УДК 004.056

СТАТИСТИЧЕСКОЕ ТЕСТИРОВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ С ПРИМЕНЕНИЕМ ЗАКОНА ПОВТОРНОГО ЛОГАРИФМА

А. И. ТРУБЕЙ, М.В. МАЛЬЦЕВ, В.Ю. ПАЛУХА, И.К. ПИРШТУК
Учреждение БГУ «НИИ прикладных проблем математики и информатики», г. Минск, 220030, Республика Беларусь

Введение

Генераторы (псевдо)случайных последовательностей являются частью многих IT продуктов и используются во многих криптографических приложениях (парольная защита, векторы инициализации, ключи блочных симметричных криптоалгоритмов, алгоритмы поточного шифрования). В этой связи к генераторам (псевдо)случайных последовательностей, применяемым для решения задач защиты информации, предъявляются особые требования. Разработано достаточно много методов статистического тестирования для определения качества генерируемых последовательностей. При этом слабым местом практически всех наборов (батарей) статистических тестов является то, что не вычисляются ошибки 2 рода. В настоящее время существует немало генераторов псевдослучайных последовательностей, реализованных в соответствии с различными принципами (в основном – на базе блочных криптоалгоритмов и криптостойких хэш-функций). Для оценки качества, применяемых в целях защиты данных, часто используется набор тестов, определяемый стандартом FIPS140-1 [1].

NIST SP800-22 (2010) [2] также предлагает современный набор тестов для генераторов (псевдо)случайных, последовательностей, чтобы обнаружить отклонения двоичной последовательности от случайной равновероятной. Батареи тестов NIST SP800-22 и FIPS140-1 имеют ограничения, связанные с ошибками 2 рода. Например, генератор G, который, в основном, генерирует (псевдо)случайные последовательности, с вероятностью β будет также генерировать смещенные на некоторую величину δ от равновероятного распределения последовательности (например, последовательности, состоящие в основном из нулей или единиц). При этом генератор G будет оцениваться как «хороший» генератор (псевдо)случайных последовательностей тестами NIST SP800-22 и FIPS140-1 (например, тестом «Монобит»), хотя выходные последовательности несложно отличить от равномерного распределения.

В докладе приводятся определенные преимущества тестирования случайных последовательностей на основе статистики повторного логарифма, имеющей аппроксимированное нормальное распределение [3, 4]. Описывается методика принятия решений на основе критерия хи-квадрат согласия.

1. Основные понятия закона повторного логарифма

Пусть независимые случайные величины $X = (x_1, \ldots, x_n)$ одинаково распределены и принимают два значения 0 и 1 с вероятностью 1/2. Тогда $S_n = \sum_n x_i$ –число успехов в схеме Бернулли с вероятностью успеха 1/2. Э. Борель доказал, что при $n \to \infty$ с вероятностью 1 $\frac{S_n}{n} \to \frac{1}{2}$.

Впоследствии (1914) Г. Харди и Дж. Литлвуд (G. Hardy, J. Littlewood) показали, что почти наверное $\limsup_{n\to\infty}\frac{\left|S_n-\frac{n}{2}\right|}{\sqrt{n\ln n}}<\frac{1}{\sqrt{2}}$. Затем А Я. Хинчин (1924) доказал более сильный результат, называемый законом повторного логарифма [5]:

$$P\left\{\limsup_{n\to\infty}\frac{\left|S_n-\frac{n}{2}\right|}{\sqrt{n\ln\ln n}}=\frac{1}{\sqrt{2}}\right\}=1 \qquad (1)$$

2. Статистические процедуры множественной проверки гипотез. Оценка вероятности ошибки 2 рода

Каждый статистический критерий предназначен для проверки конкретной гипотезы H_0 и позволяет выявить только определенные типы отклонений от H_0. Поэтому для проверки набора гипотез и увеличения точности обнаружения альтернативных гипотез при тестировании последовательностей следует использовать набор статистических критериев, а для принятия итогового решения – процедуры множественной проверки гипотез. Двухэтапные процедуры позволяют увеличить мощность критерия и удерживать вероятность ошибки 2 рода на должном уровне.

Вероятность ошибки второго рода вычисляют для каждого вида проверки гипотез по-разному. Получим выражение для вычисления ошибки второго рода для проверки двусторонней гипотезы о равенстве математического ожидания распределения μ некоторой величине (дисперсию σ^2 полагаем известной).

