

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

**А. И. МИТЮХИН**

УЧЕБНОЕ ПОСОБИЕ  
**«ТЕОРИЯ ИНФОРМАЦИИ»**

МИНСК 2016

# ЧАСТЬ 1. ИНФОРМАЦИЯ. ЭФФЕКТИВНОЕ КОДИРОВАНИЕ

## ВВЕДЕНИЕ

Теория информации это раздел науки, возникший в середине прошлого века. Умение применять на практике результаты теории информации стало важным для специалиста, создающего современные инфокоммуникационные системы. Теория информации возникла из статистической теории связи. Лишь частично отвечая на вопросы о путях и способах технической реализации аппаратуры, эта теория позволяет вычислить эффективность системы передачи, хранения, обработки и распределения информации, определить максимально возможную эффективность системы. Для широкого класса информационных задач при определенных знаниях или допущениях относительно статистики шумов стало возможным построение аппаратуры, работающей на основе оптимального приема кодированных сигналов. Такие сигналы строятся на основе составляющих теории информации – теории эффективного кодирования, теории помехоустойчивого кодирования и теории криптографического кодирования.

После того, как выбрана и обоснована конкретная схема оптимального приемника, фильтра, обнаружителя и т. п. возникают вопросы выбора метода кодирования, класса кода, способа защиты информации от несанкционированного доступа к ней.

Большая часть передаваемых, хранимых, распределяемых, преобразуемых данных соответствует звуковой, графической или видеоинформации. Увеличиваются технические затраты на хранение данных, предъявляются более высокие требования по экономии канального частотного ресурса. Алгоритмы теории информации позволяют уменьшить объем данных, используемых для представления информации.

Задача теории информации – при известной статистике шумов выбрать такое множество передаваемых сигналов, чтобы правдоподобие правильного декодирования принимаемых сообщений было максимальным. При этом важно найти не только хороший код, но и эффективный алгоритм декодирования.

Теория помехоустойчивых кодов (кодов, контролирующих ошибки) является одной из ветвей теории цифровой обработки сигналов (ЦОС). Существует тесная связь теории информации – кодирования и теории ЦОС. Но данные дисциплины развивались различными путями: одна разрабатывалась в основном алгебраистами, а другая – в основном инженерами. Первые результаты по теории информации появились в конце 40-х годов в работах К. Шеннона (Shannon Claude, амер. ученый), Голея (M. J. E. Golay, амер. ученый) и Р. Хэмминга (R. Hamming, амер. ученый). Можно определить следующие основные исторические этапы развития теории информации:

– 1948 г., К. Шеннон сформулировал и доказал теоремы кодирования для дискретного канала. К. Шеннон показал, что с каждым каналом передачи информации связано число  $C$ . Это число определяет пропускную способ-

ность канала и измеряется в битах в секунду. Если требуемая от информационной системы скорость передачи информации  $R_i$  (измеряемая в битах в секунду) меньше  $C$ , то используя коды, контролирующие ошибки, для данного канала можно построить такую информационную систему, что вероятность ошибки на выходе декодера будет сколь угодно мала;

- 1950 г., Р. Хэмминг описал класс кодов, исправляющих независимые одиночные ошибки;

- 1952 г., Д. А. Хаффмен (D. A. Huffman, амер. ученый) показал, что, разработанный им алгоритм эффективного кодирования позволяет строить класс оптимальных префиксных кодов;

- 1960 г., Р. К. Боуз (R. C. Bose, инд.-амер. ученый), Д. К. Рой-Чоудхури (D. K. R-Chaudhari, инд.-амер. ученый) и независимо Р. К. Хоквингем, 1959 (R. C. Hocquenghem, фран. ученый) открыли двоичные коды, исправляющие кратные независимые ошибки (коды Боуза- Чоудхури- Хоквингема (БЧХ-коды));

- 1963 г., И. С. Рид (I. S. Reed, амер. ученый) и Г. Соломон (G. Solomon, амер. ученый) предложили модификацию БЧХ-кодов для недвоичных каналов (коды Рида-Соломона (РС-коды)). Эти коды нашли применение для исправления пакетов и модулей ошибок;

- (1960 – 1970) г., с появлением микросхем средней степени интеграции началось практическое воплощение методов теории информации в каналах с большим уровнем помех. Применялись низкоскоростные коды максимальной длины (М-последовательности), коды Рида-Маллера (РМ-коды) и др. Кроме того, были разработаны новые эффективные алгоритмы декодирования (Питерсон, Берлекэмп, Мэсси и др.).

- 1977 г., разработан метод сжатия на основе словаря А. Лемпелем (Abraham Lempel, изр. ученый) Я. Зивом (Jacob Ziv, изр. ученый). Метод является основой алгоритмов сжатия ZIP, ARJ, gzip и др.

Методы теории информации используются:

- для защиты данных в памяти вычислительных устройств, для передачи данных в вычислительных системах (такие системы очень чувствительны к очень малой доле ошибок, т. к. даже одиночная ошибка может нарушить всю программу вычислений);

- цифровых оптических дисках (компакт-дисках);

- в системах со сжатием данных;

- в системах связи с ограничением на передаваемую мощность, например, в системах ретрансляции через спутник, где увеличение мощности обходится очень дорого;

- в системах цифрового телевидения; обработки изображения;

- в системах передачи информации разного назначения, например, в системах с пакетной коммутацией и разделением во времени, где длинные двоичные сообщения разделяются на пакеты, и пакет передается в отведенное временное окно. Из-за нарушения синхронизации пакеты могут быть утеряны. Кодирование позволяет обеспечить надежную синхронизацию в такой системе.

Кодирование применяется для защиты специальных радиотехнических систем гражданского и военного назначения, например, радиолокационных и радионавигационных систем, систем видеонаблюдения от воздействия:

Алгоритмы теории информации – это защита информационных систем от случайного и несанкционированного доступа к информации; повышение надежности радиотехнических и вычислительных устройств, делая их нечувствительными к отказам и сбоям.

## 1. МОДЕЛЬ КАНАЛА ПЕРЕДАЧИ, ХРАНЕНИЯ, ОБРАБОТКИ И РАСПРЕДЕЛЕНИЯ ИНФОРМАЦИИ

Обобщенная модель канала имеет вид, представленный на рис.1.1.

Рис. 1.1. Обобщенная модель канала передачи информации

сообщений.

Кодер первичного кода представляет информацию в форме, позволяющей упростить дальнейшую обработку.

Кодер источника сообщения предназначен для устранения информационной избыточности. Он позволяет:

- более эффективно использовать частотный ресурс;
- повысить скорость передачи информации.

Корректирующий кодер вводит информационную избыточность в передаваемое сообщение с целью обнаружения и (или) исправления ошибок сравнительно небольшой кратности (число ошибок  $t = 1, 2, 3, 4$ )

Кодер канала (помехоустойчивый кодер) также предназначен для минимизации влияния помех на передаваемую информацию и в большинстве своем используется в специальных радиотехнических системах:

- системах дальнего космоса;
- спутниковых навигационных системах (например, GPS "NAVSTAR" (Global Positioning System "NAVSTAR"), "ГЛОНАСС" (глобальная навигационная спутниковая система));
- системах скрытной связи;
- радиолокационных системах дальнего обнаружения целей, системах наведения на цель с повышенной точностью (точное оружие);
- системах мобильной и фиксированной связи третьего поколения CDMA (Code Division Multiple Access – кодовое разделение каналов, множественный доступ).

Модулятор преобразует множество дискретных сигналов канального кодера в непрерывные сигналы, которые передаются по каналам.

Физической средой передачи информации – каналом может служить:

- радиоканал;
- проводной канал;
- оптический канал;
- магнитная лента;
- компакт-диск;
- запоминающее устройство (ЗУ) и т. п.

*Замечание.* Если в качестве канала передачи использовать ЗУ, то это канал передачи информации во времени в отличие, например, от радиоканала передачи информации в пространстве.

В канале формируется смесь сигнала и помехи вида

$$y(t) = x(t)\mu(t) + n(t),$$

где –  $x(t)$  передаваемый непрерывный сигнал,  $\mu(t)$  мультипликативная помеха,  $n(t)$  аддитивная помеха (как правило, шум с гауссовским распределением).

Декодер источника восстанавливает ту избыточность, которая была ранее устранена на передающей стороне.

### *Замечания*

1. Техническая реализация составляющих рис. 1.1 с номерами 3, 4, 5, 6, 10, 11, 12, 13 осуществляется на цифровой элементной базе.

2. В элементе 1 производится дискретизация по времени и квантование по уровню входной аналоговой реализации (сообщения) с формированием символов в двоичном или  $q$ -ичном алфавите.

## **1.2. Эталонная модель взаимосвязи открытых систем**

Для построения эффективных информационных систем для каналов с различной средой необходимо использовать и другие модели. Наиболее известна так называемая эталонная модель взаимосвязи открытых систем, где в обобщенном виде рассмотрены функции, выполняемые на различных уровнях. Модель представляет семиуровневую архитектуру.

1. На физическом уровне реализуется цифровой канал.

2. На канальном уровне реализуется процедуры кодирования по сжатию, шифрованию, помехоустойчивому кодированию информации.

3. На сетевом уровне реализуется передача информации от источника к адресату.

4. Транспортный уровень управляет сквозной передачей пакетов, с коррекцией ошибок.

5. Сеансовый уровень контролирует соединения между оконечными системами.

6. На уровне представления выполняются операции сжатия данных, защиты информации, преобразования форматов для обеспечения эффективного и безопасного взаимодействия.

7. Прикладной уровень предоставляет различные сетевые службы.

## **1.3. Первичное кодирование информации**

Дискретный поток двоичных символов, сформированный аналого-цифровым преобразователем (АЦП), из-за недопустимых частотных амплитудных и др. искажений непригоден для передачи, хранения и обработки информации. Поэтому двоичный код с выхода источника преобразуется в первичный код.

Первичное кодирование может осуществляться на основе весовых кодов, построенных с использованием двоичной, восьмеричной, шестнадцатеричной и двоично-десятичной системы счисления.

### **1.3.1. Рефлексные коды**

Первичное кодирование может осуществляться также на основе невесовых кодов. Наиболее типичным представителем таких кодов является код Грея. Основным свойством кода Грея является то, что разряды кодовых слов не имеют весов и любые два соседних кодовых слов различаются только в

одном разряде. Отсюда следует, что при ошибочном приеме соседнего кодового слова ошибочно будет принят только в один разряд в слове. Код Грея обладает свойством минимизации ошибок.

Перекодирования двоичных последовательностей требуется выполнять также для борьбы с нарушением синхронизации в синхронных телекоммуникационных системах. Наиболее распространенным кодом, используемым для повышения надежности синхронизации, является код FOMOT (Four Mode Ternary). Он относится к классу троичных, где используется принцип чередования алфавитов.

## 2. КАЧЕСТВЕННАЯ И КОЛИЧЕСТВЕННАЯ ОЦЕНКА ИНФОРМАЦИИ

Под термином «информация» понимаются сведения, известия, которые описывают некоторое событие, руководство к действию, или свойство какого-либо объекта. Эти сведения могут быть представлены определенными символами, буквами алфавита, или в каком-то другом виде, например, изображением объекта «интереса», словами, и пр. Формой представления информации является сообщение. В конкретных информационных системах сообщение может использоваться, передаваться, становиться объектом хранения, распределения, преобразования.

Н. Виннер (N. Wiener, амер. ученый) определил информацию как объект нематериальной природы [1]: «Информация есть информация, а не материал или энергия». В основе теории информации является положение о том, что любой источник информации можно описать вероятностными категориями, которые могут быть измерены. Каждое сообщение содержит в себе определенную информацию. Однако одни сообщения переносят больше информации, чем другие. Если в прогнозе погоды сообщается, что 1 января температура воздуха в Минске достигнет  $+20^{\circ}\text{C}$ , то это сообщение характеризуется очень большим количеством информации. Такое событие является неожиданным, редким, вероятность  $p$  его появления стремится к нулю,  $p \rightarrow 0$ . В сообщении, что 1 января температура воздуха в Минске ожидается  $-5^{\circ}\text{C}$  не является неожиданным. Вероятность  $p$  его появления стремится к единице,  $p \rightarrow 1$ . В сообщении о высоковероятном событии содержится мало информации. Вероятность события является мерой его неожиданности (неопределенности) и связана с количественной мерой информации. Можно предположить, что количество информации о событии, обратно величине вероятности его появления, т. е.

$$I \sim \log \frac{1}{p} \sim -\log p, \quad (2.1)$$

где  $I$  – количество информации, полученное с появлением сообщения с вероятностью  $p$ . Чтобы определить понятие количества информации, как измеряемой величины, необходимо вначале рассмотреть канал передачи информации и свойства источника информации.

## 2.1. Дискретный источник информации без памяти

Дискретный источник информации без памяти  $X$  в дискретный момент времени  $i$  формирует символ  $x_i$  случайной последовательности символов. Выход источника есть случайная величина. Множество исходных символов  $X = \{x_1, x_2, \dots, x_m\}$  называется алфавитом источника  $X$ , а элементы  $x_i$  – буквами или символами. Символами источника могут быть буквы, цифры или некие абстрактные знаки. Каждый  $n$ -ый символ из конечного алфавита  $X = \{x_1, x_2, \dots, x_m\}$  источника появляется на выходе с вероятностью  $p_i$ . Вероятности появления символов источника задаются в виде множества  $\{p_1, p_2, \dots, p_m\}$ . Все вероятности символов в сумме должны давать значение 1, т. е.

$$\sum_{i=1}^m p_i = 1,$$

где  $m$  – число различных символов множества определяет размерность используемого алфавита.

Например, если  $X = \{x_1 = a, x_2 = b, x_3 = c\}, X = \{a, b, c\}, m = 3, \{p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{6}\}$ .

Практически  $m \geq 2$ . Алфавиту  $X = \{0, 1\}, m = 2$ , соответствует двоичный источник без памяти. Выходными символами источника являются символы  $x_1 = 0$  и  $x_2 = 1$ . Обозначим  $p_1$  вероятность появления символа  $x_1$ . Тогда выражение  $p_2 = (1 - p_1)$  представляет вероятность появления символа  $x_2 = 1$ .

*Определение 2.1.* Два источника  $X = \{x_1, x_2, \dots, x_m\}$  и  $T = \{t_1, t_2, \dots, t_u\}$  являются независимыми, если совместная вероятность  $p_{x,t}$  каждой пары  $(x, t)$  событий  $x \in X, t \in T$  равна произведению

$$p_{x,t} = p_x p_t,$$

где  $p_x$  и  $p_t$  вероятности появления символов источников  $X$  и  $T$ .

### 2.1.1. Блоковый источник информации

*Определение 2.2.* Если выходом источника являются последовательности (блоки) из  $n$  одиночных статистически независимых символов алфавита  $X = \{x_1, x_2, \dots, x_m\}$ , то такой источник называется источником с  $n$ -кратным расширением  $X^n$  источника  $X$ .

*Замечание.* Источник  $X^n$  называют также блоковым источником.

На выходе блокового источника можно сформировать  $m^n$  символов. Например, для  $X = \{0, 1\}, n = 2$  возможное множество символов источника с 2-кратным расширением источника  $X$  есть

$$X^2 = \{c_1, c_2, c_3, c_4\},$$

где  $c_1 = (00), c_2 = (01), c_3 = (10), c_4 = (11)$



образуют множество, состоящее из  $m^n = 2^2 = 4$ -х символов блочного источника

## 2.2. Канал передачи информации

Передача информации, формируемой источником, осуществляется посредством использования канала передачи информации. Канал – это некоторая физическая среда, соединяющая источник информации с получателем. Примеры каналов: проводная телефонная линия, среда распространения электромагнитных волн (радиоканал), используемая, например, при соединении компьютера, имеющем Wi-Fi-адаптер, с сетью Internet. CD (компакт-диск) это тоже канал в виде ЗУ и пр. На рис. 2.1. изображена обобщенная математическая модель системы передачи информации, включающая канал.

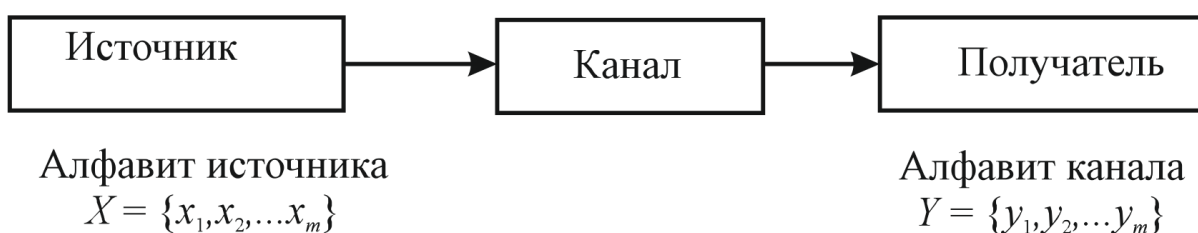


Рис. 2.1. Модель системы передачи информации

*Замечание.* Эффективность каналов передачи (хранения) информации возрастает с переходом на недвоичные символы.

## 2.3. Дискретный канал без памяти

Алфавиты передаваемых и принимаемых символов сообщения должны совпадать. Однако, из-за воздействия помех (шума) полученный символ может отличаться от переданного. В этом случае принимаемые символы называют алфавитом канала. На рис. 2.1 они обозначены как  $Y = \{y_1, y_2, \dots, y_m\}$  и приемник можно так же считать источником информации. Каналы с шумами характеризуется условными вероятностями  $p(y|x)$  для всех  $x \in X$  и  $y \in Y$ .

*Определение 2.3.* Условная вероятность  $p(y|x)$  понимается как вероятность того, что на выходе канала (входе приемника) появился символ  $y$ , при условии, что на выходе источника  $X$  был сформирован символ  $x$ .

*Замечание.* Условные вероятности  $p(y|x)$  называются вероятностями перехода канала.

*Определение 2.4.* Если имеется конечное число входов и выходов канала, и принимается, что вероятность  $p(y|x)$  не зависит от вероятностей появления предыдущих символов входа, то канал называется дискретным каналом без памяти.

В ряде приложений, например, связанных с задачами обнаружения сигналов на фоне помех, условную вероятность  $p(y|x)$  называют апостери-

орной (послеопытной) вероятностью. Вероятность  $p(y|x)$  определяет степень правдоподобия приема символа  $x$ , если был принят символ  $y$ .

Дискретный канал с алфавитом символов источника (входом канала)  $X = \{x_1, x_2, \dots, x_m\}$  и символов источника (выходом канала)  $Y = \{y_1, y_2, \dots, y_m\}$  описывается диаграммой переходных вероятностей  $p(y|x)$ . На рис. 2.2. показано графическое описание дискретного канала с двумя символами на входе и выходе канала.

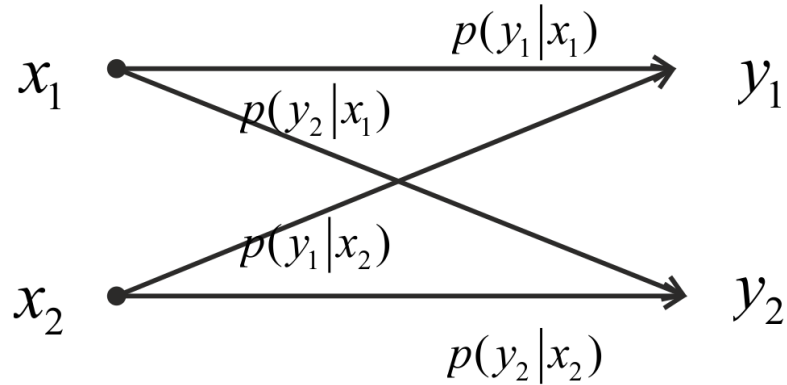


Рис. 2.2. Модель дискретного канал канала передачи информации с набором вероятностей перехода

### 2.3.1. Характеристика дискретного канала без памяти

Предположим, что событие  $X$  происходит на множестве всех возможных элементарных событий (выборочном пространстве)  $\{X\}$ . Напомним, событие является подмножеством выборочного пространства. Аналогично, событие  $Y$  является подмножеством выборочного пространства  $\{Y\}$ . Пусть  $p\{Y \cap X\}$  – это вероятность того, что произойдет как событие  $Y$ , так и событие  $X$ . Множество  $\{X \cap Y\}$  определяется результатом операции пересечения  $\cap$  над множествами  $\{X\}$  и  $\{Y\}$ . Необходимо определить вероятность события  $Y$  при условии, что событие  $X$  уже произошло. В этом случае выборочным пространством события  $Y$  служит пространство события  $X$ , а событие  $Y$  ограничивается множеством  $\{X \cap Y\}$ . Тогда вероятность наступления  $Y$  при условии, если событие  $X$  произошло, определяется как

$$p\{Y|X\} = \frac{p\{X \cap Y\}}{p\{X\}}. \quad (2.2)$$

Аналогично, справедливо выражение

$$p\{X|Y\} = \frac{p\{Y \cap X\}}{p\{Y\}}. \quad (2.3)$$

Вероятность  $p\{X \cap Y\}$  тождественна совместной вероятности  $p\{X, Y\}$  появления двух событий  $X$  и  $Y$ . Так как операция пересечения удовлетворяет

аксиоме коммутативности  $X \cap Y = Y \cap X$ , тогда из (2.2) и (2.3) следуют равенства:

$$\begin{aligned} p\{X \cap Y\} &= p\{Y|X\}p\{X\}; \\ p\{X, Y\} &= p\{Y, X\} = p\{Y|X\}p\{X\}; \end{aligned} \quad (2.4)$$

$$\begin{aligned} p\{Y \cap X\} &= p\{X|Y\}p\{Y\}; \\ p\{Y, X\} &= p\{X, Y\} = p\{X|Y\}p\{Y\}. \end{aligned} \quad (2.5)$$

Так как  $p\{X, Y\} = p\{Y, X\}$ , то

$$p\{Y|X\}p\{X\} = p\{X|Y\}p\{Y\}. \quad (2.6)$$

Решая уравнение (2.6) относительно  $p\{Y|X\}$ , получаем

$$p\{X|Y\} = \frac{p\{Y|X\}p\{X\}}{p\{Y\}}. \quad (2.7)$$

Формула (2.7) известна как теорема Байеса (Bayes' theorem). Эта формула имеет важное прикладное значение. Оценивая на выходе канала вероятность  $p\{Y\}$  формирования выходных символов, имея априорные значения вероятностей  $p\{X\}$  символов входа канала и зная свойства канала (переходные вероятности  $p\{Y|X\}$ ), можно найти вероятность  $p\{X|Y\}$  получения символов источника  $X$  на приемной стороне.

*Замечание.* Оптимальная обработка сигналов реализуется с использованием алгоритма (2.7).

Пусть на входе дискретного канала формируются множество несовместных событий  $X = \{x_1, x_2, \dots, x_m\}$  таких, что одно из них непременно произойдет. Напомним, два события  $x_i$  и  $x_j$  несовместны, если взаимно исключают друг друга, т. е.  $x_i \cap x_j = \emptyset, i \neq j$ . Все элементарные события являются взаимоисключающими. Каждое элементарное событие принадлежит одному и только одному  $x_i$  из множества  $X = \{x_1, x_2, \dots, x_m\}$ .

Аналогично, на выходе канала (входе приемника) формируются несовместные события  $Y = \{y_1, y_2, \dots, y_m\}$ .

Очевидно, объединение  $(x_1 \cup x_2 \cup \dots \cup x_m)$  всех  $x_i \in X$  дает пространство элементарных событий  $\{X\}$ . Объединение  $(y_1 \cup y_2 \cup \dots \cup y_m)$  всех  $y_i \in Y$  дает пространство  $\{Y\}$ . Тогда любое событие  $y_i$  на выходе канала может осуществиться только одновременно с некоторым событием  $x_j$ . Символически это можно записать как

$$y_i = (y_i x_1 \cup y_i x_2 \cup \dots \cup y_i x_m) \quad (2.8)$$

Поскольку события  $yx_i$  и  $yx_j$  попарно несовместны ( $yx_i \cap yx_j = \emptyset, i \neq j$ ), их вероятности складываются. Из (2.8) вероятность некоторого события  $y_i$  равна

$$p(y_i) = p(y_ix_1 \cup y_ix_2 \cup \dots \cup y_ix_m) \rightarrow p(y_ix_1) + p(y_ix_2) + \dots + p(y_ix_m). \quad (2.9)$$

Рассмотрим последнее выражение на примере.

*Пример 2.1.* Имеется дискретный канал с входным источником  $X = \{x_1, x_2\}, x_1 = 0, x_2 = 1$  и выходным источником  $Y = \{y_1, y_2\}, y_1 = 0, y_2 = 1$ . Напомним, приемник (выход канала) можно считать источником информации.

Из выражения (2.9) получаются значения вероятностей  $y_1$  и  $y_2$ :

$$p(y_1) = p(y_1x_1) + p(y_1x_2); \quad (2.10)$$

$$p(y_2) = p(y_2x_1) + p(y_2x_2). \quad (2.11)$$

Например, если на вход канала поступил символ  $x_2$ , тогда вероятности  $p(y_1)$  и  $p(y_2)$  на выходе канала определяются соответственно совместной вероятностью  $p(y_1x_2)$  и  $p(y_2x_2)$  символов  $y_1, y_2$  и  $x_2$ :

$$p(y_1) = p(y_1x_2);$$

$$p(y_2) = p(y_2x_2).$$

Подставляя выражение совместной вероятности (2.5) в (2.10) и (2.11) получаем формулы:

$$p(y_1) = p(y_1|x_1)p(x_1) + p(y_1|x_2)p(x_2), \quad (2.12)$$

$$p(y_2) = p(y_2|x_1)p(x_1) + p(y_2|x_2)p(x_2). \quad (2.13)$$

Вероятность  $p(y_i)$  выхода канала (входа приемника) и распределение вероятностей входа канала (выхода источника  $X$ ) связано следующим выражением:

$$p(y_i) = \sum_{j=1}^2 p(y_i|x_j)p(x_j). \quad (2.14)$$

Таким образом, если известны значения вероятности символов источника и переходные характеристики дискретного канала без памяти, можно вычислить вероятности символов на выходе канала.

