РАСПОЗНАВАНИЕ И ОЦЕНИВАНИЕ ПАРАМЕТРОВ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ

В. Ю. Палуха

ВВЕДЕНИЕ

Современная криптография использует генераторы случайных и псевдослучайных последовательностей [1]. При проведении испытаний средств криптографической защиты информации возникают задачи распознавания генераторов, т.е. отнесения (классификации) наблюдаемой выходной последовательности генератора $x_1, \dots, x_T \in V = \{0, 1\}$ некоторой конечной длительности T к одному из L ($2 \leq L < +\infty$) классов $\Omega_1, \dots, \Omega_L$, а также оценивания параметров генератора после того, как была установлена его принадлежность к конкретному классу.

Авторами в [2] уже был предложен подход к построению информативных признаков для распознавания генераторов. В данном докладе в качестве признаков предлагается использовать энтропийные характеристики генераторов. Что касается задачи оценивания параметров, то в данном докладе предложено аппроксимировать генераторы марковскими моделями и оценивать параметры аппроксимационной модели для построения оценок параметров исходного генератора.

1. ПОСТРОЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ НА ОСНОВЕ ПРИРАЩЕНИЯ ЭНТРОПИИ

В [3] авторами уже были построены признаки на основе энтропии в пространстве весов Хэмминга. В [4] предложено использовать приращение энтропии при увеличении длины рассматриваемого фрагмента.

Пусть наблюдается последовательность $x_1, \dots, x_T \in V = \{0, 1\}$. Энтропия Шеннона для пересекающихся фрагментов $(x_{t+1}, \dots, x_{t+n})$ длины n с распределением вероятностей $p_{i_1, \dots, i_n} = P\{x_{t+1} = i_1, \dots, x_{t+n} = i_n\}$, не зависящим от $t \in \mathbb{N} \cup \{0\}$, определяется следующей формулой [1]:

$$h(n) = - \sum_{i_1, \dots, i_n} p_{i_1, \dots, i_n} \ln p_{i_1, \dots, i_n} \quad (1)$$

Для удобства «зациклим» последовательность до длины $T + n - 1$: $x_{T+1} = x_1, \dots, x_{T+n-1} = x_{n-1}$. Обозначим: $i = \sum_{j=1}^n 2^{j-1} i_j$ – представление числа $i \in \{0, 1, \dots, 2^n - 1\}$ в двоичной системе счисления,

$p_i(n) = P\{\sum_{j=1}^n 2^{j-1} x_j = i\} = p_{i_1, \dots, i_n}, i = 0, \dots, 2^n - 1$. Построим частотные ста-

тистические оценки распределения вероятностей $\{p_i(n)\}, i = 0, \dots, 2^n - 1$:

$$\hat{p}_i(n) = \frac{1}{l} \sum_{k=1}^l \delta_{\bar{X}^{(k)}, i}, \quad \bar{X}^{(k)} = \sum_{j=1}^n 2^{j-1} x_j^{(k)}, \quad \delta_{\bar{X}^{(k)}, i} = \begin{cases} 1, & \bar{X}^{(k)} = i; \\ 0, & \bar{X}^{(k)} \neq i. \end{cases} \quad (2)$$

Используя подстановочный принцип, с помощью оценок (2) построим статистическую оценку энтропии (1):

$$\hat{h}(n) = - \sum_{i=0}^{2^n-1} \hat{p}_i(n) \ln \hat{p}_i(n). \quad (3)$$

Обозначим приращение энтропии (3)

$$g(n) = \hat{h}(n+1) - \hat{h}(n), \quad g(0) = \hat{h}(1). \quad (4)$$

Теорема. Если $T, n \rightarrow \infty, T / 2^n \rightarrow \lambda > 0$ и верна гипотеза $H_* = \{\{x_t\}$ есть РРСП}, то $g(n)$ (4) имеет асимптотически нормальное распределение с асимптотическим математическим ожиданием [1]

$$E_{H_*}\{g(n)\} = g_*(n) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k \log(k+1)}{k!} (e^{-\lambda} 2^k - 1). \quad (5)$$

