CRYPTOLOGY

УДК 004 (056.5+421.5)

ЛЮДМИЛА КОВАЛЬЧУК,
НАТАЛІЯ КУЧИНСЬКА

МЕТОДИКИ ПЕРЕВІРКИ НЕЗАЛЕЖНОСТІ СТАТИСТИЧНИХ ТЕСТІВ

Необхідною умовою стійкості криптосистеми є наявність певних криптографічних властивостей у псевдовипадкового генератора, що використовується в ній. Тому і перед розробником і перед користувачем такої системи постають питання перевірки якості генератора або його окремих послідовностей. Основним сучасним методом перевірки якості псевдовипадкової послідовності є застосування наборів статистичних тестів. На даний час існує декілька наборів статистичних тестів, серед них можна виділити найбільш широко вживаний пакет NIST. Вибір статистичних тестів та компонування набору є складною задачею, оскільки тести мають не тільки перевіряти близькість псевдовипадкової послідовності до випадкової, але й проводити таку перевірку ефективно, тобто за прийнятний час. Однак, такого компромісу досягти достатньо складно, і більшість з сучасних наборів статистичних тестів містять залежні тести, що невиправдано збільшує час їх роботи. Отже, одним з головних питань при побудові набору та використанні статистичних тестів залишається їх незалежність. Тобто для створення ефективного набору тестів мають використовуватись тести між якими відсутні статистичні залежності, і при цьому їх кількість має залишатись достатньо повною. Зазначимо, що на даний час більшість питань щодо формування набору тестів, визначення їх кількості, вибору значення помилки першого роду, тощо, вирішуються інтуїтивно та емпірично. В статті представлено огляд існуючих методик перевірки незалежності статистичних тестів оцінки якості псевдовипадкової послідовності та запропоновано нову математично обґрунтовану методику, яка може застосовуватись для довільної кількості тестів та довільної кількості випадкових послідовностей, і має переваги у швидкодії та простоті реалізації. В роботі також представлено результати проведених експериментальних досліджень щодо застосування нової методики до набору тестів, що входять до складу пакету NIST, та побудовано набір незалежних тестів, придатних для довільного алфавіту.

Ключові слова: генератор псевдовипадкових послідовностей, псевдовипадкова послідовність, якість псевдовипадкової послідовності, статистичні тести оцінки якості псевдовипадкової послідовності, незалежність тестів.

Постановка проблеми. Необхідною умовою стійкості криптосистеми є наявність певних криптографічних властивостей у генератора, що в ній використовується. До цих властивостей, зокрема, відносяться: рівноймовірність та незалежність символів вихідної послідовності; непередбачуваність; великий період (більший за 2^{80} біт). У зв'язку з цим перед розробником або користувачем будь-якої криптосистеми постають питання як про оцінку якості генератора, так і про оцінку якості його окремих послідовностей.

Аналіз останніх досліджень і публікацій. Основним методом перевірки якості генератора є застосування до нього певних статистичних критеріїв. Існує досить багато наборів таких критеріїв; найбільш популярними серед них є NIST та DIEHARD. Однак доцільно зазначити, що на даний час більшість питань щодо формування набору тестів, визначення їх кількості, вибору значення помилки першого роду та інші питання вирішуються інтуїтивно та емпірично. Актуальність даної роботи підтверджується і підвищеним інтересом до задачі формування набору статистичних тестів за останні кілька років. Наприклад, у роботі [1] розглядається питання обчислення загальної помилки 1-го роду для пакету тестів, відповідь

на яке практично неможливо отримати, якщо тести є незалежними. У роботах [2], [3] вивчається питання попарної незалежності тестів з набору NIST з використанням експериментальних даних. Зазначимо, що теоретичні результати, отримані у даній роботі, повністю перекривають і узагальнюють результати цих робіт і дають відповіді на поставлені там запитання.

Метою даної статті є отримання математично обґрунтованих способів формування набору статистичних критеріїв. На основі запропонованих методик в роботі буде проаналізовано набір тестів, що входять до складу пакету NIST, а також побудовано набір незалежних тестів, придатних для довільного алфавіту.

Виклад основного матеріалу дослідження.

1. Аналіз попередніх результатів.

Розглянемо методики перевірки незалежності статистичних тестів, які раніше розкривались у літературі.

1.1. Перевірка незалежності тестів згідно NIST 800-221 та NIST 800-22 (1a). Вперше питання необхідності перевірки незалежності статистичних тестів для усунення "надлишкових" тестів піднімається у стандарті NIST 800-221 "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications" [4]. Також про те, що потрібно позбавлятись від надлишкових тестів, йдеться у однойменному оновленні цього стандарту від 2010 року NIST 800-22 (1a) [5]. Проте процедура перевірки незалежності тестів, тобто, їх некорельованості, в жодній версії цього стандарту не обґрунтована та не формалізована. Для перевірки незалежності тестів пропонується деяка процедура, опис якої є доволі емпіричним та незрозумілим. Далі опишемо цю процедуру .

Спочатку для великої кількості послідовностей будується матриця

$$Z = \left(z_{ij}\right)_{i=1,\,j=1}^{l,1000},$$

де $z_{ij} = p_{ij}$ для [4] та $z_{ij} = \Phi^{-1}\left(p_{ij}\right)$ для [2];

Φ – функція нормального розподілу;

p_{ij} – P-величина, що відповідає результату проходження j-ою послідовністю i-го тесту;

l – кількість тестів ($l = 189$ для [4] та $l = 161$ для [5]).

