УДК 004.056.5:519.254

МЕТОДИКА НАХОЖДЕНИЯ ЭТАЛОННЫХ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ, ПОЛУЧАЕМЫХ ПРИ СТАТИСТИЧЕСКОМ ТЕСТИРОВАНИИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ КЛЮЧЕЙ

Н.Г. КИЕВЕЦ, А.И. КОРЗУН

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 31 марта 2013

Предложена методика нахождения эталонных законов распределения вероятностей *P-value*, получаемых при статистическом тестировании. Проведен анализ результатов тестирования по частотному тесту и тесту на подпоследовательности одинаковых бит системы NIST последовательностей ключей при использовании равномерного закона в качестве эталонного закона распределения вероятностей *P-value* и при использовании найденных законов распределения *P-value*.

Ключевые слова: последовательность ключей, статистическое тестирование.

Введение

Практически значимой является задача тестирования последовательностей ключей, применяемых в криптографических системах защиты информации. Получение ключей с хорошими статистическими свойствами не может гарантировать даже использование качественного генератора случайных чисел (ГСЧ). В связи с этим все ключи перед их использованием требуется подвергать статистическому тестированию.

При относительно небольшой длине ключей их статистическое тестирование может привести к неверным результатам, так как распределение тестовой статистики обычно сравнивается не с действительным, а приблизительным законом распределения. При обобщении результатов тестирования вероятность принять неверное решение увеличивается. Тесты системы NIST [1] предполагают использование для тестирования последовательностей длиной от 100 бит. Относительно небольшие длины практически используемых ключей, превышающие 100 бит, равны 128 и 256 бит. Ключи длиной 128 бит используются в симметричных алгоритмах шифрования IDEA и AES, ключи длиной 256 бит используются в симметричном алгоритме шифрования ГОСТ 28147-89 [2]. Представляет интерес интерпретация результатов тестирования последовательностей ключей длиной 128 и 256 бит.

Анализ результатов тестирования по методике NIST

Исследование по каждому из тестов последовательности ключей с обобщением результатов по методике NIST [1] включает следующие этапы.

1. Тестирование по тесту каждого ключа последовательности.

2. Формирование массива значений вероятности *P-value*, полученных при тестировании ключей последовательности.

3. Подсчет частот попадания значений *P-value* в интервалы L: $l_1 = [0;0,1]$; $l_2 = (0,1;0,2]$; $l_3 = (0,2;0,3]$; $l_4 = (0,3;0,4]$; $l_5 = (0,4;0,5]$; $l_6 = (0,5;0,6]$; $l_7 = (0,6;0,7]$; $l_8 = (0,7;0,8]$; $l_9 = (0,8;0,9]$; $l_{10} = (0,9;1]$.

4. Расчет случайной величины $\chi^2 = \sum_{i=1}^{K+1} \frac{(m_i - s \cdot p_i)^2}{s \cdot p_i}$, где $K+1$ – количество интервалов значений *P-value*, m_i – частота попадания значений *P-value* в интервал l_i, s – количество значений *P-value*, равное количеству ключей в последовательности, p_i – вероятность попадания значения *P-value* в интервал l_i. В случае равномерного распределения *P-value* для всех интервалов l_i принимают значения $p_i = 1/10$.

5. Расчет вероятности, характеризующей степень соответствия эмпирического закона распределения вероятностей *P-value* эталонному закону:

$$P\text{-}value_T = \left(\int_{\chi^2/2}^{\infty} t^{\frac{K}{2}-1} e^{-t} dt \right) / \left(\int_{0}^{\infty} t^{\frac{K}{2}-1} e^{-t} dt \right), \quad (1)$$

где $K = 9$ – число степеней свободы распределения «хи-квадрат».

6. Сравнение значения $P\text{-}value_T$ с уровнем значимости $\alpha = 0{,}0001$. Если $P\text{-}value_T \geq \alpha$, то делается вывод о том, что последовательность ключей успешно прошла тестирование. Если $P\text{-}value_T < \alpha$, то делается вывод о том, что последовательность ключей не прошла тестирование по данному тесту.

