УДК 003.26+519.2

О ТЕСТИРОВАНИИ ВЫХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ НА ОСНОВЕ ЧАСТОТНЫХ СТАТИСТИК

М.В. МАЛЬЦЕВ, Ю.С. ХАРИН

Научно-исследовательский институт
прикладных проблем математики и информатики
Минск, 220030, Республика Беларусь
maltsew@bsu.by, kharin@bsu.by

Введение

Генераторы случайных и псевдослучайных числовых последовательностей (далее – генераторы) являются важнейшими элементами систем криптографической защиты информации. Последовательности, созданные генераторами, используются для формирования ключей, векторов инициализации, параметров криптографических алгоритмов и протоколов. Использование генераторов с уязвимостями снижает безопасность криптосистем, упрощает задачу злоумышленников по вычислению секретных параметров. Известным примером такой уязвимости является предсказуемость генератора в ранней версии браузера Netscape Navigator, что позволило значительно сократить время, затрачиваемое на перебор ключей шифрования [1]. Таким образом, актуальной является задача оценки качества выходных последовательностей генераторов. Для решения данной задачи широко используются методы теории вероятностей и математической статистики.

1. Математическая модель «чистой случайности»

Результатом работы стойкого генератора должна являться «чисто случайная» последовательность, математической моделью которой является равномерно-распределенная случайная последовательность (РРСП) [2]. Принимающая значения из N-элементного множества A случайная последовательность $\{x_t \in A : t \in \mathbb{N}\}$, называется РРСП, если для нее выполняются следующие два свойства:

1. Для любого $n \in \mathbb{N}$ и для любых $1 \le t_1 < t_2 < \ldots < t_n$ случайные величины $x_{t_1}, \ldots, x_{t_n}$ независимы в совокупности.

2. Для любого $t \in \mathbb{N}$ случайная величина x_t имеет равномерное распределение вероятностей на множестве A, т.е.

$$\mathrm{P}\{x_t = i\} = \frac{1}{N}, i \in A.$$

Для оценки качества выходных последовательностей генераторов применяется статистическое тестирование. Статистические тесты, объединяемые в наборы (батареи), проверяют, соответствуют ли свойства наблюдаемой последовательности модели РРСП. Одно из важнейших свойств РРСП – равномерное распределение фрагментов длины s (s-грамм):

$$\mathrm{P}\{x_t = i_1, \ldots, x_{t+s-1} = i_s\} = \frac{1}{N^s}; i_1, \ldots, i_s \in A; s = 1, 2, \ldots; t = 1, 2, \ldots$$

Например, в простейшем случае при s=1 равномерность распределения 1-грамм проверятся с помощью частотного побитового теста (frequency (monobit) test), входящего в батарею NIST [3].

2. Тестирование *s*-мерной равномерности

Статистические тесты, проверяющие равномерность s-мерного распределения, основаны на вычислении частот s-грамм в наблюдаемой последовательности. Существуют два подхода к вычислению данных частот: по пересекающимся или по непересекающимся фрагментам последовательности. Пусть $X_1^T = (x_1, \ldots, x_T) \in A^T$ – выходная последовательность генератора длины $T = ms < \infty$, $\mathbf{1}\{B\}$ – индикаторная функция события B ($\mathbf{1}\{B\} = 1$, если B наступает, $\mathbf{1}\{B\} = 0$ в противном случае), тогда частоты s-грамм $I_1^s = (i_1, \ldots, i_s) \in A^s$, вычисленные по непересекающимся фрагментам X_1^T, имеют следующий вид:

$$\nu(I_1^s) = \sum_{t=1}^{m} \mathbf{1}\{X_{(t-1)s+1}^{ts} = I_1^s\} = \sum_{t=1}^{m} \mathbf{1}\{x_{(t-1)s+1} = i_1, \ldots, x_{ts} = i_s\}, \quad (1)$$

а частоты s-грамм I_1^s, вычисленные по пересекающимся фрагментам X_1^T, имеют вид:

$$\mu(I_1^s) = \sum_{t=1}^{T-s+1} \mathbf{1}\{X_t^{t+s-1} = I_1^s\} = \sum_{t=1}^{T-s+1} \mathbf{1}\{x_t = i_t, \ldots, x_{t+s-1} = i_s\}. \quad (2)$$

