РусКрипто 20. Бабаш А.В.

НИУ ВШЭ, РЭУ им. Г.В. Плеханова.

Об одной атаке на модель шифров гаммирования

Обозначим через M(d) множество всех d-грамм содержательных (имеющих смысл) текстов в алфавите I некоторого естественного языка. На этом множестве определим бинарное отношение σ, которое назовем частичной конкатенацией. Предполагаем, что для любого $m \in M(d)$ существует $m' \in M(d)$, для которого имеет место: $m \sigma m'$. Таким образом, мы можем строить набора элементов $m_1 \sigma m_2 \sigma m_3 ... \sigma m_{nd}$ длины кратной d и набора вида $m_1 \sigma m_2 \sigma m_3 ... \sigma m_{(n-1)d} \sigma \tilde{m}_t$ длины (n-1)d+t с длиной t начального набора символов I последней d-граммы. Далее мы опускаем знак частичной конкатенации и пишем $m_1 m_2 m_3 ... m_{L-1} \tilde{m}_L$. Таким образом, нами определено множество содержательных текстов M(L), $L \geq d$ конечной длины L выбранного языка.

Пусть Z, Y, некоторые алфавиты, |Z|=|Y|=|I|. Для множества K ключей шифра используем обозначение Z^L, а для множества шифрованных текстов Y^L. Для определения функции шифрования F введем вспомогательную функцию $f : Z \times I \to Y$ **биективную по каждой переменной.**

Для $k \in K,\ x \in X,\ X = I^L,\ k = z_1 z_2 ... z_L,\ x = i_1 i_2 ... i_L$

$$F(k, x) = F_k(x) = f(z_1, i_1) f(z_2, i_2) ... f(z_L, i_L) = y_1 y_2 ... y_L$$

Функции расшифрования обозначим через

$$F^{-1}(k, x) = F_k^{-1}(x) = f^{-1}(z_1, y_1) f^{-1}(z_2, y_2) ... f^{-1}(z_L, y_L) = i_1 i_2 ... i_L .$$

Таким образом, нами определена модель шифров гаммирования.

Предположим, что M(d) – известное множество отрезков открытого текста длины d, $d \le D = vd + r,\ r < d$**, где D определено выбранным значением вероятности P(D,L) наличия по крайней мере двух одинаковых D-грамм в случайно и равновероятно выбранной последовательности символов** $z_1 z_2 ... z_L$.

Задача состоит в определении, по крайней мере, двух отрезков длины D открытого текста $m = i_1 i_2 \dots i_L$ по заданному шифрованному тексту. Предполагается, что ключи для шифрования выбираются случайно из Z^L. При решении задачи приводимый метод заканчивает свою работу успешно.

Первый этап. Упорядочим все отрезки длины D шифрованного текста $y_1, y_2, \dots y_L$: $D(j) = y_j y_{j+1} \dots y_{j+D-1}$ по увеличению индекса j. Проведем опробование №1 всех пар отрезков $(D(j_1),(D(j_2)),\ 1 \le j_1 < j_2$. Для каждой пары отрезков $(D(j_1),(D(j_2)),\ 1 \le j_1 < j_2$ будем решать систему из двух уравнений

$$\begin{array}{l} f(z_{j(1)}, i_{j(1)}) f(z_{j(1)+1}, i_{j(1)+1}) \dots f(z_{j(1)+D-1}, i_{j(1)+D-1}) = y_{j(1)} y_{j(1)+1}, \dots y_{j(1)+D-1} \\ f(z_{j(2)}, i_{j(2)}) f(z_{j(2)+1}, i_{j(2)+1}) \dots f(z_{j(2)+D-1}, i_{j(2)+D-1}) = y_{j(2)}, y_{j(2)+1}, \dots y_{j(2)+D-1} \end{array} \qquad \textbf{(1)}$$

относительно неизвестной пары отрезков $(i_{j(1)} i_{j(1)+1} \dots i_{j(1)+D-1};\ i_{j(2)} i_{j(2)+1} \dots i_{j(2)+D-1})$ открытых текстов из M(D) при известной правой части (1).

Будем искать сначала возможные решения системы (1) при D=d. Назовем ее системой уравнений (2).