По методике NIST проведено обобщение результатов тестирования по частотному тесту и тесту на подпоследовательности одинаковых бит двух последовательностей по 8000 ключей длины 128 бит, полученных из двух ЭПК (электронных пластиковых карт). В табл. 1 представлены результаты тестирования. Так как все полученные значения $P\text{-}value_T = 0 < 0{,}0001$, ясно, что ни одна последовательность ключей не прошла тестирование ни по одному тесту. Для наглядности представления полученных данных в табл. 1 содержатся значения относительной частоты попадания значений *P-value* в интервал P^*. Значения величины P^* рассчитываются по формуле: $p_i^* = m_i / s = m_i / 8000$.

Таблица 1**. Результаты тестирования последовательностей ключей длины 128 бит, полученных из двух ЭПК**

i	l_i	Частотный тест				Тест на подпоследовательности одинаковых бит			
		ЭПК №1		ЭПК №2		ЭПК №1		ЭПК №2	
		m_i	p_i^*	m_i	p_i^*	m_i	p_i^*	m_i	p_i^*
1	[0-0,1]	700	0,0875	738	0,0922	823	0,1029	834	0,1043
2	(0,1-0,2]	730	0,0913	750	0,0938	785	0,0981	838	0,1048
3	(0,2-0,3]	1151	0,1439	1140	0,1425	903	0,1129	930	0,1163
4	(0,3-0,4]	760	0,0950	766	0,0957	789	0,0986	786	0,0983
5	(0,4-0,5]	885	0,1106	909	0,1136	781	0,0976	771	0,0964
6	(0,5-0,6]	961	0,1201	1016	0,1270	837	0,1046	862	0,1078
7	(0,6-0,7]	0	0	0	0	565	0,0706	576	0,0720
8	(0,7-0,8]	1051	0,1314	1024	0,1280	881	0,1101	831	0,1039
9	(0,8-0,9]	1159	0,1449	1085	0,1356	945	0,1181	878	0,1098
10	(0,9-1]	603	0,0754	572	0,0715	691	0,0864	694	0,0867
χ^2		1304,4225		1256,2775		134,8825		116,0475	
P-value		0		0		0		0	

По данным табл. 1 построены гистограммы относительных частот P^* для каждой из четырех ключевых последовательностей. Гистограммы представлены на рис. 1, на котором уровни $P^* = 0{,}1$ показывают ожидаемые значения величины P^* для каждого интервала.

