УДК 004.042

Е.В. Игоничкина

Статистический анализ поточных шифров

Перечислены наиболее популярные наборы статистических тестов. Описана методика тестирования ПСП, генерируемых поточными шифрами, и способы интерпретации полученных результатов.

Почти все применяемые на практике шифры характеризуются как условно надежные, поскольку они могут быть в принципе раскрыты при наличии неограниченных вычислительных возможностей. Абсолютно надежные шифры нельзя разрушить даже при использовании неограниченных вычислительных возможностей. Существует единственный такой шифр, применяемый на практике, – это одноразовый блокнот, который был изобретен в 1917 г. М. Моборном и Г. Вернамом.

Классический одноразовый блокнот – это блокнот, составленный из отрывных страниц, на каждой из которых напечатана таблица со случайными числами (ключами). Блокнот выполняется в двух экземплярах: один используется отправителем, а другой – получателем. Каждый символ ключа используется только один раз и только в одном сообщении. Шифрование представляет собой сложение (по модулю 26, если используются латинские буквы; по модулю 2, если данные представлены в двоичном виде, и т.п.) символа открытого текста и символа ключа из одноразового блокнота. Закончив шифровать сообщение, отправитель уничтожает использованные страницы блокнота. Каждое новое сообщение использует новые символы ключа.

Поточный шифр по своей сути пытается имитировать концепцию одноразового блокнота. Поточные шифры не являются абсолютно надежными, как одноразовый блокнот, но их удобнее использовать на практике, так как они используют короткий ключ для генерации псевдослучайной последовательности (ПСП) – гаммы или ключевого потока (keystream). Такая ПСП должна быть практически неотличимой от истинно случайной.

Существует множество статистических тестов, позволяющих обнаружить различные типы неслучайности, которые могут существовать в ПСП. Например, частотный тест (frequency test) определяет, является ли число единиц и нулей в двоичной ПСП приблизительно таким же, как в истинно случайной последовательности, т.е. количество единиц и нулей должно быть примерно одинаковым. Наиболее известными являются статистические тесты, описанные в книгах Д. Кнута [1] и А. Менезиса [2], а также наборы статистических тестов Crypt-X [3], Diehard [4] и NIST [5].

В своей книге [1] Д. Кнут описывает несколько эмпирических тестов: frequency, serial, gap, poker, coupon collector's, permutation, run, maximum-of-t, collision, birthday spacings, serial correlation.

Набор статистических тестов Diehard содержит следующие тесты [4]: birthday spacings, overlapping permutation, binary rank test for 31x31 matrices, binary rank test for 32x32 matrices, binary rank test for 6x8 matrices, monkey, bitstream, count-the-1's test on a stream of bytes, count-the-1's test for specific bytes, parking lot test, minimum distance test, 3dspheres, sqeeze, overlapping sums, runs, craps.

Набор статистических тестов Crypt-X содержит следующие тесты [3]: frequency, bi-nary derivative, change point, runs, sequence complexity и linear complexity тесты.

В своей книге [2] А. Мензис и др. описывают 5 основных тестов: frequency (monobit), serial (two-bit), poker, runs, autocorrelation, а также универсальный тест Мауэра.

Набор статистических тестов NIST включает в себя следующие статистические тесты [5]: frequency (monobit), frequency test within a block, runs, test for the longest-run-of-ones in a block, binary matrix rank, discrete fourier transform (spectral), non-overlapping template matching, overlapping template matching, maurer's "universal statistical", lempel-ziv compression, linear complexity, serial, approximate entropy, cumulative sums (cusums), random excursions, random excursions variant.

Основным принципом тестирования является проверка нулевой гипотезы H_0: тестируемая последовательность случайна. Альтернативной гипотезой H_a является гипотеза о том, что последовательность не случайна. По результатам каждого теста нулевая гипотеза либо принимается, либо отвергается.

