Теоретическая часть.

Дискретный канал без шума

Телетайп и телеграф - два простых примера дискретных каналов для передачи информации. В общем случае будем называть дискретным каналом систему, посредством которой последовательность выборов из набора элементарных символов $S_1, \ldots, S_n$ может быть передана из одной точки в другую. Каждый из этих символов S_i предполагается имеющим некоторую протяженность во времени t_i (не обязательно одинаковую для различных символов). Не требуется возможность передачи произвольной последовательности символов, вполне допустим вариант с ограниченным набором разрешенных последовательностей. Это - возможные сигналы в канале. К примеру, в телеграфии такими символами являются: *(1)* точка, соответствующая замыканию линии на единичное время и размыканию той же длительности, *(2)* тире, соответствующее замыканию линии на три единицы времени и размыканию на одну, *(3)* промежуток между буквами, соответствующий, к примеру, размыканию линии на три единицы времени, и *(4)* промежуток между словами, соответствующий размыканию линии на шесть единиц времени. Мы можем ограничить возможные последовательности символов, запретив к примеру последовательные промежутки (так как два следующих друг за другом буквенных промежутка идентичны промежутку между словами). Вопрос, который мы сейчас рассмотрим, состоит в том, как определить пропускную способность такого канала.

В случае телетайпа, в котором все символы имеют одну и ту же длительность и разрешена любая комбинация 32-х символов, ответ прост. Каждый символ представляет собой пять бит информации. Если система связи передает n символов в секунду, естественно сказать, что пропускная способность канала $5n$ бит в секунду. Это не значит, что телетайп всегда передает информацию с такой скоростью - это лишь максимально возможный темп, и будет или нет он достигаться - зависит от источника информации, подаваемой в канал (см. далее)

Для более общего случая различной длины символов и ограничений на разрешенные последовательности дадим следующее определение:

***Определение**: Пропускная способность C канала дается выражением*

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T}$$

где $N(T)$ - число возможных сигналов длительности T.

Легко видеть, что в случае телетайпа этот результат сводится в предыдущему. Можно показать, что этот предел в большинстве интересующих нас случаев существует и является конечным. Предположим, что разрешены все

последовательности символов $S_1, \ldots, S_n$ длительностью $t_1, \ldots, t_n$ соответственно. Какова пропускная способность канала? Если $N(t)$ - число последовательностей длительности t,

$$N(t) = N(t - t_1) + N(t - t_2) + \ldots + N(t - t_n).$$

Полное число равно сумме чисел последовательностей, оканчивающихся на $S_1, S_2, \ldots, S_n$, то есть $N(t - t_1), N(t - t_2), \ldots, N(t - t_n)$, соответственно. Согласно широко известному результату дискретного анализа, $N(t)$ при больших t асимптотически стремится к X_0^t, где X_0 - наибольший действительный корень характеристического уравнения:

$$X^{-t_1} + X^{-t_2} + \ldots + X^{-t_n} = 1$$

и следовательно

$$C = \log X_0.$$

При наличии ограничений на разрешенные последовательности символов мы зачастую также можем получить разностное уравнение такого типа и найти C из характеристического уравнения. В вышеупомянутом случае телеграфной системы

$$N(t) = N(t-2) + N(t-4) + N(t-5) + N(t-7) + N(t-8) + N(t-10)$$

что получается при учете последнего и предпоследнего символов. Следовательно, C равно $-\log \mu_0$, где μ_0 - положительный корень уравнения $1 = \mu^2 + \mu^4 + \mu^5 + \mu^7 + \mu^8 + \mu^{10}$. Решая его, находим $C = 0.539$.

Достаточно общим видом ограничений на возможные последовательности символов может быть следующий. Рассмотрим множество возможных состояний $a_1, a_2, \ldots, a_m$. В каждом состоянии могут передаваться лишь некоторые символы из набора $S_1, \ldots, S_n$ (различные подмножества в различных состояниях). При передаче одного из этих символов состояние изменяется в зависимости как от предыдущего состояния, так и от переданного символа. Простым примером такой системы является телеграф, который может находиться в одном из двух состояний в зависимости от того, был ли последним переданным символом 'пробел'. Если да, то может быть посланы только точка или тире, и состояние изменится на противоположное. Если же нет, может быть послан произвольный символ, и состояние изменится, если посланный символ - 'пробел'. Эта ситуация иллюстрируется линейным графом, изображенным на рис.2.

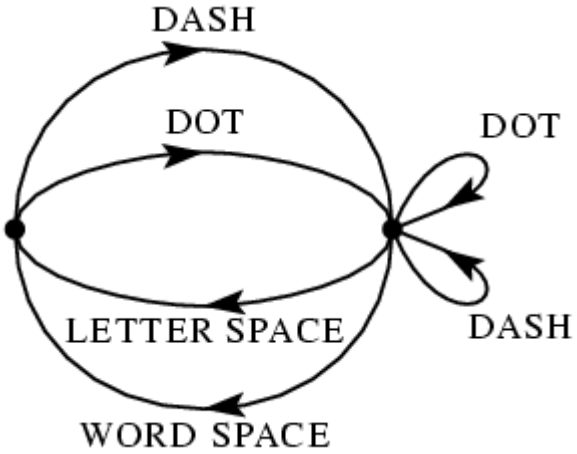


Рис. 2. Графическое представление ограничений на последовательности телеграфных символов.

Вершины графа соответствуют двум состояниям системы, а линии - разрешенным для передачи символам. В Приложении 1 показано, что при таких ограничениях на разрешенные к передаче символы C существует и может быть рассчитана в соответствии со следующим результатом:

Теорема 1. *Пусть* $b_{ij}^{(s)}$ *- длительность символа* s*, разрешенного в состоянии* i *и приводящего к переходу системы в состояние* j*. Тогда пропускная способность канала* C *равняется* $\log W$*, где* W *- наибольший действительный корень детерминистического уравнения*

$$\left| \sum_s W^{-b_{ij}^{(s)}} - \delta_{ij} \right| = 0$$

где $\delta_{ij} = 1$ *при* $i = j$ *и равно нулю иначе.*

К примеру, в случае телеграфа (рис. 2) детерминант равен

что при раскрытии дает вышеприведенное уравнение для данного случая.

Дискретный источник информации

Мы видели, что при достаточно общих условиях логарифм числа возможных сигналов в дискретном канале линейно растет со временем. Пропускная способность канала может быть охарактеризована темпом этого роста, числом бит в секунду, необходимых для передачи данного конкретного сигнала.

Рассмотрим теперь источник информации. Как он может быть описан математически, и сколько бит информации в секунду он производит? Важным тут является знание о статистике этого источника, что позволяет понизить требуемую пропускную способность канала выбором соответствующей кодировки - представления информации. В телеграфии, к примеру, передаваемые сообщения состоят из последовательностей букв. Эти последовательности, однако, не являются совершенно случайными. В общем слуяае они образуют предложения и имеют статистическую природу, скажем,

английского языка. Буква **E**встречается чаще **Q**, последовательность **TH** - чаще, чем **XP**. Наличие такой структуры позволяет получить выигрыш во времени передачи сообщения (или пропускной способности канала), соответствующим образом его кодируя. Именно такой подход используется в телеграфии, где самый короткий символ - точка - исплльзуется для наиболее часто используемой английской буквы **E**, в то время как самые редкие - **Q**, **X**, **Z** - представляются более длинными последовательностями точек и тире. Еще больший выигрыш во времени передачи достигается в некоторых коммерческих системах кодирования, в которых наиболее распространенные фразы и выражения заменяются четырех- или пятибуквенными комбинациями. Использующиеся в настоящее время стандартизованные поздравительные и приветственные телеграммы также позволяют кодировать церые предложения достаточно короткими последовательностями чисел. в настоящее время

Можно считать, что дискретный источник формирует сообщение символ за символом, выбирая их в соответствии с некоторой вероятностью, зависящей как от номера символа, так и от предыдущих выборов. Физическая система, или же математическая модель системы, генерирующей такую последовательность символов в соответствии с набором вероятностей, представляет собой стохастический процесс (См., например, S. Chandrasekhar, ``Stochastic Problems in Physics and Astronomy,'' *Reviews of Modern Physics*, v. 15, No. 1, January 1943, p. 1.). Таким образом, мы можем рассматривать дискретный источник как стохастический процесс, и наоборот, любой стохастический процесс, генерирующий дискретную последовательность символов из ограниченного множества можно считать дискретным источником. Это включает в себя:

- Естественные языки, такие, как английский, немецкий и китайский.
- Непрерывные источники информации, которая может быть дискретизована в результате некоторой операции, например, оцифрованная речь в PCM-передатчике, или дискретный телевизионный сигнал.
- Чисто математические случаи, когда мы абстрактно определяем некоторый стохастический процесс, генерирующий последовательность символов. Дадим несколько примеров этого типа.
  - **(A)** Пусть у нас есть пять букв **A**, **B**, **C**, **D**, **E**, которые выбираются с равной вероятностью 0.2, причем каждый из этих выборов не зависит от предыдущих. Тогда типичной последовательностью символов будет, к примеру,
  - B D C B C E C C C A D C B D D A A E C E E A
  - A B B D A E E C A C E E B A E E C B C E A D.

Данная последовательность была построена с использованием таблицы случайных чисел(Kendall and Smith, *Tables of Random Sampling Numbers*, Cambridge, 1939.).

  - **(B)** Используя те же самые пять букв, но с вероятностями 0.4, 0.1, 0.2, 0.2, 0.1 соответственно, получаем следующее типичное сообщение:
  - A A A C D C B D C E A A D A D A C E D A
  - E A D C A B E D A D D C E C A A A A A D.

- **(C)** Более сложная структура может быть получена при отказе от взаимной независимости отдельных процедур выбора - то есть тогда, когда каждый последующий символ зависит от предыдущих. В простейшем случае каждый символ зависит лишь от предыдущего и не зависит от более ранних. Статистическая структура тогда может быть представлена набором вероятностей перехода $p_i(j)$, то есть вероятностей того, что за символом i будет следовать символ j. Вторым равноценным методом описания такой структуры являются вероятности 'диграмм' (двухбуквенных комбинаций $i\ j$) $p(i,j)$. Частоты встречаемости букв $p(i)$ (вероятность буквы i), вероятности перехода $p_i(j)$ и вероятности диграмм $p(i,j)$ связаны следующими выражениями:

Для примера возьмем три буквы **A**, **B**,**C** с таблицами вероятностей:

$p_i(j)$		j		
		A	B	C
i	A	0	$\frac{4}{5}$	$\frac{1}{5}$
	B	$\frac{1}{2}$	$\frac{1}{2}$	0
	C	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{10}$

i	$p(i)$
A	$\frac{9}{27}$
B	$\frac{16}{27}$
C	$\frac{2}{27}$

$p(i,j)$		j		
		A	B	C
i	A	0	$\frac{4}{15}$	$\frac{1}{15}$
	B	$\frac{8}{27}$	$\frac{8}{27}$	0
	C	$\frac{1}{27}$	$\frac{4}{135}$	$\frac{1}{135}$

Типичным сообщением от такого источника будет:

A B B A B A B A B A B A B A B B B A B B B B B A B A B A B A B A B B B A C A C A B B A B B B B A B B A B A C B B B A B A.

Следующим по сложности будет учет частот встречаемости триграмм, то есть ситуация, при которой выбор текущего символа зависит лишь от двух предшествующих. В данной ситуации требуется знание вероятностей триграмм $p(i,j,k)$, или, что то же самое, вероятностей переходов $p_{ij}(k)$. При дальнейшем обобщении можно получить и сигналы с более сложной структурой. Так, в общем случае требуется знание набора вероятностей n-грамм $p(i_1,i_2,\ldots,i_n)$ или же вероятностей перехода $p_{i_1,i_2,\ldots,i_{n-1}}(i_n)$.

- **(D)** Можно также определить стохастический процесс, генерирующий текст, являющийся последовательностью ``слов''. Пусть есть пять букв **A**, **B**, **C**, **D**, **E** и 16 ``слов'' языка с соответствующими вероятностями:

.10 A .16 BEBE .11 CABED .04 DEB
.04 ADEB .04 BED .05 CEED .15 DEED
.05 ADEE .02 BEED .08 DAB .01 EAB
.01 BADD .05 CA .04 DAD .05 EE

- Пусть эти слова выбираются независимо и разделяются пробелом. Тогда типичным сообщением будет:

- DAB EE A BEBE DEED DEB ADEE ADEE EE DEB
- BEBE BEBE BEBE ADEE BED DEED DEED CEED
- ADEE A DEED DEED BEBE CABED BEBE BED DAB
- DEED ADEB.
- Если все слова имеют ограниченную длину, данный процесс сводится к одному из вышеописанных, однако описание его будет проще в терминах структуры и вероятностей слов. Можно пойти далее и ввести вероятности перехода между словами, и так далее.

