УДК 004.312

СРАВНЕНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СЛУЧАЙНЫХ ЧИСЕЛ, ФОРМИРУЕМЫХ ПРИ ПОСТОБРАБОТКЕ СХЕМАМИ НА БАЗЕ МНОГОКАНАЛЬНОГО СИГНАТУРНОГО АНАЛИЗАТОРА И SHA-1

Петровский Д.А., Иванюк А.А.

Белорусский государственный университет информатики и радиоэлектроники, Минск, Республика Беларусь, petrovky.dmitr@gmail.com

Аннотация: проведено сравнение характеристик последовательностей случайных чисел, прошедших постобработку схемами на базе модифицированного многоканального сигнатурного анализатора и криптографической хеш-функцией SHA-1. Показано, что модифицированная схема обеспечивает более высокую энтропию по Шеннону и успешно проходит тесты NIST 800-22 при разрядности 48–512 бит, не уступая SHA-1 по доле пройденных тестов. При этом она требует на 15% меньше LUT-блоков, в 3 раза меньше триггеров и обладает в 5 раз большей расчетной тактовой частотой при реализации на программируемой логической интегральной схеме xc7z010-1clg400c-1. Результаты подтверждают перспективность схемы для энергоэффективных генераторов истинно случайных чисел в устройствах интернета вещей.

Ключевые слова: случайные данные, постобработка случайных чисел, многоканальный сигнатурный анализатор.

I. ВВЕДЕНИЕ

Генераторы истинно случайных чисел (TRNG, от англ. True Random Number Generator) являются фундаментальным компонентом современных криптографических систем, протоколов безопасного обмена ключами и IoT-устройств. Проектирование TRNG регламентируется, в частности, рекомендациями National Institute of Standards and Technology (NIST), которые определяют его структуру как комбинацию источников энтропии (ES), блока тестирования (Testing Unit) для каждого независимого однобитного ES и схемы постобработки (Post-Processing Unit). Обобщённая структура TRNG показана на рисунке 1.

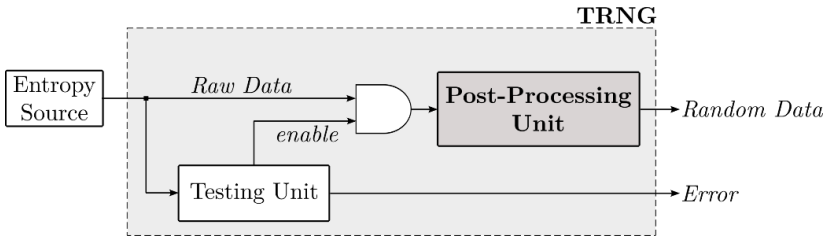


Рисунок 1. Обобщенная структура одноканального TRNG

Ядром TRNG является ES, физические процессы которого часто обладают смещённым распределением или остаточной предсказуемостью. Традиционно в качестве схем постобработки используются алгоритмические корректирующие схемы (корректор Фон Неймана, XOR-корректор, LFSR-структуры) и криптографические примитивы (хеш-функции, блочные шифры) [1]. Последние, в частности SHA-1, обеспечивают высокую энтропию и успешное прохождение тестов NIST 800-22, однако их аппаратная сложность, значительные затраты на кристалле и высокое энергопотребление делают их непригодными для мобильных и IoT-устройств. Базовые LFSR-архитектуры (SISA, от англ. Single-Input Signature Analyzer, MISA, от англ. Multi-Input Signature Analyzer) отличаются компактностью, но уступают в статистическом качестве при обработке последовательностей с низкой энтропией.

Актуальной задачей является проектирование схемы постобработки, сочетающей криптографическое качество с минимальными аппаратными затратами. В работе рассматривается модифицированная архитектура MISA с коэффициентом временного сжатия и механизмом пространственного эквивалента временной пересинхронизации. Будем называть такую схему MISA*. Цель исследования – сравнительный анализ MISA* с эталонной реализацией на базе SHA-1 по энтропии Шеннона и результатам тестов NIST 800-22.

