УДК 004.056.57:656.2

О.А. Нємкова[1]

[1]Університет банківської справи (м. Київ) Національного банку України Львівський інститут банківської справи, доцент кафедри економічної кібернетики

ЗАСТОСУВАННЯ RS–АНАЛІЗУ ДЛЯ ПЕРЕВІРКИ ЯКОСТІ ГЕНЕРАТОРІВ ПСЕВДОВИПАДКОВИХ ПОСЛІДОВНОСТЕЙ

У роботі застосовано RS-аналіз до деяких генераторів псевдовипадкових послідовностей та розраховані значення коефіцієнту Хьорста. Встановлена відповідність результатів аналізу до статистичних властивостей генераторів. Запропоновано використовувати RS-аналіз для тестування генераторів псевдовипадкових послідовностей на наявність персистентності, іншими словами, перевіряти генератори на придатність для застосування у криптографії.

Ключові слова: генератор псевдовипадкової послідовності, коефіцієнт Хьорста, потокове шифрування, лінійний конгруентний генератор, персистентність.

Вступ

Відкритість сучасних соціально-економічних інформаційних систем, за допомогою яких відбувається обробка, збереження та передавання конфіденційної або таємної інформації, потребує застосування криптографічних перетворень масивів даних. Обсяг таких даних може бути різним – невеликим для інформаційного обміну між платіжною системою та індивідуальним користувачем і дуже великим при передачі звукових, тим більш, відео файлів. Якість криптографічного перетворення повинна бути дуже високою, тому що переважно це стосується передачі та збереження банківської інформації, конфіденційних баз даних мобільних операторів, медичних та фармацевтичних компаній, військових розробок та інших даних, що пов'язані з державною таємницею. Все це вимагає в ідеалі необмежених по довжині псевдовипадкових послідовностей. Зростаюча кількість кібератак за останні роки підтверджує висновок про необхідність надійного захисту, чого можна досягти лише за допомогою шифрування даних.

Останнім часом можна почути пропозиції шифрувати значну кількість інформації, яка не тільки передається назовні з локальної корпоративної мережі, але і зберігається на жорстких дисках в системі. Особливо це стосується баз даних конфіденційної інформації. На сьогодні таке шифрування не застосовується, тому інформація, що міститься в багатьох базах даних, може бути прочитана спеціально написаними для цього програмами – павуками. У продажі є такі програмні комплекси (один з них пропонує російська фірма Аналитические бизнес решения), які мають відповідно написані павуки для різноманітних баз даних. Такі програмні комплекси позиціонуються на ринку як такі, що збирають економічну інформацію про конкретну людину. Але зрозуміло, що з їх допомогою можна збирати будь-яку конфіденційну інформацію.

В деяких випадках потрібне швидке криптографічне закриття інформації. Зазвичай для цього використовують потокове шифрування, яке базується на генерації високоякісних псевдовипадкових послідовностей. Перевагами потокового шифрування є його відносна простота, швидкість та відсутність розмноження помилок. Процес шифрування полягає у генерації гами та подальшого накладання її на потік даних. Стійкість шифрів,

утворених за допомогою гами, суттєво залежать від її стохастичних властивостей, а також від довжини періоду гами. Слід зазначити, що в залежності від способу накладання, гама генерується або у вигляді двійкових послідовностей нулів та одиниць, або як білий шум з заданою амплітудою. Перший спосіб використовується при подальшому застосуванні операції прямої суми шифрованого повідомлення з гамою.

Важливість тематики обумовила побудову достатньо великої кількості генераторів псевдовипадкових послідовностей. Генерація псевдовипадкових послідовностей відбувається, як правило, алгоритмічно, за певними правилами. Такі послідовності мають більшу чи меншу довжину, або період, після якої вони починають повторюватись. До того ж такі послідовності можуть виявитись криптографічно нестійкими. Тому при створенні чергового псевдовипадкового генератора важливо довести, що він генерує послідовність, яка наближається до випадкової.

Тестування псевдовипадкових послідовностей

На сьогоднішній день не існує єдиного універсального та перевіреного на практиці критерію або методики для визначення якості гами та її придатності для шифрування. Для непередбачуваності гами прийнято вважати, щоб її період був набагато більшим за довжину послідовності даних, що шифруються, а різноманітні комбінації бітів визначеної довжини були рівномірно розповсюджені по всій її довжині. В основному користуються статистичними методами перевірки якості псевдовипадкової послідовності, хоча є також графічні методи (гістограма розподілу елементів послідовності, перевірка серій, перевірка на монотонність, автокореляційна функція, графічний спектральний тест та ін.).В останніх оцінку дає людина на основі суб'єктивного сприйняття результату, тому такі методи не знаходять широкого застосування.