В общем случае, для источников  $X = \{x_1, x_2, \dots, x_m\}$  и  $Y = \{y_1, y_2, \dots, y_m\}$  (для входа и выхода канала) формула (2.14) примет вид

$$p(y_i) = \sum_{j=1}^m p(y_i|x_j)p(x_j). \quad (2.15)$$

*Пример 2.2.* Имеется дискретный канал с входным источником  $X = \{x_1, x_2\}$ ,  $x_1 = 0$ ,  $x_2 = 1$  и выходным источником  $Y = \{y_1, y_2\}$ ,  $y_1 = 0$ ,  $y_2 = 1$ . Символы источника  $X$  появляются с вероятностью  $p(x_1) = 0,9$  и  $p(x_2) = 0,1$ . В канале имеются шумы. Переходные вероятности канала соответственно равны:

$$p(y_1|x_1) = 0,99; p(y_2|x_1) = 0,01; p(y_2|x_2) = 0,95; p(y_1|x_2) = 0,05.$$

1. Определить вероятности появления символов на выходе канала с шумами.

Решение. По формуле (2.14) получаем:

$$\begin{aligned} p(y_1) &= p(y_1|x_1)p(x_1) + p(y_1|x_2)p(x_2) = \\ &= 0,99 \cdot 0,9 + 0,05 \cdot 0,1 = 0,896; \end{aligned}$$

$$\begin{aligned} p(y_2) &= p(y_2|x_1)p(x_1) + p(y_2|x_2)p(x_2) = \\ &= 0,01 \cdot 0,9 + 0,95 \cdot 0,1 = 0,104. \end{aligned}$$

2. Найти вероятность  $p\{x_i|y_i\}$  получения символов источника  $X$  на приемной стороне в условиях присутствия шумов в канале.

Решение. По теореме Байеса (2.7)

$$p(x_i|y) = \frac{p(y_j|x_i)p(x_i)}{p(y_j)}$$

получаем:

$$p(x_1|y_1) = \frac{p(y_1|x_1)p(x_1)}{p(y_1)} = \frac{0,99 \cdot 0,9}{0,896} = 0,9944.$$

$$p(x_2|y_2) = \frac{p(y_2|x_2)p(x_2)}{p(y_2)} = \frac{0,95 \cdot 0,1}{0,104} = 0,9134.$$

Как видно, в канале с шумом значение вероятности правильного приема символа зависит от степени его повторяемости. Чем чаще он повторяется, тем достовернее прием.

Значения вероятностей ошибок в реальных каналах зависят от многих факторов:

- свойств физических каналов;
- свойств сигналов, которые являются физическими переносчиками сообщений;
- метода обработки сигналов на приемной стороне и пр;
- отношения мощности сигнала к мощности шума  $\frac{S}{N}$  на выходе канала передачи информации.

Величину отношения мощности сигнала к мощности шума часто выражают в логарифмическом масштабе

$$\frac{S}{N} = (10 \log_{10} \frac{S}{N}) \text{ dB}.$$

Например, для цифрового телевизионного вещания с хорошим качеством стандартные значения отношения  $\frac{S}{N}$  составляют (60 – 70) dB. В этом случае отношения  $\frac{S}{N}$  превышает величину  $10^6$ .

## 2.4. Количественная оценка информации

Понятие количества информации, предложенное К. Шенноном в 1948 году, определяется при выполнении трех аксиом.

1. Информация события (символа)  $x_i \in X$ , появляющегося с вероятностью  $p_i$ , имеет положительное значение

$$I(p_i) \geq 0.$$

2. Аксиома суммируемости информации.

Если независимые события  $(x_i, x_j)$  появляются с вероятностью  $p_i$  и  $p_j$ , то вероятность совместного события  $x_i$  и  $x_j$  равна  $P(x_i, x_j) = p_i \cdot p_j$ . Напомним, если исход одного события не влияет на исход другого, то такие события называются независимыми.

*Пример 2.3.* Пусть двоичный источник без памяти  $X = \{0,1\}$ , формирует символ  $x_1$  с вероятностью  $p = 0,2$  и символ  $x_2$  с вероятностью  $(1 - p) = 0,8$ . Вероятность появления сообщения вида  $(x_2 x_1 x_1 x_2 x_1) = (10010)$  равна

$$P(x_2 x_1 x_1 x_2 x_1) = P(10010) = p^3 (1 - p)^2 = 0,2^3 (1 - 0,2)^2 = 0,00512.$$

Совместная информация двух независимых событий  $(x_i, x_j)$  с вероятностью совместного события  $P(x_i, x_j) = p_i \cdot p_j$  равна сумме их информаций

$$I(p_{i,j}) = I(p_i) + I(p_j).$$

Если Вы получили сообщение о том, что 1 июня температура воздуха в Минске достигнет  $+20^\circ\text{C}$  и что экзамен состоится в аудитории 505-3 – это независимые события. Содержание этого сложного сообщения равняется сумме информации о погоде и экзамене.

3. Информация является непрерывной функцией вероятности события.

*Определение 2.5.* Количество информации, передаваемое источником

при появления одного символа  $x_i$  с вероятностью  $p$ , равно

$$I = \log \frac{1}{p} = -\log p. \quad (2.16)$$

Для логарифма может быть использовано основание 10, основание 2, основание  $e$  натуральных логарифмов. Разные основания только изменяют единицы меры информации. Измерение объема информации по формуле (2.14) впервые было предложено Р. В. Л. Хартли (амер. ученый) в 1928 году. При использовании логарифмов с основанием 10 количество информации измеряется в единицах Хартли.

*Пример 2.4.*

$p = 10^{-5}$ , тогда  $I = \log_{10} \frac{1}{p} = -\log_{10} 10^{-5} = 5$  единиц информации Хартли.

$p = 10^{-1}$ ,  $I = \log_{10} \frac{1}{p} = -\log_{10} 10^{-1} = 1$  единица информации Хартли.

$p = 1$ ,  $I = \log_{10} \frac{1}{p} = -\log_{10} 1 = 0$  единиц информации Хартли.

При использовании логарифмов с основанием 2 количество информации измеряется в битах.

*Пример 2.5.*

$p = \frac{1}{2}$ , тогда  $I = \log_2 \frac{1}{1/2} = -\log_2 \frac{1}{2} = 1$  бит;

$p = \frac{1}{32}$ , тогда  $I = \log_2 \frac{1}{1/32} = -\log_2 \frac{1}{32} = 5$  бит.

*Пример 2.6.* Пусть передается сообщение  $c = (x_1 x_2 x_3 x_4 x_5) = (10010)$ , составленное из независимых символов  $x_i \in \{0,1\}$ . События  $x_i$  появляются с вероятностью  $p_i = \frac{1}{2}$ . Количество информации в этом сообщении равно

$$I = \log_2 \frac{1}{P(x_1 x_2 x_3 x_4 x_5)} = \log_2 \frac{1}{(\frac{1}{2})^5} = -\log_2 2^{-5} = 5 \text{ бит.}$$

Полученное значение соответствует сумме информаций 5 независимых событий

$$I(c) = I(p_1) + \dots + I(p_5) = 5 \text{ бит.}$$

На рис. 2.3 показан график, характеризующий количественное изменение  $I$  в зависимости от вероятности события.

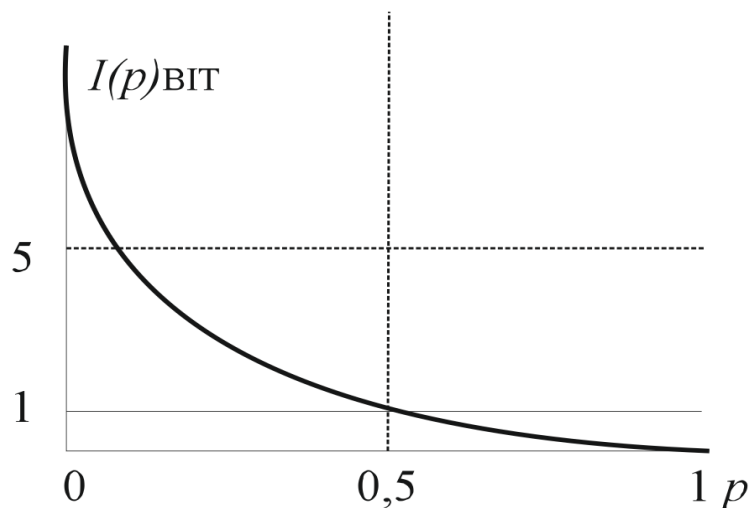


Рис. 2.3. Информация события  $x \in X$ , появляющегося с вероятностью  $p$ ,

Как видно, с уменьшением вероятности появления события или увеличением его неопределенности, количество информации возрастает. Определение информации можно трактовать как некоторое отражение возникновения событий.

#### Упражнения

2.1. Вычислить количество информации выдаваемой источником, если размерность алфавита  $X = \{x_1, x_2, \dots, x_6\}$  равна  $m = 6$ . Вероятность появления события

$p_1 = 0,05$ ;  $p_2 = 0,15$ ;  $p_3 = 0,05$ ;  $p_4 = 0,4$ ;  $p_5 = 0,2$ ;  $p_6 = 0,15$ .

2.2. Вычислить количество информации выдаваемой источником, если размерность алфавита  $X = \{x_1, x_2, \dots, x_m\}$  равна  $m = 3$ . Вероятность появления события  $p_1 = 0,15$ ;  $p_2 = 0,5$ ;  $p_3 = 0,35$ .

## 2.5. Энтропия

Пусть двоичный дискретный источник без памяти  $X = \{x_1, x_2\}$ ,  $m = 2$  формирует символ  $x_1 = 0$  с вероятностью  $p$  и символ  $x_2 = 1$  с вероятностью  $(1 - p)$ . Если получен символ  $x_1$ , то это сообщение оценивается количеством информации, равным

$$I(x_1) = -\log p.$$

Аналогично, при приеме символа  $x_2$  количество полученной информации определяется как

$$I(x_2) = -\log(1 - p).$$

Одной из характеристик двоичного дискретного источника без памяти является среднее количество (ожидаемое количество) информации выдаваемой источником. Так как такой источник формирует случайные события, то математическое ожидание определяется по формуле



$$E(I) = \sum_{i=1}^m p_i I(x_i) = \sum_{i=1}^2 p_i I(x_i) = p_1 I(x_1) + p_2 I(x_2) = \\ = -p \log p - (1-p) \log(1-p).$$

*Пример 2.7.* Пусть  $p = 0,2$ .  $I(x_1) = -\log 0,2 = 2,3219$  бит,  $I(x_2) = -\log 0,8 = 0,3219$  бит. Полученные значения  $I(x_i)$  соответствуют точкам на графике, показанном на рис. 2.3.

Среднее значение количества информации источника

$$E(I) = 0,2 \cdot 2,3219 + 0,8 \cdot 0,3219 = 0,7219 \text{ бита.}$$

*Определение 2.6.* Энтропия источника информации  $H$  – это средняя информация, полученная для всех возможных событий.

Энтропия источника (*пример 2.7*) равна  $H = 0,7219$  бит/символ. В этом примере энтропия источника информации определяется как математическое ожидание количества информации

$$E(I) = H.$$

Для дискретного источника двух независимых событий  $X = \{0,1\}$  с вероятностями  $p$  и  $(1-p)$  энтропия определяется как

$$H = -p \log_2 p - (1-p) \log_2 (1-p). \quad (2.17)$$

*Замечание.* Энтропия двоичного источника, вычисляемая по формуле (2.17) называется функцией (формулой) Шеннона.

На рис. 2.4 показан график энтропии двух событий как функция вероятности. Максимальное значение энтропии равно 1 бит/символ, когда  $p_1 = p_2 = \frac{1}{2}$ .

$$H = -p \log_2 p - (1-p) \log_2 (1-p) = \\ = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ бит/символ.}$$

Это соответствует наибольшей неопределенности для двух событий. Для значения  $p$  равного нулю или единице события имеют полную определенность, и никакая информация не передается.

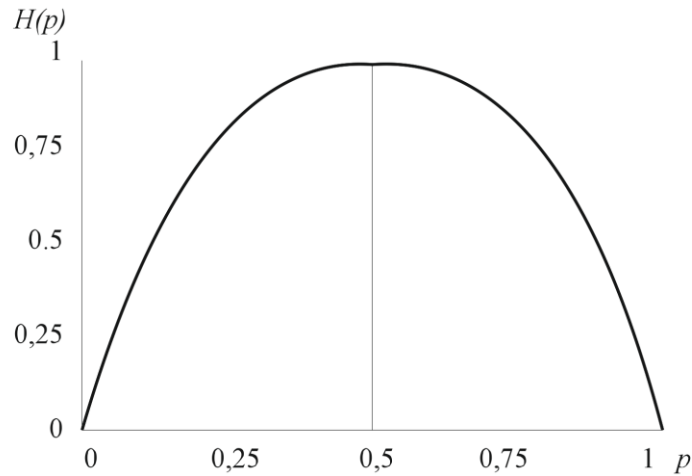


Рис. 2.4. Энтропия двоичного источника

**Определение 2.7.** Энтропия дискретного источника без памяти с символами алфавита  $X = \{x_1, x_2, \dots, x_m\}$  и соответствующими вероятностями  $p_1, p_2, \dots, p_m$  равна

$$H = H(p_1, p_2, \dots, p_m) = \sum_{i=1}^m -p_i \log p_i. \quad (2.18)$$

*Замечания.*

1. Величину (2.18) называют также неопределенностью источника.

2. Из *Определения 2.7.* следует понятие энтропии как среднее количество информации, приходящейся на один символ источника.

**Пример 2.8.** Вычислим энтропию источника с алфавитом из четырех символов  $X = \{x_1, x_2, x_3, x_4\} = \{a, b, c, d\}$  с вероятностями  $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, p_4 = \frac{1}{8}$ .

*Решение.*

$$\begin{aligned} H &= \sum_{i=1}^4 -p_i \log_2 p_i = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{8} \log_2 \frac{1}{8} + \frac{1}{8} \log_2 \frac{1}{8}\right) = \\ &= \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1,75 \text{ бит/символ.} \end{aligned}$$

### 2.5.1. Свойства энтропии

1. Энтропия  $H(p_1, p_2, \dots, p_m) = \sum_{i=1}^m -p_i \log p_i$  является неотрицательной непрерывной функцией вероятностей событий  $p_1, p_2, \dots, p_m$ .

Доказательство неотрицательности  $H(p_1, p_2, \dots, p_m) \geq 0$  очевидно. Так как  $\log p_i \leq 0$ , то  $-\log p_i \geq 0$  для всех значений  $i = 1, 2, \dots, m$ .

2. Для дискретного источника без памяти с равной вероятностью  $p_i = \frac{1}{m}$  энтропия увеличивается с увеличением размерности алфавита  $m$ .

Пример 2.9. Вычислим энтропию источника с алфавитом из четырех символов  $X = \{x_1, x_2, x_3, x_4\} = \{a, b, c, d\}$  с равными вероятностями  $p_1 = \frac{1}{4}, p_2 = \frac{1}{4}, p_3 = \frac{1}{4}, p_4 = \frac{1}{4}$ .

Решение.

$$H = \sum_{i=1}^4 -p_i \log_2 p_i = -\left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 2 \text{ бит/символ.}$$

В примере 2.8, где разные значения вероятностей  $p_i$ ,  $H = 1,75$  бит/символ.

**Теорема 2.1.** Если все события имеют одинаковую вероятность  $p_1 = \dots = p_i, \dots = p_m$ , энтропия дискретного источника без памяти максимальна и равна

$$H_0 = \log_2 m \text{ бит/символ.}$$

В этом случае неопределенность источника максимальна и источник передает максимально возможное среднее количество информации, приходящее на один символ (см. рис. 2.3 и пример 2.9).

*Определение 2.8.* Величина  $H_0$  определяет емкость дискретного источника как системы хранения информации.

3. Источник без памяти с разными значениями вероятности появления символов алфавита обладает энтропией, меньшей  $\log_2 m$ .

Сравнивая источник примера 2.8, где  $m = 4$ ,  $H = 1,75$  бит/символ с источником такого же размера, но с одинаковыми значениями вероятностями  $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ , имеем

$$H = 1,75 < H_0 = \log_2 m = 2.$$

4. Энтропия блокового источника равна

$$H' = nH,$$

где  $H$  – энтропия источника одиночных символов.

*Пример 2.10.*

1) Источник формирует символы  $X = \{x_1, x_2\} = \{0, 1\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ . Размерность алфавита  $m = 2$ .

Энтропия источника одиночных символов равна

$$H = \sum_{i=1}^2 -p_i \log_2 p_i = \frac{1}{3} \cdot 1,585 + \frac{2}{3} \cdot 0,585 = 0,918 \text{ бит/символ.}$$

2) Имеется блоковый источник  $X^2 = \{c_1, c_2, c_3, c_4\}$ ,  $n = 2$ . Символы  $c_1 = (00), c_2 = (01), c_3 = (10), c_4 = (11)$  получены расширением источника одиночных символов  $X = \{x_1, x_2\} = \{0, 1\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ .

Вычислить энтропию источника  $X^2$ . Решение.

1. Вычисляем вероятности появления символов источника  $X^2$ . Вероятность совместного события  $x_i$  и  $x_j$  равна  $P(x_i, x_j) = p_i \cdot p_j$ . Поэтому

$$\begin{aligned} P(c_1) &= p_1 \cdot p_1 = \frac{1}{9}, \\ P(c_2) &= p_1 \cdot p_2 = \frac{2}{9}, \\ P(c_3) &= p_2 \cdot p_1 = \frac{2}{9}, \\ P(c_4) &= p_2 \cdot p_2 = \frac{4}{9}. \end{aligned}$$

2. Энтропия источника равна

$$\begin{aligned} H' &= \sum_{i=1}^4 -P(c_i) \log_2 P(c_i) = \frac{1}{9} \cdot 0,585 + \frac{2}{9} \cdot 2,1699 + \frac{2}{9} \cdot 2,1699 + \\ &\quad + \frac{4}{9} \cdot 1,1699 = 1,83 \text{ бит/символ}. \end{aligned}$$

Как видно, энтропия блокового источника определяется свойством 4,

$$H' = nH = 2 \cdot 0,918 = 1,83 \frac{\text{бит}}{\text{символ}}.$$

## 2.6. Относительная избыточность источника

*Определение 2.9.* Избыточность дискретного источника без памяти  $X = \{x_1, x_2, \dots, x_m\}$  – это разность между емкостью  $H_0$  источника и энтропией источника

$$R = H_0 - H. \quad (2.19)$$

*Определение 2.10.* Относительной избыточностью источника называется величина

$$r = \frac{R}{H_0} = 1 - \frac{H}{H_0}. \quad (2.20)$$

*Пример. 2.11.* Используя данные примера 2.8 (источник  $X = \{x_1, x_2, x_3, x_4\}$  с разными вероятностями),  $H = 1,75$  и значение  $H_0 = 2$  для такого же источника, но с одинаковыми вероятностями, получаем величину избыточности

$$R = 2 - 1,75 = 0,25.$$

Относительная избыточность источника равна

$$r = 1 - \frac{1,75}{2} = 0,125 \cong 12,5\%.$$

Упражнения

2.3. Вычислить энтропию дискретного источника без памяти с символами алфавита  $X = \{a, b\}$  с вероятностью  $p_a = \frac{6}{8}, p_b = \frac{1}{4}$ .

2.4. Вычислить энтропию дискретного источника без памяти с символами алфавита  $X = \{a, b, c\}$  с вероятностью  $p_a = \frac{1}{2}, p_b = \frac{1}{3}, p_c = \frac{1}{6}$ .

2.5. Источник формирует следующие символы  $X = \{x_1, x_2, \dots, x_6\} = \{A, K, N, D, E, !\}$ . Вероятности символов задаются множеством:  $\{p_1 = 0,05, p_2 = 0,15, p_3 = 0,05, p_4 = 0,4, p_5 = 0,2, p_6 = 0,15\}$ .

2.5.1. Вычислить энтропию дискретного источника.

2.5.2. Вычислить емкость дискретного источника.

2.5.3. Вычислить избыточность дискретного источника.

2.5.4. Вычислить относительную избыточность дискретного источника.

### 3. ЭФФЕКТИВНОЕ КОДИРОВАНИЕ ДЛЯ ДИСКРЕТНОГО ИСТОЧНИКА БЕЗ ПАМЯТИ

#### 3.1. Условия взаимной однозначности алфавитного кодирования

Кодирование связано с преобразованием выходных символов дискретного источника (событий источника) в последовательность символов заданного кодового алфавита. Каждому событию соответствует один символ. На рис. 3.1. изображена математическая модель системы передачи информации с кодированием.

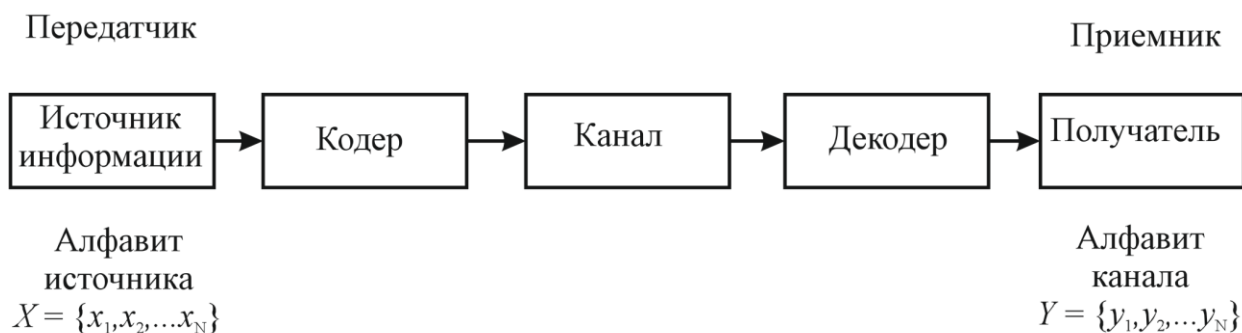


Рис. 3.1. Система передачи информации с кодированием

**Определение 3.1.** Код источника – это множество дискретных последовательностей всех событий, представленных символами кодового алфавита.

В качестве кодового алфавита могут использоваться символы двоичного  $\{0,1\}$  или бинарного алфавита  $\{1, -1\}$ . Например, слово  $x = (x_1 x_2 \dots x_7) = (-1 - 1 - 111 - 11)$  представлено символами бинарного алфавита источни-

ка,  $m = 2$ . В этом случае размерность кодового алфавита равна двум. На практике размерность кодового алфавита может быть большей.

*Определение 3.2.* Последовательности символов называются кодовыми словами или кодовыми векторами.

В результате кодирования осуществляется однозначное присвоение кодовых слов символам источника. Недопустимо присваивать символам одинаковые кодовые слова. Код должен удовлетворять условию несингулярности, когда каждое кодовое слово соответствует уникальному символу дискретного источника. Если код сингулярный, можно однозначно определить соответствующий символ источника по его кодовому слову.

Как правило, кодовые слова получаются  $n$ -кратным расширением источника одиночных символов.

*Определение 3.3.* Длина кода  $n$  (значность) – число символов кодового слова.

Параметр  $n$  определяет следующие особенности кодов. Коды бывают:

- равномерные (блоковые),  $n = \text{const}$ ;
- неравномерные,  $n = \text{var}$ .

Говорят, что двоичное слово  $x = (x_1 x_2 \dots x_n) = (1110010)$  имеет длину  $n = 7$ , слово  $l = (abbcab)$  длиной  $n = 6$ .

Код можно представить в виде списка, в котором каждому кодовому слову однозначно соответствует символ источника.

Пример 3.1. Пусть для передачи сообщения "DANKE" используется следующие кодовые слова равномерного кода:

- $A \rightarrow (000)$ ;
- $K \rightarrow (010)$ ;
- $N \rightarrow (001)$ ;
- $D \rightarrow (111)$ ;
- $E \rightarrow (100)$ .

Для построения этого кода использовались символы двоичного источника  $X = \{0,1\}$ . Кодированному сообщению "DANKE" соответствует последовательность независимых символов  $x_i \in \{0,1\}$ ,

$$DANKE \rightarrow (111000001010100).$$

Пусть символы источника  $x_i$  появляются с вероятностью  $p_i = \frac{1}{2}$ . Количество информации в этом сообщении равно

$$I = \log_2 \frac{1}{P(x_1 \dots x_{15})} = \log_2 \frac{1}{(\frac{1}{2})^{15}} = -\log_2 2^{-15} = 15 \text{ бит.}$$

Пример 3.2. В случае неравномерного кодирования используем код со словами:

$$\begin{aligned}A &\rightarrow (00); \\K &\rightarrow (10); \\N &\rightarrow (010); \\D &\rightarrow (110); \\E &\rightarrow (111).\end{aligned}$$

Сообщению "DANKE" соответствует последовательность двоичных символов

$$DANKE \rightarrow (1100001010111).$$

Количество информации в этом сообщении равно

$$I = \log_2 \frac{1}{P(x_1 \dots x_{13})} = \log_2 \frac{1}{(\frac{1}{2})^{13}} = -\log_2 2^{-13} = 13 \text{ бит.}$$

Из примеров 3.1 и 3.2 следует очевидный вывод: неравномерное кодирование более эффективно. Для передачи сообщения "DANKE" с использованием неравномерного кода потребовалось 13 бит. При равномерном кодировании того же сообщения затрачено 15 бит.