II. ОПИСАНИЕ СХЕМЫ ПОСТОБРАБОТКИ

Базовая MISA [2] – многовходовой регистр сдвига с линейной обратной связью (LFSR, от англ. Linear Feedback Shift Register) для параллельной обработки g независимых однобитных потоков с

формированием n-разрядного слова за такт. Классические конфигурации требуют выполнения условия $g = n$, что предполагает эквивалентное число ES и блоков тестирования. В архитектуре MISA* намеренно введено условие $g < n$, что снижает аппаратные затраты и упрощает топологию подключения ES. Однако при $g < n$ возникает побочный эффект: при последовательном сжатии старшие разряды выходного слова заполняются младшими разрядами предыдущего такта, что дублирует символы и снижает энтропию. Для устранения дублирования применяется пересинхронизация: тактовая частота MISA в j раз выше частоты обновления данных от ES. Параметр j – коэффициент сжатия во времени: за период одного входного символа схема выполняет j тактов, многократно применяя матрицу связи и обратной связи к одному значению, что устраняет линейные корреляции. При аппаратной реализации непрерывная пересинхронизация избыточна. Работа схемы формализуется матричным рекуррентным уравнением состояния LFSR. Замена временных итераций пространственным преобразованием переводит модель в схему, выполняющую эквивалентные вычисления за один такт, за счёт роста числа итераций с g до $j \cdot g$ и циклического обращения к входным каналам (по модулю g). Это исключает дополнительные регистры синхронизации и буферы задержки, а вычислительная модель сводится к комбинационной логике. Поведение модифицированной схемы описывается выражением:

$$\mathbf{D}(k+1) = \mathbf{C}^{j \cdot g} \cdot \mathbf{D}(k) \oplus \left[\sum_{i=1}^{g \cdot j} \left(\mathbf{C}^{j \cdot g} \cdot \mathbf{X}_{i \bmod g}(k) \right) \right],$$

$$\mathbf{C} = \begin{bmatrix} \alpha_1 & a_2 & \cdots & a_n \\ \hline & & & \\ \multicolumn{4}{c}{\mathbf{I}_{n-1}} & \begin{matrix}0\\ \vdots \\ 0\end{matrix} \end{bmatrix}; \mathbf{D}(k) = \begin{bmatrix} d_1(k) \\ d_2(k) \\ \vdots \\ d_n(k) \end{bmatrix}; \mathbf{X}_i(k) = \begin{bmatrix} x_i(k) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (1)$$

где $\mathbf{C}$ – матрица сдвига и обратной связи LFSR, включающая коэффициенты характеристического полинома α_i (где $i = \overline{1,n}$); $\mathbf{D}(k)$ – n-разрядное двоичное слово, являющееся состоянием MISA* в k-й момент времени; x_i – значение символа на входе MISA* в k-й момент времени, которое можно представить в виде вектора $\mathbf{X}_i(k)$.

MISA* и SHA-1 различаются типом преобразований. Постобработка SHA-1 основана на нелинейных криптографических операциях (циклические сдвиги, сложение по модулю, логические функции) и каскадной или конвейерной структуре: это даёт высокую стойкость, но требует значительных аппаратных затрат. MISA* использует только линейные операции над фиксированным примитивным неприводимым полиномом, что обеспечивает архитектурную простоту и высокую тактовую частоту, но требует строгого контроля параметров g, n и j для достаточной энтропии. Далее в работе количественно оценивается, как выбранное соотношение параметров MISA* позволяет достичь статистического качества, сопоставимого с нелинейным хешированием SHA-1.

III. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В экспериментальной части было проведено сравнение двух подходов к постобработке случайных последовательностей: на основе MISA* и с использованием криптографической хэш-функции SHA - 1 [3]. Эффективность оценивалась по энтропии Шеннона и набору статистических тестов NIST 800-22. Исходными (необработанными) данными служили три набора: CIFAR-10 (цветные RGB-изображения, преобразованные в битовый поток; исходная энтропия 0,9889), Arbiter PUF (данные схемотехнической реализации ФНФ типа арбитр; исходная энтропия 0,9454) и BAD (программно сгенерированные данные; исходная энтропия 0,2476)