Такие искусственные языки полезны для формулировки простых задач и примеров различных возможностей. Мы можем также приближенно описать некоторый естественный язык серией простых искусственных. Нулевым приближением такого рода будет модель с равновероятными независимыми буквами, следующим шагом может служить учет различных вероятностей букв в естественном языке (Вероятности букв, диграмм и триграмм даются в *Secret and Urgent* by Fletcher Pratt, Blue Ribbon Books, 1939. Частоты встречаемости отдельных слов приведены в *Relative Frequency of English Speech Sounds,* G. Dewey, Harvard University Press, 1923.). Так, в первом приближении для английского языка буква **E** будет выбираться с вероятностью 0.12, а **W** - 0.02, но не будет никакого влияния букв друг на друга, и не будет заметна тенденция к образованию наиболее частых буквосочетаний, таких как **TH** или **ED**. Во втором приближении необходимо учесть структуру диграмм - после выбора одной из букв следующая выбирается уже согласно соответствующей условной вероятности, что требует знания частоты диграмм $p_i(j)$. На следующем шаге вводится структура триграмм - каждая буква зависит уже от двух предшествующих.

Постепенное приближение к английскому языку

Для иллюстрации того, как эта серия приближений описывает естественный язык, были построены типичные последовательности приближений для английского языка. Во всех случаях использовался 27-символьный ``алфавит'' - 26 букв и пробел.

- Нулевое приближение (символы независимы и равновероятны).
- XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.
- Первое приближение (символы независимы, но частоты из соответствуют английскому тексту).
- OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.
- Второе приближение (учтена диграммная структура английского языка).
- ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE
- AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.
- Третье приближение (учтена триграммная структура).
- IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES
- OF THE REPTAGIN IS REGOACTIONA OF CRE.
- Четвертое приближение. Вместо дальнейшего использования тетраграмм, $\cdots$, n-грамм проще перейти к словарной структуре. В данном

приближении слова независимы, но их частоты соответствуют английскому тексту.

- REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL
- HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE
- LINE MESSAGE HAD BE THESE.
- Приближение второго порядка с использованием слов - учтены их вероятности перехода.
- THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER
- OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME
- OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

С каждым шагом заметно возрастает сходство с английским языком. Заметим, что приведенные примеры сохраняют достаточно похожую на оригинал структуру и далеко за рамками соответствующего приближения. Так, второе приближение дает приемлемый текст на уровне двухбуквенных сочетаний, но четырехбуквенные комбинации в нем также являются достаточно схожими с английским языком; фигурирующие в последнем примере комбинации четырех и более слов также вполне могут встречаться в реальных предложениях. Так, фраза из 10 слов ``attack on an English writer that the character of this" не является совсем уж бессмысленной. Видно, что достаточно сложный стохастический процесс может являться вполне удовлетворительной моделью дискретного источника.

Два первых примера построены с использованием книги случайных чисел и таблицы частот букв. Похожий метод можно было бы использовать и для последующих, так как частоты диграмм и триграмм известны, однако мы воспользовались иным, более простым подходом. Для построения второго приближения, например, открывалась случайным образом книга и выбиралась случайная буква, которая записывалась; затем книга открывалась на другой странице и искалась первая двухбуквенная комбинация, начинающаяся с уже отобранной буквы, и записывалась следующая буква. Затем процедура повторялась. Похожий подход использовался и для остальных примеров. Было бы интересным построить также и следующие приближения, однако это требует слишком больших затрат времени и сил.

Представление марковского процесса в виде графа

Стохастические процессы описанного выше типа в математике известны как марковские процессы и достаточно широко освещены в литературе (за подробным изложением отсылаем читателя к книге M. Frechet, *Methode des fonctions arbitraires. Theorie des evenements en chane dans le cas d'un nombre fini d'etats possibles*, Paris, Gauthier-Villars, 1938.). В общем случае их можно описать следующим образом. Существует конечное число возможных состояний системы $S_1, S_2, \ldots, S_n$, кроме того, известен набор вероятностей перехода $p_i(j)$ системы из состояния S_j в состояние S_i. Чтобы сделать такой марковский процесс источником информации, достаточно предположить, что он выдает по

одной букве в момент каждого перехода. Различные состояния системы в таком случае соответствуют ``остаточному влиянию" предшествующих букв.

Эту ситуацию можно представить в виде графов, как показано на рис.3, 4 и 5.

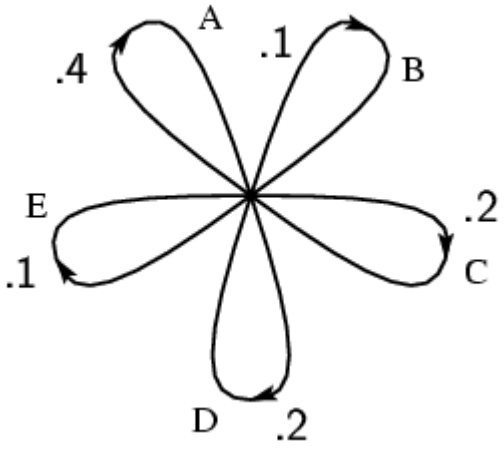


Рис.3. Граф, соответствующий источнику из примера **B**.

Состояния здесь представлены вершинами графа, а вероятности и соответствующие им буквы приведены рядом с соответствующими линиями переходов. Рис. 3 соответствует примеру **(B)** главы 2, а рис. 4 - примеру **(C)**

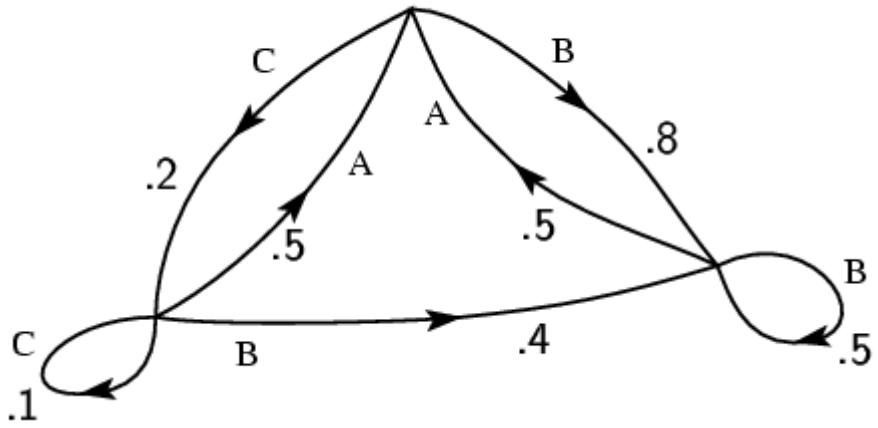


Рис.4. Граф, соответствующий источнику из примера **C**.

У системы, изображенной на рис. 3, есть лишь одно состояние, так как выбираемые буквы независимы. На рис. 4 число состояний равно числу букв; в случае учета триграммной структуры их число равно n^2, что соответствует возможным парам предшествующих букв. На рис. 5 изображен граф, соответствующий структуре слов из примера **(D)** (*S* обозначает символ пробела).

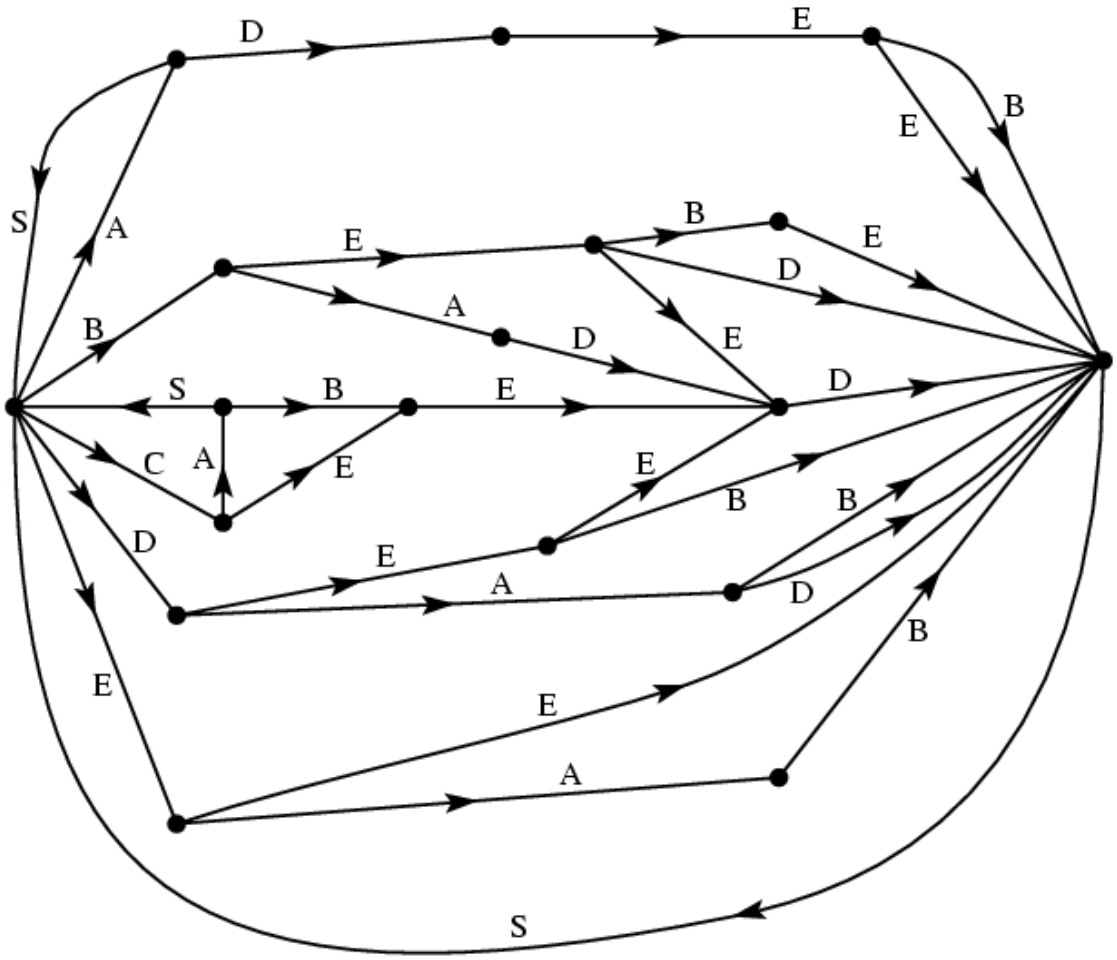


Рис.5. Граф, соответствующий источнику из примера **D**.

Эргодические и смешанные источники

Как уже было отмечено, дискретный источник в наших задачах может рассматриваться как марковский процесс. Среди всех возможных марковских процессов можно выделить группу со свойствами, важными для теории связи. В этот класс входят так называемые ``эргодические" процессы, и мы будем называть соответствующие источники эргодическими. Хотя строгое определение эргодического процесса достаточно запутанно, основная идея проста. Для эргодического процесса все сгенерированные последовательности обладают одинаковыми статистическими свойствами, то есть, к примеру, частоты встречаемости букв, диграмм и тек далее, оцененные по отдельным последовательностям, сходятся с ростом длины выборок к определенным пределам, не зависящим от последовательности. На самом деле это верно не для всякой последовательности, однако множество последовательностей, для которых это не выполняется, имеет меру ноль (то есть обладает нулевой вероятностью). Грубо свойство эргодичности означает статистическую однородность.

Все вышеприведенные примеры искусственных языков являются эргодическими. Это связано со структурой соответствующих графов. Если граф

обладает двумя следующими свойствами, соответствующий процесс будет эргодическим:

- Граф не состоит из двух изолированных частей **A** и **B**, таких, что невозможно перейти с вершин одной части на вершины другой по линиям графа в разрешенном направлении, и с вершин второй - на вершины первой.
- Назовем замкнутые серии ребер графа, которые можно обойти в разрешенном направлении, ``контуром". Длиной контура назовем число ребер в нем. Тогда на рис. 5 серия ребер **BEBES** образует контур длины 5. Вторым требуемым свойством является равенство наибольшего общего делителя длин всех контуров на графе единице.

