Ссылка на статью:
// Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ. Электрон. журн. 2017, № 1, с. 22-28.

Поступила: 06.09.2017
Окончательный вариант: 10.10.2017

УДК 004.05

Методы тестирования генераторов случайных и псевдослучайных последовательностей

Григорьев А. Ю. [1,*]

[*]als73@mail.ru
[1]УлГУ, Ульяновск, Россия

В данной статье рассмотрены существующие методы оценки качества случайных последовательностей бит. Описаны применяемые пакеты статистических тестов, а также новый тест «Стопка книг».

Ключевые слова: *генераторы случайных последовательностей, последовательность бит, статистические тесты.*

Введение

В современном мире большое значение имеют генераторы случайных и псевдослучайных последовательностей бит. Они широко применяются в различных задачах: моделирование. выборочные метод, численный анализ, программирование, криптография. Случайные и псевдослучайные последовательности имеют огромную роль для криптографии – от их качества зависит секретность информации. Поэтому задача создания хороших генераторов и эффективных методов их оценки представляет большой интерес.

В обычной жизни случайные числа можно получить подбрасывая монетку или игральные кости, вытягивая карты из колоды, шары из урны и т.д. Однако в современных компьютерных системах применяются другие методы. Далее представлены существующие варианты генераторов:

- Генератор истинно случайных последовательностей биты (ГСП). Источником случайности является физический процесс. Это может быть шум в электронных компонентах, радиоактивный распад, счетчик физических частиц, атмосферный шум, измеренный радиоприемником и т.д.[1]. Как правило такие генераторы представ-

ляют собой отдельные аппаратные модули для вычислительной машины, хотя некоторые современные компьютеры обладают встроенными аппаратными генераторами. ГСП наиболее часто применяются в криптографии (ещё их называют криптографически стойкими генераторами случайных последовательностей бит - КГСП).

- Генератор псевдослучайных последовательностей (ГПСП). В таких генераторах источником случайности является некий детерминированный алгоритм. В действительности псевдослучайные последовательности вообще не являются случайными. Они вычисляются с помощью алгоритма на основе некоторого начального числа. И зная начальное число, можно предсказать все последующие псевдослучайные числа. Стоит отметить криптографически стойкие ГПСП, которые имеют более жесткие требования. Криптографические генераторы ПСП (КГПСП) создаются с использованием функций поточных шифров, блочных шифров, односторонних функций и блоков стохастического преобразования.
- Комбинированный метод применяется для криптографии. Здесь начальное состояние генератора берётся из ГСП (т.е. из физического источника), а затем при помощи КГПСП формируются случайные последовательности.

К криптографическим генераторам применяются особые требования (в отличии от обычных генераторов): хорошие статистические свойства (формируемые последовательности не должны отличаться от истинно случайных), большой период формируемой последовательности (относится к КГПСП), вычислительно невозможно предсказать предыдущие значения генератора имея фрагмент его показаний, генерируемые последовательности чисел должны быть независимы[2].

В таблице 1 приведены преимущества и недостатки ГСЧ над ГПСП[3].

Таблица 1. Преимущества и недостатки ГСЧ над ГПСЧ

Преимущества	Недостатки
не периодичный	медленный
предсказуемость случайных чисел не основана на знании предыдущих значений	громоздкий для установки и запуска
не существует никаких зависимостей	случайные последовательности не воспроизводимы
высокий уровень безопасности	дорогой
не основан на алгоритмах	возможность влияния на показания

1. Виды тестов

Для проверки качества генераторов применяются различные тесты. В мире нет какого-либо единственного «официального» набора критериев, который бы оценивал, насколько данные случайные последовательности бит применимы именно для конкретной области применения. Существуют различные тесты, которые оценивают, насколько исследуемая последовательность бит «похожа» или «не похожа» на действительно случайную. Методы оценки качества генераторов случайных и псевдослучайных последовательностей можно разделить на две группы:

1) **Графические тесты.** Свойства последовательностей отображаются в виде графических зависимостей, по виду которых делают выводы о свойствах исследуемой последовательности.

К данной категории можно отнести следующие тесты: гистограмма распределения элементов последовательности, распределение на плоскости, проверка на монотонность и т.д.

На рисунке 1 представлен пример графического теста «распределение на плоскости». Слева результат тестирования отрицательный, справа – положительный.

