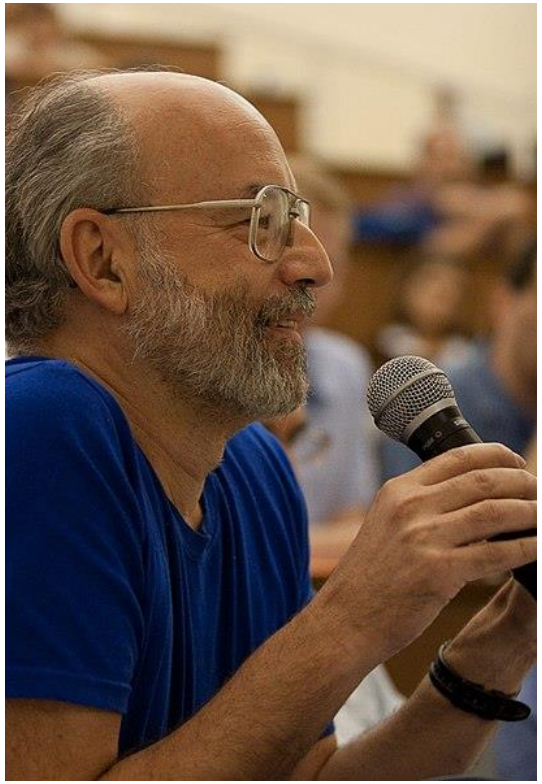
Шифр RSA. Односторонняя функция с «лазейкой»

R – Rivest

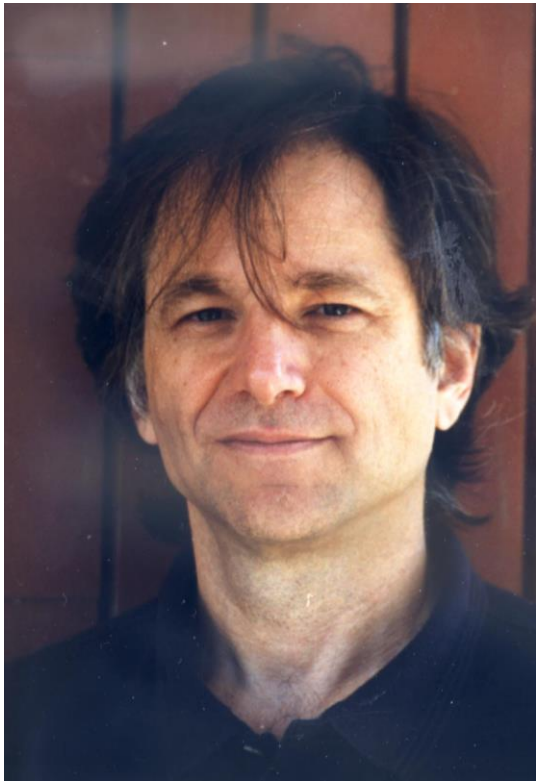
S – Shamir

A – Adleman

Ronald Rivest



Adi Shamir



Leonard Adleman

1977 – разработан алгоритм RSA

В августе 1977 года в колонке «Математические игры» Мартина Гарднера в журнале Scientific American появилось первое описание криптосистемы RSA.

Читателям также было предложено дешифровать английскую фразу, зашифрованную описанным алгоритмом:

9686	9613	7546	2206
1477	1409	2225	4355
8829	0575	9991	1245
7431	9874	6951	2093
0816	2982	2514	5708
3569	3147	6622	8839
8962	8013	3919	9055
1829	9451	5781	5154

1993 г. - запуск проекта распределённых вычислений с координацией через электронную почту по решению головоломки.

Исходное сообщение:

«THE MAGIC WORDS ARE SQUEAMISH OSSIFRAGE»

«Волшебные слова — это брезгливый ягнятник»

9686	9613	7546	2206
1477	1409	2225	4355
8829	0575	9991	1245
7431	9874	6951	2093
0816	2982	2514	5708
3569	3147	6622	8839
8962	8013	3919	9055
1829	9451	5781	5154

Описание алгоритма

Эта система базируется на следующих двух фактах из теории чисел:

1) задача проверки числа на простоту является сравнительно легкой;

2) задача разложения чисел вида $n = pq$ (p и q — простые числа) на множители является очень трудной, если мы знаем только n, а p и q — большие числа (это так называемая задача факторизации).

