УДК 621.391:004.73

Ю. Г. Лега, *д.т.н., профессор,*
Э. В. Фауре, *к.т.н., доцент,*
А. А. Лавданский
Черкасский государственный технологический университет

ТЕХНОЛОГИЯ ГЕНЕРАЦИИ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С БОЛЬШОЙ РАЗРЯДНОСТЬЮ ЧИСЕЛ

У статті розглядається задача формування рівномірно розподіленої випадкової послідовності чисел великої розрядності шляхом конкатенації слів, утворених незалежними генераторами рівномірно розподілених випадкових чисел. Для будь-якої системи числення визначено математичне очікування, дисперсія, функція розподілу та діапазон рівномірного розподілу конкатенації чисел залежно від розрядностей і діапазонів розподілу слів, що входять в конкатенацію. Розроблено структурну схему генератора рівномірно розподілених випадкових чисел, сформованих шляхом конкатенації чисел меншої розрядності, породжених незалежними генераторами рівномірно розподілених випадкових чисел.

Ключові слова: *генератор випадкових чисел, рівномірний розподіл, конкатенація.*

In the article the problem of formation of uniformly distributed random number sequence with large word length by concatenation of words formed by independent uniformly distributed random numbers generators is considered. An expectation value, variance, distribution function and uniform distribution range of numbers concatenation according to their capacity and distribution range is determined for any used number system. Structural scheme of the generator of uniformly distributed random numbers formed by concatenation of numbers of less capacity generated by independent uniformly distributed random numbers generators is developed.

Key words: *random numbers generator, uniform distribution, concatenation.*

Введение. Генераторы случайных чисел (ГСЧ) находят широкое применение во многих отраслях человеческой деятельности. В настоящее время известно немалое количество как детерминированных алгоритмов генерации случайных (точнее, псевдослучайных) чисел, так и алгоритмов, использующих внешние источники энтропии [1, 2].

Однако для всех алгоритмов характерна ограниченная разрядность внутреннего состояния генератора и состояния его выхода. Увеличение разрядности неизбежно ведет к усложнению алгоритма ГСЧ и, как следствие, к снижению его производительности.

Главным образом, указанный недостаток касается ГСЧ с аппаратной реализацией или программных ГСЧ, разрядности внутреннего и выходного состояний которых ограничены, например, разрядностью вычислительной платформы.

Целью настоящей работы является разработка метода формирования равномерно распределенной последовательности случайных чисел, разрядность которых выходит за пределы разрядности вычислительной платформы ГСЧ.

Постановка задачи. Для достижения поставленной цели необходимо решить научно-техническую задачу определения закона распределения композиции чисел двух или нескольких первичных ГСЧ меньшей разрядности.

В качестве композиции чисел в настоящей работе рассматривается их конкатенация.

Первичные источники случайных чисел являются независимыми генераторами, формирующими по одному случайному числу за некоторый, единый для всех первичных ГСЧ, интервал времени.

Решение задачи. В работе [3] отмечено, что под операцией конкатенации, в зависимости от используемых объектов, понимается объединение нескольких подстрок в одну строку, объединение подслов (старшей и младшей части) в одно слово, объединение двух последовательностей в одну.

Кроме того, в [3] показано, что результатом конкатенации циклов, порождаемых генератором конгруэнтных чисел, является сверхцикл генератора, содержащий в себе все элементы из генерируемого множества чисел. Закон распределения чисел внутри сверхцикла является равномерным.

В настоящем исследовании интерес представляет закон распределения чисел, образованных конкатенацией слов, формируемых двумя или несколькими независимыми первичными генераторами равномерно распределенных случайных чисел.

Слова на выходе первичного генератора представляют собой целые неотрицательные числа некоторого диапазона и разрядности.

Для определения закона распределения конкатенированных слов сформулируем и докажем следующее утверждение.

