УДК 004.05 +519.2

СТАТИСТИЧЕСКОЕ ТЕСТИРОВАНИЕ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ В СИСТЕМАХ ЗАЩИТЫ ИНФОРМАЦИИ

акад. НАН Беларуси, д-р физ.-мат. наук, проф. Ю. С. ХАРИН,
канд. физ.-мат. наук, доц. В. Ю. ПАЛУХА,
канд. физ.-мат. наук, доц. М. В. МАЛЬЦЕВ
(НИИ прикладных проблем математики и информатики БГУ,
г. Минск, Беларусь)

Аннотация. *Рассматривается применение вероятностно-статистических методов для анализа качества генераторов случайных и псевдослучайных числовых последовательностей, используемых в системах защиты информации. Для решения данной задачи используются малопараметрические марковские модели и энтропийные характеристики.*

Ключевые слова: *статистическое тестирование, криптографический генератор, малопараметрическая марковская модель, энтропийный профиль.*

Введение. Критически важными элементами систем защиты информации являются генераторы случайных и псевдослучайных числовых последовательностей. Последовательность, вырабатываемая стойким генератором, не должна отличаться по своим свойствам от равномерно-распределенной случайной последовательности (РРСП). Основным методом оценки качества генераторов является статистическое тестирование. Известные наборы (батареи) тестов обладают рядом недостатков и ограничений: они проверяют простую нулевую гипотезу, не фиксируют семейство альтернатив, могут не обнаруживать сравнительно простые зависимости [1]. В связи с этим актуальной является разработка методов и алгоритмов, позволяющих более эффективно выявлять зависимости в выходных последовательностях генераторов. Направлениями, показавшими свою эффективность на практике, являются статистическое тестирование на основе сложных малопараметрических марковских моделей [2] и на основе энтропийных характеристик [3].

1. Статистическое тестирование на основе малопараметрических марковских моделей. Известной малопараметрической моделью является разработанная в Белорусском государственном университете цепь Маркова порядка s с r частичными связями [4]. В настоящей статье представлено обобщение данной модели для векторной цепи Маркова с $m \geq 2$ компонентами. Обозначим: $A = \{0, 1, ..., N-1\}$ – множество мощности $|A| = N \geq 2$; $m \in \mathbb{N}$ – размерность состояния цепи Маркова,

$J_i = (j_{i_1}, \ldots, j_{i_m}) \in A^m,\ i \in \mathbb{N},$ – m-мерный целочисленный вектор; $J_a^b = (J_a, J_{a+1}, \ldots, J_b)$ – упорядоченный набор m-мерных векторов; $\{x_t = (x_{t_1}, \ldots, x_{t_m}) \in A^m : t \in \mathbb{N}\}$ – однородная векторная цепь Маркова порядка s с пространством состояний A^m с матрицей вероятностей одношаговых переходов $P = (p_{J_1^s, J_{s+1}})$:

$$p_{J_1^s, J_{s+1}} = P\{x_t = J_{s+1} \mid x_{t-1} = J_s, \ldots, x_{t-s} = J_1\},\ J_1, \ldots, J_{s+1} \in A^m,\ t \in \{s+1, s+2, \ldots\}. \quad (1)$$

Такую цепь Маркова будем обозначать ВЦМ(s) – векторная цепь Маркова порядка s.

Число независимых элементов матрицы P, равное $N^{ms}(N^m - 1)$, возрастает экспоненциально при увеличении s, и применение этой модели на практике возможно лишь при небольших значениях параметров. В связи с этим, построена модификация ВЦМ(s), для которой условное распределение вероятностей определяется лишь некоторыми «значимыми» компонентами предыдущих векторов-состояний. Обозначим:

$$M_r = \{(k_1, l_1), (k_2, l_2), \ldots, (k_r, l_r)\} \subseteq M_* = \{(k, l) : 1 \le k \le s, 1 \le l \le m\} -$$

множество, представляющее собой упорядоченный в лексикографическом порядке набор $1 \le r \le sm$ различных значений пар индексов, причем $k_1 = 1$. Множество M_r называется шаблоном связей или просто шаблоном. Определим также функцию-селектор $S_{M_r}(J_t, \ldots, J_{t+s-1}) = (j_{t+k_1-1, l_1}, \ldots, j_{t+k_r-1, l_r}),\ t \in \mathbb{N},$ которая в соответствии с шаблоном M_r «вырезает» r компонент из множества ms компонент $\{j_{u,l} :\ t \le u \le t+s-1,\ 1 \le l \le m\}$.

