METHOD OF STATISTICAL TESTS INDEPENDENCE CHECKING

L.V. Skrypnik, L.V. Kovalchuk, V.T. Bezditnyi
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, UKRAINE
e-mail: lv_kov_crypto@mail.ru

Abstract

The problem of building of sets of independent statistical tests for estimating of cryptographic qualities of random sequences and random number generators is considered and previous results are generalized. The theorem is proved which allows us to construct the methodic of statistical tests independence checking. Practical results are obtained using the methodic.

1 Introduction

The necessary condition for cryptosystem security is usage of pseudorandom/random generator with definite cryptographic properties for generating cryptosystem's parameters and key data. The main method for checking up these properties is testing its output sequences with statistical tests from the some set [1], [2]. And rather important question is: to use only independent tests, so as the results of each test don't duplicate the results of others. The reasons for this are: to minimize the testing time and to be able to calculate the general error of first kind for hypothesis of sequence randomness.

For the first time the definition of statistical tests independence was introduced in [3], than a lot of works showed it's practical usefulness and convenience. For example, it was shown that the property of independence of statistical tests doesn't depend on different parameters, such as alphabet and length of sequences, level of significance, else. The procedure of checking up tests independence according to the mentioned definition [3] is fast and simple, but needs the "almost perfect generator" (in practice was used the adopted generator from [4], Appendix A). Sometimes such a demand is rather troublesome. In this article the procedure is described for checking up the tests independence using a random/pseudorandom generator with unknown output distribution.

2 Independence of statistical tests (main definitions)

Let T_1 and T_2 be some statistical tests with levels of significance α_1 and α_2 respectively, the random values ξ_1 and ξ_2 are the indicators that some random value sequence X passed tests T_1 and T_2 respectively (i.e., the basic test basic hypothesis is adopted):

$$\xi_1 = \mathbf{1}\{X\text{passed the test } T_1\},$$

$$\xi_2 = \mathbf{1}\{X\text{passed the test } T_2\}.$$

Definition 1. *Tests T_1 and T_2 are called independent (at a given levels of significance of α_1 and α_2) if random values ξ_1 and ξ_2 are independent.*

By this definition tests independence means that the result of the test T_1 does not depend on the result of the test T_2. The condition in definition 1 is weaker than the condition of test statistics independence. Similarly we can specify any number of independent statistical tests.

Definition 2. *The tests $T = \{T_j\}_{1\le j\le N}$ with given levels of significance $\{\alpha_j\}_{1\le j\le N}$ are called independent if the correspondent indicators $\xi_1, \ldots, \xi_N$ are independent in common.*

As numerous experimental results show, the property of independence of statistical tests doesn't depends on the tests parameters, such as alphabet of sequences, length of sequences, levels of significance, etc.

So far as there is no way to give a rigorous proof of the dependence or independence of the random values $\xi_1, \ldots, \xi_N$, the hypothesis of the tests independence will be checked with statistical methods.

Let now and later $\{T_j\}_{1\le j\le N}$ be the set of statistical tests with the levels of significance, $\{\alpha_j\}_{1\le j\le N}$, respectively.

Definition 3. *With each set of tests $T = \{T_j\}_{1\le j\le N}$ we associate some vector $t = (t_1, \ldots, t_N)$, $t_i \in \{0, 1\}$, $1 \le j \le N$, which is called a template of passing of the tests $\{T_j\}_{1\le j\le N}$, and the element t_j is called a template of passing of the test T_j.*

Let X be a random element on A^l and the random values ξ_j are defined as follows:

$$\xi_j(X) = \begin{cases} 1, & \text{if } X \text{passed the test } T; \\ 0, & \text{otherwise.} \end{cases}$$

If the probability distribution of the random element X is given by some pseudo-random (random) generator with some probability characteristics, the distribution of random values ξ_j will be also defined.

Definition 4. *We say that sequence X passes the tests set $T = \{T_j\}_{1\le j\le N}$ according to the template $t = (t_1, \ldots, t_N)$, if $\forall j = \overline{1, N}$: $\xi_j(X) = t_j$.*

Set the following designations:

$$\eta_j^{t_j}(X) = \mathbf{1}\{\xi_j(X) = t_j\}, j = \overline{1, N}, \tag{1}$$

$$\eta^{t_1\ldots t_N}(X) = \mathbf{1}\{\xi_1(X) = t_1, \ldots \xi_N(X) = t_N\} = \prod_{j=1}^{N} \mathbf{1}\{\xi_j(X) = t_j\} = \prod_{j=1}^{N} \eta_j^{t_j}(X), \tag{2}$$

$$p_j^{t_j}(X) = P\{\xi_j(X) = t_j\} = P\{\eta_j^{t_j}(X) = 1\}, j = \overline{1, N}, \tag{3}$$

$$p^{t_1\ldots t_N}(X) = P\{\eta_1^{t_1}(X) = 1, \ldots \eta_N^{t_N}(X) = 1\}. \tag{4}$$

We will assume that the generator that produces random value sequences has the next properties:

(P1): the generator is a stationary source of random value sequences;

(P2): the sequence X_i, $i \geq 1$ can be chosen so that they are independent random elements of A^l.

