

ЗАСЕДАНИЕ № 2
СОВРЕМЕННЫЕ МЕТОДЫ И СРЕДСТВА КРИПТОГРАФИЧЕСКОЙ
ЗАЩИТЫ ИНФОРМАЦИИ

УДК 004.421.6:519.23

АКТУАЛЬНЫЕ НАПРАВЛЕНИЯ ОЦЕНИВАНИЯ СТОЙКОСТИ
КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ
И ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Ю. С. ХАРИН

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

Введение

Наряду с разработкой государственных стандартов криптографической защиты информации актуальным направлением криптологии является разработка и анализ качества криптографических генераторов случайных и псевдослучайных последовательностей [1, 2]. Генерация РРСП представляет собой важную задачу не только для генерации «гаммы» в поточных криптосистемах, но и в других системах криптографической защиты информации для выработки ключей и векторов инициализации, одноразовых чисел (Nonce), в криптографических протоколах и схемах ЭЦП. В криптографии генерация случайности сводится к задаче генерации равномерно распределенной случайной последовательности (РРСП) $x_t \in \{0, 1\}$, $t = 1, 2, \dots$, для которой при любом $s = 1, 2, \dots$ s -мерное распределение вероятностей является равномерным: $\mathbf{P}\{x_1 = j_1, \dots, x_s = j_s\} = 2^{-s}$. Для анализа качества выходных последовательностей длины T криптографических генераторов используются наборы (так называемые «батареи») статистических тестов проверки нулевой гипотезы $H_0 = \{x_t \text{ есть РРСП}\}$ против альтернативы $H_1 = \overline{H_0}$ [1 – 3].

Проблемы статистического тестирования и подходы к их разрешению

Проведенный в [2] обзор существующих тестов выявляет следующие недостатки: 1) многие из тестов ориентированы на проверку лишь частных случаев альтернативы $\overline{H_0}$; 2) многие тесты вообще не фиксируют семейство альтернатив и не имеют оценок мощности w_T ; 3) при включении нескольких тестов в «батарею» не удастся учесть вероятностную зависимость тестовых статистик; 4) известны примеры генераторов псевдослучайных последовательностей, «уверенно» преодолевающих батареи тестов [4]. Особенно значимой критике подвергнута «батарея» тестов NIST STS [4 – 7], что привело к решению Национального института США по стандартизации о начале «ревизии» «батареи тестов» [8].

Для частичного преодоления указанных недостатков предлагается: строить статистические тесты s -мерной равномерности для заданных семейств альтернатив $\overline{H_0}$, использовать энтропийные профили [9], находить асимптотические оценки мощностей известных тестов [10] и применять сложные марковские модели [11 – 15].

Существующие статистические тесты и построенные на их основе батареи обладают еще одним общим существенным недостатком: некорректность модели простой нулевой гипотезы H_0 и необходимость использования сложной нулевой гипотезы при статистическом тестировании.

Этот недостаток приводит к следующему парадоксу: при увеличении длины $T \rightarrow +\infty$ наблюдаемых последовательностей, порождаемых реально существующими генераторами случайности, вероятность принятия альтернативы $\overline{H_0}$ стремится к единице: $w_T \rightarrow 1$. Иначе говоря, «увеличением длины исследуемой последовательности T можно забраковать любой реальный генератор случайности». Объяснение этого «парадокса» состоит в следующем. Реальные генераторы случайности – не идеальны, они отклоняются от гипотезы H_0 по своему вероятностному распределению на некоторую величину $\varepsilon > 0$. Все тесты, включаемые в батареи тестов, как известно, обладают оптимальным свойством проверки простой гипотезы H_0 против сложной альтернативы $\overline{H_0}$ – свойством состоятельности: $w_T \rightarrow 1$ при $T \rightarrow +\infty$. Таким образом, этот «парадокс» порожден некорректностью математической модели нулевой гипотезы H_0 . Для того, чтобы избежать этого «парадокса» нулевая гипотеза должна быть сложной: она должна задавать величину допуска $\varepsilon > 0$ отклонений от простой гипотезы H_0 .

