УДК 003.26-021.333

ИДЕАЛЬНАЯ КРИПТОСИСТЕМА. ШИФР ВЕРНАМА.

Ципко А.И., студент

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Баркова Е.А. – канд. физ.-мат. наук, доцент

Аннотация. Рассмотрена концепция идеальной криптосистемы и приведены доказательства ее существования. Продемонстрирована работа шифра Вернама, представлены преимущества и недостатки этой системы, предложены перспективы развития.

На протяжении всего существования человечества необходимо было передавать секретные сообщения. Идея создания невзламываемого шифра многократно разбивалась об острые умы криптоаналитиков, но в конце концов удалось сформулировать концепцию идеальной криптосистемы, которую невозможно взломать. Настоящая работа посвящена поиску доказательств существования идеальной криптосистемы, перечню требований к ней и демонстрации работы совершенно-криптостойкого шифра.

Как известно, энтропия - характеристика системы, определяющая меру ее неопределенности и вычисляющаяся по формуле [1]:

$$H(X) = -\sum_{i=1}^{n} P(x_i \in X) * log_2(P(x_i \in X), \quad (1)$$

где основание логарифма - система счисления подсчета энтропии (в нашем случае энтропия считается в битах).

Идеальная (совершенная) криптосистема - метод шифрования, шифротекст которого не раскрывает абсолютно никакой новой информации о сообщении. Исходя из данного определения вытекает следующее уравнение:

$$H(M|C) = H(M), \quad (2)$$

неопределенность сообщения независима от шифротекста. Или иначе:

$$P(M|C) = P(M), \quad (3)$$

апостериорная вероятность сообщения из множества M, при условии знания шифротекста $c_i \in$C равна априорной. Взяв это условие за основу, начнем выводить требования для его выполнения.

Воспользуемся теоремой Байеса, позволяющей взглянуть иначе на определение идеальной криптосистемы в вероятностном виде:

$$P(M|C) = \frac{P(C|M)P(M)}{P(C)} \rightarrow P(C|M) = P(C). \quad (4)$$

Результатом является независимость некоторого шифротекста из множества C от любого из сообщений. Для дальнейшей работы необходимо определить, что некоторый шифротекст $c_i \in$C является результатом работы некоторой функции E (применение метода шифрования):

$$c_i = E_k(m).$$

Применим закон полной вероятности [2] для условной вероятности P(C|M):

$$P(c_t \in \mathrm{C}|m_t \in \mathrm{M}) = \sum_{i=1}^{n} P(c_t \in \mathrm{C}|k_i, m_t \in \mathrm{M}) * P(k_i|m_t \in \mathrm{M}), \quad (5)$$

где c_t - это шифротекст, который получается применением к сообщению m_t некоторого ключа. Благодаря этому действию мы можем связать вероятность появления c_t при условии знания m_t с ключом $k_i \in K$.

Пусть далее ключ выбирается независимо от сообщений:

$$P(k_i|\mathrm{M}) = P(k_i), \tag{6}$$

и данная криптосистема является детерминированной, то есть одна и та же пара сообщения и ключа всегда даст один и тот же шифротекст. И в обратную сторону, используя получившийся шифротекст и ключ, который использовался, мы однозначно восстановим исходное сообщение. Тогда формула (5) примет новый вид:

$$P(c_t \in \mathrm{C}|m_t \in M) = \sum_{i=1}^{n} P(c_t \in \mathrm{C}|k_i, m_t \in M) * P(k_i), \tag{7}$$

Как было сказано ранее, наша система детерминирована, тогда

$$P(c_t \in \mathrm{C}|k_i, m_t \in \mathrm{M}) = 1$$

при условии, что k_i - это ключ, применив который к m_t мы получим c_t, назовем его альфа ключ, а также

$$P(c_t \in \mathrm{C}|k_i, m_t \in \mathrm{M}) = 0$$

для всех остальных случаев.