Для проверки гипотезы используется статистика $S=\frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}$, которая имеет стандартное нормальное распределение.

Чтобы найти ошибку второго рода β при заданной ошибке 1 рода α, необходимо предположить, что гипотеза H_0: $\mu = \mu_0$ не верна и соответственно истинное среднее значение распределения $\mu = \mu_0 + \delta$.

В этом случае статистика S будет иметь нормальное распределение $N\left(\frac{\delta\sqrt{n}}{\sigma};1\right)$, то есть ее график будет смещен вправо на $\frac{\delta\sqrt{n}}{\sigma}$. $S\in\left[-S_{\alpha/2};S_{\alpha/2}\right]$

Эти значения соответствуют границам доверительного интервала. $S_{\alpha/2}$ – это верхний $\alpha/2$ квантиль стандартного нормального распределения.

На основании вышеизложенного легко показать, что ошибка второго рода в терминах стандартного нормального распределения имеет следующий вид:

$$\beta=\Phi\left(S_{\alpha/2}-\frac{\alpha\sqrt{n}}{\sigma}\right)-\Phi\left(-S_{\alpha/2}-\frac{\alpha\sqrt{n}}{\sigma}\right). \qquad (2)$$

Это справедливо и для $\delta < 0$. В случае распределения Бернулли:

$$\mu=p=\mu_0+\delta=\frac{1}{2}+\delta;\ \ \sigma=\sqrt{p(1-p)}=\sqrt{\left(\frac{1}{2}+\delta\right)\left(1-\frac{1}{2}-\delta\right)}.$$

Для теста «Монобит» в случае гипотезы H_0: $p = 1 / 2$ и альтернативы H_1: $p \neq 1 / 2$ вероятность ошибки второго рода имеет вид:

$$\beta = \Phi\left(\frac{S_+ - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{S_- - np}{\sqrt{np(1-p)}}\right), \quad (3)$$

где $S_\pm = \frac{n}{2} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{n}{4}}$ – граничные значения теста.

Из формулы (3) видно, что вероятность ошибки 2 рода β зависит от размера выборки ***n*** и уровня значимости α. Чем больше размер выборки, тем меньше ошибка второго рода (при прочих равных).

3.Двухэтапная процедура поверки гипотез с применением статистики закона повторного логарифма

Очевидно, что бинарную последовательность можно представить в виде реализации схемы Бернулли. Для последовательности $x \in \Sigma^n$ определим:

$$S(n) = \sum_{i=0}^{n-1} x_i; \quad S(n)^* = \frac{2S(n) - n}{\sqrt{n}}, \quad (4)$$

где Σ^n – множество двоичных последовательностей длины *n*.

Закон больших чисел гласит, что для случайной последовательности x $\frac{S(n)}{n} \to \frac{1}{2}$, что соответствует частотному тесту («Монобит») в NIST SP800-22. Закон повторного логарифма дает оптимальную верхнюю оценку $\sqrt{2\ln\ln n}$ для колебаний $S(n)^*$. Основываясь на этом факте, мы будем использовать следующую статистику:

$$S_{зпл}(n) = \frac{S(n)^*}{\sqrt{2\ln\ln n}} = \frac{2S(n) - n}{\sqrt{2n\ln n\ln n}}. \quad (5)$$

В соответствии с предельной теоремой Муавра-Лапласа для заданных z_1 и z_2:

$$\lim_{m\to\infty} P\left[z_1 \le S(n)^* \le z_2\right] = \Phi(z_2) - \Phi(z_1).$$

Распределение, порожденное $S_{зпл}$, определяет вероятностную меру на вещественной прямой *R*. Пусть $\mathcal{R} \in \Sigma^n$ – набор из *m* последовательностей со стандартным определением вероятности на нем. То есть, для каждой последовательности $x_0 \in \mathcal{R}$ положим $P[x = x_0] = 1/m$. Тогда каждый набор $\mathcal{R} \in \Sigma^n$ порождает вероятностную меру $\mu_n^{\mathcal{R}}$ на *R* $\mu_n^{\mathcal{R}}(I) = P\left[S_{зпл}(n) \in I, x \in \mathcal{R}\right]$ для каждого измеримого по Лебегу множества *I* на *R*. Для $U = \Sigma^n$ введем μ_n^U – соответствующую вероятностную меру, порожденную равномерным распределением. По определению, если $\mathcal{R}_n$ – это совокупность всех последовательностей длины *n*, сгенерированных генератором (псевдо)случайных чисел, то разница между μ_n^U и $\mu_n^{\mathcal{R}}$ незначительна.

