УДК 519.72
DOI https://doi.org/10.32838/2663-5941/2021.2-1/31

Мошенський А.О.
orcid.org/0000-0002-4584-4958
Національний університет харчових технологій

Савченко Ю.Г.
orcid.org/0000-0002-1799-6700
Національний технічний університет
«Київський політехнічний інститут імені Ігоря Сікорського»

Гуйда О.Г.
orcid.org/0000-0002-2019-2615
Таврійський національний університет імені В.І. Вернадського

КОМБІНАТОРНІ ЗАСОБИ ПОКРАЩЕННЯ ХАРАКТЕРИСТИК ПСЕВДОВИПАДКОВИХ ЧИСЛОВИХ ПОСЛІДОВНОСТЕЙ

У статті розглянуто задачу покращення параметрів і властивостей псевдовипадкових числових послідовностей (ПЧП) за рахунок комбінаторного «змішування» фрагментів вихідних послідовностей, отриманих за допомогою традиційних генераторів на основі регістрів зі зворотними зв'язками. Враховуючи, що ПЧП визначають у багатьох застосуваннях найбільш важливі параметри електронної системи, наприклад телекомунікаційної мережі загалом. У цьому випадку ефективність захисту інформаційного обміну критично залежить від властивостей ключа шифрування, яким здебільшого є ПЧП. Тому актуальність задачі не викликає сумнівів. Для її розв'язання визначені підходи для кількісної оцінки якості ПЧП.

Розглянуто основні методи розв'язання задачі та показано, що найбільш універсальним є підхід на основі обчислення ентропії ПЧП, тобто, по суті її випадковості (непередбачуваності) Однак точне обчислення цього показника пов'язано зі значними труднощами через необхідність врахування у більшості реальних застосувань ймовірностей (частот) усіх чисел послідовності. Можна стверджувати, що при будь-якому підході збільшення довжини ПЧП покращує її характеристики з точки зору наближення її до параметрів істинно випадкової послідовності.

У статті для досягнення цієї мети пропонується при генерації ПЧП використовувати кілька традиційних генераторів на регістрах зсуву зі зворотними зв'язками, а кінцеву ПЧП формувати шляхом «склеювання» фрагментів, отриманих від регістрів. Власне процедура склеювання реалізується деяким умовним «міксером» (програмним або апаратним), який по черзі зчитує фрагменти та записує їх у підсумкову ПЧП. Таку процедуру можна застосовувати багаторазово, змінюючи розміри фрагментів і стартові налаштування регістрів. Важливим є те, що при такому об'єднанні фрагментів їхні статистичні характеристики (ймовірнісний розподіл символів і груп символів) зберігаються у підсумковій ПЧП.

Ключові слова: *псевдовипадкова послідовність, регістр зсуву, скремблер, криптозахист.*

Постановка проблеми. Нині процедури генерації та використання псевдовипадкових числових послідовностей (далі – ПЧП) стали де факто обов'язковим компонентом сучасних інформаційних технологій. Це можна пояснити досить широкою сферою їх застосування. Для прикладу можна назвати кілька прикладів таких сфер:

– Використання ПЧП як ключів у шифруванні та дешифруванні повідомлень у телекомунікаційних системах.

– Застосування ПЧП як тестових сигналів для контролю та діагностування технічного стану цифрової апаратури у сигнатурних методах технічного обслуговування.

– Моделювання дестабілізуючих впливів довкілля або інших випадкових факторів на технічні об'єкти при їх випробуваннях.

Це лише кілька характерних прикладів, але вони дозволяють визначити вимоги до ПЧП з точки зору їхніх властивостей і характеристик, які бажано забезпечити при генерації таких послідовностей. Перші два приклади, мабуть, є найбільш характерними і дозволяють стверджувати, що:

1. Процедура створення (генерації) ПЧП обов'язково повинна бути детермінованою, тобто такою, що багаторазове її застосування завжди дає однаковий результат. У випадку, наприклад, коли ПЧП є ключем шифрування у відправника повідомлення, точно такий же ключ повинен бути створений отримувачем повідомлення (передавання ключа по відкритому каналу виключено за визначенням).

При сигнатурному діагностуванні цифрової апаратури та ж сама вимога випливає з необхідності забезпечити повторюваність результатів діагностичного експерименту при формуванні сигнатур справного технічного стану і типових порушень працездатності.

З іншого боку ПЧП за своїми статистичними характеристиками повинна максимально наближатися до істинно випадкової послідовності чисел, тобто появи будь-яких конкретних послідовностей із заданої множини (діапазону) повинні мати однакову імовірність (бути непередбачуваними).

