Telecommunications and Radio Engineering, 70(4):325-333 (2011)

COMPARATIVE ANALYSIS OF PSEUDORANDOM NUMBER GENERATION IN THE UP-TO-DATE WIRELESS DATA COMMUNICATION

L.O. Kirichenko, R.I. Tsekhmistro♦, O.Y. Krug, & A.W. Storozhenko

Kharkov National University of Radio Engineering and Electronics, 14, Lenin Ave, Kharkiv, 61166, Ukraine

♦Address all correspondence to R.I. Tsekhmistro E-mail: tapr@kture.kharkov.ua

The paper considers generation of pseudorandom number sequences derived by applying recurrent algorithm and chaotic reflections. The results of an operative implementation of these algorithms at the up-to-date microprocessor devices are presented. The generated sequences have been compared in terms of their compliance with the independence and evenness criteria.

KEY WORDS: *the Lehmer algorithm, chaotic reflection, hardware implementation of algorithms, Wi-Fi technology*

1. INTRODUCTION

Many experiments and analyzing techniques require generating of number sequences of a purely random nature. Such sequences can be applied in the problems of cryptography, navigation, radio engineering, location of remote and fast-moving objects, in stochastic computations and in a great variety of other implementations. Recently, the Wi-Fi (Wireless Fidelity) technology became very wide spread, enabling a wireless data communication with a high accuracy. The specification of this technique is described in the well-known standard IEEE 802.11x, regulating the data interchange procedures over the radio channels in the wireless local area networks (WLAN).

Standard IEEE 802.11x specifies two transmission methods of the spread spectrum signal: the Frequency Hopping Spread Spectrum (FHSS) and Direct Sequence Spread Spectrum (DSSS). The FHSS technology uses the whole range (2.4 GHz), which is divided into 79 channels, 1 MHz each. The receiver and transmitter are switched to the carrier frequencies of the channels in-series according to the pseudorandom law. The carrier frequencies are modulated through the two-level Gaussian frequency selection. In case of wide-band (pseudo-noise) signals, the signal spectrum is extended by adding

pseudorandom bit chains, so-called chips, to every data bit of the transmitted signal. The frequency 2.4 GHz is used both by Bluetooth devices and those using standard Wi-Fi/IEEE 802.11b. In order to eliminate interference with Wi-Fi devices, Bluetooth uses a frequency hopping spread spectrum signaling method [1].

These circumstances leave no doubt in the necessity of development and hardware implementation of efficient algorithms of pseudorandom number generation. Among the variety of distribution, the most relevant is the even one, suitable for deriving any other distribution. Thus, pseudorandom process generation is virtually reduced to deriving a set of evenly distributed random numbers.

There is a great variety of analytical methods of generating even pseudorandom numbers, e.g. computer-aided choosing the "mean product", deductions, mixing, etc. All of them are some kind of a recurrent ratio, where every next value is derived from the previous one or several numbers [2]. The most popular recent trend is searching for alternative ways of pseudorandom number generation, e.g. the method of determined chaotic reflections [3].

The goal set in this study is a practical generation of pseudorandom numbers obtained through various algorithms, by using the state-of-the-art 8-bit microcontrollers with the Harvard accumulator architecture, and property analysis of the obtained pseudorandom sequences.

The applied value of such research is self-evident, since even the most efficient algorithm may turn unacceptable because of the huge data arrays requiring to much memory. The generation time of specified number of samplings is also relevant.

2. ALGORITHMS OF PSEUDORANDOM SEQUENCE GENERATION AND THEIR PRACTICAL IMPLEMENTATION

It is impossible to imagine the practical aspect of developing and implementing telecommunication devices requiring pseudorandom signaling without up-to-date microprocessors. This imposes additional restrictions on the random number generation algorithms to be considered efficient, since small size and processing speed are usually the demands of the first priority. It means limitation on the software memory and data size. Logically, the quantity of random numbers should be limited, and the existing algorithms are subject to a check to confirm their compliance with the above criteria.

One of popular algorithms has been suggested by Lehmer and is known as the linear congruent method, as it fully satisfies the above conditions [4,5]. This algorithm has four parameters: m is the modulus (base of the system), $m > 0$, a is a multiplier factor, $0 \le a < m$, c is an increment, $0 \le c < m$, X_0 is the initial value, or kernel, $0 \le X_0 < m$.

The random number sequence $\{X_n\}$ is derived from the following iteration equality:

$$X_{n+1} = (aX_n + c) \bmod m .$$

If values of m, a and c are integers, the result will be a integer sequence within the range $0 \le X_n < m$. This is the sequence we need to select the next frequency from the range in service. The iteration intervals can be set with the internal microcontroller timer.

