$p^{T}_{\Delta^{(i)}_{вх}}$. Отсюда находим нижнее значение вероятности характеристики Ω для сеансовых ключей $K \in K^E$: $p^{\min}_{\Omega} = p^{\max}_{\Omega} \cdot \left(p^{T}_{\Delta^{(i)}_{вх}}\right)^{r-n^{\Omega}_{j}} = 2^{-45} \cdot \left(\frac{3}{4}\right)^{45-8} \approx 2^{-45} \cdot 2^{-15,4} = 2^{-60,4}$. Получается, что $p^{\min}_{\Omega} > p^{\min}$, и для любого $K \in K^E$ возможно выполнение дифференциального криптоанализа.

Для экспериментальной проверки полученных результатов была выбрана 16-цикловая характеристика, совпадающая с первой половиной Ω:

$$\Omega^{16} = \left(\Delta^{\Omega^{16}}_{вх} = 3\ 0303\ 0030\,0000, \Delta^{\Omega^{16}}_{вых} = 3\ 0000\,0000, \Delta^{\Omega^{16}}_{i} = \Delta^{\Omega}_{i}, 2 \le i \le 16\right). \quad (6)$$

С учетом маски подключей M_{Ω} (табл. 4) был сформирован сеансовый ключ $K^{э} \in K^E$ (табл. 6). После выполнения $2 \cdot 3{,}57 \cdot 10^{9}$ шифрований на сеансовом ключе $K^{э}$, было найдено 207 верных пар характеристики (6), что даёт экспериментальную вероятность $p^{э}_{\Omega^{16}} \approx 5{,}8 \cdot 10^{-8} = 2^{-24,04}$, находящуюся в пределах $p^{\min}_{\Omega} = p^{\max}_{\Omega^{16}} \cdot \left(\frac{3}{4}\right)^{23-n^{\Omega^{16}}_{j}} = 2^{-30,9} \le p^{э}_{\Omega^{16}} \le p^{\max}_{\Omega^{16}} = 2^{-r_{\Omega^{16}}} = 2^{-23}$. При этом $p^{э}_{\Omega^{16}}$ достаточно близко к теоретически полученному значению вероятности $p_{\Omega^{16}} = 2^{-(23+1.28)} = 2^{-24,28}$.

Отсюда следует, что экспериментальные данные полностью подтверждают приведенные выше оценки.

Таблица 6 – Пример сеансового ключа ГОСТ 28147-89, построенного с учётом масок разности M_{Ω}

K_0	K_1	K_2	K_3
F7C0080B	C0E81014	40D801B5	FC081290
K_4	K_5	K_6	K_7
9FA403C0	800C07D4	40E40009	E7C00408

Таким образом, требования [2] допускают использование заполнения узлов замены, при котором ГОСТ 28147-89 является уязвимым для дифференциального криптоанализа (сложность поиска верной пары составляет величину от 2^{45} до 2^{60} для не менее чем 2^{156} сеансовых ключей). Однако, из-за низкой вероятности получения таких подстановок, весьма жёстких условий проведения атаки и её высокой вычислительной сложности рассмотренная методика не может стать реальной угрозой безопасности алгоритма ГОСТ 28147-89. Вместе с тем факт наличия теоретической атаки актуализирует вопрос дальнейшего совершенствования критериев отбора подстановок для ГОСТ 28147-89.

Литература: 1. Олейников Р. В. Дифференциальный криптоанализ алгоритма шифрования ГОСТ 28147-89. // Радиотехника. – 2001. – № 119. С. 146–152. 2 . Холоша А. А. Об одном подходе к анализу качества блока подстановки битовых векторов. // Збірник наукових праць інституту проблем моделювання в енергетиці НАНУ. Вип. 2. Львів: Світ. – 1998. с. 59–74. 3. M. Matsui. Linear Cryptanalysis Method for the DES Cipher. Lecture Notes in Computer Science, Advances in Cryptology, proceedings of Eurocrypt '93, 1993. pp. 27–41. 4. C. Charnes, L. O'Connor, J. Pieprzyk, R. Safavi-Naini, Y. Zheng. Further comments on the Soviet Encryption Algorithm. Advances in Cryptology – EUROCRYPT'96, Springer-Verlag, Berlin, 1996, pp. 433–438.

