0 is satisfied with the assumed significance level. The significance level α specifies the discard probability of H_0 hypothesis, if this hypothesis is true. The significance level α is usually assumed as a value from range 0,001 to 0,1 [3], [10].

Incorrect results' interpretation leads to two testing errors [3], [10]:

1. Type I error – the H_0 hypothesis is assumed as false, while it is true, as a result the random sequence is discarded,
2. Type II error – the H_0 hypothesis is assumed as true, while it is false, as a result the sequence with statistical defects is considered as a fully random sequence.

From both testing errors mentioned above, the type II error is worse. It effects with the use of unsuitable statistical parameters' generator, which decreases the security of the system that uses this generator.

4. The NIST STS-2.1.1. test suite

The statistical test suite NIST STS-2.1.1, by the National Institute of Standard and Technology (NIST) is one of the most important tools for safety in information technology, considering the generating of random and pseudorandom sequences [10]. This package consists of 15 statistical tests. Detailed documentation regarding the operation and math description of used tests is available [10]. Source codes in C language are also available as well as prepared executable application for Windows users [11].

The NIST STS 2.1.1 allows to test the sequences that are written onto a hard disk file or one of 9 implemented generators. In this paper only files sequences testing of the disk is used. Authors of the NIST package recommend that the input sequence contained in the input file should have length between 10^3 and 10^7 bits.

The NIST test suite, as a result of each statistical test, returns a number called p-value [10]. The particular statistical test is considered as passed when p-value of a statistical test is bigger than the statistical significance level. In the case of NIST test suite version STS-2.1.1 the statistical significance level was set to $\alpha = 0{,}01$.

All statistical tests' results are written into text files in a main directory of the NIST test suite. A special software was developed for further analysis. This software reads all the p-values, written in text files, and it returns them as convenient spread sheet format.

5. Statistical measurements of LFSR generators

Two test sequences, used as a test sequences, were generated by two 32-bits LFSR generators. Feedback polynomials that describe both LFSR generators were primitive [1] and were described by the following formulas:

$$L(x) = x^{32} + x^{31} + x^{29} + x^{1} + 1, \tag{2}$$

$$L(x) = x^{32} + x^{19} + x^{18} + x^{13} + 1. \tag{3}$$

Sequences were generated by software, written especially for this purpose, in the Borland Delphi 7.0 environment. Generated sequences are of a size 110MB. It was noticed that in case of a file with a size of 110MB (32-bit sequence with length

of 28672000) the NIST STS 2.1.1 test suite was not able to accomplish the spectral test – DFT. In such cases, the input file size was not reduced, but the settings of input sequence length, set from the command line, were changed to 10^7 bytes. Other tests of the 110MB sequences were executed correctly.

Generated sequences were a subject to all statistical tests that were contained in the STS 2.1.1. In table 1 numbers and test names that were used to describe the output data shown in Figure 2 and Figure 3 are given. Data shown in Figure 2 relates to the generator described by the polynomial (2) and data shown in Figure 3 relates to the generator described by the polynomial (3).

Table 1. Numbers on test figures associated with test names

Test no.	Test name	Test no.	Test name
1	The Approximate Entropy Test	9	The Overlapping Template Matching Test
2	Frequency Test within a Block	10	The Random Excursions Test
3	The Cumulative Sums (Cusums) Test	11	The Random Excursions Variant Test
4	The Discrete Fourier Transform (Spectral) Test,	12	The Binary Matrix Rank Test
5	The Frequency (Monobit) Test	13	The Serial Test,
6	The Linear Complexity Test	14	The Runs Test
7	Tests for the Longest-Run-of-Ones in a Block	15	Maurer's "Universal Statistical" Test,
8	The Non-overlapping Template Matching Test,		

The pseudo random sequence of the LFSR generator that is described by the polynomial (2) does not fulfill the following statistical tests:

- Overlapping template matchings (test number 9),
- Ranks test (test number 12).

For both mentioned tests, p-values calculated by the NIST STS 2.1.1 test suite equals 0. Others p-values are located above the statistical significance level line ($\alpha = 0{,}01$) and these values allow to ascertain that the generator sequence passes the given statistical test.