Assigning values to parameters m, a and c is a critical point in developing a high-performance random number generator.

There are three criteria of checking the adequacy of a random number generator:

- the function should create a full cycle, i.e., use all numbers between 0 and m before starting a new cycle,
- the created sequence should be randomly appeared. The sequence is not random by nature, since its generation is a determined process, but statistical tests should be able to confirm its randomness,
- the function should prove its practical efficiency, being run in processors or microcontrollers.

The values of parameters m, a and c should be chosen to satisfy these three criteria. According to the first criterion, if m is a prime number and $c = 0$, there exist some value of a, by which the cycle, created by the function, will be $m-1$. For 32-bit arithmetic the respective prime number is $m = 2^{31} - 1$.

Evidently, m should be very large to be suitable for creating a great number of random numbers. It is assumed that m should be approximately equal to the maximum positive integer for the processor or microcontroller in question. Thus, usually m is close or equal to 2^{31} for 32-bit processors, or 2^{15} for 16-bit microcontrollers.

There are just a few values of parameter a that satisfy all three criteria. One of them is $a = 7^5 = 16807$, the one used in the computer family IBM 360. This is a widely applied generator, having stood thousands of tests, like none of other pseudorandom number generators.

The distinction of the linear congruent algorithm is that, by a appropriately chosen combination of multiplier factor and the modulus (base), the resulting number sequence does not statistically differs from a true random sequence derived from the set $1,2,...,m-1$. However, an algorithm-based sequence cannot contain any randomness, and the initial value X_0 does not matter. If the value is selected, all the rest numbers in the sequence are predetermined. This fact is always taken into account in cryptanalysis. This is a relevant aspect in our study, since for creating an adequate pseudorandom number generator we need a strictly defined cryptoalgorithm. In microprocessors, the analysis of the efficiency of pseudorandom sequence generation algorithms can rely on the estimation of the generation time (processing speed), the program and data storage required for a reliable performance of a certain algorithm.

In our study we used the microcontroller AVR-ATmega 128 (product by ATMEL) with the following characteristics:

- 128 kb of FLASH memory for programs
- 4 kb of static RAM
- 4 kb of EEPROM memory for data

The keyboard and the LCD Toshiba T6963C, our laboratory specimen is equipped with, enabled us to observe sampling of an arbitrary and fixed quantity of numbers for various generation algorithms. The 8-bit and 16-bit counter-timer and LCD allow to estimate the generation time of a fixed number of random numbers for certain algorithms, while using interruption for registering and displaying the generation time of a sequence cycle. The device implementing the linear congruent method in practice is shown in Fig. 1.

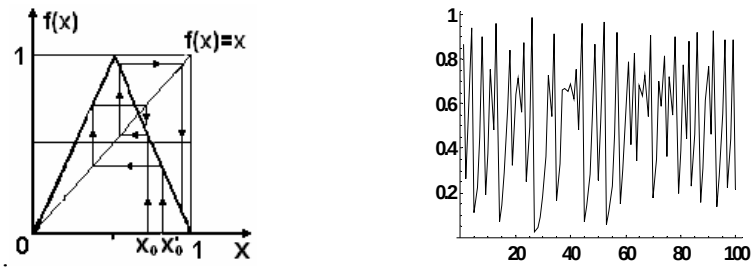
FIG. 1: Microcontroller-based (AVR-ATmega 128) laboratory breadband construction

This specimen allows to track efficiency of the method depending on the variations of the initial parameters (m , a , c and X_0). It displays the result, random signal generation time, total value of the random numbers in bytes on the liquid-crystal panel.

The above laboratory breadband construction was used to implement an alternative method of random number generation. The generator utilizes the principle of the one-dimensional iterative representation, often called the triangular representation, which is described with the formula

$$X_{n+1} = r\left(1 - 2\left|0.5 - X_n\right|\right),$$

where r is the representation parameter, while the sequence $\{X_n\}$ varies within the range of (0,1). If $r > \frac{1}{2}$, the initially close points are moving away from each other as iteration proceeds (Fig. 2(a)), and the triangular representation initiates a chaotic sequence of numbers $\{X_n\}$, which can be considered random (Fig. 2(b)).



a) b)

FIG. 2: Triangular representation and chaotic sequence of numbers

The invariant measure $\rho(x)$ defines the iteration density of representation $X_{n+1} = f(X_n),\ X_n \in [0,1]$. The invariant measure $\rho(x)$ can be considered to be an analogue of a random value distribution density. For a triangular representation by $r = 1$, obtain $\rho(x) = 1$ [3]. It means that the sequence of iterations $X_0, f(X_0), f(f(X_0)),...$ covers the range (0,1) uniformly. Thus, we can state that the triangular representation is a generator of uniformly distributed random numbers.

