

# ПРИНЦИПЫ ПОСТРОЕНИЯ АЛГОРИТМОВ ШИФРОВАНИЯ

*М.А. Протько<sup>1)</sup>, О.Ф. Борисенко<sup>2)</sup>*

1) студентка Белорусского государственного университета информатики и радиоэлектроники, г. Минск, Беларусь, [mari.protko@mail.ru](mailto:mari.protko@mail.ru)

2) к.т.н, доцент Белорусского государственного университета информатики и радиоэлектроники, г. Минск, Беларусь, [kafvm@bsuir.by](mailto:kafvm@bsuir.by)

**Аннотация:** в данной статье рассматривались основные принципы построения алгоритмов шифрования, а также дальнейшие проверки их на «секретность».

**Ключевые слова:** криптография, шифротекст, полиномиальное время, криптостойкость, принцип Керкгоффа.

## PRINCIPALS OF DERIVING ENCRYPTION ALGORITHMS

*Maria A. Protsko<sup>1)</sup>, Oleg F. Borisenko<sup>2)</sup>*

1) the student of Belarusian State University of Informatics and Radioelectronics city of Minsk, Belarus, [mari.protko@mail.ru](mailto:mari.protko@mail.ru)

2) Ph.D., associate Professor, associate Professor of the Department of Higher Mathematics, Belarusian State University of Informatics and Radioelectronics city of Minsk, Belarus, [kafvm@bsuir.by](mailto:kafvm@bsuir.by)

**Abstract:** This article explores the main principals of deriving encryption algorithms as well as further checks on their "secrecy".

**Keywords:** cryptography, ciphertext, polynomial time, cryptographic strength, Kerckhoff's principle.

Что составляет любой базовый криптографический алгоритм? По сути, это два соответствия: базовый/шифрованный текст. Связь между первым и вторым происходит по некой функции  $F$  с приблизительно следующими свойствами:

$F(B) = A$  – легко рассчитываемая функция

$B = F^{-1}(A)$  - не вычисляемая функцией доступными нам средствами.

То есть, любая наша задача будет соответствовать следующей формулировке:

Пусть  $K$  – пространство ключей,  $e$  и  $d$  – ключи шифрования и расшифрования соответственно.  $E_e$  – односторонняя функция шифрования для произвольного ключа  $e \in K$ , такая, что  $E_e(t)=c$ ,  $c \in C$ ,  $C$  – пространство шифротекстов,  $t \in T$ ,  $T$  – пространство сообщений.  $D_d$  – функция расшифрования, такая, что  $D_d(c)=t$ . Каждая пара  $(E,D)$  имеет свойство: зная  $E_e$  невозможно найти  $E_e(t)=c$ .

Чаще всего, в качестве функций используют математически трудно решаемые задачи, не решаемые простым перебором: задачи факторизации чисел, взятие корня, возведение в степень, поиск квадратичных вычетов и т.д.

Учитывая вышеупомянутую вычислительную сложность, простой перебор всех возможных значений и применение принципа индукции могут не дать верного результата. Для более качественной оценки полученных алгоритмов необходимо четко обозначить условия, из которых они вытекают, а также те параметры, которым они обязаны удовлетворять.

Основные условия, которым должен удовлетворять шифр:

1)  $T = D(E(T))$ . – где  $T$  – пространство сообщений.  $D$  и  $E$  – функции расшифрования и шифрования соответственно.

Иначе: применив функции  $D$  или  $E$  мы получим однозначный связанный текст без потери информации в обоих случаях. Мы можем выполнять эту операцию сколь угодно долго.

2)  $H(E) \geq H(T)$ . – где  $H(e)$  и  $H(T)$  – неопределенности зашифрованного и изначального сообщения.

Иными словами, вероятность предсказания следующего символа перехваченного сообщения (нахождения закономерности) значительно уменьшается при шифровании.

3) Удовлетворяет принципу Керкгоффа

То есть, даже зная процедуру (функцию  $F$ ), невозможно получить изначальный текст. Необходим ключ, чтобы это стало возможным.

При оценке шифра мною будут сделаны следующие допущения:

У злоумышленника нет никаких априорных сведений о зашифрованном сообщении (все исходные варианты равновероятны).

Отсутствуют специальные символы (для простоты понимания передаются цифры – тот же принцип должен быть применим к любому другому множеству символов)

В качестве примера возьмём один из частных случаев шифра с ассиметричным ключом.

Воспользуемся RSA алгоритмом, поскольку он достаточно надежен и прост для понимания. Надежен потому, что ассиметричные ключи, используемые им, не обязаны быть одинаково известны как адресату, так и получателю (их не нужно передавать, достаточно знать всего один). Прост потому, что основан на признаках делимости чисел и возможности нахождения простых делителей.

Для построения RSA алгоритма нам будет необходимо:

1) Выбрать два случайных простых числа  $p$  и  $q$

2) Найти модуль:  $n = p \cdot q$

3) Найти функцию Эйлера

$$\varphi(n) = (p-1) \cdot (q-1)$$

4) Найти простое число  $e$ ,  $1 < e < \varphi(n)$ , где  $e$  – взаимно простое с значением  $\varphi(n)$

5) Найти  $d$  из уравнения  $d \cdot e \equiv 1 \pmod{\varphi(n)}$

В итоге, получим пары  $(e, n)$  – открытый ключ,  $(d, n)$  – закрытый ключ.

