ЗАСЕДАНИЕ № 4

КРИПТОГРАФИЯ И ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 004.421.6: 519.23

О ПРИМЕНЕНИИ ПРОГРАММНОГО КОМПЛЕКСА «ЭАДП» ДЛЯ СТАТИСТИЧЕСКОГО ТЕСТИРОВАНИЯ ВЫХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ

В.Ю. ПАЛУХА, Ю.С. ХАРИН

НИИ прикладных проблем математики и информатики,
Белорусский государственный университет
г. Минск, 220030, Республика Беларусь

Введение

Одним из важнейших элементов систем криптографической защиты информации (СКЗИ) являются генераторы случайных и псевдослучайных последовательностей. Стойкость СКЗИ зависит от того, насколько близка генерируемая последовательность по своим свойствам к равномерно распределенной случайной последовательности (РРСП) [1], которая на практике называется «чисто случайной» последовательностью.

Для проверки качества криптографических генераторов используются статистические тесты, в которых проверяется гипотеза $H_* = \{\{x_t\}$ является РРСП$\}\}$ о том, что наблюдаемая последовательность $\{x_t\}$ является равномерно распределенной случайной последовательностью. В НИИ ППМИ разработан программный комплекс «Энтропийный анализ дискретных последовательностей» (ПК «ЭАДП») [2], в котором в качестве тестовых статистик используются статистические оценки функционалов энтропии Шеннона, Реньи и Тсаллиса (с параметром $r = 2$).

Оценки энтропии. Пусть на вероятностном пространстве (Ω, F, P) определена случайная величина x из алфавита мощности $N = 2^s$ с дискретным распределением вероятностей $p = \{p_k\}, p_k = P\{x = \omega_k\}$, $p_k \geq 0, \sum_{k=1}^{N} p_k = 1, k = 1, \ldots, N$. В таблице 1 приведены формулы функционалов энтропии, используемых в ПК «ЭАДП».

Таблица 1. – Функционалы энтропии

Энтропия Шеннона	$H(p) = -\sum_{i=1}^{N} p_i \ln p_i$
Энтропия Реньи ($r = 2$)	$H_2(p) = -\ln\left(\sum_{i=1}^{N} p_i^2\right)$
Энтропия Тсаллиса ($r = 2$)	$S_2(p) = 1 - \sum_{i=1}^{N} p_i^2$

Пусть имеется случайная последовательность $\{x_t : t = 1, \ldots, n\}$ объема n из распределения вероятностей $\{p_k\}$, по которой будет оцениваться энтропия. Частотные оценки вероятностей имеют вид

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

Рассмотрим асимптотику совместного увеличения объема выборки и сложности задачи:

$$n, N \to \infty, n/N \to \lambda, 0 < \lambda < \infty. \tag{2}$$

В асимптотике (2) для распределения вероятностей статистик $\{\nu_k\}$ справедлива аппроксимация законом Пуассона $\Pi(\lambda_k)$ с параметром $\lambda_k = np_k$. При истинной гипотезе H_* все элементарные вероятности равны: $p_k = 1/N, \quad k = 1, \ldots, N$, поэтому все частоты $\{\nu_k\}$ имеют одинаковый параметр распределения Пуассона $\lambda = n/N$.

Оценка энтропии Шеннона на основе частотных статистик (1) имеет вид:

$$\hat{H} = \hat{H}(n, N) = -\sum_{k=1}^{N} \hat{p}_k \ln \hat{p}_k = -\sum_{k=1}^{N} \frac{\nu_k}{n} \ln \frac{\nu_k}{n} = \ln n - \frac{1}{n} \sum_{k=1}^{N} \nu_k \ln \nu_k. \tag{3}$$

Теорема 1 [3]. *В асимптотике* (2) *статистика* (3) *при гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*

$$\mu_H = \ln n - e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!}, \tag{4}$$

$$\sigma_H^2 = \frac{e^{-\lambda}}{n} \sum_{k=1}^{+\infty} \frac{(k+1)\lambda^k}{k!} \ln^2(k+1) - \frac{e^{-2\lambda}}{N} \left(\sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!} \right)^2 - \\ - \frac{e^{-2\lambda}}{n} \left(\sum_{k=1}^{+\infty} \ln(k+1) \frac{\lambda^k}{k!} (k+1-\lambda) \right)^2. \tag{5}$$

Как следует из теоремы 1, в асимптотике (2) оценка (3) является смещенной, и с уменьшением λ смещение растет. Для функционалов энтропии Реньи и Тсаллиса можно построить несмещенную оценку в асимптотике (2), в том числе и при $\lambda < 1$. Для построения несмещенной оценки воспользуемся факториальной степенью $x^{\underline{2}} = x(x-1)$.