3. INDEPENDENCE AND UNIFORMITY TEST OF PSEUDORANDOM NUMBERS

Evidently, the properties of a poorly random sequence should be as follows: no correlation between numbers (independence and randomness of selective datum) and compliance with the specified distribution law, in this particular case the uniform one.

The independence (randomness) tests for the sampling data have been performed in terms of series, inversions and turning points. In every case we hypothesize (H_0) that independent outcomes of one and the same value, and set the significance level α . The brief overview of the applied factors is given below [2,6].

Criterion of series. Let us consider a sequence of N observed values of a random quantity, while every observation case is assigned to one of two mutually exclusive classes. A series is a sequence of similar observations, which is preceded with and followed by observation of the opposite class. The number of series in the observation sequence allows to define whether individual results can be considered as independent observations. Division into classes is always possible, so that $N_1 = N_2 = N/2$, e.g. by comparing the observations with the sampling median.

If the sequence of N observations consists of independent outcomes of one and the same random value, the number of series in the sequence is a random value ξ with the expectancy of $m_\xi = N/2 + 1$ and dispersion $\sigma_\xi^2 = \dfrac{N(N-2)}{4(N-1)}$. The probability distribution of the series number ξ is tabulated.

As a null hypothesis we assume that the observations are independent. To validate the hypothesis with the required significance value α, the recorded number of series should be compared with the acceptance region boundaries. If the number of series is beyond this region, the hypothesis does not qualify. Otherwise the hypothesis can be accepted with the significance value α.

Criterion of inversions. Let us consider the inequalities $x_i > x_j$ by $i < y$, i.e., inversions, in a sequence of N observations. We calculate the total number of inversions. If a sequence of N observations consists of independent outcomes of one and the same random value, the number of inversions is a random value ξ with the expectancy $m_\xi = \frac{N(N-1)}{4}$ and a dispersion $\sigma_\xi^2 = \frac{2N^3 + 3N^2 - 5N}{7^2}$. The number of inversions occurred in the recorded sequence will be characterized by a tabulated sampling distribution. To validate the hypothesis about the independence of the data with the required significance value α, the recorded value should be compared with the acceptance region boundaries. If the number of series is beyond this region, the hypothesis should be declined. Otherwise the hypothesis can be accepted with the significance value α.

Criterion of turning points. In a sequence of N observations we count the peaks and concaves, i.e., the number of inequalities $x_i < x_{i+1} > x_{i+2}$ or $x_i > x_{i+1} < x_{i+2}$. Each of such inequalities means a turning point. Let us find the total number of turning points.

If a sequence of N observations consists of independent outcomes, the number of turning points is a random value ξ with the expectancy $m_\xi = \frac{2}{3}(n-2)$ and a dispersion $\sigma_\xi^2 = \frac{16n - 29}{90}$. The number of turning points occurring in the recorded sequence tends to the normal distribution $N(m_\xi, \sigma_\xi)$. TO validate the hypothesis about the independence of data with the required significance value of α, the recorded number of turning points should be compared with the acceptance region boundaries $\lfloor m_\xi - t * \sigma_\xi m_\xi + t * \sigma_\xi \rfloor$, where $2\Phi(t) = 1 - \alpha$, and Φ is the Laplace integral. If the number of series is beyond this region, decline the hypothesis. Otherwise the hypothesis can be accepted with the significance value α. To check the uniformity of the data, the fitting criterion has been used to verify the statistic hypothesis about the distribution law. Three criteria have been used: the Pearson criterion, the Kolmogorov criterion and the Mises criterion. In each case we put forward a hypothesis H_0 about the evenness of the distribution law in the obtained sampling, and defined the significance value α. Below we give a brief overview of the applied criteria.

The Pearson criterion (criterion χ^2). The basic operational principle of this criterion consists in finding such a measure of deviation between the theoretical $F(x)$ and empirical $F_n(x)$ distribution, which would approximately obey the

distribution law χ^2. We divide the interval $[a,b]$ into l nonintersecting intervals (units). To find the common measure of deviation between $F(x)$ and $F_n(x)$ one has to compute the statistics $\chi^2 = \sum_{i=1}^{1} \frac{n_i - n \cdot p_i}{n \cdot p_i}$, where n_i is the observed rate of hit in the $i-th$ unit, n is the sample size, p_i is the theoretical probability of hitting the $i-th$ units. If the statistics value is $\chi^2 < \chi^2(k,\alpha)$, where $\chi^2(k,\alpha)$ is the tabulated value of χ^2 with the degree of freedom $k = 1-3$ and the significance value α, the hypothesis H_0 is acceptable, otherwise not.