Утверждение. Дискретная случайная величина, сформированная с помощью конкатенации k равномерно распределенных независимых дискретных случайных величин, принимающих целые неотрицательные значения, имеет также равномерное распределение.

Доказательство

Очевидно, что, если утверждение выполняется для $k=2$, то оно выполняется и для $k \geq 2$.

Пусть $k=2$, а исходные равномерно распределенные случайные величины X и Y представимы в некоторой позиционной системе счисления с основанием a:

$$X = a_{n_x-1}a_{n_x-2}\mathbf{K}\,a_1a_0 = \sum_{i=0}^{n_x-1} a_i a^i,$$

$$Y = b_{n_y-1}b_{n_y-2}\mathbf{K}\,b_1b_0 = \sum_{j=0}^{n_y-1} b_j a^j,$$

где $a_i, b_j \in [0; a-1]$.

Разрядность случайной величины X составляет n_x a-ичных разрядов, случайной величины Y – n_y a-ичных разрядов.

Конкатенация двух величин представляет собой их объединение или «склеивание», что соответствует их последовательной записи. Пусть конкатенация случайных величин X и Y происходит следующим образом:

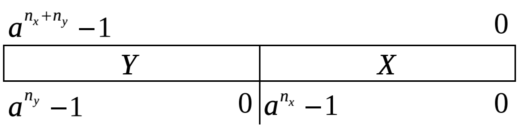

Рис. 1. Конкатенация случайных величин X и Y

Результатом конкатенации является случайная величина

$$Z = Y \cdot a^{n_x} + X = g_{n_x+n_y-1}g_{n_x+n_y-2}\mathbf{K}\,g_1g_0 =$$
$$= g_{n_x+n_y-1}a^{n_x+n_y-1} + g_{n_x+n_y-2}a^{n_x+n_y-2} + \mathbf{K} + g_1a^1 + g_0 =$$
$$= \sum_{k=0}^{n_x+n_y-1} g_k a^k,$$

где $g_k \in [0; x-1]$;

$$\begin{cases} g_k = a_k \ \text{при} \ k \in [0; m-1]; \\ g_k = b_{k-m} \ \text{при} \ k \in [m; n+m-1]. \end{cases}$$

Разрядность случайной величины Z составляет $n_x + n_y$ a-ичных разрядов.

Случайная величина $Z = Y \cdot a^{n_x} + X$ принимает некоторое значение $Z = z_i$ из множества возможных значений тогда и только тогда, когда $X = x_i$ и $Y = y_i$, причем это условие является необходимым и достаточным:

$$Z = z_i \Leftrightarrow \begin{cases} X = x_i, \\ Y = y_i. \end{cases}$$

Таким образом, вероятность того, что величина Z принимает некоторое значение z_i из множества возможных значений, определяется вероятностями $P(X = x_i)$ и $P(Y = y_i)$. В силу независимости X и Y

$$P(Z = z_i) = P(X = x_i) \cdot P(Y = y_i).$$

Если случайная величина X имеет дискретное равномерное распределение на множестве значений мощности N_x, а случайная величина Y – дискретное равномерное распределение на множестве мощности N_y, то

$$P(X = x_i) = \frac{1}{N_x},\ P(Y = y_i) = \frac{1}{N_y}, \text{ а}$$

$$P(Z = z_i) = \frac{1}{N_x N_y},$$

что и определяет закон распределения дискретной случайной величины $Z = Y \cdot a^{n_x} + X$.

Мощность множества значений случайной величины Z равняется $N_x N_y$.

Таким образом, дискретная случайная величина, сформированная с помощью конкатенации двух равномерно распределенных независимых дискретных случайных величин, принимающих целые неотрицательные значения, имеет также равномерное распределение, что и требовалось доказать.

Указанное утверждение инвариантно по отношению к области определения случайных величин X и Y.

