Уменьшение количества арифметических операций также можно достичь путем переупорядочения исходной системы [7, 8]. Этот вопрос требует дополнительных исследований и в настоящей работе рассматриваться не будет.

Список литературы

1. Хорошко В.А., Чекатков А.А. Методы и средства защиты информации. – К.: Юниор, 2003. - 501 с.
2. Бахвалов Н.С. Численные методы. – М.: Наука, 1975. – 632 с.
3. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Том 1.- М.: Наука, 1969. – 608 с.
4. Деммель Дж. Вычислительная линейная алгебра. – М.: Мир, 2001. – 430 с.
5. Ланкастер П. Теория матриц. – М.: Наука, 1978.- 280 с.
6. N.J.Higham. A survey of condition number estimation for triangular matrices.// SIAM Rev., № 29, 1987.- P. 575-596.
7. Писсанецки С. Технология разреженных матриц. – М.: Мир, 1988. – 411 с.
8. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. – М.: Мир, 1984.- 333 с.

УДК 691.321

Л.В.Ковальчук, В.Т.Бездітний

ПЕРЕВІРКА НЕЗАЛЕЖНОСТІ СТАТИСТИЧНИХ ТЕСТІВ, ПРИЗНАЧЕНИХ ДЛЯ ОЦІНКИ КРИПТОГРАФІЧНИХ ЯКОСТЕЙ ГВП.

Однією з найактуальніших задач в галузі захисту інформації є задача створення генератора випадкових (або псевдовипадкових) послідовностей (ГВП або ГПВП), що має певні криптографічні якості. Цьому питанню присвячено багато робіт, зокрема, [1-5]. Незалежно від того, за якими принципами та алгоритмами побудовано ГВЧ, оцінка його криптографічних якостей неможлива без статистичної оцінки послідовностей, що він виробляє. Для оцінки якостей послідовностей використовуються певні набори статистичних тестів (див., наприклад, [1, 6-9]). Основними вимогами до набору тестів є достатність цього набору та незалежність його складових. Питання достатності набору тестів тісно пов'язане як з умовами застосування набору, так і з можливою сферою застосування генератора і у даній роботі не розглядається. Задача перевірки незалежності статистичних тестів є актуальною не тільки при оцінці якостей генератора, але й при вирішенні багатьох інших прикладних криптографічних задач, які пов'язані з аналізом та синтезом криптографічних систем. В першу чергу це задачі оцінки *одноразових блокнотів*[1], *криптографічних алгоритмів*[2] та інших параметрів криптосистем.

Необхідність перевірки незалежності статистичних тестів, що використовуються при розв'язанні зазначених задач, полягає у наступному.

По-перше, це необхідність мінімізації кількості статистичних тестів. Набір тестів повинен бути "оптимальним" у тому розумінні, що він повинен містити суттєві тести і не містити тестів, які б повторювали інші. Наприклад, якщо відомо, що послідовність, яка пройшла тест А, з великою ймовірністю пройде тест В, то тест В можна виключити з набору для зменшення часу тестування

По друге, це необхідність оцінки загальної помилки першого роду при виборі набору тестів, яка можлива лише за умови незалежності тестів.

---

[1] В даному випадку, оцінка стійкості криптографічних перетворень зводиться до аналізу процедури генерації ключів, а фактично - фізичних датчиків випадкових чисел.

[2] Оцінці підлягають статистичні властивості виходу криптографічних перетворень, а саме функції виходу автоматної моделі криптоалгоритму.

Означення незалежності тестів

Інтуїтивно незалежність тестів може обґрунтовуватись тим, що різні тести перевіряють наявність різних властивостей у випадкової послідовності.

Незалежність тестів в методиці тестування NIST 800-22, „A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications" ([10]) перевіряється наступним чином. Будується $n \times q$ матриця з p-величин, отриманих в результаті тестування n послідовностей q тестами. До елементів матриці застосовується факторний аналіз. При цьому не зазначається, які саме фактори аналізуються, та не наводиться обґрунтування даної методики. Крім того, зрозуміло, що дана процедура є досить трудомісткою.

Тому виникає питання про імовірнісну (статистичну) незалежність тестів. Для вирішення цього питання спочатку треба дати означення незалежності для двох тестів або деякої їх сукупності.

