УДК 004.056; 004.415.24

Буценко Ю.П., к.ф.-м.н.; **Савченко Ю.Г.**, д.т.н.

ОЦІНЮВАННЯ РІВНЯ «ВИПАДКОВОСТІ» ПОСЛІДОВНОСТІ ПСЕВДОВИПАДКОВИХ БІНАРНИХ ЧИСЕЛ

Butsenko Y.P., Savchenko Y.G. Evaluation of "random" pseudorandom binary sequence of numbers.

The problem of numerical estimation of proximity of pseudorandom sequence to really random sequence is investigated. It is well known an actuality of such a problem, then, in practice, a wide range of tests is used. This tests, predominately, have as a mathematical fundament the theory of independent random trials. If we have a simplest case (p equls q). For n greater, then 100, the using of Gaussian approach and, subsequently, appropriate statistical methods is possible and obvious. At this paper the general approach is connected with an 'entropy' notion.

At proposed entropic approach, main instruments -statistical entropies for different length fragments. The level of dependence of this entropies from length of fragments is used for graphic illustration of proximity of investigated sequence to really random. Authors believe ,that the proposed method is the similar to the compression test of NISTs test suite.

Keywords: pseudorandom sequence, generator, entropy, testing

Буценко Ю.П., Савченко Ю.Г. Оцінювання рівня «випадковості» послідовності псевдовипадкових бінарних чисел.

Розглянуто задачу кількісної оцінки рівня наближеності послідовності псевдовипадкових чисел до істинно випадкової послідовності. Як основний інструмент пропонується застосувати обчислення статистичних ентропій для окремих фрагментів послідовності. Графічно залежність значень ентропій від довжини фрагментів наочно ілюструє наближеність досліджуваної послідовності до істинно випадкової. Пропонований підхід близький до тестування на стиснення в наборі тестів NIST.

Ключові слова: псевдовипадкова послідовність, ентропія, генератор, тестування

Буценко Ю.П., Савченко Ю.Г. Оценка уровня «случайности» последовательности псевдослучайных бинарных чисел.

Рассматривается задача количественной оценки степени приближения последовательности псевдослучайных бинарных чисел к истинно случайной последовательности. В качестве основного инструмента предлагается использовать энтропийный подход и вычисление соответствующих значений энтропии для фрагментов последовательности различной длины. Полученная в результате зависимость энтропии на фрагмент от длины фрагмента в графическом виде иллюстрирует степень приближения тестируемой последовательности к истинно случайной. Проведено сравнение энтропийного подхода с использованием популярных тестовNIST. Показано, что наиболее близким к энтропийному подходу из тестов NIST является тест на сжатие последовательности.

Предлагаемый подход может быть использован для тестирования генераторов последовательностей псевдослучайных чисел.

Ключевые слова: псевдослучайная последовательность, энтропия, генератор, тестирование

Вступ

Актуальність оцінки якості послідовності псевдовипадкових бінарних чисел (ПВБЧ) визначається перед усім використанням таких послідовностей як ключів при шифруванні та паролів доступу до інформаційних ресурсів і послуг. Сьогодні, коли для захисту інформаційного обміну використовуються виключно стандартні шифри (AES, DES, RSA та небагато інших), алгоритми шифрування для яких загальновідомі, саме можливість «підібрати ключі» стає вирішальним фактором безпеки. Мова йде про нейтралізацію хакерських атак та інших спроб несанкціонованого доступу до конференційної інформації. В той же час саме поняття якості зазначених об'єктів є дискусійним і не має сьогодні формального визначення. На неформальному рівні випадковість деякої послідовності чисел (не обов'язково бінарних) визначається можливістю (неможливістю) передбачити, яке число буде наступним у послідовності на основі знання попередніх чисел. Це питання трохи в інших термінах може бути поставлене так: чи є якась закономірність у розташуванні чисел у заданій послідовності? Для справжньої випадкової (істинно випадкової) послідовності такої закономірності не повинно існувати за визначенням.

Виклад основного матеріалу дослідження

При розв'язанні задачі оцінювання рівня «випадковості» послідовності ПВБЧ слід розрізняти два принципово різні випадки.

1.Отримання оцінки умовної якості ***однієї*** послідовності – дати оцінку (бажано чисельну) рівня наближеності заданої послідовності до істинно випадкової.