Если выполняется только первое из условий, а второе нарушено, т.е. наибольший общий делитель равен $d > 1$, последовательность имеет некоторую периодическую структуру. Различные последовательности распадаются на d статистически равнозначных классов, отличающихся лишь сдвигом начала (то есть тем, какая именно буква называется первой). Сдвигом от 0 до $d-1$ любая из этих последовательностей может быть сделана статистически эквивалентной любой другой. Простым примером с $d = 2$ может служить случай с тремя буквами a, b, c. За буквой a следует либо b, либо c с вероятностями $\frac{1}{3}$и $\frac{2}{3}$ соответственно. За буквами же b или c всегда следует a. Тогда типичной последовательностью будет

a b a c a c a c a b a c a b a b a c a c.

Данный случай не очень важен для нашей работы.

Если же нарушается первое условие, граф распадается на несколько подграфов, на каждом из которых это условие выполнено. Мы будем предполагать, что на каждом из них выполнено также и второе условие. В это случае мы имеем ``смешанный" источник, состоящий из нескольких чистых, соответствующих каждому из субграфов. Если L_1, L_2, $L_3, \cdots$ - чистые компоненты, можно записать

где p_i - вероятность компонента источника L_i.

Физически ситуация представляет из себя следующее. Есть несколько различных источников L_1, L_2, $L_3, \cdots$, каждый из которых статистически однороден (т.е. эргодичен). Мы не можем сказать **a priori**, который из них будет использован, но если последовательность начинается в каком-то из чистых компонентов, в дальнейшем она в нем и останется, бесконечно продолжаясь в соответствии с его статистической структурой.

Для примера возьмем два из вышеприведенных процессов и примем $p_1 = .2$ и $p_2 = .8$. Последовательность от смешанного источника

$$L = .2L_1 + .8L_2$$

может быть получена выбором L_1 либо L_2 с вероятностями 0.2 и 0.8 и генерированием последовательности в соответствием с этим выбором.

В дальнейшем мы будет предполагать источник эргодическим, если специально не указано обратное. Это предположение позволяет отождествить средние по последовательности со средними по ансамблю всех возможных последовательностей (вероятность отличия равна нулю). К примеру, относительная частота буквы **A** в некоторой бесконечной последовательности с вероятностью единица будет равна относительной частоте ее в ансамбле последовательностей.

Если P_i - вероятность состояния i и $p_i(j)$ - вероятность перехода в в состояние j, ясно, что для стационарности процесса P_iдолжна удовлетворять условию равновесия:

$$P_j = \sum_i P_i p_i(j).$$

В эргодическом случае можно показать, что при любых начальных условиях вероятности $P_j(N)$ оказаться в состоянии j после N символов стремятся к равновесному значению при $N \to \infty$.

Выбор, неопределенность и энтропия

Мы представили дискретный источник информации марковским процессом. Можем ли мы определить некоторую величину, характеризующую в некотором смысле количество информации, ``производимой" таким процессом, или, точнее, темп ``производства" информации?

Пусть у нас есть набор возможных событий с вероятностями $p_1, p_2, \ldots, p_n$. Эти вероятности известны и больше про эти события ничего не известно. Можно ли найти меру совершаемого ``выбора" одного из событий или же неопределенности получаемого результата?

Разумно потребовать от такой меры, скажем, $H(p_1, p_2, \ldots, p_n)$, следующих свойств:

- H должна быть непрерывна по p_i.
- В случае равенства всех p_i, $p_i = \frac{1}{n}$, H должна быть монотонно возрастающей функцией n. Для равновероятных событий неопределенность, или возможность выбора, возрастает с ростом числа возможных событий.
- При разбиении одного из возможных выборов на два, исходная величина H должна быть взвешенной суммой отдельных величин H. Смысл этого иллюстрируется на рис. 6.

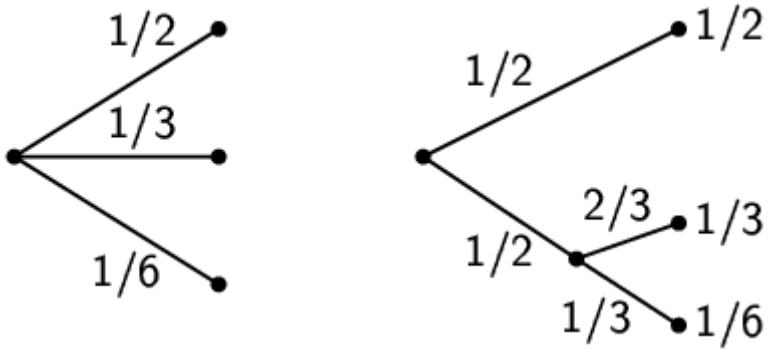


Рис.6. Разложение выбора из трех возможностей.

Слева мы имеем три возможности $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. Справа, мы сначала выбираем между двумя возможностями вероятностью $\frac{1}{2}$ каждая, и, в случае выбора второй, делаем следующий выбор с вероятностями $\frac{2}{3}$, $\frac{1}{3}$. Конечные результаты имеют те же вероятности, что и раньше. В данном конкретном случае требование сводится к

Коэффициент $\frac{1}{2}$ возникает из-за того, что второй выбор делается лишь в половине случаев.

В приложении 2 доказывается следующий результат:

Теорема 2: *Единственная H, удовлетворяющая трем вышеприведенным условиям, имеет вид:*

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

где K - некоторая положительная константа.

Эта теорема, как и предположения, требуемые для ее доказательства, никак не будут использоваться в дальнейшем. Они приведены главным образом для придания некоторой правдоподобности нашим последующим определениям. Их истинным доказательством, однако, послужат их приложения.

Величины вида $H = -\sum p_i \log p_i$ (константа K отвечает за выбор системы единиц) играют центральную роль в теории информации как меры информации, выбора и неопределенности. Такой же вид имеет выражение для энтропии в некоторых формулировках статистической механики (см., к примеру, R. C. Tolman, *Principles of Statistical Mechanics,* Oxford, Clarendon, 1938.), где p_i - вероятность найти систему в ячейке i ее фазового пространства. H тогда, к примеру, совпадает с H в знаменитой теореме Больцмана.
Назовем $H = -\sum p_i \log p_i$ энтропией набора вероятностей $p_1, \ldots, p_n$. Будем обозначать $H(x)$ энтропию случайной величины x; здесь x - не аргумент функции, а метка, призванная отличать обозначение энтропии величины x от энтропии $H(y)$ случайной величины y.

Энтропия для случая двух возможностей с вероятностями p и $q = 1 - p$,

$$H = -(p \log p + q \log q)$$

изображена на рис. 7 как функция p.

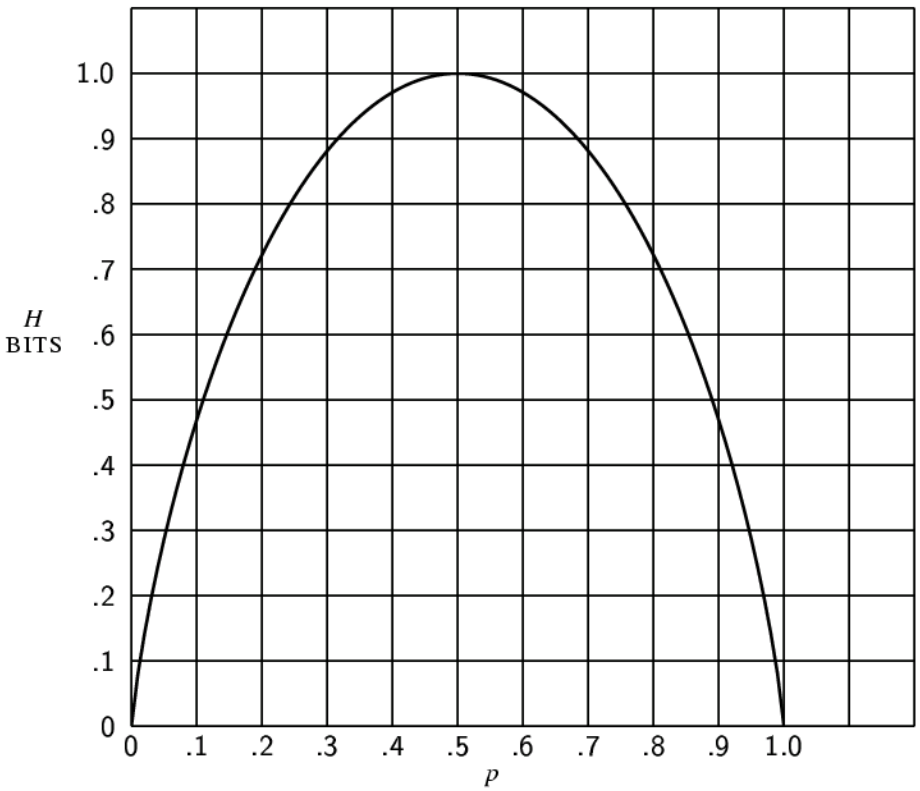


Рис.7. Энтропия для случая двух возможностей с вероятностями p и $q = 1 - p$.

Величина H обладает несколькими интересными свойствами, делающими ее приемлемой мерой выбора или информации:

1. $H = 0$ тогда и только тогда, когда все p_i за исключением одного равны нулю, а один - единице, то есть величина Hобращается в ноль лишь тогда, когда мы уверены в выборе. Во всех остальных случаях H положительна.

2. Для данного n H принимает максимальное значение тогда, когда все p_i равны между собой (и, соответственно, равны $\frac{1}{n}$). Интуитивно понятно, что это - наиболее неопределенная ситуация.

3. Рассмотрим два события, x и y, причем первому соответствует m возможностей, а второму - n. Пусть $p(i, j)$ - вероятность совместного появления i для первого и j для второго. Энтропия такого совместного события равна

$$H(x, y) = -\sum_{i,j} p(i, j) \log p(i, j)$$

в то время как

Легко показать, что

$$H(x,y) \leq H(x) + H(y)$$

причем равенство достигается лишь при независимости событий (т.е. при $p(i,j) = p(i)p(j)$). Неопределенность совместного события меньше или равна сумме неопределенностей событий по-отдельности.

4. Любое изменение, направленное на уравнивание вероятностей $p_1, p_2, \ldots, p_n$, увеличивает H. Так, если $p_1 < p_2$ и мы увеличиваем p_1, уменьшая p_2 на ту же величину, так что p_1 и p_2 оказываются ближе бруг к другу, величина H возрастает. В более общем случае, если мы производим некоторую процедуру ``усреднения" p_i вида

$$p_i' = \sum_j a_{ij} p_j$$

где $\sum_i a_{ij} = \sum_j a_{ij} = 1$ и все $a_{ij} \geq 0$, величина H возрастает (за исключением специального случая, при котором значения p_iпросто меняются местами; в этом случае H, естественно, остается неизменной).

5.Рассмотрим два случайных события, x и y, как в пункте 3, не обязательно независимых. Для любого значения i, которое может принимать x, существует условная вероятность $p_i(j)$ того, что y примет значение j. Она имеет вид

$$p_i(j) = \frac{p(i,j)}{\sum_j p(i,j)}.$$

Определим **условную энтропию** величины y, $H_x(y)$, как среднюю энтропию y при каждом значении x, взвешенную в соответствии с вероятностью получения данного значения x. Она равна

$$H_x(y) = -\sum_{i,j} p(i,j) \log p_i(j) .$$

Это величина показывает, насколько в среднем является неопределенной y, если мы знаем x. Подставляя $p_i(j)$, получаем

или

$$H(x,y) = H(x) + H_x(y).$$

Неопределенность (или энтропия) совместного события x, y равна сумме неопределенности x и неопределенности y при известной величине x.

6. Из пунктов 3 и 5 имеем

$$H(x) + H(y) \geq H(x,y) = H(x) + H_x(y).$$

Таким образом,

$H(y) \geq H_x(y)$.

Неопределенность y никогда не увеличивается при конкретизации x. Она уменьшается, если только x и y не являются независимыми событиями. В последнем же случае она остается неизменной.