Если вероятности (1) допускают следующее представление:

$$p_{J_1^s, J_{s+1}} = q_{S_{M_r}(J_1, \ldots, J_s), J_{s+1}} = q_{(j_{k_1, l_1}, \ldots, j_{k_r, l_r}), J_{s+1}},\ J_1, \ldots, J_{s+1} \in A^m,$$

где $Q = (q_{(i_1, \ldots, i_r), I_{r+1}})$ – некоторая стохастическая $N^r \times N^m$ – матрица, $i_1, \ldots, i_r \in A$, $I_{r+1} \in A^m$, то ВЦМ(s) называется векторной цепью Маркова с r частичными связями и шаблоном связей M_r (ВЦМ(s, r)). Условное распределение вероятностей состояния x_t для ВЦМ(s, r) в момент времени t зависит не от всех ms компонент s прошлых состояний, а только от r избранных компонент, которые определяются шаблоном M_r.

Разработан алгоритм идентификации ВЦМ(s, r) по реализации длины n: $X^{(n)} = (x_1, \ldots, x_n)$, $x_1, \ldots, x_n \in A^m$; построен статистический тест для обнаружения отклонений в $X^{(n)}$ от РРСП на основе ВЦМ(s, r) (гипотезе H_0 соответствует РРСП):

$$\text{принимается } \begin{cases} | H_0, \text{если } \rho_n \le \Delta, \\ | H_1 = \overline{H}_0, \text{если } \rho_n > \Delta, \end{cases} \tag{2}$$

где $\rho_n = \sum_{i_1,\ldots,i_r \in A} \sum_{l_{r+1} \in A^m} \overline{q}^2_{(i_1,\ldots,i_r),l_{r+1}} \nu^{M_r}_{s+1}\left(i_1,\ldots,i_r,l_{r+1}\right) / q^{(0)}_{(i_1,\ldots,i_r),l_{r+1}}$; $Q^{(0)} = (q^{(0)}_{(i_1,\ldots,i_r),l_{r+1}})$ – стохастическая матрица размерности $N^r \times N^m$, все элементы которой равны $1 / N^m$;

$\overline{q}^2_{(i_1,\ldots,i_r),l_{r+1}} = (q^{(0)}_{(i_1,\ldots,i_r),l_{r+1}} - \hat{q}_{(i_1,\ldots,i_r),l_{r+1}}) / \sqrt{n-s}$, $\hat{q}_{(i_1,\ldots,i_r),l_{r+1}}$ – оценки максимального правдоподобия вероятностей переходов;

$\nu^{M_r}_{s+1}(i_1, \ldots, i_r, l_{r+1})$ – частоты состояний ВЦМ(s, r);

$\Delta = G_y^{-1}(1-\alpha)$, G_y – функция стандартного χ^2-распределения с y степенями свободы, $\alpha \in (0, 1)$ – уровень значимости.

В компьютерных экспериментах с помощью алгоритма статистического тестирования, основанного на (2), выявлены отклонения от РРСП в генераторе rand стандартной библиотеки языка C – stdlib для реализации размера 10 МБ, тогда как тестирование на основе батареи NIST не выявило отклонения от «чистой случайности» в аналогичной последовательности размера 2 ГБ [5].

2. Статистическое тестирование генераторов на основе оценок энтропии. В качестве тестовых статистик могут выступать статистические оценки функционалов информационной энтропии, вычисленные по наблюдаемой двоичной последовательности. Пусть x – случайная величина из алфавита мощности $N = 2^s$ с дискретным распределением вероятностей $p = \{p_k\}, p_k = P\{x = \omega_k\}$, $\sum_{k=1}^{N} p_k = 1$, $k = 1, \ldots, N$, и пусть наблюдается случайная последовательность $\{x_t : t = 1, \ldots, n\}$ объёма n из распределения вероятностей $\{p_k\}$. Частотные оценки вероятностей имеют вид

$$\hat{p}_k = \frac{v_k}{n}, \quad v_k = \sum_{t=1}^{n} I\{x_t = \omega_k\} \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}, \quad I\{x_t = \omega_k\} = \begin{cases} |1, x_t = \omega_k; \\ |0, x_t \ne \omega_k. \end{cases}$$