Пусть наблюдается двоичная случайная последовательность длины $T = n \cdot s$, разбитая на n последовательных s -грамм ($s \in \mathbb{N}$):

$$X = X_1^T = (x_1, \dots, x_T) \equiv (X_1^s, X_{s+1}^{2s}, \dots, X_{(n-1)s+1}^T) \in V^T, V = \{0, 1\}; \quad (1)$$

эти s -граммы независимы в совокупности и одинаково распределены с некоторым распределением $p = (p_{J_1^s}) \in \mathcal{P}$:

$$\mathbf{P}\{X_{(i-1)s+1}^{is} = J_1^s\} ::= \mathbf{P}\{x_{(i-1)s+1} = j_1, \dots, x_{is} = j_s\} = p_{J_1^s}, J_1^s = (j_k) \in V^s, i = 1, \dots, n, \quad (2)$$

$$\mathcal{P} = \left\{ p = (p_{J_1^s}) : p_{J_1^s} \geq 0, J_1^s \in V^s, \sum_{J_1^s \in V^s} p_{J_1^s} = 1 \right\} - \quad (3)$$

семейство (симплекс) всевозможных s -мерных вероятностных распределений на V^s . Для упрощения обозначений (1) – (3) перейдем от мультииндекса J_1^s к одномерному индексу $k \in \{0, 1, \dots, 2^s - 1\}$:

$$k = \langle J_1^s \rangle ::= \sum_{i=1}^s j_i \cdot 2^{i-1} \in \{0, 1, \dots, K-1\}, K=2^s;$$

$$J_1^s = (j_1, \dots, j_s) = \langle k \rangle, j_{s-i+1} = \left\lceil \left(k - \sum_{l=s-i+2}^s j_l \cdot 2^{l-1} \right) / 2^{s-i} \right\rceil, i = 1, 2, \dots, s.$$

В существующих «батареях тестов» [3, 4], подвергнутых многочисленным модификациям и критике [4 – 8], используются тесты статистической проверки простой нулевой гипотезы

$$H_0 = \{p = p_0\}, \quad p_0 = (p_{0k}), \quad p_{0k} \equiv \frac{1}{K}, \quad (4)$$

против сложной альтернативы $\overline{H_0}$, приводящие к указанному выше парадоксу.

Сложная нулевая гипотеза и ее использование в тестировании

Во избежание парадокса «неадекватности модели (4) простой нулевой гипотезы» введем в рассмотрение сложную нулевую гипотезу об s -мерной равномерности:

$$H_0^\varepsilon = \{p = (p_k) \in \mathcal{P}_0^\varepsilon\}, \quad \mathcal{P}_0^\varepsilon = \left\{ p \in \mathcal{P} : \|p - p_0\| = \sqrt{\sum_{k=0}^{K-1} \left(p_k - \frac{1}{K}\right)^2} \leq \varepsilon \right\} - \quad (5)$$

пересечение симплекса $\mathcal{P}$ с гипершаром в R^K заданного радиуса ε с центром в точке p_0 s -мерного равномерного распределения; $\varepsilon \in (0, \sqrt{1-1/K})$ – достаточно малый параметр нулевой гипотезы (5), определяющий допустимые отклонения от простой гипотезы H_0 . Гипотеза H_0 является предельной по отношению к $H_0^\varepsilon : H_0^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} H_0$.

В НИИППМИ построен [13] критерий отношения правдоподобия для проверки сложных гипотез H_0^ε , $H_1^\varepsilon = \overline{H_0^\varepsilon}$ по наблюдаемой выходной последовательности генератора $X \in V^T$ длины T , определяемый решающим правилом:

$$d = d(X) = \begin{cases} 0, & \text{если } \text{sign}(\|\hat{p} - p_0\| - \varepsilon) \times (-2\tilde{\lambda}(\hat{p})/K)^{1/2} \leq \Delta_n, \\ 1, & \text{в противном случае,} \end{cases} \quad (6)$$