Пусть в нашей системе есть абоненты A, B, C, ... Каждый абонент выбирает случайно два больших простых числа P и Q. Затем он вычисляет число

$$N = PQ. \tag{2.28}$$

(Число N является открытой информацией, доступной другим абонентам.) После этого абонент вычисляет число $\phi = (P-1)(Q-1)$ и выбирает некоторое число $d < \phi$, взаимно простое с ϕ, и по обобщенному алгоритму Евклида находит число c, такое, что

$$cd \bmod \phi = 1. \tag{2.29}$$

Вся информация, связанная с абонентами и являющаяся их открытыми и секретными ключами, представлена в табл. 2.4.

Таблица 2.4. **Ключи пользователей в системе RSA**

Абонент	Секретный ключ	Открытый ключ
A	c_A	d_A, N_A
B	c_B	d_B, N_B
C	c_C	d_C, N_C

Пусть Алиса хочет передать сообщение $\boldsymbol{m}$ Бобу, причем сообщение $\boldsymbol{m}$ рассматривается как число, удовлетворяющее неравенству $\boldsymbol{m < N_B}$ (далее индекс B указывает на то, что соответствующие параметры принадлежат Бобу).

Шаг 1. Алиса шифрует сообщение по формуле

$$e = m^{d_B} \bmod N_B, \tag{2.30}$$

используя открытые параметры Боба, и пересылает e по открытой линии.

Шаг 2. Боб, получивший зашифрованное сообщение, вычисляет

$$m' = e^{c_B} \bmod N_B. \tag{2.31}$$

Теорема 2.8. *Если p и q — простые числа, $p \neq q$ и k — произвольное целое число, то*

$$a^{k\varphi(pq)+1} \bmod (pq) = a. \tag{2.12}$$

Утверждение 2.12. *Для описанного протокола $m' = m$, т.е. абонент B получает исходящее от A сообщение.*

Доказательство. По построению протокола

$$m' = e^{c_B} \bmod N_B = m^{d_B c_B} \bmod N_B.$$

Равенство (2.29) означает, что для некоторого k

$$c_B d_B = k\phi_B + 1.$$

Согласно утверждению 2.5

$$\phi_B = (P_B - 1)(Q_B - 1) = \varphi(N_B),$$

где $\varphi(\cdot)$ — функция Эйлера. Отсюда и из теоремы 2.8 следует

$$m' = m^{k\varphi(N_B)+1} \bmod N_B = m.$$

Утверждение 2.13 (свойства протокола RSA).

1) *Протокол шифрует и дешифрует информацию корректно;*

2) *злоумышленник, перехватывающий все сообщения и знающий всю открытую информацию, не сможет найти исходное сообщение при больших* P *и* Q*.*

Первое свойство протокола следует из утверждения 2.12.

Доказательство второго свойства: злоумышленник знает только открытые параметры N и d . Чтобы найти c , он должен знать значение $\varphi(N) = (P - 1)(Q - 1)$, а для этого ему требуется знать P и Q.

Он может найти P и Q, разложив N на множители, однако это трудная задача.

Отметим, что выбор больших случайных P и Q возможен за приемлемое время.

Односторонняя функция $y = x^d mod\, N$, применяемая в системе RSA, обладает так называемой «лазейкой», позволяющей легко вычислить обратную функцию $x = \sqrt[d]{y}\, mod\, N$, если известно разложение *N* на простые множители.

Действительно, легко вычислить $\varphi = (P - 1)(Q - 1)$, а затем $c = d^{-1} mod\, \varphi$.

Если P и Q неизвестны, то вычисление значения обратной функции практически невозможно, а найти *P* и *Q* по *N* очень трудно, т.е. знание *P* и *Q* — это «лазейка» или «потайной ход»).

Для схемы *RSA* важно, чтобы каждый абонент выбирал **собственную пару простых чисел *P* и *Q*, т.е. все модули N_A , N_B , N_C , . . . должны быть различны** (в противном случае один абонент мог бы читать зашифрованные сообщения, предназначенные для другого абонента).

