УДК 621.34

ИССЛЕДОВАНИЕ МЕТОДОВ КРИПТОАНАЛИЗА ГЕНЕРАТОРОВ ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Ю.А. Избенко, Я.Ю. Стасева, В.П. Кукушкин
(представил д.т.н., проф. И.Д. Горбенко)

В статье рассматриваются методы криптоанализа генераторов псевдослучайных последовательностей и даются рекомендации, которые необходимо учитывать при проектировании и конструировании генераторов псевдослучайных последовательностей.

В настоящее время всё более широкое применение находят различного рода телекоммуникационные системы управления и обмена информацией. Во многих из них в силу специфики циркулирующей информации необходимо применение систем защиты информации (СЗИ). Как и всякая система, СЗИ включает в себя различные функциональные подсистемы, эффективность работы которых определяет степень соответствия СЗИ поставленным перед ней задачами. Наиболее важной является подсистема генерации псевдослучайных чисел (ПСЧ), так как успешная атака на данную подсистему делает неуместными затраченные средства и ресурсы на другие подсистемы, например, шифрования и протоколирования.

В самом общем виде подсистема генерации псевдослучайных чисел (ПСЧ) имеет вид, представленный на рис.1.

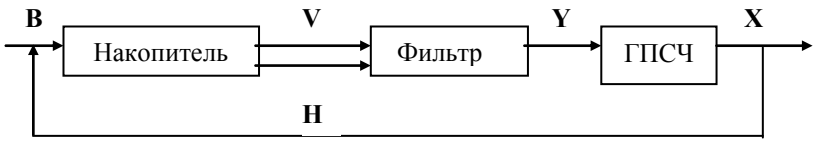

Рис. 1. Структурная схема подсистемы генерации ПСЧ

Будем полагать, что на вход накопителя будут поступать некоторые последовательности $\mathbf{B_i}$, $\mathbf{i} = \overline{\mathbf{1,m}}$, случайных чисел с распределением вероятностей, близких к равномерному распределению в интервале (0, 1), $\mathbf{B} = \{\mathbf{B_1}, \mathbf{B_2}, \ldots, \mathbf{B_m}\}$, где каждая из величин $\mathbf{B_i}$ является случайной величиной и принимает значение $\mathbf{b_i} \in \{\mathbf{0}, \ldots, \mathbf{M}\}$. Функция накопителя состоит в сборе данных последовательностей и формировании на их основе множества инициализирующих векторов $\mathbf{V} = \{\mathbf{V_1}, \mathbf{V_2}, \ldots, \mathbf{V_z}\}$ и множес-

© Ю.А. Избенко, Я.Ю. Стасева, В.П. Кукушкин, 2001

тва входных параметров ГПСЧ $\mathbf{H = \{H_1, H_2, \ldots, H_r\}}$, где каждая из величин $\mathbf{V_i}$ и $\mathbf{H_j}$ является случайной величиной, $\mathbf{i = \overline{1,Z}}$, $\mathbf{j = \overline{1,J}}$. Обычно эти значения являются результатами измерений физических процессов, взаимодействия пользователя с машиной или других внешних трудно - предсказуемых процессов. Системы реализации функций подсистемы генерации ПСЧ должны гарантировать достаточную энтропию в этих входах, чтобы сделать их непредсказуемыми для противника.

Функция фильтра состоит в отборе из поступивших множеств **V** и **H** низкоэнтропийных входных последовательностей $\mathbf{Y_i}$ во множество псевдослучайных величин $\mathbf{Y = \{Y_1, Y_2, \ldots, Y_i\}}$, $\mathbf{i = \overline{1,k}}$, $\mathbf{y_i \in \{0, \ldots, K\}}$. Отобранные $\mathbf{Y_i}$ поступают на вход генератора псевдослучайных чисел (ГПСЧ), а низкоэнтропийные последовательности отвергаются.