Статистические оценки энтропии Реньи и Тсаллиса, построенные с использованием факториальной степени, имеют вид

$$\hat{H}_2(n, N) = -\ln \left(\sum_{k=1}^{N} \hat{p}_k^2 \right) = 2 \ln n - \ln \sum_{k=1}^{N} \nu_k^{\underline{2}}, \tag{6}$$

$$\hat{S}_2(n, N) = 1 - \sum_{k=1}^{N} \hat{p}_k^2 = 1 - \frac{1}{n^2} \sum_{k=1}^{N} \nu_k^{\underline{2}}. \tag{7}$$

Теорема 2 [3]. *В асимптотике* (2) *статистика* (6) *является состоятельной оценкой энтропии Реньи и при истинной гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*:

$$\mu_{H,r} = \ln N, \quad \sigma^2_{H,2} = \frac{2}{n\lambda}. \tag{8}$$

Теорема 3 [3]. *В асимптотике* (2) *статистика* (7) *является состоятельной асимптотически несмещенной оценкой энтропии Тсаллиса и при истинной гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*:

$$\mu_{S,2} = 1 - \frac{1}{N}, \quad \sigma^2_{S,2} = \frac{2}{Nn^2}. \tag{9}$$

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

$$\text{принимается} \begin{cases} H_*, & \text{если } t_- < h < t_+; \\ \overline{H_*}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu_h \pm \sigma_h \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{10}$$

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

Вычислим нормированную статистику

$$\tilde{h} = \frac{h - \mu_h}{\sigma_h}. \tag{11}$$

Она в асимптотике (1) и при истинной гипотезе H_* имеет стандартное нормальное распределение: $\tilde{h} \sim \mathcal{N}(0,1)$. Следовательно, двустороннее p-значение для нее равно

$$p - value = 2\left(1 - \Phi\left(\left|\tilde{h}\right|\right)\right). \tag{12}$$

Пусть наблюдается двоичная последовательность $\{y_\tau\}$, $\tau = 1, \ldots, T$. Из непересекающихся подряд идущих фрагментов длины s (s-грамм) $X^{(t)} = (X_j^{(t)}) = \quad = (y_{(t-1)s+1}, \ldots, y_{ts}) \in \{0,1\}^s$, $t = 1, \ldots, n = [T/s]$, 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 (10) вычислим последовательность нормированных отклонений оценки энтропии от математического ожидания в зависимости от s, которую назовем энтропийным профилем:

$$\chi(s) = \frac{\hat{h}(s) - \mu_h(s)}{\sigma_h(s)\Phi^{-1}(1-\alpha/2)} = \frac{\tilde{h}(s)}{\Phi^{-1}(1-\alpha/2)}, \; s = s_-, \ldots, s_+. \tag{13}$$

Аналогично строятся последовательности p-значений (12).

Тестирование с помощью профиля позволяет выносить решение о принятии или отклонении гипотезы H_* на основе решающего правила (10) по совокупности значений $\chi(s)$

для различных s; такое решение видится более аргументированным, чем при принятии его по результатам применения теста (15) для отдельного значения s. Показана слабая зависимость оценок энтропии по соседним длинам фрагментов [4], тем самым получено обоснование использования энтропийных профилей.