У третьому випадку вимога детермінованості не є обов'язковою, хоча і бажаною. Можна використати, наприклад, істинно випадкові послідовності, отримані шляхом дискретизації та оцифрування шумового складника аналогового сигналу або іншим способом.

Аналіз останніх досліджень і публікацій. Вочевидь, для цілеспрямованого пошуку оптимальних за тим чи іншим критерієм способів покращення характеристик ПЧП необхідно спиратися на кількісні показники, що характеризують міру випадковості послідовності, її наближення до ідеалу, тобто істинно випадкової послідовності чисел. Але це питання досить складне і, на думку авторів, не має однозначної відповіді. І ось чому.

Що таке істинно випадкова послідовність (далі – ІВП)? На інтуїтивному рівні – це числова послідовність чисел, вибрана з деякої скінченної множини навмання (як карта з колоди), тобто ймовірність вибрати будь-яку іншу послідовність така сама і дорівнює *1/N*, де *N* – потужність множини. З точки зору теоретико-інформаційних критеріїв ентропія відповідного джерела, яке генерує такі ІВП, має бути максимально можливою, тобто

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i = \log_2 N, \qquad (1)$$

що виконується, коли $p_i = 1/N$ для усіх p_i, де p_i – ймовірність вибору *i*-ої послідовності.

Але ж бінарні послідовності, що реально застосовуються, наприклад, у реальних скремблерах для захисту інформаційного обміну, мають довжину порядка 128…512 біт, тому провести відповідний статистичний експеримент для тестування джерела, визначивши ймовірності появи кожної з $N = 2^n$ послідовностей нереально навіть за допомогою суперкомп'ютера. Тому величину ентропії $H = log_2 N$ слід сприймати скоріш як деяку абстракцію подібно поняттю, наприклад, застосовуваного у фізиці терміну «абсолютно чорне тіло».

Реальні ПЧП, що застосовуються в інформаційних технологіях, далекі від ідеалу хоча б через необхідність виконувати вимогу детермінованості алгоритму їх генерації. Ця вимога, по суті, еквівалентна використанню певної конкретної закономірності для кожної без винятку генерованої послідовності. З цього випливає, що частина послідовностей довжини n, для яких означена закономірність не виконується, мають імовірність появи, що дорівнює 0. Але ж вимога рівномірного розподілу ймовірності появи кожного з чисел у послідовності не є гарантією наближення до істинно випадкової послідовності. Наприклад, звичайний двійковий лічильник, який генерує послідовність 0, 1, 2, 3, 4, … навряд чи хто-небудь наважиться вважати генератором ПЧП, хоча кожне число на його виходах з'являється з однаковою імовірністю *1/n*, де n – розрядність лічильника.

Існують досить універсальні критерії «випадковості» бінарної послідовності чисел. Наприклад, А.М. Колмогоровим [1] було зроблене фундаментальне припущення: чим складніший опис числової послідовності, тим вона ближче до ІВП. Із цього постулату можна зробити висновок, що опис ІВП – це сама послідовність чисел, дійсно вибрана випадково.

Інший універсальний критерій сформульований Ендрю Яо [2]: для випадкової послідовності ймовірність появи в наступному біті 0 або 1 дорівнює ½ незалежно від того, які числа з'явилися раніше. Тому фундаментальним тестом на випадковість є тест на наступний біт: не повинно існувати поліноміального алгоритму, який на основі перших *k* бітів випадкової послідовності зможе передбачити *(k+1)*-ий біт із ймовірністю, більшою за 1/2.

На практиці для визначення рівня випадковості ПЧП використовують різні статистичні тести, такі як DIEHARD або NIST [3]. У 1982 році Е. Яо довів, що генератор, який пройшов тест на «наступний біт», пройде і будь-які інші тести, виконані за поліноміальний час. Але реально при оцінці рівня випадковості конкретної послідовності чисел завжди залишається велика доля невизначеності. І ось чому.

Насамперед слід зазначити, що є принципова різниця між задачею оцінки рівня випадковості окремо взятої конкретної послідовності та оцінки «якості» відповідного генератора ПЧП загалом. У першому випадку, коли для оцінки надана лише одна послідовність чисел (фактично, це послідовність двійкових чисел), її ентропію можна виміряти на основі частоти появи (ймовірностей) окремих бітів, пар бітів (00, 01, 10, 11) трійок 000, 001, 011 … і так далі. Вочевидь, для скінченої послідовності зі збільшенням довжини таких фрагментів ці частоти будуть зменшуватися, а для деяких фрагментів стануть дорівнювати 0. До речі, цей розподіл дає корисну інформацію для криптоаналізу [4].

