Бабаш А.В.
НИУ ВШЭ, РЭУ им. Г.В. Плеханова.

Дешифруем или недешифруем шифр случайного гаммирования

Для описания ШСГ фиксируем алфавит I естественного языка, алфавит K ключа и шифрованного текста Y. Положим $I = K = Y = Z / s$ - кольцо положительных вычетов по модулю s. Обозначим через $M(d) \subset I^d$ множество d-грамм содержательных текстов в алфавите I. Предполагаем, что на $M(d)$ заданы правила конкатенации d-грамм для получения содержательных (читаемых) текстов произвольной длины. Через $i + \gamma = y \bmod s$ обозначим функцию шифрования ШСГ для букв $i \in I, \gamma \in K, y \in Y$.

Для шифрования открытого текста $\Im_j^v = i_j^v i_{j+1}^v ... i_{j+d-1}^v$ из I^d выбирается случайно и равновероятно ключ $\Gamma_j^v = \gamma_j^v \gamma_{j+1}^v ... \gamma_{j+d-1}^v$ проводится операция $\Im_j^v + \Gamma_j^v = U_j^v = y_j^v y_{j+1}^v ... y_{j+d-1}^v$, где $i_{j+k}^v + \gamma_{j+k}^v = y_{j+k}^v \bmod s, \quad k \in \{0,1,...d-1\}$.

Обозначение операции обратного элемента кольца перенесем и на $(Z/s)^d$. Основой атаки является уравнение $\Im_j^v - \Im_{j'}^v = U_j^v - U_{j'}^v + q$ с тремя неизвестными, где $q = \Gamma_{j'}^v - \Gamma_j^v$.

Совершенность шифра

x – открытый текст, k – ключ, y – шифрованный текст

P(x/y)=P(x) для любых x, y - СОВЕРШЕННОСТЬ по нападению на открытый текст (К.Шеннон)

P(к/y)=P(x) для любых x, y - СОВЕРШЕННОСТЬ по нападению на ключ

Рассмотрим модель ШСГ с указанными алфавитами для шифрования содержательных текстов M={0,1}. Пусть команда 0 поступает на шифр с вероятностью p(0), а команда 1 с вероятностью $p(1) \neq p(0)$. Легко проверить, что этот шифр является совершенным по нападению на открытый текст и, следовательно, недешифруем. И несовершенным по ключу и, следовательно, дешифруем

Определение. Модель шифра является дешифруемой, если для нее найдена атака определения открытого текста по перехвату шифрованного текста с конечной трудоемкостью и надежностью, превышающей надежность метода угадывания. В противном случае, модель шифра называется недешифруемой.

Атака

Предварительный этап атаки на ШСГ. Разобьем множество упорядоченных пар элементов из $M(d)^2$ на классы $[\Im - \Im']_a$ упорядоченных пар $\Im, \Im'$ с фиксированной разностью $\Im + (-\Im') = a$, $a \in (Z / s)^d$. Через A обозначим множество индексов не пустых классов $[\Im - \Im']_a$.

Атака. Пусть неизвестный содержательный текст $i_1 i_2 ... i_L$ зашифрован на ШСГ неизвестным ключом $\gamma_1 \gamma_2 ... \gamma_L$ и получен известный шифрованный текст $y_1 y_2 ... y_L$. Задача состоит в однозначном определении какой ни будь одной пары D-грамм $\Im_j = i_j i_{j+1} ... i_{j+D-1}$, $\Im_{j'} = i_{j'} i_{j'+1} ... i_{j'+D-1}$ содержательного зашифрованного текста $i_1 i_2 ... i_L$ по шифрованному тексту $y_1 y_2 ... y_L$. Положим $2D \le L$, $D = vd + r, 0 \le r < d$, d>1. Фиксируем $q \in I^d$ и его начало $\tilde{q}$ длины r.

Шаг 1. Для каждой пары номеров j, j', $j < j'$ d-грамм $U_j, U_{j'}$ шифрованного текста проводим последовательность действий (1): $j, j' \Rightarrow b_{j,j'} \Rightarrow [j, j']_b \leftrightarrow [\Im - \Im']_{a=b_{j,j'}}$, которая означает: для пары j, j' вычисляется $b_{j,j'} = U_j - U_{j'} + q$; по значению $b_{j,j'}$ находится класс $[\Im - \Im']_a$ индексированный $a = b_{j,j'}$ и класс $[j, j']_b$ пар j, j' с данным значением $b = b_{j,j'}$. Обозначим через B множество возможных значений $b_{j,j'}$.