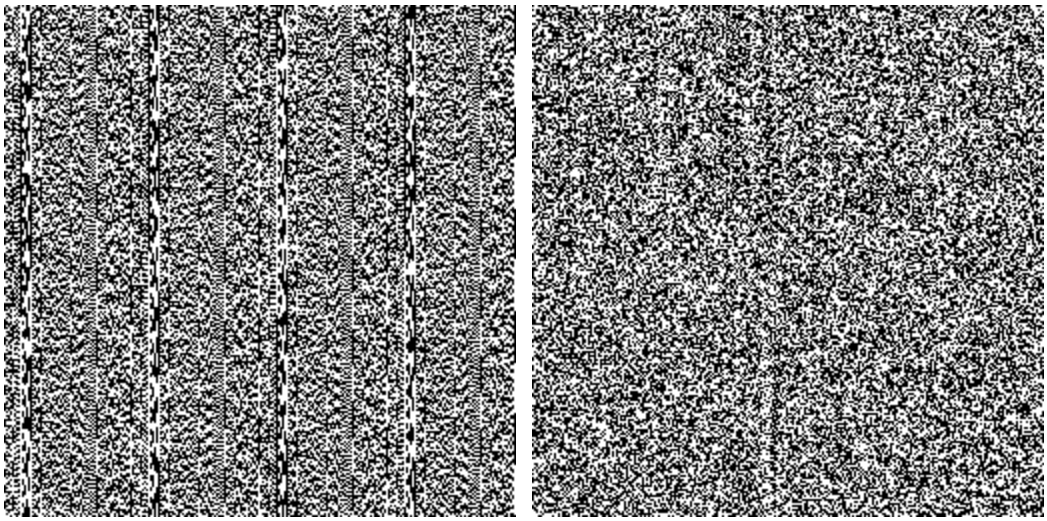

Рис. 1. Пример графического теста «распределение на плоскости»

2) **Статистические тесты.** Статистические свойства последовательностей определяются числовыми характеристиками. На основе оценочных критериев делаются заключения о степени близости свойств анализируемой и истинно случайной последовательностей. В отличие от графических тестов, где результаты интерпретируются пользователями, вследствие чего возможны различия в трактовке результатов, статистические тесты характеризуются тем, что они выдают численную характеристику, которая позволяет однозначно сказать, пройден тест или нет.

2. Статистические тесты

Для тестирования последовательностей на случайность существует большое количество алгоритмов, а для удобства проверки последовательностей уже реализованы программные продукты, содержащие в себе некоторые наборы тестов. Среди них наиболее распространены тесты NIST STS, DIEHARD, CRYPT-X, тесты Д. Кнута и другие.

Тесты Д. Кнута

Одним из первых наборов статистических тестов был предложен Д. Кнутом в 1969 году в его классической работе «Искусство программирования для ЭВМ» [4]. Тесты основаны на статистическом критерии χ^2. Вычисляемое значение статистики χ^2 сравнивается с табличными результатами, и в зависимости от вероятности появления такой статистики

делается вывод о её качестве. Среди достоинств этих тестов — небольшое их количество и существование быстрых алгоритмов выполнения. Недостаток — неопределенность в трактовке результатов.

Тесты Diehard

Тесты Diehard – набор статистических тестов для измерения качества набора случайных чисел. Они были разработаны Джорджем Марсалья в течение нескольких лет и впервые опубликованы в 1995 г. Вместе они рассматриваются как один из наиболее строгих существующих наборов тестов [5].

Тесты Crypt-X

Набор статистических тестов Crypt-X, разработанный исследователями из научно-исследовательского центра по информационной безопасности в технологическом университете Квинсленда в Австралии и является коммерческим пакетом программного обеспечения. Тесты применяются в зависимости от типа алгоритма генератора, соответственно направлены на тестирование генераторов псевдослучайных чисел. Поддерживаются потоковые шифры, блочные шифры и генераторы потока ключей. В набор включены следующие тесты: частотный, на последовательность одинаковых битов, линейная сложность, сложность последовательности, двоичная производная, изменение точки.

Стандарты и тесты NIST

В США был сделан первый шаг к стандартизации набора статистических тестов путем принятия в 1994 г. национального стандарта «Требования безопасности к криптографическим модулям» [6]. Однако требования и методика стандарта больше носят технологический характер. Они направлены на решение задачи статистического контроля псевдослучайных последовательностей, используемых в криптографических модулях, и в общем случае малопригодны к решению задачи исследования статистических свойств генераторов.