Энтропия источника информации

Рассмотрим дискретный источник информации описанного выше типа. Для каждого из состояний i существует набор вероятностей $p_i(j)$ генерации различных возможных символов j, что соответствует энторпии данного состояния H_i. Определим энтропию источника как среднее всех H_i, взвешенных согласно вероятностям соответствующих состояний:

Это энтропия в расчете на символ текста. Если марковский процесс имеет определенный темп, можно также определить энтрпоию в секунду

$$H' = \sum_i f_i H_i$$

шде f_i - средняя частота (реализаций в секунду) состояния i. Ясно, что

$$H' = mH$$

где m - среднее число символов, генерируемых за секунду. H или H' характеризуют количество информации, производимой источником в расчете на символ или на секунду. Если за основание логарифма принята двойка, они представляют собой биты на символ или биты в секунду.

Если последовательные символы независимы, то H есть просто $-\sum p_i \log p_i$, где p_i - вероятность символа i. Пусть мы имеем достаточно длинное сообщение, содержащее N символов. С высокой вероятностью в нем будет примерно $p_1 N$ вхождений первого символа, $p_2 N$ - второго, и так далее. Таким образом вероятность данного собщения будет примерно равна

или же

H , таким образом, есть логарифм обратной вероятности типичной длинной последовательности, деленный на число символов в ней. Аналогичный результат имеет место для произвольного источника. Более точно, можно сформулиовать (см. приложение 3)

Теорема 3: *Для произвольных $\epsilon > 0$ и $\delta > 0$ можно найти такое N_0, что последовательности произвольной длины $N \geq N_0$распадаются на два класса:*

- *Множество последовательностей с полной вероятностью, меньшей ϵ.*

- *Остальные последовательности, каждая из которых имеет вероятность, удовлетворяющую неравенству*

$$\left|\frac{\log p^{-1}}{N} - H\right| < \delta.$$

Иными словами, практически наверняка достаточно близко к H при достаточно больших N.

Похожий результат имеет место для числа последовательностей с различными вероятностями. Рассмотрим вновь последовательности длины N и отсортируем их по убыванию вероятности. Определим $n(q)$ как число послдовательностей, которые мы должны выбрать из данного набора, начиная с наиболее вероятной, для получения суммарной вероятности q.

Теорема 4:

$$\lim_{N\to\infty} \frac{\log n(q)}{N} = H$$

при q, не равных 0 или 1.

$\log n(q)$ может быть истолковано как число бит, требуемых для выделения последовательности, если мы рассматриваем лишь наиболее вероятные из них суммарной вероятностью q. Тогда - число бин на символ для такого выделения. Теорема гласит, что для больших N оно не зависит от q и равно H. Темп роста логарифма числа достаточно вероятных последовательностей дается H вне зависимости от того, как именно мы определяем эту ``достаточную вероятность". Благодаря этим результатам (доказываемым в приложении 3) для большинства задач можно считать, что (достаточно длинных) последовательностей всего 2^{HN}, и каждая из них имеет вероятность 2^{-HN}.

Две следующих теоремы показывают, что H и H' могут быть определены посредством пределных переходов прямо из статистики сообщения, без конкретизации различных состояний и вероятностей перехода.

Теорема 5: *Пусть $p(B_i)$ - вероятность последовательности символов B_i от источника, и пусть*

$$G_N = -\frac{1}{N}\sum_i p(B_i)\log p(B_i)$$

где суммирование производится по всем последовательностям B_i, содержащим N символов. Тогда G_N - монотонно убывающая функция N и

$$\lim_{N\to\infty} G_N = H.$$

Теорема 6: *Пусть* $p(B_i, S_j)$ *- вероятность того, что за последовательностью* B_i *следует символ* S_j*, и* $p_{B_i}(S_j) = p(B_i, S_j)/p(B_i)$ *- условная вероятность* S_j *после* B_i*. Пусть*

$$F_N = -\sum_{i,j} p(B_i, S_j) \log p_{B_i}(S_j)$$

где суммирование производится по всем блокам B_i *длиной* $N-1$ *символов и по всем символам* S_j*. Тогда* F_N *- монотонно убывающая функция* N*,*

и $\lim_{N\to\infty} F_N = H$.

Эти результаты получаются в приложении 3. Они показывают, что серия приближений к H может быть получена при рассмотрении статистической структуры последовательностей длиной $1, 2, \ldots, N$ символов. F_N является наилучшим приближением. Фактически F_N является энтропией N-того приближения для источника, описанного выше типа. Если статистическое влияние на расстояниях, больших N, отсутствует, то есть если условная вероятность последующего символа при знании $(N-1)$ предыдущего не изменяется конкретизацией любого предшествующего им, то $F_N = H$. F_N, естественно, является условной энтропией последующего симвода при знании $(N-1)$ предыдущих, тогда как G_N - энтропия на символ блоков длины N.

Назовем отношение энтропии источника к максимальному значению, которое она может принимать, будучи ограниченной теми же предшествующими символами, **относительной энтропией**. Это - наибольший возможный коэффициент сжатия при кодировании с использованием того же самого алфавита. Назовем разность единицы и относительной энтропии**избыточностью**. Избыточность английского языка, при рассмотрении статистических структур на масштабах не более восьми букв, составляет порядка 50%. Это значит, что: когда мы пишем английский текст, половина его определяется статистической структурой языка, а половина выбирается нами свободно. Оценка 50% получается несколькими независимыми методами, дающими примерно одинаковый результат, - расчетом энтропии приближений к английскому языку, удалением с последующим восстановлением части букв из английского текста, и так далее. Так, если текст может быть полностью восстановлен после удаления из него половины букв, его избыточность должна быть больше 50%. Еще один метод основан на некторвых известных результатах криптографии.

Две крайности избыточности английской прозы - Основной Английский (Бейсик Инглиш, Basic English) и книга Джеймса Джойса ``Поминки по Финнегану''. Словарь Бейсик Инглиш ограничен 850 словами, и его избыточность очень велика, что отражается в уведичении обьема сообщения при переаоде его в данный диалект. Джойс же, наоборот, расширяет лексикон, достигая таким образом уплотнения семантического содержания.

С избыточностью языка связано существование загадок-кроссвордов. При нулевой избыточности любое буквосочетание является приемлемым текстом, и любой двумерный массив букв образует кроссворд. Если же избыточность достаточно велика, язык налашаем слишком много ограничений, что делает невозможным существование больших кроссвордов. Более тщательный анализ показывает, что при случайной природе ограничений языка большие кроссворды возможны лишь при избыточности его порядка 50%. При избыточности 30% становятся возможными трехмерные кроссворды, и так далее.

Представление операций кодирования и декодирования

Мы все еще не представили математически операции, производимые передатчиком и премником при кодировании и декодировании информации. Назовем оба этих прибора $\text{дискретным преобразователями}$. Преобразователь получает на вход некоторый набор входных символов, и выдает набор выходных символов. Преобразователь может иметь внутреннюю память, так что информация на выходе будет зависеть не только от текущего поступившего на вход символа, но также и от всех предшествующих. Мы будет считать внутреннюю память конечной, то есть предполагать наличие у преобразователя конечного числа внутренних состояний m, таких, что символ на выходе является функцией символа на входе и текущего состояния. Следующее состояние является еще одной функцией этих двух величин. Таким образом, преобразователь может быть описан двумя функциями:

где

x_n - n-тый входной символ.

α_n - состояние преобразователя при получении n-того входного символа.

y_n - выходной символ (или последовательность символов), формируемый при получении символа x_n в состоянии α_n.

Если выходной символ одного из преобразователей может быть подан на вход другого, их можно соединить последовательно и получить еще один преобразователь. Если существует преобразователь, преобразующий выходную информацию первого в его же входную, первый преобразователь назовем *несингулярным*, а второй - обратным к нему.

Теорема 7: *Выходная последовательность преобразователя с конечным числом состояний является последовательностью от статистического источника с конечным числом состояний, с энтропией (в единицу времени), не большей энтропии входной последовательности. Если преобразователь является несингулярным, эти энтропии равны.*

пусть α описывает состояние источника, выдающего последовательность символов x_i, β - состояние преобразователя, дающего на выходе блоки символов y_j. Систему можно описать прямым произведением пространств их

состояний, то есть пространством пар (α, β). Соединим две точки этого пространства (α_1, β_1) и (α_2, β_2) линией, если α_1 может сгенерировать x, переводящий β_1 в β_2 так, что эта линия дает вероятность такого x. Пометим такую линию набором символов y_j, выданных преобразователем. Энтропия его может быть рассчитана как взвешенная сумма по всем состояниям. Если мы вначале просуммируем по β, результирующий член будет не большим соответствующего для α, так как энтропия не возрастает. Если преобразователь несингулярен, соединим его со входом обратного ему. Если H'_1, H'_2 и H'_3 - энтропии на выходе источника, первого и второго преобразователей соответственно, то $H'_1 \geq H'_2 \geq H'_3 = H'_1$ и, следовательно, $H'_1 = H'_2$.

Пусть мы имеем набор ограничений на возможные последовательности, которые можно предстваить в виде линейного графа, подобного изображенному на рис. 2. Если вероятности $p_{ij}^{(s)}$ приписаны всевозможным линиям, соединяющим состояние i с состоянием j, эта система становится источником. Одно из сопоставлений максимизирует энтропию на выходе (см. приложение 4).

Теорема 8: *Пусть система ограничений рассматривается как канал с пропускной способностью $C = \log W$. Если мы примем*

$$p_{ij}^{(s)} = \frac{B_j}{B_i} W^{-\ell_{ij}^{(s)}}$$

где $\ell_{ij}^{(s)}$ - длительность s-того символа, переводящего систему из состояния i в j , и B_i удовлетворяет условию

$$B_i = \sum_{s,j} B_j W^{-\ell_{ij}^{(s)}}$$

то H максимальна и равна C.

Надлеащим выбором вероятностей перехода энтропия символов в канале может быть доведена до пропускной способности канала.

Основная теорема для канала без шума

Теперь мы обоснуем интерпретацию H как темпа генерации информации, показав, что H определяет требуемую для наиболее эффективного кодирования пропускную способность канала.

Теорема 9: *Пусть энтропия источника есть H (бит на символ), а пропускная способность канала - C (бит в секунду). Тогда возможно так закодировать информацию от источника, что средний темп ее передачи по каналу будет , где ϵ может быть сделана в произвольной степени малой. Невозможно передать информацию в темпе, большем .*

Утверждение теоремы о невозможносит передачи с темпом, большим , может быть доказана с использованием того, что энтропия подаваемого в канал за секунду набора символов равна энтропии источника, так как преобразователь должен быть несингулярным, и, кроме того, эта энтропия не может быть больше пропускной способности канала. Следовательно, $H' \leq C$, и число символов в секунду $= H'/H \leq C/H$.

Первая же часть теоремы доказывается двумя различными методами. Первый заключается в рассмотрении набора всех последовательностей N символов, генерируемых источником. Для больших N мы можем разбить их на две группы: содержащую последовательности короче $2^{(H+\eta)N}$ символов и содержащую последовательности короче 2^{RN} (R здесь есть логарифм числа различных символов) с полной вероятностью меньше μ. С ростом N η и μ стремятся к нулю. Число сигналов длительности T в канале больше $2^{(C-\theta)T}$ (где θ мало) при достаточно больших T. Если мы примем

$$T = \left(\frac{H}{C} + \lambda\right)N$$

то число последовательностей символов канала для группы большой вероятности будет бостаточно большим при больших N и T (и малой λ), а кроме того, будет еще несколько последовательностей малой вероятности. Группа высокой вероятности кодируется произвольным однозначным сопоставлением с элементами этого множества. Остальные последовательности кодируются более длинными последовательностями, начинающимися и заканчивающимися элементами, не используемыми для кодирования группы высокой вероятности. Эти элементы (отдельные символы или их группы) сигнализируют начало и конец для другой кодировки. Между ними дается достаточное время для передачи сообщений малой вероятности. Это требует

$$T_1 = \left(\frac{R}{C} + \varphi\right)N$$

где φ мало. Средний темп передачи символов сообщения в секунду тогда будет больше, чем

$$\left[(1-\delta)\frac{T}{N} + \delta\frac{T_1}{N}\right]^{-1} = \left[(1-\delta)\left(\frac{H}{C} + \lambda\right) + \delta\left(\frac{R}{C} + \varphi\right)\right]^{-1}.$$

С ростом N δ, λ и φ стремятся к нулю, а темп стремится к .