Если в операции конкатенации участвуют k равномерно распределенных независимых дискретных случайных величин с мощностями множеств значений N_0, N_1, ..., N_{k-1}, то закон распределения конкатенации имеет вид:

$$P(Z = z_i) = \frac{1}{N_0 N_1 \ldots N_{k-1}}.$$

Отметим, что в большинстве случаев генераторы случайных чисел способны формировать любые значения из некоторого конечного диапазона неотрицательных целых чисел $Z_{\geq 0}$.

Определим функцию распределения конкатенации Z величин, сформированных подобными генераторами случайных чисел.

В общем случае функция распределения дискретной случайной величины Z имеет вид:

$$F(z) = \sum_{z_i < z} P(Z = z_i) = \sum_{z_i < z} \frac{1}{N_0 N_1 \ldots N_{k-1}}.$$

Для определения значения функции распределения необходимо выполнить расчет количества всех значений $z_i < z$. Определим зависимость функции распределения конкатенации случайных величин только от значения z и диапазонов определения случайных величин.

Пусть случайная величина X равномерно распределена на отрезке $[A_x, B_x]$ с мощностью множества значений $N_x = B_x - A_x + 1$, а случайная величина Y – на отрезке $\left[A_y, B_y\right]$ с мощностью множества значений $N_y = B_y - A_y + 1$.

Случайная величина Z равномерно распределена на множестве целых чисел из диапазона $\bigcup_{l=A_y}^{B_y} \left[la^{n_x} + A_x, la^{n_x} + B_x\right]$ с мощностью множества значений $N_x N_y$.

Пусть $X' = X - A_x$, а $Y' = Y - A_y$. Тогда случайные величины X' и Y' равномерно распределены на отрезках $[0, B_x - A_x]$ и $\left[0, B_y - A_y\right]$ соответственно. Случайная величина $Z' = Y' \cdot a^{n_x} + X' = Z - A_y \cdot a^{n_x} - A_x$ равномерно распределена на множестве целых чисел из диапазона $\bigcup_{l=0}^{B_y - A_y} \left[la^{n_x}, la^{n_x} + B_x - A_x\right]$.

Таким образом, в силу наличия биекции $Z \leftrightarrow Z'$ с сохранением структуры множества значений величин Z и Z', эти множества $\{z_i\}$ и $\{z_i'\}$ являются изоморфными.

Рассмотрим случайную величину $Z' = Y' \cdot a^{n_x} + X'$ и найдем для нее функцию распределения.

Определим функцию распределения $G(x', y')$ и закон распределения $g(x', y')$ системы случайных величин (X', Y'). Для этого воспользуемся рис. 2.

В соответствии с рис. 2, $G(x', y') = P(X' < x', Y' < y') = \dfrac{x'y'}{N_x N_y}$ при $x' \in [0; N_x]$, $y' \in \left[0; N_y\right]$.

Закон распределения системы дискретных случайных величин (X', Y') $g(x', y') = \dfrac{1}{N_x N_y}$.

Для нахождения закона распределения дискретной случайной величины $Z' = Y' \cdot a^{n_x} + X'$ построим на плоскости $x'0y'$ прямую $z' = y' \cdot a^{n_x} + x'$ (рис. 2).

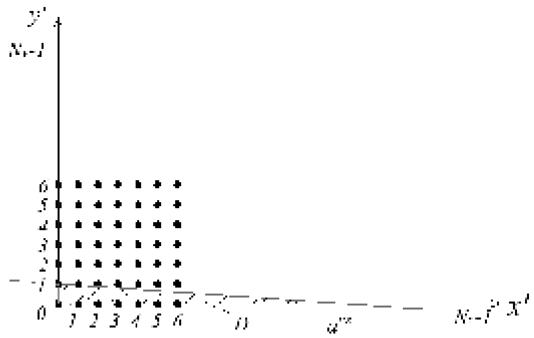

Рис. 2. Положение случайных величин на числовой оси

Обозначим D область, для которой высота поверхности $z' = y' \cdot a^{n_x} + x'$ над плоскостью $x'0y'$ меньше z. Чтобы выполнялось неравенство $F(z) = P(Z < z)$, случайная точка (x', y') должна попасть в область D. Следовательно, функция распределения величины $Z' = Y' \cdot a^{n_x} + X'$

