УДК 681.3.06

Ю.П. Буценко, Г.Н. Розоринов, Ю.Г. Савченко

ОБЩЕЕ И СЕЛЕКТИВНОЕ ТЕСТИРОВАНИЕ ПСЕВДОСЛУЧАЙНЫХ БИТОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Предлагается параметрический подход к тестированию псевдослучайных двоичных последовательностей, основанный на вычислении реальной энтропии отдельных фрагментов последовательности. Рассмотрена задача количественной оценки близости тестируемой последовательности к истинно случайной и задача выявления корреляционных связей между отдельными фрагментами. Показано, что при аппаратной реализации генератора псевдослучайных чисел чаще всего используются схемы, которые можно отнести к цифровым автоматам Мура. При реализации генератора на основе автомата Мили, схема усложняется из-за необходимости учета входных сигналов.

Ключевые слова: псевдослучайная двоичная последовательность, криптозащита, корреляция, тестирование, энтропия.

Введение

Псевдослучайные двоичные последовательности (ПСДП) широко используются в телекоммуникационных системах для криптозащиты информационного трафика при потоковом и блочном шифровании. Для создания ПСДП применяют аппаратные или программные генераторы псевдослучайных чисел (ГПСЧ). На сегодня известно достаточно большое количество реализаций таких генераторов и, естественно, возникает проблема объективной оценки их качества с точки зрения использования генерируемых последовательностей в качестве ключей при блочном шифровании или параметров генератора при скремблировании. В каждом конкретном случае для обоснованного выбора ГПСЧ проводят тестирование ПСДП "на случайность". При большой длине ПСДП процесс тестирования оказывается весьма трудоемким по временным затратам, что связано с универсальностью большинства известных тестов [1, 2, 4 - 6]. Поэтому желательно тестирование проводить целенаправленно (селективно), выбирая тот или иной тест в соответствии с некоторыми внешними требованиями, определяемыми либо областью применения ПСДП, либо алгоритмом ее генерации.

Основная часть

В 1999 г. была предложена методика, предполагавшая использование набора из 16 тестов, каждый из которых ориентирован на выявление конкретного свойства ПСДП, характерного для действительно случайной последовательности [1, 7]. Например, простейший из тестов (Monobit, Block Frequency Cumulative Sums Forward Reserve) основан на тривиальной идее подсчета относительных частот нулей и единиц в последовательности или ее фрагментах. Очевидно, что даже простая последовательность чередующихся нулей и единиц будет успешно протестирована как удовлетворительная. Следующими по сложности являются тесты типа Runs или Long Runs of Queues, контролирующие длины, возникающих в изучаемой последовательности так называемых стационарных участков (состоящих только из нулей или только единиц). Фактически, эти тесты (как и предыдущие) обнаруживают использование примитивных процедур генерации ПСДП, например, формирования последовательности путем наращивания ее длины за счет повторения стационарных фрагментов.

Группа тестов, контролирующих периодичность в последовательности (Discrete Fourier Transform, Periodic Templates, Aperiodic Templates) также ориентирована на выявление упрощенных процедур генерации, их принципиальным ограничением является ориентация на короткие периоды повторения фрагментов. Тесты, проверяющие "случайность блуждания" (Random Excursions, Random Excursions Variant) являются весьма перспективными с теоретической точки зрения. Однако, с другой стороны, способы маскирования, обеспечивающие прохождение тестов, достаточно просты, хотя и эффективны.

Тест Linear Complexity выявляет условную сложность последовательности с точки зрения ее аналитического описания, но ограничен лишь линейными формами представления.

К еще одной группе относятся тесты информационного характера (Universal Statistical Test, Approximate Entropy, Lempel-Ziv Complexity), базирующиеся на измерении количества информации, которая содержится в тестируемой ПСДП. Используется подход, основанный на измерении эффективности компрессии или сравнении частот перекрывающихся фрагментов с вероятностями таких событий для действительно случайной последовательности.