Under this conditions the secuences sequences X_i, $i \geq 1$ can be assumed as independent in common identically distributed elements of A^l.

Due to properties (P1) and (P2), $\forall i = \overline{1,n}$ the values $p_j^{t_j}(X_i)$ and $p_j^{t_1 \dots t_N}(X_i)$ do not depend on the index i; we will mark them $p_j^{t_j}$ and $p_j^{t_1 \dots t_N}$, respectively.

If the template t is fixed we will use denominators $\eta_j(X_i)$, $j = \overline{1,N}$, $i = \overline{1,n}$; $\eta(X_i)$; p_j and p instead of $\eta_j^{t_j}(X_i)$, $j = \overline{1,N}$, $i = \overline{1,n}$; $\eta^{t_1 \dots t_N}(X_i)$; $p_j^{t_j}$ and $p^{t_1 \dots t_N}$, respectively.

Also define random values

$$\theta_j^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \eta_j(X_i), j = \overline{1,N}, \tag{5}$$

$$\theta^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \eta(X_i). \tag{6}$$

Note that under the condition of tests independence the equality $p = \prod\limits_{j=1}^{N} p_j$ is true.

In our notifications the next theorem is valid.

Theorem 1. *Let some values $0 < \gamma < 1$ and $\epsilon > 0$ are given. If the tests from the set T are independent in the sense of definition 2 and*

$$n \geq \max \left\{ 9 \cdot \left(\frac{N}{2\epsilon} \Phi^{-1} \left(1 - \frac{1-\gamma}{4N} \right) \right)^2, \; 9 \cdot \left(\Phi^{-1} \left(\frac{3+\gamma}{4} \right) / \epsilon \right)^2 \right\}, \tag{7}$$

where $\Phi(\cdot)$ is a standard normal distribution function,
then the following equality is true:

$$P\{\rho^{(n)} \in (\rho_1 - \epsilon/3, \rho_2 + \epsilon/3)\} \geq \gamma,$$

where

$$\rho^{(n)} = \prod_{j=1}^{N} \theta_j^{(n)} = \frac{1}{n^N} \prod_{j=1}^{N} \sum_{j=1}^{N} \eta_i^{(j)}, \tag{8}$$

$$\rho_1 = \frac{n}{t^2+n} \left(\theta^{(n)} + \frac{t^2}{2n} - t \sqrt{\frac{\theta^{(n)}(1-\theta^{(n)})}{n} + \left(\frac{t}{2n} \right)^2} \right), \tag{9}$$

$$\rho_1 = \frac{n}{t^2+n} \left(\theta^{(n)} + \frac{t^2}{2n} + t \sqrt{\frac{\theta^{(n)}(1-\theta^{(n)})}{n} + \left(\frac{t}{2n} \right)^2} \right), \tag{10}$$

$$t = \Phi^{-1} \left(\frac{3+\gamma}{4} \right).$$

At that the length of interval $(\rho_1 - \epsilon/3, \rho_2 + \epsilon/3)$ is not greater than ϵ.

3 Results

The following results were obtained in [3].

The generator described in [4] was used. The set of 43 tests from [1] was accepted to be dependent. But the set of 4 tests (Frequency (Monobits) Test, Frequency Test within a Block, Runs Test, Test for the Longest Run of Ones in a Block) was accepted to be independent.

In this work following results were obtained.

The set of 4 tests (Frequency (Monobits) Test, Frequency Test within a Block, Runs Test, Test for the Longest Run of Ones in a Block) was tested. The template was $t = (1, 1, 1, 1)$, in other words the passing of all tests simultaneously was checked.

The generator described in [4] and the generators with some defects, such as nonuniform distribution of outs, dependence between symbols of outs and their places, dependence between symbols of outs, were used.

The set of the tests was accepted to be independent for all these generators. But trivial values were obtained when the generator with nonuniform distribution of outs was used.

Further work will deal with obtaining and researching of practical results.

References

[1] A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST 800-221.

[2] George Marsaglia, DIEHARD Statistical Tests: http://stat.fsu.edu/ geo/diehard.html.

[3] Ковальчук Л.В., Бездітний В.Т. (2006). Перевірка незалежності статистичних тестів, призначених для оцінки криптографічних якостей ГВП. *Захист інформації*, № **2(29), 2006**.

[4] Національний стандарт України. Інформаційні технології. Крипографічний захист інформації. Цифровий підпис, що грунтується на еліптичних кривих. Формування та перевіряння. ДСТУ 4145-2002.