The pseudo random sequence of the second generator, that is described by the polynomial (3), does not pass the statistical test number 12 – the matrix ranks test. Results of the test 9 and 11 are located close to the statistical significance level line. The lower region, close to the statistical significance line $\alpha = 0{,}01$, of the statistical tests results is shown in Figure 4. In this figure it can be noticed, that only the test number 12 is definitely not passed. Test number 9 and 11 are located close to the statistical significance line.

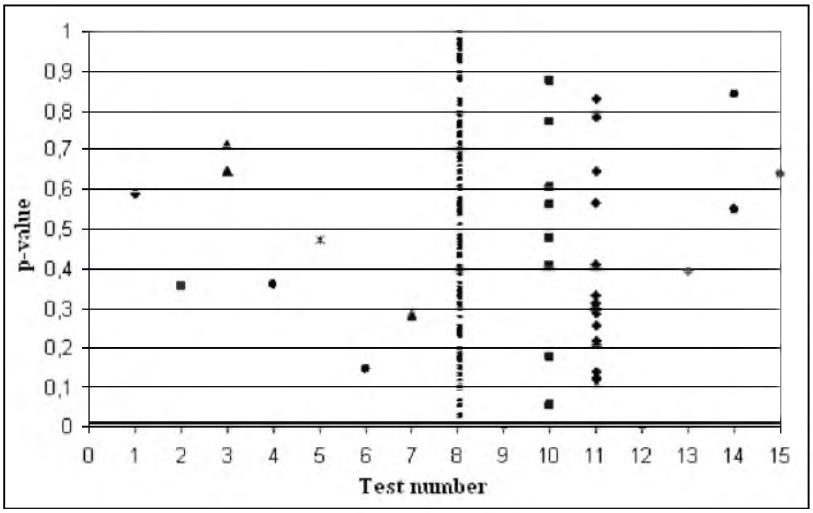

Fig. 2.Test results of the generator described by polynomial (2)

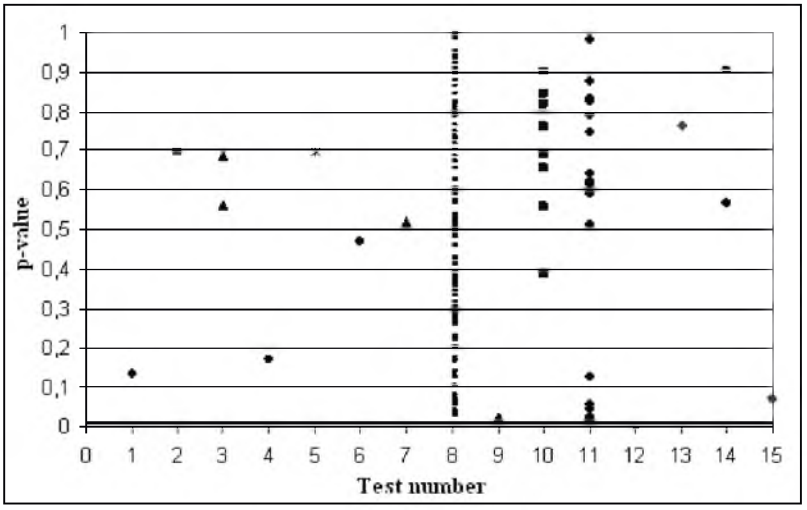

Fig. 3.Test results of the generator described by polynomial (3)

Differences of the statistical tests results are associated with the length of analyzed pseudo random sequence that equals 110 MB. The 32 bit sequence length of the LFSR generator equals:

$$l = 2^N - 1|_{N=32} = 2^{32} - 1 = 4294967295 . \quad (4)$$

The sequence with length of 4294967295 bits takes almost 4,3 Gbit on a disk file that is over 530 MB. Software generating of such long sequence is time consuming and in addition, according to the test suite documentation [10], the STS 2.1.1 test suite should not analyze sequences with lengths exceeding 100 Mbit. Also analysis of files with 110 MB in the test suite is time consuming. Analysis time, in the NIST STS 2.1.1 test suite, on a four cores Intel Core i5 2450 m (3.1 GHz) processor takes approximately 10 minutes.