Под ГПСЧ будем понимать реализуемую программно или аппаратно процедуру, производящую генерацию чисел, которые, в идеале, должны быть неотличимы от истинно случайной последовательности. В рассматриваемой модели выходом генератора является псевдослучайная величина $\mathbf{X_i}$, в которой $\mathbf{x_i \in \{0, \ldots, S\}}$, $\mathbf{i = \overline{1,s}}$, $\mathbf{X = \{X_1, X_2, \ldots, X_s\}}$. При наличии атак на ГПСЧ преобразование $\mathbf{y_i \rightarrow x_i}$ необходимо осуществлять посредством некой необратимой функцией $\mathbf{f(\,)}$ [1]:

$$\mathbf{x_i = f(y_i)}. \qquad (1)$$

В этом случае

$$\mathbf{0 < H(Y/X) < H(Y)}, \qquad (2)$$

где $\mathbf{H(Y/X)}$ – условная энтропия входной последовательности при известном выходе;

$\mathbf{H(Y)}$ – энтропия входной последовательности.

Иначе, при однозначном отображении $\mathbf{y_i \rightarrow x_i}$, имеем

$$\mathbf{H(Y/X) \rightarrow 0}.$$

В связи с этим, в рамках нашей модели будем полагать, что злоумышленнику неизвестны внутренние состояния подсистемы, которые обозначим как **S**, где $\mathbf{S \in \{V, H\}}$.

Таким образом, злоумышленник в качестве исходных данных при решении задачи криптоанализа имеет выходную псевдослучайную последовательность $\mathbf{x_i}$ и не знает внутреннее состояние **S**.

Под криптоанализом будем понимать любые методы, применяемые злоумышленником, для нахождения закономерностей в выходных последовательностях, что ведёт к вскрытию закона формирования ПСП.

В ходе решения задачи криптоанализа злоумышленник может применить следующие методы анализа:

- статистический анализ;
- корреляционный анализ;
- регрессионный анализ;

- структурный анализ;
- аналитический анализ;
- метод прямого перебора.

В ходе проведения статистического анализа злоумышленник, используя существующий математический аппарат, строит эмпирический закон распределения ПСВ **X** (эмпирическую функцию распределения)

$$\mathbf{F^*(x_i) = P\{X_1 < x_1, X_2 < x_2, \ldots, X_s\}} \quad (3)$$

и проверяет его согласие с предполагаемым теоретическим законом распределения $\mathbf{F(x_i)}$ с использованием критериев согласия, а также проверяет:

1) равномерность и однородность распределения генерируемых $\mathbf{x_i}$ с использованием критериев Пирсона χ^2, Колмогорова $\mathbf{D_n}\sqrt{n}$, Мизеса ω^2;

2) случайность распределения генерируемых $\mathbf{x_i}$ с использованием различных статистических тестов.

Поскольку в этом случае злоумышленник располагает только выходными значениями ГПСЧ, то он может определить среднее количество информации $\mathbf{I(X_i, \overline{X})}$, содержащейся в известных значениях величин относительно значения $\mathbf{X_i}$, согласно формуле

$$\mathbf{I(X_i, \overline{X}) = H(X_i) - H(X_i, \overline{X})}, \quad (4)$$

где $\mathbf{\overline{X} = (X_1, X_2, \ldots, X_{i-1})}$.

Наиболее неблагоприятный случай для него - $\mathbf{I(X_i, \overline{X}) = 0}$, следовательно, необходимо обеспечить, чтобы

$$\mathbf{(X_i) = H(X_i, \overline{X}).} \quad (5)$$

Это возможно в том случае, когда значения $\mathbf{X_i}$ образуют полную группу событий и являются независимыми друг от друга

$$\mathbf{p(X_1, X_2, \ldots, X_i) = p(X_1), p(X_2), \ldots, p(X_i)}. \quad (6)$$

Кроме этого необходимо обеспечить независимость величин **Y** и **X**, т.е.

$$\mathbf{p(X_i, Y_j) = p(X_i), p(Y_j)}, \quad \mathbf{i = \overline{1,s}, j = \overline{1,k}}, \quad (7)$$

а также независимость величин $\mathbf{S \in \{V, H\}}$:

$$\mathbf{p(V_i, H_j) = p(V_i), p(H_j)}, \quad \mathbf{i = \overline{1,Z}, j = \overline{1,J}}, \quad (8)$$