Правила кодирования должны отвечать следующим требованиям.

1. Необходимо добиваться высокой вероятности однозначного (правильного) декодирования исходной информации дискретного источника по закодированной последовательности.
2. Число символов кода, требуемого на один символ источника должно быть минимальным.

### 3.2. Эффективное кодирование

Основная идея эффективного кодирования базируется на использовании коротких кодовых слов для событий, характеризующихся высокой вероятностью. В этом случае будет уменьшаться длина закодированных сообщений. При этом должно обеспечиваться однозначное декодирование (желательно, без введения дополнительных символов – меток синхронизации между кодовыми словами).

*Определение 3.4.* Код является эффективным, если он имеет наименьшую возможную среднюю длину кодового слова.

Многие алгоритмы эффективного кодирования в качестве однозначно декодируемого кода основываются на применении префиксных моментальных кодов.

*Определение 3.5.* Префиксный код – это множество кодовых слов, в

котором каждое кодовое слово не совпадает с началом более длинного слова.

### 3.2.1. Моментальные коды

*Определение 3.6.* Если процесс однозначного декодирования каждого кодового слова осуществляется сразу же после приема всех символов кодового слова и принимается решение о соответствующем символе источника, то код с таким свойством называется моментальным.

Коды, рассмотренные в примерах 3.1 и 3.2, относятся к моментальным. Критерием моментальности кода является то, что ни одно слово, не совпадает с началом более длинного кодового слова.

#### 3.2.1.1. Код с запятой

Примером моментального кода служит код с запятой. Нуль в конце слова означает запятую, разделяющую кодовые слова. При получении нуля принимается решение о декодировании соответствующего символа. На рис. 3.2 показан пример множества слов кода с запятой. При получении нуля моментально принимается решение о декодировании соответствующего символа.

$$\begin{aligned}c_1 &\rightarrow (0); \\c_2 &\rightarrow (10); \\c_3 &\rightarrow (110); \\c_4 &\rightarrow (1110); \\c_5 &\rightarrow (11110).\end{aligned}$$

Рис. 3.2. Код с запятой

Например, последовательность  $c = (1111001101101011100)$  декодируется как  $c_5c_1c_3c_3c_2c_4c_1$ . Очевидно, введение разделительного символа уменьшает эффективность кодирования. На практике желательно использовать коды без запятой, когда при установленной синхронизации возможна передача кодированной информации без специального разделения кодовых слов.

Кодовая конструкция моментального кода иллюстрируется с помощью кодового дерева.

#### 3.2.1.2. Кодовое дерево

Кодовое дерево имеет начальную точку отсчета (корень). Из этой точки изображаются одна или две ветви. Ветвям присваиваются значения символов 0 и 1. Слева располагаются ребра, соответствующие символу 0, справа – ребра, соответствующие символу 1. Ветви заканчиваются узлами. Затем из этих узлов строятся еще одна или две ветви и т.д. пока не будет изображен конечный узел, соответствующий каждому символу источника – кодовому



слову. Кодовое слово получается в результате движения по ветвям от корня и записи символов 0 и 1 ветвей. Для наглядности изобразим кодовое дерево для кодового алфавита  $\{0,1\}$ . Для однозначно декодируемого кода не должно быть узлов на более длинном пути, заканчивающемся узлом, т.е. кодовым словом.

Пример 3.3. Пусть символы источника задаются множеством  $X = \{A, K, N, D, E\}$ . Для кодирования источника используем префиксный код со словами:

$A \rightarrow (00);$   
 $K \rightarrow (10);$   
 $N \rightarrow (010);$   
 $D \rightarrow (110);$   
 $E \rightarrow (111).$

На рис. 3.3 показано кодовое дерево неравномерного кода. Черные точки (конечные узлы дерева) соответствуют словам кода.

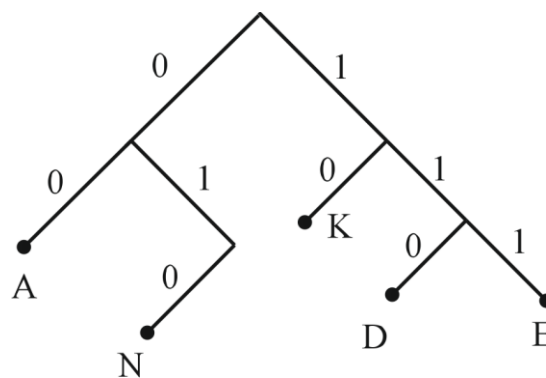


Рис. 3.3. Кодовое дерево неравномерного симплексного кода

Изображение кодового дерева равномерного кода из примера 3.1,  $(A \rightarrow (000), K \rightarrow (010), N \rightarrow (001), D \rightarrow (111), E \rightarrow (100))$ , показано на рис. 3.4.

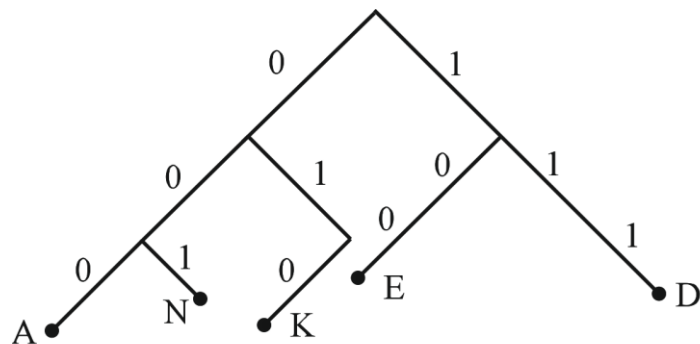


Рис. 3.4. Кодовое дерево равномерного кода

Изображение множества слов кода с запятой  $\{c_1 \rightarrow (0), c_2 \rightarrow (10), c_3 \rightarrow (110), c_4 \rightarrow (1110), c_5 \rightarrow (11110)\}$  показано на рис. 2.5.

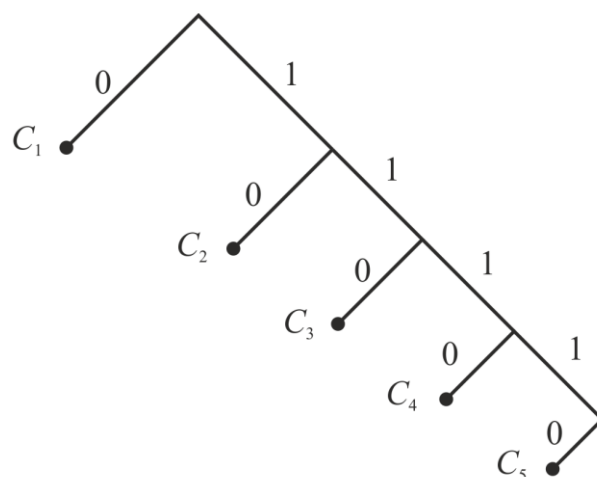


Рис. 3.5. Кодовое дерево кода с запятой

### Упражнения

3.1. Построить кодовое дерево кода  $X = \{x_1, x_2, \dots, x_8\}$ .

$x_1 = (01), x_2 = (00), x_3 = (111), x_4 = (110), x_5 = (100), x_6 = (1011),$   
 $x_7 = (10101), x_8 = (10100).$

3.2. Построить кодовое дерево кода  $X = \{x_1, x_2, \dots, x_{11}\}$ .  $x_1 = (0001), x_2 = (001), x_3 = (01), x_4 = (010), x_5 = (0111), x_6 = (0110),$   
 $x_7 = (1000), x_8 = (1001), x_9 = (101), x_{10} = (110), x_{11} = (111).$

3.3.1. Является ли следующий код:

$A \rightarrow (01),$   
 $K \rightarrow (10),$   
 $N \rightarrow (011),$   
 $D \rightarrow (101)$

однозначно декодируемым?

3.3.2. Построить кодовое дерево этого кода.

### 3.2.2. Неравенство Крафта

Для ответа на вопрос, будет ли предлагаемое множество кодовых слов кода при декодировании точно соответствовать исходной информации источника, применяется неравенство Крафта.

*Определение 3.7.* Для построения однозначно декодируемого  $q$ -ичного кода, содержащего  $m$  кодовых слов с длинами  $n_1, n_2, \dots, n_m$ , необходимо и достаточно, чтобы выполнялось неравенство Крафта

$$\sum_{i=1}^m q^{-n_i} \leq 1. \quad (3.1)$$

где  $q$  обозначает число символов кодового алфавита.

Если используются кодовые символы двоичного алфавита ( $q = 2$ ), неравенство Крафта записывается как

$$\sum_{i=1}^m q^{-n_i} = \sum_{i=1}^m 2^{-n_i} = 2^{-n_1} + 2^{-n_2} + \dots + 2^{-n_m} \leq 1.$$

Если используются кодовые символы двоичного алфавита ( $q = 2$ ) для формирования равномерного кода состоящего из  $m$  слов значностью  $n$ , неравенство Крафта имеет следующий вид:

$$\sum_{i=1}^m 2^{-n} \equiv m2^{-n} \leq 1. \quad (3.2)$$

Например, для  $m = 3$ ,  $n = 2$  имеем:

$$\sum_{i=1}^3 2^{-n_i} = 2^{-n_1} + 2^{-n_2} + 2^{-n_3} = 2^{-2} + 2^{-2} + 2^{-2} = 3 \cdot 2^{-2} = \frac{3}{4} < 1.$$

Рассмотрим примеры применения неравенства Крафта.

Пример 3.4. Используются следующие кодовые слова длиной  $n = 3$  равномерного кода:

$A \rightarrow (000);$

$K \rightarrow (010);$

$N \rightarrow (001);$

$D \rightarrow (111).$

$E \rightarrow (100).$

Удовлетворяет ли код неравенству Крафта?

Решение. Так как  $n_1 = n_2 = \dots = n_5 = n = 3$ ,

$$m2^{-n} = 5 \cdot 2^{-3} = \frac{5}{8} \leq 1.$$

Данный код однозначно декодируемый.

*Замечание.* Для равномерного кода двоичного алфавита максимальное число кодовых слов равно

$$m = 2^n.$$

Это значение позволяет закодировать  $2^n$  символов блокового источника.

Пример 3.5. Пусть для кодирования используется префиксный код со словами:

$A \rightarrow (00);$

$K \rightarrow (10);$

$N \rightarrow (010);$

$D \rightarrow (110);$

$E \rightarrow (111).$

Удовлетворяет ли код неравенству Крафта?

Решение. Так как  $n_1 = n_2 = 2, n_3 = n_4 = n_5 = 3, m = 5$ ,

$$\sum_{i=1}^5 2^{-n_i} = \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{7}{8} \leq 1.$$

Данный код также однозначно декодируемый.

Упражнения

3.4. Является ли код  $X = \{x_1, x_2, \dots, x_8\} =$

$x_1 = (01), x_2 = (00), x_3 = (111), x_4 = (110), x_5 = (100), x_6 = (1011),$   
 $x_7 = (10101), x_8 = (10100)$

однозначно декодируемым?

3.5. Является ли код  $X = \{x_1, x_2, \dots, x_{11}\}$ .

$x_1 = (0001), x_2 = (001), x_3 = (01), x_4 = (010), x_5 = (0111), x_6 = (0110),$   
 $x_7 = (1000), x_8 = (1001), x_9 = (101), x_{10} = (110), x_{11} = (111).$

однозначно декодируемым?

3.6. Из каких следующих значений длин кодовых слов можно построить однозначно декодируемый код?

3.6.1.  $n_1 = 2, n_2 = 2, n_3 = 2, n_4 = 3, n_5 = 3.$

3.6.2.  $n_1 = 1, n_2 = 2, n_3 = 3, n_4 = 3, n_5 = 8.$

3.6.3.  $n_1 = 1, n_2 = 2, n_3 = 3, n_4 = 4, n_5 = 4.$

### 3.2.3. Средняя длина кодового слова

*Определение 3.8.* Мерой эффективности кода является его средняя длина кодовых слов

$$L_n = \sum_{i=1}^m P_i l_i, \quad (3.3)$$

где  $m$  – число символов источника с  $n$ -кратным расширением источника одиночных символов,  $P_1, \dots, P_m$  – вероятности символов источника с  $n$ -кратным расширением,  $l_1, \dots, l_m$  – длина соответствующих кодовых слов.

Пример 3.6. Пусть для передачи сообщения "DANKЕ" используется следующие кодовые слова равномерного кода:

$A \rightarrow (000);$

$K \rightarrow (010);$

$N \rightarrow (001);$

$D \rightarrow (111);$

$E \rightarrow (100).$

Для построения этого кода использовались символы двоичного источника  $X = \{0,1\}$ . Пусть код характеризуется вероятностями  $P_1 = P_2 = P_3 = P_4 = P_5 = \frac{1}{5}$ . Код имеет среднюю длину

$$L_n = \sum_{i=1}^5 \frac{1}{5} l_i = \frac{1}{5} 3 + \dots + \frac{1}{5} 3 = \frac{15}{5} = 3.$$

Пример 3.7. Пусть для передачи сообщения "DANKЕ" используются следующие кодовые слова неравномерного кода:

$$\begin{aligned}A &\rightarrow (00); \\K &\rightarrow (10); \\N &\rightarrow (010); \\D &\rightarrow (110); \\E &\rightarrow (111).\end{aligned}$$

Для построения этого кода использовались символы двоичного источника  $X = \{0,1\}$ . Пусть код характеризуется вероятностями  $P_1 = P_2 = P_3 = P_4 = P_5 = \frac{1}{5}$ . Код имеет среднюю длину

$$L_n = \sum_{i=1}^5 \frac{1}{5} l_i = \frac{1}{5} 2 + \frac{1}{5} 2 + \frac{1}{5} 3 + \frac{1}{5} 3 + \frac{1}{5} 3 = 2,6.$$

Возникает важная задача эффективного кодирования – построения кода с минимально возможными длинами кодовых слов для передачи информации или определения среднего числа бит для кодирования источника.

### 3.2.4. Средняя длина кодового слова и энтропия

Понятие энтропии источника как среднее количество информации, передаваемое одним символом источника, отражает связь величины энтропии с величиной средней длины слов эффективного кода. Это отражение выражается двумя соотношениями.

*Соотношение 1.* Неравенство длины кода.

Средняя длина  $L$  двоичного однозначно декодируемого кода удовлетворяет неравенству

$$L \geq H. \quad (3.4)$$

Выражение (3.4) определяет минимально достижимую среднюю длину эффективного кода.

Пример 3.8. Пусть используется префиксный код со словами:

$$\begin{aligned}A &\rightarrow (00); \\K &\rightarrow (10); \\N &\rightarrow (010); \\D &\rightarrow (110); \\E &\rightarrow (111).\end{aligned}$$

Вероятности символов источника характеризуются множеством  $\{P(A), \dots, P(E)\} \rightarrow \{p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, p_4 = \frac{1}{16}, p_5 = \frac{1}{16}\}$ . Энтропия источника равна:

$$H = \sum_{i=1}^5 -p_i \log_2 p_i = \frac{1}{2} \cdot 1 + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \frac{1}{16} \cdot 4 + \frac{1}{16} \cdot 4 = \\ = \frac{7}{4} = 1,875 \text{ бит/символ.}$$

Средняя длина кодового слова равна:

$$L = \sum_{i=1}^m p_i l_i = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 3 + \frac{1}{16} \cdot 3 = \frac{9}{4} = 2,25.$$

В рассмотренном примере

$$L = 2,25 > H = 1,875.$$

Средняя длина кодового слова удовлетворяет соотношению 1  
 $L \geq H$ .

*Соотношение 2.* Пределы средней длины кодового слова.

Можно получить кодирование источника двоичным кодом, у которого средняя длина кодового слова удовлетворяет выражению

$$H \leq L \leq H + 1. \quad (3.5)$$

При этом выражение

$$L \leq H + 1 \quad (3.6)$$

определяет максимально возможное значение средней длины двоичного эффективного кода. Если вернуться к примеру 2.8, для рассмотренного кода имеем

$$H = 1,875 \leq L = 2,25 \leq H + 1 = 1,875 + 1.$$

## 4. ТЕОРЕМА ШЕННОНА О КОДИРОВАНИИ ДЛЯ КАНАЛА БЕЗ ШУМА (первая теорема Шеннона)

Если в канале передачи информации вероятность ошибки  $p \rightarrow 0$ , основное требование к информационной системе – это представление символов источника в максимально компактной форме. Первая теорема Шеннона определяет минимально достижимую длину кодового слова на символ источника. Учитывая статистические свойства источника, можно более эффективно передавать (обрабатывать) информацию. Если высоковероятным символам поставить в соответствие более короткие длины кодовых слов, а маловероятным символам – слова большей длины, в этом случае достигается увеличение скорости передачи информации. Количество передаваемой информации за единицу времени также увеличится.

### 4.1. Энтропия блокового источника

Пусть имеется источник блоковых символов  $X^n$ . Эти блоки имеют длину  $n$ . Первая теорема Шеннона утверждает, что если  $n \rightarrow \infty$ , можно произвести кодирование блоковых символов источника  $X^n$  кодом, у которого средняя длина кодового слова будет приближаться к энтропии  $H$  источника  $X$ .

Если рассматривать блоки символов длиной  $n$  как новые независимые события с энтропией  $H' = nH$  (свойство 4 энтропии блокового источника) в соответствии с выражением (3.5) можно записать

$$H' \leq L \leq H' + 1, \quad (4.1)$$

где  $L$  – средняя длина слова источника  $X^n$ . Это же значение  $L$  определяет среднюю длину кодового слова последовательности  $n$  символов источника  $X$ .

Перепишем (4.1) как

$$nH \leq L \leq nH + 1.$$

где  $H$  – энтропия источника одиночных символов. Полученное выражение преобразуем к виду

$$H \leq \frac{L}{n} \leq H + \frac{1}{n}, \quad (4.2)$$

где  $\frac{L}{n}$  – средняя длина слова на один символ источника  $X$  (на одно событие).

Таким образом, расширяя источник одиночных символов, т.е. увеличивая значение  $n \rightarrow \infty$  – длину блока, можно закодировать символы блокового источника кодовыми словами со средней длиной на символ источника  $X$ , приближающейся к значению энтропии источника  $X$ .

### 4.2. Первая теорема Шеннона

Кодируя блоковые последовательности источника без памяти  $X = \{x_1, x_2, \dots, x_m\}$  с энтропией  $H$ , можно построить  $q$ -ичный префиксный код, в котором средняя длина кодового слова удовлетворяет выражению

$$\lim_{n \rightarrow \infty} \frac{L_n}{n} = H,$$

где  $L_n$  – средняя длина кодовых слов префиксного кода.

*Замечание.* Составляющая  $\frac{L_n}{n} = H + \frac{1}{n}$  выражения (4.2) определяет максимальное значение средней длины кодового слова префиксного кода.

*Вывод.* Эффективное кодирование информации требует использования источника с  $n$ -кратным расширением источника  $X$ .

Пример 4.1. Вычислим энтропию двоичного источника с символами алфавита  $X = \{a, b\}$  с вероятностью  $p_1 = \frac{7}{8}, p_2 = \frac{1}{8}$ .

Решение. Применяя формулу Шеннона (1.3), получаем

$$H = -p_1 \log_2 p_1 - p_2 \log_2 p_2 = -\frac{7}{8} \log_2 \frac{7}{8} - \frac{1}{8} \log_2 \frac{1}{8} =$$

$$= \frac{7}{8} \cdot 0,1926 + \frac{1}{8} \cdot 3 = 0,54 \text{ бит/символ.}$$

Источник дает менее 1 бита информации на символ.

Символы источника кодируются словами  $a \rightarrow 0, b \rightarrow 1$ . Длина кодовых слов равна  $l_1 = l_2 = 1$ . Средняя длина слова кода равна

$$L = \sum_{i=1}^2 p_i l_i = \frac{7}{8} \cdot 1 + \frac{1}{8} \cdot 1 = 1.$$

Пример 4.2. Для увеличения количества передаваемой информации, используем расширение источника одиночных символов примера 4.1. Построим блоковый источник  $X^2 = \{(aa), (ab), (ba), (bb)\} = \{c_1, c_2, c_3, c_4\}$ .

Решение. Вероятности появления символов с 2-кратным расширением источника одиночных символов равны:

$$P_1 = p_1 p_1 = \frac{49}{64},$$

$$P_2 = p_1 p_2 = P_3 = p_2 p_1 = \frac{7}{64},$$

$$P_4 = p_2 p_2 = \frac{1}{64}.$$

Для кодирования блокового источника применим префиксный код

$$\begin{aligned} (aa) &\rightarrow 0; \\ (ab) &\rightarrow 10; \\ (ba) &\rightarrow 110; \\ (bb) &\rightarrow 111. \end{aligned}$$



Средняя длина  $L_n$  двоичного однозначно декодируемого кода равна:

$$L_n = \sum_{i=1}^4 P_i l_i = \frac{49}{64} \cdot 1 + \frac{7}{64} \cdot 2 + \frac{7}{64} \cdot 3 + \frac{1}{64} \cdot 3 = \frac{87}{64}.$$

Средняя длина слова на один символ источника равна:

$$\frac{L_n}{n} = \frac{\frac{87}{64}}{2} = \frac{87}{128} = 0,68.$$

Как видно, средняя длина кода источника  $X = \{a, b\}$  уменьшилась с  $L = 1$  до

$$\frac{L_n}{n} = 0,68 > H = 0,54.$$

Энтропия блокового источника равна

$$H' = \sum_{i=1}^4 -P(c_i) \log_2 P(c_i) = \left(-\frac{49}{64} \log_2 \frac{49}{64}\right) + \left(-\frac{7}{64} \log_2 \frac{7}{64}\right) + \left(-\frac{7}{64} \log_2 \frac{7}{64}\right) + \left(-\frac{1}{64} \log_2 \frac{1}{64}\right) = 1,08 \frac{\text{бит}}{\text{символ}(H')}.$$

Заметим, энтропия блокового источника в  $n$  раз больше энтропии соответствующего источника одиночных символов

$$H' = n H = 2 \cdot 0,54 = 1,08 \frac{\text{бит}}{\text{символ}(H')}.$$

Пример 4.3. Источник формирует символы  $X = \{x_1, x_2\} = \{a, b\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ . Размерность алфавита  $m = 2$ . Энтропия источника равна

$$H = \sum_{i=1}^2 -p_i \log_2 p_i = \frac{1}{3} \cdot 1,585 + \frac{2}{3} \cdot 0,585 = 0,918 \text{ бит/символ}.$$

Символы источника кодируются словами  $a \rightarrow 0, b \rightarrow 1$ . Длина кодовых слов равна  $l_1 = l_2 = 1$ .

Пример 4.4. Для увеличения количества передаваемой информации, используем источник одиночных символов примера 4.3. Построим блоковый источник  $X^2 = \{(aa), (ab), (ba), (bb)\} = \{c_1, c_2, c_3, c_4\}$ . Выход этого источника принимает одно из состояний множества  $C = \{c_1, c_2, c_3, c_4\}$ .

1. Вычислим вероятности появления символов множества  $C = \{c_1, c_2, c_3, c_4\}$  как независимых событий источника  $X = \{x_1, x_2\}$  с вероятностями символов  $x_1 = a \rightarrow p_1 = \frac{1}{3}$  и  $x_2 = b \rightarrow p_2 = \frac{2}{3}$ .

$$P(c_1) = p_1 \cdot p_1 = \frac{1}{9}, P(c_2) = p_1 \cdot p_2 = \frac{2}{9},$$

$$P(c_3) = p_2 \cdot p_1 = \frac{2}{9}, P(c_4) = p_2 \cdot p_2 = \frac{4}{9}.$$

Для кодирования блокового источника применим следующий равномерный блоковый код:

$$\begin{aligned}(aa) &\rightarrow 00; \\ (ab) &\rightarrow 10; \\ (ba) &\rightarrow 01; \\ (bb) &\rightarrow 11.\end{aligned}$$

2. Средняя длина  $L_n$  двоичного однозначно декодируемого кода равна:

$$L_n = \sum_{i=1}^m P(c_i) l_i = \frac{1}{9} \cdot 2 + \frac{2}{9} \cdot 2 + \frac{2}{9} \cdot 2 + \frac{4}{9} \cdot 2 = 2.$$

3. Средняя длина на один символ источника равна:

$$\frac{L_n}{n} = \frac{2}{2} = 1.$$

Как видно, средняя длина кода источника  $X = \{a, b\}$  не уменьшилась

$$\frac{L_n}{n} = 1 > H = 0,918,$$

т.к. в этом примере использовалось равномерное кодирование источника.

4. Энтропия блокового источника равна

$$\begin{aligned}H' = \sum_{i=1}^4 -P(c_i) \log_2 P(c_i) &= \frac{1}{9} \cdot 0,585 + \frac{2}{9} \cdot 2,1699 + \frac{2}{9} \cdot 2,1699 + \\ &+ \frac{4}{9} \cdot 1,1699 = 1,83 \frac{\text{бит}}{\text{символ}(H')}.\end{aligned}$$

Энтропия блокового источника в 2 раза больше энтропии соответствующего источника одиночных символов.

$$H' = n H = 2 \cdot 0,918 = 1,83.$$

Следует отметить, что применение рассмотренного метода эффективно-го представления информации приводит к увеличению сложности технической реализации кодирования, увеличения сложности декодирования, увеличения времени на процесс декодирования.

#### Упражнения

4.1. Источник формирует символы  $X = \{x_1, x_2\} = \{a, b\}$  с вероятностями  $\{p_1 = \frac{9}{10}, p_2 = \frac{1}{10}\}$ . Имеется блоковый источник с двукратным расширением  $X^2 = \{(aa), (ab), (ba), (bb)\} = \{c_1, c_2, c_3, c_4\}$ . Для кодирования блокового источника применяется префиксный код:

$$\begin{aligned}c_1 &\rightarrow (0); \\ c_2 &\rightarrow (10); \\ c_3 &\rightarrow (110); \\ c_4 &\rightarrow (111).\end{aligned}$$

4.1.1. Вычислить энтропию источника.