2.Оцінювання деякої скінченої множини послідовностей, отриманих від генератора (програмного або апаратного) послідовностей ПВБЧ. Необхідно оцінити усереднену якість генерованих послідовностей, тобто, по суті, якість генератора як джерела, наприклад, ключів для шифрування.

Перший випадок вже за самою постановкою задачі є досить проблематичним з точки зору коректності. Справа в тому, що будь-яка ***одна*** послідовність чисел сама по собі є випадковою, якщо не відома множина, з якої вона вибрана. Тобто випадковість (якість) однієї послідовності можливо оцінити лише на основі її внутрішніх властивостей. І тут виникає питання: яких саме властивостей? Найбільш відомим тестом з точки зору внутрішніх властивостей є, мабуть, перевірка на визначення значення наступного в послідовності біта на основі всіх попередніх значень. Якщо визначити це значення з ймовірністю, відмінною від ½, неможливо, то така послідовність є випадковою. Це строго доведено ще у 1983 році Андре Яо [1]. Але, на жаль, до цього часу не запропоновано конструктивної процедури проведення такої перевірки.

В узагальненому вигляді оцінювання рівня випадковості заданої послідовності є задачею виявлення внутрішніх закономірностей, притаманних їй, зокрема кореляційних зав'язків між значеннями окремих бітів або груп бітів у послідовності. Такі закономірності повинні, як слід очікувати, обов'язково проявити себе вже на статистичному рівні, тобто у вигляді нерівномірності появи окремих груп бітів у послідовності.

Для прикладу розглянемо числову послідовність натурального ряду $1,2,\ldots,N$. Зрозуміло, що вона не є випадковою, а кореляційний зв'язок між елементами послідовності можна побачити на статистичному рівні, якщо розглядати відносні частоти появи *сусідніх пар* чисел у послідовності. Аналогічно для послідовності

$$1,1,2,4,8,16,32,\ldots$$

легко виявити закономірність: кожен наступний елемент є сумою всіх попередніх елементів. У загальному випадку закономірності, притаманні конкретній послідовності, можуть бути досить складними та проявляються лише на статистичному рівні.

Уявімо, що проведено такий статистичний експеримент: досліджувана послідовність записується у деякий регістр і шляхом послідовних зсувів та лічильників фіксуються кількості бітових фрагментів різної довжини. Як результат отримаємо умовний частотний спектр досліджуваної послідовності у вигляді

$$f_0, f_1, f_{00}, f_{01}, f_{10}, f_{11}, f_{000}, f_{001},\ldots f_{111},\ldots, f_{11\ldots10}, f_{11\ldots11},$$

де індекс при символі f відповідає тому фрагменту послідовності, який з'явиться в умовному «вікні» при зсувах регістру. Отримані значення f є відображенням частот (або ймовірностей), з якими відповідні фрагменти з'являються в послідовності. Тоді можна обчислити статистичні ентропії повідомлень за умови, що вони начебто передаються як одно- двох-…. k -бітові слова

$$H_1 = -\left(\frac{f_0}{N}\log_2\frac{f_0}{N} + \frac{f_1}{N}\log_2\frac{f_1}{N}\right),$$

$$H_2 = -\frac{1}{2}\left(\frac{f_{00}}{N}\log_2\frac{f_{00}}{N} + \frac{f_{01}}{N}\log_2\frac{f_{01}}{N} + \frac{f_{10}}{N}\log_2\frac{f_{10}}{N} + \frac{f_{11}}{N}\log_2\frac{f_{11}}{N}\right),$$

.....

$$H_k = -\frac{1}{k}\sum_{00\ldots0}^{11\ldots1}\frac{f_{ij\ldots g}}{N}\log_2\frac{f_{ij\ldots g}}{N},$$

де k – довжина відповідного фрагменту.

У більшості реальних випадків, зокрема для генераторів послідовностей, що використовують регістри зсуву із зворотними зв'язками по модулю 2 функція $H_k = \varphi(k, N)$ є спадаючою та має такий характерний вигляд (рис. 1).