Далі в [4] перевіряється, що рядки цієї матриці є лінійно незалежними, і на основі цього робиться висновок про незалежність тестів. У [5] аналізуються стовпчики цієї матриці, що відповідають результатам проходження j-ою послідовністю усіх тестів. Для аналізу потрібно оцінити деяку величину, що у дослівному перекладі означає "найбільшу мінливість" (мовою оригіналу – "the largest variability"), і яка певною мірою характеризує залежність тестів. Проте ні формули, за якою ця величина обчислюється, ні правила прийняття рішень щодо незалежності чи залежності тестів при певних значеннях цієї величини у стандарті не наведено. Лише зазначено, що автори стандарту проаналізували тести та вважають їх незалежними.

Тому за результатами аналізу документу NIST 800-22 (1a) можна зробити наступні висновки:

1) з метою виключення "надлишкових" тестів з набору, призначеного для тестування генераторів випадкових послідовностей (ГВП) або генераторів псевдовипадкових послідовностей (ГПВП), необхідно перевіряти тести цього набору на незалежність;

2) автори документу NIST 800-22 (1a) не надали пояснення запропонованої методики, достатнього для розуміння принципів її роботи, а також не навели ні її обґрунтування, ні формалізації;

3) питання розробки та удосконалення відповідної методики досі залишається відкритим.

1.2. Методика перевірка незалежності тестів, що використовує центральну граничну теорему. Ця методика була вперше запропонована та обґрунтована у роботі [6], а потім удосконалена та узагальнена у роботах [7], [8]. Проаналізуємо її.

Нехай A і B – деякі тести. Випадкові величини ξ_A та ξ_B – індикатори проходження послідовністю тестів A та B відповідно, тобто

$$\xi_A = \mathrm{I}\begin{Bmatrix}\text{послідовність}\\ \text{пройшла тест } A\end{Bmatrix},\ \xi_B = \mathrm{I}\begin{Bmatrix}\text{послідовність}\\ \text{пройшла тест } B\end{Bmatrix}.$$

Означення 1. Тести A та B назвемо незалежними (відносно деякого фіксованого генератора G), якщо індикатори ξ_A, ξ_B статистично незалежні.

Іншими словами, незалежність тестів означає, що результат застосування тесту А для послідовності не залежить від результату застосування тесту В. Дане означення більш слабке, ніж означення незалежності відповідних статистик. Зауважимо, що залежність від генератора у означенні 1 є суттєвою.

Аналогічно можна визначити набір незалежних тестів.

Означення 2. Тести з набору $\{A_i\}$ назвемо незалежними (відносно деякого фіксованого генератора G), якщо незалежні у сукупності відповідні індикатори.

Позначимо через ζ_A, ζ_B статистики, які обчислюються в тестах A і B, K_A, K_B – критичні області тестів, відповідно. Тоді розподіл величин ξ_A, ξ_B є наступним:

$$P(\xi_A = 1) = P(\zeta_A \notin K_A),\ P(\xi_A = 0) = P(\zeta_A \in K_A),$$
$$P(\xi_B = 1) = P(\zeta_B \notin K_B),\ P(\xi_B = 0) = P(\zeta_B \in K_B).$$

Тому для незалежності ξ_A, ξ_B необхідно і достатньо, щоб були незалежними події $\{\zeta_A \in K_A\}$ та $\{\zeta_B \in K_B\}$. Тобто незалежність тестів еквівалентна незалежності подій, що полягають у попаданні статистик у критичні області.

Нехай потрібно перевірити незалежність набору з N тестів, де i-й тест має рівень значимості α_i, $i = \overline{1, N}$. Для перевірки незалежності набору використовується ГВП, який має необхідні статистичні якості. Це означає, що даний генератор повинен пройти тестування набором тестів, для якого перевіряється незалежність, та результати тестування повинні бути оцінені за методикою, вказаною в [4], [5]. За цих умов можна вважати, що цей генератор виробляє послідовності, які є реалізаціями незалежних, рівноймовірно розподілених на заданому алфавіті випадкових величин. При цьому залежність чи незалежність тестів не впливають на якість оцінки ГВП, а впливають лише на час проведення тестування.

Припустимо отримано n послідовностей з ГВП (довжини послідовностей повинні бути придатними для проведення тестування).

Теорема 1 ([6])**.** Покладемо $\eta_j = 1$, якщо j-та послідовність пройшла всі тести, та $\eta_j = 0$ у протилежному випадку, тобто якщо j-та послідовність не пройшла хоча б один тест, $j = \overline{1, n}$. Позначимо через η частку послідовностей, які пройшли всі тести: $\eta = \frac{1}{n} \cdot \sum_{j=1}^{n} \eta_j$. Покладемо

$$a = \prod_{i=1}^{N} (1 - \alpha_i),\ \sigma^2 = \frac{1}{n} \cdot \left(a - a^2\right). \tag{1}$$

Тоді, за умови незалежності тестів набору, випадкова величина $\frac{\eta - a}{\sigma}$ має асимптотично стандартний нормальний розподіл, тобто

$$P\left(\frac{\eta - a}{\sigma} < x\right) \approx \Phi(x),$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ – функція стандартного нормального розподілу.

Дана теорема дає можливість побудувати ефективну та математично обґрунтовану методику перевірки незалежності тестів, що застосовуються для оцінки якостей ГВП.

Методика 1. Перевірка незалежності тестів з використанням Центральної

граничної теореми (ЦГТ). Нехай для оцінки якостей ГВП застосовується набір з N тестів, які мають один і той же рівень значимості α. Використовуючи результати попереднього пункту, можна запропонувати наступну методику для перевірки незалежності тестів цього набору.