Другим способом такого кодирования (и. следовательно, доказательства теоремы) можно описать следующим образом. Отсортируем сообщения длины N по убыванию вероятности, $p_1 \geq p_2 \geq p_3 \cdots \geq p_n$. Пусть - суммарная вероятность вплоть до p_s (не включая p_s). Вначале закодируем в двоичную систему. Двоичным кодом для сообщения s будет представление P_sдвоичным числом. Это представление размещается на m_s позициях, где m_s - целое число, удовлетворяющее условию

$$\log_2 \frac{1}{p_s} \leq m_s < 1 + \log_2 \frac{1}{p_s}.$$

Таким образом, сообщения с высокой вероятностью представляются более коротким кодом, чем маловероятные. Из этих неравенств получаем

$$\frac{1}{2^{m_s}} \leq p_s < \frac{1}{2^{m_s-1}}.$$

Код для P_s будет отличаться от всех последующих в одном или более из составляющих его m_s символов, так как все остальные P_i как минимум на $\frac{1}{2^{m_s}}$ больше, и их двоичные представления различаются в первых m_s позициях. Следовательно, все коды различны, и восстановление сообщения по его коду возможно. Если последовательности канала сами по себе не являются двоичными, они могут быть сведены к ним произвольным образом так, чтобы двоичный код однозначно преобразовывался бы в сигнал, приемлемый для канала.

Среднее число H' двоичных цифр на символ исходного сообщения может быть легко оценено. Имеем

$$H' = \frac{1}{N} \sum m_s p_s.$$

В то же время

$$\frac{1}{N} \sum \left(\log_2 \frac{1}{p_s} \right) p_s \leq \frac{1}{N} \sum m_s p_s < \frac{1}{N} \sum \left(1 + \log_2 \frac{1}{p_s} \right) p_s$$

и следовательно

$$G_N \leq H' < G_N + \frac{1}{N}$$

С ростом N G_N стремится к H, энтропии источника, и H' стремится к H.

Осюда видно, что неэффективность кодирования, то есть использование не всех из N символов, может быть доведена до $\frac{1}{N}$ с добавлением разницы между истинной энтропией H и ее оценкой G_N по последовательностям длины N. Процент времени, избыточного по отношению к идеальному случаю, соответственно, может быть сделан меньшим

$$\frac{G_N}{H} + \frac{1}{HN} - 1.$$

Этот метод кодирования практически совпадает с предложенным независимо R. M. Fano (Technical Report No. 65, The Research Laboratory of Electronics, M.I.T., March 17, 1949.), который заключается в сортировке сообщений длины N по убыванию их вероятности. В дальнейшем эти сообщения делятся на две группы

с как можно более близкими полными вероятностями. Для сообщений первой группы первой двоичной цифрой будет 0, для второй - 1. Группы таким же образом делятся далее, вплоть до того, что в каждой группе остается по одному сообщению. Легко видеть, что с небольшой разницей (в основном в последнем знаке) это аналогично процессу, описанному нами выше.

Обсуждение и примеры

В электротехнике для получения максимальной передачи мощности от генератора к нагрузке генератор должен обладать такими свойствами, чтобы с точки зрения нагрузки его сопротивление было равно сопротивлению ее. В нашем случае ситуация примерно аналогична. Преобразователь, который занимается кодированием сообщения, должен соответствовать источнику в статистическом смысле. Источник и преобразователь с точки зрения канала должны обладать статистическими свойствами источника, максимизирующего энтропию в канале. Содержанием теоремы 9 является то, что, хотя идеальное соответствие в общем случае невозможно, мы можем приближаться к нему с какой угодно степенью точности. Отношение действительного темпа передачи к пропускной способности канала C может быть названо *эффективностью системы кодирования*. Она, естественно, равна отношению действительной энтропии символов в канале к максимально возможному значению.

В общем случае идеальное или почти идеальное кодирование требует длительных задержек и передатчике и приемнике. В рассматриваемом случае отсутствия шума главной функцией этих задержек является достаточно тщательное сопоставление вероятностей последовательностей соответствующих длин. При хорошем кодировании логарифм совместной вероятности длинного сообщения должен быть пропорционален длительности соответствующего сигнала, точнее

$$\left|\frac{\log p^{-1}}{T} - C\right|$$

должно быть мало практически для всех длинных сообщений.

Если источник генерирует только одно фиксированное сообщение, его энтропия равна нулю, и канал не требуется. К примеру, вычислительная машина для расчета последовательных цифр числа π выдает детерминированную последовательность без случайной составляющей. Для передачи такого сигнала нет нужды в канале - достаточно построить вторую аналогичную машину. Однако, иногда это непрактично. В таком случае мы можем пренебречь некоторой информацией о структуре сигнала. Можно рассматривать цифры числа π как случайную последовательность и разработать систему для передачи произвольной цифровой информации. Аналогично мы могли бы воспользоваться лишь частью нашего знания статистической структуры английского языка. Для такого случая можно рассмотреть источник с максимальной энтропией, удовлетворяющий выбранным нами статистическим

условиям. Энтропия такого источника определяет пропускную способность канала, которая является необходимой и достаточной. Так, в случае передачи числа π единственным статистическим условием является то, что все символы выбираются из набора $0, 1, \ldots, 9$; для английского текста мы можем оставить лишь требование соответствия частот букв. Тогда источник с максимальной энтропией является первым приближением к английскому языку, и его энтропия определяет требуемую пропускную способность канала.

В качестве простого примера вышеизложенных результатов рассмотрим источник, генерирующий последовательность независимых букв, выбранных из **A**, **B**, **C**, **D** с вероятностями $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{8}$ соответственно. Тогда

Следовательно, для такого источника можно разработать систему представления сообщений двоичным кодом со степенью сжатия вплоть до $\frac{7}{4}$ двоичных цифр на символ. В данном примере это предельное значение может быть достигнуто при использовании следующего кода (полученного методом, использованным во втором варианте доказательства теоремы 9):

A	0
B	10
C	110
D	111

Среднее число двоичных знаков, использованных для кодирования последовательности длиной N символов будет

Легко видеть, что двоичные знаки 0 и 1 имеют вероятности $\frac{1}{2}$ каждый, следовательно, H для закодированной последовательности равно одному биту на символ. Так как в среднем мы имеем $\frac{7}{4}$ двоичных знаков на исходный символ, энтропия в единицу времени остается неизменной. Наибольшей возможной энтропией исходного набора будет $\log 4 = 2$, достигаемая при вероятностях букв A, B, C, D, равных $\frac14$ каждая. Таким образом, относительная энтропия равна $\frac{7}{8}$Двоичная последовательность может быть отображена на множество исходных символов при помощи следующей таблицы:

00	A'
01	B'
10	C'
11	D'

Такое двойное преобразование кодирует исходное сообщение теми же символами, но с коэффициентом сжатия $\frac{7}{8}$.

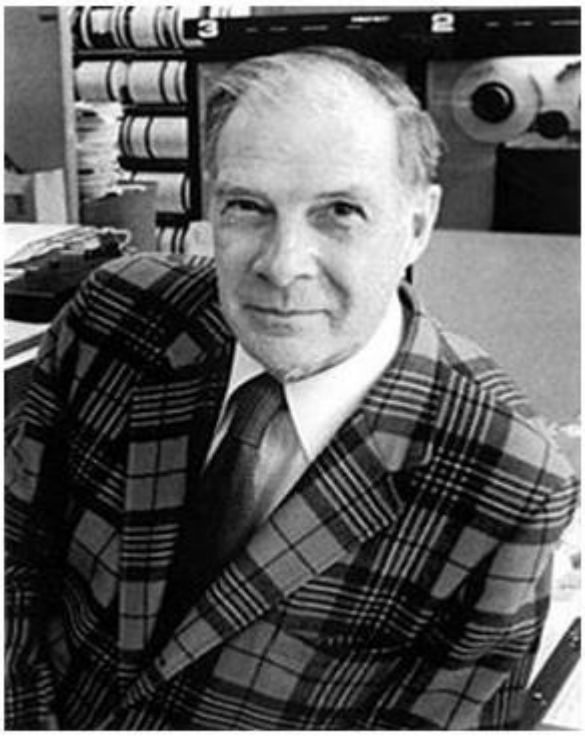
В качестве еще одного примера рассмотрим источник, выдающий буквы **A** и **B** с вероятностями p для **A** и q для **B**. При $p \ll q$

В такой ситуации можно построить достаточно хорошую систему кодирования сообщений в двоичном канале, посылая некоторую специльную последовательность, скажем, **0000**, для маловероятного символа **A**, и затем число, характеризующее количество следующих за ним символов **B**. Это количество может быть охарактеризовано двоичным представлением при удалении всех чисел, содержащих специальную последовательность. В нашем случае все числа вплоть до 16 представляются как обычно, 16 представляется следующим числом, не содержащим специальной последовательности, а именно $17 = 10001$, и так далее.

Можно показать, что при $p \to$ система кодирования стремится к идеальной при надлежащем выборе специальной последовательности.

Код Хэмминга

Наиболее распространенным систематическим линейным блочным кодом является код Хэмминга.



Richard Wesley Hamming
1915 – 1998

Рис.8 Хэмминг.

К нему относятся *коды с минимальным кодовым расстоянием* $\boldsymbol{d_{min}=3}$*, способные исправлять однократные ошибки.*

При передаче кодового слова по каналу связи возможно возникновение однократной ошибки в любом из его элементов. Количество таких

ситуаций $C_n^1 = n$. Таким образом, для того чтобы определить место возникновения ошибки, количество комбинаций проверочных элементов 2^r должно быть не меньше количества возможных ошибочных ситуаций в коде плюс ситуация, когда ошибка не возникает, т. е. должно выполняться неравенство

$$2^r \geq n+1.$$

Из этого неравенства следует *минимальное соотношение проверочных и информационных разрядов, необходимое для исправления однократных ошибок*

$$2^r - 1 = n.$$

Для расчёта основных параметров кода Хэмминга *можно задать количество проверочных элементов* **r**, *тогда длина* кодовых слов $n \leq 2^r - 1$, *а количество информационных элементов* $k = n - r$. Соотношения между **r**, **n** и **k** приведены в следующей таблице (*табл. 3.3*.)

Таблица 3.3

k	1	1	2	3	4	4	5	6	7	8	9	10	11	11	12	12
r	2	3	3	3	3	4	4	4	4	4	4	4	4	5	5	6
n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Характерной особенностью проверочной матрицы кода с $d_{min}=3$ является то, что ее столбцы – различные ненулевые комбинации длины **r**.

Хэммингом *предложено располагать столбцы проверочной матрицы* $H_{(n-k)\times n}$ *так, чтобы* **i***-й столбец матрицы и номер разряда кодовой комбинации отвечали двоичному представлению числа* **i**. *Тогда* синдром исправления однократных ошибок ***будет двоичным представлением номера разряда, в котором произошла ошибка***. *Для этого проверочные разряды должны находиться не в правой части кодового слова, а в позициях, номера которых являются степенью двойки, т. е.* $2^0, 2^1, 2^2, \ldots, 2^{r-1}$.

Например, для $r=3$ проверочная матрица кода Хэмминга имеет вид

$$H_{3\times 7} = \begin{pmatrix} 0001111 \\ 0110011 \\ 1010101 \end{pmatrix}.$$

Для данного линейного блочного (4, 7)-кода первый, второй, четвертый разряды (u_1, u_2, u_4) будут***проверочными***, а третий, пятый, шестой и седьмой разряды (u_3, u_5, u_6, u_7) – ***символами сообщения*** в том же порядке, что и кодируемое сообщение, т.е (m_1, m_2, m_3, m_4) соответственно.

Таким образом, *для* (k, n)-кода Хэмминга *в каждом* кодовом слове $u=(u_1, u_2, u_3, u_4, \ldots, u_8, \ldots, u_n)$, $r=n-k$*битов с номерами степени* **2** *являются* ***проверочными****, а остальные –* ***битами сообщения****, т.е. кодирование осуществляется так*:

$$E(m_1, m_2, \ldots, m_k)= (u_1, u_2, \ldots, u_n)=(r_1, r_2, m_1, r_3, m_2, m_3, m_4, r_4, m_5, m_6, \ldots, m_k).$$

Проверочная матрица (k, n)-кода Хэмминга *составляется из* $n=2^r-1$ *строк и* r *столбцов и представляет собой двоичные комбинации числа* i, *где* i – *номер столбца проверочной матрицы (разряда кодовой комбинации)*.