$$\tilde{\lambda}(\hat{p}) = H(\hat{p}) - \ln K + \varepsilon K \|\hat{p} - p_0\| - \frac{K\varepsilon^2}{2}, \quad \hat{p} = (\hat{p}_k), \quad \hat{p}_k = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_{(i-1)s+1}^{is} = k\},$$

$$H(\hat{p}) = -\sum_{k=0}^{K-1} \hat{p}_k \ln \hat{p}_k \geq 0 - \text{подстановочная оценка энтропии Шеннона } s\text{-граммы};$$

$$\Delta_n = \frac{\Phi^{-1}(1-\alpha_0)}{\sqrt{n}} \sqrt{\frac{1}{K} + \varepsilon(e_K^+ - \varepsilon)}, \quad e_K^+ = \frac{1-2/K}{\sqrt{1-1/K}} \leq 1,$$

$\alpha_0 \in (0, 1)$ – задаваемый асимптотический размер теста; $\Phi^{-1}(\cdot)$ – квантильная функция стандартного нормального закона; $\mathbf{1}\{C\}$ – индикатор события C .

Мощность правила (6) для частной альтернативы $h^{\varepsilon_+} = \{p \in S_{\varepsilon_+}\}$ при $\varepsilon_+ > \varepsilon$:

$$w(\varepsilon_+) = \Phi \left(\sqrt{\frac{n}{K}} \frac{K(\varepsilon_+ - \varepsilon)}{\sqrt{1 + K\varepsilon_+(e_K^+ - \varepsilon_+)}} - \sqrt{\frac{1 + K\varepsilon(e_K^+ - \varepsilon)}{1 + K\varepsilon_+(e_K^+ - \varepsilon_+)}} \Phi^{-1}(1-\alpha_0) \right), \quad (7)$$

$$e_K = \frac{1}{\varepsilon^3} \sum_{k=0}^{K-1} \left(p_k - \frac{1}{K} \right)^3, \quad |e_K| \leq e_K^+ \leq 1.$$

Построенное правило (6) состоятельно: $w(\varepsilon_+) \xrightarrow{T \rightarrow +\infty} 1$ для любого $\varepsilon_+ > \varepsilon$.

Компьютерные эксперименты с решающим правилом (6)

При $s = 8$ (тестирование байтовой равномерности), $K = 2^s = 256$, $\varepsilon = 0.9/K = 0.00352$ имитировалось по 100 реализаций двоичной последовательности (1) – (3) для $T \in \{2^{19}, 2^{20}, 2^{21}\}$, $\alpha_0 \in \{0.05, 0.1\}$. Графики зависимости оценки вероятности принятия альтернативы $\hat{P}_1$ от $\|p - p_0\|$ представлены для $T \in \{2^{19}, 2^{20}\}$ при $\alpha = 0.05$ на рис. 1, 2 соответственно.

