МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Кузбасский государственный технический университет имени Т. Ф. Горбачева»

Кафедра информационной безопасности

Составители
Е. В. Прокопенко
И. В. Чичерин

ТЕОРИЯ ИНФОРМАЦИИ

Методические материалы

Рекомендованы учебно-методической комиссией специальности 10.05.03 Информационная безопасность автоматизированных систем в качестве электронного издания для использования в образовательном процессе

Кемерово 2018

Рецензенты

Стенин Д. В.– кандидат технических наук, доцент директор ИИТМА
Сыркин И. С. – кандидат технических наук, доцент кафедры информационных и автоматизированных производственных систем

Прокопенко Евгения Викторовна
Чичерин Иван Владимирович

Теория информации: методические материалы [Электронный ресурс] для обучающихся специальности 10.05.03 Информационная безопасность автоматизированных систем очной формы обучения / сост. Е. В. Прокопенко, И. В. Чичерин; КузГТУ. – Электрон. издан. – Кемерово, 2018.

Введение

Теория информации была создана и развивалась как наука, направленная на решение проблем связи. В настоящее время теория информации является составной частью кибернетики, которая, по выражению А.Н. Колмогорова, «занимается изучением систем любой природы, способных воспринимать, хранить, перерабатывать информацию и использовать ее для управления и регулирования».

На сегодняшний день достаточно четко определилась прикладная сторона теории информации – информационная техника, направленная на использование основных положений теории при создании конкретных технических устройств. К информационной технике относятся средства, служащие для восприятия, подготовки, кодирования, передачи, переработки, хранения и представления информации, поступающей от объекта наблюдения.

Теория информации и информационная техника являются сравнительно новыми отраслями, получающими наибольшее развитие на этапе применения вычислительной техники и систем управления, ибо без информации нет управления.

1. Теория информации как наука. Источники сообщений

Теория информации – раздел прикладной математики, радиотехники (теория обработки сигналов) и информатики, относящийся к измерению количества информации, её свойств и устанавливающий предельные соотношения для систем передачи данных. Как и любая математическая теория, теория оперирует математическими моделями, а не реальными физическими объектами (источниками и каналами связи). Использует, главным образом, математический аппарат теории вероятностей и математической статистики.

Основные разделы теории информации – кодирование источника (сжимающее кодирование) и канальное (помехоустойчивое) кодирование. Теория информации тесно связана с информационной энтропией, коммуникационными системами, криптографией и другими смежными дисциплинами.

Теория информации началась с работ, написанных в конце двадцатых годов XX столетия. Еще в 1928 г. американский ученый Р. Хартли предложил логарифмическую меру оценки количества информации. В 1933 г. была опубликована работа советского ученого В. А. Котельникова, в которой было фактически заложено начало общей теории передачи сообщений.

Наиболее бурное развитие теория информации получила после опубликования в 1947–1948 гг. классических работ американского математика и инженера К. Шеннона. Большой вклад внесли в развитие теории информации американский ученый Н. Винер, советские ученые А.Я. Хинчин, А.Н. Колмогоров.

Основоположник теории информации Клод Шеннон определил информацию, как снятую неопределенность. Точнее сказать, получение информации – необходимое условие для снятия неопределенности. Неопределенность возникает в ситуации выбора. Задача, которая решается в ходе снятия неопределенности – уменьшение количества рассматриваемых вариантов (уменьшение разнообразия), и в итоге выбор одного соответствующего ситуации варианта из числа возможных. Снятие неопределенности дает возможность принимать обоснованные решения и действовать. В этом управляющая роль информации.

Представьте, что вы зашли в магазин и попросили продать вам жевательную резинку. Продавщица, у которой, скажем, 16 сортов жевательной резинки, находится в состоянии неопределенности. Она не может выполнить вашу просьбу без получения дополнительной информации. Если вы уточнили, скажем, – «*Orbit*», и из 16 первоначальных вариантов продавщица рассматривает теперь только 8, вы уменьшили ее неопределенность в два раза (забегая вперед, скажем, что уменьшение неопределенности вдвое соответствует получению 1 бита информации). Если вы просто указали пальцем на витрине, – «вот эту!», то неопределенность была снята полностью.