Например, для $r=2, 3, 4$ проверочные матрицы кода Хэмминга имеют вид

$$H_{2\times 3}=\begin{pmatrix}0&1&1\\1&0&1\end{pmatrix},\quad H_{3\times 7}=\begin{pmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{pmatrix},\quad H_{4\times 15}=\begin{pmatrix}0&0&0&0&0&0&0&1&1&1&1&1&1&1&1\\0&0&0&1&1&1&1&0&0&0&0&1&1&1&1\\0&1&1&0&0&1&1&0&0&1&1&0&0&1&1\\1&0&1&0&1&0&1&0&1&0&1&0&1&0&1\end{pmatrix}.$$

Синдром, *определяющий* систему проверочных уравнений *кода, находится из уравнения* $u' H^T=0$.
Например, для $r=3$ система проверочных уравнений будет следующей:

$$\begin{cases}\underline{u_4}+u_5+u_6+u_7=0,\\ \underline{u_2}+u_3+u_6+u_7=0,\\ \underline{u_1}+u_3+u_5+u_7=0.\end{cases}$$

Отсюда проверочные разряды (контрольные суммы) находятся как

$$\begin{cases}r_1=u_1=u_3+u_5+u_7=m_1+m_2+m_4,\\ r_2=u_2=u_3+u_6+u_7=m_1+m_3+m_4,\\ r_3=u_4=u_5+u_6+u_7=m_2+m_3+m_4.\end{cases}$$

Чтобы ***закодировать*** *сообщение* m, *в качестве* u_i, *где* i *не равно степени* **2**, *берутся соответствующие биты сообщения, а* ***проверочные разряды*** *с индексами степени* **2** *находятся из* ***системы проверочных уравнений кода***. *В каждое уравнение системы входит только одна****контрольная сумма***.

Пример 1 Закодируем сообщение $m=(0\ 1\ 1\ 1)$ (4, 7)-кодом *Хэмминга*.

Кодовым словом данного кода будет последовательность (***u*₁ *u*₂ 0 *u*₄ 1 1 1**), где ***u*₁, *u*₂, *u*₄** – контрольные суммы ***r*₁, *r*₂, *r*₃**.
Из системы проверочных уравнений находим контрольные суммы:

$$r_1 = u_1 = u_3+u_5+u_7 = m_1+m_2+m_4 = 0+1+1 = 0,$$
$$r_2 = u_2 = u_3+u_6+u_7 = m_1+m_3+m_4 = 0+1+1 = 0,$$
$$r_3 = u_4 = u_5+u_6+u_7 = m_2+m_3+m_4 = 1+1+1 = 1.$$

Таким образом, кодовым словом будет последовательность (0001111).

Декодирование кода Хэмминга происходит по следующей схеме. *Определяется* синдром *принятой последовательности* $\boldsymbol{S}=\boldsymbol{y}'\boldsymbol{H}^{\mathrm{T}}$, *где* $\boldsymbol{H}^{\mathrm{T}}$ *- транспонированная проверочная матрица кода;* $\boldsymbol{y}$ *– принятый вектор. Если синдром равен нулевому вектору, то считается, что слово передано без ошибок, иначе значение синдрома соответствует двоичному представлению номера разряда, в котором произошла ошибка. В этом случае нужно изменить значение в ошибочном разряде, считая разряды слева направо, начиная с 1.*

Пример 2 Сообщение кодируется (4, 7)-кодом *Хэмминга*. Принята последовательность ***y***=(0011111). Декодируем принятый вектор.

Определяем синдром принятого вектора:

$$\boldsymbol{y}'\boldsymbol{H}^{\mathrm{T}} = (0011111)' \begin{pmatrix} 0001111 \\ 0110011 \\ 1010101 \end{pmatrix}^{\mathrm{T}} = (0\ 1\ 1) = 3_{10},$$

т. е. ошибка произошла в третьем разряде.

Исправляем ошибку, изменяя значение в третьем бите

(00<u>1</u>1111) ® (0001111).

Переданное сообщение декодируется как

$$D(u_1, u_2, u_3, u_4, u_5, u_6, u_7) = D(0001111) = (0111).$$

Порождающей матрицей (k, n)-кода Хэмминга *является матрица* (***k*×*n***), *в которой столбцы с номерами не степенями* ***2*** *образуют единичную подматрицу, а остальные столбцы соответствуют проверочным уравнениям кода. Такая матрица при кодировании будет копировать биты сообщения в позиции, не степени 2, и заполнять другие позиции кода согласно системе вычисления контрольных разрядов.*

Пример 3 Система проверочных уравнений (4, 7)-кода *Хэмминга* следующая:

$$r_1 = u_1 = u_3+u_5+u_7 = m_1+m_2+m_4,$$
$$r_2 = u_2 = u_3+u_6+u_7 = m_1+m_3+m_4,$$
$$r_3 = u_4 = u_5+u_6+u_7 = m_2+m_3+m_4.$$

Соответственно порождающая матрица данного кода имеет вид

$$G_{4\times 7} = \begin{pmatrix} 1\,1\,1\,0\,0\,0\,0 \\ 1\,0\,0\,1\,1\,0\,0 \\ 0\,1\,0\,1\,0\,1\,0 \\ 1\,1\,0\,1\,0\,0\,1 \end{pmatrix}.$$

Код Вермана

В криптографии **шифр Вернама** известен также как **«схема одноразовых блокнотов»**. Решение является системой симметричного шифрования, которая была изобретена в 1917 году сотрудниками AT&T Мейджором Джозефом Моборном и Гильбертом Вернамом.

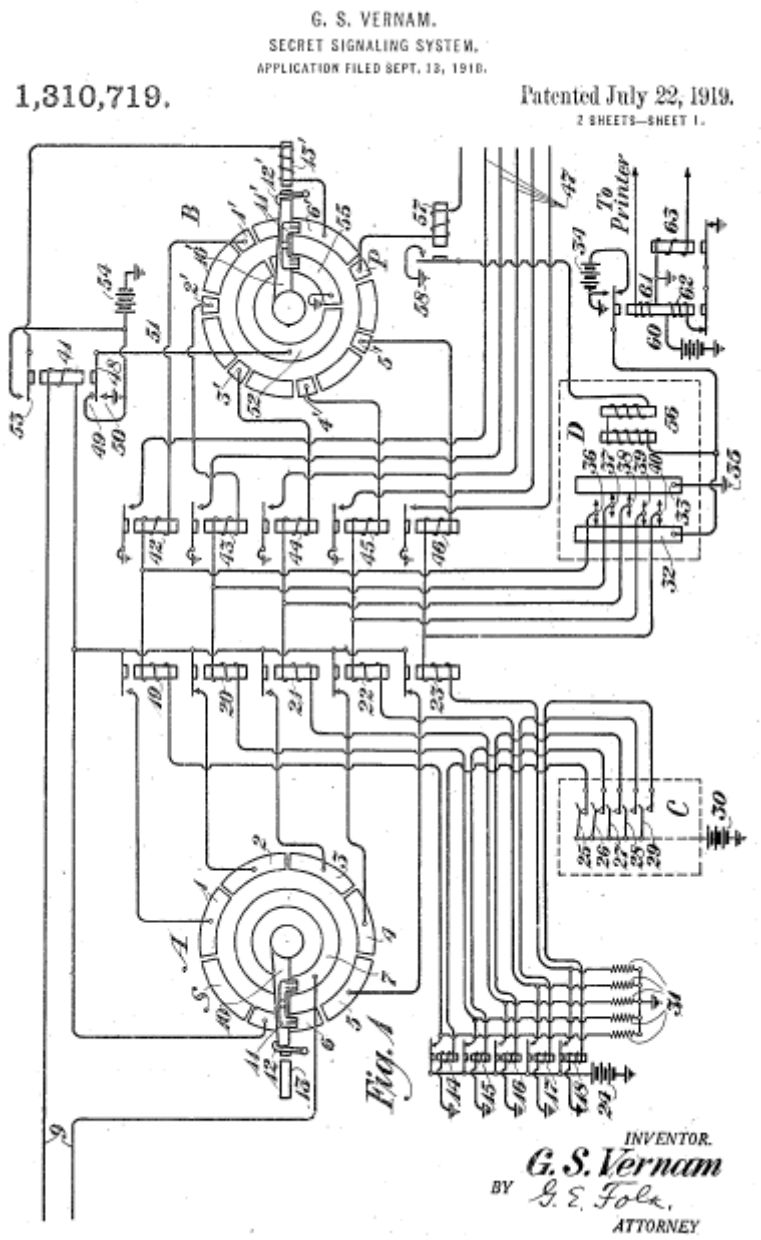


Рис. 9. Патент на систему Вермана.

В 1949 годах была опубликована работа Клода Шеннона, где Шеннон доказал абсолютную стойкость шифра Вернама. В этой работе Шеннон показал, что не существует других шифров с подобными свойствами и его выводом стало следующее утверждение: **шифр Вернама – самая безопасная криптосистема из всех имеющихся.**
Однако, следует заметить, что для того, чтобы шифр действительно был стойким, необходимо выполнение следующих трех правил:

1. Ключ для шифрования выбирается случайным образом.

2. Длина ключа должна быть равна длине открытого текста.

3. Ключ должен использоваться **ТОЛЬКО** один раз.

А теперь поподробней о самом шифре и процессе шифрования. Так как этот шифр был придуман для компьютерных систем, то следует заметить,

что базируется он на двоичной арифметике. Надеюсь что вы знакомы с ней ;). Основным объектом рассмотрения в данном методе шифрования является логическая операция **XOR** (взаимоисключающее ИЛИ). Таким образом, так как у нас двоичная арифметика, то все операции будут осуществляется над нулем (0) и единицей (1). Логическая операция **XOR**, в отличие от операции **OR**, при логическом сравнении 0 и 1 дает 1, при сравнении 1 с 1 дает 0, а при 0 с 0 дает 0. Следовательно, если мы выполним операцию **XOR** над числами 10110 и 11010, то получим: 10110 **xor**11010 = 01100. Надеюсь что принцип работы операции**XOR** понятен.

Далее, так как шифр работает с двоичной системой исчисления, необходимо понимать, что буквы — это всего лишь некоторая интерпретация числа, то есть число является кодом символа некоторой таблицы кодировок. К примеру, наиболее популярные таблицы кодировок это: ANSI, ASCII и UTF(unicode). Естественно, что в каждой таблице один и тот же символ может иметь разный код, поэтому, во избежании путаницы, имейте это ввиду и используйте одну и ту же кодировку при шифровании и дешифровании. Замечу, что использовать можно и свою (придуманную) кодировку. Кроме того, данный шифр может использоваться не только на компьютерах. Его можно применить и к тексту написанному на бумаге. Только перед применением надо сделать некие преобразования. Таким образом, перед тем, как осуществить шифрование, необходимо перевести все символы в их однозначную числовую интерпретацию. Если Вы решили применить шифр в компьютерных системах, то для вас уже существуют соответствующие кодировки и язык программирования, выбранный Вами, скорее всего поддерживает явное или неявное преобразование. И вам остается только произвести над каждой парой операцию **XOR**. В различных языках это операция определяется по разному, приведу пример: для pascal/Delphi/Assembler — **xor**, C/C++ —^. Однако, если же Вы решили применить шифр Вернама к письменному тексту, то, к примеру, дайте каждой букве используемого вами алфавита, соответствующий ей порядковый номер в двоичной системе исчисления. Например, если вы используете русский алфавит (без учета буквы Ё) то это будет выглядеть так: а -> 00000, б -> 00001, в -> 00010, г -> 00011, … я -> 111111. Тем самым мы определили свою таблицу кодировки. После этого, написав сообщение и придумав ключ, преобразуйте каждый символ в их числовое значение, соответствующее вашей таблице кодировки, и после этого осуществляйте операцию **XOR** над каждой соответствующей парой. Так как данный метод шифрования является симметричным, следовательно, применив операцию **XOR**к каждой паре символов шифр-текста (шифрограммы) и ключа, мы получим открытый текст.сибо, объявление скрыто.

Потоковые шифры на базе сдвиговых регистров активно использовались в годы войны, ещё задолго до появления электроники. Они были просты в проектировании и реализации.

В 1965 году Эрнст Селмер, главный криптограф норвежского правительства, разработал теорию последовательности сдвиговыхрегистров. Позже Соломон Голомб, математик Агентства Национальной Безопасности США, написал книгу под названием «Shift Register Sequences»(« Последовательности сдвиговых регистров»), в которой изложил свои основные достижения в этой области, а также достижения Селмера.