УДК 681.3.067:681.3.016

ОСНОВНАЯ ПРОБЛЕМА ТЕСТИРОВАНИЯ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ ЧИСЕЛ

Виктор Куценко, Тарас Левченко, Николай Миронов, Виктор Мясоедов
Научно-технический комплекс "Импульс", г. Киев

Аннотация: **Применяемые обычно критерии оценки псевдослучайных последовательностей приводят к отбраковке слишком большого числа последовательностей. Это приводит к усложнению схем генерации и замедляет шифрование. В статье предложен критерий оценки, аналогичный**

критерию χ^2. Введенный критерий связан с основной проблемой тестирования числовых генераторов. Для проверки пригодности этого критерия подходит описываемая модель численных экспериментов.
Summary: **The used usually criteria of a rating of random sequences result to drop too large number of sequences. It results in complication of the schemes of generation and slows down encryption. In clause the criterion of a ratings similar χ^2 is offered. The entered criterion is connected to the basic problem of testing of numerical generators. For check of suitability of this criterion the described model of numerical experiments approaches.**
Ключевые слова: **Равномерность распределения, статистические критерии, шифрование, статистическое тестирование, случайные числа, абсолютно стойкие алгоритмы, генератор Фибоначчи.**

I Введение

Область, пограничная между классической и дискретной математикой, мало изучена. Статистические исследования выхода генераторов случайных чисел (ГСЧ) предоставляют уникальные возможности изучения этой области в процессе численных экспериментов.

Гипотезы о функции распределении выхода генераторов оценивают по критерию χ^2, основанному на статистике Пирсона

$$\chi^2 = N * \sum_{i=1}^{n} (\frac{\nu_N(\alpha_i, \alpha_{i+1})}{N} - p_i)^2 / p_i. \tag{1}$$

По этому критерию непосредственный выход ГСЧ обычно не позволяет принять гипотезу о равномерности функции распределения [1]. Возможен выбор другого статистического критерия, основанного на предположении о гладкости функции распределения и замене в статистике Пирсона вектора частот независимых событий вектором отношений разностей функции распределения в соседних узлах сетки интервалов к длине интервалов. Точнее, из критериев равномерности [2] распределения любых последовательностей чисел в интервале [0, 1],

$$\lim_{N\to\infty} \frac{\nu_N(\alpha_i, \alpha_{i+1})}{N} = \alpha_{i+1} - \alpha_i \tag{2}$$

и

$$\lim_{N\to\infty} \frac{s_N(\alpha_i, \alpha_{i+1})}{N} = \frac{1}{2}(\alpha_{i+1} - \alpha_i) \tag{3}$$

следует критерий

$$\lim_{N\to\infty} \left[\frac{s_N(\alpha_i, \alpha_{i+1})}{\nu_N(\alpha_i, \alpha_{i+1})} - \frac{1}{2} \frac{s_N(\alpha_i, \alpha_{i+1})}{N} \right] = \alpha_i. \tag{4}$$

Эти критерии допускают обобщение на случай любых гладких функций распределения числовых последовательностей. Для равномерных распределений имеются строгие критерии, отличающиеся от критериев (2), (3). Однако для них непосредственный переход к функции равномерного распределения в правой части (4) затруднён.

Аналогом статистики Пирсона в критерии (4) является выражение

$$\overline{\chi}^2 = N * n * \sum_{i=0}^{n-1} (DF_{i+1} - DF_i - \frac{1}{n})^2, \tag{5}$$

в котором

$$DF_i = \left[\frac{s_N(\alpha_i, \alpha_{i+1})}{\nu_N(\alpha_i, \alpha_{i+1})} - \frac{1}{2} \frac{s_N(\alpha_i, \alpha_{i+1})}{N} \right], \tag{6}$$

а $\frac{s_N(\alpha_i, \alpha_{i+1})}{\nu_N(\alpha_i, \alpha_{i+1})}$ – средние значения чисел из сплошной выборки длиной N исходной числовой последовательности $\{\mathbf{t}\}$ в интервалах равномерной сетки α_i, $i=0, 1, ..., n$.