Рассмотрим асимптотику соразмерного увеличения объёма выборки и мощности алфавита:

$$n,\ N \to \infty,\ \frac{n}{N} \to \lambda,\ 0 < \lambda < \infty. \tag{3}$$

В таблице приведены формулы вычисления оценок энтропии Шеннона, Реньи и Тсаллиса, для которых в [3] при истинной гипотезе H_0 в асимптотике (3) доказана асимптотическая нормальность, а также параметры асимптотически нормального распределения. Для построения несмещённых оценок функционалов энтропии Реньи и Тсаллиса используется факториальная степень $x^{\underline{2}} = x(x-1)$.

Таблица. – Оценки функционалов энтропии и параметры их распределения

Тип	Оценка	Мат. ожидание	Дисперсия
Шеннон	$\hat{H} = \ln n - \frac{1}{n}\sum_{k=1}^{N} v_k \ln v_k$	$\mu_H = \ln n - e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\ln(k+1)\lambda^k}{k!}$	$\sigma_H^2 = \frac{e^{-\lambda}}{n}\sum_{k=1}^{+\infty}\frac{(k+1)\lambda^k}{k!}\ln^2(k+1) - \frac{e^{-2\lambda}}{N}\left(\sum_{k=1}^{+\infty}\frac{\ln(k+1)\lambda^k}{k!}\right)^2 - \frac{e^{-2\lambda}}{n}\left(\sum_{k=1}^{+\infty}\ln(k+1)\frac{\lambda^k}{k!}(k+1-\lambda)\right)^2$
Реньи	$\hat{H}_2 = 2\ln n - \ln\sum_{k=1}^{N} v_k^{\underline{2}}$	$\mu_{H,2} = \ln N$	$\sigma_{H,2}^2 = \frac{2}{n\lambda}$
Тсаллис	$\hat{S}_2 = 1 - \frac{1}{n^2}\sum_{k=1}^{N} v_k^{\underline{2}}$	$\mu_{S,2} = 1 - \frac{1}{N}$	$\sigma_{S,2}^2 = \frac{2}{Nn^2}$

Пусть $\alpha \in (0,1)$ – уровень значимости, $\hat{h}$ – статистическая оценка энтропии Шеннона, Реньи или Тсаллиса, μ_h и σ_h^2 – асимптотические математическое ожидание и дисперсия этих оценок при истинной гипотезе H_0. Вычислим $\hat{h}$ для наблюдаемой последовательности. Решающее правило, основанное на статистике $\hat{h}$, имеет вид [3]:

$$\text{принимается}\begin{cases} \mid H_0, & \text{если } t_- < \hat{h} < t_+; \\ \mid \overline{H_0}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu_h \pm \sigma_h \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{4}$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона.

Вычислим нормированную статистику $\tilde{h} = \left(\hat{h} - \mu_h\right)/\sigma_h$ которая в асимптотике (3) и при истинной гипотезе H_0 имеет стандартное нормальное распределение: $\tilde{h} \sim \mathcal{N}(0,1)$. Следовательно, двустороннее p-значение для неё равно

$$p - value = 2\left(1 - \Phi\left(\left|\tilde{h}\right|\right)\right). \tag{5}$$

Пусть наблюдается двоичная последовательность $\{y_\tau\}$, $\tau = 1, \ldots, T$. Из непересекающихся фрагментов длины s (s-грамм) $X^{(t)} = (X_j^{(t)}) = (y_{(t-1)s+1}, \ldots, y_{ts}) \in \{0,1\}^s$, $t = 1, \ldots, n = [T/s]$, сформируем новую последовательность $\{x_t\}$ из алфавита мощности $N = 2^s$ по правилу $x_t = \sum_{j=1}^{s} 2^{j-1} X_j^{(t)} + 1$. На основе критерия (4) вычислим последовательность нормированных отклонений оценки энтропии от математического ожидания в зависимости от s, которую назовём энтропийным профилем:

$$\chi(s) = \frac{\hat{h}(s) - \mu_h(s)}{\sigma_h(s)\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)} = \frac{\tilde{h}(s)}{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}, s = s_-, \ldots, s_+. \tag{6}$$

Аналогично строятся последовательности p-значений (3).