Однако этого не требуется от второго открытого параметра *d* . **Параметр *d* может быть одинаковым у всех абонентов.**

Часто рекомендуется выбирать $d = 3$ (при соответствующем выборе *P* и *Q*). Тогда шифрование выполняется максимально быстро, всего за два умножения.

Пример 2.17. Допустим, Алиса хочет передать Бобу сообщение $m = 15$. Пусть Боб выбрал следующие параметры:

$$P_B = 3,\ Q_B = 11,\ N_B = 33,\ d_B = 3$$

(3 взаимно просто с $\varphi(33) = 20$). Найдем c_B с помощью обобщенного алгоритма Евклида:

$$c_B = 7$$

(проверим: $3 \cdot 7 \bmod 20 = 1$). Кодируем m по формуле (2.30):

$$e = 15^3 \bmod 33 = 15^2 \cdot 15 \bmod 33 = 27 \cdot 15 \bmod 33 = 9.$$

Число 9 Алиса передает Бобу по открытому каналу связи. Только Боб знает $c_B = 7$, поэтому он декодирует принятое сообщение, используя (2.31):

$$m' = 9^7 \bmod 33 = \left(9^2\right)^2 \cdot 9^2 \cdot 9 \bmod 33 = 15^2 \cdot 15 \cdot 9 \bmod 33 = 15.$$

Рассмотренная система невскрываема при больших P и Q, но обладает следующим недостатком: A передает сообщение B , используя открытую информацию абонента B (числа N_B и d_B). **Злоумышленник не может читать сообщения, предназначенные для *B* , однако он может передать сообщение к *B* от имени *A*.**

Избежать этого можно, используя более сложные протоколы, например, следующий.

A хочет передать B сообщение m . Сначала A вычисляет число

$$e = m^{c_A} \bmod N_A .$$

Злоумышленник не может этого сделать, так как c_A секретно. Затем A вычисляет число

$f = e^{d_B} \bmod N_B$ и передает f к B.

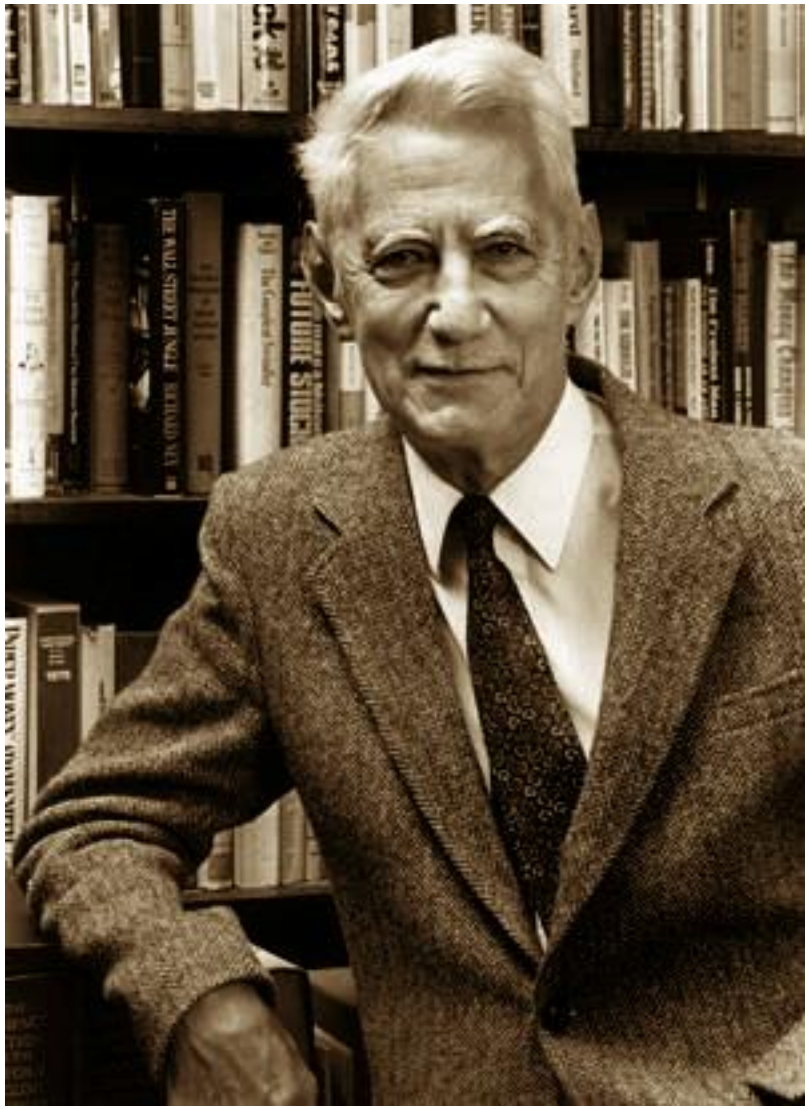
B получает f и вычисляет последовательно числа

