УДК 004.942

ГЕНЕРАТОРИ ПСЕВДОВИПАДКОВИХ ПОСЛІДОВНОСТЕЙ ДЛЯ КРИПТОГРАФІЧНИХ СИСТЕМ

Назаров Є.О., Чернишова А.В., Губенко Н.Є.
Донецький національний технічний університет

Розглянуто алгоритмічні методи генерації псевдовипадкових послідовностей, чисел (ГВЧ) для генерації гам в криптографічних системах. Розроблено алгоритм і програмна реалізація лінійного конгруентного методу ГВЧ. Проведено дослідження ефективності лінійних конгруентних методів ГВЧ на основі статистичних критеріїв узгодження та серій.

Вступ

Криптографічні методи захисту інформації - це спеціальні методи шифрування, кодування або іншого перетворення інформації, в результаті якого її отримання стає недоступним без пред'явлення ключа криптограми й оберненого перетворення. Криптографічний метод захисту, безумовно, самий надійний метод, тому що охороняється безпосередньо сама інформація, а не доступ до неї.

Основна проблема класичної криптографії довгий час полягала у необхідності генерування непередбачуваних послідовностей чисел великої довжини із застосуванням короткого випадкового ключа. Для її розв'язання широко використовуються генератори псевдовипадкових послідовностей. Істотний прогрес у розробці та аналізі таких генераторів було досягнуто лише на початку шістдесятих років двадцятого століття. Однак і на сьогоднішній день створення якісного крипостійкого ГВЧ залишається предметом численних досліджень [1].

У даній статті розглянуті різні алгоритмічні способи отримання псевдовипадкових послідовностей та проаналізовано ефективність їх застосування в криптографічних системах для перетворення повідомлення в шифровку.

Застосування генераторів випадкових чисел у криптографії

Одержувані програмно з ключа, випадкові або псевдовипадкові ряди чисел називаються гамою, за назвою букви грецького алфавіту, якою в математичних записах позначаються випадкові величини. За способом отримання генератори випадкових чисел діляться на класи: фізичні; табличні; алгоритмічні. Фізичне моделювання випадковості за допомогою таких фізичних явищ, як радіоактивне випромінювання, дробовий шум у електронній лампі або тунельний пробій напівпровідникового стабілітрона не дають справжніх випадкових процесів. Хоча відомі випадки вдалих застосувань їх у генерації ключів, наприклад, в російському криптографічному пристрої «Криптон». Тому замість фізичних процесів для генерації гами застосовують програми для ЕОМ, які називаються генераторами випадкових чисел, але насправді видають детерміновані числові ряди, які тільки здаються випадковими за своїми властивостями. Можна сформулювати основні вимоги до криптографічного генератору псевдовипадкової послідовності або

гами: період гами повинен бути достатньо великим для шифрування повідомлень різної довжини, гама повинна бути важко передбачуваною, генерування гами не повинно бути пов'язане з великими технічними і організаційними труднощами.

Алгоритмічні ГВЧ

Заслуга конструювання довгих псевдовипадкових рядів з хорошими статистичними властивостями повністю належить криптографії. Але області застосування цих методів значно ширше.

Числа, які генеруються за допомогою ГВЧ, є псевдовипадковими і кожне наступне генероване число залежить від попереднього:

$$r_{i+1}=f(r_i).$$

Послідовності, що складаються із таких чисел, можуть утворювати петлі, тобто, як правило, існує цикл, що повторюється нескінченне число разів. Повторювані цикли називаються періодами. Перевагою даних ГВЧ є їх швидкодія, генератори практично не вимагають великих ресурсів пам'яті, компактні. Недоліки: числа не можна в повній мірі назвати випадковими, оскільки між ними існує залежність, а також наявність періодів в послідовності квазі случайних чисел.

Найбільш поширеними алгоритмічними методами отримання ГВЧ є:

- метод серединних квадратів;
- метод серединних добутків;
- метод перемішування;
- лінійний конгруентний метод.

Лінійний конгруентний метод генерації псевдовипадкових чисел

Лінійний конгруентний метод (ЛКМ) є однією з найбільш уживаних в даний час процедур, що імітують отримання випадкових чисел. Цей метод широко використовується в криптографії в якості генератора гами для ключів, головним чином, за рахунок простоти реалізації і хороших статистичних властивостей, продемонстрованих на численних емпіричних тестах.

