

Формальные модели шифров (К. Шеннон)

Шеннон К. Теория связи в секретных системах
// В кн.: Работы по теории информации и кибернетике.
– М.:ИЛ, 1963.

Обозначения:

X – конечное множество возможных *открытых текстов*

K – конечное множество возможных *ключей*

Y – конечное множество возможных *шифрованных текстов*

$E_k: X \rightarrow Y$ – правило зашифрования на ключе $k \in K$

$\{E_k: k \in K\}$ – обозначим через E

$\{E_k(x) : x \in X\}$ – обозначим через $E_k(X)$

$D_k: E_k(X) \rightarrow X$ – правило расшифрования на ключе $k \in K$

$\{D_k: k \in K\}$ – обозначим через D

Определение 1

Шифром (шифрсистемой) назовем совокупность

$$\Sigma_A = (X, K, Y, E, D)$$

введенных множеств, для которых выполняются следующие свойства:

1) для любых $x \in X$ и $k \in K$ выполняется равенство $D_k(E_k(x)) = x$;

2) $Y = \bigcup_{k \in K} E_k(X)$.

Σ_A – алгебраическая модель шифра

$P(X)$, $P(K)$ – априорные распределения вероятностей на множествах X и K соответственно.

Т.е.

для любого $x \in X$ определена вероятность $p_X(x) \in P(X)$,

для любого $k \in K$ определена вероятность $p_K(k) \in P(K)$,

причем выполняются равенства

$$\sum_{x \in X} p_X(x) = 1$$

и

$$\sum_{k \in K} p_K(k) = 1$$

$$\Sigma_B = (X, K, Y, E, D, P(X), P(K))$$

Σ_B – вероятностная модель шифра

Замечание

В большинстве случаев множества X и Y представляют собой объединения декартовых степеней некоторых множеств A и B соответственно:

$$X = \bigcup_{i=1}^L A^i, \quad Y = \bigcup_{i=1}^{L_1} B^i$$

A – алфавит открытого текста

B – алфавит шифрованного текста

Пусть $X = Y = \bigcup_{i=1}^L A^i$, $K \subseteq S(A)$,

где $S(A)$ – симметрическая группа подстановок множества A .

Для любого ключа $k \in K$, открытого текста $x=(x_1, x_2, \dots, x_l)$ и шифрованного текста $y=(y_1, y_2, \dots, y_l)$ правила зашифрования и расшифрования шифра простой замены в алфавите A определяются формулами:

$$E_k(x) = (k(x_1), k(x_2), \dots, k(x_l)),$$

$$D_k(y) = (k^{-1}(y_1), k^{-1}(y_2), \dots, k^{-1}(y_l)),$$

где k^{-1} – подстановка,
обратная к k .

Замечание:

В общем случае для шифра простой замены $X = \bigcup_{i=1}^L A^i$, $Y = \bigcup_{i=1}^L B^i$,
причем $|A| = |B|$, а K представляет собой множество биекций
множества A на множество B .

(k^{-1} – биекция, обратная к k)

Пусть $X = Y = A^L$, и пусть $K \subseteq S_L$

где S_L – симметрическая группа подстановок множества $\{1, 2, \dots, L\}$.

Для любого ключа k , открытого текста $x = (x_1, x_2, \dots, x_L)$ и шифрованного текста $y = (y_1, y_2, \dots, y_L)$ правила зашифрования и расшифрования *шифра перестановки* в алфавите определяются формулами:

$$E_k(x) = (x_{k(1)}, x_{k(2)}, \dots, x_{k(L)}),$$

$$D_k(y) = (y_{k^{-1}(1)}, y_{k^{-1}(2)}, \dots, y_{k^{-1}(L)}),$$

где k^{-1} – подстановка, обратная к k .

Пусть $n = pq$, где p и q – простые числа.

Пусть $X = Y = Z_n$ – кольцо вычетов по модулю n .

Положим $K = \{(n, p, q, a, b) : a, b \in Z_n, n = pq, ab \equiv 1 \pmod{\varphi(n)}\}$,

где φ – функция Эйлера.