Якщо розглядати статистики, які обчислюються в тестах, як випадкові величини, то вони, взагалі кажучи, не будуть статистично незалежними у сукупності. Розглянемо тести з набору NIST. Нехай $\{\varepsilon_i\}_{i=1}^{n}$ – випадкова послідовність, яку виробляє ГВЧ, $n=N \cdot M$, де N, M – параметри тесту частот в середині блоків. Статистика монобітного частотного тесту (Frequency Test) $\zeta_1 = \dfrac{\left|2 \cdot \sum_{i=1}^{n} \varepsilon_i - n\right|}{\sqrt{n}}$, тесту частот в середині блоків (Frequency Test within a Block) $\zeta_2 = \dfrac{1}{M} \cdot \sum_{i=1}^{N} \left(2 \cdot \sum_{j=1}^{M} \varepsilon_{(i-1)M+j} - M \right)^2$ (див. [8]). Вочевидь, $P\left(\zeta_1 = \dfrac{|n-2|}{\sqrt{n}}, \zeta_2 = n \right) = 0$ тому, що $\zeta_1 = \dfrac{|n-2|}{\sqrt{n}}$, коли послідовність $\{\varepsilon_i\}_{i=1}^{n}$ або містить рівно одну одиницю, або містить рівно один нуль, а $\zeta_2 = n = N \cdot M$, коли послідовність складається або з усіх одиниць, або з усіх нулів. В той же час,

$$P\left(\zeta_1 = \frac{|n-2|}{\sqrt{n}} \right) = \frac{n}{2^{n-1}}, \; P(\zeta_2 = n) = \frac{1}{2^{n-1}},$$

$$P\left(\zeta_1 = \frac{|n-2|}{\sqrt{n}}, \zeta_2 = n \right) \neq P\left(\zeta_1 = \frac{|n-2|}{\sqrt{n}} \right) \cdot P(\zeta_2 = n).$$

Тому випадкові величини ζ_1 та ζ_2 статистично залежні, отже, набір статистик тестів NIST не є набором незалежних у сукупності випадкових величин.

Аналогічна ситуація виникає для багатьох інших тестів, тому було б нераціонально визначати незалежність статистичних тестів через незалежність випадкових величин, що відповідають їх статистикам.

Нехай A і B – деякі тести. Випадкові величини ξ_A та ξ_B – індикатори проходження послідовністю тестів A та B відповідно, тобто $\xi_A = \mathrm{I}\left\{ \begin{matrix} \textit{послідовність} \\ \textit{пройшла тест } A \end{matrix} \right\}$, $\xi_B = \mathrm{I}\left\{ \begin{matrix} \textit{послідовність} \\ \textit{пройшла тест } B \end{matrix} \right\}$.

Означення. Тести А та В назвемо незалежними, 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 статистик.

Аналогічно можна означити набір незалежних тестів.

Означення. Тести з набору $\{A_i\}$ назвемо незалежними, якщо незалежні у сукупності відповідні індикатори.

Позначимо через ζ_A, ζ_B статистики, які обчислюються в тестах А і В, K_A, K_B – критичні області тестів, відповідно.

Тоді розподіл величин ξ_A, ξ_B наступний:

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

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.

Оскільки не є можливим строго довести залежність або незалежність ξ_A та ξ_B як випадкових величин, гіпотезу про їх незалежність слід перевіряти статистичними методами.

Допоміжні результати

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 ГВЧ, 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 [10]. Залежність чи незалежність тестів не впливають на якість оцінки ГВЧ, вони впливають лише на час проведення тестування.

Нехай отримано n послідовностей з ГВЧ (довжини послідовностей повинні бути придатними для проведення тестування).

Теорема:

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

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

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.

Доведення:

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

$$P(\eta_i^{(j)}=0)=\alpha_i,\ P(\eta_i^{(j)}=1)=1-\alpha_i.$$

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, ГВЧ має необхідні статистичні якості, тоді згенеровані 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 Центральну граничну теорему ([11]), для досить великих значень n одержуємо:

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

Теорему доведено

Методика перевірки незалежності тестів, що застосовуються для оцінки якостей ГВЧ

Нехай для оцінки якостей ГВЧ 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. Використовуючи n послідовностей, отриманих з ГВЧ, обчислити наступні величини: $a=p=(1-\alpha)^N$, $q=1-(1-\alpha)^N$, $\sigma^2=\frac{p\cdot q}{n}$.

3. Відповідно до попереднього пункту, обчислити величини 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.

Зауваження.

1) Ймовірність "відбраковки" набору з незалежних тестів не перевищує β.

2) Процедури, описані у п.п. 5 та 6, можна замінити на наступні:

5'. Визначити константу C з умови: $\Phi(C) = 1 - \alpha/2$, 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.

Приклади застосування методики перевірки незалежності тестів.