Правая часть первого уравнения системы уравнений (2) известна. Проведем опробование №2 всех отрезков $\vec{\mathrm{i}} \in M(d)$. **Для каждого отрезка** $\vec{\mathrm{i}} = (\mathrm{i}(\mathrm{j}(1))\mathrm{i}(\mathrm{j}(1)+1)\ldots\mathrm{i}(\mathrm{j}(1)+d-1)$ **из первого уравнения системы (2) находим часть ключа** $\vec{z} = (z(j(1))z(\mathrm{j}(1)+1)\ldots z(\mathrm{j}(1)+d-1)$**, на которой расшифровываем набор** $y_{j(2)}y_{j(2)+1}\ldots y_{j(2)+d-1}$**. Это можно сделать в силу биективности функции f(z,i) по каждой переменной. Расшифрованная d-грамма**

$$\tilde{\mathrm{i}} = \mathrm{i}(\mathrm{j}(2))\mathrm{i}(\mathrm{j}(2)+1)\ldots\mathrm{i}(\mathrm{j}(2)+d-1)$$

проверяется на принадлежность множеству M(d). Если она принадлежит M(d), то найдено одно из решений $(\vec{\mathrm{i}},\tilde{\mathrm{i}})$ **системы (2). В противном случае решение с первой компонентой** $\vec{\mathrm{i}}$ **не существует.**

Второй этап. Если на первом этапе решений не найдено, или найдено только одно решение, то для D=vd, v>1 решений не существует. Метод закончил свою работу неуспешно. В противном случае, будем искать решения системы уравнений (1) для D=2d относительно отрезков открытых текстов длины 2d с учетом найденного множества W(d) решений длины d системы (2). Отметим, что найденные на первом этапе решения индексированы своими номерами расположений в открытом тексте

$$\vec{\mathrm{i}}(j_1)=i(\mathrm{j}_1)\mathrm{i}(\mathrm{j}_1+1)\ldots\mathrm{i}(\mathrm{j}_1+d-1);\tilde{\mathrm{i}}(j_2)=\mathrm{i}(\mathrm{j}_2)\mathrm{i}(\mathrm{j}_2+1)\ldots\mathrm{i}(\mathrm{j}_2+d-1) \qquad \mathbf{(3)}$$

и наша задача состоит в том, чтобы узнать, возможно ли их продолжение до длины 2d. Данная пара отрезков открытого текста (3) может быть продолжена только в том случае, если в множестве W(d) найдется следующее решение

$$\vec{\mathrm{i}}(j_1+d)=(i(\mathrm{j}_1+d)\mathrm{i}(\mathrm{j}_1+d+1)\ldots\mathrm{i}(\mathrm{j}_1+2d-1);\ \tilde{\mathrm{i}}(j_2+d)=\mathrm{i}(\mathrm{j}_2+d)\mathrm{i}(\mathrm{j}_2+d+1)\ldots\mathrm{i}(\mathrm{j}_2+2d-1)).$$

В случае не пустого множества решений $W(d)$ **для каждого решения из W(d) ищем возможное продолжение до длины 2d и так далее. Пусть** $W(\mathrm{v}'d)$ **последнее не пустое множество решений длины** $v'd$. **Если** $\mathrm{v}'<\mathrm{v}$, **то метод завершил свою работу.**

Третий этап. Пусть Пусть $W(\mathrm{v}d)$ последнее не пустое множество решений. Будем искать максимальное число r', при котором множество $W(vd+r')$ решений системы (1) не пустое. Если найденное ниже $\mathrm{r}'<\mathrm{r}$, то метод завершит свою работу. С этой целью для каждого решения из W(vd) ищем в этом же множестве решение $(\vec{\mathrm{i}}(j_1+vd+1);\tilde{\mathrm{i}}(j_2+vd+1))$. Если оно найдено, то данное решение может быть продолжено до решения из множества W(vd+1). При этом решения $(\vec{\mathrm{i}}(j_1+vd);\tilde{\mathrm{i}}(j_2+vd))$ и $(\vec{\mathrm{i}}(j_1+vd+1);\tilde{\mathrm{i}}(j_2+vd+1))$. имеют общую часть $(\vec{\mathrm{i}}(j_1+vd+1);\tilde{\mathrm{i}}(j_2+vd+1))\in W(vd-1)$. Продолжая итеративно процесс получения множеств решений, удлиненных на единицу, получаем в случае $\mathrm{r}'=\mathrm{r}$ искомое множество решений $W(vd+r)$. Надежность предлагаемой атаки не меньше вероятности P(D,L).

Трудоемкость метода не превышает величины

$$\frac{(\mathrm{L}-D+1)(\mathrm{L}-D)}{2}|M(d)|+\sum_{c=2}^{v+1}\left(\frac{(\mathrm{L}-D+1)(\mathrm{L}-D)|M(d)|^2}{2|I|^d}\right)^c+\sum_{c=1}^{r}\left(\frac{(\mathrm{L}-D+1)(\mathrm{L}-D)|M(d)|^2}{2|I|^d}\right)^{v+c}$$

- Вопросы?