Действительно, так как к определенному шифротексту для некоторого исходного сообщения приводит лишь один единственный ключ, то заменяя только что-то одно из пары сообщение-ключ (либо ключ, либо сообщение) мы никогда снова не получим этот шифротекст, только если поменять и сообщение, и ключ, что не имеет никакого смысла. Тогда, рассматривая все возможные ключи, примененные к разным сообщениям, получим, что одна часть всех возможных сумм уравнения (7) будет равна нулю, а другая будет равна сумме вероятностей появления альфа ключа:

$$P(c_t \in \mathrm{C}|m_t \in M) = \sum_{k:E_k(m_t)=c_t}^{n} P(k). \tag{8}$$

Далее необходимо, чтобы для любого m_t уравнение было верным. Этого можно добиться полным контролем распределения вероятностей множества ключей, синхронизирующимся с множеством сообщений, что довольно сложно. Есть еще один вариант, полностью противоположный - абсолютная непредсказуемость - равномерное распределение вероятностей множества ключей. Для второго варианта достаточно лишь добиться для всех ключей равновероятности - одно условие для всего множества. Выбирая этот способ, мы получаем следующие требования:

1. Шифрующий ключ используется всегда один раз.
2. Множество ключей K обладает равномерным распределением вероятностей.

Эти два требования обеспечивают выполнение равенства (4), так как никакое сообщение никогда не изменит вероятность появления шифротекста.

Для дальнейшей работы введем определение [1; 3]:
Условная энтропия - величина, рассчитываемая по формуле

$$H(X|Y) = -\sum_{k=1}^{p} P(y_k \in Y) \sum_{i=1}^{n} P(x_i \in X|y_k \in Y) * log_2 P(x_i \in X|y_k \in Y), \tag{9}$$

преобразуем полученное равенство занесением $P(y_k \in Y)$ под знак первой суммы, умножением на условную вероятность $P(x_i \in X|y_k \in Y)$. В результате получим совместную вероятность $P(x_i \in X, y_k \in Y)$. Следовательно формула (9) примет новый вид:

$$H(X|Y) = -\sum_{i=1,k=1}^{n,p} P(x_i \in X, y_k \in Y) * log_2 P(x_i \in X|y_k \in Y). \quad (10)$$

По теореме умножения распишем совместную вероятность трех событий:

$$P(X,Y,Z) = P(X|Y,Z) * P(Y|Z) * P(Z). \quad (11)$$

Найдем совместную энтропию для трех событий:

$$H(X,Y,Z) = -\sum P(X,Y,Z) log_2 P(X,Y,Z) =$$

$$= -\sum P(X,Y,Z) log_2 P(X|Y,Z) - \sum P(X,Y,Z) log_2 P(Y|Z) - \sum P(X,Y,Z) log_2 P(Z). \quad (12)$$

Рассмотрим слагаемые детально по отдельности:

$$-\sum P(X,Y,Z) log_2 P(Z) = -\sum_z \sum_y \sum_x P(X,Y,Z) log_2 P(Z) = \sum_z log_2 P(Z) \sum_{x,y} P(X,Y,Z).$$

Найдем маргинальную вероятность [4] события Z, суммируя совместную вероятность $P(X,Y,Z)$ по X и Y, тогда получим:

$$-\sum_z P(Z) log_2 P(Z) = H(Z),$$

- по определению энтропии.

$$-\sum P(X,Y,Z) log_2 P(X|Y,Z) = H(X|Y,Z),$$

- по уравнению (10).

Для оставшегося слагаемого воспользуемся сначала маргинализацией, затем уравнением (10):

$$-\sum P(X,Y,Z) log_2 P(Y|Z) = H(Y|Z).$$

Таким образом было доказано, что разложение энтропии некоторой системы событий* X на сумму энтропий новых преобразованных систем событий $Y_1, Y_2 \dots Y_n$ повторяет преобразование X при разложении ее вероятности на множители с эквивалентными системами событий $Y_1, Y_2 \dots Y_n$. Назовем это свойствами связи энтропии и вероятностей. Это объясняется особенностями логарифма, присутствующего в формуле энтропии.

*Будем называть системой событий любое множество событий, пересечение и/или объединение множеств событий.