Представим ключ $k \in K$ в виде $k = (k_z, k_p)$, где $k_z = (n, b)$ и $k_p = (n, p, q, a)$ – ключи зашифрования и расшифрования соответственно.

Правила зашифрования и расшифрования шифра RSA определим для $x \in X$ и $y \in Y$ формулами:

$$E_{k_z}(x) = x^b \pmod{n},$$

$$D_{k_p}(y) = y^a \pmod{n}.$$

Математические модели открытого текста

«Вероятностная модель k -го приближения»:

Пусть $P^{(k)}(A)$ представляет собой массив, состоящий из приближений для вероятностей $p(b_1 b_2 \dots b_k)$ появления *k -грамм* $b_1 b_2 \dots b_k$ в открытом тексте, $k \in N$, $A = \{a_1, \dots, a_n\}$ – алфавит открытого текста, $b_i \in A$, $i = \overline{1, k}$.

$c_1, c_2, \dots, c_k, c_{k+1}, \dots$ – посл-сть знаков алфавита A , которую генерирует источник «*открытого текста*», в которой:

$p(c_1 c_2 \dots c_k) \in P^{(k)}(A)$ – вер-сть появления *k -граммы* $c_1 c_2 \dots c_k$

$p(c_2 c_3 \dots c_{k+1}) \in P^{(k)}(A)$ – вер-сть появления *k -граммы* $c_2 c_3 \dots c_{k+1}$

...

«Вероятностная модель 1-го приближения»:

(позначная модель открытого текста)

c_1, c_2, \dots

$p(c_i) \in P^{(1)}(A)$ — вероятность появления знака c_i , $i = 1, 2, \dots$

$p(c_1 c_2 \dots c_l) = \prod_{i=1}^l p(c_i)$ — вероятность появления текста
 $c_1 c_2 \dots c_l$.

(Каждый знак появляется независимо от других знаков)

«Вероятностная модель 2-го приближения» :

(пространство элементарных исходов Марковского)

c_1, c_2, \dots

$p(c_1) \in P^{(1)}(A)$ – вероятность появления знака c_1 ,

$p(c_i/c_{i-1}) = \frac{p(c_{i-1}c_i)}{p(c_{i-1})}$ – вероятность появления знака

$c_i, i = 2, 3, \dots,$

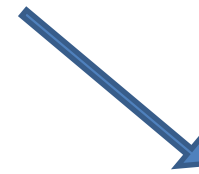
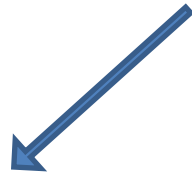
где $p(c_{i-1}c_i) \in P^{(2)}(A), p(c_{i-1}) \in P^{(1)}(A), i = 2, 3, \dots$

$p(c_1c_2 \dots c_l) = p(c_1) \prod_{i=2}^l p(c_i/c_{i-1})$ – вероятность

появления текста $c_1c_2 \dots c_l$.

(Каждый следующий знак зависит от предыдущего)

Критерии распознавания открытого текста



Стандартные методы различения статистических гипотез

$$A = \{a_1, \dots, a_n\},$$

$$P(A) = (p(a_1), \dots, p(a_n)),$$

$c_1 c_2 \dots c_l$ – является ли *открытым текстом* ?

H_0 – гипотеза о том, что данная последовательность – *открытый текст*,

H_1 – альтернативная гипотеза (послед-сть – *случайная равновероятная*).

Применяется *наиболее мощный критерий* различения двух простых гипотез, который дает *лемма Неймана-Пирсона*.

Ош. 1-го рода: $\alpha = p\{H_1/H_0\}$!!! *min* !!!

Ош. 2-го рода: $\beta = p\{H_0/H_1\}$

Наличие в открытых текстах некоторых запретов (критерий запретных k -грамм)

Отбирается некоторое число S редких k -грамм, которые объявляются запретными. Просматривая последовательно k -грамму за k -граммой анализируемой последовательности $c_1 c_2 c_3 \dots c_l$, объявляем ее случайной, как только в ней встретится одна из запретных k -грамм, и открытым текстом в противном случае.

...