$$u = f^{c_B} \bmod N_B \text{ и } w = u^{d_A} \bmod N_A .$$

В результате абонент *B* получает сообщение $\boldsymbol{w = m}$. Как и в исходной схеме *RSA*, **злоумышленник не может прочитать переданное сообщение**, но здесь, в отличие от *RSA*, **он не может также послать сообщение от имени *A*** (поскольку не знает секретного $\boldsymbol{c_A}$).

Здесь мы встречаемся с новой ситуацией. *B* знает, что сообщение пришло от *A*, т.е. *A* как бы «подписал» его, зашифровав своим секретным $\boldsymbol{c_A}$. Это пример так называемой ***электронной*** или ***цифровой подписи***. Она — одно из широко используемых на практике изобретений современной криптографии.

ТЕОРЕТИЧЕСКАЯ СТОЙКОСТЬ КРИПТОСИСТЕМ



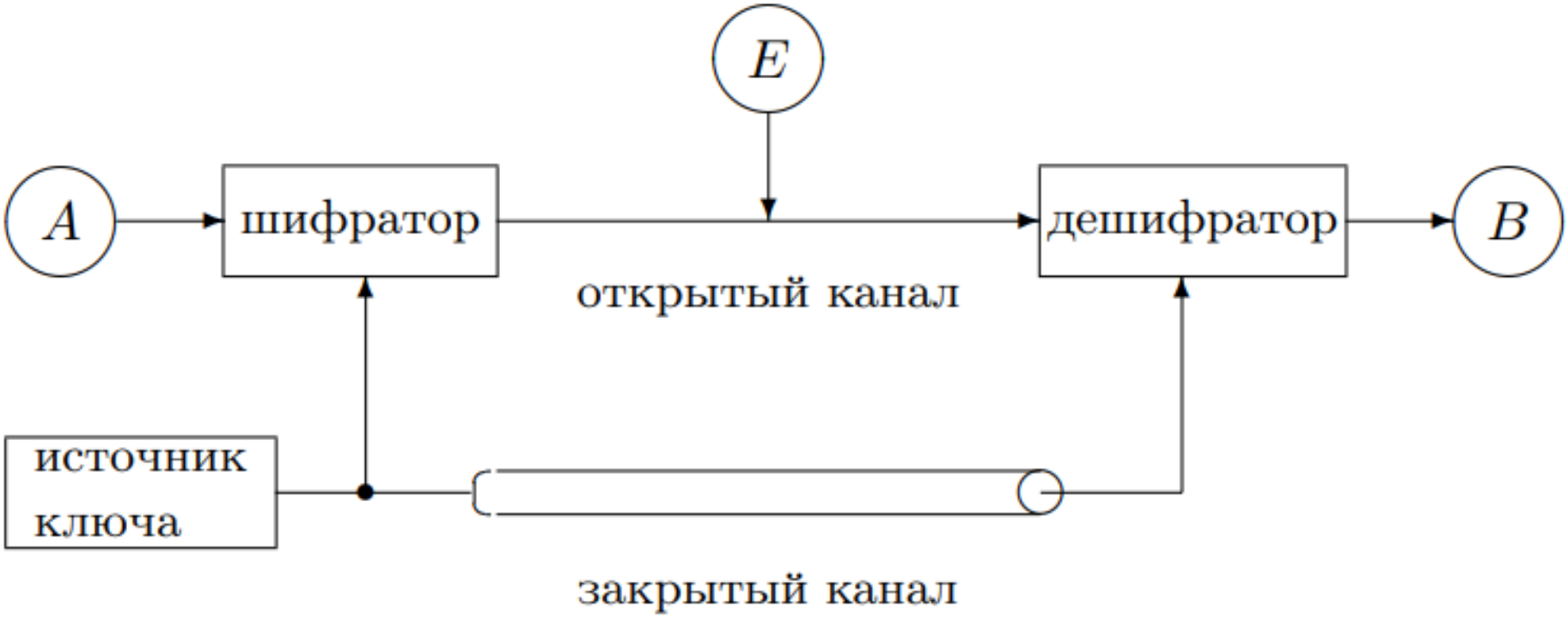
1949 г.

