Лекция 1. История шифрования. Понятие абсолютной стойкости. Шифр одноразового блокнота.

14 сентября 2020 г.

История криптографии насчитывает около 4 тысяч лет, и до 70х годов 20 века она занималась вопросами шифрования и расшифровки сообщений, преимущественно в дипломатической и военной областях.

Первейшей задачей было обеспечить недоступность содержимого сообщения для перехвата. С этой целью установленные участники обмена информацией (скажем, штаб и войско в поле) заранее вырабатывают общий секрет, который называется ключом. Затем сообщение шифруется ключом на одной стороне, получается шифротекст, который с помощью ключа расшифровывается на другой. Без знания ключа неприятелю, который перехватил сообщение, должно быть невозможно его расшифровать (в идеале).

Рассмотрим несколько исторических шифров:

1. Шифр Цезаря

Берем алфавит, и заменяем букву той, что идет третьей за ней (a -> D), (b -> E), ..., (x -> A), (y -> B), (z -> C), если в русском алфавите, то (а -> Г) и.т.д.

Пример: расшифруйте “ГХГНСЕГХЯ”.

2. Шифр сдвига (shift cipher):

Обобщение шифра Цезаря (не имеющего ключа). Размер сдвига теперь определяется ключом k. Для шифра Цезаря k = 3.

Шифр сдвига тривиально ломается перебором. Ясно, что маленькое пространство ключа является проблемой. Что мы можем сделать?

3. Шифр замены (подстановки)

Мы заменяем букву алфавита любой другой согласно таблицы замены (скажем, (a -> X), (b -> M), ...). Ключом является таблица замены. Количество возможных комбинаций для русского алфавита $33! \approx 2^{122}$. Однако, шифр тривиально взламывается частотным анализом.

4. Полиалфавитные шифры. Шифр Виженера.

ключ - это слово длиной N, и каждый mod N символ открытого текста сдвигается на соответствующее символу ключа значение. Взламывается тривиально, если длина шифротекста много больше длины ключа.

Принцип Керкгоффса (из книги “Военная криптография» (издана в 1883 году)”) - секретным является лишь ключ, а не алгоритм.

Для того, чтобы выработать идеальный шифр, введем определания.

Мы будем описывать шифр тремя алгоритмами, генерации ключа, шифровки текста и расшифровки шифротекста:

- $k \leftarrow Gen()$
- $c := Enc_k(m)$
- $m := Dec_k(m)$

Мы хотим иметь абсолютную корректность шифра: $Dec_k(Enc_k(m)) = m$

Как теперь определитиь стойкость шифра?

Предположим, что штаб может послать войску лишь одно из двух сообщений: “атаковать” или “отступать” (оба, между прочим, состоят из 9 букв!), и это известно неприятелю. Также неприятелю известны вероятности появления сообщений, скажем, $Pr[m =$ “атаковать” $= 0.6]$, $Pr[m =$ “отступать” $= 0.4]$. Мы хотим, чтобы неприятель не получил из изучения шифротекста ничего нового.

Более формально, атакующий не может извлечь из шифротекста никакой информации о сообщении, которая не была известна априори (до перехвата шифротекста).

Что при этом может делать атакующий?

- Атака на основе шифротекста (ciphertext-only attack) - все, что есть у атакующего, это шифротексты.
- Атака на основе открытых текстов (Known-plaintext attack) - атакующий способен узнать одну или более пар сообщение-шифротекст, с целью получить информации о сообщениях, зашифрованных в других шифротекстах (тем же ключом).
- Атака на основе подобранного открытого текста (Chosen-plaintext attack, CPA) - атакующий способен выбрать текст и получить соответствующий шифротекст.
- Атака на основе подобранного шифротекста (Chosen-ciphertext attack, CCA) - атакующий способен выбрать шифротекст для расшифровки, с целью извлечь информацию о сообщениях, зашифрованных в других шифротекстах (тем же ключом).

Определим теперь безопасность шифра еще более формально, для атак на основе шифротекста (атак перехвата).