Используя частоты (1) и (2), строятся статистические оценки вероятностей s-грамм $p(I_1^s) = \mathrm{P}\{X_t^{t+s-1} = I_1^s\}$:

$$\hat{p}(I_1^s) = \frac{\nu(I_1^s)}{m}, \ \hat{q}(I_1^s) = \frac{\mu(I_1^s)}{T-s+1}.$$

Если верна гипотеза H_0={ X_1^T является РРСП}, то частоты $\nu(I_1^s)$ – суммы независимых, одинаково распределенных случайных величин, математическое ожидание и ковариации которых имеют вид:

$$\mathrm{E}\{\nu(I_1^s)\} = mp(I_1^s), \ \mathrm{cov}\{\nu(I_1^s), \nu(J_1^s)\} = mp(I_1^s)(\mathbf{1}\{I_1^s = J_1^s\} - p(J_1^s)),$$

а распределение вероятностей $\hat{p}(I_1^s)$ сходится при $m \to \infty$ к нормальному распределению с математическим ожиданием $p(I_1^s)$ и дисперсией

$$\frac{p(I_1^s)(1-p(I_1^s))}{m}.$$

Математическое ожидание оценок $\hat{q}(I_1^s)$ также, очевидно, равно $p(I_1^s)$. Кроме того, нормированные уклонения оценок

$$\overline{q}(I_1^s) = \sqrt{T-s+1}(\hat{q}(I_1^s) - p(I_1^s))$$

имеют асимптотическое нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей $\Sigma = \Sigma(I_1^s, J_1^s)$ [4]:

$$\Sigma(I_1^s, J_1^s) = \frac{\sigma(I_1^s, J_1^s)}{T-s+1} + O(1/T^2),$$

Величины $\sigma(I_1^s, J_1^s)$ имеют вид [5]:

$$\sigma(I_1^s, J_1^s) = p(I_1^s)(\mathbf{1}\{I_1^s = J_1^s\} - p(J_1^s)) + p(I_1^s)S(I_1^s, J_1^s) + p(J_1^s)S(J_1^s, I_1^s), \quad (3)$$

где

$$S(I_1^s, J_1^s) = \sum_{i=1}^{\infty} (p^{(i)}(I_1^s, J_1^s) - p(J_1^s)),$$

$p^{(i)}(I_1^s, J_1^s) = \mathrm{P}\{X_{t+i}^{t+i+s-1} = J_1^s, X_t^{t+s-1} = I_1^s\}$ – вероятность перехода из состояния I_1^s в состояние J_1^s за i шагов.

В задачах тестирования генераторов A = {0, 1}; тогда, если верна гипотеза H_0, то величины (3) принимают вид:

$$\sigma(I_1^s, J_1^s) = \frac{1}{2^s}(\mathbf{1}\{I_1^s = J_1^s\} - \frac{1}{2^s} + \sum_{i=1}^{s-1} \frac{1}{2^i}(\mathbf{1}\{I_{i+1}^s = J_1^{s-i}\} + \mathbf{1}\{J_{i+1}^s = I_1^{s-i}\}) - \frac{s-1}{2^{s-1}}).$$

Использование пересекающихся s-грамм для построения статистических тестов позволяет использовать больше информации о выходной последовательности, но требует более сложного теоретического обоснования.

3. Компьютерный эксперимент

Проиллюстрируем использование частот s-грамм, вычисляемых по пересекающимся и по непересекающимся фрагментам, при вычислении s-мерной энтропии двоичной последовательности:

$$\mathrm{H}_s = -\sum_{I_1^s \in \{0,1\}^s} p(I_1^s) \log_2 p(I_1^s).$$

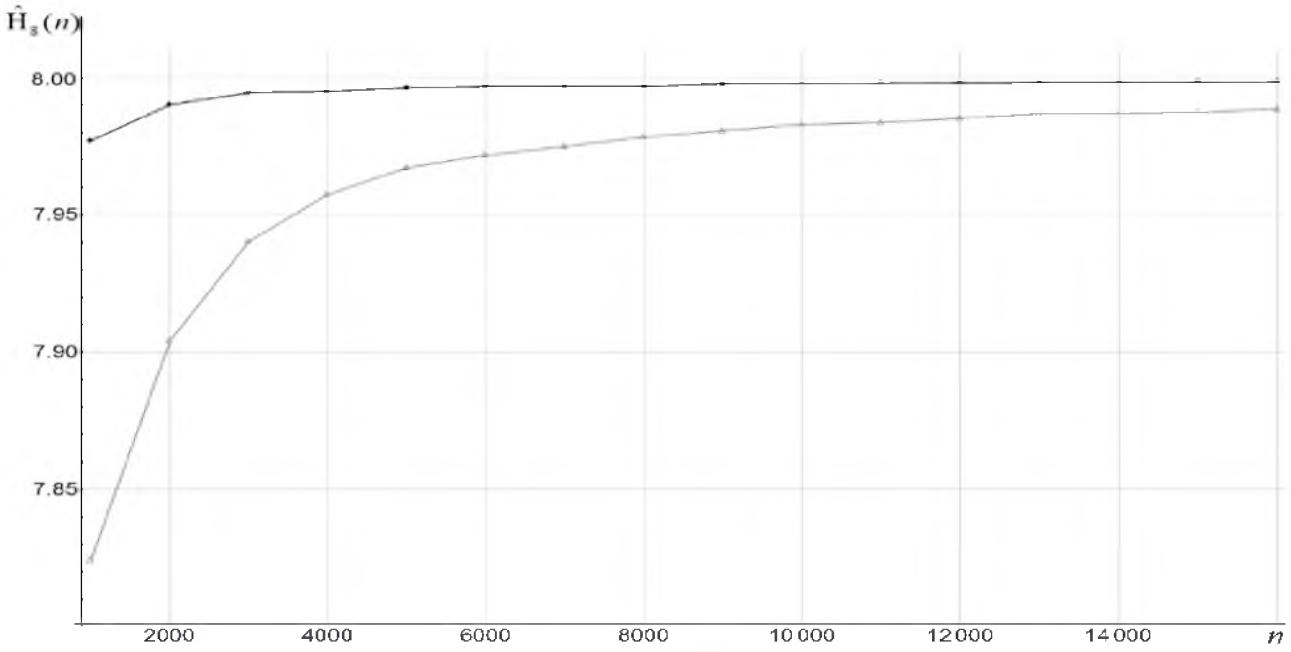

Рисунок 1. – Сравнение оценок 8-мерной энтропии, построенных по пересекающимся и по непересекающимся фрагментам двоичной последовательности

Для РРСП $H_s = s$, поэтому значимое отклонение от этой величины для выходной последовательности сигнализирует об уязвимости генератора. Рисунок 1 иллюстрирует поведение оценок энтропии $\hat{H}_s(n)$, построенных по пересекающимся (верхняя линия на графике) и непересекающимся (нижняя линия на графике) фрагментам длины 8 двоичной последовательности длины n = 1000, 2000,…, 16000, полученной с помощью алгоритма СТБ 34.101.31 (BelT) в режиме счетчика. График иллюстрирует превосходство в точности оценок, построенных по пересекающимся фрагментам.

Литература

1. Goldberg, I. Randomness and the Netscape browser / I. Goldberg, D. Wagner // Dr. Dobb's Journal. – 1996. – Vol. 1. – P. 66–71.

2. Криптология / Ю.С.Харин [и др.]. – Минск: БГУ, 2013. – 512 с.

3. A statistical test suite for random and pseudorandom number generators for cryptographic applications [Electronic resource]: National Institute of Standards and Technology Special Publication 800-22 Rev. 1a, 2010.

4. Maltsew, M.V. On testing randomness using frequency estimators / M.V. Maltsew, Yu.S. Kharin // 13-th International Conference Computer Data Analysis and Modeling: Stochastics and data science. Minsk. 2022. – P 112–115.

5. Billingsley, P. Statistical methods in Markov chains / P. Billingsley // The Annals of Mathematical Statistics. – 1961. – Vol. 32. – P. 12–40.