Большую популярность потоковым шифрам принесла работа Клода Шеннона, опубликованная в 1949 году, в которой Шеннон доказал абсолютную стойкость шифра Вернама (также известного, как одноразовый блокнот). В шифре Вернама ключ имеет длину, равную длине самого передаваемого сообщения. Ключ используется в качестве гаммы, и если каждый бит ключа выбирается случайно, то вскрыть шифр невозможно (т.к. все возможные открытые тексты будут равновероятны). До настоящего времени было придумано немало алгоритмов потокового шифрования. Такие как: A3, A5, A8, RC4,PIKE, SEAL, eSTREAM.

Режим гаммирования для поточных шифров

Простейшая реализация поточного шифра изображена на рисунке. Генератор гаммы выдаёт ключевой поток (гамму): $k_1, k_2, k_3, \ldots, k_L$. Обозначим поток битов открытого текста $m_1, m_2, m_3, \ldots, m_L$. Тогда поток битов шифротекста получается с помощью применения операции XOR: $c_1, c_2, c_3, \ldots, c_L$, где $c_i = m_i \oplus k_i$.

Расшифрование производится операцией XOR между той же самой гаммой и зашифрованным текстом: $m_i = c_i \oplus k_i$.

Очевидно, что если последовательность битов гаммы не имеет периода и выбирается случайно, то взломать шифр невозможно. Но у данного режима шифрования есть и отрицательные особенности. Так ключи,

сравнимые по длине с передаваемыми сообщениями, трудно использовать на практике. Поэтому обычно применяют ключ меньшей длины (например, 128 бит). С помощью него генерируется псевдослучайная гаммирующая последовательность (она должна удовлетворять постулатам Голомба). Естественно, псевдослучайность гаммы может быть использована при атаке на поточный шифр

Протокол Диффи-Хелмана.

Рис. 10. Изобретатели алгоритма Диффи-Хелмана.

Описание алгоритма

1). Пользователь A, который хочет получать шифрованные сообщения, должен выбрать какую-нибудь функцию FK с секретом K. Он сообщает всем заинтересованным (например, публикует) описание функции FK в качестве своего алгоритма шифрования. Но при этом значение секрета K он никому не сообщает и держит в секрете. Если теперь пользователь B хочет послать пользователю A защищаемую информацию x X, то он вычисляет y = FK(x) и посылает y по открытому каналу пользователю A. Поскольку A для своего секрета K умеет инвертировать FK, то он вычисляет x по полученному y. Никто другой не знает K и поэтому в силу свойства в) функции с секретом не сможет за полиномиальное время по известному шифрованному сообщению FK(x) вычислить от- крытый текст x. Описанную систему называют криптосистемой с открытым ключом, поскольку алгоритм шифрования FK является общедоступным или открытым. В последнее время такие криптосистемы еще называют асимметричными, поскольку в них есть асимметрия в алгоритмах. Для асимметричных систем алгоритм шифрования общеизвестен, но восстановить по нему алгоритм дешифрования за полиномиальное время

невозможно. Описанную выше идею Диффи и Хеллман предложили использовать также для электронной подписи сообщений, которую невозможно подделать за полиномиальное время. Пусть пользователю A необходимо подписать сообщение x. Он, зная секрет K, находит такое y, что FK(y) = x, и вместе с сообщением x посылает y пользователю B в качестве своей электронной подписи. Пользователь B хранит y в качестве доказательства того, что A подписал сообщение x. Сообщение, подписанное электронной подписью, можно представлять себе как пару (x, y), где x — сообщение, y — решение уравнения FK(y) = x, FK : X → Y — функция с секретом, известная всем взаимодействующим абонентам. Из определения функции FK очевидны следующие полезные свойства электронной подписи: 1) подписать сообщение x, т. е. решить уравнение FK(y) = x, может только абонент — обладатель данного секрета K; другими словами, под- делать подпись невозможно; 2) проверить подлинность подписи может любой абонент, знающий открытый ключ, т. е. саму функцию FK; 3) при возникновении споров отказаться от подписи невозможно в силу ее неподделываемости; 4) подписанные сообщения (x, y) можно, не опасаясь ущерба, пере- сылать по любым каналам связи. Кроме принципа построения криптосистемы с открытым ключом, Диффи и Хеллман в той же работе предложили еще одну новую идею — открытое распределение ключей. Они задались вопросом: можно ли организовать такую процедуру взаимодействия абонентов A и B по от- крытым каналам связи, чтобы решить следующие задачи: 1) вначале у A и B нет никакой общей секретной информации, но в конце процедуры такая общая секретная информация (общий ключ) у A и B появляется, т. е. вырабатывается; 2) пассивный противник, который перехватывает все передачи ин- формации и знает, что хотят получить A и B, тем не менее не может восстановить выработанный общий ключ A и B. Диффи и Хеллман предложили решать эти задачи с помощью функ- ции $F(x) = \alpha^x \bmod p$, § 4. Новые направления 23 где p — большое простое число, x — произвольное натуральное число, α — некоторый примитивный элемент поля GF(p). Общепризнанно, что инвертирование функции $\alpha^x \bmod p$, т. е. дискретное логарифмиро- вание, является вычислительно трудной математической задачей. (По- дробнее см. главу 4.) Сама процедура или, как принято говорить, протокол выработки общего ключа описывается следующим образом. Абоненты A и B независимо друг от друга случайно выбирают по одному натуральному числу — скажем x_A и x_B. Эти элементы они дер- жат в секрете. Далее каждый из них вычисляет новый элемент: $y_A = \alpha^{x_A} \bmod p$, $y_B = \alpha^{x_B} \bmod p$. (Числа p и α считаются общедоступными.) Потом они обмениваются этими элементами по каналу связи. Теперь абонент A, получив y_B и зная свой секретный элемент x_A, вычисляет новый элемент: $y^{x_A}_B \bmod p = (\alpha^{x_B})^{x_A} \bmod p$. Аналогично поступает абонент B: $y^{x_B}_A \bmod p = (\alpha^{x_A})^{x_B} \bmod p$. Тем самым у A и B появился общий элемент поля, равный $\alpha^{x_A x_B}$. Этот элемент и объявляется общим ключом A и B. Из описания протокола

видно, что противник знает p, α, αxA , αxB , не знает xA и xB и хочет узнать α xAxB .

RSA

Система RSA была предложена в 1978 г. и в настоящее время является наиболее распространенной системой шифрования с открытым ключом.

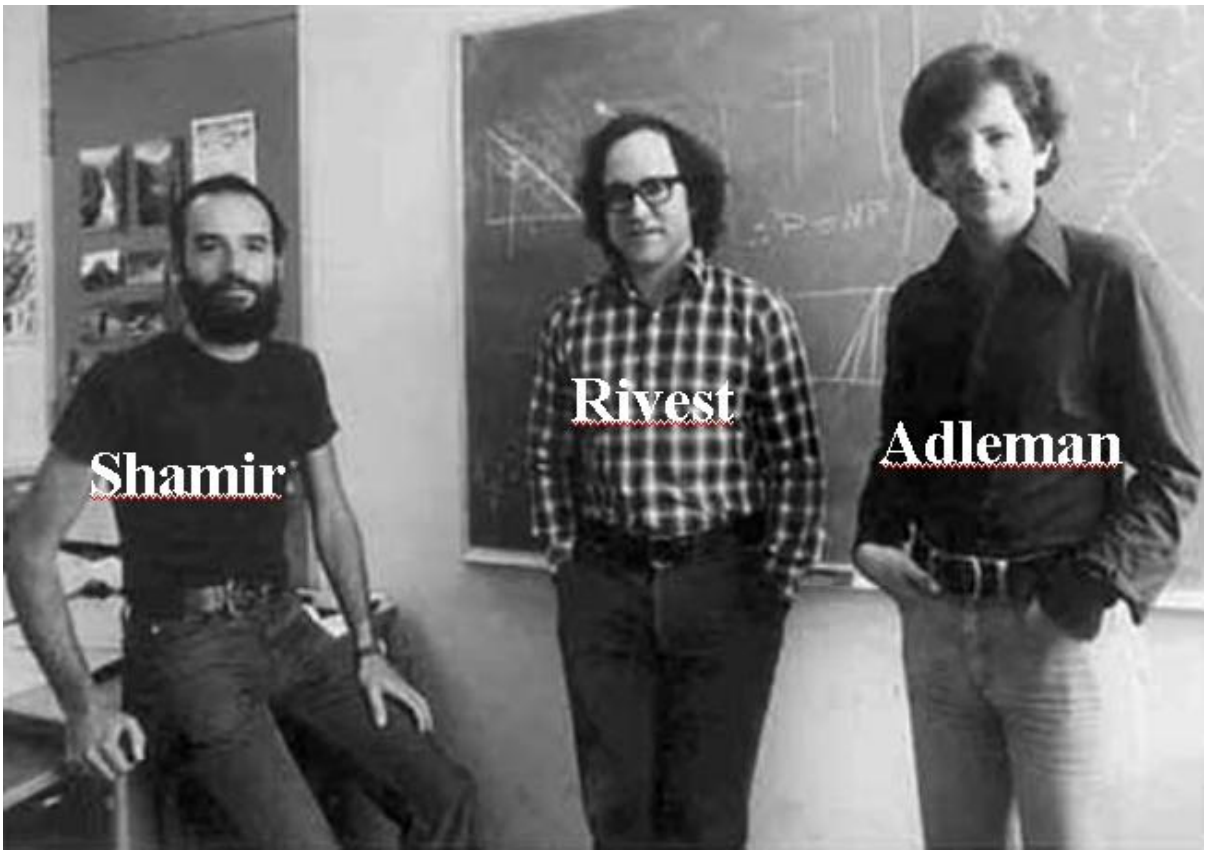


Рис. 11. Изобретатели алгоритма RSA

Пусть n = p • q — целое число, представимое в виде произведения двух больших простых чисел p,q.Выберем числа e и d из условия

e•d°1(mod j(n)),

где j(n) = (p – 1) • (q–1) – значение функции Эйлера от числа n. Пусть k=(n,p,q,e,d)– выбранный ключ, состоящий из открытого ключа $k_з$ =(n,e) и секретного ключа k_p = (n,p,q,d). Пусть M– блок открытого текста и C – соответствующий блок шифрованного текста. Тогда правила зашифрования и расшифрования определяются формулами:

$C = E_k(M) = M^e (\mathrm{mod}\ n)$, $D_k(C) = C^d (\mathrm{mod}\ n)$.

Заметим, что в соответствии с (2) $D_k(C) = M$. Это вытекает из следующих рассуждений. Для любого целого числа M и любого простого p справедливо сравнение

$M^{p} \circ M(\mathrm{mod}\ p)$.

В самом деле

$M^{p} - M \circ 0(\mathrm{mod}\ p)$

или сравнению

$M(M^{p-1} - 1) \circ 0(\mathrm{mod}\ p)$

Если НОД(M,p)=p, то p делит M, и поэтому $M \circ 0(\mathrm{mod}\ p)$.

Если же НОД(M,p) = 1, то, согласно малой теореме Ферма, $M^{p-1} \circ l(\mathrm{mod} p)$, откуда также следует.

Согласно, существует целое число r, такое, что $e \times d = r \times j(п) + 1$. Отсюда получаем следующую цепочку равенств и сравнений:

$C^{d} = (M^{e})^{d} = M^{e \times d} = M^{r \times j(п) + 1} = M^{r \bullet (p-l)(q-l)+l} = M^{r \bullet p(q-l)} \times M^{-r \bullet q+r+l} =$

$= M^{r \bullet (q-l)} \times M^{-r \bullet q+r+l)} = M^{r \bullet (q-l)} \times M^{-r \bullet (q+r+l} \circ M(\mathrm{mod}\ p)$.

Аналогично можно показать, что

$C^{d} = M(\mathrm{mod}\ p)$.

Поскольку p и q — разные простые числа, то на основании известных свойств сравнений, получаем:

$C^{d} = M(\mathrm{mod}\ n)$.

Отсюда и следует корректность определения. Для того чтобы лучше представить себе технические детали, возникающие при зашифровании и расшифровании, приведем пример работы с RSA.

Конрольная работа.
Контрольная работа состоит из 3-х заданий, которые необходимо оформить в виде сайта, размещенного в Интернет на одном из бесплатных хостингом (somee.com, smartasp.net? gear.host и др.). Вариант выдается преподавателем.