$$F(z') = P((X', Y') \subset D) = \sum_{(x', y') \subset D} g(x', y') =$$

$$= \sum_{p=0}^{\frac{z' - |z'|_{a^{n_x}}}{a^{n_x}}} \sum_{r = pa^{n_x}}^{\min\left(pa^{n_x} + N_x - 1; z'\right)} \frac{1}{N_x N_y} =$$

$$= \frac{1}{N_x N_y} \left(\frac{z' - |z'|_{a^{n_x}}}{a^{n_x}} \cdot N_x + |z'|_{a^{n_x}} \right).$$

Таким образом, функция распределения конкатенации случайных величин X' и Y'

$$F(z') = \frac{1}{N_x N_y} \left(\frac{z' - |z'|_{a^{n_x}}}{a^{n_x}} \cdot N_x + |z'|_{a^{n_x}} \right).$$

В силу изоморфизма $\{z_i\}$ и $\{z_i'\}$ функция распределения конкатенации $Z = Y \cdot a^{n_x} + X$ случайных величин X и Y, равномерно распределенных на отрезках $[A_x, B_x]$ и $[A_y, B_y]$, соответственно,

$$F(z) = \frac{1}{N_x N_y}\left(\frac{z' - |z'|_{a^{n_x}}}{a^{n_x}} \cdot N_x + |z'|_{a^{n_x}}\right),$$

где $z' = z - A_y \cdot a^{n_x} - A_x$,

$N_x = B_x - A_x + 1$,

$N_y = B_y - A_y + 1$.

Пусть случайная величина X_i с разрядностью n_i на выходе генератора *Гi* имеет равномерное распределение на отрезке $[A_i, B_i]$, $i \in [0, k-1]$. Тогда числа, образованные путем конкатенации чисел на выходе источников *Гi* вида

X_{k-1}	$\ldots$	X_1	X_0

равномерно распределены на множестве целых чисел из диапазона

$$\bigcup_{l_{k-1}=A_{k-1}}^{B_{k-1}} \ldots \bigcup_{l_2=A_2}^{B_2} \bigcup_{l_1=A_1}^{B_1} \left[\begin{matrix} l_{k-1}a^{p_{k-1}} + \ldots + l_2 a^{p_2} + l_1 a^{p_1} + A_0, \\ l_{k-1}a^{p_{k-1}} + \ldots + l_2 a^{p_2} + l_1 a^{p_1} + B_0 \end{matrix}\right]$$

где $p_m = \sum_{l=0}^{m-1} n_l$ $(m > 0)$, $p_0 = 0$.

Функция распределения конкатенации k равномерно распределенных случайных величин

$$F(z) = \frac{1}{N_0 N_1 \ldots N_{k-1}}\left(\begin{matrix}\sum_{i=1}^{k-1} \frac{|z'|_{a^{p_{i+1}}} - |z'|_{a^{p_i}}}{a^{p_i}} \cdot N_i + \\ + |z'|_{a^{p_1}}\end{matrix}\right),$$

где $N_i = B_i - A_i + 1$,

$$z' = z - \sum_{j=0}^{k-1} A_i \cdot a^{p_j}.$$

Определим математическое ожидание и дисперсию случайной величины $Z = \sum_{i=0}^{k-1} X_i \cdot a^{p_i}$.