При этом требуется проверка «малости» разностей $DF_i - \alpha_i$ и возможность их интерпретации как нормально распределённой "ошибки измерения" функции распределения.

Предварительные наблюдения статистических явлений, обнаруживаемых в эмпирическом материале, были осуществлены с помощью компьютерной программы в НТК «Импульс». Эти наблюдения показали сходство поведения критериев χ^2 и $\bar{\chi}^2$ в зависимости от параметров выбранного генератора. При этом значения критерия $\bar{\chi}^2$, вообще говоря, оказались меньшими значений критерия χ^2. Имеются также некоторые основания для утверждения о том, что ошибки измерения функции распределения выхода датчика случайных чисел распределены нормально.

Предметом рассмотрения является экспериментальная модель эмпирического измерения функции распределения, обеспечивающая возможность решения основной проблемы статистического тестирования генераторов случайных чисел.

II Формулировка проблемы

Корректная формулировка проблемы статистического тестирования генераторов случайных чисел для шифрования данных заключается в следующем: пусть $p(x)$ – плотность распределения вероятности символа x в естественном тексте, причем x может быть символом естественного или "компьютерного" алфавита, $\bar{p}(\bar{x})$ – плотность распределения вероятности появления байта $\bar{x}$ в зашифрованном тексте. По эмпирическим (или, в случае бесконечной длины текста, близким к теоретическим) оценкам плотностей требуется определить функцию распределения вероятности появления шифрующих байтов на выходе генератора случайных чисел, определить тип генератора и его параметры, в том числе начальные значения параметров.

С этой точки зрения равномерность распределения шифрующих байтов представляется не столь существенной, как гладкость функции их распределения. В этом смысле критерий $\bar{\chi}^2$ предпочтительнее классического критерия χ^2. Несколько больший объём вычислений значений критерия $\bar{\chi}^2$ компенсируется большей толерантностью к эмпирическим последовательностям случайных чисел. Проверка гипотезы о нормальности распределения «ошибок измерения» эмпирической функции распределения может быть осуществлена с помощью обычного критерия χ^2, играющего роль «заглушки» в потенциально бесконечной цепочке применения критерия $\bar{\chi}^2$.

Строго говоря, «гладкость» функции распределения выхода генератора зависит от дискретности шкалы генерируемых чисел. Поэтому имеет смысл независимое статистическое тестирование генераторов случайных чисел.

III Модель тестирования

Для конкретности изложения модель тестирования ориентирована на генератор Фибоначчи [3]. Поэтому параметры генератора Фибоначчи, в т. ч. дискретные параметры, для других генераторов должны быть заменены специфическими параметрами этих генераторов. Для включения в модель тестирования задачи статистического оценивания [4] в настоящее время нет достаточных оснований.

Целью выполнения работ в модели тестирования является изучение зависимости критериев χ^2 и $\bar{\chi}^2$, равно как и статистических характеристик эмпирической функции распределения (6), от параметров v_1, v_2, L генератора Фибоначчи – $\mathbf{u}_1 = v_1$, $\mathbf{u}_2 = v_2$; $\mathbf{t} \equiv \mathbf{u}_1 + \mathbf{u}_2 \pmod{2^L}$, $\mathbf{u}_1 = \mathbf{u}_2$, $\mathbf{u}_2 = \mathbf{t}$. При этом выход генератора представлен как скользящая последовательность чисел $\{\mathbf{t}\}_{K,k}$, $k = 0, 1, \ldots, N-K$, из интервала [0, 1], а значения критериев вычисляются по равномерной сетке из n интервалов. "Ошибки измерения" оцениваются по относительной величине. Гипотеза об их нормальном распределении проверяется в равномерной сетке из m интервалов по критериям χ^2 и $\bar{\chi}^2$. Рассматриваемая зависимость должна быть представлена в некоторой парадигме (базе данных) первичных записей и в наглядном виде, позволяющем делать непосредственные выводы для теоретических исследований применения датчика Фибоначчи в технологии защиты данных, с возможностью дополнения журнала наблюдений характеристиками случайных последовательностей из других источников.

Круг численных экспериментов в данной модели ограничивается выяснением зависимости статистических характеристик от положения последовательности {**t**} в пирамидальной подрешётке целочисленной решётки $\{v_1\}\times\{v_2\}\times\{L\}$.