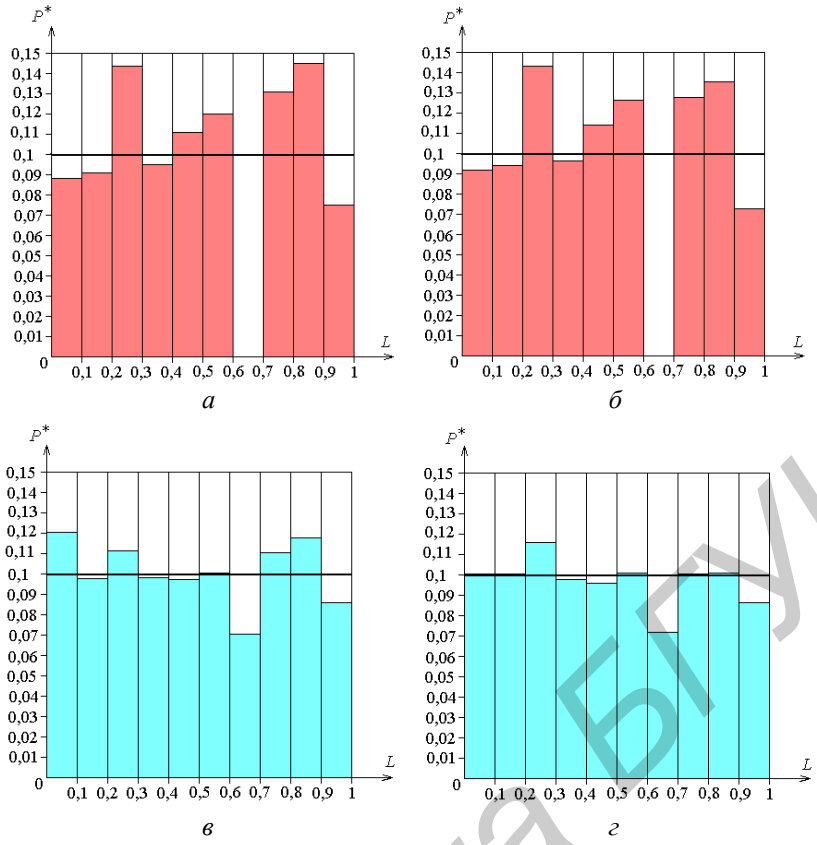


Рис. 1. Гистограммы относительных частот $P*$ для: *а* – последовательности ключей, полученной из ЭПК №1 и протестированной по частотному тесту; *б* – последовательности ключей, полученной из ЭПК № 2 и протестированной по частотному тесту; *в* – последовательности ключей, полученной из ЭПК № 1 и протестированной по тесту на подпоследовательности одинаковых бит; *г* – последовательности ключей, полученной из ЭПК №2 и протестированной по тесту на подпоследовательности одинаковых бит

Значения $P\text{-}value_T = 0$ свидетельствуют о том, что либо полученные ключи не пригодны для использования, либо методология имеет изъян. Поскольку ключи были получены от физических ГСЧ, нулевые результаты заставили усомниться в равномерности распределения значений *P-value*.

Было решено подвергнуть сомнению равномерность распределения значений $P-value$, получаемых при тестировании последовательностей ключей длиной 128 бит по частотному тесту и тесту на подпоследовательности одинаковых бит по следующим причинам.

1. Получены отрицательные результаты при тестировании всех последовательностей ключей, вырабатываемых физическими ГСЧ.

2. В тестах используется аппроксимация фактических законов распределения количества единиц или количества непрерывных подпоследовательностей бит в ключе нормальным законом.

3. Гистограммы на рис. 1 показывают сходство распределения *P-value* для разных ключевых последовательностей при тестировании по одному и тому же тесту.

Таким образом, показано, что следует провести корректировку эталонного закона распределения значений *P-value*, получаемых при тестировании последовательностей ключей длины 128 бит при тестировании по частотному тесту и по тесту на подпоследовательности одинаковых бит.

Методика нахождения эталонного закона распределения значений *P-value*

Для нахождения закона распределения значений *P-value* при тестировании по некоторому тесту ключей длины n бит необходимо протестировать полный набор 2^n ключей и определить долю значений *P-value*, принимающих значения из каждого из десяти интервалов

l_i ($i = \overline{1,10}$). Такая задача является трудновыполнимой из-за больших временных затрат, необходимых для тестирования полного набора ключей, который при длине ключа 128 бит равен $2^{128} \approx 3{,}4028 \cdot 10^{38}$, при длине 256 бит – $2^{256} \approx 1{,}1579 \cdot 10^{77}$ ключей.

Обычно возможное количество значений вероятности *P-value*, получаемых при тестировании ключей по данному тесту при заданной длине n, бывает существенно меньше полного набора ключей. Для снижения временных затрат при нахождении эталонных законов распределения вероятностей *P-value* предложена следующая методика.

1. Формируется массив всех возможных значений *P-value* для данного теста при заданной длине ключа, которые делятся на 10 групп в зависимости от принадлежности к интервалу l_i ($i = \overline{1,10}$).

2. Определяется вероятность появления каждого значения вероятности *P-value*.

3. Для каждой группы значений *P-value* рассчитывается вероятность попадания в данную группу вероятности *P-value*.

Рассмотрим приложение предложенной методики к частотному тесту и тесту на подпоследовательности одинаковых бит.

В частотном тесте значение вероятности *P-value* определяется по формуле [1]: $P\text{-}value = erfc\left(|S_n| / \sqrt{2n}\right)$, где $erf(x)$ – дополнительная функция ошибок; S_n – сумма элементов последовательности, полученной из исходной последовательности нулей и единиц путем замены элемента «0» на элемент «–1»; n – длина битовой последовательности. Количество возможных значений величины $|S_n|$ равно $n/2+1$. Следовательно, имеем $n/2+1$ различных значений вероятности *P-value*. Вероятность получения значения вероятности *P-value* при заданном значении S_n определяется по формулам [3]: $p(S_n) = \begin{pmatrix} n \\ \left|\frac{n-S_n}{2}\right| \end{pmatrix} / 2^{n-1}$, если $S_n \neq 0$, и $p(S_n) = \frac{n!}{(n/2)! \cdot (n/2)! \cdot 2^n}$, если $S_n = 0$.

