Криптография, Лекция № 5

6 октября 2014 г.

Доведем рассуждения с прошлой лекции. В общих чертах: генератор случайных числе - функция, котора берет короткое число независимых случайных битов и выдает длинную строку, чтобы отличить которую нужно затратить много ресурсов.

Theorem 1. *(Голдрайх-Левин)*
Если $f : \{0,1\}^n \to \{0,1\}^n$ - односторонняя перестановка, то $g : \{0,1\}^{2n} \to \{0,1\}^{2n}$, $g(x,y) = (f(x), y)$ - тоже односторонняя перестановка, а $h(x,y) = x \odot y$ - трудный бит для g.

Theorem 2. *(О кодах Адамара или о списочном декодировании)*
Пусть $h_x(y) = x \odot y$ - код Адамара слова $x \in \{0,1\}^n$. Пусть $\bar{h}(y)$ - такая функция, что доля y удовлетоворяющих $\bar{h}(y) = h_x(y)$ больше либо равна $\frac{1}{2}$. Тогда существует вероятностный алгоритм, работающий $poly(\frac{n}{\varepsilon})$, имеющий произвольный доступ к $\bar{h}$, выдающий список слов длины n, с вероятностью $\geq \frac{1}{2}$ содержит x.

Почему это называется декодированием списка? Вообще кодирование нужно для исправления ошибок. Код Адамара - длинный код. Точного полиномиального алгоритма декодирования не существует. h - код, $\bar{h}$ - испорченный код.

Theorem 3.
Из теоремы о списочном декодировании выводится теорема Голдрайха-Левина.

Доказательство.
Пусть существует схема, которая с вероятностью $\geq \frac{1}{2}+\varepsilon$ по $(f(x), y)$ находит $x \odot y$.

$$Pr_y\{C(f(x), y) = x \odot y\} \geq \frac{1}{2} + \varepsilon$$

$$Pr_y\{\bar{C}(y) = h_x(y)\} \geq \frac{1}{2} + \varepsilon$$

$\bar{C}$ - искаженный код Адамара слова x.

Обратитель f:

1. Применяет алгоритм списочного декодирования для $\bar{h} = \bar{C}$ и ε

2. Вычисляет значения f на всех элементах полученного списка

3. Если $f(\bar{x}) = f(x)$, то выводит $\bar{x}$, иначе выводит что угодно.

Утверждается, что этот обратитель будет достаточно успешным. Он точно будет работать полиномиальное время (он обращается к схеме, нужно переделать алгоритм в полиномиальную схему). Нужно понять, с какой вероятностью он будет обращать. Эта вероятность будет равна $\frac{\varepsilon}{4}$. Ибо если декодирование успешно, то нужный x будет найден и вероятность успеха декодирования будет $\frac{1}{2}$. Нужно посчитать долю x, для которых

$$Pr_y\{\bar{C}(y) = h_x(y)\} \geq \frac{1}{2} + \frac{\varepsilon}{2}$$

Эта доля $\geq \frac{\varepsilon}{2}$, иначе

$$Pr_y\{C(f(x), y) = h_x(y)\} < \frac{\varepsilon}{2} \cdot 1 + 1 \cdot (\frac{1}{2} + \frac{\varepsilon}{2}) = \frac{1}{2} + \varepsilon$$

Получили противоречие с предположением.

$\geq \frac{\varepsilon}{2}$ - вероятность хорошего x;

$\geq \frac{1}{2}$ - вероятность успешного обращения хорошего x

Из последних двух оценок поулчаем $\geq \frac{\varepsilon}{4}$ - вероятность успешного обращения x. □

Lemma 1. *(ЗБЧ для попарно независимых случайных величин)*
Пусть $\xi_1, \ldots, \xi_N$ - попарно независимые одинакого распределенные, бернулиевские случайные величины. $E\xi_i = \frac{1}{2} - \varepsilon$. Тогда

$$Pr\{\xi_1 + \ldots \xi_N \geq \frac{1}{2}N\} \leq \frac{1}{\varepsilon^2}\frac{1}{N}$$

Доказательство.
Следствие попарной независимости:

$$D(\xi_1 + \ldots + \xi_N) = D\xi_1 + \ldots + D\xi_N$$

$$E(\xi_1 + \ldots + \xi_N) = (\frac{1}{2} - \varepsilon)N$$

Применим неравенство Чебышёва:

$$Pr\{\xi_1 + \ldots + \xi_N \geq E(\xi_1 + \ldots + \xi_N) + \varepsilon N\} \leq \frac{D(\xi_1 + \ldots + \xi_N)}{(\varepsilon N)^2} \leq \frac{N \cdot \frac{1}{4}}{(\varepsilon N)^2} = \frac{1}{4N\varepsilon^2}$$

□

Доказательство. (теоремы о списочном декодировании)

$$h_x(y) = h_x(y + r) + h_x(r)$$

Отсюда еще можно заметить, что если бы было испорчено не более четверти битов, то можно было $h_x(y), h_x(r)$ заменить на $\bar{h}$ и выбирать максимум.

$$h_x(y) = majority_r(\bar{h}(y+r) + h_x(r))$$

Можно взять выборку из попарно независимых r_i. Как их построить? Есть небольшое количество случайных вектор-строк длины n: $u_1, \ldots, u_s$.
$r_1, \ldots, r_{2^s-1}$ - все нетривиальные суммы u_j. Они будут попарно независимыми равномерно распределенными. Представим r_i в виде вектор-столбца R. $R = A \cdot U$.

$h_x(u_1 + u_2 + u_s) = h_x(u_1) + h_x(u_2) + h_x(u_s)$. То есть $h_x(r_i)$ полностью определяется $h_x(u_1), \ldots, h_x(u_s)$.

k-ый бит x: $x_k = h_k(e_k) = x \odot e_k$

Алгоритм восстановления x (списком):

1. Для всех строк длины $b \in \{0,1\}^s$
2. Для всех $k = 1, \ldots, n$

   $$x_k = majority|_{i=1,\ldots,2^s-1}(\bar{h}(e_k + r_i) + h_x(r_i))$$

   где $h_x(r_i)$ посчитан при условии $h_x(u_1) = b_1, \ldots, h_x(u_s) = b_s$
3. Добавить x в список.

Введем случайные перменные ξ_i - индикатор $\bar{h}(y + r_i) = h_x(y + r_i)$. r_i - случайные, следовательно $Pr\{\xi_i = 1\} \geq \frac{1}{2} + \varepsilon$. Посчитаем вероятность плохого исхода: $\leq \frac{1}{4\varepsilon^2} \cdot \frac{1}{2^s-1} \cdot n$. Хотим, чтобы последняя величина была меньше $\frac{1}{2}$. Откуда $s \simeq \log \frac{n}{\varepsilon^2}$. Весь алгоритм, как легко проверить, полиномиальный. $\square$