Математическое ожидание случайной величины X_i в силу ее равномерного распределения

$$M(X_i) = \frac{A_i + B_i}{2}.$$

С учетом того, что $M(X_j \cdot a^m) = M(X_j) \cdot a^m$, а величины, входящие в конкатенацию, независимы,

$$M(Z) = M\left(\sum_{i=0}^{k-1} X_i \cdot a^{p_i}\right) = \sum_{i=0}^{k-1}\left(M(X_i) \cdot a^{p_i}\right)$$

или

$$M(Z) = \sum_{i=0}^{k-1}\left(\frac{A_i + B_i}{2} \cdot a^{p_i}\right).$$

Дисперсия случайной величины X_i в силу ее равномерного распределения

$$s^2(X_i) = \frac{N_i^2 - 1}{12}.$$

Учитывая, что случайные величины X_i, $i \in [0, k-1]$, независимы, их ковариации равны нулю, а

$$s^2(Z) = s^2\left(\sum_{i=0}^{k-1} X_i \cdot a^{p_i}\right) = \sum_{i=0}^{k-1}\left(s^2(X_i) \cdot a^{2p_i}\right)$$ или

$$s^2(Z) = \frac{1}{12}\sum_{i=0}^{k-1}\left(\left(N_i^2 - 1\right) \cdot a^{2p_i}\right).$$

Покажем возможность использования операции конкатенации в задачах генерации последовательностей с большой разрядностью чисел. Для этого сформулируем следствие для доказанного выше утверждения.

Следствие. Слово, сформированное с помощью конкатенации k слов, образованных k независимыми первичными генераторами равномерно распределенных случайных чисел, имеет равномерное распределение.

Таким образом, структурная схема генератора равномерно распределенных случайных чисел может быть представлена в виде, изображенном на рис. 3.

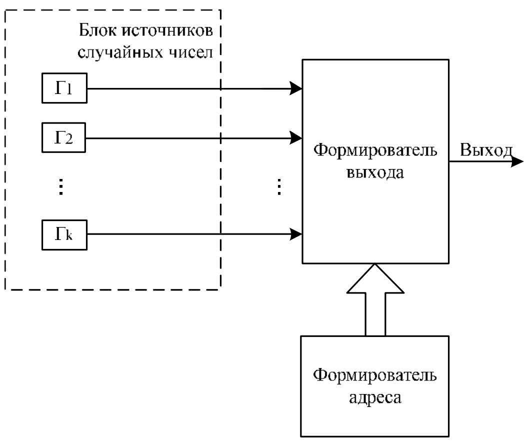

Рис. 3. Структурная схема генератора равномерно распределенных случайных чисел

Работа генератора происходит следующим образом.

Каждый из k независимых первичных генераторов равномерно распределенных случайных чисел *Гi* формирует одно собственное слово разрядности n_i на временном интервале формирования составного слова генератора *T*.

Формирователь выхода образует выходное слово путем конкатенации k слов с выходов k источников. Формирователь адреса служит для управления порядком расстановки слов генераторов в составном результирующем слове.

При рассмотрении источников случайных чисел, выполненных на базе комбинационных автоматов, представляется целесообразным использовать предложенные в [4, 5] идеи выделения одной комбинационной части для всех источников и общего блока памяти. Структурная схема такого устройства представлена на рис. 4.

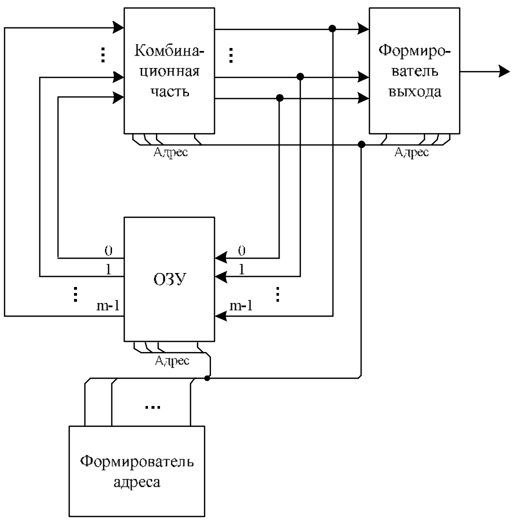

Рис. 4. Структурная схема генератора равномерно распределенных случайных чисел с выделением комбинационной части и памяти

Комбинационная часть выполняет операции, предусмотренные алгоритмом формирования случайных чисел. Заметим, что все известные способы генерации случайных чисел используют итерационный процесс. Поэтому на каждом этапе вычисления числа на вход комбинационной части поступает результат предыдущего вычисления генератора, хранимый в памяти. После вычисления текущего слова оно подается на формирователь выхода и записывается в ячейку памяти в качестве текущего состояния генератора.