Формально у другому випадку можна використати аналогічний підхід, об'єднавши («склеївши» умовно) деяку скінченну множину послідовностей. Але виникає питання щодо обґрунтованості такого переходу з точки зору коректної оцінки якості окремих послідовностей.

Хоча нині відома велика кількість тестів на випадковість, мабуть, найбільш природним і інтуїтивно зрозумілим із них залишається тест, що використовує оцінку ентропії ПЧП, тобто обчислення рівня інформаційної надлишковості. Наприклад, до ПЧП застосовують процедури архівації, а за рівнем досягнутого стиснення (компресії) оцінюють якість послідовності. Виходячи з (1), збільшивши кількість доданків у цьому співвідношенні, можна покращити й оцінку якості ПЧП. З цього випливає тривіальний висновок: чим більша довжина послідовності N, тим краще.

Виклад основного матеріалу дослідження. Наразі оприлюднена сотня (без перебільшення) наукових статей і патентів з описом способів і методів збільшення довжини ПЧП. Більшість із них ґрунтується на використанні як генераторів ПЧП регістрів зсуву зі зворотним зв'язком по модулю 2 (Linear feedback shift register-LFSR) (рис. 1).

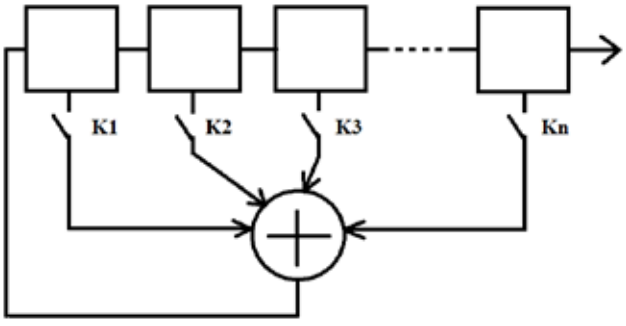

Рис. 1. Генератор на основі регістра зсуву зі зворотними зв'язками по модулю 2

На цьому рисунку наведено типову схему такого генератора. Генератори на основі LFSR без проблем реалізуються апаратно та програмно і генерують послідовності довжиною 2^n-1 двійкових п-розрядних чисел без їх повторення у послідовності, що задовольняє вимоги більшості застосувань. Так, вже при помірній довжині регістру п = 32 довжина генерованої послідовності складає $N = 2^{32}-1$ 32-розрядних комбінацій, що задовольняє будь-які вимоги застосувань у скремблерах або сигнатурному діагностуванні цифрових пристроїв.

Але з точки зору вимог криптоаналізу послідовності генеровані LFSR можуть бути досить легко «зламані» у результаті так званих алгебраїчних атак [4]. Ідея такої атаки полягає в обчисленні параметрів фільтра, який її генерує. Дійсно, LFSR повністю і однозначно визначений п-коефіцієнтами примітивного поліному, що його утворює, та *n* розрядною стартовою комбінацією, з якої починається генерація. Не заглиблюючись в алгебраїчні деталі, можна стверджувати, що для обчислення коефіцієнтів полінома треба записати та розв'язати *n* лінійних відносно операції складання по модулю 2 незалежних рівнянь (це необхідна умова, але не завжди достатня).

Можна також використати інший шлях: перебором можливих значень коефіцієнтів знайти такі, що відповідають значенням заданої числової послідовності. До речі, перебір може бути відносно невеликим, якщо перебирати лише серед коефіцієнтів примітивних поліномів заданого ступеня (за приблизною оцінкою таких поліномів). Саме ці обставини є головною вадою при використанні L як генераторів ПЧП.

Суттєвого збільшення довжини послідовності можна досягти [5], застосувавши процедуру, за якої після закінчення періоду генерації для одного примітивного полінома у роботу системи включався б інший. Таким чином для системи певної розрядності період генерації збільшувався б удвічі. Отже, можливо створювати системи, в яких період генерації залежав би тільки від кількості примітивних поліномів. Але таке вдосконалення не вирішує проблеми захисту від алгебраїчної атаки загалом, хоча і збільшує трудомісткість криптоаналізу.