Программный комплекс «ЭАДП». Разработанный в НИИ ППМИ программный комплекс «ЭАДП» реализует критерий (10). В начале работы необходимо выбрать файл с последовательностью, порядок Big Endian или Little Endian, диапазон s и функционалы энтропии. Вычисляемые значения добавляются на экран в режиме реального времени. Имеется возможность изменять уровень значимости α без пересчета оценок энтропии и переключаться на различные режимы отображения: непосредственно оценки энтропии h, нормированные значения (13), p-значения (12). Помимо вывода самих значений в консоль, программа отображает графики зависимостей этих величин от длины фрагмента s. Главное окно программного комплекса с результатами работы представлено на рисунке 1.

Протоколы тестирования с помощью ПК «ЭАДП». Работа ПК «ЭАДП» продемонстрирована путем анализа последовательностей, полученных генераторами случайных и псевдослучайных чисел.

На рисунке 1 отображен энтропийный профиль Реньи последовательности speedtest-500MB.bin физического генератора [5]. Длина последовательности $T = 125 \cdot 2^{25}$. Длина фрагмента $s \in [1,35]$. Уровень значимости $\alpha = 0{,}05$. Как видно из рисунка 1, оценки энтропии попадают в интервал допустимых значений при всех значениях s, кроме $s = 26$. Для рассматриваемого генератора гипотеза H_* принимается.

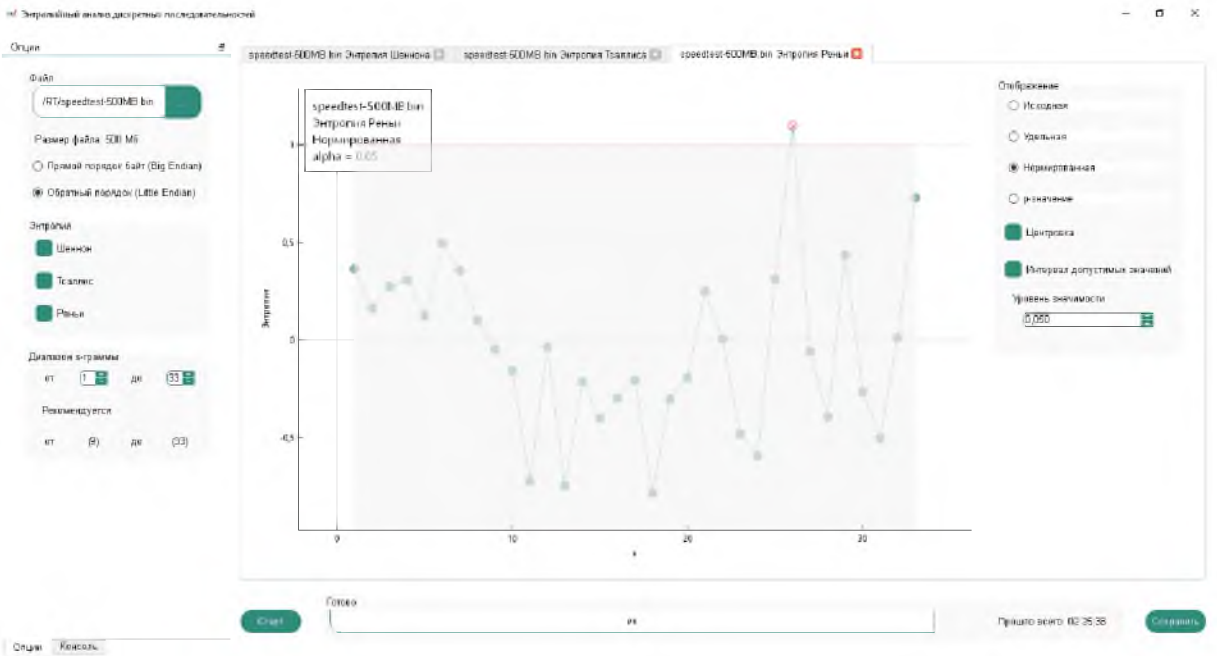

Рисунок 1. – Программный комплекс «ЭАДП». Нормированная энтропия Реньи последовательности speedtest-500MB.bin физического генератора

На рисунке 2 приведен протокол тестирования последовательности алгоритма шифрования BelT [6] в режиме счетчика с нулевыми ключом и синхропосылкой. Длина последовательности $T = 2^{30}$. Длина фрагмента $s \in [1,32]$. Уровень значимости $\alpha = 0{,}05$. Отображены p-значения критерия на основе оценок энтропии Тсаллиса. Как видно из рисунка 2, оценки энтропии попадают в интервал допустимых значений при всех значениях s, кроме $s = 1$ и $s = 3$. Для рассматриваемого генератора гипотеза H_* принимается.