Как было сказано выше, для случайной последовательности распределение статистики $S(n)^*$ может быть аппроксимировано нормальным распределением с математическим ожиданием 0 и дисперсией 1 с ошибкой, не превышающей $1/n$. Другими словами, меру μ_n^U можно рассчитать следующим образом:

$$\mu_n^U\left((-\infty, z]\right) = \Phi\left(z\sqrt{2\ln\ln n}\right) = \sqrt{2\ln\ln n}\int_{-\infty}^{z}\phi\left(y\sqrt{2\ln\ln n}\right)dy. \quad (6)$$

При проверке гипотезы H_0 одним из общепринятых подходов к увеличению точности обнаружения альтернативных гипотез является двухэтапный анализ. На первом этапе двухэтапной процедуры по каждой выборке вычисляется статистика критерия. На втором этапе,

как правило, проверяется гипотеза согласия последовательности статистик, вычисленных на первом этапе, с теоретическим распределением. Мы будем проверять гипотезу согласия последовательности значений статистики $S_{зпл}(n)$ с нормальным распределением на вещественной прямой $R\ (-\infty,\infty)$.

Процедура принятия решения основана на критерии хи-квадрат согласия. Для построения критерия будем использовать в качестве дискретного разбиения вещественной прямой R, множество Ѣ, определяемое следующим образом:

$$\text{В} = \cup I = \{(-\infty,-1),[1,\infty)\} \bigcup \{[0.05r-1, 0.05r-0.95) : 0 \le r \le 39\}. \quad (7)$$

Чтобы оценить генератор G, необходимо:

1. Осуществить генерацию набора $\mathcal{R} \in \Sigma^n$ из m = 10000 последовательностей возможно большей длины.
2. На первом этапе двухэтапной процедуры вычислить значения статистики $S_{зпл}(n)$ по всем m последовательностям.
3. На втором этапе двухэтапной процедуры сравнить вероятностные меры $\mu_n^{\mathcal{R}}$ и μ_n^{U} Для сравнения будем использовать следующую статистику χ^2 согласия [7]:

$$\chi^2 = \sum_{j=1}^{|\text{B}|} \frac{\left[\nu_n^{\mathcal{R}}(I_j) - m p_n^{\text{U}}(I_j)\right]^2}{m p_n^{\text{U}}(I_j)}, \quad (8)$$

где $\nu_n^{\mathcal{R}}(I_j)$ – частоты попадания значений статистики $S_{зпл}(n)$ в интервал I_j по всем m последовательностям; $p_n^{\text{U}}(I_j)$ – теоретические вероятности попадания $S_{зпл}(n)$ в интервал I_j. То есть, необходимо проверить гипотезу о том, что выборка $\left\{\nu_n^{\mathcal{R}}(I_1),\ldots,\nu_n^{\mathcal{R}}(I_j),\ldots,\nu_n^{\mathcal{R}}(I_{|\text{B}|})\right\}$ извлечена из некоторой нормальной совокупности с математическим ожиданием 0 и среднеквадратичным отклонением $\sigma = \frac{1}{\sqrt{2\ln\ln n}}$.

Для проверки гипотезы проведено тестирование 10 000 последовательностей объемом 0,5 MB каждая, выработанных соответственно линейным конгруэнтным генератором и физическим генератором «Ключ-BC» (суммарный объем – 5 GB для каждого генератора). При сравнении выборок с нормальным распределением по критерию χ^2 получим следующие P-значения (после объединения групп):

- для линейного конгруэнтного генератора – P = 0.00028;
- для физического генератора «Ключ-BC» – P = 0.69373.

Это означает, что для последовательностей, вырабатываемых физическим генератором, выполняется гипотеза H_0 согласия с моделью независимых симметричных испытаний Бернулли, в то время как последовательности, вырабатываемые линейным конгруэнтным генератором, гипотезе H_0 не соответствуют. При тестировании последовательностей меньшей длины отклонений от гипотезы H_0 как для линейного конгруэнтного генератора, так и для физического генератора не выявлено. Результаты тестирования последовательностей объемом 5 GB приведены на рисунках 1, 2.