Будем предполагать, что выходная последовательность генератора $x_t \in V$ является случайной последовательностью на некотором вероятностном пространстве (Ω, F, P) . Пусть наблюдается статистика $g(n)$ при различных длинах фрагмента $n \in [n_-, n_+], 0 \leq n_- < n_+$; $g_*(n) = E_{H_*}\{g(n)\}$ – математическое ожидание этой статистики при истинной гипотезе H_* . В качестве признака предлагается использовать уклонение статистики $g(n)$ от математического ожидания (5) в l_1 – метрике:

$$\rho = \frac{1}{n_+ - n_- + 1} \sum_{n=n_-}^{n_+} |g(n) - g_*(n)| \quad (6)$$

Кратко опишем проведенный эксперимент. Классы Ω_1 и Ω_2 – самосжимающиеся генераторы [5] с различными характеристическими многочленами порождающего регистра сдвига. В качестве признака использовалось значение статистики (6). Были заданы параметры $n_- = 0, n_+ = 19$. В результате применения байесовского решающего правила [6] была получена оценка безусловной вероятности ошибки 0.

2. СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ГЕНЕРАТОРА

Пусть на вероятностном пространстве $\{\Omega, \mathcal{F}, \mathbb{P}\}$ определена эргодическая цепь Маркова s -го порядка (ЦМ(s)) $x_t \in V = \{0, 1\}, t \in \mathbb{N}$ [6]:

$$P\{x_{t+1} = i_{t+1} \mid x_t = i_t, \dots, x_1 = i_1\} = P\{x_{t+1} = i_{t+1} \mid x_t = i_t, \dots, x_{t-s} = i_{t-s}\}, t > s.$$

Обозначим условное распределение вероятностей одношаговых переходов

$$p_{i_1, \dots, i_{s+1}} = P\{x_{t+s} = i_{s+1} \mid x_t = i_1, \dots, x_{t+s-1} = i_s\}, \sum_{i_{s+1}} p_{i_1, \dots, i_{s+1}} = 1, i_1, \dots, i_{s+1} \in V,$$

которое в силу однородности цепи Маркова не зависит от $t \in \mathbb{N}$.

На практике модель полностью связанной цепи Маркова высокого порядка используется редко из-за экспоненциального роста числа параметров модели (2^s) с ростом порядка s . Вместо них используются малопараметрические модели. Одной из таких моделей является цепь Маркова s -го порядка с r частичными связями (ЦМ(s, r)), предложенная в 2004 г. [7]:

$$p_{i_1, \dots, i_{s+1}} = q_{i_{m_1}, \dots, i_{m_r}, i_{s+1}}, i_1, \dots, i_{s+1} \in V, M = \{m_1, \dots, m_r\}, I = \{1, \dots, s\}, M \subseteq I.$$

Вероятности одношаговых переходов $q_{i_{m_1}, \dots, i_{m_r}, i_{s+1}}, i_1, \dots, i_{s+1} \in V$, образуют матрицу вероятностей одношаговых переходов $Q \in \mathbb{R}^{2^r \times 2}$.

Одними из самых простых в реализации и удобных в использовании генераторов являются регистры сдвига с обратной связью. Эти генераторы характеризуются состоянием $(y_1 y_2 \dots y_s)$ и функцией обратной связи $f(y_1, y_2, \dots, y_s)$, $y_i \in V = \{0, 1\}, i \in \{1, \dots, s\}$. Приведем алгоритм выработки последовательности регистром сдвига с обратной связью.

- Вход: начальное состояние $(x_0 x_1 \dots x_{s-1})$, длина T .
- Шаг 0: $(y_1 y_2 \dots y_s) := (x_1 x_2 \dots x_s)$.
- Шаг $i, i = 1, \dots, T$:
 - $x_i := y_1; c := f(y_1, y_2, \dots, y_s)$.
 - Для $j = 1, \dots, s-1: y_j := y_{j+1}$.
 - $y_s := c$.
- Выход: $x_1, \dots, x_T$.