В целом, статистические тесты NIST (Национального института стандартов и технологий), получившие на сегодня наибольшее распространение и признание, являются хорошим инструментом для сравнительной оценки различных ПСДП в некоторой условной системе координат. Однако, строго говоря, полученные оценки можно рассматривать как в определенной мере субъективные, поскольку они жестко привязаны к выбранному набору тестов. Наиболее перспективным, на наш взгляд, является информационный подход и поэтому может быть поставлена задача построения некоторой универсальной процедуры вычисления таких характеристик ПСДП, которые позволили бы оценить, насколько конкретная битовая последовательность близка к действительно случайной. При этом истинно случайной будем считать такую последовательность, в которой любой фрагмент произвольной длины появляется примерно с одинаковой частотой. Например, представим себе, что фрагмент ПСДП наблюдается через некоторое воображаемое "окно" шириной S бит, и для истинно случайной последовательности все 2^S разновидностей фрагмента являются равновероятными.

С учетом этих предположений возникают следующие задачи.

Задача А. Пусть задана конкретная битовая последовательность $W = (w_1, w_2, ..., w_n)$. Необходимо оценить *количественно*, насколько эта последовательность близка к действительно случайной. То есть речь идет о попытке объективной оценки качества конкретной ПСДП.

В других случаях, например, относящихся к криптоанализу, постановка задачи может быть конкретизирована за счет некоторой априори известной дополнительной информации.

Задача В. Как и в предыдущем случае пусть задана конкретная ПСДП. Кроме того, известен общий алгоритм формирования последовательности. Например, предположительно известно, что генератором ПСДП является регистр сдвига с обратными связями по модулю 2 (LFSR – Linear Feedback Shift Register), а тестирование должно подтвердить или опровергнуть это предположение и выявить (обнаружить) корреляционные зависимости между отдельными фрагментами ПСДП. Полученный результат в этом случае может быть использован для организации соответствующей атаки, то есть определения конкретных обратных связей в регистре генератора и его начальные установки.

Принимаемое предположение о наличии дополнительной информации о ГПСЧ основано на фундаментальном принципе Опоста Керкгоффса, который очень кратко сформулировал К. Шеннон: "Противник знает все, кроме ключа".

Очевидно, что при решении задачи ***В*** целесообразно использовать эту дополнительную информацию с тем, чтобы уменьшить трудоемкость вычислений при тестировании.

Основной и, во многих случаях, доступной является информация о классе аппаратных или программных средств, используемых для генерации ПСДП. Чаще всего известно, например, что в качестве генератора ПСДП используется линейный фильтр с обратными связями по модулю 2. В этом случае конкретная ПСДП однозначно может быть вычислена с помощью следующих параметров: длина регистра d; коэффициенты многочлена, задающего конкретный вид обратных связей регистра - $b_0, b_1, b_2, ..., b_d$; стартовое слово (комбинация), задающее начальное состояние регистра (то есть, по сути, первые d бит ПСДП).

Переходя к формальной постановке указанных задач, можно рассуждать так.

Если тестируемая последовательность действительно случайная, то вероятность появления любой другой последовательности такой же длины равна $\frac{1}{2^n}$. Энтропия источника таких сообщений максимальна, то есть $H_{\max} = \log_2 2^n = n$, а при отклонении от равномерного распределения $H_{реал} = -\sum p_i \log p_i$, где p_i - вероятность появления сообщения.

В рассматриваемом случае, когда все сообщения равновероятны $p_i = \frac{1}{2^n}$.

Естественно, измерение фактической энтропии на основе статистического эксперимента для реальных значений n вряд ли осуществимо. К тому же, в достаточно типичных случаях для тестирования предъявляется лишь одна ПСДП. Поэтому следует отталкиваться от некоторой гипотетической, но на самом деле легко практически реализуемой процедуры. Реальная ПСДП, конечно же, *не случайная.* Поэтому $H_{реал} < H_{\max}$. Найти точное значение $H_{реал}$ для реальных генераторов не представляется возможным. Даже при весьма скромных (но реальных) значениях $n = 256...512$ подсчитать частоты появления каждой из $2^{256}...2^{512}$ возможных ПСДП нереально за любое разумное время.