Для подальшого вдосконалення процедури генерації пропонується підхід, коли в схемі використовується кілька регістрів (мінімум два), а для ускладнення криптоаналізу формування ПЧП здійснюється шляхом об'єднання окремих фрагментів різних послідовностей, що генеруються

різними регістрами. Для цього використовується перемикач виходів регістрів (рис. 2), який за певною програмою вибирає фрагменти числових послідовностей із кожного з регістрів, формуючи таким чином результуючу ПЧП. Наприклад, якщо використовується два регістра R1 та R2, які генерують відповідно числа a1, a2, a3, … та в1, в2, в3, …, то перемикач за відповідною програмою може сформувати такі варіанти ПЧП:

a1, в1, a2, в2, a3, в3 …;
a1, в2, в3, a2, в4, в5 …;
a1, в3, a3, в5, a5, в7 …;
……

Очевидно, що комбінаторне різноманіття таких варіантів практично нескінченне, що дозволяє не лише суттєво збільшувати за своїм бажанням довжину ПЧП, а й ставить під сумнів успішність алгебраїчних атак. Принаймні трудомісткість таких процедур збільшується багаторазово.

Визначимо, наскільки зростає кількість можливих варіантів генерованих ПЧБ при застосуванні пропонованого підходу. Так, нехай на входи умовного «міксера» надходять фрагменти окремих ПЧД із *m* регістрів і кожен регістр задає (вносить) у підсумкову ПВП фрагмент у середньому довжиною *s* байтів. Тоді підсумкова послідовність буде довшою мінімум у *m* разів.

Власне процедура утворення підсумкової ПЧП реалізується деяким умовним «міксером» (апаратним або програмним), який по черзі зчитує фрагменти із *m* регістрів і записує їх у підсумкову ПЧП. Таку процедуру можна застосовувати багаторазово, змінюючи розмір фрагментів При цьому кожного разу відбувається відповідне збільшення довжини послідовності. Важливо також, що при об'єднанні фрагментів їхні статистичні характеристики (ймовірнісний розподіл символів і груп символів) зберігається у підсумковій ПЧП.

Висновки. Таким чином, пропонована інформаційна технологія є, по суті, засобом формування та керування параметрами ПЧД. Зокрема, технологія дозволяє:

1) змінювати в широкому діапазоні довжину ПЧП залежно від вимог застосування;

2) регулювати склад і довжину фрагментів підсумкової ПЧП.

Важливо також, що параметри формування ПЧП можна оперативно змінювати, наприклад, для кожного сеансу інформаційного обміну, що збільшує ефективність захисту.

Список літератури:

1. Колмогоров А.Н. Три подхода к определению понятия «количество информации». *Проблемы передачи информации*, 1964, Т. 1(1). С. 3–11.

2. Statistical Testing of Random Number Generators / Proceedings of the 22-nd National Information Systems Security Conference, 10/99.

3. 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, вип. 2. С. 206–214.

4. Савченко Ю.Г., Малогулко Р.В. Вдосконалення генераторів ПВП та їх застосування в системах скремблер-дескремблер телекомунікаційних пристроїв. *Наукові записки УНДІЗ*, № 6(8), 2008.

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

Moshenskyi A.O., Savchenko Yu.H., Huida O.H. COMBINATORIAL MEANS OF IMPROVING THE CHARACTERISTICS OF PSEUDO-RANDOM NUMERICAL SEQUENCES

The article considers the problem of improving the parameters and properties of pseudo-random numerical sequences (PRS) by combinatorially "mixing" fragments of the original sequences obtained using traditional generators based on feedback registers. Given that PRSs determine in many applications the most important parameters of the electronic system, such as the telecommunications network as a whole. In this case, the effectiveness of information exchange protection is critically dependent on the properties of the encryption key, which is mostly PRS.

Therefore, the relevance of the problem is not in doubt. To address this, approaches have been identified to quantify the quality of PRSs. The main methods of solving the problem are considered and it is shown that the most universal approach is based on the calculation of the entropy of PRS, ie, essentially its randomness (unpredictability frequencies) of all sequence numbers. It can be argued that in any approach, increasing the length of the PRS improves its shape by characteristics in terms of approaching it to the parameters of a truly random sequence.

In the article to achieve this goal it is proposed to use several traditional generators on shift registers with feedback, and the final PRS to form by "gluing" the fragments obtained from the registers. Actually, the

gluing procedure is implemented by some conditional "mixe" (software or hardware), which in turn reads the fragments and writes them to the final PRS. This procedure can be applied repeatedly by resizing fragments and starting register settings. Importantly, with such a combination of fragments, their statistical characteristics (probabilistic distribution of symbols and groups of symbols) are stored in the final PRS.

Key words: *pseudo-random sequence, shift register, scrambler, cryptosecurity.*