Comparative Statistical Analysis of Pseudorandom Binary Sequences by 53-bit Internal State Size Generator

Tatiana Menshikh, Ilya Piatrenka, Vitaliy Khazan, Stanislav Derechennik

***Abstract*—**In this paper we present a comparative statistical analysis of the subset of pseudorandom binary sequence generators. The Micali-Schnorr, G-SHA-1 generators and unique 53-bit internal state size generator are compared using NIST Statistical Test Suite. For practical purposes statistical tests allow extracting data on the generator producing a truly random sequence. 15 tests for sequences were carried out and analyzed.

***Keywords*—** Random number generation, testing, statistical analysis.

I. INTRODUCTION

Pseudorandom binary sequence (PRBS) generators may be used in many applications and they are central to cryptographic protocols and algorithms development. PRBS-generators testing is important both theoretically and practically. To analyze if the sequence is random, it is nessesary to determine the quality of generators. Randomness is a probabilistic observation, so the sequence is evaluated on the basis of the probability theory. A variety of statistical tests can be applied to pseudorandom sequences. The best known are the NIST suites of statistical tests (NIST STS) [1], Donald Knuth's test [2], the DIEHARD [3] and the Crypt-XS [4].

The NIST suite of statistical tests in the package include the folowing 15 tests: frequency, block frequency, cumulative sums (2 subtests), runs, long runs, Marsaglia's rank, spectral (based on the Discrete Fourier Transform), non-overlapping template matchings (148 subtests), overlapping template matchings, Maurer's universal statistical, approximate entropy (based on the work of Pincus, Singer and Kalman), random excursions (due to Baron and Rukhin, 8 subtests), random excursions variant (18 subtests), linear complexity, and serial (2 subtetsts) [1]. The basis ideas of these tests are those of Donald Knuth from his book [2], and also the DIEHARD tests developed by George Marsaglia [3]. However, all NIST tests are in the uniformed computational technique consisting of four steps. Firstly, a specific null hypothesis is formulated as the main hypothesis. Secondly, P-value statistics are calculated for the fixed block length of pseudorandom binary sequences. Thirdly, the uniformity of the resulting P-value distribution in the range [0,1] is verified. Forthly, the relative frequencies of the P-value exceeding significance level (α) with the subsequent verification of this value within the confidence interval $[\beta_{min}, \beta_{max}]$ depending on α. If P-value is within this interval, then null hypothesis is not rejected, and it means that the sequence passed the test successfully.

To compute probabilistic P-values, there are used some special functions in NIST STS: Standard normal (Cumulative Probability Distribution),

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-u^2/2) du \quad (1)$$

T. J. Menshikh Brest State Technical University, Brest, Belarus (e-mail: menshikh93@bk.ru).
I. A. Piatrenka Warsaw University of Life Sciences, Warsaw, Poland (e-mail: petrenko.ilia@gmail.com).
V. L. Khazan, Omsk State Technical University, Omsk, Russia (e-mail: vlhazan@yandex.ru).
S. S. Derechennik Brest State Technical University, Brest, Belarus (e-mail: cm@bstu.by).

Complementary error function,

$$erfc(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} \exp(-u^2) du \tag{2}$$

Incomplete gamma function,

$$Q(a,x) = 1 - P(a,x) \equiv {\Gamma(a,x)}\Big/{\Gamma(a)} \equiv \frac{1}{\Gamma(a)} \int_x^{\infty} \exp(-t) \cdot t^{a-1} dt \tag{3}$$

where $Q(a,0) = 1$, $Q(a,\infty) = 0$, and a is a variable parameter computed for each test individually [1].

II. PRELIMINARIES

Three generators were chosen for testing. The first one is a linear congruential generator previously developed and described by authors – the 53-bit internal state size generator G53 [5], [6]. Second and third generators are cryptographically secure G-SHA-1 and Micali-Schnorr (MS) ones [1]. 53-bit generator produces pseudorandom number sequence in the range [0, 1], as illustrated in Fig. 1, and each pseudorandom number x_i for this number sequence is changed to binary b_i

$$b_i = fix(2 \cdot x_i) \tag{4}$$

where *fix* is a function which rounds $2 \cdot x_i$ value to the nearest integer toward zero.