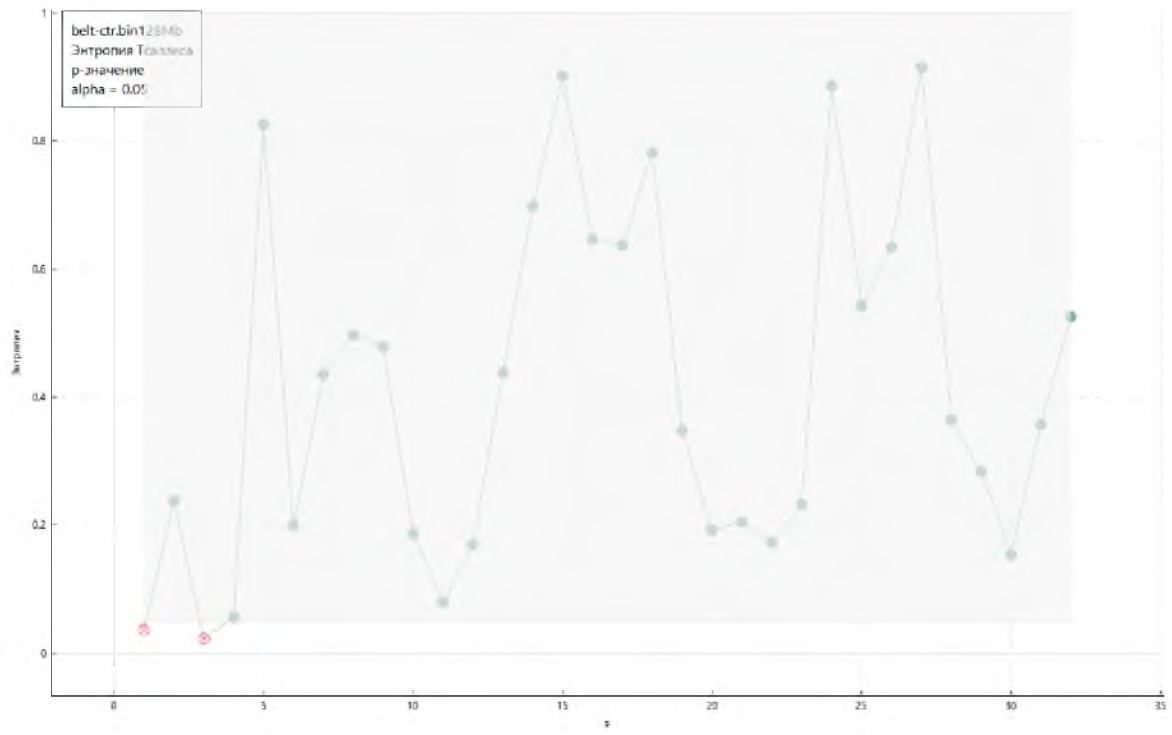

Рисунок 2. – p-значение теста на основе энтропии Тсаллиса последовательности BelT в режиме счетчика

На рисунке 3 приведен протокол тестирования последовательности самосжимающего генератора [1] на основе регистра сдвига с линейной обратной связью с характеристическим многочленом $x^{10} + x^3 + 1$. Длина последовательности $T = 25 \cdot 2^{25}$. Длина фрагмента $s \in [1, 25]$. Уровень значимости $\alpha = 0{,}05$. Отображены оценки энтропии Шеннона. Как видно из рисунка 3, начиная с $s = 7$ оценки энтропии не попадают в интервал допустимых значений. Кроме того, наблюдается закономерность в значениях оценок. Для рассматриваемого генератора гипотеза H_* отклоняется.