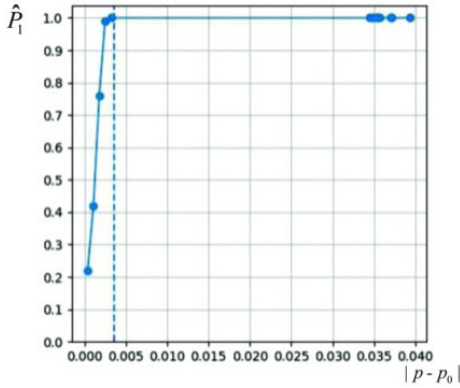


Рисунок 1. $\alpha_0 = 0.05$, $T = 2^{19}$

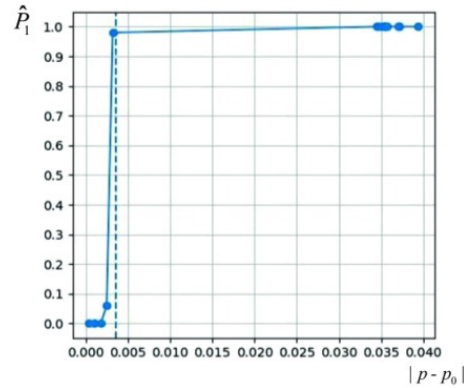


Рисунок 2. $\alpha_0 = 0.05$, $T = 2^{20}$

Рисунки 1, 2 иллюстрируют состоятельность решающего правила (6). Видно также, что при увеличении длительности T последовательности X график зависимости $\hat{P}_1$ от $\|p - p_0\|$ приближается к ступенчатой функции со ступенькой в точке $\|p - p_0\| = \varepsilon$: при $\|p - p_0\| < \varepsilon$ величина $\hat{P}_1 < \alpha_0$, а при $\|p - p_0\| > \varepsilon$, величина $\hat{P}_1$ приближается к единичному значению.

Оценивание уровней безопасности (стойкости) генераторов

Построенный в [13] критерий перспективно использовать для статистического анализа выходных последовательностей **криптографического** генератора с целью установления его «уровня безопасности (стойкости)». Введем, для примера, два уровня безопасности $u \in \{1, 2\}$:

$$u = \begin{cases} 1, & \text{если } \varepsilon_1 \leq \|p - p_0\| < \varepsilon_2, \\ 2, & \text{если } 0 \leq \|p - p_0\| < \varepsilon_1, \end{cases}$$

где $0 < \varepsilon_1 < \varepsilon_2 < 1$ – два заданных критических значения отклонения распределения вероятностей s -грамм $p = (p_k)$ от равномерного распределения $p_0 = (p_{0k})$. Определим сложные гипотезы $H_0^{\varepsilon_1} \subset H_0^{\varepsilon_2}$. С помощью теста из [13] по X_1^T последовательно проверяем пары гипотез $(H_0^{\varepsilon_1}, \overline{H_0^{\varepsilon_1}})$, $(H_0^{\varepsilon_2}, \overline{H_0^{\varepsilon_2}})$. Из (8) получаем оценки для уровня безопасности:

$\hat{u} = 1$, если $H_0^{\varepsilon_2}$ верна, но $H_0^{\varepsilon_1}$ не верна; $\hat{u} = 2$, если $H_0^{\varepsilon_1}$ верна; если $H_0^{\varepsilon_2}$ не верна, то генератор бракуется.

Для статистического тестирования s -мерной равномерности в НИИ ППМИ разработан программный комплекс ЭАДП (Энтропийный Анализ Дискретных Последовательностей) [9], проиллюстрированный на рис. 3.

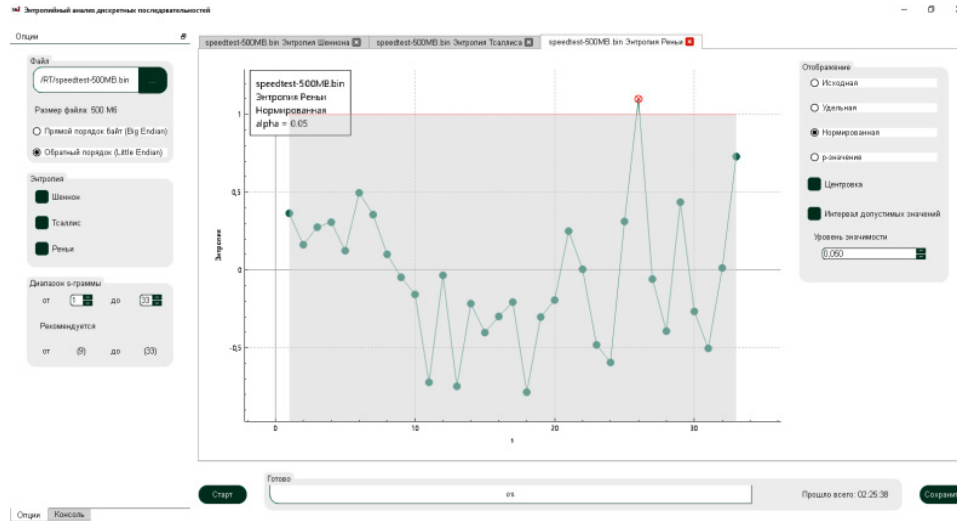


Рисунок 3 – Скриншот Программного Комплекса ЭАДП

Применение сложных марковских моделей для статистического анализа выходных последовательностей криптографических генераторов