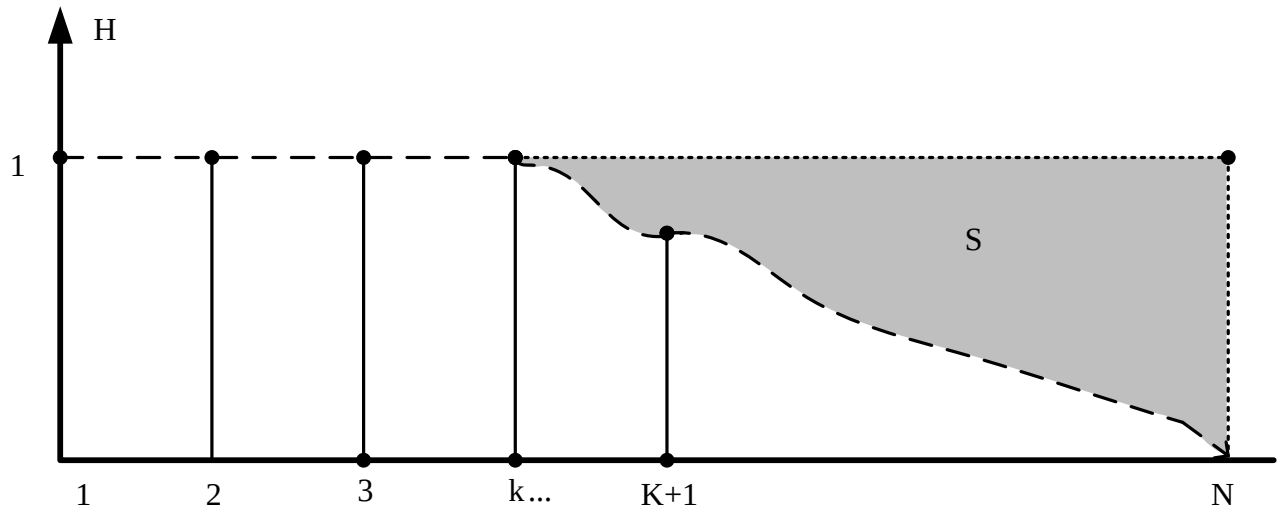

Рис. 1

Цей вигляд можна пояснити таким чином.

Генератори, побудовані на базі регістрів із зворотними зв'язками, використовують для утворення цих зв'язків коефіцієнти примітивних поліномів у полі GF2, що гарантує разом із іншими умовами генерацію послідовностей максимальної довжини, а це, в свою чергу, забезпечує появу в регістрі *всіх* можливих комбінацій розрядності k та всіх можливих їх фрагментів. Тому відповідні статистичні ентропії мають максимальне значення для всіх значень k, які менше за довжину регістру. Далі, після цієї точки відбувається природне зменшення ентропії за рахунок зменшення комбінаторного різноманіття можливих фрагментів (не всі фрагменти довжини, більшої за k може містити генерована послідовність). Мінімальне значення ентропії, очевидно, відповідає випадку, коли $k = N$.

У другому випадку, коли необхідно обчислити деяку усереднену оцінку рівня випадковості сукупності послідовностей, можна спочатку обчислити таку оцінку для кожної з послідовностей, а потім знайти їх середнє арифметичне, вважаючи, що вибираються вони випадковим чином з рівними ймовірностями. Можна очікувати, що загальний вигляд залежності ентропії від довжини фрагментів залишиться приблизно таким самим, як і для однієї послідовності.

Зазначимо, що в усіх відношеннях аналогічно генератору випадкових 0 та 1 є реалізація найпростішої схеми випробувань Бернулі ($p = q = 1/2$). Така схема давно та, можна сказати, вичерпно досліджена. Основними при цьому є формула Бернулі (для імовірності кількості «успіхів» у n випробуваннях), локальна та інтегральна формули Муавра-Лапласа як асимптотичні варіанти попередньої. Зазначимо, що на практиці рекомендується використовувати останні дві формули, починаючи з $n = 100$ (у нашому випадку це довжина послідовності ПВБЧ), причому при збільшенні n точність формул зростає. До того ж саме при $p = q = 1/2$ точність максимальна. Закладену у формулах можливість апроксимувати центровану та нормовану певним чином кількість 0 (або, що теж саме, 1) у послідовності за допомогою випадкової величини, розподіленої за стандартним нормальним законом використовують такі тести NIST [2], як частотний побітовий, частотний блочний, тест на однакові біти, що йдуть підряд, на найдовшу послідовність одиниць. Як і слід було очікувати, їх рекомендується використовувати при $n \geq 100$. Більш витонченим, можна сказати, є ранговий тест, що ґрунтується на роботах І.М. Коваленка [3]. По суті, він еквівалентний дослідженню, чи не є група m послідовних символів із n лінійною комбінацією попередніх аналогічних фрагментів довжини m