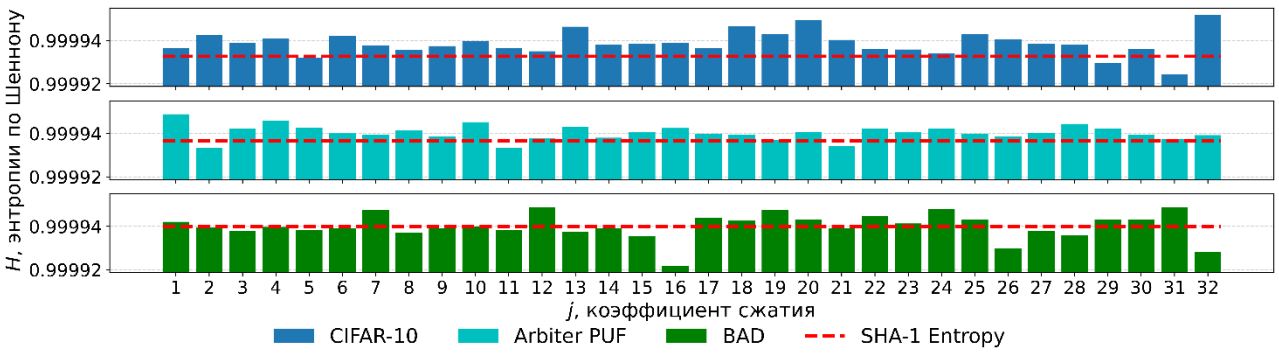


Рисунок 2. Энтропия обработанных последовательностей схемой на базе MISA* с различным j

На первом этапе сравнивались значения энтропии Шеннона для схемы MISA* и SHA-1 при идентичных параметрах: произвольное начальное состояние, число входных каналов $g = 32$, разрядность схемы $n = 512$ и разрядность обрабатываемого слова 32. Для схемы MISA* применялся порождающий полином $\varphi(x) = x^{512} + x^{510} + x^{507} + x^{504} + 1$ с коэффициентом сжатия $j = \overline{1{,}32}$.

Как следует из рисунка 2, на наборах данных с высокой энтропией CIFAR-10 и Arbiter PUF, схема на базе MISA*, при сопоставимых со схемой SHA-1 параметрах, обеспечивает более высокое численное значение энтропии обработанных последовательностей независимо от коэффициента сжатия. На наборе с искусственно низкой энтропией BAD, схема на базе MISA* показала сравнимый результат с SHA-1.

В завершение было проведено сопоставление результатов набора статистических тестов NIST 800–22 для данных, обработанных программными моделями схем MISA* с различной разрядностью, а также схемой SHA-1. Тестированию подвергались обработанные последовательности длиной $8x10^7$ бит, разбитые на 80 подпоследовательностей по 10^6 бит каждая. Данные были получены с помощью программного стенда для экспериментов с набором CIFAR-10. Итоги статистических тестов приведены в таблице 1.

Таблица 1. Результаты NIST тестирования последовательностей, обработанных схемами на базе SHA-1 и MISA* с различной разрядностью

Название теста	P-value						
	Исходные данные	MISA*					SHA-1
		48	64	128	256	512	
Frequency	н/п	0,1056	0,4372	**0,5091***	0,4145	0,2133	0,0449
BlockFrequency	н/п	0,6890	0,8119	0,1742	0,3306	0,2932	**0,8555***
CummulativeSums	н/п	0,3115	0,7399	0,3924	0,5341	0,1315	**0,8755***
Runs	н/п	0,6371	0,2429	0,6371	0,7146	**0,7887***	0,4145
LongestRun	н/п	0,1056	0,4606	0,7646	0,7399	**0,9535***	0,0032
Rank	н/п	0,7146	0,6890	**0,8555***	0,3924	0,6371	0,4846
FFT	н/п	0,0043	0,1137	0,1223	0,7146	**0,7887***	0,4372
NonOverlapingTemplate	н/п	0,7146	0,9535	0,8755	0,8555	**0,9733***	0,8555
OverlapingTemplate	н/п	0,6371	**0,7646***	0,3504	0,2429	0,3924	0,4606
Universal	н/п	**0,8555***	0,0274	0,4846	0,1626	0,6890	0,8119
ApproximateEntropy	н/п	0,5091	**0,9411***	0,6111	0,5852	0,7887	0,5595
RandomExcursions	н/п	0,1742	**0,8513***	0,4372	0,8343	0,3041	0,8343
RandomExcursionsVariant	н/п	0,3306	0,8165	0,5852	0,1056	0,4372	**0,8343***
Serial	н/п	**0,9914***	0,8119	0,4846	**0,9914***	0,9270	0,9270
LinearComplexity	0,000012	0,6631	0,6111	0,3115	**0,9114***	0,0138	0,2932
Обозначение – н/п – тест не пройден							
Обозначение – ***** – наибольшее значение P-value в строке							

Как следует из таблицы 1, все схемы постобработки обеспечили улучшение статистических характеристик выходных последовательностей относительно исходных данных набора CIFAR-10. По доле успешно пройденных тестов для обработанных последовательностей схемы MISA* не уступают схеме SHA-1. Для всех рассмотренных разрядностей MISA* тестирование завершилось успешно, что свидетельствует о равномерности распределения формируемой последовательности, а также об отсутствии в ней предсказуемых шаблонов и автокорреляций.