Теория вероятностно-статистического анализа дискретных временных рядов $x_t \in A$ глубоко развита лишь для «непрерывных» временных рядов, когда $A \subset R^m$ – подмножество ненулевой меры Лебега, а для дискретных временных рядов, когда A – дискретное множество (в нашем случае $A = V = \{0, 1\}$ – двоичный алфавит) лишь начинает развиваться. При этом главная теоретическая проблема состоит в моделировании стохастической зависимости большой глубины s в $\{x_t\}$.

Универсальной моделью для описания стохастических зависимостей высокого порядка является предложенная Дж. Дубом однородная цепь Маркова ($MC(s)$) достаточно большого порядка $s \in \mathbb{N}$ на вероятностном пространстве $(\Omega, F, \mathbf{P})$, определяемая порождающим уравнением для условного распределения вероятностей: $\mathbf{P}\{x_t = j_t | F_{t-1}\} = \mathbf{P}\{x_t = j_t | X_{t-s}^{t-1} = J_{t-s}^{t-1}\} = p_{J_{t-s}^{t-1}, j_t}$, $t \in \mathbb{Z}$, где $F_{t-1} = \sigma(\{x_\tau : \tau \leq t-1\})$ – σ -алгебра, $X_{t-s}^{t-1} = (x_{t-s}, \dots, x_{t-1}) \in A^s$, $J_{t-s}^{t-1} = (j_{t-s}, \dots, j_{t-1}) \in A^s$. Эта модель определяется матричным параметром – $(s+1)$ -мерной матрицей переходных вероятностей $P = (p_{J_1^{s+1}})$, $J_1^{s+1} \in A^{s+1}$. Для $MC(s)$ число независимых параметров $D_{MC(s)} = N^s(N-1) = O(N^{s+1})$ увеличивается экспоненциально с ростом порядка s . Для преодоления «проклятия размерности» мы предлагаем использовать малопараметрические модели, для которых матрица P имеет параметрический вид: $P = (p_{J_1^{s+1}}) = P_\alpha = :: (p_\alpha(J_1^{s+1}))$, $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$, где α – векторный параметр некоторой малой размерности $d \ll D_{MC(s)}$.

Мы предлагаем четыре основных подхода к построению малопараметрических цепей Маркова высокого порядка [11 – 15]: 1) сжатие множества возможных значений элементов

матрицы P ; 2) использование параметрических семейств стандартных дискретных распределений вероятностей; 3) использование искусственных нейронных сетей для аппроксимации зависимости от предыстории; 4) подход на основе достаточных статистик и информационной геометрии.

На основе этих подходов удалось построить следующие новые малопараметрические вероятностные модели, применимые для двоичных временных рядов $x_t \in V$: цепь Маркова порядка s с r частичными связями $MC(s, r)$ [5]; цепь Маркова условного порядка $MCCO(s, L)$; двоичная условно нелинейная авторегрессионная модель $BCNAR(s)$ [11]; биномиальная условно нелинейная авторегрессионная модель $BiCNAR(s)$ [12]; семибиномиальная условно нелинейная авторегрессионная модель $SBiCNAR(s)$; нейросетевая модель [14]; семейство малопараметрических моделей на основе экспоненциальных семейств вероятностных распределений и достаточных статистик [15].

Разработан метод статистического оценивания параметров α построенных малопараметрических моделей на основе условных многомерных частот, позволяющий строить состоятельные статистические оценки в явном виде. Доказана состоятельность и асимптотическая нормальность построенных статистических оценок $\hat{\alpha}$ при возрастании длины T наблюдаемого дискретного временного ряда. Количественная оценка стойкости криптографического генератора определяется по величине уклонения матрицы $P_{\hat{\alpha}}$ от матрицы P^0 для РПСР. Приводятся результаты компьютерных экспериментов.

