Об экстракции случайности в квантовых генераторах

к.ф.-м.н. Миронкин Владимир Олегович
НИУ ВШЭ

Суперкомпьютерные дни в России

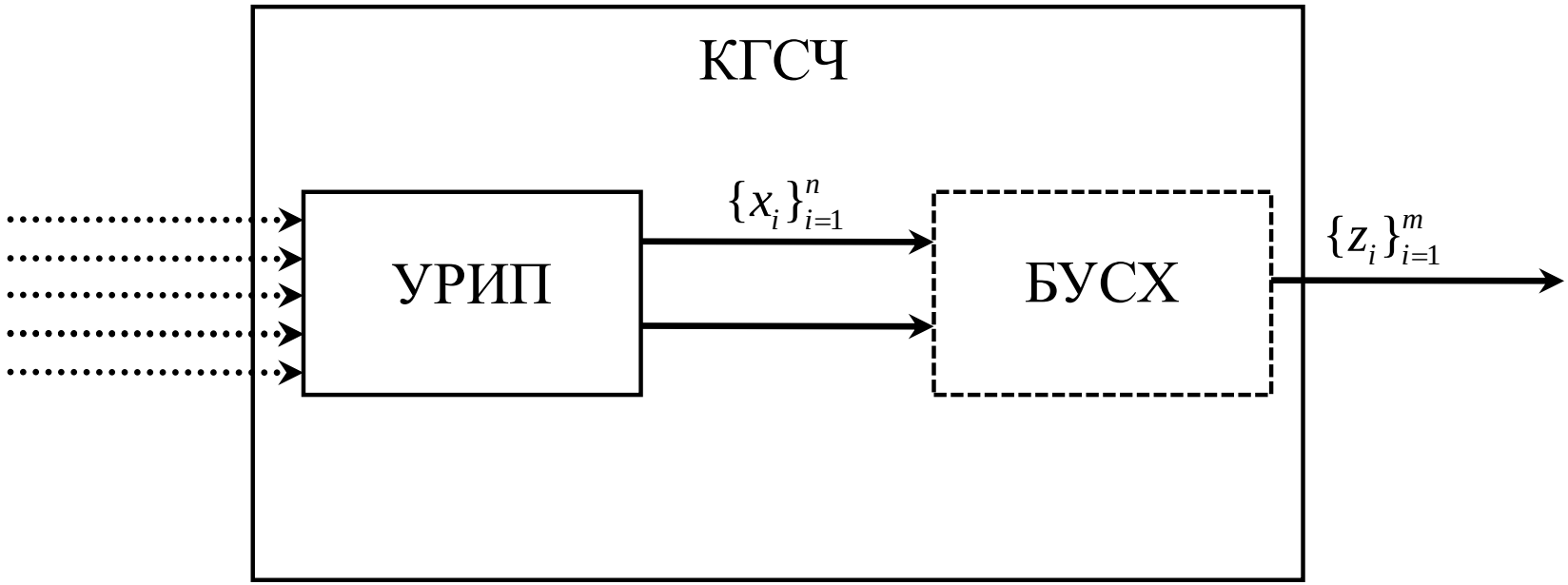
27 сентября 2022 г.

Принцип формирования случайности КГСЧ

Квантовый генератор случайных чисел (КГСЧ) формирует 2 типа данных:

- ***исходная*** (сырая) двоичная последовательность $\{x_i\}_{i=1}^{n}$, $n \in \mathbb{N}$, – набор дискретных значений, соответствующих результатам регистрации излучения фотонов с использованием матрицы фотодетекторов;
- ***выходная*** двоичная последовательность $\{z_i\}_{i=1}^{m}$, $m \in \mathbb{N}$, – результат применения алгоритма улучшения статистических характеристик к исходной последовательности.

Общая схема функционирования КГСЧ



Здесь

- УРИП – устройство регистрации исходного процесса;
- БУСХ – блок улучшения статистических характеристик, реализующий алгоритм типа Элайеса, Фон Неймана.

Условия применимости БУСХ указанного типа

1. **На входе БУСХ:**

   исходная последовательность $\{x_i\}_{i=1}^{n}$ представляет собой реализацию схемы Бернулли с вероятностью успеха $p \in (0, 1)$

2. **На выходе БУСХ:**

   выходная последовательность $\{z_i\}_{i=1}^{m}$ представляет собой реализацию равновероятной схемы Бернулли

Примеры БУСХ

- *детерминистические экстракторы;*
- *алгоритм Фон Неймана;*
- *алгоритм Элайеса.*

Поаккуратнее с энтропией!

В условиях стационарности исходной последовательности и, как следствие, выходной последовательности уместно говорить об энтропии двух дискретных процессов $H_{\overline{x}}$ и $H_{\overline{z}}$.

БУСХ предназначен для «выжимки» энтропии.

Но!!!

Вывод о качестве $\{z_i\}_{i=1}^m$ нельзя делать на основании оценки $H_{\overline{z}}$ даже при очень большом объеме. Зубков А.М. показал, что линейные рекурренты и периодические последовательности могут иметь энтропию близкую к 1.

Зубков А.М., "Энтропия как характеристика качества случайных последовательностей", Матем. вопр. криптогр., 12:3 (2021), 31–47.