Для анализа прохождения псевдослучайными последовательностями статистического теста используются различные подходы, например такие, как критерий по пороговому уровню, доверительные интервалы и критерий с использованием значения вероятности [6]. Наиболее гибким из перечисленных является третий подход, который заключается в вычислении для последовательности s тестовой статистики $c(s)$ и соответствующего ей значения вероятности (P-$value$). Правило принятия решения в данном случае формулируется следующим образом: для фиксированного уровня значимости α двоичная последовательность s не проходит статистический тест, если $P\text{-}value < \alpha$.

Для статистического анализа генераторов ПСП используется следующая стратегия [5]:

1. Выбирается генератор, который будет тестироваться. Генератор должен производить двоичную последовательность нулей и единиц длиной n.

2. Для выбранного на предыдущем шаге генератора конструируется набор, состоящий из m двоичных последовательностей, каждая длиной n бит. Размер набора m должен иметь порядок α^{-1}, т.е. для $\alpha = 0{,}01$ набор должен состоять по крайней мере из 100 последовательностей.

3. Выполняется набор статистических тестов. Каждый статистический тест оценивает каждую n-битную последовательность и вырабатывает одно или более значений *P-value*. На основе этих значений *P-value* может быть сделано заключение, касающееся качества тестируемой последовательности.

4. Для каждого статистического теста производится набор значений *P-value*, соответствующий набору из m последовательностей. Для фиксированного уровня значимости ожидается, что определенный процент значений *P-value* будет указывать на то, что последовательность не проходит тестирование. Например, для уровня значимости $\alpha = 0{,}01$ ожидается, что около 1% последовательностей провалит тестирование. Последовательности проходят статистические тесты всякий раз, когда *P-value* $\geq \alpha$, и не проходят во всех остальных случаях.

В результате тестирования может произойти три типичных исхода:

1. Анализ *P-value* не указывает на отклонение от случайности.
2. Анализ ясно указывает на отклонение от случайности.
3. Анализ является недостаточным.

В руководстве NIST предлагается два способа интерпретации результатов [5]:

1. Исследование доли последовательностей, прошедших статистические тесты.
2. Исследование распределения значений *P-value*, чтобы проверить равномерность их распределения.

В случае, когда любой из этих подходов терпит неудачу, т.е. соответствующая нулевая гипотеза должна быть отклонена, должны быть проведены дополнительные численные эксперименты на других образцах генератора, чтобы определить, было ли это событие статистическим отклонением или это было доказательством неслучайности.

Выполняя статистический анализ поточного шифра, необходимо помнить, что он не является генератором ПСП в чистом виде, так как кроме наличия хороших статистических свойств должен обладать стойкостью к различным криптоатакам. Статистический анализ ПСП, генерируемых поточным шифром, можно рассматривать как первый этап всестороннего анализа поточного шифра.

Литература

1. Кнут Д. Искусство программирования для ЭВМ. – Т. 2. Получисленные алгоритмы. – М.: Мир, 1977. – 727 с.

2. Menezes A., van Oorshot P., Vanstone S. Handbook of Applied Cryptography. – CRC Press, 1997

3. Statistical test suite Crypt-X // http://www.isi.qut.edu.au/resources/cryptx

4. The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness // http://www.stat.fsu.edu/pub/diehard/

5. A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications. NIST Special Publication 800-22. May 15, 2001.

6. J. Soto Statistical Testing of Random Number Generators // http://csrc.nist.gov/rng/nissc-paper.pdf

---

Игоничкина Екатерина Викторовна
ГОУ ВПО Томский государственный университет систем управления и радиоэлектроники
Аспирант, ассистент кафедры радиотехнических систем
Тел.: 413-608, 413-469.
Эл. почта: iev109@mail.ru.

E.V. Igonichkina

The statistical analysis of stream ciphers

The most popular sets of statistical tests are listed. Methodology of testing of pseudorandom sequences generated by stream ciphers and ways of interpretation of the received results are described.