Рассмотрим следующие преобразования, руководствуясь свойствами связи и приведенными определениями:

$$H(K|C) = H(K|C) + H(M|K,C) = H(M,K|C),$$

где $H(M|K,C) = 0$

$$H(M,K|C) = H(M|C) + H(K|M,C) = H(K|C),$$

$H(K|M,C) \geq 0$, т.к. по определению энтропия не может быть меньше нуля, следовательно

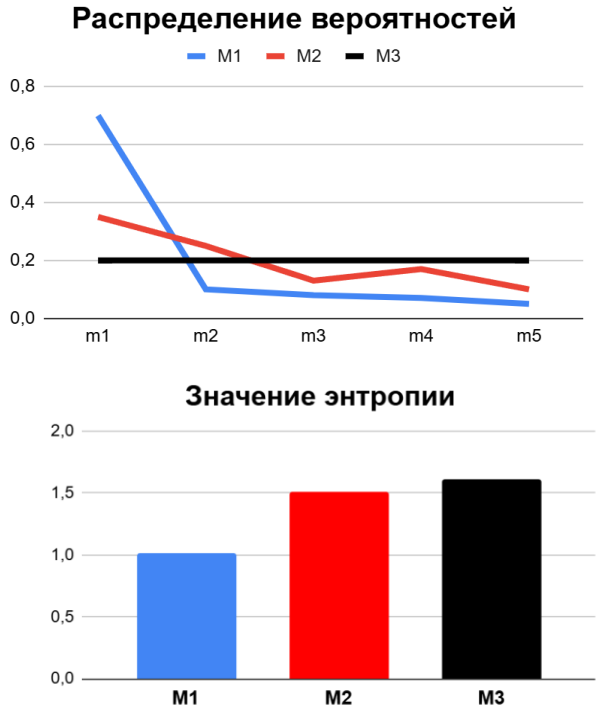
$$H(K|C) \geq H(M|C) \rightarrow H(K) \geq H(M). \quad (13)$$

Результатом этих преобразований является следующее утверждение: “Энтропия ключа должна быть не меньше энтропии сообщения”.

Как уже было определено, распределение вероятностей ключа должно быть равномерным, тогда неопределенность ключа полностью зависит от мощности множества ключей, которая зависит от длины ключа.

$$H(K) = -\sum \frac{1}{|K|} log_2 \frac{1}{|K|} = |K| * \frac{1}{|K|} * log_2 |K| = log_2 |K|. \quad (14)$$

Рассмотрим динамику изменения энтропии в зависимости от распределения вероятностей этого множества событий на примере из пяти элементов для трех множеств:



По графикам видно, что энтропия максимальна при равномерном распределении вероятностей. Тогда для анализа энтропии сообщения неравенства (13) достаточно рассмотреть максимальную энтропию сообщения, то есть когда все сообщения равновероятны:

$$H(M)_{max} = -\sum \frac{1}{|M|} log_2 \frac{1}{|M|} = |M| * \frac{1}{|M|} * log_2|M| = log_2|M|. \quad (15)$$

На основании последних двух уравнений делаем вывод, что длина ключа должна быть не меньше длины сообщения:

$$H(K) \geq H(M) \rightarrow L(K) \geq L(M). \quad (16)$$

Только что мы вывели третье требование к системе. Таким образом совершенная криптосистема - это такая криптосистема, которая удовлетворяет следующим требованиям:

1. Шифрующий ключ используется всегда один раз.
2. Множество ключей K обладает равномерным распределением вероятностей.
3. Длина ключа должна быть не меньше длины сообщения.

