УДК 004.93:004.4’244

ПОСТПРОЦЕССИНГОВАЯ ОБРАБОТКА ДАННЫХ С АППАРАТНОГО ГЕНЕРАТОРА СЛУЧАЙНЫХ ЧИСЕЛ

ПИКУЗА М. О.

Белорусский государственный университет информатики и радиоэлектроники
(г. Минск, Республика Беларусь)

E-mail: maksimpikuza@gmail.com

Аннотация. Рассмотрены и программно реализованы методы постпроцессинговой обработки, такие как метод исключающего ИЛИ, метод фон Неймана, метод H-функции, метод S-блоков, метод конечных разностей. Проведено тестирование с использованием наборов статистических тестов NIST исходной и программно обработанных последовательностей с опытного образца квантового генератора случайных чисел. Показаны результаты тестирования и сделан вывод о наиболее эффективном методе постпроцессинговой обработки.

Abstract. Methods of post-processing processing, such as the XOR method, the von Neumann method, the H-function method, the S-box method, and the finite difference method, are considered and software implemented. Testing was carried out using sets of NIST statistical tests of the original and software-processed sequences from a prototype quantum random number generator. The results of testing are shown and a conclusion is made about the most effective method of post-processing processing.

Введение

В качестве источника случайности используют аппаратные генераторы случайных чисел (ГСЧ). Основой случайности в таких генераторах являются хаотически изменяющиеся параметры физических процессов, такие как тепловой и квантовый шум. Воздействие внешних факторов может значительно ухудшить статистические характеристики аппаратных ГСЧ, что ограничивает сферы его использования. Для улучшения статистических характеристик ГСЧ используются различные методы постпроцессинговой обработки [1].

Методы постпроцессинговой обработки данных с ГСЧ

Существует множество методов постобработки, которые позволяют улучшить статистические характеристики случайной последовательности, полученной с аппаратного ГСЧ. Применение методов постобработки приближают энтропию к идеальному значению и позволяет устранить смещение распределения. Недостатком большинства методов постобработки является уменьшение длины выходной последовательности. Рассмотрим некоторые из методов постпроцессинговой обработки.

Метод исключающего ИЛИ (XOR). Входной поток случайных чисел разбивается на блоки по два бита, после чего над каждым блоком проводится операция исключающее ИЛИ, результат которой записывается в выходной поток.

Метод фон Неймана (Von Neumann). Входной поток случайных чисел разбивается на блоки по два бита. Если блок равен 01, то в выходной поток записывается 0, если блок равен 10, то в выходной поток записывается 1, если блок равен 00 или 11, то в выходной поток не записывается ничего.

Метод H-функции (H function). Входной поток случайных чисел разбивается на блоки по 16 бит, каждый блок разбивается на два байта: A1 и A2. Далее применяется H-функция, которая вычисляется как: $H(A1, A2) = A1 \oplus RL(A1, 1) \oplus A2$, где $\oplus$ - операция исключающего ИЛИ, RL(A1, 1) - операция циклического сдвига влево на 1 байта A1. Результат выполнения H-функции размером 8 бит записывается в выходной поток в двоичном виде.

Метод S-блоков (S-box). Входной поток случайных чисел разбивается на блоки по 48 бит, каждый блок разбивается на 8 частей по 6 бит. Каждой из частей соответствует своя заранее заданная таблица (S1-S8) из 4-х строк (0-3) и 16-ти столбцов (0-15), в каждой из ячеек таблицы содержатся числа от 0 до 15, при чем в строках числа не повторяются. Из 6-ти бит каждой части

получается необходимая позиция в таблице согласно следующему принципу: номер строки образуется из старшего и младшего бита (5-й, 0,й), номер столбца образуется из оставшихся 4-х бит (4-й, 3-й, 2-й, 1-й). Число из полученной ячейки размером 4 бита записывается в выходной поток в двоичном виде [2].

Метод конечных разностей (Finite differences). Входной поток случайных чисел разбивается на блоки по 8 бит., т.е. по байтам. Каждый байт преобразуется в число с плавающей запятой. Далее производится вычисление конечной разности 47-го над полученной последовательностью байтов. Результат вычисления преобразуется в положительное целое число и представляется в двоичном виде. От этого двоичного числа берется 45 младших бит и записывается в выходной поток [3].

Проверка эффективности методов постпроцессинговой обработки

Для проверки эффективности методов постпроцессинговой обработки было написано специализированное программное обеспечение, реализующее рассмотренные ранее методы постобработки. Данная программа позволяет открыть файл с исходной последовательностью случайных чисел, применить выбранный метод постобработки и сохранить полученную последовательность чисел в новый файл.