Формирователь выхода генерирует выходную последовательность путем конкатенации слов от каждого генератора. Данные с формирователя адреса служат для извлечения из памяти промежуточного состояния генератора, преобразования данных в комбинационной части, а также для формирования составного слова в формирователе выхода.

В простейшем случае формирователь адреса представляет собой циклический счетчик с количеством состояний, равным количеству исходных слов конкатенации. В иных случаях необходимо учитывать как количество генераторов, так и разрядность генерируемых ими слов.

Заметим, что в качестве компонентов конкатенации при формировании равномерно распределенной последовательности случайных чисел можно использовать реализации не нескольких, а одного случайного процесса. Это становится возможным вследствие выполнения условия k-распределения [1] случайной последовательности длины N для всех $k \le \log_b N$, $b = a^n$ (или условия многомерной однородности [6]).

В случае объединения k_0 слов генератора в одно слово для равномерного распределения конкатенации достаточно, чтобы поток слов генератора обладал свойством k_0-мерной однородности. Однако чтобы последовательность полученных составных слов удовлетворяла условию случайности по [1] и была равномерно распределена в $\log_{b^{k_0}} \dfrac{N}{k_0}$ измерениях, необходимо, чтобы исходная последовательность длины N была k-распределена для всех $k \le \log_{b^{k_0}} \dfrac{N}{k_0} + k_0$.

Отметим также, что в случае использования единственного первичного потока псевдослучайных чисел может возникнуть проблема увеличения периода исходной последовательности, что является отдельным направлением исследований.

Полученные результаты: Разработанный метод позволяет формировать равномерно распределенную случайную последовательность чисел разрядности $\sum_{i=0}^{k-1} n_i$ на основе конкатенации k слов разрядности n_i $\left(i \in [0, k-1]\right)$ от независимых источников равномерно распределенных случайных чисел.

При равномерном распределении чисел на выходе i-го первичного генератора на отрезке $[A_i, B_i]$ $\left(i \in [0, k-1]\right)$ конкатенация чисел с выходов генераторов равномерно распределена на множестве чисел диапазона

$$\bigcup_{l_{k-1}=A_{k-1}}^{B_{k-1}} \mathbf{K} \bigcup_{l_2=A_2}^{B_2} \bigcup_{l_1=A_1}^{B_1} \left[\begin{matrix} l_{k-1}a^{p_{k-1}} + \mathbf{K} + l_2 a^{p_2} + l_1 a^{p_1} + A_0, \\ l_{k-1}a^{p_{k-1}} + \mathbf{K} + l_2 a^{p_2} + l_1 a^{p_1} + B_0 \end{matrix}\right]$$

где a – основание системы счисления,

$$p_m = \sum_{l=0}^{m-1} n_l \ \ (m > 0),\ p_0 = 0 .$$

Таким образом, в зависимости от разрядностей слов исходных источников случайных чисел и диапазонов их распределения можно легко определить разрядность и диапазон равномерного распределения конкатенации этих слов. Указанную операцию можно произвести и в обратном направлении – в зависимости от требуемых значений разрядности и диапазона распределения составных чисел можно определить параметры исходных последовательностей чисел.

