УДК 681.3.06

М. А. КРИВОШЛЫК, Ю. И. ГОРБЕНКО, В. Н. ВЕРВЕЙКО

АЛГОРИТМЫ И СРЕДСТВА ТЕСТИРОВАНИЯ СЛУЧАЙНЫХ И ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Введение

Безопасность большинства криптографических систем зависит от способов формирования и использования ключей и параметров. Предельные характеристики стойкости достигаются в случае, если ключи и параметры выбираются случайно, равновероятно и независимо из полного пространства. Для выполнения этих условий используются генераторы случайных и псевдослучайных последовательностей. Случайный генератор – это устройство или алгоритм, который выдает последовательность статистически независимых символов с основанием алфавита m. Псевдослучайным генератором (ПСГ) называется детерминированный алгоритм (совокупность алгоритмов, устройство, совокупность алгоритмов и устройств), который, используя действительно случайную последовательность (СП) длиной k, формирует на ее основе псевдослучайную последовательность длины $l>>l_0$, где l_0 – допустимая длина. Вход в ПСГ называется начальным состоянием или зерном *(seed)*, на выходе ПСГ формируется псевдослучайная последовательность. Основным применением генераторов ПСП и СП в криптографии является формирование ключей, параметров и синхромаркеров.

Чтобы убедиться, что генератор безопасен, он должны быть подвергнут ряду статистических испытаний с целью подтверждения у сформированной им последовательности таких характеристик, которые ожидаются у случайных последовательностей. Псевдослучайные последовательности оцениваются с использованием ряда количественных показателей. Основными показателями являются[1,2]:

- случайность, равновероятность, независимость, однородность;
- период построения l_n ПСП;
- основание алфавита m;
- вероятность перекрытия в пространстве или во времени двух сегментов Y_r и Y_μ;
- структурная скрытность (эквивалентная сложность) S_l последовательности Y;
- энтропия источника начальных значений (seed);
- расстояние равнозначности l_0 конкретной последовательности Y_v;
- безопасное время генератора ПСЧ $t_Б$
- сложность I_Y формирования последовательности Y;
- длина параметров обратной связи (выход-вход) генератора;

Тесты, которые мы рассмотрим, помогают обнаружить уязвимые места, которые может иметь генератор. Анализ свойств ПСП и СП может быть выполнен путем исследования выходной последовательности генератора с использованием статистических тестов. Каждый статистический тест позволяет определить, обладает ли последовательность такими свойствами, какими обладает истинно случайная последовательность. Если последовательность не проходит хотя бы один из тестов, генератор либо может быть отвергнут как не обеспечивающий свойств случайности, либо может быть подвергнут дальнейшему тестированию. Если тестируемая последовательность проходит все статистические тесты, с определенной вероятностью генератор признается случайным. Более точно термин “признается” следует понимать как “не отвергается”.

При построении ключей одной из основных задач является получение случайных или псевдослучайных последовательностей, которые неотличимы от случайных, обладают большим периодом и другими свойствами, перечисленными выше. Перед использованием сгенерированной последовательности чисел $X = \{x_1, x_2, \ldots, x_n\}$ необходимо убедиться, что случайная величина X обладает равномерным законом распределения, её реализации случайны и независимы.

Статистическая гипотеза H представляет собой предположение относительно распределения случайной величины. Проверка статистической гипотезы представляет собой процедуру, позволяющую на основании значений случайной переменной сделать вывод о справедливости или ошибочности с определенной вероятностью выдвинутой гипотезы. Важным при гипотетическом тестировании является понятие уровня значимости α.

Уровень значимости α проверки статистической гипотезы H представляет собой вероятность того, что мы отвергнем гипотезу H, являющуюся на самом деле истинной.

Правильный выбор уровня значимости α для проверки является очень важной задачей. Если мы возьмем α слишком большим, то велика вероятность того, что мы отвергнем гипотезу, являющуюся на самом деле истинной. С другой стороны, если мы возьмем α слишком маленьким, то велика вероятность того, что мы примем гипотезу, являющуюся на самом деле ложной. На практике обычно используются значения уровня значимости от 0,001 до 0,05.

