УДК 004.942

ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ В КРИПТОГРАФИЧЕСКИХ СИСТЕМАХ

Губенко Н.Е., Назаров Е.А.
Донецкий национальный технический университет, г. Донецк
Кафедра компьютерных систем мониторинга
E-mail: gubenko@cs.dgtu.donetsk.ua, ianazard@gmail.com

Аннотация

Губенко Н.Е., Назаров Е.А. Эффективность использования генераторов псевдослучайных чисел в криптографических системах. *Рассмотрены алгоритмические методы генерации псевдослучайных последовательностей (ГСЧ) для генерации гамм в криптографических системах. Разработан алгоритм и программная реализация линейного конгруэнтного метода ГСЧ. Проведено исследование эффективности линейных конгруэнтных методов ГСЧ на основе статистических критериев согласия и серий.*

Общая постановка проблемы

Криптографические методы защиты информации – это специальные методы шифровки, кодировки или другого превращения информации, в результате которого ее содержание становится недоступным без предъявления ключа криптограммы и обратного превращения. Криптографический метод защиты, безусловно, самый надежный метод, потому что охраняется непосредственно сама информация, а не доступ, к ней.

Основная проблема классической криптографии долгое время заключалась в трудности генерирования непредсказуемых последовательностей чисел большой длины с применением короткого случайного ключа. Для ее решения широко используются генераторы псевдослучайных последовательностей. Существенный прогресс в разработке и анализе таких генераторов был достигнут лишь к началу шестидесятых годов двадцатого столетия. Однако и на сегодняшний день создание качественного криптостойкого ГСЧ остается предметом многочисленных исследований [1].

В данной статье рассмотрены различные алгоритмические способы получения псевдослучайных последовательностей и проанализирована эффективность их применения в криптографических системах для преобразования сообщения в шифровку.

Применение генераторов случайных чисел в криптографии

Получаемые программно из ключа, случайные или псевдослучайные ряды чисел называются гаммой, по названию γ – буквы греческого алфавита, которой в математических записях обозначаются случайные величины. По способу получения генераторы случайных чисел делятся на классы: физические; табличные; алгоритмические. Физическое моделирование случайности с помощью таких физических явлений, как радиоактивное излучение, дробовой шум в электронной лампе или туннельный пробой полупроводникового стабилитрона не дают настоящих случайных процессов. Хотя известны случаи удачных применений их в генерации ключей, например, в российском криптографическом устройстве «Криптон». Поэтому вместо физических процессов для генерации гаммы применяют программы для ЭВМ, которые называются генераторами случайных чисел, но на самом деле выдающие детерминированные числовые ряды, которые только кажутся случайными по своим свойствам. Можно сформулировать основные требования к криптографическому генератору псевдослучайной последовательности или гаммы: период гаммы должен быть достаточно большим для шифрования сообщений различной длины, гамма должна быть трудно предсказуемой, генерирование гаммы не должно быть связано с большими техническими и организационными трудностями.

Алгоритмические ГСЧ

Заслуга конструирования длинных псевдослучайных рядов с хорошими статистическими свойствами полностью принадлежит криптографии. Но области применения этих методов значительно шире.

Числа, генерируемые с помощью ГСЧ, являются псевдослучайными и каждое последующее сгенерированное число зависит от предыдущего:

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

Последовательности, составленные из таких чисел, могут образовывать петли, то есть, как правило, существует цикл, повторяющийся бесконечное число раз. Повторяющиеся циклы называются периодами. Достоинством данных ГСЧ является их быстродействие, генераторы практически не требуют больших ресурсов памяти, компактны. Недостатки: числа нельзя в полной мере назвать случайными, поскольку между ними имеется зависимость, а также наличие периодов в последовательности квазислучайных чисел. Наиболее распространенными алгоритмическими методами получения ГСЧ являются:
метод серединных квадратов; метод серединных произведений; метод перемешивания; линейный конгруэнтный метод.

Линейный конгруэнтный метод генерации псевдослучайных чисел

Линейный конгруэнтный метод (ЛКМ) является одной из наиболее употребительных в настоящее время процедур, имитирующих получение случайных чисел. Этот метод широко используется в криптографии в качестве генератора гаммы для ключей, главным образом, за счет простоты реализации и хороших статистических свойств, продемонстрированных на многочисленных эмпирических тестах.

В ЛКМ используется операция $mod(x,y)$, возвращающая остаток от деления первого аргумента на второй. Каждое последующее случайное число рассчитывается на основе предыдущего случайного числа по следующей рекуррентной формуле:

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

где M – модуль ($0 < M$);
k – множитель ($0 \le k < M$);
b – приращение ($0 \le b < M$),
r_0 – начальное значение ($0 \le r_0 < M$).