У ЛКМ використовується операція mod(x,y), що повертає залишок від ділення першого аргументу на другий. Кожне наступне випадкове число розраховується на основі попереднього випадкового числа за наступною рекурентною формулою:

$$r_{i+1} = \mathrm{mod}(k \cdot r_i + b, \mathrm{M}), \qquad (1)$$

де M – модуль ($M > 0$); k – множник ($0 \le k < M$); b – приріст ($0 \le b < M$); r_0 – початкове значення ($0 \le r_0 < M$).

Послідовність випадкових чисел, отриманих за допомогою даної формули, називається лінійною конгруентною послідовністю. Багато авторів називають лінійну конгруентну послідовність при b=0 мультиплікативним конгруентним методом, а при $b \neq 0$ – змішаним конгруентним методом [2].

Для якісного генератора потрібно підібрати відповідні коефіцієнти. Необхідно, щоб число M було досить великим, так як період не може мати більше M елементів. З іншого боку, операція ділення, що використовується в цьому методі, є досить повільною операцією, тому для двійкової обчислювальної машини логічним буде вибір $M=2^N$,

оскільки в цьому випадку знаходження залишку від ділення зводиться всередині ЕОМ до двійкової логічної операції «AND».

Також широко поширений вибір найбільшого простого числа M, меншого, ніж $M=2^N$: в спеціальній літературі доводиться, що в цьому випадку молодші розряди одержуваного випадкового числа r_{i+1} поводяться так само випадково, як і старші, що позитивно позначається на всій послідовності випадкових чисел в цілому. В якості прикладу можна навести одне з чисел Мерсенна, що дорівнює $M=2^{31}–1$.

Однією з вимог до лінійних конгруентних послідовностей є якомога більша довжина періоду. Довжина періоду залежить від значень і співвідношень між параметрами M,k і b та встановлюється наступним твердженням [2-3].

Лінійна конгруентна послідовність, що визначається параметрами M, k, b і r0 має період довжиною, якщо M:

1) числа b і M є взаємно простими;
2) число k-1 кратне p для кожного простого p, що є дільником M;
3) число k-1 кратне 4, якщо M кратне 4.

Для дослідження властивостей ГВЧ використовувався лінійний конгруентний метод з наступними параметрами: M=2^N, k=3+8q (або k=5+8q), b=0 , r0 - непарне.

Було встановлено, що ряд псевдовипадкових чисел, що генеруються на основі наведених даних, буде повторюватися через кожні M/4 чисел. Число q задається довільно перед початком обчислень, однак при цьому слід мати на увазі, що ряд справляє враження випадкового при великих k (а, значить, і q).

Результат можна дещо поліпшити, якщо b - непарне і k=1+4q - в цьому випадку ряд буде повторюватися через кожні M чисел:

$M = 2^{31} – 1$, k = 1220703125, b = 7, $r_0 = 7$.

Генератор випадкових чисел, що використовує дані з прикладу, буде видавати випадкові неповторювані числа з періодом, рівним 7 мільйонам.

Перевірка якості роботи генератора випадкових чисел

Від якості роботи ГВЧ залежить якість роботи всієї криптографічної системи в цілому і точність отриманих результатів. Тому згенеровані на основі лінійного конгруентного методу вибіркові послідовності, випадкові числа, породжувані ГВЧ, досліджувалися з цілого ряду статистичних критеріїв.

По-перше, було проведено дослідження згенерованих послідовностей на випадковість і незалежність на основі непараметричного критерію серій.

По-друге, здійснювалася перевірка рівномірності отриманого експериментального розподілу на основі наступних підходів:

– відповідності точкових вибіркових оцінок істинним значенням параметрів рівномірного закону розподілу;
– частотного тесту;
– критерію узгодження «χ^2-квадрат» або Пірсона.

Для того, щоб отримані випадкові послідовності мали рівномірний закон розподілу, ГВЧ повинен видавати значення статистичних параметрів близькі до наступних, характерних для рівномірного випадкового закону:

1) $m_r = \frac{\sum_{i=1}^{n} r_i}{n} \approx 0.5$ – математичне очікування;

2) $D_r = \frac{\sum_{i=1}^{n} (r_i - m_r)^2}{n-1} \approx \frac{1}{12}$ – дисперсія;

3) $\sigma_r = \sqrt{D_r} \approx 0.2887$ – середньоквадратичне відхилення.

Коефіцієнт асиметрії повинен наближено дорівнювати нулю.