Математическая статистика дает нам возможность построить статистические тесты для проверки гипотез о равномерности, случайности и независимости случайных величин. Для этих целей можно использовать критерий χ^2 Пирсона. В США принята методика тестирования ПСП и СП, базирующаяся на критерии χ^2 Пирсона. Она зарегистрирована как FIPS 140-1. Недостатком этой методики является то, что с ее использованием можно протестировать ПСП или СП строго определенной длины – 20000 битов. Целью настоящей статьи является разработка методики проверки на случайность ПСП и СП произвольной длины (ограниченной только снизу), а также рассмотрение как частного случая методики, определенной FIPS 140-1.

Базовые статистические вероятностные тесты

В качестве базовых рекомендуется использовать следующие тесты[1,2]:

- частотный (монобитный) тест;
- тест двухбитных серий;
- тест Поккера;
- общий тест серий;
- автокорреляционный тест.

1.1 Монобитный тест

Цель этого теста состоит в том, чтобы определить, является ли количество «0» и «1» в последовательности s приблизительно таким, каким оно ожидается для случайной последовательности. Пусть n_0 и n_1 обозначают количество нулей и единиц в s, соответственно. Тогда параметр ПСП

$$X_1 = \frac{(n_0 - n_1)^2}{n}$$

подчиняется χ^2 - распределению с одной степенью свободы (если $n \geq 10$).

1.2 Тест двухбитных серий

Цель этого теста состоит в том, чтобы определить, является ли число появлений 00, 01, 10 и 11 как последовательностей s приблизительно таким же, как ожидается для случайной последовательности. Пусть n_0 и n_1 обозначают количество нулей и единиц в s, соответственно, и пусть n_{00}, n_{01}, n_{10}, n_{11} обозначают число появлений 00, 01, 10, 11 в s, соответственно. Заметим, что $n_{00} + n_{01} + n_{10} + n_{11} = (n-1)$, так как подпоследовательности могут перекрываться. Используемый статистический параметр равен

$$X_2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2\right) - \frac{2}{n}\left(n_0^2 + n_1^2\right) + 1$$

и подчиняется χ^2 - распределению с двумя степенями свободы (если n≥21).

1.3 Тест Покера

Пусть m будет положительным целым числом, таким, что $\left\lfloor \frac{n}{m} \right\rfloor \geq 5 \cdot 2^m$, и пусть $k = \left\lfloor \frac{n}{m} \right\rfloor$. Разделим последовательность s на k неперекрывающихся частей, каждая длиной m, и пусть n_i будет числом появлений i-го типа последовательности длины m, $1 \leq i \leq 2^m$. Тест Покера определяет, действительно ли каждая последовательность длиной m появляется приблизительно столько же раз в s, сколько ожидается для случайной последовательности. Используемый статистический параметр равен

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k,$$

что приблизительно соответствует χ^2 - распределению с 2^m -1 степенями свободы. Заметим, что тест Покера является обобщением частотного теста: установка $m = 1$ в тесте Покера дает частотный тест.

1.4 Тест серий

Тест серий позволяет определить, действительно ли число серий либо нулей, либо единиц различных длин в последовательности s такое же, как ожидается для случайной последовательности. Ожидаемое число интервалов (или блоков) длиной i в случайной последовательности длиной n равно $e_i = (n - i + 3)/2^{i+2}$. Пусть k будет равным наибольшему целому числу i, для которого $e_i \geq 5$. Пусть Bi, Gi будут числом блоков нулей и единиц, соответственно, длиной i в s для каждого i, $1 \leq i \leq k$.

Используемый статистический параметр равен

$$X_4 = \sum_{i=1}^{k} \frac{(B_i - e_i)^2}{e_i} + \sum_{i=1}^{k} \frac{(G_i - e_i)^2}{e_i}$$

и приблизительно соответствует χ^2 - распределению с $2k$ -2 степенями свободы.