Последовательность случайных чисел, полученных с помощью данной формулы, называется линейной конгруэнтной последовательностью. Многие авторы называют линейную конгруэнтную последовательность при $b = 0$ мультипликативным конгруэнтным методом, а при $b \neq 0$ – смешанным конгруэнтным методом [2].

Для качественного генератора требуется подобрать подходящие коэффициенты. Необходимо, чтобы число M было довольно большим, так как период не может иметь больше M элементов. С другой стороны, деление, использующееся в этом методе, является довольно медленной операцией, поэтому для двоичной вычислительной машины логичным будет выбор $M = 2^N$, поскольку в этом случае нахождение остатка от деления сводится внутри ЭВМ к двоичной логической операции «AND».

Также широко распространен выбор наибольшего простого числа M, меньшего, чем 2^N: в специальной литературе доказывается, что в этом случае младшие разряды получаемого случайного числа r_{i+1} ведут себя так же случайно, как и старшие, что положительно сказывается на всей последовательности случайных чисел в целом. В качестве примера можно привести одно из чисел Мерсенна, равное $M = 2^{31} - 1$.

Одним из требований к линейным конгруэнтным последовательностям является как можно большая длина периода. Длина периода зависит от значений и соотношений между параметрами: M, k и b и устанавливается следующим утверждением [2-3].

Линейная конгруэнтная последовательность, определенная числами M, k, b и r_0, имеет период длиной M, если:

1) числа b и M взаимно простые;

2) $k - 1$ кратно p для каждого простого p, являющегося делителем M;

3) $k - 1$ кратно 4, если M кратно 4.

Для исследования свойств ГСЧ использовался линейный конгруэнтный метод со следующими параметрами: $M = 2^N$, $k = 3 + 8 \cdot q$ (или $k = 5 + 8 \cdot q$), $b = 0$, r_0 – нечетно.

Было установлено, что ряд псевдослучайных чисел, генерируемых на основе приведенных данных, будет повторяться через каждые $M/4$ чисел. Число q задается произвольно перед началом вычислений, однако при этом следует иметь в виду, что ряд производит впечатление случайного при больших k (а значит, и q).

Результат можно несколько улучшить, если b – нечетно и $k = 1 + 4 \cdot q$ – в этом случае ряд будет повторяться через каждые M чисел: $M = 2^{31} - 1$, $k = 1\,220\,703\,125$, $b = 7$, $r_0 = 7$.

Генератор случайных чисел, использующий данные из примера, будет выдавать случайные неповторяющиеся числа с периодом, равным 7 миллионам.

Проверка качества работы генератора случайных чисел

От качества работы ГСЧ зависит качество работы всей криптографической системы и точность полученных результатов. Поэтому полученные на основе линейного конгруэнтного метода выборочные последовательности, случайные числа, порождаемые ГСЧ, исследовались по целому ряду статистических критериев.

Во-первых, было проведено исследование генерируемой последовательности на случайность и независимость на основе непараметрического критерия серий.

Во-вторых, осуществлялась проверка равномерности полученного экспериментального распределения на основе следующих подходов:

- соответствия точечных выборочных оценок истинным значениям параметров равномерного закона распределения;
- частотного теста;
- критерия согласия «χ^2-квадрат» или Пирсона.

Для того, чтобы полученные случайные последовательности имели равномерный закон распределения, ГСЧ должен выдавать значения статистических параметров близкие к следующим, характерных для равномерного случайного закона:

1) $m_r = \dfrac{\sum_{i=1}^{n} r_i}{n} \approx 0.5$ – математическое ожидание;

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

3) $\sigma_r = \sqrt{D_r} \approx 0.2887$ – среднеквадратичное отклонение.

Коэффициент асимметрии должен быть приближенно равен нулю.

Следующим этапом проверки на равномерность распределения было использование частотного теста. Частотный тест позволяет выяснить, сколько чисел попало в интервал $(m_r - c_r; m_r + c_r)$, то есть $(0.5 - 0.2887, 0.5 + 0.2887)$ или, в конечном итоге, $(0.2113, 0.7887)$. Так как $0.7887 - 0.2113 = 0.5774$, заключаем, что в хорошем ГСЧ в этот интервал должно попадать около 57.7% из всех выпавших случайных чисел (см. рис. 1). Также необходимо учитывать, что количество чисел, попавших в интервал (0; 0.5), должно быть примерно равно количеству чисел, попавших в интервал (0.5; 1).