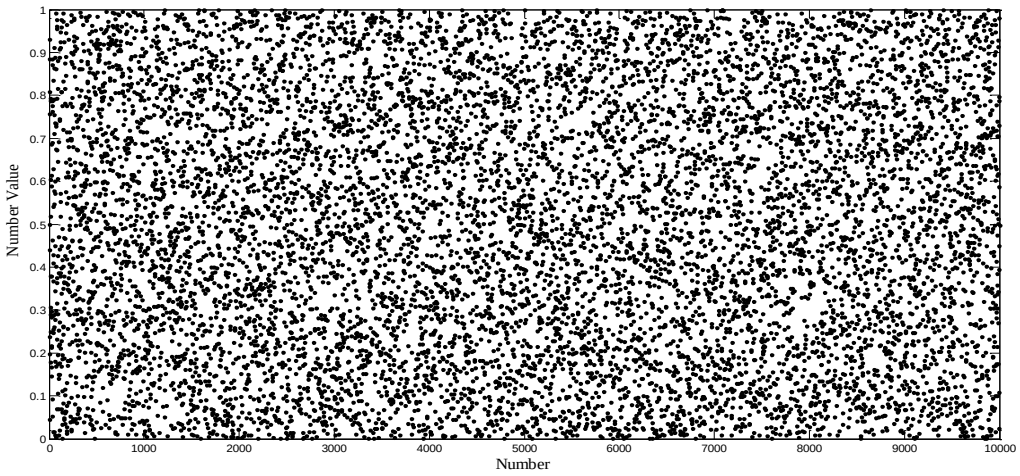

Fig. 1 The plot of pseudo-random numbers for 53-bit Generator

These generators have long M-sequences. For example, G53 generator have the period $M = 2^{53} - 1$. For each generator testing, the sequence length is $N = 2^{20}$ bit, the sample size is $n = 256$, the significance level $\alpha = 0.01$.

The parameteres for NIST tests are shown in Table I.

TABLE I
NIST TEST PARAMETERS

Statistical Test	Block length, bit
Block Frequency	16384
Non-overlapping Template	9
Overlapping Template	9
Approximate Entropy	10
Serial	16
Linear Complexity	500

Two basic cheking methods of P-value distribution uniformity are used. First one is estimation of expectation (mean), variance estimate computing, and their comparison with theoretical: mean $m = 0.5$, variance $v = 0.0833$. Second one is chi-square using P-value histogram analysis.

III. ANALYSIS AND COMPARISON

The comparative analysis results for tested generators as percent deviations of mean and variance from theoretical results are shown in Table II.

TABLE II
MEAN AND VARIANCE PERCENT DEVIATION

Statistical Test	MS		G-SHA-1		G53	
	$m(\%)$	$v(\%)$	$m(\%)$	$v(\%)$	$m(\%)$	$v(\%)$
Frequency	4.34	4.68	4.50	12.00	8.66	6.48
Block frequency	4.70	4.80	1.48	4.80	3.44	0.72
Cumulative sums	4.28	1.32	0.54	5.40	8.78	0.12
Runs	5.80	7.20	2.06	8.76	6.76	2.52
Long runs	5.22	1.20	1.44	1.56	3.74	6.96
Rank	3.36	1.68	3.6	0.84	0.66	2.52
Spectral	1.06	10.68	0.56	1.20	2.50	6.72
Non-overlapping template	0.10	0.84	0.04	0.36	0.22	0.48
Overlapping template	3.30	16.21	3.42	1.08	2.34	3.84
Universal statistical	0.30	0.84	4.34	0.84	8.84	2.28
Approximate entropy	0.28	7.68	0.02	0.48	12.34	22.2
Random excursions	0.04	0.84	1.44	0.24	1.78	0.00
Random excursions variant	1.22	2.88	0.72	4.56	0.36	2.16
Serial	3.70	13.69	4.10	0.24	9.58	2.28
Linear complexity	6.54	8.52	1.94	7.32	7.92	7.56

The minimum location parameter (expectation) deviations are observed for G-SHA-1 generator – 8 from 15 tests, for Micali-Schnorr generator – 4 from 15 tests, and for G53 generator – 3 from 15 tests. The minimum scale parameter (variance) deviations are noticed in 7 from 15 tests for G-SHA-1 generator, 5 from 15 tests for G53 generator, and 3 from 15 tests for Micali-Schnorr generator.

To show the distribution uniformity, histograms are made for one of the tests, non-overlapping template matching test, as illustrated in Fig. 2.