В 1999 г. специалистами NIST (Национальный институт стандартов и технологий (НИСТ) США), в рамках проекта AES (Advanced Encryption Standard) был разработан набор статистических тестов «NIST STS» (NIST Statistical Test Suite) [7] и предложена методика проведения статистического тестирования ГСП (ГПСП), ориентированных на использование в задачах криптографической защиты информации, которая, на взгляд многих специалистов в данной области, на настоящий момент наилучшим образом отвечает потребностям всех заинтересованных сторон. Пакет NIST STS включает в себя 15 статистических тестов, которые разработаны для проверки гипотезы о случайности двоичных последовательностей произвольной длины, порождаемых ГСП или ГПСП. Все тесты направлены на выявление различных дефектов случайности. В таблице 2 приведён список тестов и определяемый дефект.

Таблица 2. Список тестов NIST STS

Название теста	Определяемый дефект
Частотный тест	Слишком много нулей или единиц
Частотный тест в блоках	Слишком много нулей или единиц в блоках
Проверка кумулятивных сумм	Слишком много нулей или единиц в начале последовательности
Проверка «дырок» в подпоследовательностях	Отклонение в распределении последовательностей единиц
Проверка «дырок»	Большое (малое) число подпоследовательностей нулей и единиц свидетельствует, что колебание потока бит слишком быстрое (медленное)
Проверка рангов матриц	Отклонение распределения рангов матриц от соответствующего распределения для истинно случайной последовательности, связанное с периодичностью последовательностей
Проверка непересекающихся шаблонов	Непериодические шаблоны встречаются слишком часто
Проверка пересекающихся шаблонов	Слишком часто встречаются m-битные последовательности единиц
Универсальный статистический тест Маурера	Сжимаемость (регулярность) последовательности
Проверка случайных отклонений	Отклонение от распределения числа появлений подпоследовательностей определённого вида
Разновидность проверки случайных отклонений	Отклонение от распределения числа появлений подпоследовательностей определённого вида
Проверка аппроксимированной энтропии	Неравномерность распределения m-битных слов. Малые значения означают высокую повторяемость
Проверка серий	Неравномерность распределения m-битных слов
Линейная сложность	Отклонение от распределения линейной сложности для конечной длины подстроки
Дискретное преобразование Фурье	Поиск повторяющихся шаблонов

В 2012 году Национальный Институт Стандартов и Технологий (NIST) выпустил серию рекомендаций [8-10] по использованию генераторов случайных чисел. В документе [8] приведены рекомендации по получению случайных последовательностей использую КГПСП, в [9] – рекомендации для источников энтропии для ГСП, в [10] – рекомендации по конструированию ГСП.

Тест «Стопка книг».

В 2004 году российскими учеными Б.Я. Рябко и А.И. Пестунов был разработан новый статистический тест «Стопка книг» [11]. В работе [12] А.И.Миненко показал эффективность этого теста в сравнении с тестами NIST STS. Далее представлено описание статистического теста.

Пусть некоторый источник порождает буквы из алфавита $A = \{a_1, a_2, \dots, a_S\}$, $S > 1$ и требуется по выборке $x_1, x_2, \dots, x_n$ проверить гипотезу H_0 $p(a_1) = p(a_2) = \dots = p(a_S) = 1/S$ против альтернативной гипотезы H_1, являющейся отрицанием H_0.

При тестировании по предлагаемому методу буквы алфавита А упорядочены и занумерованы в соответствии с этим порядком от 1 до S. Причём этот порядок меняется после анализа каждого выборочного значения x_i следующим образом: буква x_i, которую мы обозначаем через a, получает номер 1, номера тех букв, которые были меньше, чем номер a, увеличиваются на 1, а у остальных букв номера не меняются. Как в стопке книг, если считать, что номер книги совпадает с положением в стопке. Книга извлекается и кладётся наверх. Её номер становится первым; книги, которые первоначально были над ней, сдвигаются вниз, а остальные остаются на месте. Основная идея метода – подсчитывается не частота встречаемости букв в выборке $x_1, x_2, \dots, x_n$, а частота встречаемости номеров букв (при описанном упорядочивании). В том случае, когда выполнена гипотеза H_1, вероятность (и частота встречаемости в выборке) некоторых букв больше $1/S$, и их номера в среднем будут меньше, чем у букв с меньшими вероятностями. Другими словами, книги, к которым обращаются чаще, проводят в верхней части стопки значительно большее время, чем остальные. Следовательно, вероятность обнаружить требуемую книгу в верхней части стопки больше, чем в нижней. Если же выполнена гипотеза H_0, то, очевидно, вероятность появления в выборке буквы с любым номером равна $1/S$.