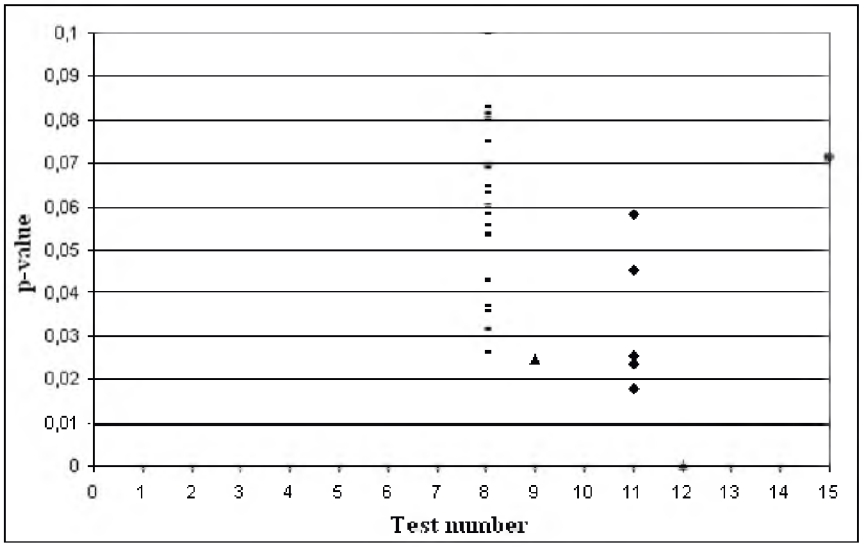

Fig. 4. Lower part of the figure 3

The binary matrixes test also appears in the DIEHARD statistical test suite. In paper [12] LFSR generator statistical tests results, conducted in DIEHARD, are described. The analyzed generator also does not pass the binary matrix ranks test, with size 31x31 and 32x32. The binary matrix ranks, size 6x8, DIEHARD test is passed.

6. Summary

In this paper the statistical tests results of the LFSR generator were described. Researches were performed in the NIST STS 2.1.1 test suite that is one of the most popular tools for pseudo random signal analysis. Two pseudo random test

sequences were generated with a software implementation of two LFSR generators that were described with different polynomials.

Paper describes the NIST STS 2.1.1 test suite, interpretation of the results and limits in pseudo random testing in this suite. Discrepancies in the results of the two tested pseudo random signal sequences, generated by the same type generators were also discussed. Obtained results were described and were presented in form of a chart. The NIST STS test suite will be used in a future work as a tool for analysis of the Dynamic Linear Feedback Shift Register sequences.

References

[1] Schneier B.: *Kryptografia dla praktyków*, Vol. 2, WNT, Warszawa 2002.
[2] Wesołowski K.: *Systemy radiokomunikacji ruchomej*, Wydawnictwo Komunikacji i Łączności, Warszawa 2006.
[3] Zwierko A.: *Testowanie generatorów pseudolosowych – wybrane programowe pakiety testów statystycznych*, VII Krajowa Konferencja Zastosowań Kryptografii, Warszawa, maj 2003, ss:1-20.
[4] Soto J.: *Statistical Testing of Random Number Generators*, National Institute of Standards & Technology, Proceedings of the 22nd National Information Systems Security Conference, 10/99, pp:1-12.
[5] Czernik P.: *Metodyka testowania bezpieczeństwa generatorów liczb pseudolosowych w systemach pomiarowo-sterujących*, Prace Instytutu Lotnictwa, Kwartalnik naukowy 6/2009 (201), ss: 20-34.
[6] Kotulski Z.: *Generatory liczb losowych: algorytmy, testowanie, zastosowania*, Matematyka Stosowana 2,2001, ss:1-9
[7] Chen L, Gong G.: *Communication Systems Security*, Appendix A, 2008, pp:1-20.
[8] Golomb S. W.: *Shift Register Sequences*, Laguna Hills, CA Aegean. Park Press, 1982.
[9] Walczak J., Stępień R.: *Modeling of the pseudo random signal generators using digital filters*. Proceedings of XXXIII conference IC-SPETO, 2010, pp: 85-86.
[10] Rukhin A i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, National Institute of Standards and Technology, rev 1a, april 2010.
[11] Strona internetowa pakietu STS 2.1.1 http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html
[12] Stępień R., Walczak J.: Analiza właściwości statystycznych sygnałów pseudolosowych generatorów zbudowanych na rejestrach przesuwnych, Zeszyty Naukowe Politechniki Poznańskiej „Electrical Engineering", zeszyt 73, 2013 ss: 65-70.