Наведена у попередньому пункті методика була застосована до різних піднаборів NIST ([10]). Слід зазначити наступне. Набір NIST складається з 16 основних тестів, деякі з них, в свою чергу, розгалужуються на підтести (тести шаблонів, накопичених сум та інші), у результаті чого загальна кількість різних тестів дорівнює 161. Використовувався псевдовипадковий генератор, описаний у додатку А до ДСТУ 4145-2002.

Для перевірки незалежності було використано 500 послідовностей довжиною 125000 біт кожна.

Приклад 1.

n=43 (для тестів шаблонів роздалися лише шаблони довжиною 2; для тесту m-грам розглядалися лише m-грами довжиною 2). Рівні значимості тестів обирались однаковими: α=0,01. Рівень значимості для тесту перевірки незалежності обрано β=0,01.

За цими даними згідно до методики було обчислено наступні параметри:

a=0,6491026284, σ=0,02134330838,

Δ=(0,5941259092; 0,7040793476).

За результатами тестування вказаних послідовностей отримано значення η=0,506. Згідно нашої методики, тести з набору NIST слід вважати залежним. Але слід зауважити, що статистика відрізняється від нижньої границі інтервалу лише у другому знаку, тому надлишковість набору незначна. Крім того, можливо, слід обрати більш слабкі вимоги до незалежності. Тобто розширити границі інтервалу Δ, поклавши β=0,001.

Приклад 2.

n=17 (без варіаційного тесту випадкових блукань та тесту випадкових блукань, але з двома тестами частот неперіодичних двобітових шаблонів, які не перетинаються, з двома тестами частот біграм, що перетинаються, з двома тестами накопичених сум).

Рівні значимості тестів обирались однаковими: α=0,01. Рівень значимості для тесту перевірки незалежності обрано β=0,01.

За цими даними згідно до методики було обчислено наступні параметри:

a=0,8429431964, σ=0,01627205986,

Δ=(0,8010291448; 0,884857242).

За результатами тестування вказаних послідовностей отримано значення η=0,87. Згідно нашої методики, тести з даного піднабору слід вважати незалежним.

Приклад 3.

n=4 (монобітний, частоти в блоках, серій, серій максимальної довжини).

Рівні значимості тестів обирались однаковими: α=0,01. Рівень значимості для тесту перевірки незалежності обрано β=0,01.

За цими даними згідно до методики було обчислено наступні параметри:

a=0,96059601, σ=0,00870000725897,
Δ=(0,9381842553; 0,9830075947).

За результатами тестування вказаних послідовностей отримано значення η=0,948. Згідно нашої методики, тести з даного піднабору слід вважати незалежним.

Зрозуміло, що вибір незалежних тестів для перевірки необхідних статистичних гіпотез є складною прикладною задачею, яка в деяких випадках може не вирішуватись однозначно. Запропонований метод дозволяє шляхом практичних експериментів відповісти на досить широкий спектр питань, що лежать у сфері застосування статистичних критеріїв.

Список літератури

1. Knuth D.E. The Art of Computer Programming, Addison-Wesley, Reading, Massachussetts, 2nd edition, 1973.

2. Blum L., Blum M., Shub M. "A simple unpredictable pseudo-random number generator" // SIAM Journal on Computing, 15, 1986, p. 364-383.

3. Marsaglia W. "A high performance encryption algorithm", Computer Sequritu: F Global Chfllenge, Proc. of the Second IFIP International Conference on Computer Sequrity, p. 557-570, North-Holland, 1984.

4. B.S. Kaliski Jr. "A pseudo-random bit generation base on elliptic logarithms" // Advances in Cryptology – CRYPTO'86 (LNCS 263), p. 84-103, 1987.

5. Ковальчук Л.В. О периодах выходных последовательностей некоторых псевдослучайных генераторов, построенных на основе односторонних функций. // "Безопасность информации в информационно-телекоммуникационных системах", III Международная научно-практическая конференция, Киев, апрель, 2000 г., с. 56-59.

6. Гулак Г.Н., Ковальчук Л.В. Різні підходи до визначення випадкових послідовностей. // НТЗ Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. – вип. 3, Київ, 2001, с. 127-133.

7. Beker H., Piper F. Cipher systems: The Protection of Communications, John Wiley&Sons, New York, 1998. 8. Marsaglia G. "A current view of random number generation" // L. Billard, editor, Computer Science fnd Statistics: Proceedings of the Sixteen Symposium on the Interface, p. 3-10, North-Holland, 1985.

8. Kimberley M. "Comparison of two statistical tests for keystream sequences" // Electronic letters, 23 (April 9, 1987), p. 365-366.

9. „A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", NIST 800-221.

10. Ширяев А.Н. Вероятность. – М., «Наука», - 1989-640.