4.1.2. Вычислить энтропию блокового источника.

4.1.3. Вычислить среднюю длину слова декодируемого кода.

4.1.4. Вычислить среднюю длину слова на один символ источника  $X$ .

4.2. Источник формирует символы  $X = \{x_1, x_2\}$  с вероятностями  $\{p_1 = \frac{9}{10}, p_2 = \frac{1}{10}\}$ . Имеется блочный источник с трехкратным расширением  $X^3 = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$ . Для кодирования блочного источника применяется префиксный код:

$c_1 \rightarrow (1);$   
 $c_2 \rightarrow (011);$   
 $c_3 \rightarrow (010);$   
 $c_4 \rightarrow (001);$   
 $c_5 \rightarrow (00011);$   
 $c_6 \rightarrow (00010);$   
 $c_7 \rightarrow (00001);$   
 $c_8 \rightarrow (00000).$

4.2.1. Вычислить энтропию источника.

4.2.2. Вычислить энтропию блочного источника.

4.2.3. Вычислить среднюю длину слова декодируемого кода.

4.2.4. Вычислить среднюю длину слова на один символ источника  $X$ .

### 4.3. Сжатие данных

В настоящее время большая часть передаваемых, хранимых, распределяемых, преобразуемых данных соответствует звуковой, графической или видеоинформации. В этом случае реализация современных инфокоммуникационных систем неизбежно усложняется. Увеличиваются технические затраты на хранение данных, предъявляются более высокие требования по экономии канального частотного ресурса. Сжатие данных позволяет уменьшить объем данных, используемых для представления информации. Даже при постоянном росте емкости хранения данных и пропускной способности каналов сжатие остается необходимым и существенным компонентом информационных технологий. Различают два основных метода кодирования данных используемых для сжатия: кодирование с потерями и кодирование без потерь.

Идея сжатия основывается на возможности устранения или уменьшения избыточности передаваемых (хранимых) данных. Сигнал, несущий информацию, можно сжать путем удаления из него имеющейся избыточности.

Различают два вида избыточности.

1. Статистическая избыточность, связанная с корреляцией и предсказуемостью обрабатываемых данных. Такая избыточность может быть полностью устранена без потери информации и исходные данные могут быть полностью восстановлены.

2. Видео-аудио (субъективная) избыточность, которую можно устранить с некоторой потерей информации, сравнительно мало влияющей на качество воспроизводимого изображения или звука.

Например, в обычных телевизионных каналах, передача видеoinформации осуществляется в полосе 6 МГц. Хотя видеоспектр имеет значительно большее значение, часть спектральных составляющих отбрасывается. В этом случае исходная видеoinформация не может быть полностью восстановлена.

Примером сжатия с потерями, когда отбрасывается несущественная информация, может служить система цифрового покомпонентного телевидения (Digital Video Broadcasting, DVB). В покомпонентном телевидении разделённые сигналы яркости  $Y$  и цветоразностные сигналы  $R - Y$  и  $B - Y$  квантуются на 256 уровней. Длина кодового слова, соответствующего каждому уровню яркости равна  $n = 8$ . При дискретизации телевизионного сигнала с частотой Найквиста-Котельникова ширина полосы частот цифрового полного телевизионного сигнала составляет величину

$$W = f_{d_Y} n + f_{d_R} n + f_{d_B} n = 13,5 \cdot 8 + 6,75 \cdot 8 + 6,75 \cdot 8 = 216 \text{ МГц},$$

где частота дискретизации яркостного канала  $f_{d_Y} = 13,5$  МГц, частота дискретизации цветоразностных сигналов  $f_{d_R} = f_{d_B} = 6,75$  МГц.

Скорость передачи информации достигает значения  $R = 216 \frac{\text{Мбит}}{\text{с}}$ . В этом случае возникает цифровой поток объемом  $216 \frac{\text{Мбит}}{\text{с}}$ .

Телевидение высокой четности для формата HDTV (High Definition Television) имеет примерно удвоенную разрешающую способность по горизонтали и как минимум, удвоенную разрешающую способность по вертикали и соответствующее увеличение объема цифрового информационного потока до значения  $\approx 576 \frac{\text{Мбит}}{\text{с}}$ . Стандартный телевизионный каналный частотный ресурс находится в диапазоне 48 – 862 МГц. Очевидно, без эффективного сжатия невозможно реализовать технологию многоканальной передачи информации в формате HDTV. В этой технологии сжатие (кодирование) видеосигнала реализуется на основе стандарта MPEG (Motion Pictures Experts Group – разработан Экспертной группой по вопросам движущихся изображений). Кодирование основано на удалении, невоспринимаемой органами зрения, части видеосигнала (отдельных спектральных составляющих).

Технология сжатия может осуществляться также за счет применения эффективного кодирования источников. В этом случае под избыточностью понимают частое повторение символов информационного сообщения, повторяемость слов, предложений самих сообщений. Алгоритмы сжатия получаются за счет использования кодирования данных без потерь. Примером такого алгоритма является эффективное кодирование источника с помощью кода Хаффмена.

Эффективность сжатия оценивается коэффициентом

$$K = \frac{N-M}{N}, \quad (4.3)$$

где  $N$  обозначает затраты на передачу (хранение) данных без сжатия,  $M$  – затраты на передачу (хранение) данных со сжатием.

Например, для передачи яркостей всех пикселей фрагмента изображения размером  $8 \times 8$  требуется  $N = 2^9 = 512$ . Здесь значение яркости каждого пикселя кодируется двоичным кодом длиной 8. Пусть с использованием метода сжатия по стандарту *MPEG-2* получено  $M = 128$ . Тогда эффективность сжатия

$$K = \frac{512-128}{512} = 0,75,$$

что соответствует 75%.

Преобразуем формулу (4.3) следующим образом:

$$\begin{aligned} KN &= N - M, \\ N &= KN + M, \end{aligned}$$

$$\frac{N}{M} = \frac{KN + M}{M}.$$

В этом случае отношение вида  $\frac{N}{M}$  характеризует выигрыш в записи данных. Для рассматриваемого примера выигрыш равен

$$\frac{N}{M} = \frac{KN + M}{M} = \frac{0,75 \cdot 512 + 128}{128} = 4.$$

**Пример 4.5.** Оценить эффективность сжатия стереофонического аудио усовершенствованной системы кодирования звука *MPEG-2 AAC* (Advanced Audio Coding; стандарт разработан в 1998 году в Институте интегральных схем Фраунгофера – *IS-A*, Германия) со скоростью 112 Кбит/с. В качестве сравнения рассмотрим кодирование по стандарту стереофонического аудио компакт дисков (*CD-Audio*), где сжатия практически нет. Стандартная частота дискретизации  $f_d$  непрерывного аудио сигнала для *CD*  $f_d = 44,1$  КГц. Длина кодовых слов  $n = 16$ .

**Решение.** Скорость передачи стереоданных аудио- *CD* определяется как

$$R = 2 \cdot f_d \cdot n = 2 \cdot 44,1 \cdot 16 = 1411,2 \text{ Кбит/сек.}$$

Эффективность сжатия

$$K = \frac{1411,2 - 112}{1411,2} \cong 0,92.$$

Выигрыш в записи данных

$$\frac{N}{M} = \frac{1411,2}{112} = 12,6.$$

*Замечание.* Сравнительно высокая степень сжатия методом кодирования *AAC* отражается на качественных характеристиках воспроизводимого аудио сигнала.

Одним из первых методов сжатия без потерь является метод Шеннона-Фано.

### 4.3.1. Коды Шеннона-Фано

Код по своему построению удовлетворяет свойству префикса. Средняя длина кодов моментального кода Шеннона-Фано приближается к границе, определяемой энтропией. Однако, этот метод не всегда дает оптимальное пространство кодовых слов. Алгоритм необязательно минимизирует среднюю длину слова.

### 4.3.2. Энтропийное кодирование методом Хаффмена

Ранее было показано, чем более неравномерно распределены вероятности появления символов источника, тем меньше энтропия. Так, в примере 4.1 для  $X = \{a, b\}$ ,  $p_1 = \frac{7}{8}$ ,  $p_2 = \frac{1}{8}$ , было получено  $H = 0,54$  бит/символ. Для другого источника (пример 4.3,  $X = \{x_1, x_2\} = \{a, b\}$ ,  $p_1 = \frac{1}{3}$ ,  $p_2 = \frac{2}{3}$ .) энтропия источника  $H = 0,918$  бит/символ. С уменьшением неопределенности появления случайного события энтропия уменьшается.

Из первой теоремы Шеннона средняя длина кодового слова источника находится в диапазоне

$$H \leq L \leq H + 1.$$

Так как энтропия – это мера количества информации, то чем меньше энтропия, тем эффективнее будет кодирование неравномерными словами. Такой метод эффективного кодирования называется энтропийным. Один из основных методов энтропийного кодирования известен как кодирование Хаффмена (D.A. Huffman, амер. ученый). В 1952 году Хаффмен показал, что, разработанный им алгоритм эффективного кодирования позволяет строить класс оптимальных префиксных кодов без запятой, т.е. неравномерных кодов,  $n = \text{var}$ . Эти коды позволяют кодировать символы источника с минимальной избыточностью. Кодирование Хаффмена формирует оптимальный код для дискретных источников без памяти. Средняя длина  $L$  кодового слова кода Хаффмена приближается к энтропии источника  $H$ . В настоящее время коды Хаффмена используются при цифровой обработке изображений и звуков. Они составляют основную часть стандартов сжатия изображений и звука MPEG, H.264. JPEG (Joint Photographic Experts Group – разработан Объединенной группой экспертов по обработке фотографических изображений).

### 4.3.3. Алгоритм Хаффмена

Алгоритм включает в себя выполнение ряда итерационных действий.

1. Упорядочение. Расставить символы источника в порядке уменьшения их вероятностей.

2. Редукция. Объединить два символа с наименьшими вероятностями в один символ.

3. Переупорядочение. Расставить символы в порядке уменьшения их вероятностей.

4. Продолжить процессы 2 и 3 до тех пор, пока все символы не будут объединены. В случае, когда несколько символов имеют одинаковые вероятности, объединяются те из них, которые имели до этого меньшее число объединений.

5. Кодирование. Начать с последнего объединения. Приписать первой компоненте составного символа значение 0, а второй компоненте значение 1. Продолжать этот процесс до тех пор, пока все символы не будут закодированы.

Пример 4.6. Источник формирует следующие символы  $X = \{x_1, x_2, \dots, x_6\} = \{A, K, N, D, E, !\}$ . Вероятности символов задаются множеством:

$\{p_1 = 0,05, p_2 = 0,15, p_3 = 0,05, p_4 = 0,4, p_5 = 0,2, p_6 = 0,15\}$ .

Построить код Хаффмена.

Решение. Алгоритм реализуется по схеме, приведенной на рис. 4.1. Упорядочение.

<i>D</i>	<i>E</i>	<i>K</i>	!	<i>A</i>	<i>N</i>
0,4	0,2	0,15	0,15	0,05	0,05

Редукция.

<i>D</i>	<i>E</i>	<i>K</i>	!	<i>A N</i>
0,4	0,2	0,15	0,15	0,1

Редукция.

<i>D</i>	<i>E</i>	<i>K</i>	! <i>A N</i>
0,4	0,2	0,15	0,25

Упорядочение.

<i>D</i>	! <i>A N</i>	<i>E</i>	<i>K</i>
0,4	0,25	0,2	0,15

Редукция.

<i>D</i>	! <i>A N</i>	<i>E K</i>
0,4	0,25	0,35

Упорядочение.

<i>D</i>	<i>E K</i>	! <i>A N</i>
0,4	0,35	0,25

Редукция.

<i>D</i>	<i>E K</i> ! <i>A N</i>
0,4	0,6

Упорядочение.

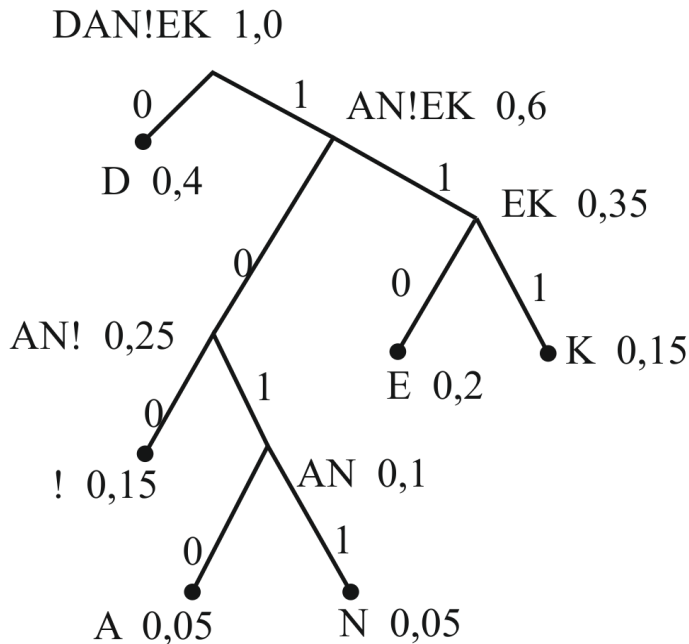
<i>E K</i> ! <i>A N</i>	<i>D</i>
0,6	0,4

Редукция.

<i>DEK</i> ! <i>A N D</i>
---------------------------

Рис. 4.1. Последовательность шагов алгоритма Хаффмена

Далее изобразим кодовое дерево, как показано на рис. 4.2.

Рис. 4.2. Кодовое дерево источника  $X = \{A, K, N, D, E, !\}$ 

Соответственно, кодовые слова кода Хаффмена имеют вид:

$A \rightarrow 1010,$   
 $N \rightarrow 1011,$   
 $! \rightarrow 100,$   
 $E \rightarrow 110,$   
 $K \rightarrow 111,$   
 $D \rightarrow 0.$

Средняя длина полученного двоичного однозначно декодируемого кода равна

$$L_n = \sum_{i=1}^6 p_i l_i = 0,05 \cdot 4 + 0,15 \cdot 3 + 0,05 \cdot 4 + 0,4 \cdot 1 + 0,2 \cdot 3 + 0,15 \cdot 3 = 2,3.$$

Энтропия источника равна

$$H = \sum_{i=1}^6 -p_i \log_2 p_i = (-0,05 \log_2 0,05) + (-0,15 \log_2 0,15) + (-0,05 \log_2 0,05) + (0,4 \log_2 0,4) + (0,2 \log_2 0,2) + (-0,15 \log_2 0,15) = 2,25 \frac{\text{бит}}{\text{символ}}.$$



Сравнивая полученные значения  $L_n$  и  $H$ , видно что код Хаффмена является эффективным. В этом случае его средняя длина удовлетворяет выражению (3.5)

$$H \leq L \leq H + 1, \\ 2,25 < 2,3 < 3,25.$$

#### 4.3.4. Эффективность кода

*Определение 4.1.* Эффективностью кода или фактором сжатия называется отношение энтропии к средней длине кода

$$\eta = \frac{H}{L_n}.$$

В примере 4.6 эффективность кода равна

$$\eta = \frac{2,25}{2,3} = 0,976.$$

#### 4.3.5. Коды Хаффмена блоковых источников

Пусть источник формирует одиночные символы  $X = \{a, b\} = \{0, 1\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ . Энтропия источника равна  $H = 0,918$  бит/символ (см. пример 4.3). Кодирование этого источника кодом Хаффмена приводит к двум словам с минимальной длиной  $L = 1$ .

1. Имеется блоковый источник с двукратным расширением  $X^2 = \{(00), (01), (10), (11)\} = \{c_1, c_2, c_3, c_4\}$ . Символы  $c_1, c_2, c_3, c_4$  получены расширением источника одиночных символов  $X = \{a, b\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ . Вероятности  $P(c_i)$  появления символов источника  $X^2$  определяются множеством

$$\{P(c_1) = \frac{1}{9}, P(c_2) = P(c_3) = \frac{2}{9}, P(c_4) = \frac{4}{9}\}.$$

Энтропия источника  $X^2$

$$H^2 = \sum_{i=1}^4 -P(c_i) \log_2 P(c_i) = 1,83 \text{ бит/символ (см. пример 1.8).}$$

Кодирование блокового источника  $X^2$  реализуем на основе алгоритма Хаффмена. Структуре алгоритма соответствует последовательность шагов показанная на рис. 4.3.

Упорядочение.

$c_4$	$c_2$	$c_3$	$c_1$
4	2	2	1
$\frac{4}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Редукция.

$c_4$	$c_2$	$c_1 c_3$
4	2	3
$\frac{4}{9}$	$\frac{2}{9}$	$\frac{3}{9}$

Упорядочение.

$c_4$	$c_1 c_3$	$c_2$
4	3	2
$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{9}$

Редукция.

$c_4$	$c_1 c_3 c_2$
4	5
$\frac{4}{9}$	$\frac{5}{9}$

Упорядочение.

$c_1 c_3 c_2$	$c_4$
5	4
$\frac{5}{9}$	$\frac{4}{9}$

Редукция.

$c_1 c_3 c_2 c_4$
1

Рис. 4.3. Последовательность шагов алгоритма Хаффмена

Далее изобразим кодовое дерево кода Хаффмена, рис. 4.4.

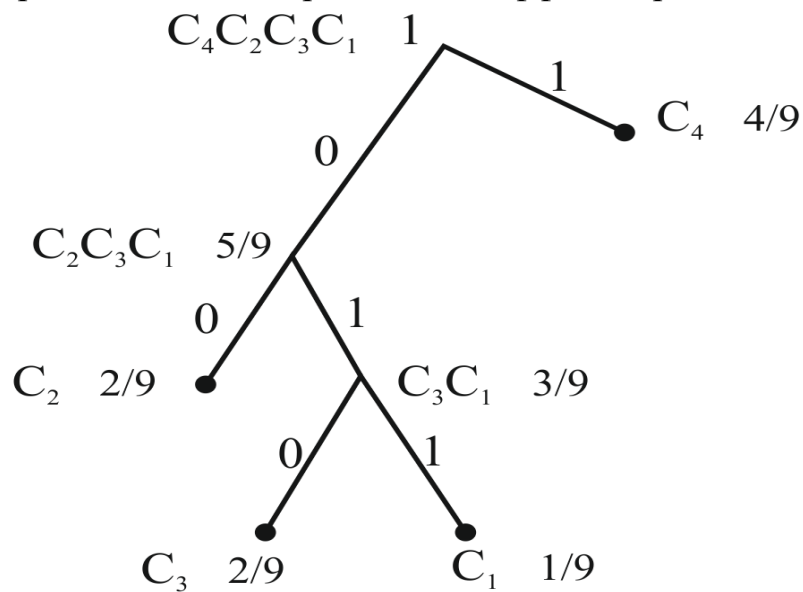


Рис. 4.4. Кодовое дерево

Соответственно, кодовые слова имеют вид:

$$c_1 \rightarrow (011),$$

$$c_2 \rightarrow (00),$$

$$c_3 \rightarrow (010),$$

$$c_4 \rightarrow (1).$$

Средняя длина полученного двоичного кода Хаффмена равна

$$L_n = \sum_{i=1}^4 P(c_i) l_i = \frac{1}{9} \cdot 3 + \frac{2}{9} \cdot 2 + \frac{2}{9} \cdot 3 + \frac{4}{9} \cdot 1 = \frac{17}{9} \cong 1,88.$$

Средняя длина слова на один символ источника равна:

$$\frac{L_n}{n} = \frac{17}{9 \cdot 2} = \frac{17}{18} \cong 0,944.$$

Как видно, средняя длина кода источника  $X = \{a, b\}$  уменьшилась с  $L = 1$  до  $\frac{L_n}{n} \cong 0,94 > H = 0,918$ .

2. Имеется блочный источник с трехкратным расширением

$$X^3 = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\} = \{(000), (001), (010), (011), (100), (101), (110), (111)\}.$$

Символы блочного источника  $X^3$  получены расширением источника одиночных символов  $X = \{0,1\}$  с вероятностями  $\{p_1 = \frac{1}{3}, p_2 = \frac{2}{3}\}$ . Вероятности  $P(c_i)$  появления символов источника  $X^3$  определяются как

$$P(c_1) = p_1 \cdot p_1 \cdot p_1 = \frac{1}{27},$$

$$P(c_2) = p_1 \cdot p_1 \cdot p_2 = \frac{2}{27},$$

$$P(c_3) = p_1 \cdot p_2 \cdot p_1 = \frac{2}{27},$$

$$P(c_4) = p_1 \cdot p_2 \cdot p_2 = \frac{4}{27},$$

$$P(c_5) = p_2 \cdot p_1 \cdot p_1 = \frac{2}{27},$$

$$P(c_6) = p_2 \cdot p_1 \cdot p_2 = \frac{4}{27},$$

$$P(c_7) = p_2 \cdot p_2 \cdot p_1 = \frac{4}{27},$$

$$P(c_8) = p_2 \cdot p_2 \cdot p_2 = \frac{8}{27}.$$

Энтропия источника  $X^3$

$$H^3 = nH = 3 \cdot 0,918 = 2,754.$$

Кодирование блочного источника  $X^3$  реализуем на основе алгоритма Хаффмена. Структуре алгоритма соответствует последовательность шагов, показанная на рис. 4.6.

Исходные данные.

$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$
$\frac{1}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{8}{27}$

Упорядочение.

$c_8$	$c_7$	$c_6$	$c_4$	$c_5$	$c_3$	$c_2$	$c_1$
$\frac{8}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{4}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$

Редукция.

$c_8$	$c_7$	$c_6$	$c_4$	$c_5$	$c_3$	$c_2c_1$
8	4	4	4	2	2	3
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Упорядочение.

$c_8$	$c_7$	$c_6$	$c_4$	$c_2c_1$	$c_5$	$c_3$
8	4	4	4	3	2	2
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Редукция.

$c_8$	$c_7$	$c_6$	$c_4$	$c_2c_1$	$c_5c_3$
8	4	4	4	3	4
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Упорядочение.

$c_8$	$c_7$	$c_6$	$c_4$	$c_5c_3$	$c_2c_1$
8	4	4	4	4	3
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Редукция.

$c_8$	$c_7$	$c_6$	$c_4$	$c_5c_3c_2c_1$
8	4	4	4	7
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Упорядочение.

$c_8$	$c_5c_3c_2c_1$	$c_7$	$c_6$	$c_4$
8	7	4	4	4
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Редукция

$c_8$	$c_6c_4$	$c_5c_3c_2c_1$	$c_7$
8	8	7	4
$\overline{27}$	$\overline{27}$	$\overline{27}$	$\overline{27}$

Редукция.

$c_8$	$c_6c_4$	$c_7c_5c_3c_2c_1$
8	8	11
$\overline{27}$	$\overline{27}$	$\overline{27}$

Упорядочение.

$c_7c_5c_3c_2c_1$	$c_8$	$c_6c_4$
11	8	8
$\overline{27}$	$\overline{27}$	$\overline{27}$

Редукция.

$c_7c_5c_3c_2c_1$	$c_8c_6c_4$
$\frac{11}{27}$	$\frac{16}{27}$

Упорядочение.

$c_8c_6c_4$	$c_7c_5c_3c_2c_1$
$\frac{16}{27}$	$\frac{11}{27}$

Редукция.

$c_8c_6c_4c_7c_5c_3c_2c_1$
1

Рис. 4.6. Последовательность шагов алгоритма Хаффмена

Далее представим код Хаффмена на кодовом дереве, как показано на рис. 4.7.

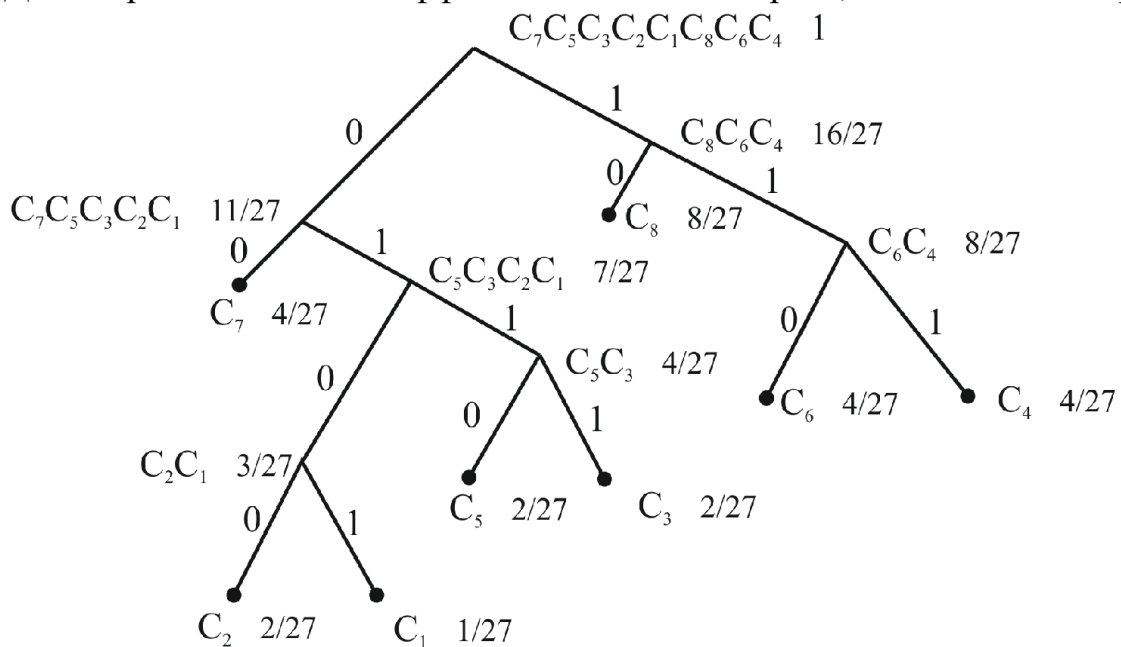


Рис. 4.7. Кодовое дерево

Соответственно, кодовые слова имеют вид:

- $c_1 \rightarrow (0101),$
- $c_2 \rightarrow (0100),$
- $c_3 \rightarrow (0111),$
- $c_4 \rightarrow (111),$
- $c_5 \rightarrow (0110),$
- $c_6 \rightarrow (110),$
- $c_7 \rightarrow (00),$
- $c_8 \rightarrow (10).$

Средняя длина полученного двоичного кода Хаффмена равна

$$L_n = \sum_{i=1}^8 P(c_i) l_i = \frac{1}{27} \cdot 4 + \frac{2}{27} \cdot 4 + \frac{2}{27} \cdot 4 + \frac{4}{27} \cdot 3 + \frac{2}{27} \cdot 4 + \frac{4}{27} \cdot 3 + \frac{4}{27} \cdot 2 + \frac{8}{27} \cdot 2 = \frac{76}{27} \cong 2,81.$$

Средняя длина слова на один символ источника равна:

$$\frac{L_n}{n} = \frac{76}{27 \cdot 3} = \frac{76}{81} \cong 0,938.$$

Как видно, средняя длина кода источника  $X = \{0,1\}$  уменьшилась с  $L = 1$  до  $\frac{L_n}{n} \cong 0,938 > H = 0,918$ .