Зі статистичних методів найбільш відомими є наступні:

- тести diehard (міцний горішок). Складаються з набору 12 різнопланових тестів для дробових псевдовипадкових чисел. Суттєва частина даних тестів використовує геометричні залежності;
- тести Кнута. Складаються з 7 тестів і базуються на статистичному критерії хі квадрат;
- статистичні тести NIST. Містять не менше 15 різноманітних тестів, мета яких у визначенні міри випадковості двійкових послідовностей.

Існують ще декілька статистичних тестів, але вони менш відомі.

Характерною особливістю перелічених тестів є суттєва кількість дій, які треба виконати, щоб переконатись в валідності досліджуваної псевдовипадкової послідовності. Звичайно, потрібний деякий час для виконання перевірки. Велика кількість тестів та їх різноманіття наводять на думку, що було б зручніше користуватись меншим числом тестів, в ідеалі взагалі одним, який би давав відповідь на питання можливості застосування псевдовипадкової послідовності в криптографії.

Спосіб перевірки послідовностей, точніше часових рядів, на випадковість був запропонований достатньо давно, майже сто років тому. Галузь знань, у якій він вперше був випробуваний, була далека від криптографії. Спосіб носить назву RS-аналіз і застосовується в наш час переважно для аналізу фінансових часових рядів [1], хоча об'єктом аналізу може бути будь-яке явище природи та суспільства. Найважливішою особливістю RS-аналізу є те, що наперед не ставиться ніяких обмежень стосовно закону розподілу ряду, який досліджується. Власне, в результаті аналізу можна стверджувати, наскільки ряд близькій до чисто випадкового. Кількісною ознакою цього слугує показник Хьорста. Для чисто випадкового процесу показник Хьорста дорівнює ½. Цікаво, що переважна більшість природніх явищ, а також величин часових процесів фінансової природи характеризуються показником Хьорста, більшим за ½. Такі процеси отримали назву процесів з довгою пам'яттю, або персистентних. Система на наступному ході ніби то пам'ятає,

що було з нею на попередніх ходах і зберігає таку тенденцію. Іншими словами, якщо відхилення від рівноваги були значними на попередніх ходах, то і на наступному ході варто очікувати значного відхилення. RS-аналіз використовують для перевірки наявності у ряді даних довготривалої залежності.

У даній роботі RS-аналіз застосовується для дослідження якості генераторів псевдовипадкових послідовностей.

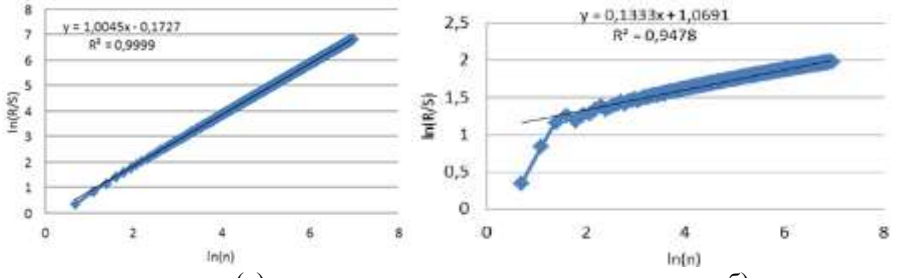
Методологія RS-аналізу

Нехай є послідовність a_n. Утворюється послідовність часткових сум A_n. Вираховуються наступні числові характеристики: сер.зн.A_n - середнє арифметичне елементів A_n, послідовність R_n – розмах накопичених сум (різниця максимального та мінімального значень часткових сум відхилення елементів a_n від середнього арифметичного A_n), послідовність S_n – середнє квадратичне відхилення an від сер.зн.A_n, послідовність $RS_n = R_n/S_n$. На площині будується множина точок $(x_n;y_n) = (\ln(n);\ln(RS_n))$. Надалі застосовується метод найменших квадратів для визначення кутового коефіцієнта тренду. Цей кутовий коефіцієнт називається коефіцієнтом Хьорста та позначається літерою H. Знання коефіцієнта Хьорста дозволяє отримати значення розмірності Мінковського d = 2 – H, [2].

Перед перевіркою гами, отриманої шляхом генерації, за методом RS-аналізу слід переконатись, що у граничних випадках, таких як лінійна або квадратична залежність, а також періодична залежність з коротким періодом виходить достовірний результат.

В результаті виконання чисельного експерименту для послідовності, значення елементів якої утворюють лінійну залежність від номеру, отримане значення коефіцієнта Хьорста дорівнює одиниці, так як і очікувалося, рис.1(а). Такий самий результат отримано для квадратичної залежності – коефіцієнт Хьорста дорівнює одиниці. Виходячи з основних положень даної теорії, слід очікувати такий самий результат у випадку монотонної визначеної залежності.