При применении описываемого теста множество всех номеров $1,\dots,S$ заранее, до анализа выборки, разбивается на $r>1$ непересекающихся частей $A_1 = 1,2, \dots, k_1, A_2 = k_1 + 1, \dots, k_2, \dots, A_r = k_{r-1} + 1, \dots, k_r$. Затем по выборке $x_1, x_2, \dots, x_n$ подсчитывается количество номеров $v^t(x_t)$, принадлежащих подмножеству A_j, которое мы обозначим через n_j, $j=1\dots r$. При выполнении гипотезы H_0 вероятность того, что $v^t(x_t)$ принадлежит множеству A_j, пропорциональна количеству его элементов, т.е. равна $|A_j|/S$, а появление каждого элемента x_j в выборке независимо. Затем по $n_1, \dots, n_r$ по критерию χ^2 проверяется гипотеза H_0^*:

$$P\{ v^t x_t \in A_j\} = |A_j|/S \qquad (1)$$

против альтернативной гипотезы H_1^*. Очевидно, при выполнении исходной гипотезы H_0 выполняется H_0^*, и наоборот, при выполнении гипотезы H_1^* выполняется H_1. Поэтому применение описанного критерия корректно. При применении критерия χ^2 вычисляется величина

$$x^2 = \sum_{j=1}^{r} \frac{(n_j - nP_j^0)^2}{nP_j^0}, \qquad (2)$$

где $P_i^0 = |A_i|/S$. Известно, что распределение случайной величины x^2 с $r-1$ степенью свободы при выполнении H_0.

Заключение

Проблема генерации случайных и псевдослучайных последовательностей, применяемых в криптографии, остаётся актуальной на сегодняшний день. Существует большое количество статистических тестов для проверки генераторов. Исследования в данной области продолжаются, и находятся более эффективные методы оценки качества генераторов случайных последовательностей.

Список литературы

1. Рябко Б. Я., Фионов А. Н. *Криптографические методы защиты информации: Учебное пособие для вузов.* М.: Горячая линия – Телеком, 2005
2. Иванов М. А. *Теория, применение и оценка качества генераторов псевдослучайных последовательностей* / М. А. Иванов, И. В. Чугунков. М.: КУДИЦ-ОБРАЗ, 2003.
3. *Random Number Generators: An Evaluation and Comparison of Random.org and Some Commonly Used Generators*. Режим доступа: https://www.random.org/analysis/Analysis2005.pdf (дата обращения: 08.08.2017)
4. Кнут Д. *Искусство программирования, том 2. Получисленные методы* / Д. Кнут. М.: Изд. дом «Вильяме», 2007.
5. Brown R. *Dieharder: A Random Number Test Suite*. Режим доступа: http://www.phy.duke.edu/~rgb/General/dieharder.php (дата обращения: 10.08.2017)
6. *Security Requirements For Cryptographic Modules*. Режим доступа: http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf (дата обращения: 10.08.2017).
7. *NIST SP 800-22 Rev. 1a. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications* / [A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, S. Vo]. National Institute of Standards and Technology, 2010.
8. *Recommendation for Random Number Generation Using Deterministic Random Bit Generators*. Режим доступа: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-90Ar1.pdf (дата обращения: 07.08.2017).
9. *Recommendation for the Entropy Sources Used for Random Bit Generation*. Режим доступа: http://csrc.nist.gov/publications/drafts/800-90/sp800-90b_second_draft.pdf (дата обращения: 10.08.2017).
10. *Recommendation for Random Bit Generator (RBG) Constructions*. Режим доступа: http://csrc.nist.gov/publications/drafts/800-90/draft-sp800-90c.pdf (дата обращения: 10.08.2017).
11. Б. Я. Рябко, А. И. Пестунов. «Стопка книг» как новый статистический тест для случайных чисел // *Пробл. передачи информ.* 2004, том 40, выпуск 1, с. 73–78

    А. И. Миненко. Экспериментальное исследование эффективности тестов для проверки генераторов случайных чисел // *Вестник СибГУТИ.* 2010, № 4, с. 36-46.