Поэтому при тестировании ограничим длину анализируемых фрагментов ПСДП.

Пусть длина фрагмента $s = 1$ (один бит). Очевидно, что в действительно случайной ПСДП $p(0) = p(1) = \frac{1}{2}$. Это легко и быстро можно проверить при тестировании, подсчитав частоты появления 0 и 1.

Пусть длина фрагмента $s = 2$. В этом случае также легко подсчитать частоты (вероятности) появления комбинаций 00, 01, 10, 11. Можно ожидать, что для "хорошей" ПСДП эти вероятности будут близки к $\frac{1}{4}$.

Пусть длина фрагмента s = 3. Частоты появления комбинаций 000, 001, …, 111 для "хорошей" ПСДП будут близки к $\frac{1}{8}$. И так далее.

Для подсчета реальной энтропии воспользуемся формулой

$$H_{реал} = -\frac{1}{s}\sum_{i=1}^{q} p_i \log_2 p_i , \qquad (1)$$

где $q = 2^s, s -$ длина окна.

В формуле (1) содержится множитель $\frac{1}{s}$, который позволяет нормировать полученное в результате эксперимента значение энтропии и "привести" его к одному биту, то есть вычисленные таким образом значения не зависят от длины окна и становятся сравнимыми друг с другом.

Можно уверенно утверждать, что рано или поздно статистический эксперимент покажет, что равномерное распределение нарушается, и $H_{реал} < H_{\max}$. К тому же, трудоемкость подсчета частот растет экспоненциально и очень скоро становится вычислительно нереализуемой. Можно предполагать, что картина изменения реальной энтропии $H_{реал}$ с увеличением длины фрагмента s будет выглядеть примерно так, как показано на рис.1.

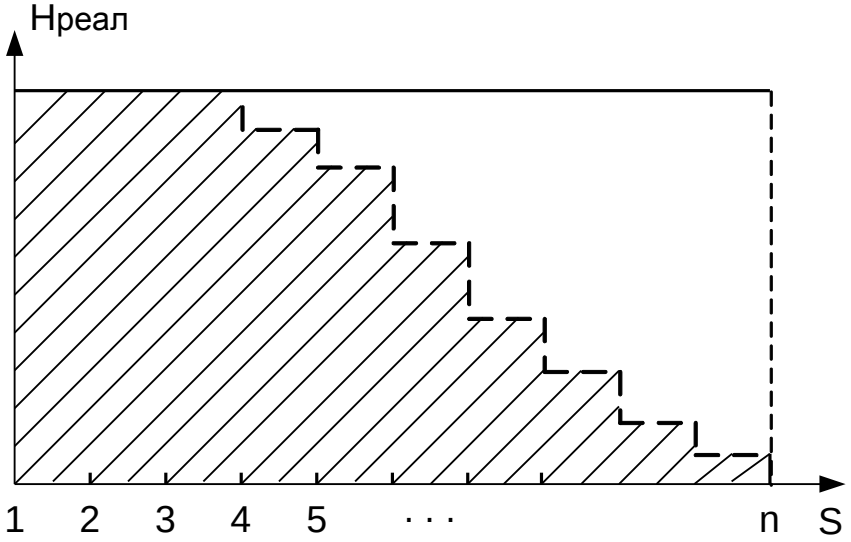

Рис. 1. Зависимость реальной энтропии от длины фрагмента

Очевидно, теоретически для действительно случайной ПСДП и фрагментов любой длины энтропия будет оставаться максимальной, и кривая ее изменения превратится в горизонтальную прямую, параллельную оси абсцисс. В качестве меры качества произвольной ПСДП можно использовать относительную площадь заштрихованной области (относительно площади всего прямоугольника).