Положим $A_d(J,J') = \bigcup_{b\in A\cap B}[j,j']_b$ и через $A_w(J,J')$ обозначим множество пар позиций с найденными на них возможными содержательными текстами длины w. Предполагается, что в дальнейшем мы используем лишь такие пары индексов. Очевидно, что $|A_d(J,J')| \le \frac{(\mathrm{L}-d+1)(\mathrm{L}-d)}{2} = T_1$. Трудоемкость шага 1 есть число операций последовательных действий по цепочкам, оно равно T_1.

Шаг 2. В последовательности действий (1) для активных пар позиций j, j' имеются и соответствия:

$$j+d, j'+d \Rightarrow b_{j+d,j'+d} \Rightarrow [j, j']_{b_{j+d,j'+d}} \leftrightarrow [\Im - \Im']_{a=b_{j+d,j'+d}}.$$

Получение возможных 2d-грамм содержательных текстов для каждого $b \in A \cap B$ и каждой пары $(j, j') \in [j, j', q]_b$ формализуем цепочкой соответствий

$$b \Rightarrow j, j' \Rightarrow bb_{j+d,j'+d} \Rightarrow [\Im - \Im']_b [\overline{\Im} - \overline{\Im}']_{b_{j+d,j'+d}}.$$

Последняя компонента цепочки есть обозначение класса 2d-грамм полученных правилами конкатенации элементов из $[\Im - \Im']_b$ с элементами из $[\overline{\Im} - \overline{\Im}']_{b_{j+d,j'+d}}$.

Трудоемкость получения 2d-грамм не больше чем

$$\sum_{b\in A\cap B_1}\ \sum_{(j,j')\in[j,j',q]_b} |[\Im-\Im']_b|\cdot|[\bar{\Im}-\bar{\Im}']_{b_{j+d,j'+d}}|\le$$

$$\le| A_d(J,J')||M(d)|^2$$

конкатенаций и в среднем левая часть не превосходит

$$A_d(J,J')\cdot(T_1\cdot\frac{1}{|I|^d})^2.$$

Продолжая аналогично будут найдены все возможные пары vd-грамм содержательного текста $\tilde{\tilde{\mathfrak{I}}}, \tilde{\tilde{\mathfrak{I}}}'$, начинающихся с позиций $(j, j') \in A_{vd}(JJ')$, либо метод закончит работу раньше не найдя приемлемых конкатенаций. Трудоемкость получения пар vd-грамм открытого текста $\tilde{\tilde{\mathfrak{I}}}, \tilde{\tilde{\mathfrak{I}}}'$ не превосходит $(v-1)(|A_{vd}(J,J')| \cdot |M(d)|^2)$ конкатенаций.

Шаг 3. Этот шаг аналогичен шагу 2. Его трудоемкость не больше $r \cdot (|A_{vd}(J, J')| \cdot |M(d)|^2)$ конкатенаций. Трудоемкость всей атаки не превосходит величины $T_1 + (v + r - 1) \cdot (|A_d(J, J')| \cdot |M(d)|^2$ в случае заданного конкретного шифртекста. а в среднем, не больше

$$T_1 + (vd + r - 1) \cdot |A_d(J, J')| \cdot (T_1 \cdot \frac{1}{|I|^d})^2 \leq$$

$$\leq T_1 \cdot (1 + (vd + r - 1) \cdot (T_1 \cdot \frac{1}{|I|^d})^2).$$

В результате атаки будет найдена по крайней мере одна пара индексов j, j' с множеством возможных пар содержательных текстов содержащим истинную пару $\Im_j, \Im_{j'}$ D-грамм. Надежность атаки равна вероятности наличия двух D-грамм $\Gamma_j, \Gamma_{j'}, j < j'$ в ключе $\Gamma_{j'} - \Gamma_j = qq...q\tilde{q}.$ Для $q = 0$ эта вероятность известна как вероятность $P(k = L - d + 1, n = \mid I \mid^D)$ двух одинаковых дней рождений.

$$P_D \le \prod_{c=0}^{D-d+1} P(b_{j+c,j'+c} \in A) \le \left(\frac{|A|}{|I|^d}\right)^{D-d+1}$$

вероятность ложного решения