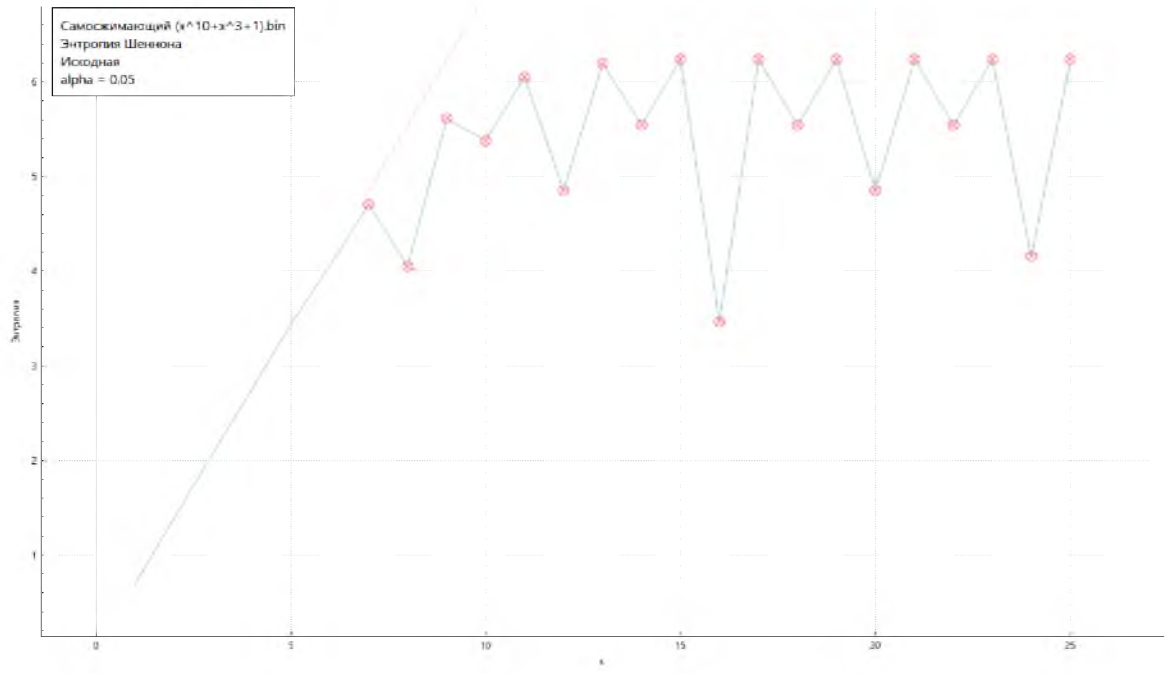

Рисунок 3. – Энтропия Шеннона последовательности самосжимающего генератора

Литература

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

2. Программный комплекс «Энтропийный анализ дискретных последовательностей» [Электронный ресурс]. – Режим доступа: http://www.product.bsu.by/katalog/informacionnie-tehnologii/informacionnie-naukoemkie-tehnologii/statistika/programmnoe-sredstvo-entropijnogo-analiza-diskretnih-sluchajnih-posledovatel_nostej/. – Дата доступа: 31.03.2023.

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. Palukha, U.Yu. On statistical testing of random and pseudorandom sequences based on entropy functionals / U.Yu. Palukha, Yu.S. Kharin, A.I. Siarheeu, A.A. Arlou // Computer Data Analysis and Modeling: Stochastics and Data Science. Proceedings of the XIII International Conference, Minsk, September 6–10, 2022 / BSU; eds.: Yu. Kharin [et al.]. – Minsk: BSU, 2022. – P. 148–162.

5. speedtest-500MB.bin [Electronic resource] // Humboldt Berlin University, Faculty of Mathematics and Natural Sciences, Department of Physics. – Mode of access: http://qrng.physik.hu-berlin.de/files/speedtest-500MB.bin. – Date of access: 31.03.2022.

6. Информационные технологии и безопасность. Алгоритмы шифрования и контроля целостности = Інфармацыйныя тэхналогіі і бяспека. Алгарытмы шыфравання і кантролю цэласнасці: СТБ 34.101.31-2020. – Взамен СТБ 34.101.31-2011; введ. 01.10.20. – Минск: Белорус. гос. ин-т стандартизации и сертификации, 2020. – 44 с.