Разработанный в НИИ ППМИ программный комплекс «Энтропийный анализ дискретных последовательностей» (ЭАДП) реализует критерий (4). Реализована возможность отображения оценок энтропии $\hat{h}$, нормированных значений (6), p-значений (5). Главное окно программного комплекса представлено на рисунке.

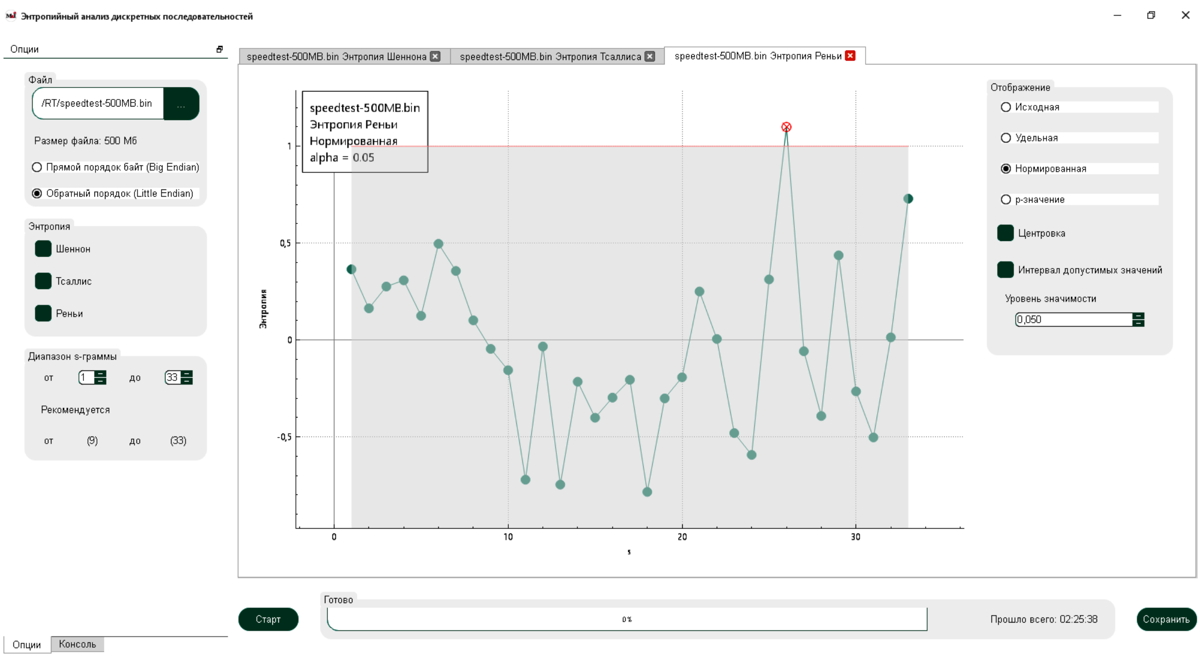

Рисунок. – Программный комплекс «ЭАДП»

ЛИТЕРАТУРА

1. Зубков, А.М. Проверка пакета статистических критериев NIST на специальных псевдослучайных последовательностях / А.М.Зубков, А.А.Серов // Математические вопросы криптографии. – 2019. – Т. 10, вып 2. – С. 89–96.

2. Kharin, Yu.S. Parsimonious models of high-order Markov chains for evaluation of cryptographic generators / Yu.S. Kharin // Математические вопросы криптографии, том 7, выпуск 2. – С. 131–142.
3. Палуха, В.Ю. Статистические тесты на основе оценок энтропии для проверки гипотез о равномерном распределении случайной последовательности / В.Ю. Палуха // Весці НАН Беларусі. Серыя фізіка-матэматычных навук. – 2017. – № 1. – С. 79–88.
4. Харин, Ю. С. Цепи Маркова с r-частичными связями и их статистическое оценивание / Ю. С. Харин // Доклады НАН Беларуси. – 2004. – Т. 48, № 1. – С. 40–44.
5. Харин, Ю.С. Применение специальных марковских моделей для оценки качества криптографических генераторов / Ю.С.Харин, М.В.Мальцев // Комплексная защита информации: материалы XXV научно-практической конференции, Россия, 15–17 сентября 2020 года. – С. 224–228.