поскольку величина **Y** является зависимой от **S**:

$$\mathbf{H(Y/S) = \frac{p(Y) \cdot p(S/Y)}{p(S)}}. \quad (9)$$

В ходе проведения корреляционного анализа злоумышленник в общем случае проверяет гипотезу о независимости генерируемых величин

$\mathbf{x_i}$ с использованием различных методов, например, с использованием критериев ранговой корреляции Спирмена $\boldsymbol{\rho}$, коэффициентов корреляции Пирсона **r** и критерия Пирсона $\boldsymbol{\chi^2}$. На практике данный метод анализа может быть успешно применён к последовательностям, поступившим к генераторам, комбинирующим выходы от нескольких криптографически слабых генераторов. Комбинирование выходов генераторов происходит посредством некой комбинирующей функции **g(.)**:

$$\mathbf{M[x^m] = g(\, x^z, x^s, \ldots, x^f\,) + I(x^m \backslash x^z, x^s, \ldots, x^f\,)}, \qquad (10)$$

где $\mathbf{x^z, x^s, \ldots, x^f}$ – выходы комбинируемых генераторов с длинами последовательностей **z, s, …, f** бит соответственно;

$\mathbf{M[x^m]}$ – результирующая последовательность длины $\mathbf{m = \ell_{max}(z, s, \ldots, f)}$ бит;

$\mathbf{I(x^m / x^z, x^s, \ldots, x^f)}$ –количество информации, которое несёт в себе $\mathbf{x^m}$ **о** $\mathbf{x^z, x^s, \ldots, x^f}$.

Данный метод использует уязвимые места в конструкции комбинирующей функции **g(.)**, которые позволяют по выходной последовательности $\mathbf{M[x^m]}$ получить информацию об отдельных входных последовательностях функции рассматривая $\mathbf{I(x^m / x^z, x^s, \ldots, x^f)}$.

В ходе проведения регрессионного анализа злоумышленник с использованием различных методов изучает изменение, в среднем, одной величины, например, $\mathbf{x_i}$, от остальных величин выходной последовательности, рассматривая соотношение

$$\mathbf{M[x_i / x_1, x_2, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_s\,] =}$$
$$\mathbf{= g_i(\, x_1, x_2, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_s\,) + h_i(x_1, x_2, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_s\,)}, \qquad (11)$$

где $\mathbf{h_i(x_1, x_2, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_s\,)}$ – поправочный член, характеризующий изменения, вносимые функцией преобразования (1). В результате анализа он может прогнозировать значение величины $\mathbf{x_i}$ по значениям других величин $\mathbf{(\, x_1, x_2, \ldots, x_{i-1}, \ldots, x_{i+1}, \ldots, x_s\,)}$. Вследствие этого можно прогнозировать выходную последовательность, не обладая при этом знаниями об архитектуре ГПСЧ, режимах его функционирования, внутреннем состоянии и не вскрывая закон формирования ПСП.

В ходе проведения структурного анализа злоумышленник, основываясь на выходных последовательностях, пытается вскрыть структуру ГПСЧ. Структурная скрытность $\mathbf{S_0}$ ГПСЧ определяется по формуле

$$\mathbf{S_0} = \frac{\boldsymbol{l_0}}{\mathbf{L}}\ , \qquad (12)$$

где $\boldsymbol{l_0}$ - количество бит, при перехвате которых злоумышленник может произвести успешный криптоанализ;

L – период последовательности.

Очевидно, что для достижения высокой структурной скрытности необходимо, чтобы $l_0 \rightarrow \max$.

Предполагая, что выходная последовательность содержит некоторое количество информации о внутреннем состоянии и структуре ГПСЧ, злоумышленник оперирует выходными значениями генератора, на основе которых рассматривает линейную сложность генератора.

Линейная сложность – длина L_c самого короткого регистра, который может сгенерировать конечную выходную последовательность, при условии, что первые L_c бит есть начальное заполнение регистра. На сегодняшний день существуют эффективные универсальные алгоритмы, позволяющие после изучения первых $2L_c$ бит восстановить порождающий регистр (алгоритм Берлекемпа-Месси).