The Kolmogorov criterion. While using this technique, the sample is expanded into variational series, instead of being divided into units. The modulus of the maximum difference $d_n = \max_{x \in \{x_n\}} \left| F(x) - F_n(x) \right|$ serves as the measure of deviation between the theoretical $F(x)$ and empirical $F_n(x)$ distribution functions of a continuous random value X.

By an unlimited number of observations n, the distribution function of the random value $d_n\sqrt{n}$ asymptotically approaches to the distribution function $K(\lambda) = P\left(d_n\sqrt{n} < \lambda\right) = \sum_{k=-\infty}^{\infty} (-1)^k \exp\left(-2k^2\lambda^2\right)$. Hence, the hypothesis H_0 can be accepted, if the statistics value is $d_n\sqrt{n} < \lambda$. With the specified significance value α, the value of λ is chosen from the ratio $\alpha = 1 - K(\lambda)$. Otherwise the hypothesis H_0 is declined.

The Mises criterion (criterion w^2). According to the Mises criterion, the mean squared deviation over all argument values $x: w_n^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, dF(x)$ is used as a measure of deviation between the theoretical $F(x)$ and empirical $F_n(x)$ distribution functions. The criterion statistics is the value of $nw_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x_i) - \frac{i - 0.5}{n} \right]^2$. If n increases without limitations, the limiting statistics distribution will be nw_n^2. By choosing the significance value α, one can determine the critical values of $nw_n^2(\alpha)$. If the actual value of nw_n^2 is equal or higher than the critical one, then according to the Mises criterion with the significance value α the hypothesis H_0 should be declined, otherwise accepted.

4. RESULTS OF TESTING

The applied laboratory device allowed us to implement the above described, as well as other algorithms of random number generation and analyze the following aspects:

- generation time estimation for random number sequences of similar length, with the amount of storage being equal,
- estimation of the amount of storage taken by the generated numbers, with their number in every algorithm being the same,
- analysis of initial parameters and their influence on each algorithm's performance,
- estimation of sample instant characteristics of random number sequences by a fixed and variable generation times.

The numerical simulation of the pseudorandom number generation has been performed using a test model, after the technique of chaotic reflection and the Lehmer algorithm. In every case, a 100 elements sampling has been generated, the independence hypothesis has been verified for every sampling, using the criteria of series, inversions and turning points; the evenness hypothesis has been verified using the Pearson, Kolmogorov and Mises criteria with the significance value of $\alpha = 0.05$.

The time of generating 100 symbols, according to the Lehmer algorithm , by the clock rate of 11.059 MHz, neglecting the time of displaying symbols on the liquid crystal monitor, was about 160 μs; with the chaotic reflection and all the rest parameters being equal, the generation time was 105 μs. However, the estimation of the generation time and the amount of storage taken by the generated numbers is strongly influenced by the algorithm parameters. These relations are the object of future research.

The independence tests have revealed that the data obtained by the triangular representation, as well as the pseudorandom numbers of embedded generators, satisfy the requirements of independence; no correlation between the numbers is revealed.
The results of evenness tests are listed in Table 1.

TABLE 1: Acceptance rate of the evenness hypothesis

	The Pearson criterion	**The Kolmogorov criterion**	**The Mises criterion**
The Lehmer algorithm	89.9	96.9	93.8
Chaotic reflection	87.3	97.2	94.5

5. CONCLUSIONS

The hardware implementation of the pseudorandom number generators by various techniques has been suggested in this paper. The tests proved the obtained numbers to be uncorrelated and even. Testing of the algorithms at a microprocessor mock-up was

included into the program of academic activity as a practical training in algorithm's implementation.

REFERENCES

1. Zhuravlev, V.I., (1986), *Search and synchronization in broadband systems*, Radio and Svyaz, Moscow: 102 p. (in Russian).
2. Yermakov, S.M., (1975), *The Monte Carlo methods and associated questions*, Nauka, Moscow: 211 p. (in Russian).
3. Schuster, H.G., (1988), *Deterministic Chaos*, Mir, Moscow: 240 p. (in Russian).
4. Nechayev, V.I., (1999), *Some issues of cryptography*, Vysshaya Shkola, Moscow: 109 p. (in Russian).
5. Sarvate D.V. and Presley, M.B., (1980), Intercorrelation properties of consequent and contiguous series, *TIIER*, **68**(5):59-88.
6. Kendall, M., (1981), *Time series*, Finansy and Statistika, Moscow: 198 p. (in Russian).
7. Tyurin, Yu.N. and Makarov, A.A., (1997), *Computer-aided statistical analysis of data*, Infra, Moscow: 528 p. (in Russian).