Было доказано, что существует криптосистема, которая является идеальной, это шифр Вернама. Мною написана программа, реализующая и демонстрирующая работу этой криптосистемы

```
Введите сообщение:
BSUIR
Вы ввели:
BSUIR

len = 48
Сообщение в двоичном коде     01000010 01010011 01010101 01001001 01010010 00001010
                              ++++++++ ++++++++ ++++++++ ++++++++ ++++++++ ++++++++
Ключ                          00001000 10101000 10000010 00000010 10011011 00100000
                              ======== ======== ======== ======== ======== ========
Шифротекст                    01001010 11111011 11010111 01001011 11001001 00101010
ЈыЧКЙ*
Зашифрованное сообщение
ЈыЧКЙ*

Шифротекст в двоичном коде    01001010 11111011 11010111 01001011 11001001 00101010
                              ++++++++ ++++++++ ++++++++ ++++++++ ++++++++ ++++++++
Ключ                          00001000 10101000 10000010 00000010 10011011 00100000
                              ======== ======== ======== ======== ======== ========
Результат                     01000010 01010011 01010101 01001001 01010010 00001010

BSUIR
```

Принцип работы предельно простой: каждый символ кодируется битами согласно таблице ASCII, далее генерируется ключ, длиной, равной длине сообщения (в данном случае + 8 бит в конце для управляющего символа). Ключ должен генерироваться равновероятно для всех возможных вариантов $2^{L(K)}$. После этого производится последовательное сложение по модулю 2 соответствующих битов сообщения и ключа. Для расшифровки к результату предыдущей операции снова применяется тот же секретный ключ.

Для практической оценки надежности данного метода представим следующее: злоумышленник пытается расшифровать перехваченное послание. Он уже знает, каким образом шифровалось полученное им сообщение. Также предположим, что у него есть бесконечное количество времени на взлом и бесконечные вычислительные мощности. Так или иначе, если не получается обнаружить закономерности или уязвимости шифра, которые помогли бы определить исходное сообщение, всегда можно перебрать все возможные варианты, чем злоумышленник и займется. Для шифротекста “ЈыЧКЙ*“ таких вариантов получится 2^{40} - столько сообщений в результате получит злоумышленник, из которых лишь 12-15 тысяч* будут иметь значение. Какое слово из 12-15 тысяч было зашифровано? Этого ни он, ни кто-либо другой никогда не узнают без альфа ключа, не говоря уже про то, что сообщение могло быть обманкой и вообще не иметь смысла. Это практическая суть идеальной криптосистемы.

Конечно, у такой системы есть и недостатки:

1. Необходимо использовать ключ, генерация которого равновероятна относительно всех возможных других ключей, чего не так просто добиться.

2. В нашей системе длина ключа равна длине сообщения. Обязательным условием конфиденциальности является секретная передача ключа. Но вот парадокс: если есть возможность по некоторому секретному каналу передать ключ, длиной, равной сообщению, то что мешает по этому же каналу сразу же передать сообщение?

В ходе выполненной работы была доказана концепция идеальной криптосистемы, разработан алгоритм, представляющий собой шифр Вернама. Данная тема представляет интерес идеей невзламываемого шифра, который гарантирует абсолютную секретность переданной информации. Впереди концепцию ждут модернизации и адаптации под реалии. В основном банально нет необходимости добиваться абсолютной криптостойкости, так как современные алгоритмы шифрования обеспечивают секретность на необходимый промежуток времени, однако все равно регулярно происходят взломы теми или иными методами. Поэтому идею идеальной криптосистемы ждет перспективное будущее, когда у людей больше не будет сложностей с хранением и передачей больших ключей. Также перспективным направлением будет создание криптосистемы, свойства которой стремятся к идеальной системе, но при этом ее работа будет менее затратной и более эффективной.

Список использованных источников:

[1] Shannon, C. E. A mathematical theory of communication // Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423, 623–656.

[2] Онлайн ресурс курс лекций “Теория вероятностей”, Райгородский А.М МФТИ.

[3] Онлайн ресурс курс лекций “Теория информации”, Григорьев А.А МФТИ.

[4] Marginal distribution, wikipedia.

UDC 003.26-021.333

THE IDEAL CRYPTO SYSTEM. THE VERNAM CIPHER.

Tsipko A.I., student

Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

Barkova E.A. – PhD in Physics and Mathematics, Associate Professor

Annotation. The concept of an ideal cryptosystem is examined, and proofs of its existence are provided. The operation of the Vernam cipher is demonstrated, the advantages and disadvantages of this system are presented, and prospects for its development are proposed.