Близьким до рангового є спектральний тест, що виявляє повторюваність фрагментів у послідовності. Тести на повторюваність шаблонів (різновиди – шаблони перетинаються або не перетинаються) виявляють можливість отримання блоків шляхом внутрішніх перестановок. До цих тестів, у свою чергу, близьким є тест на лінійну складність, який має на меті виявлення факту застосування для генерації послідовностей вже згаданих регістрів зсуву із зворотними зв'язками по модулю 2.

Значно ширшим за можливою сферою застосування є універсальний тест Мауера, що полягає у з'ясуванні можливості безвтратного стиснення послідовності. Тести кумулятивних сум та на довільні відхилення фактично ведуть до вивчення поведінки випадкових блукань по одновимірній цілочисельній решітці та спираються на відповідні граничні теореми. З точки зору пропонованого підходу особливу увагу до себе звертають тести на підпослідовності та наближену ентропію.

У загальному та практичному плані найбільш привабливим серед відомих тестів можна вважати тести на стиснення. Чим в більшій мірі може бути стиснена тестована послідовність, тим гірші її показники «випадковості». Їх аналогами з формальних тестів є тести ентропії, які спираються на той факт, що функція $E(x_1,...,x_k)=-\frac{1}{k}\sum_1^n x_k \log_2 x_k$ має за наявності рівняння зв'язку $x_1+x_2+...+x_n=1$ єдиний максимум, рівний 1. Якщо $x_1, x_2,..., x_n$ – частоти деякого повного набору можливих бінарних комбінацій у послідовності (зазвичай $n=2^k$), то розглядаються всі можливі бінарні фрагменти, то $H_k=1$ при $x_1=x_2=...=x_k=\frac{1}{2^k}$ (ідеальний випадок). В той же час ці частоти при $N\gg 1$ внаслідок існуючих граничних теорем дуже близькі до своїх теоретичних значень для майже всіх реалізацій випадкової послідовності та менші 1 для генерованих псевдовипадкових послідовностей при однакових значеннях N.

Спираючись на припущення, зроблене А.М. Колмогоровим ще в 1965 р. у класичній роботі [4], можна обережно стверджувати, що складність опису процедури генерації (формування) послідовності ПВБЧ є показником наближеності її до істинно випадкової. Можна, мабуть, сказати, що за простим алгоритмом генерації не можна сформувати «якісну» послідовність ПВБЧ. Тому виникає незвична та парадоксальна ситуація: чим складніший алгоритм генерації, тим краще! Правда, зразу ж виникає проблема: а чим вимірювати складність? Із цієї точки зору популярні алгоритми генерації, що базуються на використанні регістрів зсуву, є занадто простими і тому майже не захищені від так званих алгебраїчних атак [5]. При використанні більш складних моделей цього недоліку можна у багатьох випадках уникнути, наприклад, застосовуючи моделі цифрових автоматів із пам'яттю [6].

Насправді, якщо замислитись, що таке випадкова послідовність чисел? Мабуть, така, яку ми не взмозі передбачити, тобто **будь-яка**. А чи можливо **будь-яку** послідовність якось описати іншим способом, ніж переліком її елементів у послідовності? Хто може стверджувати, що здатен запропонувати такий спосіб? Якщо існує такий спосіб, в ентропійному вимірі це відповідає можливості стиснення первинного опису. Тобто відносна площа області «сірої зони» S на рисунку (наприклад, у відсотках до загальної площі прямокутника) і є тією кількісною мірою відхилення досліджуваної послідовності від істинно випадкової. Із такої точки зору будь-яка послідовність, сформована за допомогою детермінованої процедури принципово не може бути істинно випадковою. Але важко сподіватись, що такі загальнотеоретичні питання можуть бути вирішені у короткому повідомленні. Тому на завершення зупинимось на практичних аспектах задачі.