1. Задати β – рівень значимості для статистичної перевірки незалежності тестів цього набору (ймовірність помилки першого роду).
2. Обчислити наступні величини: $a = p = (1-\alpha)^N$, $q = 1-(1-\alpha)^N$, $\sigma^2 = \frac{p \cdot q}{n}$.
3. Використовуючи n послідовностей, отриманих з ГВП, обчислити η_j, $j = \overline{1,n}$.
4. Обчислити $\eta = \frac{1}{n} \cdot \sum_{j=1}^{n} \eta_j$.
5. Обчислити значення $erfc\left(\frac{|\eta - a|}{\sigma \cdot \sqrt{2}}\right)$.
6. Якщо $erfc\left(\frac{|\eta - a|}{\sigma \cdot \sqrt{2}}\right) \geq \beta$, то гіпотеза про незалежність тестів у наборі приймається; в іншому випадку – відхиляється.

Тут під функцією $erfc(x)$ розуміється так звана "функція помилок" або функція напівнормального розподілу.

Зауваження.

1) Імовірність "відбраковки" набору з незалежних тестів не перевищує β.
2) Процедури, описані у пунктах 5 та 6, можна замінити на наступні:

   5'. Визначити константу C за умови: $\Phi(C) = 1 - \beta/2$, де $\Phi(x)$ – функція стандартного нормального розподілу та побудувати інтервал $\Delta = (a - \sigma C, a + \sigma C)$.

   6'. Якщо $\eta \in \Delta$, то гіпотеза про незалежність тестів у наборі приймається; в іншому випадку – відхиляється.

Зазначимо, що наведена вище методика спирається лише на строго доведені твердження і є повністю обґрунтованою з математичної точки зору [6]. Дієздатність даної методики та зручність у практичному використанні підтверджена застосуваннями для перевірки багатьох відомих наборів статистичних тестів. Показовим прикладом її практичного застосування є перевірка незалежності набору тестів, що входять до пакетів [4], [5], а також до більш простого набору з шести тестів, описаних у [6], [7]. Методика є універсальною і може бути застосована до будь-якої кількості тестів будь-якої природи. Вона є адаптивною, тобто її параметри обчислюються у залежності від певних параметрів набору (кількість тестів, рівні значимості). Крім того методика є зручною, алгоритм її використання є чітким та простим для розуміння, а також досить швидкодіючим.

Проте запропонований підхід має і значні недоліки.

1. Як зазначалось раніше, ця методика може застосовуватись тільки до "ідеального" з точки зору криптографічних властивостей генератора. Наприклад, вона добре працює для генератора, описаного у додатку А ДСТУ 4145-2002 [9]. Але якщо розподіл на послідовностях, що виробляє генератор, має відхилення від "ідеального", то умови теореми порушуються.
2. При формулюванні теореми припускається, що якщо рівень значимості тесту дорівнює α, то математичне сподівання пропорції послідовностей, що проходять цей тест, дорівнює $1-\alpha$. Це припущення буде вірним у випадку, якщо всі тести, що перевіряються, базуються на стандартних граничних розподілах. Тому дана методика не може застосовуватись у багатьох випадках для тестів, що будуються з використанням комбінаторних міркувань (наприклад, тест серій максимальної довжини). Для деяких таких тестів імовірність помилки першого роду не дорівнює імовірності проходження тесту послідовністю, а є лише її верхньою оцінкою. Тому на практиці дане припущення може виявитись обтяжливим.

3. Для того, щоб наближення нормальним розподілом було коректним, кількість послідовностей повинна бути близько 10000. Тестування такої кількості послідовностей вимагає досить великого часу, особливо якщо тести є громіздкими та їх багато.

Зауважимо, що на практиці методика добре працює і в тому випадку, коли умови, названі у пунктах 2 та 3, не виконуються. Але тоді у неї немає строгого теоретичного обґрунтування, вона стає більш емпіричною.

Для подолання недоліку, зазначеного у пункті 2, методику удосконалено [7], [8] та узагальнено на випадок генератора з довільним розподілом та на випадок тестів довільної природи. Але при цьому удосконалена методика стала більш громіздкою та затратною, і вимагає ще більшу кількість послідовностей (в деяких випадках – близько 1000000). Але це не є критичним для роботи методики, хоча суттєво збільшує час її роботи.

1.3. Методика перевірки незалежності тестів, що використовує взаємну коваріацію випадкових величин. Ця методика була вперше запропонована та обґрунтована у подальших роботах авторів [7], [8]. При побудові методики перевірки незалежності статистичних тестів скористаємось теорією, викладеною в [10].

Нехай $y_1,\ldots, y_n$ – деякі випадкові величини. Позначимо y_{ij}, $i=\overline{1,n}$, $j=\overline{1,n}$ – результат j-го спостереження (вимірювання) випадкової величини y_i, $\overline{y_i}$ – її вибіркове середнє: $\overline{y_i}=\frac{1}{s}\sum_{j=1}^{s} y_{ij}$. Побудуємо матрицю вибіркової коваріації цих випадкових величин:

$$D=\left(d_{ij}\right)_{i,j=1}^{n},\ d_{ij}=\frac{1}{s-1}\sum_{k=1}^{s}\left(y_{ik}-\overline{y_i}\right)\left(y_{jk}-\overline{y_j}\right) \tag{2}$$

та їх матрицю вибіркової кореляції:

$$R=\left(R_{ij}\right)_{i,j=1}^{n},\ R_{ij}=\frac{D_{ij}}{\sqrt{D_{ii}D_{jj}}}. \tag{3}$$

Теоретично для незалежних випадкових величин R – одинична матриця і її визначник повинен бути рівний 1. Але в даному випадку є лише оцінка R, побудована на n вибірках обсягу s значень випадкових величин. Тому в якості критерію перевірки гіпотези H_0: $R=I$ беруть визначник: $V=\det R$.