Клод Шеннон

(Claude Shannon)

предложил классическую схему криптосистемы с секретным ключом

Классическая система секретной связи



Все методы шифрования можно грубо разделить на два больших класса:

1) схемы, принципиально не вскрываемые, что доказывается строго;

2) схемы, стойкость которых основана на том, что дешифрование без ключа, вообще говоря, возможно, но для этого требуется перебор очень большого количества вариантов.

Теория систем с совершенной секретностью

$M = \{M_1, M_2, M_3, \ldots, M_m\}$ — множество всех возможных сообщений (например, множество всех текстов длины не более, чем 1000 букв),

$K = \{K_1, K_2, K_3, \ldots, K_n\}$ — множество всех возможных ключей,

$E = \{E_1, E_2, \ldots, E_k\}$ — множество всех криптограмм (т.е. зашифрованных сообщений).

Зашифрованные сообщения зависят от исходного сообщения и ключа, т.е.

$$E_j = f(M_i, K_l).$$

Будем считать, что на всем множестве сообщений M задано распределение вероятностей P, т.е. определены вероятности $P(M_i), i = 1, 2, \ldots, m$.

Это априорное распределение вероятностей, которое известно и противнику.

$P(A|B)$ — условная вероятность события A при условии наступления события B .

Определение 7.1. Криптосистема называется *совершенно секретной*, если выполняется равенство

$$P\left(M_i|E_j\right) = P\left(M_i\right) \tag{7.1}$$

при всех M_i, K_l и $E_j = f(M_i, K_l)$.

Поясним это определение. Пусть Ева перехватила криптограмму E_j. Если (7.1) выполняется для всех возможных сообщений, то это означает, что она не получила никакой информации о переданном сообщении, т.е. знание E_j совершенно бесполезно для нее. Рассмотрим схематичный пример. Пусть M — множество всех сообщений из шести букв на русском языке . Пусть априори известно, что для некоторой системы

$$P(\text{сообщение} = \text{«доллар»}) = 0.00015,$$
$$P(\text{сообщение} = \text{«бутыль»}) = 0.0000012 \text{ и т.д.}$$

Допустим, мы имеем несовершенную систему, и Ева после перехвата и вычислений получила следующие данные:

$$P(\text{сообщение} = \text{«доллар»}) = 10^{-20},$$
$$P(\text{сообщение} = \text{«бутыль»}) = 0.9999.$$

Это означает, что Ева практически расшифровала сообщение: она практически уверена, что передано слово «бутыль», так как вероятность, что передано другое сообщение меньше 0.0001.

Если же для рассмотренной системы при любой перехваченной криптограмме E_j мы получаем

$$P(\text{сообщение} = \text{«доллар»}|E_j) = 0.00015,$$
$$P(\text{сообщение} = \text{«бутыль»}|E_j) = 0.0000012$$

и такие же равенства выполняются для всех остальных сообщений, то Ева вообще может не обращать внимание на перехваченный шифротекст E_j, а, например, отгадывать сообщение на основе исходных вероятностей. Другими словами, (7.1) — действительно разумное определение совершенно секретной системы.

Исследуем свойства совершенно секретной системы.