Оцінювання рівня випадковості однієї окремо взятої послідовності з практичної точки зору позбавлено, на наш погляд, сенсу, оскільки не відомо, як можна реально використати результат такого оцінювання. Інша справа, коли мова йде про результати, отримані для деякої сукупності послідовностей, сформованих реальним генератором (програмним чи апаратним). У цьому випадку кількісна оцінка може стати базою для порівняння різних алгоритмів генерації та відбору кращого з них. Окрім того, конкретний вигляд залежності $H_k=\varphi(k,N)$ надає корисну інформацію для криптоаналізу. Зокрема, значення k, при якому починається спадання залежності, у більшості випадків однозначно співпадає з довжиною

регістру зсуву при регістровій реалізації генератора або кількості елементів пам'яті при використанні для цього автоматних моделей.

Висновок

На завершення слід зауважити, що реалізація пропонованого, а також параметричного [7] підходів при реальних значеннях довжини ключів ($N = 64...256$) може бути пов'язана із обмеженнями витрат на час реалізації статистичного експерименту із досліджуваною послідовністю. Тому для частини залежності $H_k = \varphi(k, N)$ доведеться скористатися її, наприклад, лінійною апроксимацією. Але найбільш важливим у запропонованому підході, на нашу думку, є можливість отримання саме кількісної оцінки рівня випадковості послідовностей ПВБЧ та порівняння різних алгоритмів генерації.

Список використаної літератури

1. Danny Dolev, Andrew Chi-Chih Yao On the security of public key protocols / Danny Dolev, Andrew Chi-Chih Yao // - IEEE Trans. Information Theory. - 1983. - №29. - P. 198-207.

2. Потий А., Орлова С. Статистическое тестирование генераторов случайных и псевдослучайных чисел с использованием набора статистических тестов NIST STS / А. Потий, С. Орлова // Правове, нормативне та метрологічне забезпечення систем захисту інформації в Україні. - 2001. - №2. - С. 206-214

3. Коваленко И.Г. Теория вероятностей и ее применение / И.Г. Коваленко // М.: Физматлит, 1972. - 435 с.

4. Колмогоров А.Н. Три подхода к определении понятия «количество» информации / А.Н. Колмогоров // Проблемы передачи информации. - 1965. - Т.1. - С. 3-11

5. Пометун С.О. Алгебраїчні атаки на потокові шифратори як узагальнення кореляційних атак / С.О. Пометун // Системні дослідження та інформаційні технології. - 2008. - №2. - С. 29-40

6. Буценко Ю.П., Савченко Ю.Г. Автоматні моделі генераторів псевдовипадкових бінарних чисел / Ю.П. Буценко, Ю.Г. Савченко // Сучасний захист інформації. - 2016. - №1.- С. 32-37

7. Буценко Ю.П., Савченко Ю.Г. Тестування бінарних послідовностей: параметричний підхід / Ю.П. Буценко, Ю.Г. Савченко // П'ятнадцята міжнародна наукова конференція імені академіка Михайла Кравчука, Матеріали конференції, ч. 3 – Теорія ймовірностей та математична статистика. - 2013. - С. 30-32

Автори статті

Буценко Юрій Павлович - кандидат фізико-математичних наук, доцент, доцент кафедри математичного анализу та теорії ймовірностей, Національний технічний університет України «КПІ» ім. І. Сікорського, Київ, Україна. Тел. +38 050 207 34 42. E-mail: armchairdoc@yandex.ua

Савченко Юлій Григорович - доктор технічних наук, професор, професор кафедри звукотехніки та регістрації інформації, Національний технічний університет України «КПІ» ім. І. Сікорського, Київ, Україна. Тел. +38 095 838 97 69. E-mail: ssaavvaa@ukr.net

Authors of the article

Butsenko Yurij Pavlovich – Candidate of Sciences (physics and mathematics), associate professor, associate professor of Department probability theory and mathematical analysis, National Technical University of Ukraine "Igor Sikorsky KPI", Kyiv, Ukraine. Tel. +38 050 207 34 42. E-mail: armchairdoc@yandex.ua

Savchenko Yulij Grigorievich – Doctor of Sciences(technical), professor, professor of Department Audio Engineering and Information Registration, National Technical University of Ukraine "Igor Sikorsky KPI", Kyiv, Ukraine. Tel. +38 095 276 92 70. E-mail: ssaavvaa@ukr.net

Дата надходження в редакцію: 30.01.2017 р. Рецензент: д.т.н., проф. А.І. Семенко