Закон розподілу V досить складний, але при досить великих значеннях s можна використовувати його асимптотичне подання:

$$P\{-m\cdot\ln V\le\nu\}=P\left\{\chi_f^2\le\nu\right\}+\left(\gamma_2/m^2\right)\cdot\left[P\left\{\chi_{f+4}^2\le\nu\right\}-P\left\{\chi_f^2\le\nu\right\}\right]+O\left(m^{-3}\right), \tag{4}$$

де $f=\frac{n(n-1)}{2}$;

$m=s-\frac{2n+11}{6}$,

$\gamma_2=\frac{n(n-1)}{288}(2n^2-2n-13)$.

Зауважимо, що в “ідеалі” $\ln V=0$. Отже, критична область для перевірки гіпотези буде $m\cdot\ln V>\nu$ при рівні значущості $\alpha=1-P\{-m\cdot\ln V\le\nu\}$.

Вираз (4) можна використовувати, якщо

$$\left(\gamma_2/m^2\right)<1. \tag{5}$$

Помітимо, що другий доданок в (4) невід’ємний. Тому, вибравши спочатку критичну область для заданого рівня значущості та використовуючи тільки перший доданок, отримаємо в результаті дещо більший рівень значимості для даної критичної області. Якщо при обчисленні ймовірності взяти два доданка у (4), то результат буде трохи точнішим.

Використовуючи наведені результати, побудуємо методику перевірки незалежності статистичних тестів.

Методика 2. Перевірка незалежності тестів з використанням вибіркової кореляційної матриці.

1) отримаємо значення y_{ij}, $i=\overline{1,n}$, $j=\overline{1,n}$, де y_{ij} є результатом тестування i-им тестом j-ї послідовності, та обчислимо величини $\overline{y_i}=\frac{1}{s}\sum_{j=1}^{s} y_{ij}$;

2) за формулами (2) та (3) обчислимо кореляційну матрицю R;

3) обчислимо визначник V кореляційної матриці;

4) задаємо рівень значущості β, за таблицею знаходимо квантіль χ^2_β з f ступенями свободи, де $f=\frac{n(n-1)}{2}$;

5) якщо $\chi^2_\beta > -m\cdot\ln V > \nu$, то гіпотеза про незалежність тестів підтверджується, інакше відхиляється.

Методика 2 не має ряду недоліків, притаманних Методиці 1, тобто вона може застосовуватись до довільних тестів та для довільних генераторів, при цьому не є необхідним тестувати занадто велику кількість послідовностей. Всі умови використання цієї методики визначаються вимогою (5). Однак недоліком Методики 2, на відміну від Методики 1, є те, що вона перевіряє лише лінійну залежність відповідних випадкових величин.

Зазначимо, що застосування Методики 1 та Методики 2 до різних пакетів тестів, зокрема до тестів, визначених у [4] - [7] для ГВП з Додатку А у [9], дало приблизно однакові результати.

Підводячи підсумки стосовно викладених вище Методики 1 та Методики 2, а також їх переваг та недоліків, можна сказати наступне. Обидві методики є досить зручними на практиці, і умови їх коректного застосування хоч і призводять до певних обмежень та трудовитрат, проте не є занадто складними і цілком можуть забезпечуватись при практичному використанні методик. Але все одно залишається потреба у розробці та обґрунтуванні такої методики, яка матиме і більш прості умови застосування, і потребуватиме меншої кількості послідовностей, а, отже забезпечуватиме менші трудовитрати і, отже, більшу швидкодію.

2. Побудова та обґрунтування методики перевірки незалежності статистичних тестів, що використовує нерівність Чебишева. Тут і далі використовуватимемо всі позначення, введені вище для Методики 1.

2.1. Розробка та обґрунтування методики перевірки незалежності статистичних тестів, що використовує нерівність Чебишева. Сформулюємо теорему, на базі якої побудуємо методику перевірки незалежності тестів. Спочатку припустимо, що ГВП, який використовується для формування відповідних послідовностей, має хороші криптографічні якості, а рівень значимості кожного тесту дорівнює середній кількості послідовностей, що не проходять цей тест. А потім покажемо, як можна модифікувати цю методику для більш загального випадку.

Теорема 2. Покладемо $\eta_j=1$, якщо j-та послідовність пройшла всі тести, та $\eta_j=0$ у протилежному випадку, тобто якщо j-тою послідовністю хоча б один тест не пройдено, $j=\overline{1,n}$. Позначимо через η частку послідовностей, які пройшли всі тести: $\eta=\frac{1}{n}\cdot\sum_{j=1}^{n}\eta_j$.

Покладемо $a=\prod_{i=1}^{N}(1-\alpha_i)$, $\sigma^2=\frac{a(1-a)}{n}$. Тоді, якщо тести незалежні, то

$$P\{|\eta-a|>\varepsilon\}\le\frac{\sigma^2}{\varepsilon^2}.$$

Доведення: Введемо випадкові величини $\eta_i^{(j)}$ – індикатори проходження j-тою послідовністю i-го тесту, $i=\overline{1,N}$, $j=\overline{1,n}$ наступним чином:

$\eta_i^{(j)} = 1$, якщо j-та послідовність пройшла i-й тест,

$\eta_i^{(j)} = 0$, в іншому разі.