Результаты, полученные для блочных источников удобно анализировать, рассматривая данные табл. 4.1.

Табл. 4.1

Источник оди- ночных символов $X$	Блочный источник $X^2$	Блочный источник $X^3$
Символы $c_1 \rightarrow (0)$ $c_2 \rightarrow (1)$  Энтропия источника $H = 0,918$ бит/символ	Код Хаффмена $c_1 \rightarrow (110)$ $c_2 \rightarrow (10)$ $c_3 \rightarrow (111)$ $c_4 \rightarrow (0)$	Код Хаффмена $c_1 \rightarrow (0100)$ $c_2 \rightarrow (0101)$ $c_3 \rightarrow (0111)$ $c_4 \rightarrow (111)$ $c_5 \rightarrow (0110)$ $c_6 \rightarrow (110)$ $c_7 \rightarrow (00)$ $c_8 \rightarrow (10)$ .
Средняя длина кода на один символ источника $X$ 1	Средняя длина кода на один символ источника $X^2$ $\cong 0,944$	Средняя длина кода на один символ источника $X^3$ $\cong 0,938$
Эффективность кода $(\eta = \frac{H}{L_n})$ $\eta = \frac{0,918}{1} = 0,918$	Эффективность кода $\eta = \frac{1,836}{1,88} \cong 0,944$	Эффективность кода $\eta = \frac{2,754}{2,81} \cong 0,980$

#### Выводы

1. С увеличением степени блочного источника значение средней длины (на один символ блочного источника) слова кода Хаффмена уменьшается и стремится к энтропии одиночного источника.

2. С увеличением степени блочного источника увеличивается эффективность кода или сжатие источника.

3. Недостатком алгоритма Хаффмена является требование априорного знания значений вероятностей появления символов, или оценок этих вероятностей.

#### 4.3.6. Декодирование кода Хаффмена

Декодирование входного слова начинается с начальной точки отсчета дерева (корня). Если входному символу соответствует значение 1, следует двигаться по ветви с присвоенным значением 1. Если принимается 0, следует идти по ветви, соответствующей значению 0. При попадании сразу в конечный узел дерева принимается решение о принятом символе. При попадании в узел, из которого выходят две ветви, следующий принятый символ (0 или 1) указывает, по какой ветви следует двигаться. Движение по дереву продолжается до достижения конечного узла. Принимается решение о символе, соответствующем данному кодовому слову. При приеме символа следующего кодового слова процесс декодирования вновь начинается с корня дерева и т.д. Для наглядности процесса декодирования изобразим кодовое дерево декодера источника  $X = \{A, K, N, D, E, !\}$  (см. пример 4.6).

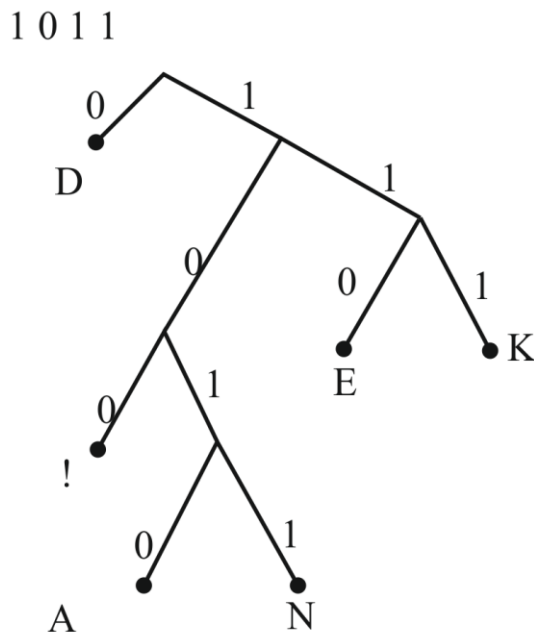


Рис. 4.8. Кодовое дерево декодера источника  $X = \{A, K, N, D, E, !\}$

Очевидно, декодирование кодовой последовательности (1011) по этому дереву приводит к символу  $N$ .

*Замечание.* Передача информации посредством эффективного кодирования требует использования каналов, характеризующихся повышенной надежностью. Вероятность ошибки в таком канале должна быть сравнительно малой. Даже одиночная ошибка в потоке кодированной информации приводит к неправильному декодированию сообщения источника. Рассмотрим это замечание на примере.

Пример 4.7. Пусть для передачи сообщения "DANKЕ!" используется код Хаффмена, полученный в примере 4.6. Код состоит из следующих слов:

$A \rightarrow 1010,$

$N \rightarrow 1011,$   
 $! \rightarrow 100,$   
 $E \rightarrow 110,$   
 $K \rightarrow 111,$   
 $D \rightarrow 0.$

Сообщению "DАНKE" соответствует поток двоичных символов

$u = 010101011111110100.$

Пусть при передаче этого потока, из-за воздействия помехи в канале, возникла одиночная ошибка в 4-м двоичном символе. На вход декодера поступила последовательность

$y = 010001011111110100.$

0 1 0 0 0 1 0 1 1 1 1 1 1 1 0 1 0 0

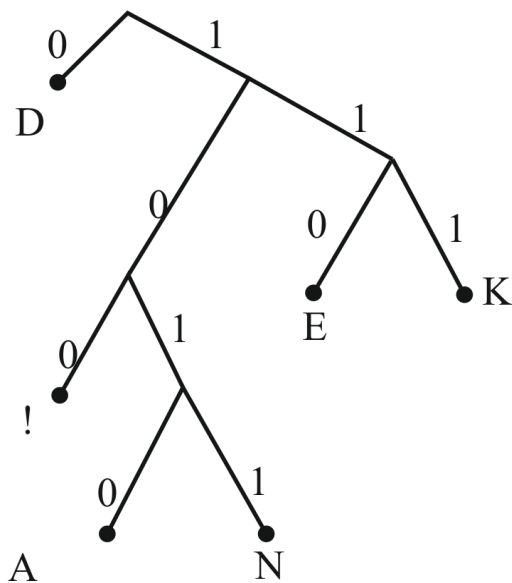


Рис. 4.9. Декодирование двоичного потока символов

Декодирование входной последовательности  $y = 010001011111110100$  с использованием кодового дерева приводит к формированию ошибочного сообщения «D!DNKE!». Легко убедиться, что структура кодового дерева отвечает следующему последовательному соответствию:  $0 \rightarrow D$ ,  $100 \rightarrow !$ ,  $0 \rightarrow D$ ,  $1011 \rightarrow N$ ,  $111 \rightarrow K$ ,  $110 \rightarrow E$ ,  $100 \rightarrow !$ .

Как видно, один бит, принятый с ошибкой, приводит к тому, что часть символов или все последующие символы будут декодированы неправильно.

Очевидными недостатками энтропийного метода кодирования являются:

- необходимо априорное знание вероятностных характеристик (статистик) символов источника;



– сжатие данных снижает избыточность и поэтому понижает надежность передачи информации. Вероятность ошибочного приема информации увеличивается.

#### 4.3.7. Адаптивный алгоритм Хаффмена

Адаптивный алгоритм эффективного кодирования реализуется на основе двух операций.

1. Вначале выполняется кодирование источника в предположении, что все символы имеют равные вероятности появления.

2. По мере накопления знаний о статистических характеристиках источника выполняется кодирование по алгоритму Хаффмена.

##### Упражнения

4.3.1. Размерность алфавита источника  $X = \{x_1, x_2, \dots, x_6\}$  равна  $m = 6$ . Построить дерево Хаффмена для следующих значений вероятностей появления символов источника:

$p_1 = 0,05; p_2 = 0,15; p_3 = 0,05; p_4 = 0,4; p_5 = 0,2; p_6 = 0,15$ .

4.3.2. Записать слова кода Хаффмена.

4.3.3. Декодировать последовательность  $X = 1110111011110101101$ , используя полученный код Хаффмена.

4.4.1. Источник символов  $\{A, K, N, D, E\}$ , характеризуется следующими вероятностями:

$p_A = \frac{1}{5}; p_K = \frac{1}{5}; p_N = \frac{1}{5}; p_D = \frac{1}{5}; p_E = \frac{1}{5}$ . Построить дерево Хаффмена

4.4.2. Записать слова кода Хаффмена.

4.5.1. Источник формирует символы  $X = \{x_1, x_2\}$  с вероятностями  $\{p_1 = \frac{9}{10}, p_2 = \frac{1}{10}\}$ . Имеется блочный источник с трехкратным расширением  $X^3 = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$ . Построить дерево Хаффмена блочного источника.

4.5.2. Записать слова кода Хаффмена.

4.6.1. Источник формирует символы алфавита  $X = \{x_1, x_2, \dots, x_8\} = \{A, K, N, D, E, !, *, J\}$  с вероятностями их появления

$\{p_1 = 0,5, p_2 = 0,1, p_3 = 0,1, p_4 = 0,1, p_5 = 0,06, p_6 = 0,04, p_7 = 0,05, p_8 = 0,05\}$ .

Построить кодовое дерево Хаффмена.

4.6.2. Запишите код Хаффмена.

4.7.1. Для передачи звука используется 8 уровней квантования. Распределению уровней отвечает гистограмма со следующими значениями:

$\{p_1 = 0,3; p_2 = 0,23; p_3 = 0,15; p_4 = 0,08; p_5 = p_6 = p_7 = p_8 = 0,06\}$ .

Построить кодовое дерево Хаффмена.

4.7.2. Записать слова кода Хаффмена.

4.7.3. Вычислить среднюю длину кода, кодирующего звук.

4.7.4. Вычислить энтропию источника звука.

4.8.1. Для передачи изображения используется 6 уровней квантования. Распределению уровней отвечает гистограмма со следующими значениями:  
 $\{p_1 = 0,27; p_2 = 0,21; p_3 = 0,23; p_4 = 0,15;$   
 $p_5 = p_6 = 0,07\}$ .

Построить кодовое дерево Хаффмена.

4.8.2. Записать слова кода Хаффмена.

4.8.3. Вычислить среднюю длину кода, кодирующего изображение.

4.8.4. Вычислить энтропию источника изображения.

#### **4.4. Универсальный алгоритм сжатия**

Реализация этого алгоритма не требует предварительных знаний статистики символов источника и по сути является адаптивным. Универсальный алгоритм сжатия информации основан на идее использования словаря последовательностей символов, слов, фраз и пр., или по-другому – образцов, встречающихся в несжатых данных. Универсальный метод сжатия присваивает кодовое слово образцам последовательных символов, встречающихся в текстах, изображениях, данных. При этом кодовое слово кода представляется определенным словарным индексом. В процессе сжатия при определении этих последовательностей (образцов) в несжатых данных они заменяются кодом, который ссылается на такую же последовательность в словаре. Чем больше объем словаря из этих последовательных символов, тем чаще они встречаются в несжатых данных, тем больше выигрыш в сжатии. Универсальный метод сжатия эффективен при архивации текстовой информации при обработке изображений, звуков и др.

##### **4.4.1. Алгоритм эффективного кодирования Лемпеля – Зива**

Метод сжатия на основе словаря был разработан А. Лемпелем (Abraham Lempel) Я. Зивом (Jacob Ziv) в 1977 году и известен как LZ77. Метод LZ77 является основой алгоритмов сжатия ZIP, ARJ, gzip и др. применяемых в компьютерах, используется в растровом файловом формате сжатия изображений PNG (Portable Network Graphic).

Многие реальные источники информационных символов характеризуются тем, что не всегда символы удовлетворяют свойству независимости. Например, в любом языке вероятность появления той или иной буквы зависит от предыдущей буквы. В этом случае говорят о межсимвольной зависимости, коррелированности. Кроме того, практически в любом тексте слова, фразы повторяются. Можно сказать, что текст состоит из образцов, которые образуют некоторый словарь. Кодирование текста можно свести к некоторому алгоритму выбора образцов из словаря.

Процесс кодирования методом Лемпеля – Зива начинается сразу после поступления выходных символов источника на вход кодера. Кодирование информации производится по следующему алгоритму.

1. Передающая сторона записывает в специальный буфер поиска то, что было уже отправлено (символ, последовательность символов, слово, фразу и пр.). Принимающая сторона также записывает то, что было уже получено для осуществления декодирования.

2. При подготовке следующего фрагмента текста передающая сторона находит в ранее переданном фрагменте образцы.

3. Далее идет процесс передачи не самих образцов, а только информации об этих образцах в виде ссылок.

4. Ссылка записывается в форме трех указателей  $(x, y, z)$ :

–  $x$  указывает относительный адрес образца в буфере поиска. Адрес определяется числом позиций символов в обратном направлении в буфере, где начинается образец;

–  $y$  обозначает длину образца – число совпадающих символов образца и символов несжатых данных;

–  $z$  обозначает следующую букву в буфере, которая отличается от продолжения фразы в словаре образцов.

Например, ссылке  $(7, 4, A)$  соответствует новый текст, состоящий из 4 букв образца, который начинается с 7-й буквы в обратном направлении буфера, и что в новом тексте следующая за образцом идет буква А.

Таким образом, кодирование осуществляется посредством использования ссылок  $(x, y, z)$ . Закодированная информация представляется последовательностью этих ссылок.

Пример 4.8. Необходимо передать сообщение: ДЕКОДИРОВАНИЕ КОДА с помощью алгоритма кодирования LZ77.

1. Процесс передачи сообщения начинается с кодирования первой буквы сообщения.

Сообщение

Д.

Кодовое слово выглядит как  $Д \rightarrow (0, 0, Д)$ . Символ Д еще не содержится в буфере поиска (словаре образцов), поэтому:

адрес  $x \rightarrow 0$ ;

длина образца  $y \rightarrow 0$ ;

следующая буква  $z \rightarrow Д$ .

Если символ не содержится в словаре образцов, он кодируется словом вида  $(0, 0, \text{символ})$ . Такое слово называется «нулевой фразой». При декодировании оно распознается по двум нулям.

2. Буфер поиска (словарь образцов) еще пуст.

Буфер поиска

–.

3. Далее идет кодирование и передача следующей буквы Е.

Формируется вновь слово «нулевая фраза»

$(0, 0, E)$ .

4. В буфере поиска уже записана буква Д.

Буфер поиска

Д.

5. Далее идет кодирование словами «нулевая фраза» других символов: К, О).

6. Далее передается ссылка (4, 1, И), которая указывает, что уже передавалась буква Д, а также показывает следующую за ней букву И. Число 4 ссылки соответствует числу позиций символов в обратном направлении в буфере поиска. Фактически эта ссылка (кодированное слово) соответствует передаче двух букв. Последующие шаги алгоритма иллюстрируются в табл. 4.2.

Таблица. 4.2.

N	Буфер поиска	Буфер для предварительной записи сообщения	Кодовое слово (x y z)
1	–	ДЕКОДИРОВАНИЕ_КОДА	(0, 0, Д)
2	Д	ЕКОДИРОВАНИЕ_КОДА	(0, 0, Е)
3	ДЕ	КОДИРОВАНИЕ_КОДА	(0, 0, К)
4	ДЕК	ОДИРОВАНИЕ_КОДА	(0, 0, О)
5	ДЕКО	ДИРОВАНИЕ_КОДА	(4, 1, И)

Число 4 кодового слова (4, 1, И) соответствует позиции в обратном направлении от текущего передаваемого символа «Д» до записанного уже ранее в буфер поиска этого символа фразы ДЕКО. Символ И (указатель z в слове (4, 1, И)) – это следующая буква кодируемого сообщения. Последующий процесс кодирования показан в продолжении таблицы.

Продолжение таблицы 4.2.

6	ДЕКОДИ	РОВАНИЕ_КОДА	(0, 0, Р)
7	ДЕКОДИР	ОВАНИЕ_КОДА	(4, 1, В)
8	ДЕКОДИРОВ	АНИЕ_КОДА	(0, 0, А)
9	ДЕКОДИРОВА	НИЕ_КОДА	(0, 0, Н)
10	ДЕКОДИРОВАН	ИЕ_КОДА	(6, 1, Е)
11	ДЕКОДИРОВАНИЕ	–	(0, 0, _)
12	ДЕКОДИРОВАНИЕ_	КОДА	(12, 3, А)
	ДЕКОДИРОВАНИЕ_КОДА		

Таким образом, передаваемому сообщению соответствует последовательность слов:

{(0, 0, Д) (0, 0, Е) (0, 0, К) (0, 0, О) (4, 1, И) (0, 0, Р) (4, 1, В) (0, 0, А) (0, 0, Н) (6, 1, Е) (0, 0, \_) (12, 3, А)}.

Метод кодирования LZ приводит к сжатию тогда, когда затраты на кодирование оказываются в среднем меньше по сравнению с кодированием кодом ASCII.

*Замечание.* ASCII-код (American Standard Code for Information Interchange – Американский стандартный код для обмена информацией) – это давно установленный стандарт кодирования букв, чисел, символов для пред-

ставления их в цифровой форме. Каждый текстовый символ обозначается числом от 32 до 127. Используется семь битов для идентификации символа и один бит добавляется таким образом, чтобы количество единиц в кодовом слове было четным. Получается блоковый равномерный код с контролем четности длиной 8 бит (байт). Например, символу DEL соответствует кодовое слово (1 1 1 1 1 1 1). Символ \$ – кодируется словом ASCII-кода как (0 0 1 0 0 1 0 0).

Сравним затраты на кодирование слова «КОДА». Фраза, состоящая из четырех букв, кодированная ASCII-кодом записывается  $N = 32$  битами. Кодовое слово (12, 3, A) → «КОДА» требует только  $M = 24$  бит. Эффективность сжатия

$$K = \frac{32 - 24}{32} = 0,25,$$

что соответствует 25%. Для длинных текстов кодирование LZ методом позволяет получить эффективность сжатия в пределах  $K = 50 - 60\%$ .

Для длинных текстов LZ77 кодирование практически полностью устраняет избыточность. В этом случае средняя длина кодового слова на один символ источника (текста) стремится к энтропии текста.

*Замечание.* В формате PNG используется разновидность кодирования LZ77, включающая кодирование Хаффмена.

#### 4.4.2. Декодирование LZ-кода

LZ –декодер осуществляет декодирование каждого кодового слова по идентичному словарю буфера поиска, который создается на приемной стороне. При поступлении на вход декодера первых 6-и кодовых слов (пример 4.8) получаем символы передаваемого сообщения, таблица 4.3

Таблица 4.3

N	Кодовое слово (x y z)	Буфер поиска	Сообщение
1	(0, 0, Д)	–	Д
2	(0, 0, Е)	Д	Е
3	(0, 0, К)	ДЕ	К
4	(0, 0, О)	ДЕК	О
5	(4, 1, И)	ДЕКО	ДИ

В слове (4, 1, И) число 4 – это указатель  $x$  позиции в обратном направлении от текущего принимаемого символа до записанного уже ранее в буфер поиска образца. Символ И (указатель  $z$  в слове (4, 1, И)) – это следующая буква декодируемого сообщения. Последующий процесс декодирования показан в продолжении таблицы.

Продолжение таблицы 4.3

6	(0, 0, Р)	ДЕКОДИ	Р
7	(4, 1, В)	ДЕКОДИР	ОВ

8	(0, 0, A)	ДЕКОДИРОВ	A
9	(0, 0, H)	ДЕКОДИРОВА	H
10	(6, 1, E)	ДЕКОДИРОВАН	IE
11	(0, 0, _)	ДЕКОДИРОВАНИЕ	_
12	(12, 3, A)	ДЕКОДИРОВАНИЕ_	КОДА
		ДЕКОДИРОВАНИЕ_КОДА	

Недостатком алгоритма LZ77 кодирования является конечный размер буферов памяти. Типовой размер памяти буфера поиска фраз составляет величину  $W_p = 2^{12} = 4096$ . Размер памяти буфера для записи кодируемой информации  $W_i = 2^4$ . Если образец повторяется, но предыдущий пример его является более удаленным в прошлом, чем длина буфера поиска, сделать ссылку становится невозможным. Кроме того, большие размеры адреса  $x$  и длины  $y$  образца могут потребовать относительно большого числа бит для их записи при формировании кодового слова (ссылки).

#### 4.4.3. Алгоритм кодирования Лемпеля – Зива – Уэлча

Модификацией алгоритма эффективного кодирования LZ77 является алгоритм LZ78, также разработанный Зивом и Лемпелем в 1978 году. В алгоритме LZ78 фактически буферы памяти не имеют конечный размер и изменена структура кодового слова. Модификацией алгоритма LZ78 является метод LZW предложенный Т. Уэлчем (Terry Welsh) в 1984 году.

Метод LZW используется в таких растровых файловых форматах сжатия изображений как:

- GIF (Graphic Interchange Format – Формат обмена графическими данными), находит применение в сети Интернет;
  - TIFF (Tagged Image File Format – Формат представления графической информации, находит применение при подготовке печатных документов).
- Кодирование по алгоритму LZW применяется в методах сжатия JPEG-LS, PNG, в векторном файловом формате PDF (Portable Document Format).

Алгоритм кодирования LZW, как и LZ77 основывается на свойстве межсимвольной зависимости символов данных, повторяемости образцов. В процессе кодирования – декодирования создается словарь образцов. Как и в LZ77, образцы соответствуют символам, словам, фразам, предложениям и пр. Процесс сжатия осуществляется посредством использования ссылок, представляемых в виде индексов. На приемной стороне также создается словарь, который используется для декодирования кодированной информации синхронно с кодером.

Отличие метода кодирования LZW от LZ77 состоит в следующем.

1. Процесс кодирования начинается с загрузки в словарь образцов (буфер поиска) некоторого множества базовых символов алфавита, слов, фраз.
2. Образцы индексируются, например, десятичным номером.

3. При передаче символов сообщения, если в словаре находится нужный образец, отправляется только его индекс.

4. Процесс формирования образцов носит динамический характер.

Пример 4.9. Необходимо передать сообщение ДЕКОДЕР\_КОДА с помощью алгоритма кодирования LZW.

1. Множество базовых символов алфавита

{Д, Е, К, О, Р, \_, А}

передаваемого сообщения состоит из 7-и символов, которые формируют начальный словарь.

2. Этим символам соответствуют индексы:

Д → 1, Е → 2, К → 3, О → 4, Р → 5, \_ → 6, А → 7.

3. Первая буква сообщения передается в виде ссылки (кодového слова) (1).

4. Так как последовательности символов ДЕ в базовом словаре нет, в словарь записывается новый образец в виде числа 8.

5. Далее передается символ Е в виде кодového слова

(2),

и последовательность символов ЕК записывается в словарь под очередным индексом 9.

6. Далее продолжаютс я однобуквенные передачи символов К, О, и записать в словарь последовательностей КО, ОД.

7. Затем передается индексом 8 сразу два символа ДЕ, так в словаре уже имеется этот образец.

По мере расширения словаря передаются все более длительные последовательности символов сообщения. Длина кодového слова также увеличивается. Процесс кодирования иллюстрируется данными табл. 4.4 и табл. 4.5.

Таблица 4.4

Индекс	Словарь образцов
1	Д
2	Е
3	К
4	О
5	Р
6	_
7	А
8	ДЕ
9	ЕК
10	КО
11	ОД
12	ДЕР
13	Р_
14	_К

15	КОД
16	ДА
17	А...

Таблица 4.5

Сообщение	Д	Е	К	О	ДЕ	Р	_	КО	Д	А
Кодовые слова	1	2	3	4	8	5	6	10	1	7

Метод кодирования LZW приводит к сжатию тогда, когда затраты на кодирование оказываются в среднем меньше по сравнению с кодированием кодом ASCII. Затраты на кодирование сообщения ДЕКОДЕР\_КОДА ASCII-кодом составляют

$$N = 12 \times 8 = 96 \text{ бит.}$$

Кодирование LZW-кодом требует использования

$$M = 11 \times 8 = 88 \text{ бит.}$$

Эффективность сжатия

$$K = \frac{96-88}{96} \cong 0,08,$$

что соответствует 8%.

Как видно, даже на малой длине сообщения достигается сжатие. В типовом случае LZW кодирование обеспечивает сжатие текстового файла исполняемого кода примерно наполовину исходного размера.

#### 4.4.4. Декодирование LZW-кода

LZW –декодер осуществляет декодирование каждого кодового слова по идентичному словарю, который создается на приемной стороне. При поступлении на вход декодера последовательности значений индексов (см. пример 4.9)

1, 2, 3, 4, 8, 5, 6, 10, 1, 7,

Для каждого индекса выбираются соответствующие последовательности символов из словаря образцов, табл. 4.6. В результате получаем принятое сообщение.