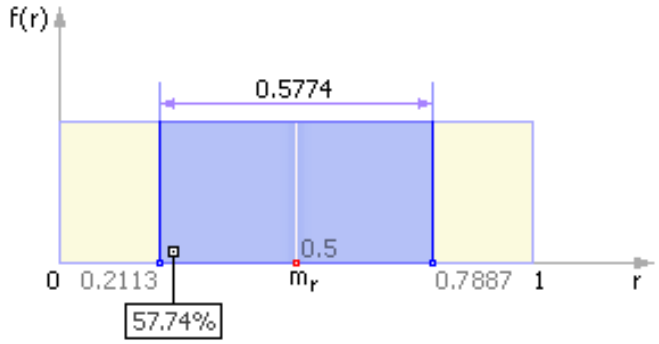

Рисунок 1 – Частотная диаграмма идеального ГСЧ
в случае проверки его на частотный тест

Критерий «хи-квадрат» (χ^2-критерий) – это один из самых известных статистических критериев; он является основным методом проверки соответствия эмпирического распределения предполагаемому теоретическому закону распределения. Для нашего случая проверка по критерию «χ^2-квадрат» позволит узнать, насколько созданный нами реальный ГСЧ близок к эталону ГСЧ, то есть удовлетворяет ли он требованию равномерного распределения или нет. Частотная диаграмма *эталонного* ГСЧ представлена на рис. 2. Так как закон распределения эталонного ГСЧ равномерный, то теоретическая вероятность p_i попадания чисел в i-ый интервал (всего этих интервалов k) равна $p_i = 1/k$. И, таким образом, в каждый из k интервалов попадет ровно по $p_i \cdot N$ чисел (N – общее количество сгенерированных чисел) [4].

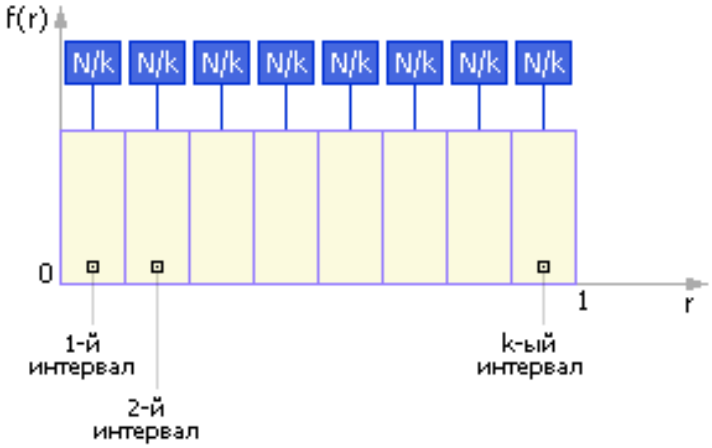

Рисунок 2 – Частотная диаграмма эталонного ГСЧ

Реальный ГСЧ будет выдавать числа, распределенные (причем, не обязательно равномерно) по k интервалам и в каждый интервал попадет по n_i чисел (в сумме $n_1 + n_2 + \ldots + n_k = N$). Рассмотрим квадраты разностей между полученным количеством чисел n_i и «эталонным» $p_i \cdot N$. Сложим их, и в результате получим:

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

Из этой формулы следует, что чем меньше разность в каждом из слагаемых (а значит, и чем меньше значение χ^{2^*}), тем сильнее закон распределения случайных чисел,

генерируемых реальным ГСЧ, тяготеет к равномерному. В предыдущем выражении каждому из слагаемых приписывается одинаковый вес, что на самом деле может не соответствовать действительности; поэтому для статистики «χ^2-квадрат» необходимо провести нормировку каждого i-го слагаемого, поделив его на Np_i:

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

При этом дополнительно надо иметь в виду, что все значения Np_i должны быть достаточно большими, например больше 5, только тогда (при достаточно большой статистической выборке) условия проведения эксперимента можно считать удовлетворительными.

Далее полученное значение статистики критерия согласия по общим правилам проверки статистических гипотез сравнивается с критическим на заданном уровне значимости α:

$$\chi^{2^*} < \chi^2_{1-\alpha}.$$

При выполнении указанного неравенства считается, что предполагаемое равномерное распределение случайных чисел не противоречит опытным данным, то есть делается вывод о пригодности генератора для использования.

Для разработанного ГСЧ на основе линейного конгруэнтного метода с указанными параметрами проведена серия статистических испытаний для различных объемов выборочных данных и различных уровней значимости $\alpha = 0{,}05;0{,}001;0{,}0005\ldots$. Проведенный эксперимент показал существенную зависимость качества полученных последовательностей случайных чисел от длины генерируемой последовательности.

Выводы

В работе рассмотрено использование и исследована эффективность линейного конгруэнтного метода генерации псевдослучайных последовательностей для генерации гамм в криптографических системах. Результаты теоретических исследований и проведенных экспериментов показали хорошие статистические свойства приведенного ЛКМ для генерации псевдослучайных последовательностей.

Перспективным направлением дальнейших исследований является применение для генерации случайных чисел методов целочисленной арифметики, в частности последовательностей Фибоначчи, свойств простых чисел, что позволит увеличить криптографическую стойкость ГСЧ.

Список литературы

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

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

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

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