ЛЕКЦИЯ 1

ВВЕДЕНИЕ В ЗАЩИТУ ИНФОРМАЦИИ

1. Защита информации

Защита информации — комплекс технических, организационных и юридических мер, направленных на обеспечение целостности информации, на обеспечение возможности передать информацию, реализацию права на доступ к информации и множество других целей.

Комплекс технических мер: набор программ, алгоритмов, технических средств, средств проверки прав доступа, как в компьютере, так и физически (замки, сейфы, ключи).

Юридические меры: меры, которые обеспечивает государство или другой правоустанавливающий орган для того, чтобы эффективно защищать информацию. Сюда можно отнести различные законы устанавливающие ответственность за нарушение режимов доступа к личной информации (закон о сохранении частной жизни, режим доступа к именам, паспортам, медицинским данным). Гражданский и уголовный кодекс, устанавливающий административную и уголовную ответственность (например за нарушение функционирования сети Интернет, за взлом программ, за взлом сайтов). Комплекс законов упрощающих доступ к информации (например закон об электронной цифровой подписи, утверждающий, что в качестве вашего согласия на выполнение какого-либо действия, например банком, может выступать не только ваша персональная подпись на бумаге, но и набор цифр, который кто-то передаст в банк от вашего имени, набор цифр считается вашей подписью).

Административные меры: меры на уровне предприятия, которое тоже организует разделение прав доступа к информации (бухгалтер имеет доступ к счетам, директор — ко всему).

Комплекс мер по защите должны обеспечивать:

- Конфиденциальность информации (злоумышленник не должен иметь возможность прочитать информацию).
- Защиту от несанкционированного изменения (злоумышленник не должен иметь возможность поменять информацию, даже если она передаётся по открытым каналам связи).
- Защиту возможности доступа к информации (злоумышленник не должен иметь возможность сделать так, что легальный пользователь не сможет передать информацию (DDOS, блокировка телевизионного сигнала, нарушение работы радиостанции), попытки прекратить доступ к информации тоже являются уголовным делом).

2. Математическая криптография

Здесь и далее будем обозначать собеседников, обменивающихся сообщениями, как Алиса, Боб, Ева и т. д.

Алиса хочет передать Бобу сообщение (см. 1.1) так, чтобы Ева его не подслушала, причём сообщение передаётся по открытому каналу связи (Ева может слушать этот канал). Надо как-то предотвратить возможность Евы прочитать сообщение.

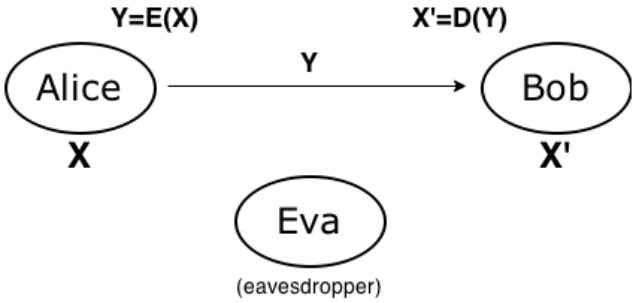

Рис. 1.1

X —случайная величина, источник сообщения. Y —случайная величина, передаваемое (зашифрованное) сообщение. X' —случайная величина, расшифрованное сообщение.

Обычно берётся сообщение X преобразуется в некоторую форму, которая пригодна для передачи по открытому каналу связи, причём сначала выполняются преобразования, которые затрудняют чтение сообщения Евой (сообщение шифруется), затем зашифрованное сообщение кодируется для удобной передачи по каналу связи.

Сразу различаем два момента: шифрование и кодирование.

Кодирование — перевод из одного алфавита в другой с целью обеспечения удобной передачи, хранения или обработки информации. Кодирование не меняет сообщение, не затрудняет его чтение, наоборот облегчает обработку сообщения (перевод сообщения из русского текста в азбуку морзе или телеграфный код).

Шифрование — подготовка сообщения таким образом, чтобы промежуточный нелегальный участник не смог его прочитать.

Соответственно, различают функции кодирования-декодирования и шифрования-дешифрования.

Модель пассивного криптоаналитика — модель в которой криптоаналитик может только слушать.

Модель активного криптоаналитика — модель в которой криптоаналитик может не только слушать, но и внедриться в канал связи, мешать, либо подменивать сообщения сторон.

Открытый текст — текст до преобразования.

Шифрованный текст (шифротекст) — текст после шифрования.

Шифрограмма — зашифрованный текст пригодный для передачи по каналу связи.

Функция шифрования — функция преобразующая открытый текст в шифрованный текст.

Ключ шифрования — дополнительный аргумент функции (используя разные ключи получаем разный шифротекст на выходе).

Ключ расшифрования — ключ, который должен использовать легальный пользователь, для того, чтобы восстановить текст.

Русские источники рекомендуют использовать слово **«расшифрование»** для операций выполняемых легальным пользователем, а **«дешифрование»** — для операций выполняемых криптоаналитиком. В лекции будет использоваться **«расшифрование»** для обозначения операций легальным пользователем, а **«криптографическая атака»** — для обозначения взлома систем (успешных и неуспешных).