В качестве источника случайной последовательности был взят опытный образец квантового ГСЧ. Данный ГСЧ работает следующим образом. Светодиод под воздействием импульсов низкой интенсивности в результате квантовых процессов генерирует поток фотонов. Эти фотоны регистрируются кремниевым фотоэлектронным умножителем, на выходе которого формируется аналоговый сигнал пропорционально количеству зарегистрированных фотонов. Далее аналоговый сигнал с помощью порогового дискриминатора и делителя частоты преобразуется в двухуровневый цифровой шум. Полученная последовательность двоичных чисел объединяется в байты и отправляется на ПЭВМ.

В ходе проверки с ГСЧ была получена и записана в файл последовательность случайных чисел. К полученной последовательности применялись рассматриваемые методы постобработки. Образованные после применения методов новые последовательности чисел записывались в отдельные файлы. Размер файлов и процент отношения размеров новых файлов к исходному указаны в таблице 1.

Таблица 1. Размеры исходного и преобразованного файлов

Метод обработки	Размер файла (байт)	Отношение к исходному размеру
Hardware RNG	11109571	100%
XOR	5554785	50%
Von Neumann	2012567	18,1%
H function	5554785	50%
S-box	7406380	66,7%
Finite differences	62491072	562,5%

Для сравнения статистических характеристик исходной последовательности и последовательностей после применения методов постобработки использовался набор статистических тестов от NIST, позволяющий исследовать различные типы отклонения от случайности [4]. Результаты тестирования файлов с последовательностями приведены в таблице 2.

Таблица 2. Результаты тестирования методов постобработки тестами NIST

Тесты \ Методы обработки	Hardware RNG	XOR	Von Neumann	H function	S-box	Finite differences
Frequency				+		+
Block Frequency						+
Cumulative Sums				+		+
Runs	+	+	+		+	+
Longest Run				+		+
Rank				+		+
FFT				+		
Non Overlapping Template						+
Overlapping Template						+
Universal						+
Approximate Entropy						
Random Excursions				+		+
Random Excursions Variant				+		+
Serial						
Linear Complexity	+	+	+	+	+	+
Σ	**2**	**2**	**2**	**8**	**2**	**12**

В таблице указаны результаты проведения тестов NIST (в столбце 1 названия тестов) над исходной последовательностью с ГСЧ (столбец 2) и последовательностями, полученными разными методами постобработки (столбцы 3-7). Успешно пройденные тесты отмечены символом «+» в соответствующих строках. В последней строке указана сумма всех успешно пройденных тестов для конкретного метода постобработки.

Из результатов тестирования видно, что большинство рассматриваемых методов уменьшают длину выходной случайной последовательностей относительно исходной. Исключением является метод конечных разностей, после применения которого длина последовательности увеличилась. Это связано с тем, что из каждых входных 8 бит алгоритм извлекает 45 бит. Также из результатов видно, что только применение метода H-функции и метода конечной разности показали увеличение количества пройденных статистических тестов, что говорит об улучшении статистических характеристик выходной последовательности. Остальные методы не смогли улучшить характеристики ГСЧ, т.к. исходная последовательность имеет достаточно плохие статистические характеристики, что связано с воздействием внешних факторов на ГСЧ.

Заключение

В ходе изучения возможности улучшения статистических характеристик аппаратного ГСЧ при помощи постпроцессинговой обработки было установлено, что наиболее эффективными методами являются метод H-функции и метод конечной разности. При этом после применения метода конечной разности длина последовательности увеличилась более чем в 5 раз, а количество пройденных статистических тестов возросло с 2 до 12. Это говорит о том, что данный метод является эффективным и его можно применять для улучшения статистических характеристик различных аппаратных ГСЧ.

Список использованных источников

1. Herrero-Collantes M. Quantum Random Number Generators / M. Herrero-Collantes, J.C. Garcia-Escartin // Reviews of Modern Physics. – 2017. – №89(1).
2. Avaroglu E., Tuncer T. A novel S-box-based postprocessing method for true random number generation / E.Avaroglu, T. Tuncer // Turk J Elec Eng & Comp Sci. – 2020. – №28. – P. 288-301.
3. Chizhevsky V.N. Symmetrization of single-sided or non-symmetrical distributions: the way to enchance a generation rate of random bits from a physical source of randomness / V.N. Chizhevsky // Phys. Rev. E. – 2010. –№82(5).
4. A Statistical Test Suitefor Random and Pseudorandom Number Generators for Cryptographic Applications / National Institute of Standards and Technology. – Gaithersburg, Maryland, 2010.