УДК 621.391.1:519.2

Уривский А.В.
ОАО «ИнфоТеКС»

Аккумулятор энтропии для генератора псевдослучайных чисел

Рассмотрим задачу получения случайных чисел в компьютерной системе (КС) для криптографических приложений. Секретные ключи и синхропосылки для шифров, случайные величины для ключевых пар асимметричных криптосистем и ЭЦП, случайные запросы для криптопротоколов и т.д. выбираются, как правило, случайно и равномерно из некоторого конечного множества, что обычно реализуется путем генерирования случайных двоичных последовательностей заданной длины.

Существует два подхода к получению случайных чисел в КС. Первый заключается в использовании аппаратных датчиков, измеряющих физически непредсказуемые процессы (обычно различного рода шумы (тепловой, дробовый и т.д.) в электронных компонентах). Второй подход состоит в использовании источников энтропии (случайности) для создания стартового значения (СЗ) для запуска криптографически стойкого генератора псевдослучайных чисел (ГПСЧ). В качестве ГПСЧ применяются как специальные генераторы (например, Blum-Blum-Shub), так и криптографические примитивы (например, блочные шифры) в подходящих режимах использования.

Основные этапы создания СЗ с применением энтропийных источников – это сбор данных, их оценка, накопление и обработка для выделения нужного числа случайных бит [1]. Источниками выступают данные о работе и внутренние состояния CPU и других процессоров, памяти и различных контроллеров системной платы, системы прерываний и сетевой активности, действия оператора КС и др. [2]. Метод сбора данных специфичен для каждого источника, но обычно затруднений не представляет.

Оценка собираемых данных заключает в оценке энтропии каждого источника. Энтропию будем понимать в шенноновском смысле и оценивать (в битах на один отсчет) с помощью формулы

$$H(S) = -\sum_{i=1}^{M} p_S(i) \log p_S(i), \qquad (1)$$

где p_S – распределение вероятностей сообщений на выходе источника S с объемом алфавита M. Оценка по формуле (1) затруднена по нескольким причинам.

- Величина M бывает слишком велика, чтобы эффективно оценить распределение p_S.

Точность оценки p_S ухудшается и в виду ограниченности интервала наблюдения.

- Многие источники в КС неэргодические, и оценка $\widetilde{p_S}$, полученная на этапе создания генератора может не иметь отношения к p_S в процессе его работы. Кроме того, источники часто нестационарные и имеют память, что ведет к необходимости определения условных распределений и к их переоценке с течением времени.
- Источники могут быть статистически зависимы. Например, входящий сетевой пакет вызывает прерывание центрального процессора. Поэтому изменение сетевой статистики вызывает изменение статистики прерываний. Для зависимых источников $S_1, S_2, \ldots$ необходимо вычислять совместную энтропию $H(S_1, S_2, \ldots)$ через оценку совместного распределения вероятностей $p_{S_1 S_2 \ldots}$, что значительно усложняет анализ.
- Часть источников может контролироваться злоумышленником, постоянно или в определенные моменты времени. Например, он может симулировать высокую сетевую активность. При оценках энтропии для контролируемых источников необходимо вводить поправки для учета доступной злоумышленнику энтропии.

С учетом указанных особенностей предлагается следующий подход к оценке энтропии. Первичная оценка производится по формуле (1) в предположении, что все источники – это независимые эргодические стационарные источники без памяти. Временные эффекты учитываются оценкой p_S в скользящем временном окне: $p_S = p_S(t)$. Периодически вычисляется величина энтропии $H(S, t_j)$ для текущего распределения $p_S(t_j)$. За текущую оценку энтропии принимается $H(S, t^*) = \min\limits_{l=1,..,w} H(S, t_{j-l})$ по последним w замерам. Для учета взаимозависимости группы источников вводится эмпирическая поправка $a < 1$, так что для каждого источника в группе $H(S) = aH(S, t^*)$. Для учета действий злоумышленника используется поправочный коэффициент $b_S < 1$, определяемый из модели злоумышленника, т.е. $H(S) = ab_S H(S, t^*)$.

Еще один метод компенсации неточностей оценок энтропии источников и активности злоумышленников, понижающий скорость создания СЗ, но повышающий надежность, состоит в следующем. Для каждого источника S отдельно подсчитывается собранная энтропия: $W(S) = v(S)H(S)$, где $v(S)$ - число сообщений (отсчетов) от S. Как только для *k* из *n* источников $W(S) > W_0$, где W_0 – порог (например, битовая длина СЗ), считается, что собрано достаточно энтропии. Величина *k* определяется как характеристиками источников, так и моделью нарушителя. Очевидно, что k должно

быть больше, чем число источников, контролируемых злоумышленником, но меньше *n,* иначе остановка любого источника приведет к недостижимости критерия.

Накопление и обработка собранных данных имеет свои особенности. У источника с сообщениями большого размера может быть очень низкая энтропия, т.е. собранный массив данных, содержащий необходимое количество энтропии, будет чрезвычайно большим. Обработка всего такого массива вызывает нежелательные задержки. Удобнее обрабатывать данные в режиме поступления, и для этого лучше всего подходит криптографическая хэш-функция (КХФ).

- КХФ сжимает поступающие данные в выход фиксированной длины. Итерационный принцип работы КХФ позволяет обрабатывать входные данные в реальном времени, что не требует наличия памяти большого объема для хранения.
- КХФ реализует эффективное накопление энтропии: независимо от сложности и запутанности распределения энтропии внутри массива данных, выход зависит от всей поступившей на вход энтропии (в пределах длины выхода функции).
- Из-за высокой нелинейности КХФ контроль над частью источников не позволит злоумышленнику подавить влияние на выход остальных источников или перевести выход в некоторое вырожденное состояние.
- При компрометации СЗ однонаправленность КХФ не позволит злоумышленнику получить дополнительную информацию о неподконтрольных ему источниках.

Таким образом, КХФ как выделяет энтропию из выборок различных по свойствам источников, так и аккумулирует ее в векторе фиксированной длины.

Рассмотренные базовые принципы сбора энтропии для создания СЗ были развиты [3] и использованы для создания ГПСЧ для сетевой принтерной карты.

СПИСОК ЛИТЕРАТУРЫ

1. *Kelsey J., Schneier B., Ferguson N.* Yarrow-160: Notes on the Design and Analysis of the Yarrow Cryptographic Pseudorandom Number Generator // Selected Areas of Cryptography. – 2000. – LNCS 1758. – P.13-33.
2. *Seznec A., Sendrier N.* HAVEGE: a user-level software heuristic for generating empirically strong random numbers // ACM TOMACS. – 2003. – V. 13, N. 4. – P.334-346.
3. *Уривский А., Чмора А., Захаров С., Некрасов М., Богачов А.* Способ и устройство формирования стартового значения для генератора псевдослучайных чисел // Заявка на патент RU 2004128637. – 2004.