Пример работы алгоритма:

Допустим, у нас есть ряд чисел, которые необходимо зашифровать

T: 2 1 4 3 8 5 6

$n = 10$

$e = 3$

$d = 3$  (случайное совпадение)

$E(m) = m^e$

$c = E(m) \pmod n$

E(m): 8 1 64 27 512 125 216

c: 8 1 4 7 2 5 6 - T'

Для получения исходного  $m$  из  $c$  те же действия, с заменой  $e$  на  $d$ :

D(c): 512 1 64 343 8 125 216

m: 2 1 4 3 8 5 6

Как видно из примера, лучше брать большие числа, т. к. 1 и 6 не меняется во всех случаях, что может оказаться слабым местом в кодировке, ведь возможно проанализировать процентное соотношение этой 1 и 6 в сообщении (в случае если передается большой буквенный текст). А известные две буквы резко увеличивают скорость расшифровки.

Но наглядности для обоснованности недостаточно. Воспользуемся формулами:

$$H(T) = - \sum_{i=1}^n p_i \log_2 p_i$$

Где  $H(T)$  – мера неопределенности сообщения/шифра.  $T$  – множество возможных отправленных сообщений,  $p_i$  – вероятность отправки соответствующего сообщения  $T_i$ ,  $n$  – количество битов шифротекста. Физический смысл найденной величины: количество битов информации, которое необходимо в среднем передать, чтобы полностью устранить неопределенность.

После перехвата полученного сообщения получим следующее значение неопределенности (в литературе называемое условным):

$$H(T|T') = - \sum_{i=1}^n p(T_i|T') \log_2 p_i(T_i|T')$$

Где  $p(T_i|T')$  – вероятность того, что исходное сообщение есть  $T_i$  при условии, что результат шифрования  $T'$ .

Отсюда найдем значимую характеристику нашего шифра:

$$I = H(T) - H(T|T')$$

Где  $I$  – информация об исходном тексте, которую злоумышленник может извлечь из перехваченного шифротекста. Необходимо обеспечить, чтобы  $H(T) \rightarrow H(T|T')$  ( $I \rightarrow 0$ ).

Т.е, чем меньше  $I$ , тем меньше вероятность однозначного дешифрования без знания ключа.

Таким образом, мы сможем математически оценить криптографическую стойкость нашего алгоритма.

Вернемся к примеру, и попытаемся его качественно оценить:

Для начала предположим, что один бит сообщения  $T$  – одна цифра. Передается по порядку следования. Никаких манипуляций, связанных с непоследовательной передачей сообщения, не было. Также не передается «шум» - какие-либо последовательности, не имеющие логического смысла.

Таким образом, человек, перехватывающий сообщение будет с большей вероятностью полагать, что операции будут производиться с цифрами от 1-9.

Посчитаем  $H(T)$  из вышеописанных условий:

Если злоумышленник не имеет никаких сведений:

$P_i=0,11111$  (1 к 9, где событие – выбор конкретной цифры.)

Поскольку мы рассматриваем априорный случай, предыдущие сообщения никак не влияют на последующие, получим:

$H(T) = - 7 * (0,11111)$

$$H(T) = - \sum_{i=1}^7 0,11111 \log_2 0,11111 = -2,4654$$

Если никаких апостериорных сведений не было получено:

$H(T) = H(T|T')$   $I=0$  – самый лучший исход, возможный только теоретически.

Если нам известно, что цифры не повторяются:

$$H(T|T') = - \sum_{i=1}^n \frac{1}{n-i+1} \log_2 \left( \frac{1}{n-i+1} \right) = -3,0518$$
$$I = -2,4654 - (-3,0518) = 0,5864$$

Сразу становится заметно, что, сделав всего одно предположение, вероятность расшифровки резко возрастает.

Поэтому, недостаточно использовать трудно вычисляемую задачу для создания шифра, нужно позаботиться и о том, чтобы  $T'$  и  $T$  не взаимоисключали друг друга ( в рассмотренном примере достаточно сразу можно заметить, что большая часть элементов  $T'$  идентична  $T$ , а также не используются цифры 7 в  $T$  и 3 в  $T'$  что даст достаточно оснований для предположения, что скорее всего в  $T'$  тройке соответствует 7 или 9, а от этого предположения, можно вывести и все остальное (составив все соответствия и выбрав наиболее вероятные))

#### Список использованных источников:

1. М.А. Протыко. Теория чисел в асимметричном шифровании. 57-я

Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР, Минск, 2021.

2. Стройникова Е.Д. Основы прикладной алгебры: учеб.-метод. пособие/ Е.Д.Стройникова.-Минск:БГУИР,2010.-120 с.:ил.

3. Горовенко Л.А. О развитии математической культуры студентов инженерного вуза // Прикладные вопросы точных наук Материалы III Международной научно-практической конференции студентов, аспирантов, преподавателей. - Армавир: РИО АГПУ, 2019. - С. 280-282.

4. Часов К.В., Горовенко Л.А. Математическая культура как неотъемлемая составляющая информационной образовательной среды инженерно-технического вуза: монография/ К.В. Часов, Л.А. Горовенко; Армавирский механико-технологический институт.- Армавир: РИО АГПУ, 2019. - 188 с.

5. Липницкий В.А Современная прикладная алгебра. Математические основы защиты информации от помех и несанкционированного доступа: учеб.-метод. пособие/ Липницкий В.А.- Минск:БГУИР, 2006. -87 с.:ил.