При дослідженні короткоперіодичної послідовності значення коефіцієнта Хьорста повинно складати малу величину, значно меншу за одиницю. Якщо скористатись аналогією з відхиленням рухомої частинки від початкової точки, то випадок лінійної залежності від номеру відповідає прямолінійному рівномірному руху, тому віддаль частинки прямо пропорційна часу. Ейнштейн використовував дану методику для виводу формули для розрахунку віддалення броунівської частинки від початку руху, його частинка рухалася хаотично – кожний рух після наступного удару не був пов'язаний з попереднім. Якщо ж частинка коливається з малою амплітудою навколо початкової точки, то її віддаль від місця початку руху не перевищує амплітуди коливань і фактично не зростає з часом. Розрахунки підтверджують цей висновок, рис.1(б).



(а) б)

Рис.1(а) значення H дорівнює 1,00 для лінійної залежності,(б) значення H дорівнює 0,13 для короткоперіодичної послідовності.

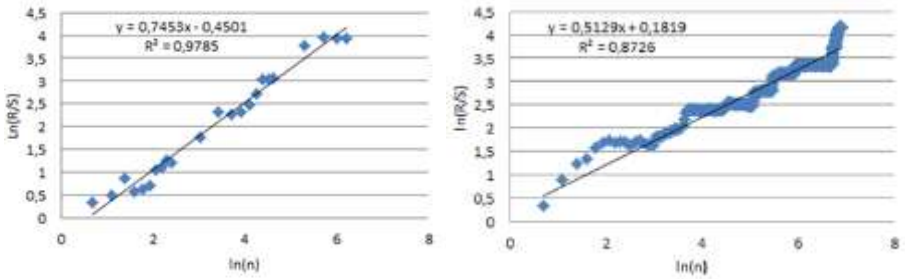
Таким чином, отримані результати для лінійної та короткоперіодичної залежності вкладаються в теорію.

Дослідження деяких псевдовипадкових послідовностей

В наш час існує достатня кількість генераторів псевдовипадкових послідовностей. Тим не менш отримати псевдовипадкову послідовність можна трьома шляхами: скористатись готовими таблицями, скористатись вбудованими в програми генераторами, або використати заданий алгоритм. В даній роботі перевірялись псевдовипадкові послідовності, взяті з трьох типів джерел.

На рис.2(а) представлені результати розрахунку коефіцієнта Хьорста для п'ятизначних випадкових чисел, взятих з таблиці 26.11 [3]. Хоча дані числа позиціонувались як випадкові, результати проведеного аналізу дають підставу стверджувати, що вони не є чисто випадковими, послідовність персистентна.

У якості другої послідовності було досліджено результат роботи генератора rnd(1). Цей генератор визначений в програмі MathCAD як такий, що генерує білий шум – випадкову послідовність з рівномірним розподілом на відрізку [0;1]. Результати перевірки представлені на рис.2(б). Можна стверджувати, що даний генератор (по критерію H) наближається до ідеального (кожне наступне число дуже слабо пов'язане з попереднім за критерієм Хьорста). Його варто використовувати для генерації псевдовипадкової послідовності в модельних експериментах. Такий процес для руху броунівських частинок говорить про те, що відстань, на яку віддаляється частинка з часом від початку руху, пропорційна квадратному кореню з часу (формула Ейнштейна).



(а) (б)

Рис.2. (а) значення H дорівнює 0,745 для послідовності, взятої з таблиць,(б) значення H дорівнює 0,513 для генератора rnd(1).

Третім досліджуваним генератором був лінійний конгруентний генератор. Як відомо, такі генератори не можуть бути використані у криптографії. Вперше лінійні конгруентні генератори були зламані Дж. Рідсом, а потім Дж. Бояр. Послідовність чисел може бути прорахована, для цього достатньо знання трьох сусідніх значень. З часом Дж. Бояр вдалося зламати квадратичні та кубічні генератори. Надалі за її ідеями були зламані будь-які поліноміальні генератори, тим самим доведена неможливість їх використання у криптографії. Тим не менш для задач математичного моделювання дані генератори широко використовуються. В свій час Національне бюро стандартів Сполучених Штатів рекомендувало таблицю констант, при використанні яких конгруентні генератори мають задовільні статистичні властивості та достатню довжину [4].

Для прикладу на рис.3 представлені результати розрахунку коефіцієнту Хьорста для лінійного конгруентного генератора.

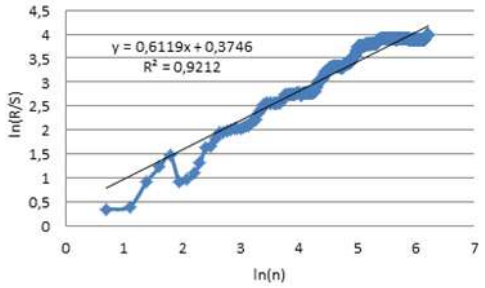


Рис.3. Значення H дорівнює 0,612 для лінійного конгруентного генератора.