В число параметров входят логические параметры генератора: исключение внутренних симметрий, выделение шифрующих слов длиной $l=L$ битов в позиции $\lambda=0$, перестановка начальных значений, а также настроечные параметры статистики N, n, m. Скользящие последовательности полагаются идентичными исходной последовательности, т. е. $K=N$, $k=0$. Зависимости представляются в виде одно- или двухпараметрических графиков. Трёхпараметрическая зависимость представляется «движением» двухпараметрической зависимости. Форма представления может быть изменена перестановкой параметров либо демонстрацией всех проекций трёхмерного движущегося изображения. Все сложные графические изображения должны быть зафиксированы в виде кадров множества фильмов, в частности, по каждой статистической характеристике. Эти фильмы должны иметь формат, пригодный для независимого просмотра. Эксперимент должен допускать перевод в фоновый режим, приостановку, остановку и возобновление. Фильмы должны позволять одиночные обращения к базе данных по значениям параметров и после дополняющего редактирования использоваться для демонстрации хода исследования и фиксации необычных моментов.

Возможность рекурсивного применения статистик (1), (5) обеспечивается виртуальным сохранением моделей случайных величин с последующим импортом. Исходные последовательности, как правило, не сохраняются, однако могут быть воспроизведены по параметрам генерации.

В дальнейшем предполагается построение регрессионных моделей и факторный анализ. Предполагается также проверка возможности статистического оценивания (группировки) параметров v_1, v_2 в зависимости от соотношения длины используемых чисел L и длины шифрующих слов l на основании статистических характеристик предположительно имеющейся выборки шифрующих слов.

Последующие замечания относятся собственно к практическим ограничениям тестирования генератора Фибоначчи.

Выход генератора с длиной чисел $L\leq 7$ не подлежит статистическому изучению, т. к. генерируемые последовательности имеют период 2^L после исключения внутренних симметрий.

При использовании длин чисел от 8 до 15 генерируются полнопериодные последовательности, т. е. $N=2^L$ или $N=3*2^{L-1}$. Статистическое изучение таких последовательностей имеет формальный характер и относится к эмпирическому рассмотрению. Здесь выясняется влияние логических параметров генератора на статистические характеристики, настраиваются параметры статистик n, m. Проверяются относительная малость "ошибок измерения" и нормальность их распределения, а также плодотворность усложнения схемы статистического анализа рекурсивностью применения статистик.

При длине чисел, большей 16-ти бит, длина генерируемых последовательностей по умолчанию не превосходит 2^{L-8}; шаг изменения v_1,v_2 должен быть достаточно малым, чтобы графические изображения зависимостей выглядели непрерывными, причём масштаб трёхмерного изображения должен быть эмпирически управляемым в паузе между "движением" поверхности, например, $\chi^2(v_1,v_2)$ = const при возрастании L. Кроме того, должен быть обеспечен доступ к базе данных параметров и результатов по "особенным" точкам (участкам) поверхности и фоновый режим эксперимента.

IV Резюме

В отличие от классической статистики исследования ГСЧ не приводят к проблемам, связанным с величиной выборки. Возможность выполнения работ в описанной модели эмпирического измерения функций распределения обеспечена арсеналом современных средств вычислительной техники. Ясно, что столь трудоёмкие исследования целесообразно применять только к абсолютно стойким алгоритмам шифрования, одним из которых является алгоритм, использующий генератор Фибоначчи.

Литература: 1. G. Marsaglia. DIEHARD Statistical tests. – Available on http://stat.fsu.edu/~geo/diehard.html. 2. Г. Полиа, Г. Сеге. Задачи и теоремы из анализа. – Т. 1, 2. – М.: Изд-во "Наука" – 1978 г. – 392 с. – С. 94. 3. В. В. Мясоедов, Золотое сечение в шифровании данных. В зб.: Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. – Науково-технічний збірник. – Випуск 4. – К.: НДЦ "Тезіс" НТУУ "КПІ". – 2002 р. – 214 с. – С. 105. 4. И. А. Ибрагимов, Статистическое оценивание, Математическая энциклопедия. – М.: ''Советская энциклопедия''. – 1985 г. – Т. 5. – 1246 с. – С. 195.