Успешная атака на криптосистему — которая позволяет сократить время, взлома системы по сравнению с атакой полным перебором. Криптографическая атака будет считаться успешной, если в результате, вместо перебора 100 миллионов ключей (полный перебор), будет достаточно перебрать 50 миллионов ключей.

3. Требования к криптосистеме

Хорошая криптосистема должна использовать мало вычислительных ресурсов, чтобы зашифровать сообщение. Легальный получатель должен тратить мало вычислительных ресурсов, чтобы расшифровать сообщение. Нелегальный пользователь, криптоаналитик, должен тратить много ресурсов, чтобы с практической точки зрения было невозможно выполнить операцию взлома за обозримое время.

Есть два варианта решения задачи:

- Сохранение функции шифрования и расшифрования в тайне.
- Сами функции сделать общедоступными, но для обеспечения защиты информации ввести дополнительный аргумент (ключ Z) и именно его хранить в секрете.

Рис. 1.2

4. Типы криптосистем

4.1. Классическая криптография (с секретным ключом):

Это симметричная криптосистема, в которой шифруем и расшифровываем одним ключом ($Z_1 = Z_2$). Или если легко получить один из другого.

4.2. Криптография с открытым ключом (на открытых ключах):

Ассиметричная криптосистема — это такая криптосистема, в которой ключ шифрования и расшифрования связаны друг с другом сложными математическими операциями.

Для криптографии с открытым ключом рекомендуется использовать термины **«открытый ключ»** (public key) и **«закрытый ключ»** (private key).

4.3. Математическая криптография

Революционные идеи Клода Шеннона состояли в том, что он предложил рассматривать всё происходящее в криптографической системе как некоторые случайные случайные процессы, а именно как преобразование случайных величин. Исходное сообщение, сообщение, передаваемое по каналу связи, и ключ являются случайными величинами. Обозначают X — открытый текст, Y — шифротекст, Z — используемый ключ.

Очевидные требования:

- Энтропия источника сообщений должна быть меньше, чем логарифм от множества всех сообщений: $H(X) \leqslant \log|M|$, где M — множество всех допустимых сообщений.
- X и Z — независимы, ключ выбирается независимо от исходного сообщения:

$$P_{XZ}(X,Z) = P_X(X)P_Z(Z),$$

$$H(XZ) = H(X) + H(Z),$$

$$I(XZ) = 0.$$

4.4. Определения из теории информации

Собственная информация — показывает, сколько информации несёт одно сообщение (вероятности для всех элементов одинаковы):

$$I(x_i) = \log \frac{1}{P(x = x_i)} = -\log P(x = x_i),$$

$$I(X) = \log \frac{1}{P(x_i)} = -\log P(x_i).$$

Взаимная информация — показывает, насколько связаны две случайные величины:

$$I(XY) = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

$$I(XY) = \sum_{x_i} \sum_{y_j} \log \frac{P(x_i y_j)}{P(x_i)P(y_j)}.$$

Энтропия:

$$H(X) = -\sum_{x_i} P(x_i) \log P(x_i).$$

Условная энтропия:

$$H(X|Y) = -\sum_{y_j} P(y_k) H(X|y_k),$$

$$H(X|Y) = \sum_{x_i} \sum_{y_k} P(x_i y_k) \log P(x_i|y_k),$$

$$H(X|Y) = -\sum_{x_i} P(x_i) \sum_{y_k} P(x_i|y_k) \log P(x_i|y_k).$$

4.5. Корректность криптосистемы

Корректная криптосистема — симметричная криптосистема, для которой выполняется ряд условий:

- Ключ шифрования не зависит от сообщения:

$$I(ZX) = 0,$$

$$H(Z|X) = H(Z),$$

- Возможность восстановления сообщения по ключу:

$$H(X|YZ) = 0.$$

4.6. Совершенная криптостойкость

Симметричная криптосистема называется **совершенно криптостойкой**, если апостериорное распределение вероятностей исходного случайного сообщения x_i при регистра-

ции случайного шифротекста y_k совпадает с априорным распределением вероятностей:

$$P(x_i|y_k) = P(x_i),$$

другими словами, взаимная информация открытого и шифротекста должна равняться нулю:

$$I(YX) = 0 = I(XY).$$

Пример 1 (Латинский квадрат) Пример совершенной криптоситемы, при равновероятном использовании ключей:

	x_1	x_2	...	x_n
z_1	y_{11}	y_{21}	...	y_{n1}
z_2	y_{12}	y_{22}	...	y_{n2}
...	...	...	...	...
z_n	y_{1n}	y_{2n}	...	y_{nn}

*

Таблица 1.1

x_i — варианты исходного текста, z_i — ключи, y_{ik} — соответствующий шифротекст. Элементы y_{ik} всех строк различны. Элементы y_{ik} всех строк есть перестановки первой строки. Каждый элемент y_{ik} встречается в столбце ровно один раз.