Таким образом, последовательно вычисленные энтропии $H_s(s=1,2,...,n)$ должны хорошо согласовываться (для случайной последовательности) с теоретической линейной зависимостью $H_{реал}=H_{\max}$. И, наоборот, существенные отклонения от такой зависимости указывают на детерминированный характер формирования последовательности. При аппроксимации такой зависимости методом наименьших квадратов получаем систему равенств

$$a\sum_{i=1}^{k} i^2 + b\sum_{i=1}^{k} i = \sum_{i=1}^{k} iH_i,$$

$$a\sum_{i=1}^{k} i + kb = \sum_{i=1}^{k} H_i.$$

При близости к вышеуказанной зависимости необходимо и достаточно, чтобы одновременно выполнялись условия:

$$\tilde{A}(k) = 12\frac{\sum_{i=1}^{k} iH_i - \frac{k+1}{2}\sum_{i=1}^{k} H_i}{ak(k^2-1)} \approx 1,$$

$$\tilde{B}(k) = \sum_{i=1}^{k} iH_i - \frac{k+1}{2}\sum_{i=1}^{k} H_i \approx 0.$$

Отметим, что критическим значением для величины $\tilde{A}(k)$ принимается в данном случае значение $1-\alpha$, то есть для $\tilde{A}(k)$ при $A(k)<1-\alpha$ принимается решение о детерминированности ПСДП. В то же время значение k^*, при котором для $A(k^*)$ начинает выполняться это неравенство, может рассматриваться как нижняя граница для выбора разрядности предполагаемого генератора тестируемой ПСДП. Аналогичным образом,

критическим значением для величины $\tilde{B}(k)$ является число $\beta > 0$, и решение принимается в случае, когда $B(k) > \beta$.

Принципиально иным является подход, основанный на статистических оценках параметров распределения количества нулей и единиц в тестируемой последовательности. Такой подход может рассматриваться как дополняющий.

Также информационным является подход, основанный на неравенстве Крамера-Рао, ограничивающем снизу дисперсию оценки параметра распределения. В рассматриваемом случае традиционным является подход, предполагающий равномерность дискретного распределения нулей и единиц в любой "выборке" из тестируемой последовательности, что исключает возможность использования упомянутого неравенства. В то же время относительная частота $\frac{k}{n}$ (например, единиц в "выборке" объема n) имеет асимптотически нормальное распределение, которое, в свою очередь, допускает использование неравенства. Таким образом, определив для вероятности p в образованной последовательностью схеме Бернулли доверительный интервал (которому указанная вероятность должна будет принадлежать с вероятностью $\gamma = 1 - \alpha$):

$$[\frac{k}{n} - t_1\sqrt{\frac{k(n-k)(1-\frac{n}{N})}{n^3}}; \quad \frac{k}{n} + t_1\sqrt{\frac{k(n-k)(1-\frac{n}{N})}{n^3}}];$$

где t_1 - соответствующий квантиль нормального распределения; N – длина анализируемой последовательности.

Аналогичный интервал может быть определен "вокруг" априори известной вероятности $p = \frac{1}{2}$ для априори известной дисперсии эмпирической частоты $\frac{pq}{n} = \frac{1}{4n}$ с помощью неравенства Чебышева и использования нижней границы дисперсии такой оценки

$$P\left\{\left|D - \frac{1}{4n}\right| \le \varepsilon\right\} \ge 1 - \frac{\overline{D}}{\varepsilon^2} = 1 - \varepsilon,$$

откуда $\varepsilon_1 = \sqrt{\frac{D}{\varepsilon}}$, где $\overline{D}$ - нижняя граница дисперсии по Крамеру-Рао. Полученный таким образом доверительный интервал $\left(\frac{1}{4n} - \varepsilon_1;\ \frac{1}{4n} + \varepsilon_1\right)$ сравнивается с интервалом $[\varphi(\alpha); \varphi(\beta)]$, где $\varphi = \frac{x(1-x)}{n}, (\alpha \vee \beta) = \frac{k}{n} \pm \sqrt{\frac{k(k-n)(1-n/N)}{n^3}}$. В случае, если при $n \to N$ отношение $\frac{\varphi(\beta) - \varphi(\alpha)}{2\varepsilon_1}$, начиная с некоторого значения n демонстрирует резкое убывание, это может указывать на искусственный характер чередования нулей и единиц в последовательности, то есть ее неслучайный характер.