Таблица 4.6

Кодовые слова	1	2	3	4	8	5	6	10	1	7
Сообщение	Д	Е	К	О	ДЕ	Р	_	КО	Д	А

#### Упражнения

4.9.1. Используйте алгоритм кодирования LZ77 для сжатия сообщения ТЕОРИЯ ИНФОРМАЦИИ – ТЕОРИЯ КОДИРОВАНИЯ.

4.9.2. Оцените эффективность сжатия.

4.10.1. Используйте алгоритм кодирования LZW для сжатия сообщения ТЕОРИЯ ИНФОРМАЦИИ – ТЕОРИЯ КОДИРОВАНИЯ.

4.10.2. Оцените эффективность сжатия.



## 5. КАНАЛЫ БЕЗ ПАМЯТИ И ПЕРЕДАЧА ИНФОРМАЦИИ

### 5.1. Двоичный симметричный канал без памяти

Двоичный симметричный канал (ДСК) является математической двоичной моделью взаимодействия двух дискретных источников без памяти, т. к. приемник можно считать источником информации. Двоичный симметричный канал является также моделью передачи информации по каналу с аддитивным белым гауссовским шумом. Графическое представление такого канала показано на рис. 5.1.

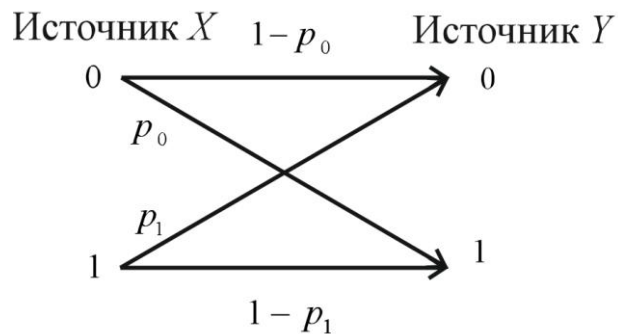


Рис. 5.1. Модель двоичного симметричного канала передачи информации

Входными символами канала являются символы  $x_1 = 0$  и  $x_2 = 1$ . Выходными символами источника являются символы  $y_1 = 0$  и  $y_2 = 1$ . Если при передаче информации выходные символы канала совпадают с входными ( $x_1 = y_1$  и  $x_2 = y_2$ ), пространство событий отражается подмножеством

$$\{00, 11\}.$$

Из-за воздействия шума возможен неправильный прием символов, когда выход канала не совпадает с входом ( $y_1 \neq x_1$  и  $y_2 \neq x_2$ ), В этом случае пространство таких событий отражается подмножеством

$$\{01, 10\}.$$

Тогда передачу информации по модели, показанной на рис. 5.1, можно отображать в виде пространства событий определяемым множеством

$$\{00, 01, 10, 11\}.$$

Это пространство событий описывается вероятностями перехода символов входа  $X$  в символы выхода  $Y$ .

Достоверной передаче символов соответствуют события  $\rightarrow$  вероятности

$$\{00 \rightarrow p(0|0), \{11 \rightarrow p(1|1)\},$$

где  $p(0|0)$  и  $p(1|1)$  – вероятности достоверной передачи символов.

Передача с ошибками характеризуется вероятностями

$$\{10 \rightarrow p(1|0), \{01 \rightarrow p(0|1)\},$$

где  $p(0|1)$  и  $p(1|0)$  – вероятности искаженной передачи символов. Если вероятности искажений  $p(0|1) \cong p(1|0) = p$ , то канал называется двоичным симметричным.

Пусть  $p$  вероятность перехода (вероятность ошибки). Тогда достоверная передача информации – это событие, происходящее с вероятностью  $(1 - p)$ . Для описания ДСК воспользуемся выражением (2.14). Оно определяет вероятность выхода (апостериорную вероятность)  $p(y_i)$  ДСК через распределение вероятностей источника  $X$  (априорные вероятности  $p(x_j)$ ) по формуле

$$p(y_i) = \sum_{j=1}^2 p(y_i|x_j)p(x_j).$$

Тогда

$$p(y_1) = p(y_1|x_1)p(x_1) + p(y_1|x_2)p(x_2),$$

$$p(y_2) = p(y_2|x_1)p(x_1) + p(y_2|x_2)p(x_2).$$

Легко увидеть, что этим выражениям соответствует матричная форма

$$\begin{bmatrix} p(y_1) \\ p(y_2) \end{bmatrix} = \begin{bmatrix} p(y_1|x_1) & p(y_1|x_2) \\ p(y_2|x_1) & p(y_2|x_2) \end{bmatrix} \begin{bmatrix} p(x_1) \\ p(x_2) \end{bmatrix}. \quad (5.1)$$

Двоичный симметричный канал определяется четырьмя вероятностями перехода для значений  $0 \leq p \leq 0,5$ .

*Замечание.* Матрица

$$\mathbf{P} = \begin{bmatrix} p(y_1|x_1) & p(y_1|x_2) \\ p(y_2|x_1) & p(y_2|x_2) \end{bmatrix}$$

называется матрицей переходных вероятностей канала (матрицей канала).

Так как для ДСК справедливо  $p(0|1) \cong p(1|0) = p$ , то

$$p(y_1|x_2) = p(y_2|x_1) = p, \quad p(y_1|x_1) = p(y_2|x_2) = (1 - p). \quad (5.2)$$

Подставляя выражения (5.2) в  $\mathbf{P}$ , распределение вероятностей выходных символов канала записывается в виде

$$\begin{bmatrix} p(y_1) \\ p(y_2) \end{bmatrix} = \begin{bmatrix} 1 - p & p \\ p & 1 - p \end{bmatrix} \begin{bmatrix} p(x_1) \\ p(x_2) \end{bmatrix}. \quad (5.3)$$

Из (5.3) определяются апостериорные вероятности ДСК. Вероятность того, что на выходе канала будет символ  $Y_1$  равна

$$p(y_1) = (1 - p)p(x_1) + pp(x_2).$$

Вероятность того, что на выходе канала будет символ  $y_1$  равна

$$p(y_2) = pp(y_1) + (1 - p)p(y_2).$$

Пример 5.1. Входными символами канала являются символы  $x_1 = 0$  и  $x_2 = 1$ . Пусть  $p = 0,05$ ,  $p(X_1) = 0,9$ ,  $p(X_2) = 0,1$ .

Матрица ДСК имеет вид

$$\mathbf{P} = \begin{bmatrix} 0,95 & 0,05 \\ 0,05 & 0,95 \end{bmatrix}.$$

Вероятность того, что на выходе канала будет символ  $y_1$  равна

$$p(y_1) = (1 - p)p(x_1) + pp(x_2) = 0,95 \cdot 0,9 + 0,05 \cdot 0,1 = 0,86.$$

Вероятность того, что на выходе канала будет символ  $y_2$ , равна

$$p(y_2) = pp(x_1) + (1 - p)p(x_2) = 0,05 \cdot 0,9 + 0,95 \cdot 0,1 = 0,14.$$

2. Найти вероятность  $p\{x|y\}$  получения символов источника  $X$  на приемной стороне

Решение. По теореме Байеса (2.7)

$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{p(y_j)}$$

получаем:

$$p(x_1|y_1) = \frac{p(y_1|x_1)p(x_1)}{p(y_1)} = \frac{0,95 \cdot 0,9}{0,86} = 0,9941.$$

$$p(x_2|y_2) = \frac{p(y_2|x_2)p(x_2)}{p(y_2)} = \frac{0,95 \cdot 0,1}{0,14} = 0,6857.$$

Как видно из примера, в условиях воздействия шума в канале значение вероятности правильного приема символа зависит от степени его повторяемости. Чем чаще он повторяется, тем с лучшими качественными характеристиками осуществляется прием. На основе этого фундаментального принципа введения избыточности (повторяемости) строятся все оптимальные алгоритмы обработки сигналов в шумах, помехоустойчивое кодирование и пр.

## 5.2. Комбинирование источников

Не теряя общего представления о комбинировании  $l$  источников, ограничимся рассмотрением двух источников,  $l = 2$ .

*Определение 5.1.* Под комбинированием источников  $X = \{x_1, x_2, \dots, x_m\}$  и  $Y = \{y_1, y_2, \dots, y_n\}$  понимают источник  $(X, Y)$ , который характеризуется множеством  $\{x, y\}$ . Источник  $(X, Y)$  включает в себя пары совместных событий  $(x, y)$  из  $X$  и  $Y$ .

Вероятности  $p(x, y)$  пар совместных событий удовлетворяют выражению

$$\sum_{i=1}^m \sum_{j=1}^u p(x_i, y_j) = 1.$$

Например, если  $X = \{x_1, x_2\}$ ,  $Y = \{y_1, y_2\}$ , комбинированный источник характеризуется совместными событиями и вероятностями  $p(x_i, y)$ .

$$(X, Y) = \{(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)\}.$$

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) &= \sum_{i=1}^2 [p(x_i, y_1) + p(x_i, y_2)] = \\ &= p(x_1, y_1) + p(x_1, y_2) + p(x_2, y_1) + p(x_2, y_2) = 1. \end{aligned}$$

### 5.3. Совместная энтропия

**Теорема 5.1.** Если источники  $X$  и  $Y$  независимы, то энтропия комбинированного источника  $(X, Y)$  равна сумме энтропий отдельных источников

$$H(X, Y) = H(X) + H(Y).$$

Доказательство. Напомним (см. определение 2.1), источники  $X = \{x_1, x_2, \dots, x_m\}$  и  $Y = \{y_1, y_1, \dots, y_u\}$  независимы, если совместная вероятность каждой пары  $(x, y)$  определяется как  $p_{x,y} = p_x p_y$ .

Пусть  $X = \{x_1, x_2\}$ ,  $Y = \{y_1, y_2\}$ . Тогда событие  $(x, y) \in \{X, Y\}$  имеет вероятность  $p_x p_y$ . Из определения понятия энтропии можно записать для комбинированного источника  $(X, Y)$  следующее выражение:

$$\begin{aligned} H(X, Y) &= - \sum_{i=1}^2 \sum_{j=1}^2 p_{x_i} p_{y_j} \log_2 p_{x_i} p_{y_j} = \\ &= - \sum_{i=1}^2 \sum_{j=1}^2 p_{x_i} p_{y_j} [\log_2 p_{x_i} + \log_2 p_{y_j}] = \\ &= - \sum_{i=1}^2 p_{x_i} \sum_{j=1}^2 p_{y_j} [\log_2 p_{x_i} + \log_2 p_{y_j}] = \\ &= - \sum_{i=1}^2 p_{x_i} \log_2 p_{x_i} \sum_{j=1}^2 p_{y_j} - \sum_{j=1}^2 p_{y_j} \log_2 p_{y_j} \sum_{i=1}^2 p_{x_i}. \end{aligned}$$

Все вероятности символов одиночных источников в сумме должны давать значение 1, поэтому  $\sum_{i=1}^2 p_{x_i} = \sum_{j=1}^2 p_{y_j} = 1$ . В результате получаем

$$H(X, Y) = -\sum_{i=1}^2 p_{x_i} \log_2 p_{x_i} - \sum_{j=1}^2 p_{y_j} \log_2 p_{y_j} = H(X) + H(Y). \quad (5.4)$$

*Утверждение 5.1.* Если источники  $X$  и  $Y$  не являются независимыми, всегда выполняется

$$H(X, Y) < H(X) + H(Y).$$

#### 5.4. Условная энтропия

*Определение 5.2.* Условная энтропия  $H(Y|X)$ — это среднее количество информации на один символ источника  $Y$ , при условии наличия символа источника  $X$ .

Пусть имеется символ  $x_i \in X$ . Для конкретного значения  $y_i \in Y$  в соответствии с определением энтропии (см. (2.18),  $H = \sum_{i=1}^m -p_i \log p_i$ ) условная энтропия записывается в виде

$$H(Y|x_i) = -\sum_{y_j \in Y} p(y_j|x_i) \log_2 p(y_j|x_i). \quad (5.5)$$

*Замечание.* Выражение (5.5) называется частной условной энтропией источника  $Y$  для состояния  $x_i$  источника  $X$ .

Среднее значение  $H(Y|x_i)$  по всем  $x_i \in X$  в соответствии с их вероятностями  $p(x_i)$  определяется по формуле

$$H(Y|X) = \sum_{x_i \in X} H(Y|x_i) p(x_i). \quad (5.6)$$

Подставив выражение (5.5) в (5.6), получаем

$$H(Y|X) = -\sum_{x_i \in X} p(x_i) (\sum_{y_j \in Y} p(y_j|x_i) \log_2 p(y_j|x_i)). \quad (5.7)$$

Очевидно, формулу (5.7) можно записать в виде

$$H(Y|X) = -\sum_{x_i \in X} (\sum_{y_j \in Y} p(y_j|x_i) p(x_i) \log_2 p(y_j|x_i)). \quad (5.8)$$

Заменяя в (5.8) выражение  $p(y_j|x_i)p(x_i)$  на совместную вероятность  $p(x_i, y_j)$  появления двух событий  $y_j$  и  $x_i$ , (формула 2.4,  $p(x_i, y_j) = p(y_j|x_i)p(x_i)$ ), получим

$$H(Y|X) = -\sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j) \log_2 p(y_j|x_i). \quad (5.9)$$

По аналогии можно записать выражение для условной энтропии  $H(X|Y)$ .

*Определение 5.3.* Условная энтропия  $H(X|Y)$  – это среднее количество информации на один символ источника  $X$ , при условии наличия символа источника  $Y$ .

$$H(X|Y) = - \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j) \log_2 p(x_i|y_j). \quad (5.10)$$

*Пример. 5.2.* Вычислим условную энтропию ДСК. Входные символы канала:  $x_1 = 0$  и  $x_2 = 1$ . Выходные символы канала:  $y_1 = 0$  и  $y_2 = 1$ . Вероятность ошибки в канале  $p = \frac{1}{8}$ .

$$\begin{aligned} \text{Решение. } H(Y|X) &= - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log_2 p(y_j|x_i) = \\ &= - \sum_{i=1}^2 [p(x_i, y_1) \log_2 p(x_i|y_1) + p(x_i, y_2) \log_2 p(x_i|y_2)]. \end{aligned}$$

Условная энтропия ДСК  $H(Y|x_1)$  равна

$$H(Y|x_1) = -[p(x_1, y_1) \log_2 p(x_1|y_1) + p(x_1, y_2) \log_2 p(x_1|y_2)]. \quad (5.11)$$

Исходя из свойства переходных вероятностей ДСК

$$p(x_i, y_i) = p(x_j|y_j) = 1 - p; p(x_i|y_j) = p(x_j|y_i) = p,$$

Выражение (5.11) примет вид

$$H(Y|x_1) = -[p(x_1|y_1) \log_2 p(x_1|y_1) + p(x_1|y_2) \log_2 p(x_1|y_2)], \quad (5.12)$$

$$H(Y|x_1) = H(Y|0) = -[(1 - p) \log_2(1 - p) + p \log_2(p)] = H.$$

Аналогично получаем условную энтропию ДСК  $H(Y|x_2)$ :

$$H(Y|x_2) = -[p(x_2|y_1) \log_2 p(x_2|y_1) + p(x_2|y_2) \log_2 p(x_2|y_2)],$$

$$H(Y|x_2) = H(Y|1) = -[p \log_2 p + (1 - p) \log_2 p(1 - p)] = H. \quad (5.13)$$

*Вывод.* Условная энтропия ДСК определяется функцией Шеннона (2.17).

Напомним, функция Шеннона характеризует дискретный источник двух независимых событий  $X = \{0,1\}$  с вероятностями  $p$  и  $(1 - p)$ .

Применяя (5.12) и (5.13), получаем

$$H(Y|0) = H(Y|1) = H = -p \log_2 p - (1-p) \log_2 (1-p) =$$

$$-\frac{1}{8} \log_2 \frac{1}{8} - \frac{7}{8} \log_2 \frac{4}{8} = 0,54 \text{ бит/символ.}$$

#### *Выводы*

1. Из-за действия шумов в канале количество принимаемой информации уменьшилось почти в два раза.

2. Условная энтропия ДСК не зависит от вероятностей символов входа канала.

*Замечание.* В рассматриваемом примере источники  $X$  и  $Y$  связаны между собой каналом с шумами. В этом случае входные символы источника  $X$  позволяют делать некоторое предположение о символах источника  $Y$ . Эта заранее получаемая информация снижает степень неопределенности источника  $Y$ , и, следовательно, уменьшение среднего ожидаемого количества полученной информации – энтропии.

### **5.5. Соотношение между совместной и условной энтропией**

Вновь обратимся к определению совместной энтропии комбинированного источника как математическое ожидание информации всех пар событий этих источников. Не теряя общего представления о комбинированном источнике, ограничимся рассмотрением двух источников  $X = \{x_1, x_2, \dots, x_m\}$  и  $Y = \{y_1, y_2, \dots, y_m\}$ . Событие  $(x, y) \in (X, Y)$  имеет совместную вероятность  $p_x, p_y$ . Из определения понятия энтропии можно записать для комбинированного источника  $(X, Y)$  следующее выражение:

$$H(X, Y) = - \sum_{i=1}^m \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j). \quad (5.14)$$

Заменяя в (5.14) совместную вероятность  $p(x_i, y_j)$  на выражение  $p(y_j|x_i)p(x_i) = p(x_i|y_j)p(y_j)$ , (см. (2.4) для совместной вероятности появления двух событий  $y_i$  и  $x_i$ ), получим

$$H(X, Y) = - \sum_{i=1}^m \sum_{j=1}^m p(x_i)p(y_j|x_i) \log_2 p(x_i)p(y_j|x_i).$$

Раскрывая логарифм произведения, преобразуем последнее выражение к виду

$$H(X, Y) = - \sum_{i=1}^m \sum_{j=1}^m p(x_i)p(y_j|x_i) (\log_2 p(x_i) + \log_2 p(y_j|x_i)) =$$

$$= - \sum_{i=1}^m \sum_{j=1}^m p(y_j|x_i) (p(x_i) \log_2 p(x_i) + p(x_i)p(y_j|x_i) \log_2 p(y_j|x_i)) =$$

$$= - \sum_{i=1}^m \left[ \sum_{j=1}^m p(y_j|x_i) (p(x_i) \log_2 p(x_i) + \sum_{j=1}^m p(x_i) p(y_j|x_i) \log_2 p(y_j|x_i)) \right].$$

Далее за знак суммирования по индексу  $j$  вынесем вероятности с индексом суммирования по  $i$ .

$$H(X, Y) = - \sum_{i=1}^m [(p(x_i) \log_2 p(x_i) \sum_{j=1}^m p(y_j|x_i) + p(x_i) \sum_{j=1}^m p(y_j|x_i) \log_2 p(y_j|x_i))].$$

Последнее выражение запишем в виде суммы двух слагаемых

$$= - \sum_{i=1}^m p(x_i) \log_2 p(x_i) \sum_{j=1}^m p(y_j|x_i) - \sum_{i=1}^m p(x_i) \sum_{j=1}^m p(y_j|x_i) \log_2 p(y_j|x_i).$$

Так как

$$\sum_{j=1}^m p(y_j|x_i) = 1 \text{ и } \sum_{i=1}^m p(x_i) = 1, \text{ получаем}$$

$$H(X, Y) = - \sum_{i=1}^m p(x_i) \log_2 p(x_i) - \sum_{j=1}^m p(y_j|x_i) \log_2 p(y_j|x_i),$$

$$H(X, Y) = H(X) + H(Y|X). \quad (5.15)$$

Аналогично, можно записать формулу

$$H(X, Y) = H(Y) + H(X|Y). \quad (5.16)$$

Совместная энтропия представляется суммой энтропии одного источника и частью другого источника.

*Вывод.* Если между источниками имеется статистическая зависимость, априорные знания свойств одного источника, приводят к уменьшению среднего количества информации на выходе этого источника.

## 5.6. Пропускная способность канала

### 5.6.1. Средняя взаимная информация

Вновь рассмотрим систему связанных источников  $X$  и  $Y$  в виде модели дискретного канала, рис. 2.1. Апостериорное знание вероятностей символов источника  $Y$  позволяет иметь предположение о вероятностях символов источника  $X$ . Апостериорная информация источника  $Y$  об источнике  $X$  уменьшает неопределенность источника  $X$  — энтропию  $H(X)$ . В терминах теории информации среднее количество информации, приходящееся на один символ



источника  $X$ , но после получения апостериорных знаний о событиях источника  $Y$  (на выходе канала) – это условная энтропия  $H(X|Y)$  („новая“ энтропия источника  $X$ ).

Получение апостериорных знаний о событиях источника  $Y$  (на выходе канала) приводит к уменьшению среднего количества информации, приходящей на символ источника  $X$ .

*Определение 5.4.* Средняя взаимная информация событий источника  $X$  при условии наличия событий источника  $Y$  (на выходе канала) равна

$$I(X; Y) = H(X) - H(X|Y), \quad (5.17)$$

где  $H(X) = -\sum_{x_i \in X} p(x_i) \log_2 p(x_i)$ ,

$H(X|Y) = -\sum_{x_i \in X} \sum_{y_j \in Y} p(x_i, y_j) \log_2 p(x_i|y_j)$ .

Формула (5.17) определяет среднее количество информации о источнике  $X$ , переданное источником  $Y$ .

Подставляя в (5.17) выражения  $H(X)$  и  $H(X|Y)$ , получаем форму среднего количества информации переданного источником  $Y$ , которая удобна для анализа системы источник – канал – источник.

$$I(X; Y) = \sum_{x_i \in X} p(x_i) \frac{1}{\log_2 p(x_i)} - \left( \sum_{y_j \in Y} \sum_{x_i \in X} p(y_j, x_i) \frac{1}{\log_2 p(x_i|y_j)} \right) \quad (5.18)$$

Из формулы (5.18) следует, что среднее количество информации, получаемое при наблюдении источника  $Y$  (выхода канала), зависит:

- от статистических характеристических источника  $X$ , т. е. от распределения вероятностей символов  $p(x_i)$  источника  $X$ ;
- от статистических характеристических канала, т. е. от условных вероятностей  $p(x_i|y_j)$  канала (матрицы канала).

В канале без шумов, когда вероятность ошибки  $p = 0$ , компонента

$$H(X|Y) = \left( \sum_{y_j \in Y} \sum_{x_i \in X} p(y_j, x_i) \frac{1}{\log_2 p(x_i|y_j)} \right)$$

выражения (5.18) равна нулю. Тогда среднее количество информации, переданного источником  $Y$ , достигает максимума

$$I(X; Y)_{max} = H(X)$$

и вся информация источника  $X$  передается на выход канала достоверно (без потерь).

При увеличении вероятности ошибки  $p$  в канале среднее количество достоверной информации  $I(X; Y)$  снижается.

Используя соотношение между совместной и условной энтропией (5.15) в форме

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

и (5.17) получаем выражение взаимной информации в виде

$$I(X; Y) = H(X) - H(X|Y) = H(X, Y) - H(X|Y) - H(Y|X). \quad (5.19)$$

### 5.6.2. Пропускная способность канала с матрицей переходных вероятностей

*Определение 5.5.* Пропускная способность  $C$  канала – это максимальная взаимная информация  $I(X; Y)_{max}$ , которая может быть достигнута в канале с матрицей  $\mathbf{P}$  переходных вероятностей.

Пропускная способность канала определяет максимальную скорость передачи информации, при которой она может передаваться без ошибок.

#### 5.6.2.1. Пропускная способность двоичного симметричного канала

Рассмотрим ДСК с источником  $X = \{x_1 = 0, x_2 = 1\}$ . Вероятности символов источника  $X$  обозначим  $p(x_1)$  и  $p(x_2)$ . Канал характеризуется вероятностью ошибок  $p_e$  и матрицей канала

$$\mathbf{P} = \begin{bmatrix} 1 - p_e & p_e \\ p_e & 1 - p_e \end{bmatrix}.$$

На выходе канала формируются символы  $Y = \{y_1 = 0, y_2 = 1\}$ . Если известны входные (априорные) вероятности ДСК, выходные (апостериорные) вероятности  $p(y_i)$  можно найти как (5.3)

$$\begin{bmatrix} p(y_1) \\ p(y_2) \end{bmatrix} = \mathbf{P} \begin{bmatrix} p(x_1) \\ p(x_2) \end{bmatrix} = \begin{bmatrix} 1 - p_e & p_e \\ p_e & 1 - p_e \end{bmatrix} \begin{bmatrix} p(x_1) \\ p(x_2) \end{bmatrix}.$$

Для определения пропускной способности ДСК необходимо вычислить среднюю взаимную информации  $I(X; Y)$ , т. е.

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

В примере 5.2 для ДСК была найдена условная вероятность  $H(Y|X) = H(p)$ . Взаимная информация равна

$$I(X; Y) = H(Y) - H(p).$$

Пропускная способность ДСК определяется как

$$C = \max \{H(Y) - H(p)\}.$$

Когда шумов в канале нет,  $x_1 = y_1, x_2 = y_2$ . Энтропия  $H(Y) = H(X) = 1 \frac{\text{бит}}{\text{символ}}$  достигает своего максимального значения в случае равенства вероятностей символов источника  $X$ . Поэтому пропускная способность ДСК составляет

$$C = \{1 - H(p)\} \frac{\text{бит}}{\text{символ}}.$$

Рассмотрим ситуации возможные при передаче информации.

1. Если ошибки не возникают,  $p_e = 0$ , условная вероятность (формула Шеннона)

$$H(Y|X) = -[p \log_2 p + (1 - p) \log_2(1 - p)] = 0.$$

Пропускная способность ДСК  $C = 1$ . В этом случае обеспечивается надежная передача информации.