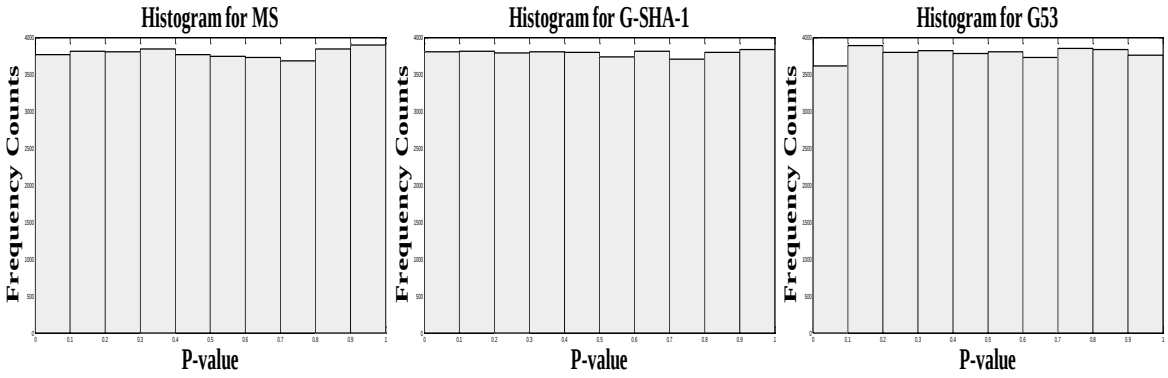

Fig. 2 Non-overlapping template matching test P-value histagrams for Generators

To estimate the P-values on the significance level $\alpha = 0.01$ (about 1% of the sequences are expected to fail), the confidence interval is defined as $\beta \pm 3\sqrt{\beta(1-\beta)/n}$, where $\beta = 1 - \alpha$, and n is the sample size. For our sample size $n = 256$, the confidence interval is 0.99 ± 0.018656. For each statistical test, the proportion of sequences that pass is computed. This is illustrated on Figs. 3–5 for all three generators (with straight horizontal line making the confidence interval lower bondary).

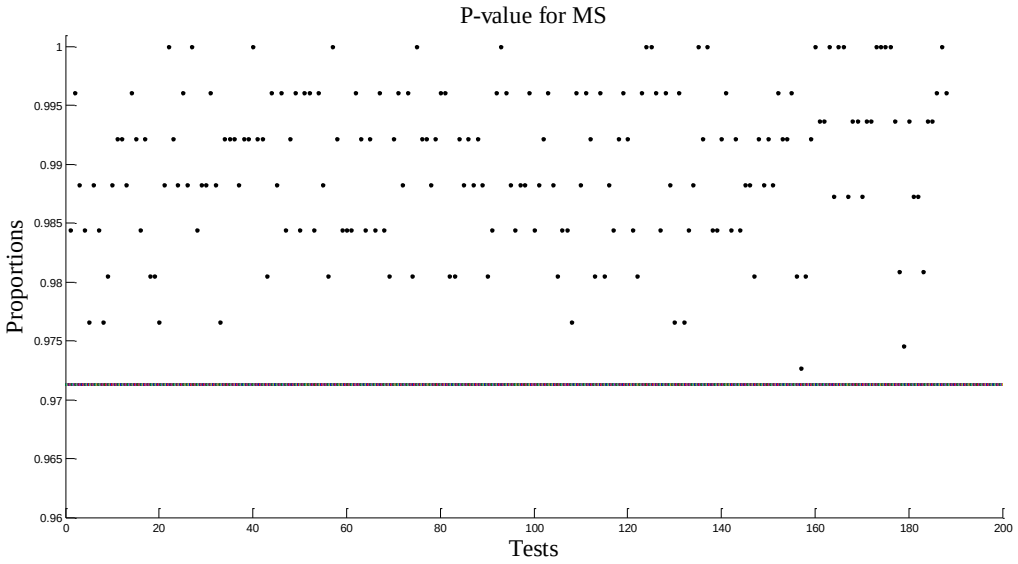

Fig. 3 P-value plot for Micali-Schnorr Generator

As can be seen in Fig. 3, the computed P-values for all tests are within the confidence interval. For G-SHA-1 generator, the significant deviation for the 44-th subtest of non-overlapping template matching test was revealed. For G53 generator, less significant deviations for the 23-th and 93-th subtests also for non-overlapping template matching test for the calculated confidence interval were revealed.