Выводы

При аппаратной реализации ГПСЧ чаще всего используются схемы, которые можно отнести к цифровым автоматам Мура, то есть к таким схемам, у которых выходные сигналы определяются внутренним состоянием автомата в текущий момент времени, а переход в следующее состояние происходит при поступлении входного сигнала в виде, например, очередного тактового импульса.

Разнообразие выходных сигналов в этом случае, очевидно, ограничено множеством внутренних состояний автомата, то есть его объемом памяти. С точки зрения тестирования ПСДП, (а это, фактически, последовательность сменяющих друг друга состояний автомата)

важным является то, что каждое последующее состояние автомата *однозначно* определяется предыдущим его состоянием, и задают этот переход соответствующие автоматные уравнения. Основная задача криптоанализа как раз заключается в выявлении пар соседних фрагментов последовательности, связанных жесткой и функциональной зависимостью [3], которая может быть обнаружена, в том числе, и при матричном подходе [8].

В случае, когда генератор реализован в виде автомата Мили, задача существенно усложняется, поскольку появляется неизвестная составляющая – это входные сигналы и внешний по отношению к генератору способ их образования.

Сформулированная в работе задача ***A*** относится, в основном, к количественной *оценке качества* различных генераторов ПСДП. Очевидно, что в этом случае для принятия решения о пригодности (или непригодности) того или иного программного или аппаратного генератора необходимо дополнительно выработать пороговый критерий с учетом реальных требований к необходимому уровню защищенности конкретного информационного ресурса.

Задача ***B*** ориентирована на выявление наличия корреляционных связей в тестируемой последовательности. В других терминах эта задача сводится к обнаружению закономерностей в процедуре формирования ПСДП, которые позволили бы в случае их обнаружения предсказать всю последовательность по ее фрагменту ограниченной длины.

Литература

1. Soto J. Randommess Testing of Advanced Encrypti on Candidate Algorithms - NIST, 1999, 743 p.

2. Потий А., Орлова С., Статистическое тестирование генераторов случайных и псевдослучайных чисел с использованием набора статистических тестов NIST STS. Правове, нормативне та метрологічне забезпечення систем захисту інформації в Україны, вип. 2, 2001, с.206-214.

3. Пометун С.О., Алгебраїчні атаки на потокові шифратори як узагальнення кореляційних атак. – Системні дослідження та інформаційні технології, №2, 2008, с. 29-40.

4. Дональд Э. Кнут, Глава 3. Случайные числа // Искусство программирования = The Art of Computer Programming. — 3-е изд. — М.: Вильямс, 2000.— Т.2. Получисленные алгоритмы. — 832 с.

5. Иванов М.А., Чугунков И.В., Глава 4. Методика оценки качества генераторов ПСП // Теория, применение и оценка качества генераторов псевдослучайных последовательностей. — М.: КУДИЦ-ОБРАЗ, 2003. — 240 с.

6. Statistical Testing of Random Number Generators // Proceedings of the 22nd National Information Systems Security Conference, 10/99.

7. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. - NIST SP 800-22.

8. Савченко Ю.Г., Чич Т.В., Матричный подход к построению и реализации алгоритмов шифрования двоичной информации в телекоммуникационных сетях // Зв'язок, №5, с. 38-41.

Надійшла 02.04.2014 р. Рецензент: д.т.н., проф. Барабаш О.В.