Посредством аналитических методов, использование которых возможно после сбора злоумышленником всевозможной информации о предполагаемой структуре генератора, режимах его функционирования и изучения динамики его работы, злоумышленник пытается либо раскрыть закон формирования ПСП, либо построить эквивалентную модель устройства, описываемую совершенно другими соотношениями и алгоритмами. В этом случае работа злоумышленника связана с большим объёмом вычислений.

Поскольку в этом случае злоумышленник располагает выходными значениями ГПСЧ **X** и предполагаемыми входными значениями **Y**, то он может определить среднее количество информации, содержащейся во множестве значений псевдослучайной величины **X** относительно возможных значений псевдослучайной величины **Y**, согласно формуле

$$\mathbf{I(Y, X) = H(Y) - H(Y/X)} \qquad (13)$$

Наиболее неблагоприятный случай для него - $\mathbf{I(Y, X) = 0}$, т.е. мы должны обеспечить, чтобы

$$\mathbf{H(Y) = H(Y/X),} \qquad (14)$$

что возможно в том случае, когда значения $\mathbf{Y_i}$ образуют полную группу событий и являются независимыми друг от друга:

$$\mathbf{p(Y_1, Y_2, \ldots, Y_i) = p(Y_1), p(Y_2), \ldots, p(Y_i)}. \qquad (15)$$

Аналогично статистическому методу, необходимо обеспечить независимость $\mathbf{S} \in \{\mathbf{V, H}\}$.

Посредством прямого перебора, используя предполагаемые сочетания возможных внутренних и начальных состояний ГПСЧ, и комбинируя с другими методами анализа, злоумышленник пытается взломать систему «грубой силой». Метод не требует никаких знаний о ГПСЧ. Недостаток метода – большая вычислительная сложность в случае большого ко-

личества внутренних и начальных состояний ГПСЧ, а также большого периода последовательности.

Разумно будет предположить, что злоумышленнику удалось, по любой причине, – временное проникновение через систему защиты, утечка из-за небрежности персонала и т.п., заполучить внутреннее состояние $\mathbf{S_i}$ в некоторый момент времени **t**.

В этом случае он может реализовать следующие стратегии.

1. ***Атака, основанная на входе***. Применяется, когда противник способен использовать значения или контроль над входами генератора для криптоанализа, т.е. отличать выход генератора от случайной последовательности. Сопоставляя скомпрометированные входные параметры генератора с соответствующими выходными параметрами, он пытается отыскать закон формирования ПСП.

2. ***Атака на расширение***. Атака пытается расширить преимущества предварительного успешного усилия, которое восстановило **S** насколько возможно. Атака успешна, когда противник способен восстановить неизвестные выходные сигналы генератора (или отличить его выход от истинно случайных значений) до того, как **S** было скомпрометировано, или восстановить выходные сигналы после того, как генератор собрал последовательность входов, которые он может лишь предполагать. Данная атака наиболее перспективна, когда генератор стартует в опасном (предсказуемом) состоянии, имеющем недостаточную стартовую энтропию.

В результате проведённого анализа при проектировании и конструировании ГПСБ необходимо учитывать следующие общие требования:

- выходные значения генератора должны быть равномерно, однородно и случайно распределены, некоррелируемы;
- наличие высокой структурной скрытности и линейной сложности порождаемых последовательностей;
- наличие корреляционного иммунитета при использовании комбинирующих узлов;
- наличие достаточной стартовой энтропии, т.е непредсказуемых внутренних состояний.

ЛИТЕРАТУРА

1. Потий А.В. Теоретические предпосылки построения устройств формирования стойких псевдослучайных последовательностей // ИУСЖТ. – 1998. – №1. – С. 53 - 59.

2. J.Kelsey, B.Schneir, D.Wagner, C.Hall. Cryptanalytic Attack on Pseudorandom Number Generators. – *www: http : // a1.cs.engr.uky.edu / kelsey.html.*

Поступила в редколлегию 22.02.2001