Заключение

В результате проведенных исследований получены следующие результаты:

1. Установлено, что уязвимости в часто используемых реализациях генераторов псевдослучайных последовательностей не выявляются батареей NIST SP800-22, в основном, из-за недостаточной длины тестируемых последовательностей.

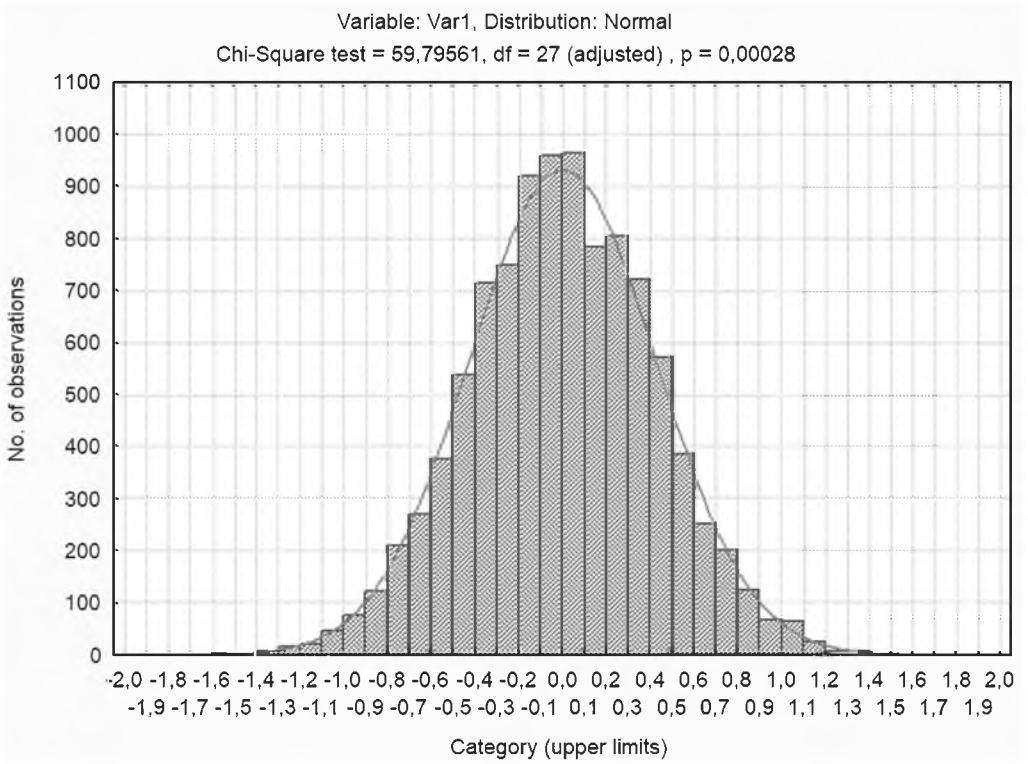

Рисунок 1. Гистограмма частот выборки линейного конгруэнтного генератора, 5 GB

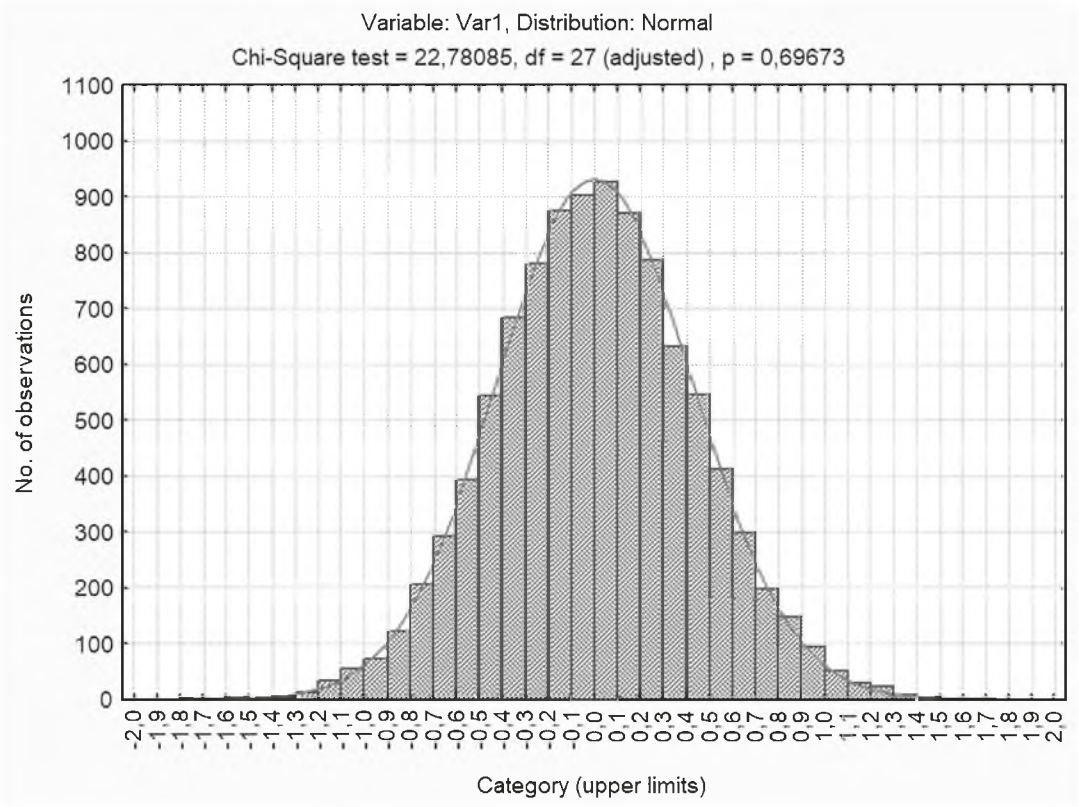

Рисунок 2. Гистограмма частот выборки физического генератора, «Ключ-ВС», 5 GB

2. Проведена оценка зависимости вероятности ошибки 2 рода от вероятности ошибки 1 рода, размера выборки, отклонения от равновероятности.
3. Построены статистики на основе закона повторного логарифма, распределение которых можно аппроксимировать нормальным законом.
4. С применением двухэтапной процедуры проверки гипотез проведено тестирование длинных последовательностей, сгенерированных конгруэнтным генератором, а также физическим генератором на основе шумового диода («Ключ-ВС») по тесту «Монобит». При этом физический генератор успешно прошел тестирование, а конгруэнтный генератор тестирование провалил.

Список литературы

1. Federal Information Processing Standards. FIPS PUB 140-1. Security Requirements for Cryptographic Modules. – 1994.
2. A statistical test suite for random and pseudorandom number generators for cryptographic applications NIST Special Publication 800-22 (with revisions dated May 15, 2001) / Andrew Rukhin at al, 2010.