В тесте на подпоследовательности одинаковых бит значение вероятности *P-value* определяется по формуле [1]: $P\text{-}value = erfc\left(\frac{|r - 2n\lambda(1-\lambda)|}{2\sqrt{2n}\lambda(1-\lambda)}\right)$, где r – количество непрерывных подпоследовательностей нулей и единиц, $\lambda = n1/n$ – доля единиц в тестируемой последовательности, содержащей $n1$ единиц. При заданной длине n значение $P-value$ полностью определяется величинами r и $n1$. Величина $r = 1$ при $|n1 - n/2| = n/2$ и величина r может принимать значения от 2 до $2 \cdot |n1 - n/2|$ при $|n1 - n/2| \neq n/2$. С учетом того, что вероятность *P-value* принимает одинаковые значения при $n1 = x$ и $n1 = |n - x|$ при одинаковых значениях r, то всего имеем $2\sum_{j=1}^{n/2-1}(2 \cdot j - 1) + n + 1$ разных возможных значений вероятности *P-value* для теста на подпоследовательности одинаковых бит. Вероятность получения данного значения *P-value*, что эквивалентно получению данных значений $n1$ и r, определяется по формулам [3, 4]:

1. $p(n1, r) = 2\begin{pmatrix} n1-1 \\ r/2-1 \end{pmatrix}\begin{pmatrix} n-n1-1 \\ r/2-1 \end{pmatrix} / \left(2^n\right)$, если r – четное число;

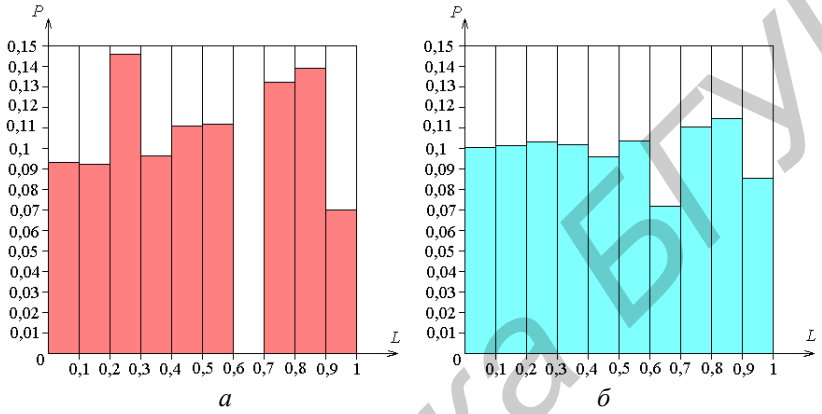
2. $p(n1, r) = \left[\begin{pmatrix} n1-1 \\ (r-1)/2 \end{pmatrix}\begin{pmatrix} n-n1-1 \\ (r-3)/2 \end{pmatrix} + \begin{pmatrix} n1-1 \\ (r-3)/2 \end{pmatrix}\begin{pmatrix} n-n1-1 \\ (r-1)/2 \end{pmatrix}\right] / \left(2^n\right)$, если r – нечетное число.

В соответствии с предложенной методикой определены вероятности P – вероятности попадания в каждый из интервалов l_i ($i = \overline{1,10}$) значений *P-value*, получаемых при тестировании

последовательностей ключей длиной 128 бит по частотному тесту и тесту на подпоследовательности одинаковых бит. Полученные значения вероятности P представлены в табл. 2. По данным табл. 2 на рис. 2 построены гистограммы вероятностей P, соответствующих действительным законам распределения *P-value* для данных тестов.