Величина $\eta_i^{(j)}$ має розподіл Бернулі: $P(\eta_i^{(j)} = 0) = \alpha_i$, $P(\eta_i^{(j)} = 1) = 1 - \alpha_i$.

Очевидно, $\eta_j = \prod_{i=1}^{N} \eta_i^{(j)}$. Якщо набір складається з незалежних тестів, то для кожного j випадкові величини $\{\eta_i^{(j)}\}_{i=1}^{N}$ незалежні в сукупності. В цьому випадку

$$P(\eta_j = 1) = \prod_{i=1}^{N} (1 - \alpha_i),\ P(\eta_j = 0) = 1 - \prod_{i=1}^{N} (1 - \alpha_i).$$

Математичне сподівання та дисперсія випадкової величини η_j знаходяться за формулами:

$$M\eta_j = \prod_{i=1}^{N} (1 - \alpha_i),\ D\eta_j = \prod_{i=1}^{N} (1 - \alpha_i) - \left(\prod_{i=1}^{N} (1 - \alpha_i) \right)^2.$$

За припущеннями, ГВП має необхідні статистичні якості, тоді генеровані ним послідовності незалежні. Отже, випадкові величини $\{\eta_j\}_{j=1}^{n}$ незалежні у сукупності. Тому

$$M\eta = \frac{1}{n} \cdot \sum_{j=1}^{n} M\eta_j = \prod_{i=1}^{N} (1 - \alpha_i) = a,$$

$$D\eta = \frac{1}{n^2} \cdot \sum_{j=1}^{n} D\eta_j = \frac{1}{n} \cdot \left(\prod_{i=1}^{N} (1 - \alpha_i) - \left(\prod_{i=1}^{N} (1 - \alpha_i) \right)^2 \right) = \sigma^2.$$

Застосовуючи нерівність Чебишева до випадкової величини η, отримуємо:

$$P\{|\eta - a| > \varepsilon\} \le \frac{\sigma^2}{\varepsilon^2}.$$

Теорему доведено.

Використовуючи наведені результати, побудуємо методику перевірки незалежності статистичних тестів. За структурою вона дуже подібна до Методики 1, але замість наближення стандартним нормальним розподілом використовуватимемо нерівність Чебишева.

Методика 3. Перевірка незалежності тестів з використанням нерівності Чебишева.

1. Задати β – рівень значимості для статистичної перевірки незалежності тестів цього набору (ймовірність помилки першого роду).
2. Обчислити наступні величини: $a = \prod_{i=1}^{N} (1 - \alpha_i)$, $\sigma^2 = \frac{a(1-a)}{n}$.
3. Використовуючи n послідовностей, отриманих з ГВП, обчислити η_j, $j = \overline{1, n}$.
4. Обчислити $\eta = \frac{1}{n} \sum_{j=1}^{n} \eta_j$.
5. Обчислити $\varepsilon = \sqrt{\frac{\sigma^2}{\beta}}$.
6. Обчислити границі довірчого інтервалу: $I_1 = \max\{0, a - \varepsilon\}$; $I_2 = \min\{1, a + \varepsilon\}$.
7. Якщо $\eta \in [I_1, I_2]$, то гіпотеза про незалежність тестів приймається, інакше – відхиляється.

Зауваження 1. Якщо всі тести з набору мають однаковий рівень значимості α і базуються на граничних розподілах, то параметр *a* у пункті 2 Методики 3 обчислюється як $a = (1 - \alpha)^N$.

Зауваження 2. Якщо ГВП має певні статистичні відхилення, або якщо у наборі є тести, в яких рівень значимості може не співпадати з імовірністю відхилення послідовності, то у пункті 2 Методики 3 замість величин α_i, $i=\overline{1,N}$, потрібно взяти величини A_i, $i=\overline{1,N}$, де A_i дорівнює відношенню кількості послідовностей, що не пройшли тест з номером i, до кількості всіх послідовностей. При цьому точність обчислень буде тим більшою, чим більша кількість послідовностей. Тому в цьому випадку рекомендується брати не менше 1000 послідовностей. При цьому необхідною умовою коректної роботи методики є умова

$$\forall i=\overline{1,N}: A_i \neq 1. \quad (6)$$

Зазначимо, що, з урахуванням зауваження 2, Методика 3 може застосовуватись до набору тестів з різними принципами побудови, а також може використовувати ГВП з різними "незначними" відхиленнями від "ідеального", де під незначними відхиленнями розуміються такі відхилення, які не порушують умову (6).

Особливою перевагою Методики 3 є те, що для її коректної роботи достатньо порівняно невеликої кількості послідовностей – вже при $n=100$ методика дає коректні результати, тобто нетривіальні довірчі інтервали.

2.2. Обчислення параметрів та приклади застосування Методики 3 до різних наборів тестів.

У табл. 1 – табл. 6 наведено приклади значень, які приймають величини a, σ, ε, I_1, I_2 при різних вихідних даних, а саме при різних значеннях n, N, α_i та β.