Механизмы контроля случайности

Замечание

В зависимости от области применения КГСЧ устанавливаются значений параметров тестирования:

- α – *уровня значимости критериев;*
- n – *длины* $\{x_i\}_{i=1}^{n}$, *определяемой соотношением*

$$n \geqslant 10^{10}.$$

- δ – *максимального допустимого отклонения частоты знака «1» в* $\{z_i\}_{i=1}^{m}$ *от* $\frac{1}{2}$;
- m – *длины* $\{z_i\}_{i=1}^{m}$, *определяемой соотношением*

$$m \geq \max\left\{\left(\frac{t_{1-\frac{\alpha}{2}}}{4\delta}\right)^2, 10^{10}\right\}.$$

Механизмы контроля случайности

Тестирование ***исходной последовательности*** осуществляется в **3** этапа:

1. проверка гипотезы независимости знаков в $\{x_i\}_{i=1}^n$;
2. проверка гипотезы однородности распределения знаков в $\{x_i\}_{i=1}^n$;
3. проверка согласия распределения числа k-грамм в $\{x_i\}_{i=1}^n$ с полиномиальным законом.

Замечание

Этапы 1-3 нацелены на проверку соответствия исходной последовательности условиям применимости БУСХ.

Механизмы контроля случайности

Тестирование ***выходной последовательности*** осуществляется в **4** этапа:

1. проверка соответствия частот знаков в $\{z_i\}_{i=1}^m$ теоретико-вероятностной модели образца КГСЧ;
2. проверка гипотезы независимости знаков в $\{z_i\}_{i=1}^m$;
3. проверка гипотезы однородности распределения знаков в $\{z_i\}_{i=1}^m$;
4. проверка согласия распределения числа k-грамм в $\{z_i\}_{i=1}^m$ с полиномиальным законом.

Замечание

Этапы 1-4 нацелены на проверку статистического качества КГСЧ.

Перспективные направления исследований

Перечень текущих проблем:

1. решения в рамках тестирования как $\{x_i\}_{i=1}^n$, так и $\{z_i\}_{i=1}^m$, принимаются по каждому тесту в отдельности, а не в совокупности;
2. длина $\{x_i\}_{i=1}^n$ существенно зависит от распределения ее знаков – критично при проведении эксплуатационных испытаний КГСЧ;
3. к исходной последовательности $\{x_i\}_{i=1}^n$ предъявляются жесткие требования;
4. в общем случае неизвестно влияние БУСХ на распределение $\{z_i\}_{i=1}^m$;
5. отсутствие рекомендаций по применению тех или иных БУСХ и по работе датчика в целом.

Продолжается борьба разработчиков с нейтрализацией эффектов повторной регистрации сигналов с целью исключения локальных участков «зависимости».

Один из механизмов:

Различные модификации прореживания исходной последовательности $\{x_i\}_{i=1}^n$.

В результате:

Существенное уменьшение производительности КГСЧ!!!

Алгоритмы Фон Неймана и Элайеса являются блочными с размерами блоков 2 и $m \geqslant 2$ соответственно.

Замечание

На практике алгоритм Элайеса использует параметр $m \in \{32, 64\}$ *в зависимости от разрядности процессора.*

Гипотеза

Для алгоритмов Фон Неймана или Элайеса требование реализации схемы Бернулли в качестве $\{x_i\}_{i=1}^{n}$ *завышено.*

Для корректной работы БУСХ оцифрованный исходный процесс может представлять собой последовательность схем серий независимых одинаково распределенных случайных величин

$$\xi_{1,1}, \xi_{1,2}, \dots \xi_{1,m}, \xi_{2,1}, \xi_{2,2}, \dots \xi_{2,m}, \dots, \tag{1}$$

где m *– размер блока алгоритма.*

Одно из направлений решения:

В случае стационарности исходного процесса в последовательности

$$x_1, x_2, x_3, \ldots, x_n$$

сохранить первые $N = m\left[\frac{n}{m}\right]$ элементов и преобразовать ее следующим образом:

$$x_1, x_{\left[\frac{n}{m}\right]+1}, \ldots, x_{(m-1)\left[\frac{n}{m}\right]+1}, \ldots, x_{\left[\frac{n}{m}\right]}, x_{2\left[\frac{n}{m}\right]}, \ldots, x_{m\left[\frac{n}{m}\right]}.$$

Свойства последовательности

- *свойство стационарности не нарушается после перестановки элементов* $\{x_i\}_{i=1}^n$;
- *при глубине зависимости в* $\{x_i\}_{i=1}^n$ *меньшей* $\left[\frac{n}{m}\right]$ *каждый из полученных блоков размера* m *соответствует корректному применению алгоритма Элайеса.*

Проблемы реализации

- *указанная процедура предобработки* $\{x_i\}_{i=1}^n$ *не может быть реализована в режиме Online (но возможны модификации);*
- *требуется хранить последовательность* $\{x_i\}_{i=1}^n$ *целиком;*
- *преобразование не гарантирует отсутствие зависимости блоков в сформированной последовательности.*

Криптографические механизмы с блочной структурой?

Алгоритм Элайеса является составной частью алгоритма оптимального кодирования с блочной структурой.

Предпосылки к применению

- *стандарты блочных алгоритмов шифрования (например, «AES»-128, «Кузнечик»-128);*
- *стандарты хеш-функций (например, «Стрибог»-512, «SHA-3»-512);*
- *латинские квадраты (подход сформулирован д.ф.-м.н. Зубковым Андреем Михайловичем).*

Спасибо за внимание!