Таблица 2. **Значения вероятностей** P

i	1	2	3	4	5	6	7	8	9	10
l_i	[0-0,1]	(0,1-0,2]	(0,2-0,3]	(0,3-0,4]	(0,4-0,5]	(0,5-0,6]	(0,6-0,7]	(0,7-0,8]	(0,8-0,9]	(0,9-1]
Частотный тест										
p_i	0,0927	0,0920	0,1463	0,0955	0,1098	0,1224	0	0,1323	0,1386	0,0704
Тест на подпоследовательности одинаковых бит										
p_i	0,1006	0,1014	0,1134	0,1022	0,0959	0,1041	0,0719	0,1103	0,1143	0,0860



а *б*

Рис. 2. Гистограммы вероятностей P: *а* – для частотного теста; *б* – для теста на подпоследовательности одинаковых бит

Из табл. 2 видно, что в случае тестирования ключей длиной 128 бит по частотному тесту невозможно получение значения *P-value* из интервала $(0{,}6;0{,}7]$. В связи с этим для частотного теста при данной длине ключа следует рассматривать не десять интервалов значений *P-value*, а девять, исключив из диапазона $[0;1]$ интервал $(0{,}6;0{,}7]$. В таком случае при расчете по формуле (1) будем иметь $K = 8$ степеней свободы распределения «хи-квадрат».

Таким образом, получены эталонные законы распределения вероятностей *P-value* для частотного теста и теста на подпоследовательности одинаковых бит при n=128 бит. Выявлено, что данные законы отличны от равномерного. Аналогичным образом получены эталонные законы распределения *P-value* для последовательностей ключей длиной 256 бит.

Интерпретация результатов тестирования с использованием полученных эталонных законов распределения вероятностей *P-value*

Для проверки качества ранее полученных последовательностей результаты тестирования по частотному тесту и тесту на подпоследовательности одинаковых бит были интерпретированы с учетом полученных эталонных законов распределения *P-value*. Результаты расчетов представлены в табл. 3.

Таблица 3. **Результаты тестирования последовательностей ключей длиной 128 бит**

Название теста	Частотный тест		Тест на подпоследовательности одинаковых бит	
№ ЭПК	1	2	1	2
χ^2	8,2468	5,2488	3,7839	9,3189
$P\text{-value}_T$	0,5095	0,8121	0,9251	0,4084

Из табл. 3 видно, что обе последовательности ключей по обоим тестам прошли тестирование, так как все значения $P\text{-}value_T > 0,0001$. Это свидетельствуют о хороших статистических свойствах исследуемых последовательностей ключей, которые выявляют частотный тест и тест на подпоследовательности одинаковых бит.

Заключение

Таким образом, предложена методика получения эталонных законов распределения значений *P-value*. Получены эталонные законы распределения значений *P-value* для частотного теста и теста на подпоследовательности одинаковых бит при длине ключей $n = 128$ бит и $n = 256$ бит. Показано, что использование равномерного закона распределения в качестве эталонного закона распределения вероятностей *P-value* при исследовании последовательностей ключей относительно небольших длин приводит к неверным выводам:

1) последовательности ключей с хорошими статистическими свойствами могут быть забракованы;

2) последовательности ключей с плохими статистическими свойствами могут успешно пройти тестирование.

Использование методики интерпретации результатов тестирования с истинными эталонными законами распределения позволяет получить достоверные результаты. Поэтому методику нахождения эталонных законов распределения вероятностей *P-value* целесообразно распространить на другие тесты.

THE TECHNIQUE OF PROBABILITY REFERENCE DISTRIBUTION LAW FINDING AT KEY SEQUENCE STATISTICAL TESTING

N.G. KIYEVETS, A.I. KORZUN

Abstract

The technique of probability reference distribution law finding at statistical testing is offered. The analysis of results of key sequence testing under the frequency test and runs test of system NIST is carried out at use of the uniform law as the probability reference distribution law and at use of the found distribution laws.

Список литературы

1. A statistical test suite for random and pseudorandom number generators for cryptographic applications. [Электронный ресурс]. – Режим доступа: http://csrc.nist.gov/publications/nistpubs/800-22-rev1a/SP800-22rev1a.pdf. – Дата доступа: 13.11.2012.
2. *Дишунян В.Л., Шаньгин В.Ф.* Электронная идентификация. Бесконтактные электронные идентификаторы и смарт-карты. М., 2004.
3. *Гмурман В. Е.* Теория вероятностей и математическая статистика. М., 1977.
4. Nonparametric Statistical Interference. [Электронный ресурс]. – Режим доступа: http://f3.tiera.ru/2/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Gibbons%20J.%20Nonparametric%20statistical%20inference%20(Dekker,%202003)(ISBN%200824740521)(O)(672s)_MVsa_.pdf. – Дата доступа: 25.02.2014.