Наступним етапом перевірки на рівномірність розподілу було використання частотного тесту. Частотний тест дозволяє з'ясувати, скільки чисел потрапило в інтервал ($m_r - \sigma_r$; $m_r + \sigma_r$), тобто (0,5 – 0,2887; 0,5 + 0,2887) чи, зрештою, (0,2113; 0,7887). Оскільки 0,7887 – 0,2113 = 0,5774, вважаємо, що у якісному ГВЧ в цей інтервал повинно потрапляти близько 57.7% з усіх випадкових чисел. Також необхідно враховувати, що кількість чисел, які потрапили в інтервал (0;0.5), приблизно дорівнює кількості чисел, що потрапили в інтервал (0.5;1).

Критерій Пірсона (χ^2-критерій) – це один з найвідоміших статистичних критеріїв; він є основним методом перевірки відповідності емпіричного розподілу передбачуваному теоретичному закону розподілу. Для нашого випадку перевірка за критерієм «χ^2-квадрат» дозволить дізнатися, наскільки створений нами реальний ГВЧ близький до еталону ГВЧ, тобто задовольняє він вимогам рівномірного розподілу чи ні. Так як закон розподілу еталонного ГВЧ рівномірний, то теоретична ймовірність pi попадання чисел в i-тий інтервал (всього цих інтервалів k) дорівнює pi=1/k. Таким чином, в кожний з k інтервалів потрапить рівно по pi*N чисел (N – загальна кількість згенерованих чисел) [4].

Реальний ГВЧ видаватиме числа, розподілені (причому, не обов'язково рівномірно) по k інтервалах і в кожен інтервал потрапить по n_i чисел (в сумі $n_1 + n_2 + \ldots + n_k = N$). Розглянемо квадрати різниць між отриманою кількістю чисел ni і «еталонною» pi*N. Складемо їх, і в результаті отримаємо:

$$\chi^{2^*} = (n_1 - p_1 N)^2 + (n_2 - p_2 N)^2 + \ldots + (n_k - p_k N)^2.$$

З цієї формули випливає, що чим менше різниця у кожному з доданків (а значить, і чим менше значення χ^{2^*}), тим сильніше закон розподілу випадкових чисел, що генеруються реальним ГВЧ, тяжіє до рівномірного. У попередньому виразі кожному з доданків приписується однакова вага, що насправді може не відповідати дійсності; тому для статистики «χ^2-квадрат» необхідно провести нормування кожного i-го доданка, поділивши його на Npi:

$$\chi^{2^*} = \sum_{i=1}^{k} \frac{(n_i - p_i N)^2}{N p_i} = \frac{1}{N} \sum_{i=1}^{k} \left(\frac{n_i^2}{p_i} \right) - N.$$

При цьому додатково треба мати на увазі, що всі значення повинні бути досить великими, наприклад більше 5, тільки тоді (при досить великій статистичній вибірці) умови проведення експерименту можна вважати задовільними. Далі отримане значення статистики критерію узгодження за загальними правилами перевірки статистичних гіпотез порівнюється з критичним на заданому рівні значущості α: $\chi^{2^*} < \chi^2_{1-\alpha}$. При виконанні зазначеної нерівності вважається, що передбачуваний рівномірний розподіл випадкових чисел не суперечить дослідженим даним, тобто робиться висновок про

придатність генератора для використання у криптографічній системі.

Для розробленого ГВЧ на основі лінійного конгруентного методу із зазначеними параметрами проведена серія статистичних випробувань для різних обсягів вибіркових даних і різних рівнів значущості α=0,05; 0,001; 0,0005, ... Проведений експеримент показав істотну залежність якості отриманих послідовностей випадкових чисел від довжини послідовності, що генерується.

Основна часть досліджень проводилась у професійному пакеті STATISTICA, що включає практично всі види статистичної обробки експерименту. Для перевірки якості ГВЧ було використано такі модулі пакету, як Nonparametric Statistics із блоками: Nonparametric stat (непараметричні методи) і Distribution fitting (підгін ймовірнісного розподілу до реальних даних), модуль Basic Statistics/Tables – ймовірнісний калькулятор (Probability calculator) та інші. Приклад застосування критерію Пірсона приведено на рис. 1-2.

