Лекция 10. Статистические модели

Цель лекции: рассмотреть один из универсальных методов криптоанализа.

План лекции:
Введение.
1 Статистические модели
Заключение
Контрольные вопросы

Ключевые слова: [выбрать самостоятельно].

Содержание лекции:

Введение

1 Статистические модели

Для того, чтобы оценить реализуемость метода, рассмотрим следующую вероятностную модель. Рассмотрим набор случайных величин

$$A = \{a, b_{ij}, \quad i = \overline{1,m}, j = \overline{1,t}\},$$

удовлетворяющих системе уравнений:

$$a + \sum_{j=1}^{t} b_{ij} = 0, \;\; i = \overline{1,m}.$$

Замечание 1. *В модели, рассмотренной в лекции 8, t = 2.*

Аналогично рассмотрим набор случайных величин

$$Z = \{z, y_{ij}, \quad i = \overline{1,m}, j = \overline{1,t}\},$$

Эти случайные величины представляют z_n. Два набора связаны следующим соотношениями

$$P(z = a) = p, \;\; P(b_{ij} = y_{ij}) = p.$$

Замечание 2. *Такие соотношения можно получить следующим образом. Так, например, в модели (2)*

$$z = a + \gamma,$$
$$y_{ij} = b_{ij} + \gamma_{ij},$$

где γ – независимые, одинаково распределенные случайные величины с P(γ = 0) = p.

Обозначим

$$b_i = \sum_{j=1}^{t} b_{ij}, \;\; i = \overline{1,m},$$

$$y_i = \sum_{j=1}^{t} y_{ij}\,,\ \ i = \overline{1,m}.$$

Положим

$$L_i = z + y_i, \quad i = 1, \ldots, m.$$

Пусть вероятность $s(t, p) = s = P(y_i = b_i)$ не зависит от i. По формуле полной вероятности получим рекурренту

$$s(t, p) = ps(t - l, p) + (l - p)(l - s(t - 1, p)),$$
$$s(1, p) = p.$$

Замечание 3. *Для t =2* $s(2, p) = p^2 + (l - p)^2$.

Обозначим через B_k события, состоящие в том, что k из m линейных форм L_i равны 0. Тогда следующий вывод «z = a» определяется апостериорной вероятностью

$$P(z = a|B_k) = \frac{\binom{m}{k} p s^k (1-s)^{m-k}}{\binom{m}{k} p s^k (1-s)^{m-k} + \binom{m}{k}(1-p) s^{m-k}(1-s)^k} = p^*.$$

Это апостериорная вероятность того, что «z = a». Аналогично

$$P(z \neq a|B_k) = \frac{\binom{m}{k} s^{m-k}(1-s)^k (1-p)}{\binom{m}{k} p s^k (1-s)^{m-k} + \binom{m}{k}(1-p) s^{m-k}(1-s)^k} = 1 - p^*.$$

Идея состоит в том, что мы интуитивно ожидаем, что p^* увеличивается по сравнению с p, если z = a и уменьшается, если $z \neq a$. Поэтому найдем математическое ожидание p^* в двух случаях z = a и $z \neq a$. При z = a

$$E_0(p^*) = E(p^*|z = a) =$$
$$= \sum_{k=0}^{m} \binom{m}{k} \frac{p s^k (1-s)^{m-k}}{p s^k (1-s)^{m-k} + (1-p)(1-s)^k s^{m-k}} s^k (1-s)^{m-k}\,.$$

Далее при $z \neq a$

$$E_1(p^*) = E(p^*|z \neq a) =$$
$$= \sum_{k=0}^{m} \binom{m}{k} \frac{p s^k (1-s)^{m-k}}{p s^k (1-s)^{m-k} + (1-p)(1-s)^k s^{m-k}} s^{m-k} (1-s)^k\,.$$

При p = 3/4, t =2, m = 20 получим

$$E_0(p^*) = 0{,}9;$$
$$E_1(p^*) = 0{,}3.$$

Осталось оценить число допустимых соотношений m как функции от длины регистра r и длины текста N. Пусть t-членное соотношение получено с использованием равенства

$$(C(x))^j = C(x^j), \quad j = 2^i, \forall\, i = 1, ..., m.$$

Тогда длина задействованного участка при i = m равна $r2^m$. Таких соотношений $N - r^2$ m > 0. Тогда общее число соотношений равно

$$T = \sum_{m=0}^{\log_2 \frac{N}{r}} (N - 2^m r) = N\left(\log_2 \frac{N}{r} + 1\right) - \sum_{m=0}^{\log_2 \frac{N}{r}} 2^m r =$$
$$= N\left(\log_2 \frac{N}{r} + 1\right) - \left(2^{\log_2 \frac{N}{r} + 1} - 1\right) r = \quad (1)$$
$$N \log_2 \frac{N}{r} + N - \left(\frac{2N}{r} - 1\right) r =$$
$$= N \log_2 \frac{N}{2r} + r - N.$$

Каждое соотношение (1) связано с t + 1 знаками последовательности z. Поэтому среднее число m соотношений на один знак равно

$$m = \frac{T(t+1)}{N} = \left(\log_2 \frac{N}{2r} + \frac{r}{N} - 1\right)(t+1).$$

Для наших приложений (r/N)(t + 1) << 1. Отсюда из формулы (1) получим приближенное равенство

$$m \approx (t+1) \log_2 \frac{N}{2r}.$$

Заключение

Контрольные вопросы

Смотри руководство по организации самостоятельной работы магистрантов.