Таблиця 1 – Значення величин a, σ, ε, I_1, I_2 при $N=6$, $\alpha_i=0{,}01$, $i=\overline{1,6}$, $\beta=0{,}01$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,94148	0,000550953	0,234724	0,706756	1
200	0,94148	0,000275476	0,165975	0,775505	1
300	0,94148	0,000183651	0,135518	0,805962	1
400	0,94148	0,000137738	0,117362	0,824118	1
500	0,94148	0,000110191	0,104972	0,836508	1
600	0,94148	9,18255E-05	0,095826	0,845655	1
700	0,94148	7,87075E-05	0,088717	0,852763	1
800	0,94148	6,88691E-05	0,082987	0,858493	1
900	0,94148	6,1217E-05	0,078241	0,863239	1
1000	0,94148	5,50953E-05	0,074226	0,867254	1

Таблиця 2 – Значення величин a, σ, ε, I_1, I_2 при $N=6$, $\alpha_i=0{,}05$, $i=\overline{1,6}$, $\beta=0{,}05$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,735092	0,001947	0,197348	0,537744	0,93244
200	0,735092	0,000974	0,139546	0,595546	0,874638
300	0,735092	0,000649	0,113939	0,621153	0,849031
400	0,735092	0,000487	0,098674	0,636418	0,833766
500	0,735092	0,000389	0,088257	0,646835	0,823349
600	0,735092	0,000325	0,080567	0,654525	0,815659
700	0,735092	0,000278	0,074591	0,660501	0,809683
800	0,735092	0,000243	0,069773	0,665319	0,804865
900	0,735092	0,000216	0,065783	0,669309	0,800875
1000	0,735092	0,000195	0,062407	0,672685	0,797499

Таблиця 3 – Значення величин a, σ, ε, I_1, I_2 при $N=6$, $\alpha_i=0{,}001$, $i=\overline{1,6}$, $\beta=0{,}001$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,994015	5,9492E-05	0,24391	0,750105	1
200	0,994015	2,9746E-05	0,17247	0,821545	1
300	0,994015	1,98307E-05	0,140821	0,853194	1
400	0,994015	1,4873E-05	0,121955	0,87206	1
500	0,994015	1,18984E-05	0,10908	0,884935	1
600	0,994015	9,91533E-06	0,099576	0,894439	1
700	0,994015	8,49886E-06	0,092189	0,901826	1
800	0,994015	7,4365E-06	0,086235	0,90778	1
900	0,994015	6,61022E-06	0,081303	0,912712	1
1000	0,994015	5,9492E-06	0,077131	0,916884	1

Таблиця 4 – Значення величин a, σ, ε, I_1, I_2 при $N=16$, $\alpha_i=0{,}01$, $i=\overline{1,16}$, $\beta=0{,}01$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,851458	0,001265	0,355637	0,495821	1
200	0,851458	0,000632	0,251473	0,599985	1
300	0,851458	0,000422	0,205327	0,646131	1
400	0,851458	0,000316	0,177818	0,673639	1
500	0,851458	0,000253	0,159046	0,692412	1
600	0,851458	0,000211	0,145188	0,70627	0,996646
700	0,851458	0,000181	0,134418	0,71704	0,985876
800	0,851458	0,000158	0,125737	0,725721	0,977194
900	0,851458	0,000141	0,118546	0,732912	0,970003
1000	0,851458	0,000126	0,112462	0,738996	0,96392

Таблиця 5 – Значення величин a, σ, ε, I_1, I_2 при $N=16$, $\alpha_i=0{,}05$, $i=\overline{1,16}$, $\beta=0{,}05$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,440127	0,002464	0,221998	0,218129	0,662124
200	0,440127	0,001232	0,156976	0,28315	0,597103
300	0,440127	0,000821	0,128171	0,311956	0,568297
400	0,440127	0,000616	0,110999	0,329128	0,551126
500	0,440127	0,000493	0,09928	0,340846	0,539407
600	0,440127	0,000411	0,09063	0,349496	0,530757
700	0,440127	0,000352	0,083907	0,356219	0,524034
800	0,440127	0,000308	0,078488	0,361639	0,518615
900	0,440127	0,000274	0,073999	0,366127	0,514126
1000	0,440127	0,000246	0,070202	0,369925	0,510329

За результатами застосування Методики 3 до пакетів статистичних тестів, описаних у [4] - [7], з використанням ГПВП з Додатку А у [9], були отримані наступні результати.

1. Для пакету NIST-I, описаному у [4], за результатами 10 експериментів при $\beta=10^{-3}$ тести завжди виявлялись залежними; при $\beta=10^{-4}$ завжди виявлялись незалежними.

Таблиця 6 – Значення величин a, σ, ε, I_1, I_2 при $N=16$, $\alpha_i=0{,}001$, $i=\overline{1,16}$, $\beta=0{,}001$ та при різних значеннях n

n	a	σ^2	ε	I_1	I_2
100	0,984119	0,000156	0,395327	0,588792	1
200	0,984119	7,81E-05	0,279539	0,704581	1
300	0,984119	5,21E-05	0,228242	0,755877	1
400	0,984119	3,91E-05	0,197664	0,786456	1
500	0,984119	3,13E-05	0,176796	0,807324	1
600	0,984119	2,6E-05	0,161392	0,822728	1
700	0,984119	2,23E-05	0,14942	0,8347	1
800	0,984119	1,95E-05	0,139769	0,84435	1
900	0,984119	1,74E-05	0,131776	0,852344	1
1000	0,984119	1,56E-05	0,125013	0,859106	1

2. Для пакету NIST-II, описаному у [5], за результатами 10 експериментів при $\beta=10^{-3}$ тести 9 разів виявлялись залежними; при $\beta=10^{-4}$ завжди виявлялись незалежними.

3. Для пакету з шести тестів, описаному у [6], [7] (а саме: критерій χ^2 для знаків, критерій χ^2 для біграм, тест кількості серій, тест серій максимальної довжини, критерій інверсій та критерій місць знаків), за результатами 10 експериментів при $\beta=10^{-3}$ тести завжди виявлялись незалежними; при $\beta=10^{-2}$ 9 разів виявлялись незалежними.