Заключение

В статье получены следующие основные результаты:

- 1) сформированы основные проблемы оценивания стойкости криптографических генераторов и подходы к их разрешению;
- 2) обоснована актуальность использования сложной нулевой гипотезы и построены статистические тесты с ее использованием;
- 3) предложены подходы к оцениванию уровней стойкости (безопасности) криптографических генераторов на основе тестирования сложных гипотез и на основе сложных марковских моделей.

Литература

1. Основы криптографии / А. П. Алферов, А. Ю. Зубов, А. С. Кузьмин [и др.]. – М. : Гелиос АРВ, 2005, 480 с.
2. Криптология / Ю. С. Харин, С. В. Агиевич, Д. В. Васильев [и др.]. – Минск : БГУ, 2023, 512 с.
3. NIST SP 800-22: Download Documentation and Software. – URL: <https://csrc.nist.gov/projects/random-bit-generation/document at ion-and-software>.
4. Zubkov, A. M., Serov, A. A. Testing the NIST Statistical Test Suite on artificial pseudorandom sequences / A. M. Zubkov, A. A. Serov // Математические вопросы криптографии. – 2019. – Vol. 10, вып. 2. – С. 89–96.
5. Харин, Ю. С., Петлицкий, А. И. Цепь Маркова s -го порядка с r частичными связями и статистические выводы о ее параметрах / Ю. С. Харин, А. И. Петлицкий // Дискретная математика. – 2007. – Т. 12, вып. 2. – С. 109–130.

6. Савелов, М. П. Предельное совместное распределение статистик критериев пакета NIST и обобщения критерия “Approximate Entropy Test” / М. П. Савелов // Дискретная математика. – 2023. – Т. 35, вып. 2. – С. 93–108.
7. Kowalska, K. A. On the revision of NIST 800-22 Test Suite. URL: <https://eprint.iacr.org/2022/540.pdf>.
8. Decision to Revise NIST SP 800-22 Rev. 1a. URL: <https://csrc.nist.gov/news/2022/decision-to-revise-nist-sp-800-22-rev-1a>.
9. Палуха, В. Ю., Харин, Ю. С., Мальцев, М. В. [и др.]. Программный комплекс для энтропийного анализа дискретных последовательностей / В. Ю. Палуха, Ю. С. Харин, М. В. Мальцев // Информационные системы и технологии. – 2022. – Т. 1. – С. 102–107.
10. Волошко, В. А., Трубей, А. И. О мощности тестов многомерной дискретной равномерности, используемых для статистического анализа генераторов случайных последовательностей / В. А. Волошко, А. И. Трубей // Журнал Белорусского государственного университета. Математика. Информатика. – 2022. – № 2. – С. 26–37.
11. Kharin, Yu. S., Voloshko, V. A., Medved, E. A. Statistical estimation of parameters for binary conditionally nonlinear autoregressive time series / Yu. S. Kharin, V. A. Voloshko, E. A. Medved // Mathematical Methods of Statistics. – 2018. – Vol. 27(2). – P. 103–118.
12. Kharin, Yu., Voloshko, V. Robust estimation for Binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies // Journal of Multivariate Analysis. – 2022. – Vol. 185(2). 104777.
13. Харин, Ю. С., Зубков, А. М. О статистической проверке сложных гипотез об s-мерном равномерном распределении вероятностей двоичных последовательностей // Дискретная математика. – 2024. – Т. 36, вып. 1. – С. 116–135.
14. Харин, Ю. С., Волошко, В. А. Об аппроксимации двоичных цепей Маркова высокого порядка малопараметрическими моделями Ю. С. Харин, В. А. Волошко // Дискретная математика. – 2022. – Т. 34, вып. 3. – С. 114–135.
15. Kharin, Yu., Voloshko, V. Statistical analysis of parsimonious high-order multivariate finite Markov chains based on sufficient statistics // Journal of Multivariate Analysis. – 2025. – Vol. 208. 105422.