2. Если  $p_e = 1$ , условная вероятность

$$H(Y|X) = -[p \log_2 p + (1 - p) \log_2(1 - p)] = 0.$$

Пропускная способность ДСК  $C = 1$ . Среднее количество передаваемой информации также оказывается равным 1 бит/символ. Но в этом случае с вероятностью единица принятый символ не равен переданному символу. Тогда нулевой символ надо читать как единичный и наоборот. Процент ошибок достигает 100%. Невозможно определить, какой символ передавался 0 или 1.

3. Если  $p_e = 0,5$ ,  $H(p) = 1$  и  $C = \{1 - H(p)\} = 0$ , передача информации по ДСК невозможна.

На рис. 5.2 показана зависимость пропускной способности ДСК от вероятности ошибочного приема.

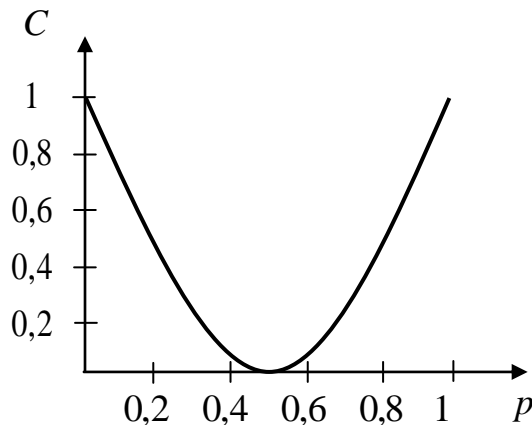


Рис. 5.1. Пропускная способность ДСК

При наличии шумов пропускная способность  $C$  в канале всегда меньше

одного бита на символ (рис. 5.1). Из рисунка видно, что с ростом  $p$  от 0 до 0,5 пропускная способность  $C$  убывает от своего максимального значения, равного 1 до 0. Среднее количество передаваемой информации оказывается равным нулю.

**Пример 5.3.** Пусть в среднем один из каждых 100 символов принимается не правильно, т. е.  $p = 0,01$ . Определить пропускную способность ДСК.

**Решение.**  $C = [1 + 0,01 \log_2 0,01 + 0,99 \log_2 0,99] = 0,98 \frac{\text{бит}}{\text{символ}}$ .

Ошибки в канале уменьшают среднее количество передаваемой информации. В примере, вместо одного бита достоверно получено 0,98 бита.

Известна другая форма записи пропускной способности ДСК.

**Определение 5.6.** Пропускная способность двоичного симметричного канала с вероятностью ошибки  $p$  равна

$$C = W[1 + p \log_2 p + (1 - p) \log_2 (1 - p)] \text{ бит/с},$$

где  $W = 1/\tau$  – тактовая частота следования информационных символов;  $\tau$  – длительность символа.

При отсутствии помех  $p = 0$ , пропускная способность достигает максимума:

$$C_{\max} = W \text{ бит/с}.$$

## 5.7. Дифференциальная энтропия

Понятие дифференциальной энтропии вводится для непрерывных источников информации. Выходом такого источника является непрерывный (аналоговый) сигнал  $x(t)$  – случайная функция от времени  $t$ . Для непрерывных представлений энтропии, вместо вероятностей символов источников рассматриваются функции распределения плотностей переменных. Введение функций распределения позволяет определить понятия энтропии, условной энтропии для непрерывных источников.

**Определение 5.7.** Дифференциальная энтропия – это среднее количество информации непрерывного источника определяется как

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx,$$

где  $f(x)$  обозначает функцию распределения вероятности случайного процесса непрерывного источника.

Основываясь на общих понятиях о дифференциальной энтропии можно утверждать, что энтропийное представление, описывающие дискретные источники без памяти и каналы, справедливы и для непрерывных источников и каналов.

## 5.8. Пропускная способность канала непрерывного канала

Качество цифровой системы передачи информации характеризуется вероятностью ошибки на бит (частотой ошибок на бит). При передаче дискретных данных по каналу с аддитивным гауссовским шумом вероятность ошибки на бит может быть уменьшена путем увеличения мощности передатчика, которая также является одной из характеристик качества системы. Лучшей из двух систем передачи данных считается та, которая достигает желаемой частоты ошибок на бит при меньшей мощности передатчика.

Сообщение из  $k$  информационных бит имеют энергию

$$E_c = \sum_{n=0}^{k-1} |x(n)|^2.$$

Энергия сигнала, соответствующая одному информационному биту, определяется соотношением

$$E_b = \frac{E_c}{k} = \frac{1}{k} \sum_{n=0}^{k-1} |x(n)|^2. \quad (5.20)$$

*Замечание.* Проверочные символы, символы синхронизации (например, начала кодового слова, строчные, кадровые или символы канального обмена и др.) не несут информации и поэтому не могут участвовать в вычислении  $E_b$ .

Для сообщений, передаваемых со скоростью  $R_i$  информационных бит/с, величина  $E_b$  определяется из выражения

$$E_b = \frac{P_c}{R_i},$$

где  $P_c$  – средняя мощность сообщения.

На вход приемника поступает также и белый шум с односторонней спектральной плотностью  $N_0$  Вт/Гц. Очевидно, что на частоту ошибок на бит влияет только отношение  $\frac{E_b}{N_0}$ . Сравнительные качественные характеристики различных способов передачи сигналов можно получить, оценивая зависимости их вероятностей ошибок на бит от отношения  $\frac{E_b}{N_0}$ . Нижняя, теоретически достижимая в цифровой системе передачи информации, граница  $\frac{E_b}{N_0}$  определяется из формулы пропускной способности непрерывного канала.

**Теорема 5.2.** Пропускная способность (Хартли – Шеннона) идеального канала равна

$$C = W \log_2(1 + P_c/P_N), \quad (5.21)$$

где  $W$  – ширина полосы частот канала;  $P_N = N_0 W$  – средняя мощность помех с нормальным законом распределения амплитуд и равномерным спектром в полосе частот канала.

### 5.8.1. Граница Шеннона

Определим границу абсолютного значения отношения  $\frac{E_b}{N_0}$ . Ширина полосы сигнала в формуле (5.21) не ограничена. Устремим  $W$  к бесконечности и найдем предельное значение пропускной способности канала  $C_\infty$ :

$$C_\infty = \lim_{W \rightarrow \infty} C = \lim_{W \rightarrow \infty} W \log_2 \left( 1 + \frac{P_c}{N_0 W} \right).$$

Обозначим  $\frac{1}{W}$  символом  $\gamma$ , тогда можно записать

$$C_\infty = \lim_{\gamma \rightarrow 0} C = \lim_{\gamma \rightarrow 0} \frac{1}{\gamma} \log_2 \left( 1 + \frac{P_c}{N_0} \gamma \right). \quad (5.22)$$

Функция (5.22) в точке  $\gamma = 0$ , принимая вид  $0/0$ , не определена. Для раскрытия неопределенности и вычисления предела воспользуемся правилом Лопиталя. Продифференцируем числитель и знаменатель (5.22) по  $(1 + \frac{P_c}{N_0} \gamma)$ .

$$C_\infty = \lim_{\gamma \rightarrow 0} \frac{\log_2(1 + \frac{P_c}{N_0} \gamma)'}{\gamma'}. \quad (5.23)$$

Числителю (5.23) соответствует выражение

$$\log_2(x)' = \log_2 e.$$

Дифференцирование знаменателя приводит к

$$\frac{d\gamma}{d(1 + \frac{P_c}{N_0} \gamma)} = \frac{N_0}{P_c}.$$

Формулу (5.23) можно записать как

$$C_\infty = \lim_{W \rightarrow \infty} C = \frac{P_c}{N_0} \log_2 e = 1,443 \frac{P_c}{N_0}. \quad (5.24)$$

Выражения (5.24) и (5.22) определяют границу Шеннона. Подставив в последнее выражение значение мощности информационных бит

$$R_i E_b = P_c,$$

получим граничное значение  $\frac{E_b}{N_0}$  для максимальной скорости передачи информации  $R_i = C_\infty$ .

$$R_i \leq 1,443 \frac{P_c}{N_0} = 1,443 R_i \frac{E_b}{N_0},$$

$$1 \leq 1,443 \frac{E_b}{N_0},$$

$$0,69 \cong \frac{1}{1,443} \leq \frac{E_b}{N_0}.$$

*Замечание.* Нужно иметь в виду, что формула Шеннона справедлива

только тогда, когда передаваемый сигнал образует аддитивную смесь с белым шумом. Кроме того, по своим статистическим свойствам сигнал подобен нормальному стационарному шуму с заданной средней мощностью и равномерной спектральной плотностью внутри полосы частот  $W$ .

**Теорема 5.3.** В системе, передающей информацию в условиях белого гауссовского шума с односторонней спектральной плотностью  $N_0$  необходимо, чтобы энергия на бит удовлетворяла неравенствам:

$$E_b \geq 0,69 N_0;$$

$$\frac{E_b}{N_0} \geq 0,69 \cong -1,6 \text{ dB}.$$

*Вывод.* Для передачи одного информационного бита необходимо, чтобы отношение энергии на бит  $E_b$  к спектральной плотности мощности шума  $N_0$  было, как минимум 0,69.

Формула (5.21) приводит к очень важному заключению: для случая малого отношения

$$\frac{P_c}{P_N} = \frac{\text{сигнал}}{\text{шум}} \ll 1$$

на входе приемника пропускная способность канала

$$C_\infty = 1,443 \frac{P_c}{N_0}$$

не зависит от ширины его пропускания, а определяется средней мощностью передаваемого сигнала и спектральной плотностью мощности шума (мощности, приходящейся на единицу полосы).

Теорема 5.3 определяет нижний предел допустимого отношения  $\frac{E_b}{N_0}$ . Верхний предел установлен экспериментально. При отношении сигнал/шум  $\frac{E_b}{N_0} \geq 12 \text{ dB}$  обеспечивается практически безошибочная передача информации. Следовательно, практическая необходимость применения кодирования и выбора соответствующего кода возникает лишь в том случае, когда соотношение  $\frac{E_b}{N_0}$  лежит в диапазоне минус 1,6 dB плюс 12 dB.

Формулу (5.21) можно записать в следующем виде:

$$I = WT \log_2(1 + P_c/P_N), \quad (5.25)$$

где  $I$  характеризует максимальное количество информации, передаваемое по каналу за время  $T$ .

Из (5.22) следует, что при уменьшении отношения  $(P_c/P_N)$  или  $\frac{E_b}{N_0}$  можно

сохранить количество передаваемой информации, расширяя полосу сигнала или увеличивая время передачи. Выражение (5.22.) имеет фундаментальное значение для теории кодирования.

## 5.9. Статистические характеристики каналов

Проектирование, разработка инфокоммуникационных систем различного назначения требует априорного знания значений вероятностей ошибок при передаче информации в том или ином канале. Статистика ошибок в различных каналах к настоящему времени исследована достаточно полно. Опубликованы результаты экспериментов, касающиеся измерения частоты ошибок на информационный бит и характера их группирования в каналах. Вероятность появления ошибок на выходе соответствующего канального приемника по данным отечественной и зарубежной литературы находится в пределах:

радиорелейного  $p_f = 10^{-4} - 10^{-5}$ ;  
 телефонного  $p_f = 10^{-3} - 10^{-5}$ ;  
 магнитной ленты  $p_f = 10^{-4} - 10^{-5}$ ;  
 телеметрического  $p_f = 10^{-6} - 10^{-10}$ ;  
 оптического диска  $p_f = 10^{-5} - 10^{-6}$ ;  
 космического телеметрического  $p_f = 10^{-12} - 10^{-23}$ .

В телеметрическом канале космического корабля многоразовых полетов (Space Shuttle) с помощью (127, 120) БЧХ-кода достигается вероятность ложного приема командной информации менее  $p_f = 6,62 \cdot 10^{-23}$ .

Для цифровых устройств различают следующие вероятности ошибок:

- вероятность ошибки из-за дефекта (отказа) элементов;
- вероятность ошибки из-за сбоя элементов.

Дефект – отказ какого-либо элемента, местоположение которого известно, а состояние не изменяется при входных воздействиях. Покажем это на примере матрицы оперативного ЗУ. Дефектной является третья ячейка памяти, рис. 5.2.

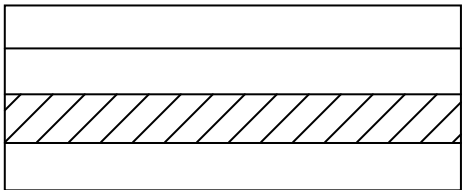
Записываемая информация	Дефект (состояние "1" 3-ей ячейки)	Считываемая информация
0		0
1		1
0		1
1		1

Рис. 5.2

Сбой это перемежающийся переход состояния элемента из правильно-



го в неправильное и обратно (возможен и при входном воздействии).

Для полупроводниковой памяти вероятность ошибки из-за отказа

$p_{fa} = 10^{-5} - 10^{-6}$ , а вероятность ошибки из-за сбоев элементов

$p_{fs} = 10^{-4} - 10^{-5}$ . Для того чтобы гарантировать высокую надежность памяти необходимо иметь вероятность ошибки из-за отказа  $p_{fa} = 10^{-7} - 10^{-8}$ .

Упражнения

5.1. Определить пропускную способность ДСК с вероятностью  $p = 10^{-3}$ .

5.2. Определить пропускную способность непрерывного канала, если  $\frac{P_c}{N_0} = 0,1$ .

5.3. Показать, что отношение  $P_c/P_N$  определяет параметр  $(2^{\frac{R_i}{W}} - 1)$  системы передачи информации.

5.4.1. Вычислить отношение сигнал/шум по мощности на выходе космического канала Марс – Земля.

Средняя мощность сигнала на входе приемника определяется соотношением

$$S = \frac{S_s G_s A}{4\pi D^2} \frac{1}{B},$$

где  $S_s$  – средняя мощность сигнала передатчика;

$G_s$  – коэффициент усиления антенны передатчика;

$D$  – расстояние до приемника;

$A$  – эффективная площадь антенны приемника;

$B$  – коэффициент потерь мощности сигнала на входе приемника, учитывающий потери мощности (влияние ионосферы, тропосферы, неравномерности диаграмм направленности антенны передатчика и приемника, и др.).

Экспериментальные исследования космического канала показали, что величина  $B$  находится в диапазоне 1,2 – 2.

Расстояние  $D \approx 400$  млн. км.

Пусть  $S_s = 60$  Вт. Если принять коэффициент полезного действия передатчика 10% (что реально для несущей частоты передатчика космического аппарата порядка  $f = 1000$  МГц), передатчик должен иметь блок энергоснабжения мощностью 600 Вт. Для обеспечения такого расхода электричества на Марсе потребуются солнечные батареи с площадью панелей  $\approx 15$  м<sup>2</sup>.

Коэффициент  $G_s$  зависит от размеров антенны с параболическим рефлектором передатчика космического аппарата. Пусть диаметр антенны равен 1,5 м. В этом случае можно иметь коэффициент усиления  $G_s \approx 200$ .

Из-за большого расстояния между передатчиком и приемником, и необходимостью иметь приемлемое значение средней мощности сигнала на входе приемника, на Земле используются приемные антенны (антенные поля) большой площади. Пусть  $A = 600$  м<sup>2</sup>.

Исследования космических каналов большой протяженности показали, что основной помехой в них является белый шум со спектральной плотностью мощности

$$N_0 = kT,$$

где  $k = 1,38 \cdot 10^{-23} \frac{\text{Дж}}{\text{град}}$  – постоянная Больцмана,

$T$  – шумовая температура всех источников помех (собственные шумы устройств космического аппарата, Галактика, яркие звезды, Солнце, Луна, Земля, атмосфера и др.). На частоте  $f = 1000$  МГц шумовая температура  $T \approx 50^\circ\text{K}$ .

Пусть ширина полосы частот  $W$  канала равна  $W = 20$  КГц.

5.4.2. Вычислить пропускную способность космического канала Марс – Земля.

## **6. ВВЕДЕНИЕ В ЗАЩИТУ ИНФОРМАЦИИ**

Защита информации – комплекс мероприятий, направленных на обеспечение информационной безопасности.

Под информационной безопасностью понимают защищенность информации и поддерживающей инфраструктуры от случайных и преднамеренных воздействий естественного или искусственного характера, которые могут нанести неприемлемый ущерб пользователям информации и поддерживающей инфраструктуре. Информационная безопасность не сводится только к защите от несанкционированного доступа к информации, это более широкое понятие. Субъект информационных отношений может пострадать (понести убытки, моральный ущерб) не только от несанкционированного доступа к информации, но и от повреждения элементов информационной системы.

Информационная безопасность в значительной степени зависит от надежности поддерживающей инфраструктуры к которой можно отнести системы электро-, водо- и теплоснабжения, средства коммуникации, обслуживающий персонал и др.

В определении информационной безопасности употреблено понятие «неприемлемый ущерб». Например, недопустимым ущербом являются: нанесения вреда здоровью, окружающей среде, урон, нанесенный стране и пр. Часто порог неприемлемости имеет материальное (денежное) выражение. Застраховаться от всех видов ущерба невозможно. Тогда целью защиты информации становится уменьшение размеров ущерба до допустимых значений.

### **6.1. Составляющие информационной безопасности**

Спектр интересов субъектов, связанных с использованием инфокоммуникационных систем разделяют на следующие категории:

- обеспечение доступности;
- обеспечение конфиденциальности информационных ресурсов и поддерживающей инфраструктуры;
- обеспечение целостности информации.

Информационные системы необходимы для получения информационных услуг. Если услуги становятся по разным причинам недоступными, то наносится ущерб субъектам информационных отношений.

*Определение 6.1.* Доступность – это возможность получить информационную услугу за приемлемое время.

*Определение 6.2.* Конфиденциальность – это статус, представленный данным и определяющий требуемую степень их защиты.

*Определение 5.3.* Секретность – это понятие, которое употребляется по отношению к отдельным лицам, которые имеют право объявлять информацию закрытой, т. е. подлежащей защите.

*Определение 5.4.* Под целостностью понимают защищенность информации от разрушения и несанкционированного изменения.

Целостность подразделяют на:

- статическую, понимаемую как неизменность информационных объектов;
- динамическую, относящуюся к правильному выполнению сложных действий (транзакций).

Рецептура лекарств, характеристики комплектующих изделий, описание хода технологического процесса, база данных землетрясений, данные аэро – , космического дистанционного зондирования участков Земли – все это примеры информации, нарушение целостности которой может привести к неприемлемому ущербу. Преднамеренное искажение информации – это также нарушение целостности.

Введенные категории информационной безопасности рассматривают относительно независимо. Считается, что если все три категории реализуются, то обеспечивается информационная безопасность. В этом случае субъектам информационных отношений не будет нанесен неприемлемый ущерб.

*Замечание.* К поддерживающей инфраструктуре применимы те же требования целостности и доступности, что и к информационным системам.

## **6.2. Информационные угрозы и атаки**

*Определение 6.5.* Угроза – это потенциальная возможность определенным образом нарушить информационную безопасность.

*Определение 6.6.* Попытка реализации угрозы называется атакой, а тот, кто предпринимает такую попытку, называется злоумышленником.

Угроза является следствием наличия уязвимых мест в защите информационных систем (таких, например, как возможность доступа посторонних лиц к важному оборудованию).

*Определение 6.6.* Промежуток времени от момента, когда появляется возможность использовать слабое место в защите, и до момента, когда пробел ликвидируется, называется окном опасности.

Некоторые угрозы существуют в силу самой природы современных информационных систем. Например, угрозы отключения электричества или выхода параметров источника напряжения за допустимые пределы существуют в силу зависимости аппаратного обеспечения информационных систем от надежности и качественных характеристик электропитания. Иметь представление о возможных угрозах, а также об уязвимых местах защиты необ-

ходимо для того, чтобы выбирать эффективные и наиболее экономичные средства обеспечения информационной безопасности. Угрозы классифицируют по следующим критериям:

- по составляющим информационной безопасности (доступность, целостность, конфиденциальность);
- по компонентам информационных систем (аппаратура, поддерживающая инфраструктура, данные, программы);
- по способу осуществления: случайные и преднамеренные.

Случайные угрозы могут быть обусловлены физическими воздействиями стихийных природных явлений, не зависящих от человека. К угрозам случайного характера также относятся аварийные ситуации на объекте размещения информационной системы. Аварийные ситуации это – отказы аппаратуры системы, пожары, наводнения, ураганы, разряды атмосферного электричества и др.

Преднамеренные угрозы направлены против элементов и подсистем, образующих информационную систему.

### **6.3. Модели разграничения доступа к информации в инфокоммуникационных системах**

Устранение или уменьшение преднамеренных угроз основывается на использовании определенных моделей разграничения доступа к информации. Модели строятся с учетом следующих возможных злоумышленных действий:

- несанкционированный доступ к информации и ознакомление с хранящейся и циркулирующей в информационной системе конфиденциальной информацией;
- доступ локальных пользователей к информации, на работу с которой они не имеют полномочий;
- несанкционированное копирование сведений: данных и программ;
- кража физических носителей информации и оборудования, приводящая к утрате информации;
- умышленное уничтожение информации;
- несанкционированная модификация документов и баз данных;
- фальсификация сообщений;
- дезинформация, т. е. навязывание ложного сообщения и пр.

Конкретные модели разграничения доступом к информации в инфокоммуникационных системах должны учитывать следующие угрозы доступности.

1. Самыми частыми и самыми опасными (с точки зрения размера ущерба) являются случайные (непреднамеренные) ошибки лиц, обслуживающих информационные системы. По некоторым источникам, до 65% потерь – это следствие непреднамеренных ошибок (ошибки в программе), вызвавшие крах системы.

2. Отказ информационной системы, повреждение аппаратуры, разрушение данных (например, мощный кратковременный импульс способен разрушить данные на магнитных носителях);

3. Программные атаки на доступность, когда используется агрессивное потребление ресурсов (полосы пропускания сетей, вычислительной способности процессора или оперативной памяти);

4. Внедрение в атакуемые системы вредоносного программного обеспечения.

5. Отказ поддерживающей инфраструктуры (нарушение работы (случайное или преднамеренное)), системы связи, электропитания, террористический акт.

6. Стихийные бедствия. По статистике на долю огня, воды, землетрясений, ураганов и пр. приходится 13% потерь, нанесенных информационным системам.

7. Отказ пользователей (невозможность работать с информационной системой в силу отсутствия подготовки, технической поддержки, справочной литературы и др.).

#### **6.4. Методы разграничения доступа и способы их реализации**

Общий подход по методам разграничения доступа к информации и способам их реализации основывается на стандартах информационной безопасности. Стандарты описывают средства, с помощью которых обеспечивается информационная безопасность.

Исторически первым стандартом, получившим широкое распространение и оказавшим влияние на базу стандартизации информационной безопасности во многих странах мира, стал стандарт Министерства обороны США «Критерии оценки доверенных компьютерных систем». Этот стандарт, называемый по цвету обложки «Оранжевая книга», был опубликован в 1983 году. Особенностью стандарта является то, что рассмотрению подлежат так называемые доверенные системы, т. е. информационные системы, которым можно оказать определенную степень доверия, с точки зрения информационной безопасности. «Оранжевая книга» поясняет понятие безопасной системы, которая управляет с помощью соответствующих средств, доступом к информации так, что только авторизованные лица или процессы, действующие от их имени, получают право читать, записывать, создавать и удалять информацию.

*Определение 6.7.* Доверенная система – это информационная система, использующая достаточные аппаратные и программные средства, чтобы обеспечить одновременную обработку информации разной степени секретности группой пользователей без нарушения прав доступа.

*Замечание.* Безопасность и доверие оцениваются с точки зрения управления доступом к данным, что является одним из средств обеспечения конфиденциальности и целостности (статической). Например, руководство режимного предприятия в первую очередь заботится о защите от несанкционированного доступа к информации.

рованного доступа к информации, т. е. конфиденциальности. В частности, правила определяют, в каких случаях пользователь может работать с конкретными данными.

Если понимать политику безопасности узко, то это правила разграничения доступа к информации («Можно читать только то, что положено» или «Субъект может читать информацию из объекта, если уровень секретности субъекта не ниже, чем у объекта»).

Стандартные средства осуществляют разграничение доступа к информации на следующих уровнях.

1. Законодательный. В области информационной безопасности законы реально работают через нормативные акты, подготовленные соответствующими ведомствами.

2. Административный. Практическое осуществление административных мер защиты информации связано с ограничением доступа людей к аппаратуре, компьютерам, программам, обрабатываемой информации, данным и пр. На этом уровне устанавливаются способы доступа к информации и условия ее распространения, регламентируются процедуры выдачи допусков к данным. Часть из этих правил определяются законами и нормативными актами. Но большинство правил определяет организация на основе приказов и инструкций. При введении административных мер допуска возникают определенные проблемы. Реализация этих мер создает неудобства для пользователей. Эффективность административных мер может свестись к нулевой: список паролей будет лежать под стеклом, дверь то запирается, то открыта и т. д.

3. Процедурный. На этом уровне выделяют такие меры доступа как:

- управление персоналом, физическая защита;
- технический.

Реализация меры по управлению персоналом строится на двух принципах:

- разделение обязанностей;
- минимизация привилегий.

Принцип разделение обязанностей предписывает распределение ответственности. Например, один инженер разрабатывает схему процессора, а другой разрабатывает схему слежения за задержкой и пр.

Принцип минимизации привилегий предписывает выделять пользователям информационной системы только те права, которые необходимы для выполнения своих обязанностей. Назначение этого принципа – уменьшить ущерб от случайных или умышленных действий.

Основной принцип физической защиты доступом формулируется как «непрерывность защиты в пространстве и времени». Для физической защиты окон опасности быть не должно. Под физической защитой здесь понимается отражение попыток несанкционированного доступа к данным. Средства физического управления доступом:

- охрана;

- двери с замками;
- телекамеры;
- датчики движения и др.