Цікаво зазначити, що застосування Методики 3 для зазначеного вище набору з шести тестів навіть для ГПВП з ”незначними” відхиленнями (у сенсі умови (6)) дає приблизно ті ж самі результати.

Висновки. В статті наведено огляд всіх методик перевірки незалежності статистичних тестів, призначених для перевірки якості ГВП, ГПВП та їх послідовностей. Одну з них запропоновано та обґрунтовано вперше. Проведено порівняльний аналіз всіх розглянутих методик, наведено їх переваги та недоліки. За результатами порівняльного аналізу показано, що нова методика (Методика 3) має найбільше переваг як у швидкодії та простоті реалізації, так і у можливості застосування для широкого спектра тестів та генераторів.

Також у роботі проведено експериментальні дослідження стосовно застосування нової методики до двох пакетів тестів. Дослідження підтвердили зручність використання та коректність запропонованої методики.

Отже, основним результатом роботи є математично обґрунтована методика, що може використовуватись для довільної кількості тестів на довільній кількості випадкових послідовностей. Методика добре працює як на якісних генераторах, так і на генераторах зі слабкими статистичними властивостями.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] A.L. Kostevich, and A.V. Shilkin, “Analysis of tests for randomness based on universal predictors: Bernoulli trials case”, in *Proc. 9th International conference Computer data analysis and modeling: complex stochastic data and systems*, Minsk, 2010, pp. 36-39.

[2] A. Doğanaksoy, F. Sulak, M. Uğuz, O. Şeker, and Z. Akcengiz “Mutual correlation of NIST statistical randomness tests and comparison of their sensitivities on transformed sequences”, *Turkish Journal of Electrical Engineering and Computer Sciences*, № 25 (2), pp. 655-665, 2017. doi: 10.3906/elk-1503-214.

[3] F. Sulak, M. Uğuz, O. Kocak, and A. Doğanaksoy “On the independence of statistical randomness tests included in the NIST test suite”, *Turkish Journal of Electrical Engineering and Computer Sciences*, № 25, pp. 3673-3683, 2017. doi: 10.3906/elk-1605-212.

[4] National Institute of Standards and Technology. (2010, April 27). *NIST 800-22, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. [Online]. Available: https://www.ipa.go.jp/files/000011794.pdf Accessed on: Sept. 25, 2017.

[5] National Institute of Standards and Technology. (2010, August 11). *NIST 800-22, Rev. 1a, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications.* [Online]. Available: https://www.ipa.go.jp/files/000011794.pdf Accessed on: Sept. 25, 2017.

[6] Л. Ковальчук, та В. Бездітний, "Перевірка незалежності статистичних тестів, призначених для оцінки криптографічних якостей ГПВ", *Захист інформації*, №2 (29), с. 18-23, 2006.

[7] Л. Ковальчук, В. Бездітний, та Л. Скрипник, "Методика перевірки незалежності статистичних тестів при невідомих статистичних властивостях генератора послідовностей", Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні, Вип. 1 (20), с. 66-71, 2010.

[8] L. Skrypnik, L. Kovalchuk, and V. Bezditnyi, "Method of statistical tests independence checking", in *Proc. 9th International conference Computer data analysis and modeling: complex stochastic data and systems*, Minsk, 2010, с. 71-74.

[9] Держстандарт України. (2002, Груд. 28). *ДСТУ 4145, Інформаційні технології. Криптографічний захист інформації. Цифровий підпис, що ґрунтується на еліптичних кривих. Формування та перевірка*. Київ, 2003, 40 с.

[10] Т. Андерсон, Введение в многомерный статистический анализ. Москва, Россия: Физматгиз, 1963.

Стаття надійшла до редакції 30 вересня 2017 року.

REFERENCE

[1] A.L. Kostevich, and A.V. Shilkin, "Analysis of tests for randomness based on universal predictors: Bernoulli trials case", in *Proc. 9th International conference Computer data analysis and modeling: complex stochastic data and systems*, Minsk, 2010, pp. 36-39.

[2] A. Doğanaksoy, F. Sulak, M. Uğuz, O. Şeker, and Z. Akcengiz "Mutual correlation of NIST statistical randomness tests and comparison of their sensitivities on transformed sequences", *Turkish Journal of Electrical Engineering and Computer Sciences*, № 25 (2), pp. 655-665, 2017. doi: 10.3906/elk-1503-214.

[3] F. Sulak, M. Uğuz, O. Kocak, and A. Doğanaksoy "On the independence of statistical randomness tests included in the NIST test suite", *Turkish Journal of Electrical Engineering and Computer Sciences*, № 25, pp. 3673-3683, 2017. doi: 10.3906/elk-1605-212.

[4] National Institute of Standards and Technology. (2010, April 27). *NIST 800-22, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. [Online]. Available: https://www.ipa.go.jp/files/000011794.pdf Accessed on: Sept. 25, 2017.

[5] National Institute of Standards and Technology. (2010, August 11). *NIST 800-22, Rev. 1a, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications.* [Online]. Available: https://www.ipa.go.jp/files/000011794.pdf Accessed on: Sept. 25, 2017.

[6] L. Kovalchuk, and V. Bezditnyi, "Inspection of statistical tests independence intended for PRNG cryptographic qualities evaluation", *Ukrainian Information Security Research Journal*, № 2 (29), pp. 18-23, 2006.

[7] L. Kovalchuk, V. Bezditnyi, and L. Skrypnyk, "The method of statistical tests independence checking for sequence generator unknown statistical properties", Legal, normative and metrological provision of the information security system in Ukraine, № 4 (21), pp. 35-41, 2010.