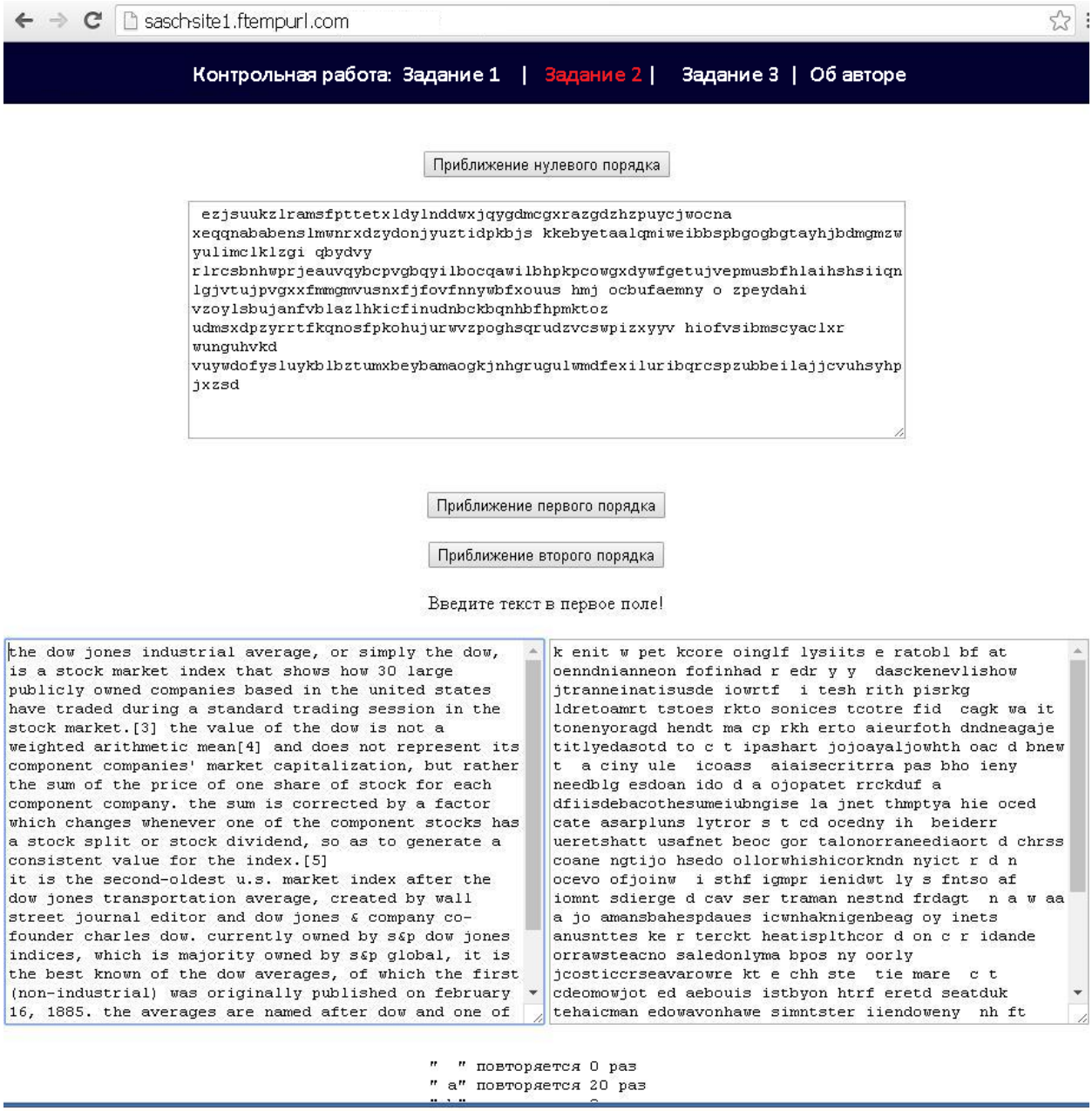


Рис. 3.1.Внешний вид контрольного задания.

Задание 1. Энтропия.
Решить задачу.

Вариант 1.
Определить количество информации, содержащееся в телевизионном сообщении, которое длится 1 с. Число элементов разложения в одной строке равно 600. Число строк равно 600. Число градаций яркости равно 128. Число кадров в секунду равно 25.

Вариант 2.
Найти среднее количество информации по Шеннону в системе со следующим вероятностным распределением $p\ (1/2;1/4;1/4)$.

Вариант 3.
Какое максимальное количество информации по Шеннону содержит система со следующим вероятностным распределением $p\ (0{,}2;0{,}8)$.

Вариант 4.
Какова энтропия системы, представляющей собой телефонную станцию, обслуживающую 300 абонентов, если вероятность позвонить любому абоненту в течение часа работы равна 0.01?

Вариант 5.
Вычислить энтропию источника и его избыточность, если алфавит состоит из независимых букв с вероятностями 0,4; 0,25; 0,2; 0,1; 0,05.

Вариант 6.
Построить код Хаффмана и вычислить его эффективность для источника с вероятностями букв 7/16; 5/16; 3/16; 1/16.

Вариант 7.
Задано десятичное число 13. Закодировать соответствующее двоичное число кодом Хэмминга (7, 4).

Задание 2.

Написать программу аппроксимации по Шеннону

Вариант 1. Символьная аппроксимация (приближение) 0-го, 1-го и 3-го порядка.

1. Zero-order approximation (symbols independent and equiprobable).

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZL-HJQD.

Рис. 2.1. Аппроксимация 0-го порядка

Вариант 2. Символьная аппроксимация (приближение) 0-го, 1-го и 2-го порядка.

2. First-order approximation (symbols independent but with frequencies of English text).

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

Рис. 2.2. Аппроксимация 1-го порядка

Вариант 3. Символьная аппроксимация (приближение) 0-го, 2-го и 3-го порядка.

3. Second-order approximation (digram structure as in English).

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

Рис. 2.3. Аппроксимация 2-го порядка

Вариант 4. Аппроксимация (приближение) слов 1-го порядка

5. First-order word approximation. Rather than continue with tetragram, . . . , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

Рис. 2.4. Аппроксимация слов 1-го порядка

Вариант 5. Аппроксимация (приближение) слов 2-го порядка.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

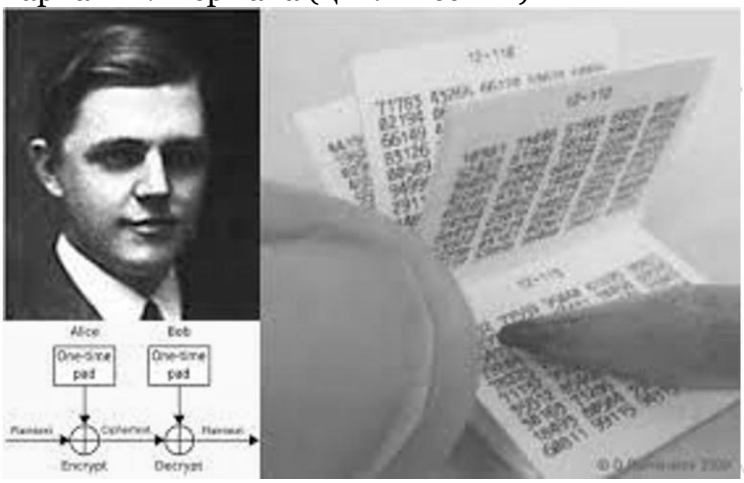
   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Рис. 2.5. Аппроксимация слов 2-го порядка

Задание 3

Реализовать алгоритмы шифрования

Вариант 1. Вермана (циклический)



Conversion Table

A	= 1	K	= 11	U	= 21
B	= 2	L	= 12	V	= 22
C	= 3	M	= 13	W	= 23
D	= 4	N	= 14	X	= 24
E	= 5	O	= 15	Y	= 25
F	= 6	P	= 16	Z	= 26
G	= 7	Q	= 17		
H	= 8	R	= 18		
I	= 9	S	= 19		
J	= 10	T	= 20		

Рис.3.1. Код Вермана.

Вариант 2. Диффи-Хеллмана.

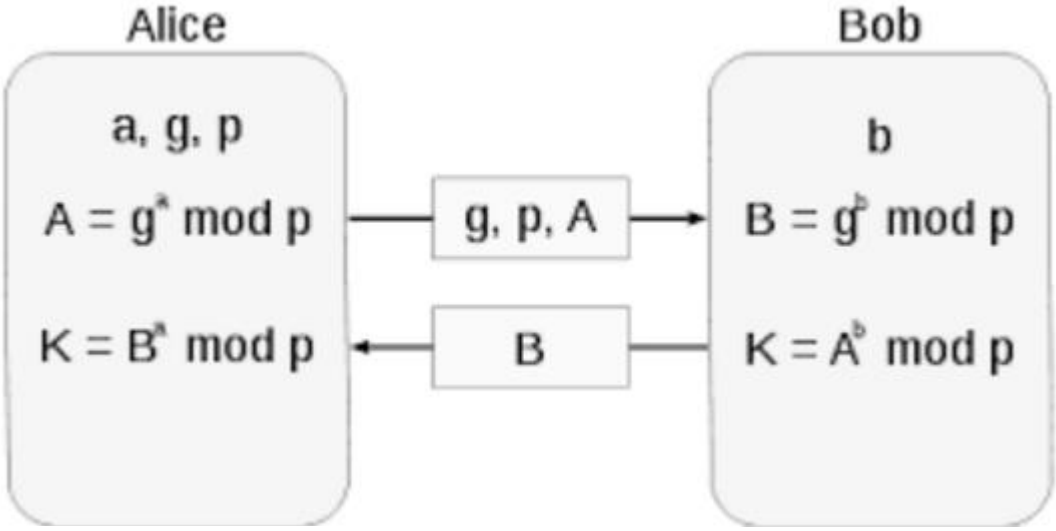


Рис.3.2. Протокол Диффи-Хеллмана.

Вариант 3. RSA.

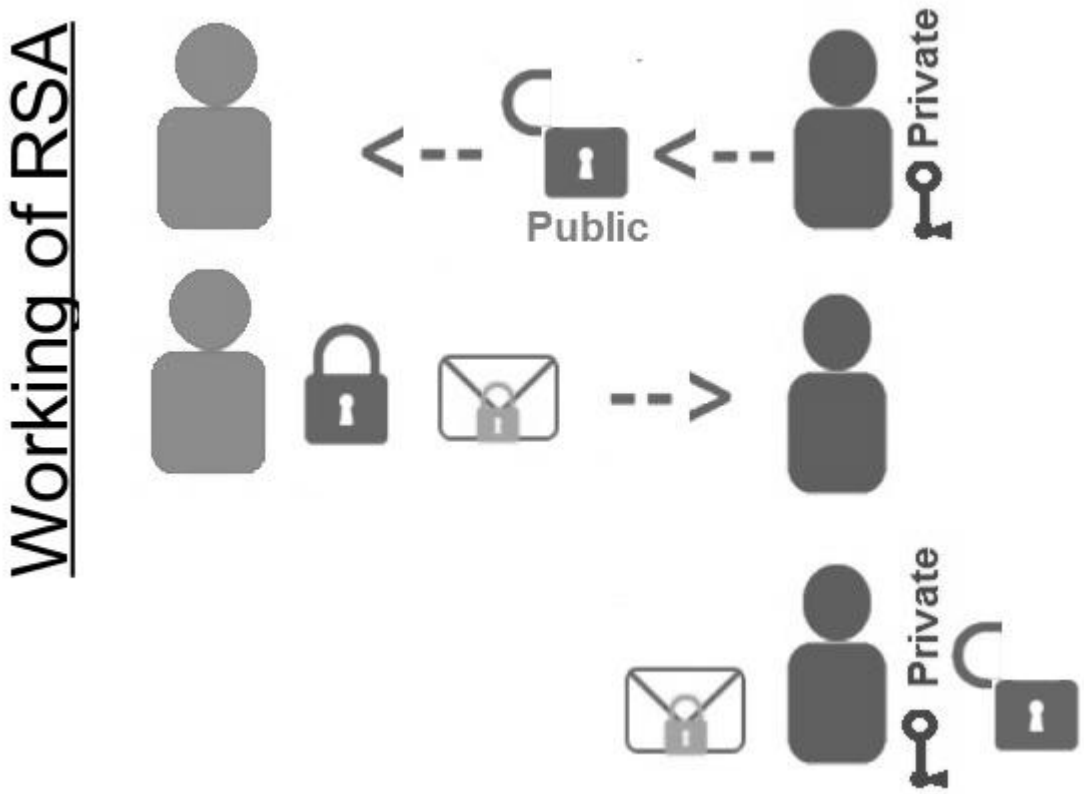


Рис.3.2. Протокол RSA.