3. Yongge Wang On the Design of *LIL* Tests for (Pseudo) Random Generators and Some Experimental Results / Dept. SIS, UNC Charlotte Charlotte, NC 28223, USA. – [Electronic resource]. – 2014. – Mode of access: https://arxiv.org/abs/1401.3307. – Date of access: 01/03/2020.
4. Yongge Wang, Tony Nicol On statistical distance based testing of pseudorandom sequences and experiments with PHP and Debian OpenSSL / Computers & Security. – Volume 53. – September 2015. – P 44–64.
5. Khintchin, A Über einen Satz der Wahrscheinlichkeitsrechnung / Fundamenta Mathematicae. – 1924. – 6. – P. 9–20.
6. Feller W. Introduction to probability theory and its applications, vol. I. New York: John Wiley & Sons, Inc.; 1968.
7. Крамер, Г. Математические методы статистики / Г. Крамер – М.: Мир, 1976.

УДК 004.056.53

ФУНКЦИОНАЛЬНЫЕ МОДЕЛИ ПРОЦЕССОВ РЕАЛИЗАЦИИ УГРОЗ БЕЗОПАСНОСТИ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ БАЗОВОЙ СИСТЕМЫ ВВОДА-ВЫВОДА

И.С.ГЕФНЕР[1], В.В.ТЕКУНОВ[2], Ю.К.ЯЗОВ[2]
ФСТЭК России, г. Москва
ФАУ «ГНИИИ ПТЗИ ФСТЭК России» г. Воронеж, Российская Федерация

Для оценки возможностей реализации угроз необходимо определить для каждой угрозы состав, последовательность и время выполнения действий (команд, процедур, функций и т.п.), составляющих содержание процесса реализации этой угрозы. Такие данные необходимы и при анализе возможностей парирования угроз и оценке эффективности мер защиты. Ранее при