Конкатенация Z случайных величин X_i обладает следующими характеристиками:

– математическое ожидание

$$M(Z)=\sum_{i=0}^{k-1}\left(\frac{A_i+B_i}{2}\cdot a^{p_i}\right);$$

– дисперсия

$$s^2(Z)=\frac{1}{12}\sum_{i=0}^{k-1}\left(\left(N_i^2-1\right)\cdot a^{2p_i}\right),$$

$N_i=B_i-A_i+1$;

– функция распределения

$$F(z)=\frac{1}{N_0N_1\ldots N_{k-1}}\left(\sum_{i=1}^{k-1}\frac{|z'|_{a^{p_{i+1}}}-|z'|_{a^{p_i}}}{a^{p_i}}\cdot N_i+|z'|_{a^{p_1}}\right),$$

$$z'=z-\sum_{j=0}^{k-1}A_i\cdot a^{p_j};$$

– закон распределения

$$P(Z=z_i)=\frac{1}{N_0N_1\ldots N_{k-1}}.$$

Заметим, что важными требованиями для первичных источников случайных чисел в разработанном методе являются требования независимости и равномерного распределения формируемых ими последовательностей. Поэтому представляется целесообразным отдельно исследовать статистические свойства последовательности конкатенаций слов первичных ГСЧ, выполненных на базе комбинационных автоматов.

Выводы: Выполненное исследование позволяет сформулировать следующие выводы:

– разработан метод формирования равномерно распределенной случайной последовательности чисел, основанный на конкатенации слов, образованных независимыми первичными генераторами равномерно распределенных случайных чисел, который позволяет без уменьшения продуктивности генератора формировать последовательность равномерно распределенных случайных чисел с разрядностью, выходящей за пределы разрядности вычислительной платформы первичных генераторов;

– определены для любой используемой системы счисления математическое ожидание, дисперсия, функция распределения и диапазон равномерного распределения конкатенации чисел, формируемых независимыми первичными генераторами равномерно распределенных случайных чисел, в зависимости от их разрядности и диапазона распределения;

– разработана структурная схема генератора равномерно распределенных случайных чисел, сформированных путем конкатенации чисел меньшей разрядности, порожденных независимыми первичными генераторами равномерно распределенных случайных чисел;

– разработана структурная схема генератора равномерно распределенных случайных чисел, сформированных путем конкатенации чисел меньшей разрядности, порожденных независимыми первичными генераторами, которая отличается выделением комбинационной части и памяти, что позволяет повысить эффективность используемого оборудования.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Кнут Д. Э. Искусство программирования / Дональд Э. Кнут. – М.: Вильямс, 2007. – Том 2. Получисленные алгоритмы. – 832 с.
2. Иванов М. А. Теория, применение и оценка качества генераторов псевдослучайных последовательностей / М. А. Иванов, И. В. Чугунков. – М.: КУДИЦ-ОБРАЗ, 2003. – 240 с. – (СКБ – специалисту по компьютерной безопасности).
3. Митянкина Т. В. Рандомизация последовательности конгруэнтных чисел / Т. В. Митянкина, В. В. Швыдкий, А. И. Щерба // Вестник Инженерной академии Украины. – 2008. – № 2. – С. 107–111.
4. Швыдкий В. В. Групповая обработка сигналов многоканальными конечными автоматами. / В. В. Швыдкий // Техника средств связи: сб. – 1978. – Вып. 4 (25). – С. 58–64. – (Серия ТПС).
5. Фауре Э. В. Устройство формирования остатков в многоканальных помехоустойчивых кодеках / Э. В. Фауре, Д. В. Фауре, Р. О. Бивзюк // Вісник Хмельницького національного університету. – 2010 – № 4. – С. 75–78.
6. Marsaglia G. A Current View of Random Number Generators / G. Marsaglia // Computer Science and Statistics: Proceedings of the Symposium on the Interface, 16th, Atlanta, Georgia, March 1984 (Ed. L. Billard). – New York: Elsevier, 1985.

Стаття надійшла до редакції 07.08.2012.