В процессе исследований оценивались аппаратные затраты схем постобработки на программируемых логических интегральных схемах (ПЛИС) (табл. 2). Для схемs MISA* было разработано параметризуемое описание на HDL-языке SystemVerilog. Сравнение не включало сопутствующие аппаратные затраты на блоки тестирования.

Таблица 2. Оценка аппаратных затрат схем TRNG для кристалла ПЛИС xc7z010-1clg400c-1

Тип постобработки	Аппаратные затраты		Максимальная частота
	LUT-блоков	Триггеров	МГц
MISA*	1352	512	502
SHA-1	1553	1531	108

Анализ данных табл. 2 показывает, что схема MISA* требует меньше аппаратных затрат ПЛИС по сравнению с SHA-1, LUT-блоков меньше на 15%, триггеров меньше в три раза и расчетная максимальная тактовая частота в пять раз выше. Предполагается, что данная схема при равной тактовой частоте будет обладать наименьшим энергопотреблением из-за меньшего использования аппаратных ресурсов. Это делает ее подходящей для применений, требующих энергоэффективности и компактности, таких как простые схемы TRNG в устройствах интернета вещей.

IV. ЗАКЛЮЧЕНИЕ

В работе проведено сравнение последовательности случайных чисел после двух схем постобработки: MISA* и эталона на SHA-1. Эксперимент показал, что при равных параметрах MISA* даёт более высокую энтропию по Шеннону чем SHA-1. Тесты NIST 800-22 показали сопоставимое качество обеих схем: все конфигурации MISA* (48–512 бит) прошли испытания, тогда как исходные данные не прошли ни одного. Ранее выявленное ограничение — провал теста Rank [4] при разрядности менее 32 бит — устранено увеличением разрядности. Оценка аппаратных затрат для ПЛИС: MISA* требует на 15 % меньше LUT - блоков, втрое меньше триггеров и даёт в пять раз более высокую расчётную тактовую частоту по сравнению с SHA-1. Это делает её перспективной для энергоэффективных TRNG, включая устройства интернета вещей.

БЛАГОДАРНОСТЬ

Авторы выражают искреннюю благодарность резиденту ПВТ компании «Инженерный Центр Ядро», которая является одним из центров разработки YADRO, за предоставленное оборудование для проведения экспериментов в рамках работы совместной учебной лаборатории с Белорусским государственным университетом информатики и радиоэлектроники.

ЛИТЕРАТУРА

[1] Fischer, V. True Random Number Generators in FPGAs / V. Fischer // Security Trends for FPGAs: From Secured to Secure Reconfigurability / B. Badrignans, J.-L. Danger, V. Fischer [et al.]. – Dordrecht : Springer Netherlands, 2011. – P. 101–135. – DOI: 10.1007/978-94-007-1338-3_5.

[2] Ярмолик, В. Н. Построение многоканальных сигнатурных анализаторов / В. Н. Ярмолик // Автоматика и телемеханика. – 1985. – № 1. – С. 127–132

[3] Jun, B. The Intel Random Number Generator / B. Jun, P. Kocher // Cryptograph Research Inc., White Paper. – 1999. – Vol. 27. – P. 1–8

[4] Петровский, Д. А. Анализ характеристик схемы постобработки последовательности случайных чисел на основе многоканального сигнатурного анализатора / Д. А. Петровский, А. А. Иванюк // Информационные технологии и системы 2024 (ИТС 2024) = Information Technologies and Systems 2024 (ITS 2024) : материалы международной научной конференции, Минск, 20 ноября 2024 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2024. – С. 93–94.

COMPARISON OF RANDOM NUMBER SEQUENCES GENERATED VIA POST-PROCESSING USING MULTI-CHANNEL SIGNATURE ANALYZER AND SHA-1 BASED SCHEMES

D. Petrovsky, A. Ivaniuk

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus,
petrovky.dmitr@gmail.com

Abstract: a comparison was conducted between the characteristics of random number sequences subjected to post-processing using schemes based on a modified multi-channel signature analyzer and the SHA-1 cryptographic hash function. It is shown that the modified scheme yields higher Shannon entropy and successfully passes NIST 800-22 tests for bit widths ranging from 48 to 512 bits, matching SHA-1 in terms of the proportion of tests passed. Furthermore, when implemented on the xc7z010-1clg400c-1 FPGA, the scheme requires 15% fewer LUTs and three times fewer flip-flops, while achieving a five times higher estimated clock frequency. These results demonstrate the scheme's potential for use in energy-efficient true random number generators within Internet of Things (IoT) devices.

Keywords: random data, random number post-processing, multi-input signature analyzer.