Выделяют четыре вида охранных мер:

- охрана границ территории (зоны, окружающей здание);
- охрана самого здания;
- охрана входов в здание;
- охрана критических зон.

Для защиты границ территории используют ограды, инфракрасные или СВЧ детекторы, а также замкнутые телевизионные системы.

Для защиты здания оно должно быть построено из прочных материалов и иметь толстые стены. Здание фирмы IBM, например, имеет стены из железобетона толщиной 33 см.

Для обнаружения проникновения злоумышленника в критическую зону используют системы сигнализации (системы наблюдения за входом в помещение). К наиболее распространенным системам сигнализации относятся следующие.

1. Фотометрические системы обнаруживают изменение уровня освещенности.
2. Звуковые, ультразвуковые, СВЧ системы обнаружения реагируют на изменение частоты сигнала, отраженного от движущегося объекта.
3. Акустосейсмические (вибрационные) системы обнаруживают шум и вибрации.
4. Системы, реагирующие на приближение к объекту, обнаруживают нарушение структуры электромагнитного или электростатического поля.

Кроме того, посредством физической защиты реализуется управление носителями.

## **6.5. Обеспечение целостности данных в инфокоммуникационных системах и сетях**

С целью нарушения статической целостности злоумышленник может:

- ввести неправильные данные;
- изменить данные;
- разрушить информацию деструктивными программными воздействиями (компьютерными вирусами и пр.).

Угрозами динамической целостности являются:

- нарушение атомарности сложных действий;
- переупорядочение;
- кража;
- дублирование данных или внесение дополнительных сообщений (сетевых пакетов и пр.).

Кроме того, случайные ошибки пользователей системы, обслуживающего персонала, грозят повреждением аппаратуры, разрушением программ и пр.

Выделяют следующие направления деятельности, относящиеся к обес-



печению целостности данных.

1. Необходима поддержка пользователей – это консультирование, связанное с информационной безопасностью. Целесообразно фиксировать вопросы пользователей, чтобы выявлять их типичные ошибки.

2. Для обеспечения целостности и доступности поддерживающей инфраструктуры нужно защищать оборудование от краж и повреждений, выбирать оборудование с максимальным временем наработки на отказ, дублировать узлы, иметь запасные части.

3. Поддержка программного обеспечения – необходимо следить за тем, какое программное обеспечение установлено на компьютерах; необходим контроль неавторизованного доступа к программам и их изменениям.

Конфигурационное управление контролирует и фиксирует изменения в программной конфигурации. Фиксация изменений позволяет восстановить конфигурацию после аварии. Необходима защита от случайных, непродуманных модификаций. Предусмотреть возможность возврата к прошлой программной работающей версии.

Резервное копирование необходимо для восстановления программ и данных после аварии. Должны быть созданы полные эталонные копии программ системы. Копии размещаются в месте, защищенном от несанкционированного доступа.

Управление носителями, посредством физической защиты, обеспечивает целостность хранящейся вне компьютерных систем. Под физической защитой здесь понимается не только отражение попыток несанкционированного доступа, но и предохранения от вредных влияний окружающей среды (влага, жара, холод, магнетизм).

Документирование. В виде документов оформляется журнала учета носителей информации, план восстановления данных после аварии.

Для построения надежной защиты информации современная информационная система должна включать в себя и такие сервисы безопасности как:

- помехоустойчивое кодирование;
- криптографическое кодирование (шифрование).

## **6.6. Общие сведения по классической криптографии**

### **6.6.1. Криптографическое кодирование**

*Определение 6.8.* Криптография – это раздел прикладной математики, изучающий модели, методы, алгоритмы, программные и аппаратные средства преобразования информации.

Криптографическое кодирование (шифрование) относится к традиционным сервисам безопасности. Это наиболее мощное средство обеспечения информационной безопасности. Криптографическое кодирование необходимо для реализации трех сервисов безопасности:

- шифрования;
- контроля целостности;
- аутентификации.

*Определение 6.9.* Шифрование – это кодирование (преобразование) исходного текста, который носит название открытого текста, в зашифрованный текст (криптограмму) с помощью секретного ключа.

Процесс создания криптограммы записывается как

$$C = E_k(m),$$

где  $m$  (message) – открытый текст,  $E$  (encryption) – шифрующая функция (кодирование, преобразование),  $k$  – секретный ключ,  $C$  – шифротекст.

Ключ  $k$  определяет также процесс, обратный шифрованию, который называют расшифрованием (дешифрованием)

$$m = D_k(C),$$

где  $D$  (decryption) – дешифрующая функция (декодирование, обратное преобразование).

При этом должно выполняться тождество

$$m = D_k(C) = D_k(E_k(m)).$$

*Замечание.* Алгоритмы шифрования  $E$  и дешифрования  $D$  открыты, и секретность исходного текста  $m$  в шифротексте  $C$  зависит от ключа  $k$ .

Существуют два основных алгоритма шифрования.

В первом, процессы шифрования – дешифрования используют один и тот же секретный ключ отправителем и получателем информации. Этот метод используется в, так называемых, симметричных криптосистемах. В них ключ поставляется абонентам специальным конфиденциальным способом.

Второй метод шифрования основан на двух ключах. Первый, – открытый ключ, доступный всем пользователям информационной системы, применяется при шифровании. Второй ключ, математически связанный с первым – секретный. Он нужен при расшифровании текста. Такие криптосистемы называются асимметричными, или криптосистемами с открытым ключом.

*Определение 6.10.* Криптосистема – это система, реализованная программно, аппаратно или программно – аппаратно и осуществляющая криптографическое преобразование информации.

Аппаратная реализация имеет существенную стоимость, однако ей присущи и преимущества:

- высокая производительность;
- сравнительная простота;
- защищенность.

Программная реализация криптосистемы более практична, допускает гибкость в использовании.

#### **6.6.1.1. Требования к криптосистемам защиты информации**

Основными требованиями являются:

- функциональное преобразование сообщения  $E_k(m)$  и обратное преобразование зашифрованного сообщения  $D_k(C)$  должны быть сравнительно легко вычислимы;
- не зная ключ  $k$ , невозможно за заданное (реальное) время вычислить сообщение  $m$  по шифротексту  $C = E_k(m)$ . Не должно быть простых зависимостей между используемыми ключами;
- изменение длины ключа не должно вести к качественному ухудшению алгоритма шифрования;
- число операций, необходимых для определения ключа по фрагменту шифрованного текста должно быть не меньше общего числа ключей;
- число операций, необходимых для дешифрования информации путем перебора всевозможных ключей должно иметь строгую нижнюю оценку. При этом следует учитывать вычислительные возможности по числу операций в единицу времени современных суперкомпьютеров, возможности использования сетевых вычислений;
- знание алгоритма шифрования не должно влиять на надежность защиты.

#### **6.6.2. Криптоанализ**

Криптоанализ занимается задачами, обратными задачам криптографии. Основной задачей специалиста криптоаналитика является поиск ключа. Ему могут представиться следующие возможности для атаки:

- получен лишь зашифрованный текст  $C = E_k(m)$ ;
- известны незашифрованный и зашифрованный тексты;
- имеется возможность выбрать пространство сообщений  $\{m\}$  и пространство шифротекстов  $\{C\}$ , т. е. иметь пару  $\{m, C\}$ .

Криптоанализ и криптография развиваются параллельно. Криптографы пытаются создать такую криптосистему, которая была бы стойкой ко всем известным в данный момент методам криптоанализа.

Эффективность шифрования зависит от сохранения тайны ключа и криптостойкостью шифра.

*Определение 6.11.* Криптостойкостью называется характеристика шифра, определяющая его стойкость криптоанализу (к дешифрованию) без знания ключа.

Имеется несколько основных показателей криптостойкости, среди которых:

- количество всех возможных ключей;
- среднее время, необходимое для криптоанализа.

В практических применениях ограничиваются алгоритмами, обеспечивающими вычислительную стойкость, т. е. такими алгоритмами, которые теоретически раскрываемы, но требуют для осуществления криптоанализа значительных вычислительных затрат, например, работы 10 млн. компьютеров в течение 10000 лет.

История развития криптологии имеет большую продолжительность. При археологических раскопках в Месопотамии был найден, относящийся к 20 веку до н. э. один из самых древних шифротекстов. Он был написан клинописью на глиняной дощечке и содержал коммерческую тайну: рецепт глазури для покрытия гончарных изделий. В 17 веке кардинал Ришелье создал первую в мире шифрослужбу. Задачами криптологии занимались такие известные ученые как Ньютон, Лейбниц, Эйлер, Гаусс и др. В развитии криптологии принято выделять три этапа.

Первый этап – с древних времен до 1949 года, характеризовался частными, узкоспециальными и вычислительно простыми алгоритмами криптографии и криптоанализа. Этот этап называют этапом докомпьютерной криптологии.

Второй этап – с 1949 до 1976 года, когда К. Шеннон опубликовал работу «Теория связи в секретных системах». В этот период проводились большие исследования с использованием компьютеров. Основным потребителем результатов криптологии являлась связь для военных и дипломатических организаций, поэтому криптология была закрытой наукой.

Третий этап – с 1976 года по настоящее время, называют периодом открытой криптологии. Его принято отсчитывать с момента публикации работы американских математиков У. Диффи (W. Diffie) и М. Хеллмена (M. Hellman) «Новые направления в криптографии». В этой работе показано, что секретная передача информации возможна без предварительной передачи ключа. Особенностью этого этапа стало применение криптографии в банковском деле, компьютерных сетях и др. приложениях. В развитие криптологии вкладываются значительные государственные средства. Например, в США ежегодные расходы на криптологию составляют порядка 18 – 20 млрд. долларов.

Криптология строится на базе таких дисциплин как теория вероятностей, математическая статистика, алгебра, теория чисел, теория алгоритмов и сложность вычислений. Процесс шифрования осуществляется на специализированных компьютерах.

## **6.7. Алгоритмы блочного шифрования**

Блочное шифрование используется в симметричных криптосистемах. Блочный шифр обрабатывает блок открытого текста фиксированной длины. Процесс шифрования текста  $m$  записывается как

$$C = E_k(m)$$

где  $C$  – это блок шифротекста,  $k$  – секретный ключ, шифрующая функция. В этом алгоритме зашифрованный первый блок сообщения далее используется для шифрования следующего. Шифрующие процедуры одного типа чередуются с процедурами другого типа. В качестве простых шифров могут быть использованы подстановки  $S$ , перестановки  $L$  и линейные преобразования  $T$ . Секретный ключ может использоваться при осуществлении всех процедур. Блочные шифры используют многократное повторение операций преобразования, называемое раундом шифрования.

На этом принципе работает известный стандарт шифрования данных DES и его модификация AES. DES использует ключи размером 64 бит с эффективной длиной 56 бит. Всего создается  $K_{DES} = 2^{56} \approx 7,2 \times 10^{16}$  ключей. Система AES имеет три варианта размеров ключей 128, 192 и 256 битов. Длина ключа 128 бита позволяет формировать  $K_{AES} = 2^{128} = 3,4 \times 10^{38}$  ключей. Система AES имеет в  $(\frac{K_{AES}}{K_{DES}}) \approx 10^{21}$  раз больше ключей, чем DES. Если предположить, что можно проверить все ключи DES за одну секунду, то при такой скорости  $R = 2^{56} \frac{\text{бит}}{\text{с}}$ , для тестирования всех ключей блочного шифрования AES потребовалось 149 триллионов лет (возраст вселенной 13,7 миллиарда лет).

Недостатком блочного шифрования является обмен ключами между отправителем и получателем. Передача ключей в практическом аспекте уязвима для перехвата.

## 6.8. Ассиметричные алгоритмы шифрования

Теоретико-числовые алгоритмы являются основой современной криптографии и криптографических систем с открытым ключом. В этом случае программное обеспечение, программный интерфейс криптографических систем с шифрующими алгоритмами строится на базе конечных алгебраических структур – групп, колец, полей.

Первой известной криптографической системой с открытым ключом является система, созданная в Массачусетском технологическом институте в 1978 году. Система известна и названа по фамилиям авторов (R. L. Rivest, A. Shamir, L. Alldeman) как RSA-криптосистема. Особенностью математического алгоритма системы является то, что криптосистему создает не отправитель сообщения, а получатель. Алгоритм шифрования основывается на задаче RSA.

### 6.8.1. Задача RSA

Предположим, что произвольный получатель А информации разрешает всем желающим передавать ему секретные сообщения. Т. е. он выступает в качестве получателя сообщения. Получатель А случайным образом выбирает

два больших простых числа  $p$  и  $q$ , причем  $p \neq q$ . Числа  $p$  и  $q$  выбираются порядка не менее чем  $2^{256}$ . Эти числа являются секретными.

Получатель А вычисляет число  $N = p \cdot q$ . Число  $N$  называется модулем алгоритма и является несекретным. Таким образом, всем пользователем системы известно число  $N$ , но не известны сомножители –  $p$  и  $q$ .

Решение задачи RSA (дешифрации) сведется к поиску простых делителей  $p$  и  $q$  числа  $N$ . Криптостойкость системы обосновывается сложностью решения задачи факторизации очень больших чисел в произведение простых чисел.

Алгоритм RSA использует понятие функции Эйлера числа  $N$  и теорему Эйлера.

*Определение 6.12.* Количество положительных целых чисел меньших  $M$  и взаимно простых с  $M$  называется функцией Эйлера  $\varphi(M)$ , или тотиент-функцией Эйлера. Функция Эйлера  $\varphi(M)$  – это количество вычетов по модулю  $M$ .

**Теорема 6.1.** Если  $p$  – простое число, то  $\varphi(p) = p - 1$ .

Так как пара простых чисел  $p$  и  $q$  известна получателю А, используя свойство мультипликативности функции Эйлера, он легко может вычислить значение функции  $\varphi(N)$ .

$$\varphi(N) = \varphi(p \cdot q) = \varphi(p)\varphi(q) = (p - 1)(q - 1).$$

Получатель А публикует также число  $E$ , т. е.  $E$  является несекретным. Число  $E$  выступает в качестве открытого ключа криптосистемы RSA и используется для шифрования данных. Число  $E$  называется шифрующей экспонентой. Выбор получателем А числа  $E$  должен удовлетворять двум условиям:

$$1 < E \leq \varphi(N) = (p - 1)(q - 1); \quad (6.1)$$

$$\text{НОД}(E, \varphi(N)) = \text{НОД}(E, (p - 1), (q - 1)) = 1. \quad (6.2)$$

Следовательно, число  $E$  и число  $\varphi(N)$  должны быть взаимно простыми. Число  $E$  выбирается случайным образом, но часто  $E$  равно числу Ферма

$$E = 2^{2^t} + 1, t \in Z_N.$$

Например, открытый ключ  $E$  может быть равен таким числам Ферма:

$$5, 17, 257, 65537, \dots$$

*Определение 6.13.* Пара  $(E, N)$  называется открытым ключом RSA (RSA public key).

**Теорема 6.2.** (Теорема Эйлера). Если  $\alpha$  и  $M$  – взаимно простые числа, т. е.  $\text{НОД}(\alpha, M) = 1$ , то

$$\alpha^{\varphi(M)} \equiv 1 \pmod{M}. \quad (6.3)$$

Получатель А пересылает отправителю В пару чисел  $(E, N)$  по несекретному (незащищенному) каналу. Таким образом, всем желающим передавать получателю А секретную информацию доступна пара  $(E, N)$ .

Исходный текст  $m$  переводится в числовую форму (шифруется) по формуле

$$m^E \equiv C \bmod N. \quad (6.4)$$

В результате текст представляется криптограммой  $C = Z_N \in \{1, 2, \dots, N - 1\}$  в виде одного большого числа. Затем это число разбивается на блоки так, что каждый из них представляется в виде числа  $m_i \in \{0, 1, 2, \dots, N - 1\}$ . Метод (6.4) шифрования (кодирования) текста в числовую форму является несекретным.

Так как по алгоритму RSA число  $E$  и число  $\varphi(N)$  должны быть взаимно простыми, по теореме Эйлера справедливо выражение

$$E^{\varphi(N)} \equiv 1 \bmod \varphi(N) = E^{\varphi((p-1)(q-1))} \equiv 1 \bmod (p-1)(q-1). \quad (6.5)$$

Обозначим функцию Эйлера  $\varphi(\varphi(N)) = \varphi((p-1)(q-1))$  через  $x$ , тогда выражение (6.5) запишем в виде

$$E^x \equiv 1 \bmod (p-1)(q-1). \quad (6.6)$$

Выражение (6.6) можно записать в виде

$$E \cdot E^{x-1} \equiv 1 \bmod (p-1)(q-1). \quad (6.7)$$

Обозначим  $E^{x-1} = d$ . С учетом этого получаем

$$E \cdot d \equiv 1 \bmod (p-1)(q-1), \quad (6.8)$$

где  $d$  – это секретный ключ.

Секретный ключ  $d$ , применяется для дешифрования криптограммы по формуле

$$m \equiv C^d \bmod N.$$

Таким образом, секретными данными RSA системы является тройка чисел  $(d, p, q)$ .

Нахождение  $d$  получателем А для дешифрации криптограммы сводится к вычислению числа обратному числу  $E$ . Необходимо найти такое число  $d \in \{1, 2, \dots, N - 1\}$  для которого выполняется сравнение (6.8). Так как получателю А известны числа  $p, q, E$ , сравнение разрешимо единственным образом, поскольку  $\text{НОД}(E, (p-1)(q-1)) = 1$ .

*Замечание*

1. Число  $d$  можно легко вычислить, используя алгоритм Евклида или другие быстрые алгоритмы нахождения обратных чисел.

### 6.8.1.1. Алгоритм вычисления обратных чисел по теореме Эйлера

Используя формулу (6.3), можно записать

$$\alpha^{-1} \cdot \alpha^{\varphi(M)} \equiv \alpha^{-1} \cdot 1 \pmod{M}.$$

Тогда число, обратное числу  $\alpha$  равно

$$\alpha^{-1} \equiv \alpha^{\varphi(M)-1}. \quad (6.9)$$

Эффективное применение выражения (6.9) требует нахождения значения порядка  $n$  элемента  $\alpha$ , когда выполняется

$$\alpha^n \equiv 1 \pmod{M}.$$

**Теорема 6.3.** Порядок  $n$  числа  $\alpha$  должен быть делителем функции Эйлера  $\varphi(M)$ .

*Пример 6.1.* Пусть  $\alpha = 5$ ,  $M = 31$ . Найти  $\alpha^{-1}$ .

Решение. Так как  $M$  – простое число,  $\varphi(M) = M - 1 = 30$ .

$$\alpha^{-1} = ((\alpha^{\varphi(M)-1})) = ((\alpha^{(M-1)-1})) = ((\alpha^{30-1})) = \alpha^{29}.$$

Находим порядок числа 5:

$$5^3 = 125 \equiv 1 \pmod{31}, n = 3. \text{ Число 3 является делителем числа } \varphi(31) = 30.$$

Обратное число

$$5^{-1} = 5^{29} = 5^{27} \cdot 5^2 = ((5^3)^9 \cdot 5^2) = 25.$$

Действительно,

$$5 \cdot 5^{-1} = 5 \cdot 25 \equiv 1 \pmod{31}.$$

*Пример 6.2.* Пусть  $\alpha = 6$ ,  $M = 31$ . Найти  $6^{-1}$ .

Решение. Функция Эйлера  $\varphi(M) = M - 1 = 30$ .

Обратное число  $6^{-1} = ((\alpha^{30-1})) = 6^{29}$ .

Хотя ближайшие числа 2, 3, 5, делят число 30, они не являются порядком  $n$  числа 6. Только число 6 является порядком  $n$  числа 6. Действительно

$$6^6 = 46656 \equiv 1 \pmod{31}.$$

Обратное число

$$6^{-1} = ((6^{29})) = ((6^{24} \cdot 6^5)) = ((6^6)^4 \cdot 6^5) = 6^5 = 7776 \equiv 26 \pmod{31}.$$



Действительно,

$$6 \cdot 26 \equiv 1 \pmod{31}.$$

### 6.8.2. Последовательность шагов криптоалгоритма RSA

*Пример 6.3.* Зашифровать слово РУХ с помощью алгоритма RSA.

Решение. Действия получателя *A* шифрованной информации.

1. Выбираем  $p = 3$  и  $q = 7$ .

2. Вычисляем модуль  $N = pq = 3 \cdot 7 = 21$ .

3. Вычисляем значение функции Эйлера для  $N = 21$ ,

$$\varphi(N) = \varphi(21) = (p - 1)(q - 1) = 2 \cdot 6 = 12.$$

4. Выбираем в качестве открытого ключа  $E$  произвольное число с учетом выполнения условий:

$$1 < E \leq \varphi(N); \text{НОД}(E, \varphi(N)) = 1, \text{НОД}(E, 12) = 1. \text{ Пусть } E = 11.$$

5. Из выражения (6.8) вычисляем значение секретного ключа  $d$ , используя алгоритм нахождения обратных чисел по теореме Эйлера:

$$E \cdot d \equiv 1 \pmod{\varphi(N)},$$

$$11 \cdot d \equiv 1 \pmod{12}.$$

$$d = E^{-1} \equiv 11^{-1} \pmod{12}.$$

Число является 2 порядком  $n$  числа 11, так как соблюдается условие

$$\alpha^n \equiv 1 \pmod{M}.$$

$$11^2 = 121 \equiv 1 \pmod{12}.$$

Из алгоритма нахождения обратных чисел по теореме Эйлера

$$\alpha^{-1} = ((\alpha^{\varphi(M)-1}))$$

сначала находим  $\varphi(12) = 4$ .

Число, обратное  $E$  равно

$$d = E^{-1} = ((E^{\varphi(12)-1})) = ((E^3)),$$

$$11^{-1} = ((11^3)) = ((11^2 \cdot 11)) = 11.$$

Проверим правильность полученного значения  $d$ :

$$E \cdot d \equiv 1 \bmod 12 = 11 \cdot 11 \equiv 1 \bmod 12.$$

6. Получатель А зашифрованной информации пересылает отправителю пару открытых чисел ( $N = 21, E = 11$ ).

*Действия отправителя зашифрованной информации.*

7. Представляем шифруемое сообщение РУХ в виде последовательности целых чисел в диапазоне  $M = 0, 1, \dots, N - 1$ .

Пусть буква Р записывается числом 1, буква У – числом 2, буква Х – числом 3. Сообщению РУХ соответствует последовательность (блоки) чисел 123. Текст сообщения в виде блоков записывается как  $m_1 m_2 m_3$ , где

$$m_1 = 1, m_2 = 2, m_3 = 3.$$

8. Шифруем текст, используя ключ  $E = 11$  и  $N = 21$  по формуле (6.4)

$$C_i \equiv m_i^E \bmod N \equiv m_i^{11} \bmod 21.$$

Шифротекст состоит из следующих чисел:

$$C_1 = 1^{11} \bmod 21 \equiv 1;$$

$$C_2 \equiv 2^{11} \bmod 21 \equiv 2048 \bmod 21 \equiv 11;$$

$$C_3 \equiv 3^{11} \bmod 21 \equiv 12.$$

*Действия получателя по дешифрации криптограммы*

9. Дешифрация криптограммы  $C_1 C_2 C_3 = 1, 11, 12$  производится с использованием секретного ключа  $d = 11$  по формуле

$$m_i = C_i^d \bmod N = C_i^{11} \bmod 21.$$

В результате получаем:

$$m_1 = 1^{11} \bmod 21 \equiv ((1));$$

$$m_2 = 11^{11} \bmod 21 \equiv ((2));$$

$$m_3 = 12^{11} \bmod 21 \equiv ((3)).$$

Исходное сообщение: РУХ.

Упражнения

6.1. Пусть  $\alpha = 2, M = 9$ . Найти порядок элемента  $\alpha$ .

6.2. Пусть  $\alpha = 4, M = 17$ . Найти порядок элемента  $\alpha$ .

6.3. Пусть  $\alpha = 7, M = 17$ . Найти  $7^{-1}$ .

6.4. Пусть  $\alpha = 10, M = 23$ . Найти  $10^{-1}$ .

6.5. Создайте RSA-криптосистему, используя  $N = 2 \cdot 13$ , вычислив публичный ключ  $(N, E)$  и секретный ключ  $d$ . Используйте эти ключи, для кодирования сообщения «МІНСК».

6.6. Создайте RSA-криптосистему, используя  $N = 3 \cdot 17$ , вычислив публичный ключ  $(N, E)$  и секретный ключ  $(d, N)$ . Используйте эти ключи, для кодирования сообщения «ІІТ».

#### Литература

1. Кудряшов, Б. Д. Теория информации: Учебник для вузов. – СПб.: Питер, 2009.
2. Теория прикладного кодирования: Учеб. пособие. В 2т./ В. К. Конопелько, А. И. Митюхин и др.; Под ред. проф. В. К. Конопелько.– Мн.: БГУИР, 2004.
3. Ватолин Д, Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. – М. : Диалог-МИФИ, 2002.
4. Луенбергер Д. Дж. Информатика.– Москва: Техносфера, 2008.
5. Митюхин, А. И., Пачинин В.И. Элементы алгебраических структур теории кодирования: учеб. пособие / А. И. Митюхин, Пачинин В. И. – Минск: БГУИР, 2012.
6. Вернер М. Основы кодирования. Учебник для вузов. Москва: Техносфера, 2004.
7. Андерсон Дж. А. Дискретная математика и комбинаторика: Пер. с англ.– М.: Вильямс, 2004.
8. Лидл Р., Нидеррайдер Г. Конечные поля: В 2т. – М.: Мир, 1988.
9. Хаггарти Р. Дискретная математика для программистов. Москва: Техносфера, 2005.