В [8] предложен алгоритм статистического оценивания параметров модели ЦМ(s, r). Он позволяет построить оценки порядка цепи Маркова $s \in [s_-, s_+]$, числа частичных связей $r \in [r_-, r_+]$, шаблона M и матрицы Q . Вычислительная сложность алгоритма оценивания M и Q при известных r и s и длине последовательности T равна $O(2^{r+1}s^{r-1} + Ts^r)$.

Оценка матрицы вероятностей одношаговых переходов Q , построенная в результате работы алгоритма оценивания параметров ЦМ(s, r), может быть использована для построения таблицы истинности булевой

функции обратной связи. Она строится следующим образом. Матрица Q имеет размерность $2^r \times 2$. Каждая строка матрицы $(q_{i_{m_1}, \dots, i_{m_r}, 0} \ q_{i_{m_1}, \dots, i_{m_r}, 1})$ содержит вероятности генерации 0 или 1 в зависимости от предыстории $i_{m_1}, \dots, i_{m_r}$. Значение функции определяется формулой (7):

$$f(i_{m_1}, \dots, i_{m_r}) = \begin{cases} 0, & q_{i_{m_1}, \dots, i_{m_r}, 0} > q_{i_{m_1}, \dots, i_{m_r}, 1}; \\ 1, & q_{i_{m_1}, \dots, i_{m_r}, 0} \leq q_{i_{m_1}, \dots, i_{m_r}, 1}. \end{cases} \quad (7)$$

Заметим, что если мы имеем дело с генератором без искажений, то строки матрицы будут иметь вид $(0 \ 1)$ либо $(1 \ 0)$. Это значит, что вектором значений булевой функции является второй столбец матрицы Q .

Была написана программа, которая преобразует оценку матрицы Q , полученную в результате работы алгоритма оценивания параметров ЦМ(s, r), в таблицу истинности. Также была написана программа, которая восстанавливает булеву функцию по таблице истинности.

Кратко опишем проведенный эксперимент. Реализованы шесть регистров сдвига с нелинейной обратной связью с функциями из [9]. С помощью каждого из регистров получены по 5 последовательностей длины $T = 10000$, которые отличались начальным заполнением регистра. Во всех 30 случаях исходная функция была восстановлена верно.

Литература

1. Харин Ю. С., Берник В. И., Матвеев Г. В., Агиевич С. В. Математические и компьютерные основы криптологии / Минск, 2003.
2. Харин Ю. С., Палуха В. Ю. Информативные признаки для статистического распознавания криптографических генераторов // Информатика. 2013. №3. С. 126–138.
3. Палуха В. Ю., Харин Ю. С. Об информативных признаках для статистического распознавания криптографических генераторов // Информационные компьютерные технологии: проектирование, разработка, применение: сборник научных статей. Гродно, 2013. С. 121–127.
4. Rukhin A. L. Approximate Entropy for Testing Randomnesses / Интернет-адрес: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.6174&rep=rep1&type=pdf>
5. Meier W., Staffelbach O. Analysis of pseudo random sequences generated by cellular automata // Advances in Cryptology: Eurocrypt '91. Brighton, 1991. P. 186–199.
6. Харин Ю. С., Зуев Н. М., Жук Е. Е. Теория вероятностей, математическая и прикладная статистика / Минск, 2011.
7. Харин Ю. С. Цепи Маркова с r частичными связями и их статистическое оценивание // Докл. НАН Беларуси. 2004. Т. 48. № 1. С. 40–44.
8. Харин Ю. С., Петлицкий А. И. Цепь Маркова s -го порядка с r частичными связями и статистические выводы о ее параметрах // Дискретная математика. 2007. Т. 19. № 2. С. 109–130.
9. Dubrova E. A List of Maximum Period NLFSRs // Cryptology ePrint Archive, Report 2012/166 (2012). Интернет-адрес: <http://eprint.iacr.org/2012/166>