Зададим пространство ключей буквой $\overline{K}$, пространство сообщений буквой $\overline{M}$, пространство шифротекстов буквой $\overline{C}$.

Без потери общности предполагаем, что $Gen()$ возвращает случайный результат (с равномерным распределением), $k \leftarrow \overline{K}$, тогда $\overline{K}$ и $\overline{M}$ независимы.

0.1 Абсолютная стойкость

Определение: схема шифрования (Gen, Enc, Dec) является абсолютно стойкой, если для любого распределения в $\overline{M}$, любого сообщения m из $\overline{M}$, и любого шифротекста $c \in \overline{C}$, для которого $Pr[C = c] > 0$: $Pr[M = m|C = c] = Pr[M = m]$

Альтернативная формулировка:

$$Pr[Enc_K(m) = c] = Pr[Enc_K(m') = c]$$

Теорема 1: схема шифрования (Gen, Enc, Dec) является абсолютно стойкой, если и только если для любого распределения в , любых сообщений m, m' из $\overline{M}$, и любого шифротекста $c \in \overline{C}$, альтернативная формулировка верна.

Доказательство (ветки "если"):

- 1) если $Pr[M = m] = 0$, $Pr[M = m|C = c] = 0 = Pr[M = m]$.
- 2) рассмотрим случай, когда $Pr[M = m] > 0$.

  $Pr[C = c|M = m] = Pr[Enc_K(M) = c|M = m] = Pr[Enc_K(m) = c]$

  . Пусть

  $$\delta_c = Pr[C = c|M = m] = Pr[Enc_K(m) = c]$$

  . Если альтернативная формулировка верна, то

  $$Pr[Enc_K(m') = c] = \delta_c$$

  . По теореме Байеса:

$$Pr[M = m|C = c] = \frac{Pr[C = c|M = m] \cdot Pr[M = m]}{Pr[C = c]} = \frac{Pr[C = c|M = m] \cdot Pr[M = m]}{\sum_{m' \in \overline{M}} Pr[C = c|M = m'] \cdot Pr[M = m']} = \frac{\delta_c \cdot Pr[M = m]}{\sum_{m' \in \overline{M}} \delta_c \cdot Pr[M = m']} = \frac{Pr[M = m]}{\sum_{m' \in \overline{M}} Pr[M = m']} = Pr[M = m] \quad (1)$$

Абсолютная неразличимость:

Схема шифрования обладает абсолютной неразличимостью, если:

$$\Pr\left[\begin{array}{c} (m_0, m_1) \in \overline{M} \leftarrow Adv \\ b \leftarrow \{0,1\} \\ k \leftarrow Gen() \\ c \leftarrow Enc_k(m_b) \\ b' \leftarrow Adv(c) \\ b = b' \end{array}\right] = \frac{1}{2}.$$

Теорема 2: схема шифрования является абсолютно стойкой, если и только если она является абсолютно неразличимой.

0.2 Теорема Шеннона

Теорема 3: пусть схема шифрования (Gen, Enc, Dec) обладает свойством $|\overline{M}| = |\overline{K}| = |\overline{C}|$. *Схема является абсолютно стойкой если и только если:*

1. *Каждый возможный ключ* $k \in \overline{K}$ *выбран алгоритмом Gen с вероятностью* $\frac{1}{|\overline{K}|}$.

2. *Для каждого* $m \in \overline{M}$ *и каждого* $c \in \overline{C}$ *существует единственный и уникальный ключ* $k \in \overline{K}$, *такой, что* $Enc_k(m) = c$.

0.3 Шифр одноразовых блокнотов

Шифр одноразовых блокнотов можно считать обобщением алгоритма Виженера для того случая, когда длина ключа равна длине текста. Вместо операции сдвига обычно берется операция исключающего ИЛИ (XOR), которую мы будем обозначать как $\oplus$. Ключ, сообщение, и шифротекст (k, m, c) мы тгда полагаем битовыми строками длины l. $Enc_k(m) = k \oplus m$, $Dec_k(m) = k \oplus c$

Теорема 4: Шифр одноразовых блокнотов является абсолютно стойким.