1.4 Автокорреляционный тест

Цель автокорреляционного теста состоит в том, чтобы проверить степень связи между последовательностью s и апериодическим ее сдвигом. Пусть d будет фиксированным целым числом, $1 \leq d \leq \lfloor n/2 \rfloor$. Число битов в последовательности s не совпадает с их числом в d-сдвигах, и равно $A(d) = \sum_{i=0}^{n-d-1} s_i \oplus s_{i+d}$. Используемый статистический параметр равен

$$X_5 = 2\left(A(d) - \frac{n-d}{2}\right)/\sqrt{n-d}$$

и приблизительно соответствует $N(0,1)$ нормальному распределению, если n-$d \geq 10$. Так как малые значения $A(d)$ являются столь же маловероятными, как и большие, то должен использоваться двусторонний тест.

2. Пример решения задачи тестирования последовательности

Рассмотрим последовательность s длины $n = 128$, полученную путем четырехкратного копирования следующей последовательности:

0110 0100 0111 1010 1100 1000 1111 0101.

Проверим эту последовательность по статистическим критериям:

2.1 Частотный (монобитный) тест

$n = 128$, $n_0 = 60$, $n_1 = 68$, тогда $X_1 = \frac{(60-68)^2}{128} = 0{,}5$. Поскольку $X_1 < X_{1пр} = \chi^2(1;0.05) = 3{,}84$ [2] и $X_1 = 0{,}5 < 3{,}84$, то тест пройден, и гипотеза не отвергается.

2.2 Двухбитный тест серий

$n_{00} = 24$, $n_{01} = 36$, $n_{10} = 35$, $n_{11} = 32$ и значение статистического параметра

$X_2 = \frac{4}{127}\left(24^2 + 36^2 + 35^2 + 32^2\right) - \frac{2}{127}\left(60^2 + 68^2\right) + 1 = 2{,}295$. Поскольку

$X_2 < X_{2пр} = \chi^2(2;0{,}05) = 5{,}99$ [2] и $X_2 = 2{,}295 < 5{,}99$, то тест пройден, и гипотеза не отвергается.

2.3 Тест Покера

Пусть $m = 3$ и $k = 42$. Блоки 000, 001, 010, 011, 100, 101, 110, 111 появляются 4, 6, 5, 6, 5, 6, 4, 6 раз

соответственно, а значение статистического параметра

$$X_3 = \frac{2^3}{42}\left(\sum_{i=1}^{8} 4^2 + 6^2 + 5^2 + 6^2 + 5^2 + 6^2 + 4^2 + 6^2\right) - 42 = 1{,}048.$$

Поскольку $X_3 < X_{3пр} = \chi^2(7;0{,}05) = 14{,}076$ и $X_3 = 1{,}048 < 14{,}076$, то тест пройден, и гипотеза не отвергается.

2.4 Тест серий

Здесь $e_1 = 16{,}25$, $e_2 = 8{,}0625$ и $k = 2$. Имеется 20, 8 блоков единиц длиной 1, 2 соответственно, и 20, 8 интервалов нулей длиной 1, 2 соответственно. Значение статистического параметра X_4 равно 1,73.

Поскольку $X_4 < X_{4пр} = \chi^2(2;0{,}05) = 5{,}99$ и $X_4 = 1{,}73 < 5{,}99$, то тест пройден, и гипотеза не отвергается.

2.5 Автокорреляционный тест

Если $d = 8$, то $A(8) = 61$. Значение статистического параметра

$$X_5 = 2\left(61 - \frac{128-8}{2}\right)/\sqrt{128-2} = 0{,}18$$

Так как $X_{5пр} \le X_5 < X_{5пр} = N(0;1)$ и X_5 находится в допустимом интервале, $-2{,}17 \le 0{,}18 \le 2{,}17$, то тест пройден и гипотеза о нормальном законе распределения не отвергается.

Таким образом, рассмотренная ПСП проходит все использованные тесты по критерию χ^2, и гипотеза о том, что она обладает свойствами случайной последовательности не отвергается. Дело в том, что выбранная 32-разрядная последовательность является линейной рекуррентной последовательностью максимального периода с добавленным одним битом, и она должна успешно проходить тесты на псевдослучайность.