Для перевірки поведінки лінійного конгруентного генератора було згенеровано декілька послідовностей для рекомендованих значень констант a, b, m та розраховано коефіцієнти Хьорста для кожної з них. Результати розрахунків наведено у таблиці 1.

Таблиця 1

Коефіцієнт Хьорста H для лінійного конгруентного генератора в залежності від довжини послідовності та констант a, b, m

a	b	m	Переповнення при	H
106	1283	6067	2^{20}	0,612
211	1663	7875	2^{21}	0,653
421	1663	7875	2^{22}	0,617
939	1399	6655	2^{23}	0,679
859	2531	11979	2^{24}	0,760
141	28411	134456	2^{25}	0,648
1255	6173	29282	2^{26}	0,537
1021	24631	116640	2^{27}	0,649
1277	24749	117128	2^{28}	0,684
2311	25367	120050	2^{29}	0,545
3613	45289	214326	2^{30}	0,563
8121	28411	134456	2^{31}	0,546
9301	49297	233280	2^{32}	0,626
2416	374441	1771875	2^{33}	0,578
17221	107839	510300	2^{34}	0,786
84589	45989	217728	2^{35}	0,577

Значення коефіцієнту Хьорста із таблиці 1 підтверджують висновок про наявність залежності в досліджуваних послідовностях. Відхилення від величини ½ не можна списати за рахунок точності обчислень. Відзначимо, що за допомогою застосованого аналізу можна підібрати значення констант, що дають значення коефіцієнта Хьорста ближчим до ½, дозволяючи будувати послідовності з непоганими властивостями для інших прикладних задач окрім криптографічного захисту.

Таким чином, результати аналізу підтверджують висновок про наявність залежності між елементами послідовності для лінійних конгруентних генераторів. Це дає підстави зробити наступний висновок: RS-аналіз дає змогу виявити нестохастичність генератора псевдовипадкових послідовностей.

Висновки

RS-аналіз застосовано до деяких генераторів псевдовипадкових послідовностей – більш детально досліджені лінійні конгруентні генератори. Для досліджуваних генераторів розраховані значення коефіцієнту Хьорста. Встановлена відповідність результатів аналізу до статистичних властивостей генераторів. Запропоновано використовувати RS-

аналіз для тестування генераторів псевдовипадкових послідовностей на стохастичність (відсутність персистентності). Також слід відзначити, що в вище перелічених статистичних тестах не застосовується коефіцієнт Хьорста.

Література

1. Петерс Э. Фрактальний анализ финансових рынков. Применение теории хаоса в инвестициях и экономике. – М.: Интернет-трейдинг, 2004. – 304 с.
2. Ширяев А.Н. Основы стохастической финансовой математики. – М.: Фазис, 1998. – Т.1. – 512 с.
3. Справочник по специальным функциям с формулами, графиками и математическими таблицами. Под редакцией М Абрамовица и И. Стиган. – М.: Наука, главная редакция физико-математической литературы, 1979. – 832 с.
4. Есин В.И., Кузнецов А.А., Сорока Л.С. Безопасность информационных систем и технологий. – Х.: ООО «ЭДЭНА», 2010. – 656 с.

Надійшла до редколегії 06.04.2013 р.

***Рецензент**:* Бондарєв А.П., доктор технічних наук, професор кафедри ТРР Національного університету «Львівська політехніка», м. Львів

А.А. Немкова

ПРИМЕНЕНИЕ RS-АНАЛИЗА ДЛЯ ПРОВЕРКИ КАЧЕСТВА ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

В работе применен RS-анализ в некоторых генераторов псевдослучайных последовательностей и рассчитаны значения коэффициента Херста. Установлено соответствие результатов анализа статистических свойств генераторов. Предложено использовать RS-анализ для тестирования генераторов псевдослучайных последовательностей на наличие персистентности, другими словами, проверять генераторы на пригодность для применения в криптографии.

Ключевые слова: генератор псевдослучайной последовательности, коэффициент Херста, потоковое шифрование, линейный конгруентный генератор, персистентность.

A.A. Nemcova

APPLICATION OF RS-ANALYSIS FOR QUALITY CONTROL OF GENERATORS PSEUDO-RANDOM SEQUENCE

The paper used RS-analysis of some of pseudorandom sequences and calculated values of the Hurst exponent. A correspondence analysis of statistical properties of the generators. Proposed to use the RS-analysis testing of pseudorandom sequences for the presence of persistence, in other words, to check the generators as suitable for use in cryptography.

Keywords: pseudo-random sequence generator, Hurst coefficient, streaming encryption, the linear congruential generator, persistence.