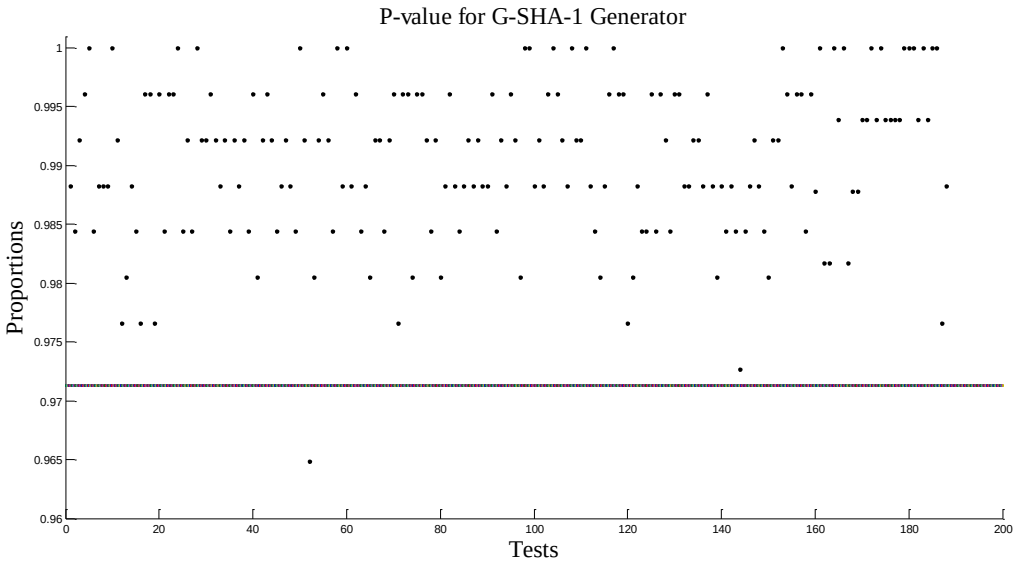

Fig. 4 P-value plot for G-SHA-1 Generator

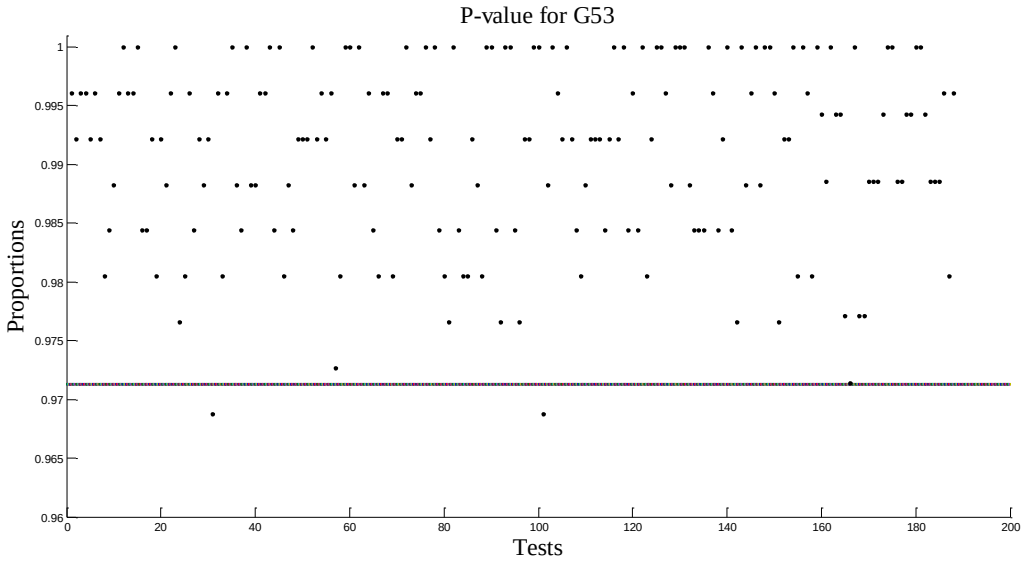

Fig. 5 P-value plot for G53 Generator

Therefore, the operating results for all three generators can be considered as satisfactory.

IV. Conclusion

The presented results of the pseudorandom binary sequence generators comparative analysis are showing that 53-bit generator can compete with the higher class generators, at least with G-SHA-1, for requirements comparable with ones researched (sequences length $N = 2^{20}$ bit and the sample size $n = 256$). In the same time G-53 is obviously faster and simpler in implementation due to its congruency, and this fact makes it a viable choice for a large subset of practical tasks.

References

[1] A. Rukhin, J. Soto, et al, "Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," NIST, US, 2010.

[2] D. E. Knuth, “The Art of Computer Programming, Seminumerical Algorithms,” 3nd ed. vol. 2, Addison Wesley, Reading. Massachusetts, 1998.
[3] G. Marsaglia, “A current view of random number generators,” Proc. Comput. Sci. Statistics: Sixteenth Symp. Interface, Mar., 1985.
[4] W. Caelli, “Crypt x package documentation”, Tech. Rep. Information Security Research, 1992.
[5] V. L. Khazan, “Mathematical Models of Discrete Decameter-Wave Communication Channels,” Publishing House of Omsk State Technical University [in Russian], Omsk, 1998.
[6] T. J. Menshikh, “Pseudorandom number generators for cryptographic secure of communication channels,” in Proc. 9th Young Scientists and Students Conf. [in Russian], Brest, 2015, pp. 31–34.