3. Тестирование источников случайных и псевдослучайных последовательностей на основе методики американского федерального стандарта FIPS 140-1

В американском федеральном стандарте FIPS 140-1 используется четыре статистических теста на случайность: монобитный тест, блочный тест, тест серий и тест длин серий. В этих тестах для удовлетворительных значений статистических параметров задаются границы. Отдельная битовая строка длиной 20000 битов, получаемая из генератора, подвергается проверке по каждому из четырех названных тестов. Если какой-нибудь из тестов не пройден, то считается, что генератор не прошел тестирование. FIPS 140-1 рекомендуется применять для технологического тестирования аппаратных генераторов случайных чисел.

Периодом периодической последовательности s называется наименьшее положительное число n, для которого s периодическая. Если s — периодическая последовательность периода n, то циклом s является подпоследовательность s.

Пусть s — последовательность. Серией s называется подпоследовательность s, состоящая из последовательных 0 или 1. Серия, состоящая из 0, называется интервалом, а серия, состоящая из 1, называется блоком.

3.1. Монобитный тест

Цель этого теста состоит в том, чтобы определить, является ли количество нулей и единиц в последовательности s приблизительно таким, каким оно ожидается для случайной последовательности. Пусть n_1 обозначает количество нулей (или единиц) в s. Число должно удовлетворять условию $9654 < n_1 < 10346$.

3.2. Блочный тест

Пусть m положительное целое число, такое, что $\left\lfloor \frac{n}{m} \right\rfloor \ge 5 \cdot (2^m)$, и пусть $k = \left\lfloor \frac{n}{m} \right\rfloor$. Разобьем последовательность s на k непересекающихся подпоследовательностей, каждая длиной m, и пусть n_i бу-

дет числом появлений *i-го* типа последовательности длиной *m*. Блочный тест определяет, действительно ли последовательности длиной *m* появляются в *s* приблизительно столько же раз, сколько ожидается для случайной последовательности. Для применения критерия используется расчет параметра

$$X_3 = \frac{2^m}{k}\left(\sum_{i=1}^{2^m} n_i^2\right) - k \ ,$$

что приблизительно соответствует χ^2 распределению с 2^m-1 степенями свободы. Статистический параметр, задаваемый уравнением, вычисляется для $m = 4$. Статистика должна удовлетворять условию $1{,}03 < X_3 < 57{,}4$.

3.3 Тест серий

Цель теста серий состоит в том, чтобы определить, действительно ли число серий различных длин в последовательности *s* такое же, как ожидается для случайной последовательности. Пусть *k* равно наибольшему количеству битов в последовательности. Для каждого *i* от 1 до k подсчитывается число интервалов и блоков длиной *i* (в целях упрощения теста серии длиной, больше 6, рассматриваются как серии длиной 6). Тест серий пройден, если количество серий нулей и единиц последовательности находится в пределах соответствующего интервала, заданного табл. 1.

Таблица 1

Длина серии	Требуемый интервал
1	2267-2733
2	1079-1421
3	502-748
4	223-402
5	90-223
6	90-223

3.4 Тест длин серий

Цель теста длин серий состоит в том, чтобы определить, действительно ли максимальная длина серии в последовательности *s* такая же, как ожидается для случайной последовательности. Тест длины серий пройден, если длина любой серии анализируемой последовательности не превышает 34.

Заключение

Использование критерия χ^2 позволяет проверить ПСП и СП на степень их "похожести" на случайную последовательность.

При создании стандарта статистического тестирования Украины в качестве базовых можно использовать частотный тест, двухбитный тест, тест Покера, общий тест серий и автокорреляционный тест. Эти тесты могут применяться при разработке новых программных и аппаратных генераторов ПСП и СП, а также в качестве технологического теста.

Предлагаемый набор тестов является более расширенным по сравнению со стандартом FIPS 140-1 и может применяться для ПСП и СП неограниченных длин.

Список литературы: 1. *FIPS PUB 140-1.* Cryptographic modules security requirements // NIST, 1993. **2.** *Менезис А., Ван Оршот П., Ватсон С.* Прикладная криптография Гл. 5 // CRC Press, 1996. **3.** *ANSI X9.17.* "American National Standard for Financial Institution Key Management (Wholesales)" American Bankers Association, 1985.

Харьковский государственный технический университет радиоэлектроники

Поступила в редколлегию 04.04.2001