[8] L. Skrypnik, L. Kovalchuk, and V. Bezditnyi, “Method of statistical tests independence checking”, in *Proc. 9th International conference Computer data analysis and modeling: complex stochastic data and systems*, Minsk, 2010, c. 71-74.
[9] Governmental Standard of Ukraine. (2002, Dec. 28). *GSTU 4145, Information technology. Cryptographic information protection. Elliptic curves digital signature. Formation and verification.* Kyiv, 2003, 40 p.
[10] T. Anderson, *Introduction to Multidimensional Statistical Analysis*. Moskow, Russia: Fizmatgiz, 1963.

ЛЮДМИЛА КОВАЛЬЧУК,
НАТАЛИЯ КУЧИНСКАЯ

МЕТОДИКИ ПРОВЕРКИ НЕЗАВИСИМОСТИ СТАТИСТИЧЕСКИХ ТЕСТОВ

Необходимым условием стойкости криптосистемы является наличие определенных криптографических свойств у псевдослучайного генератора, который используется в ней. Поэтому и перед разработчиком и перед пользователем такой системы стоят вопросы проверки качества генератора и его отдельных последовательностей. Основные современные методы проверки качества псевдослучайной последовательности сводятся к применению наборов статистических тестов. На данный момент времени существует несколько наборов статистических тестов, среди которых можно выделить наиболее широко используемые наборы статистических тестов NIST. Выбор статистических тестов и создание набора представляет собой сложную задачу, поскольку тесты должны не только проверять насколько псевдослучайная последовательность близка к случайной, но и проводить такую проверку эффективно, то есть за приемлемое время. К сожалению, компромисс в таком случае достигается достаточно тяжело и большинство современных наборов содержат тесты, имеющие статистические зависимости, что неоправданно увеличивает время работы таких наборов. Таким образом, одним из главных вопросов при построении набора и использования статистических тестов остается их независимость. Поэтому для создания эффективного набора должны использоваться тесты, между которыми отсутствуют статистические зависимости и, при этом, количество тестов должно оставаться достаточно полным. Отметим, однако, что сейчас вопросы формирования набора тестов, определение их количества, выбора значения ошибки первого рода и другие решаются интуитивно и эмпирически. В данной статье представлен обзор существующих методик проверки независимости статистических тестов оценки качества псевдослучайной последовательности и предлагается новая, математически обоснованная, методика, которая может быть применена для произвольного количества тестов и произвольного количества случайных последовательностей. Предлагаемая методика имеет преимущества в быстродействии и простоте реализации. В работе так же представлено результаты проведенных экспериментальных исследований применения новой методики к пакету тестов.

Ключевые слова: генератор псевдослучайной последовательности, псевдослучайная последовательность, качество псевдослучайной последовательности, статистические тесты оценки качества псевдослучайной последовательности, независимость тестов.

LIUDMYLA KOVALCHUK,
NATALIIA KUCHYNSKA

METHODS OF STATISTICAL TESTS INDEPENDENCE VERIFICATION

The necessary condition for the cryptosystem security is the certain cryptographic properties of pseudorandom number generator used in it. Therefore, both the developer and the user of such system are faced with the quality checking issues of the generator or its individual sequences. The main modern methods for pseudorandom sequence quality testing are reduced to the statistical randomness tests use. At the moment there are several statistical test suites, among which the most widely used is NIST statistical randomness test suite. The statistical tests choice and the suite creation is a complex

task, since the tests should not only verify propinquity of pseudorandom sequence to true random sequence, but also perform such task effectively. Unfortunately, a compromise is achieved quite hard in this case and modern suites also have statistical dependencies, which unreasonably increase such suites operating time. Thus, one of the main question of tests suite construction and the statistical tests using is statistical independence of tests from this suite. To create an effective suite, tests without statistical dependencies should be used and, at the same time, the tests set should remain sufficiently complete. However, modern questions of forming texts suite, its number determination, type I error value, etc. are solved intuitively and empirically. This article provides the existing evaluation methods overview of statistical randomness tests independence verification and proposes a new, mathematically grounded method, which can be applied to arbitrary tests number and arbitrary random sequences number. The proposed method has advantages in speed and implementation. The paper also presents the experimental research results of the new method application to the statistical randomness test suite.

Keywords: pseudorandom sequences generator, pseudorandom sequence, pseudorandom sequence quality, statistic tests of pseudorandom sequence quality assessment, test independence.

Людмила Василівна Ковальчук, доктор технічних наук, професор, професор кафедри математичних методів захисту інформації, Фізико-технічний інститут Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна.

E-mail: lusi.kovalchuk@gmail.com

Наталія Вікторівна Кучинська, кандидат технічних наук, доцент кафедри інформаційної безпеки, Фізико-технічний інститут Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна.

E-mail: n.kuchinska@gmail.com.

Людмила Васильевна Ковальчук, доктор технических наук, професор, профессор кафедры математических методов защиты информации, Физико-технический институт Национального технического университета Украины "Киевский политехнический институт имени Игоря Сикорского", Киев, Украина.

Наталия Викторовна Кучинская, кандидат технических наук, доцент кафедры информационной безопасности, Физико-технический институт Национального технического университета Украины "Киевский политехнический институт имени Игоря Сикорского", Киев, Украина.

Liudmyla Kovalchuk, doctor of technical sciences, professor, department Mathematical Methods of Information Security professor, Institute of Physics and Technology National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute", Kyiv, Ukraine.

Nataliia Kuchynska, candidate of technical sciences, department of information security associate professor, Institute of Physics and Technology National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute", Kyiv, Ukraine.