Теорема 7.1. *Если система является совершенно секретной (выполняется* (7.1)*), то справедливо равенство*

$$P\left(E_j|M_i\right) = P\left(E_j\right) \tag{7.2}$$

при всех i *и* j *. Верно и обратное утверждение: если* (7.2) *выполняется, то система совершенно секретна.*

Доказательство. По определению условной вероятности

$$P\left(A|B\right) = \frac{P(AB)}{P(B)},$$

при $P(B) \neq 0$. Поэтому при $P(E_j) \neq 0$ можно записать

$$P(M_i|E_j) = \frac{P(M_iE_j)}{P(E_j)} = \frac{P(M_i)P(E_j|M_i)}{P(E_j)}.$$

Принимая во внимание (7.1), получаем

$$P\left(M_i|E_j\right) = \frac{P(M_i|E_j)P(E_j|M_i)}{P(E_j)},$$

т.е.

$$\frac{P(E_j|M_i)}{P(E_j)} = 1.$$

Таким образом, (7.2) доказано. Обратное утверждение теоремы доказывается «обратным ходом» приведенных равенств .

Шифр Вернама

1917 г.

Гилберт Вернам

(Gilbert Vernam)

изобрёл систему автоматического шифрования телеграфных сообщений

$$0 \oplus 0 = 0,$$
$$0 \oplus 1 = 1,$$
$$1 \oplus 0 = 1,$$
$$1 \oplus 1 = 0.$$

Это операция – исключающее ИЛИ – записывается двумя обозначениями: «⊕» и XOR – английским сокращением от исключающее ИЛИ (англ. eXclusive OR). В отношении сложения по модулю 2 интересен еще один факт – обратный к нему, т.е. восстанавливающей операцией является кроме вычитания по модулю 2 еще и сама эта операция, т. е. для любых X и Y выполняется

$$(X \oplus Y) \oplus Y = X$$

Пусть множество сообщений M состоит из слов двоичного алфавита длины n, т.е. всего сообщений не более, чем 2^n. В шифре Вернама множество ключей также состоит из слов той же длины n и каждый ключ используется с вероятностью $1/2^n$. Другими словами, все ключи используются с одинаковой вероятностью.

Пусть необходимо зашифровать сообщение $\bar{m} = m_1 m_2 \dots m_n$ и пусть выбран ключ $\bar{k} = k_1 k_2 \dots k_n$. Тогда зашифрованное сообщение $\bar{e} = e_1 e_2 \dots e_n$ получается по формуле:

$$e_i = m_i \oplus k_i, \tag{7.3}$$

где $i = 1, 2, \dots, n$ и $\oplus$ обозначает сложение по модулю 2. Другими словами, сообщение шифруется по схеме

$$\begin{array}{cccccc} & m_1 & m_2 & \dots & m_n \\ \oplus & k_1 & k_2 & \dots & k_n \\ \hline & e_1 & e_2 & \dots & e_n \end{array}.$$

Так как сложение и вычитание по модулю 2 совпадают, то легко видеть, что дешифрование осуществляется по формуле

$$m_i = e_i \oplus k_i. \tag{7.4}$$

Пример 7.1. Пусть $\bar{m} = 01001$, $\bar{k} = 11010$. Тогда получаем $\bar{e} = 10011$. Сложив $\bar{e}$ с $\bar{k}$, восстанавливаем $\bar{m}$.

Теорема 7.2. *Шифр Вернама является совершенно секретной криптосистемой.*

Доказательство. Согласно теореме 7.1 достаточно доказать справедливость (7.2). Имеем

$$
\begin{aligned}
P(E_j|M_i) &= P(e^n|m^n) = \\
&= P(k_1 = e_1 \oplus m_1;\ k_2 = e_2 \oplus m_2; \ldots; k_n = e_n \oplus m_n) = \\
&= P(\text{ключ} = k_1 \ldots k_n) = 2^{-n}
\end{aligned}
$$

(в последнем равенстве мы использовали то, что, по предположению, все ключи равновероятны).

Найдем $P(E_j)$. По формуле полной вероятности

$$P(E_j) = \sum_{i=1}^{2^n} P(M_i)P(E_j|M_i).$$

Учитывая, что $P(E_j|M_i) = 2^{-n}$, получаем

$$P(E_j) = 2^{-n} \sum_{i=1}^{2^n} P(M_i).$$

Так как сумма вероятностей всех возможных сообщений равна 1, получаем

$$P(E_j) = 2^{-n}.$$

Таким образом, справедливо (7.2). Теорема доказана.