Проблема подобных криптосистем состоит в том, что энтропия множества ключей должна быть не меньше энтропии множества сообщений. Другими словами, размер ключа должен быть не меньше размера текста. Это увеличивает размер передаваемой информации вдвое.

4.7. Расстояние единственности

Расстояние единственности — количество символов шифротекста, которое должен получить криптоаналитик, чтобы установить, какой именно из ключей использовался. Uоворя в рамках метематической модели, это такое количество элементов шифротекста (u), которое надо получить, чтобы условная энтропия ключа относительно шифротекста стала равна нулю ($f_u = 0$):

$$f_0 = H(Z),$$

$$f_1 = H(Z|y_1) \leqslant f_0,$$

$$f_2 = H(Z|y_1y_2) \leqslant f_1 \leqslant f_0,$$

$$...$$

$$f_n = H(Z|y_1y_2 \ldots y_n) \leqslant fn-1 \leqslant \ldots \leqslant f_2 \leqslant f_1 \leqslant f_0.$$

Если криптосистема абсолютно надёжна, то расстояние единственности равно бесконечности (по определению).

Лемма 1

Условная энтропия шифротекста по ключу равна энтропии открытого текста.

$$H(Y|Z) = H(X),$$

где X — открытый текст, Z — ключ, Y — шифротекст, а **криптосистема корректна**.

Док-во:

$$H(ZYX) = H(Z) + H(Y|Z) + H(X|ZY) = H(Z) + H(Y|Z),$$

так как в корректной криптосистеме расшифрование однозначно.

$$H(ZXY) = H(Z) + H(X|Z) + H(Y|XZ) = H(Z) + H(X) + 0,$$

так как ключ и открытый текст независимы, а шифрование однозначно.

$$H(Y|Z) = H(X)$$

4.8. «Линеаризация» свойств криптосистемы

А теперь «линеаризуем» криптосистему. Рассмотрим ситуацию, когда длинное сообщение подаётся маленькими частями, по буквам. Сделаем предположение, что криптосистема линейна. Энтропия открытого текста и условная энтропия шифротекста по ключу линейно увеличивается с ростом количества смволов:

$$\forall y_i \in Y : P(y_i) = \frac{1}{L},$$

$$H(y_1 y_2 \dots y_n) \approx n \log L,$$

$$H(y_1 y_2 \dots y_n|Z) \approx \frac{n}{N} H(X),$$

где L — размер алфавита (открытого текста и шифротекста), N — количество символов, n — количество принятых символов.

4.9. Набор преобразований для «линеаризованных» систем

$$H(y_1 y_2 \dots y_n Z) = H(y_1 y_2 \dots y_n) + H(Z|y_1 y_2 \dots y_n),$$

$$H(y_1 y_2 \dots y_n Z) \approx n \log L + f_n,$$

$$H(y_1 y_2 \dots y_n Z) = H(z) + H(y_1 y_2 \dots y_n|Z),$$

$$H(y_1 y_2 \dots y_n Z) = H(z) + \frac{n}{N} H(X),$$

$$n \log L + f_n \approx H(z) + \frac{n}{N} H(X),$$

$$n \log L - \frac{n}{N} H(X) \approx H(Z) - f_n.$$

4.10. Расстояние единственности

Данные выше соотношения позволяют вычислить расстояние единственности, зная характеристики открытого текста и ключа:

$$n \log L - \frac{n}{N} H(X) \approx H(Z) - f_n,$$

$$n \approx \frac{H(Z) - f_n}{\log L - \frac{H(X)}{N}} = \frac{H(Z) - f_n}{\log L (1 - \frac{H(X)}{N \log L})}.$$

Если $\exists u : f_u = 0$, то для «линеаризованной» криптосистемы:

$$u \approx \frac{H(Z)}{\log L - \frac{H(X)}{N}} = \frac{H(Z)}{\log L (1 - \frac{H(X)}{N \log L})}.$$

4.11. Избыточность источника

$$\rho = 1 - \frac{H(X)}{N \log L},$$

ρ называется избыточностью источника X, и отражает, сколько лишней информации содержится в сообщениях. Часто виновниками избыточности являются правила орфографии, пунктуации и стилистика. Если источник текста не имеет избыточности, то есть $\rho = 0$, то $u \longrightarrow \infty$.

В среднем, в русском и английском языке избыточность текста — 1,5 бита на символ. Избыточность 8-ми битового текста составляет: $\rho \approx 1 - \frac{1,5}{\log_2 2^8} \approx 0,81$, из-за этого расстояние единственности равно длине ключа плюс 20%. Это означает, что в современных криптосистемах, таких как AES, достаточно текста размером 156 байт, чтобы определить, правильно ли расшифрован текст. То есть, с теоритической точки зрения математической криптографии, все современные шифры ущербны — они подвержены атаке полным перебором. Поэтому современные криптосистемы сосздаются такими, чтобы осуществить перебор ключей было невозможно с практической точки зрения. Именно из таких соображений и существуют современные криптосистемы.