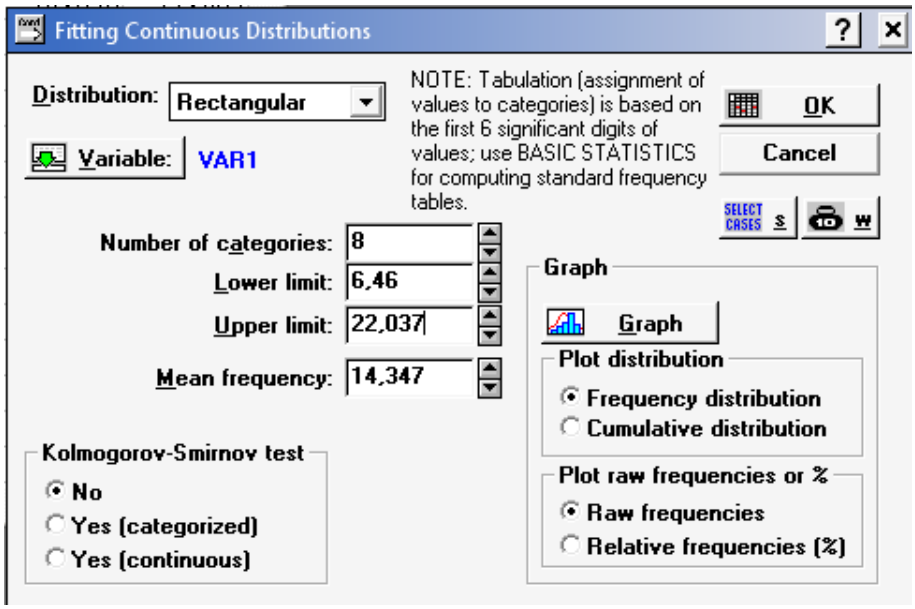

Рисунок 1. Вікно модулю Distribution fitting (підбір рівномірного закону розподілу)

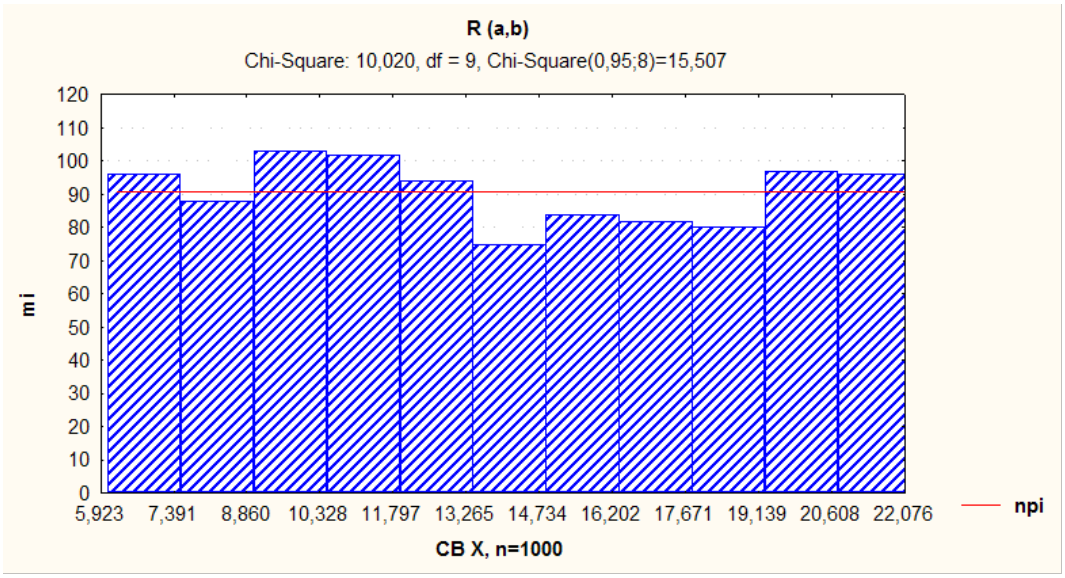

Рисунок 2. Іллюстрація виконання досліджень за критерієм Пірсону у модулі Distribution fitting в пакеті STATISTICA

Висновки

У роботі розглянуто використання та досліджено ефективність лінійного конгруентного методу генерації псевдовипадкових послідовностей для генерації гам в криптографічних системах. Результати теоретичних досліджень і проведених експериментів продемонстрували добрі статистичні властивості наведеного ЛКМ для генерації псевдовипадкових послідовностей.

Перспективним напрямом подальших досліджень є застосування для генерації випадкових чисел методів цілочисельної арифметики, зокрема послідовностей Фібоначчі, властивостей простих чисел, що дозволить збільшити криптографічну стійкість ГВЧ.

Література

[1] Фергюсон Н. Практическая криптография / Н. Фергюсон, Б. Шнайдер. Пер. с англ. – М.: Издательский дом «Вильямс», 2005. – 424 с.

[2] Кнут Д. Искусство программирования, том 2. Получисленные методы / Д. Кнут. – М.: Изд. дом «Вильямс», 2007. – 832 с.

[3] Соболь И.М. Численные методы Монте-Карло / И.М. Соболь. М.: Наука, 1977. – 327 с.

[4] Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер. – М.: Юнити-Дана, 2000. – 543 с.