Ситуация максимальной неопределенности предполагает наличие нескольких равновероятных альтернатив (вариантов), т.е. ни один из вариантов не является более предпочтительным. Причем, чем больше равновероятных

вариантов наблюдается, тем больше неопределенность, тем сложнее сделать однозначный выбор и тем больше информации требуется для этого получить.

Минимальная неопределенность равна 0, т.е. эта ситуация полной определенности, означающая что выбор сделан, и вся необходимая информация получена. Распределение вероятностей для ситуации полной определенности выглядит так: {1, 0, …0}.

Величина, характеризующая количество неопределенности в теории информации обозначается символом *H* и имеет название энтропия, точнее информационная энтропия.

Энтропия (*H*) – мера неопределенности, выраженная в битах. Так же энтропию можно рассматривать как меру равномерности распределения случайной величины.

Количество информации *I* и энтропия *H* характеризуют одну и ту же ситуацию, но с качественно противоположенных сторон. *I* – это количество информации, которое требуется для снятия неопределенности *H*. По определению Леона Бриллюэна информация есть отрицательная энтропия (негэнтропия).

Когда неопределенность снята полностью, количество полученной информации *I* равно изначально существовавшей неопределенности *H*.

При частичном снятии неопределенности, полученное количество информации и оставшаяся неснятой неопределенность составляют в сумме исходную неопределенность.

Система передачи информации – совокупность технических средств для передачи информации от источника к приемнику информации.

Любая ЭВМ также представляет собой информационную систему, которую можно рассматривать как канал преобразования информации.

Задачей системы передачи информации является воспроизведение с заданной точностью наиболее экономичным методом сообщения выработанного источником информации.

От источника к приемнику информация передается в виде сообщений.

Сообщение – это предназначенное для передачи высказывание, текст, изображение, физический предмет или поступок.

Сообщение выступает в качестве материальной оболочки для представления информации при передаче.

Источник информации выдает сообщение или последовательность сообщений. Сообщения могут иметь непрерывный или дискретный характер.

Дискретными называются сообщения, которые представляются последовательностью из конечного числа отдельных, резко различимых элементов, между которыми нет промежуточных значений, т. е. дискретная информация представляется в виде конечной совокупности символов (печатные тексты и документы, состояния цифровых автоматов и т. д.).

Непрерывные сообщения характеризуются тем, что два нетождественных сообщения могут отличаться, сколь угодно мало друг от друга. Непрерывные сообщения можно преобразовать в дискретные.

Для передачи на расстояние сообщение преобразуется в сигнал. Процесс преобразования сообщения в сигнал состоит из трех этапов (операций): преобразование, кодирование, модуляция. В процессе преобразования сообщение, которое может иметь любую физическую природу (изображение, звук и т.д.), преобразуется в первичный сигнал. В телефонии микрофон преобразует звуковые волны (давление) в электрический ток микрофона. В телеметрии датчики преобразуют изменение физических величин (температура, давление и т.д.) в электрические.

Сигнал – это физический процесс, некоторая характеристика которого несёт информационный смысл.

Например, световой сигнал (поток света) характеризуется яркостью, цветом, поляризационными свойствами, направлением распространения и др. Информацию может нести как одна из этих характеристик, так и одновременное сочетание нескольких характеристик.

Сигнал возникает в природе при взаимодействии материальных объектов и несёт в себе информацию об этом взаимодействии. Сигнал способен перемещаться, распространяться в некоторой материальной среде, тем самым обеспечивая пространственный перенос информации от объекта (источника события) к субъекту (наблюдателю). Материальная среда, в которой распространяется сигнал, называется носителем сигнала.

Сигналы различаются, прежде всего, по своей физической природе. Примеры: световой сигнал, звуковой, электрический, радиосигнал и др.

В зависимости от порождающего их источника сигналы бывают естественные или искусственные.

Естественные сигналы возникают в силу того, что где-то в живой или неживой природе взаимодействуют материальные объекты. Это естественный процесс, никак не связанный с деятельностью человека. Примеры: свечение Солнца, пение птиц, распространение запаха цветов.

Искусственные сигналы инициируются человеком или возникают в технических системах, созданных человеком. Примеры: радиосигналы; сигнальная ракета или костёр; сигнал светофора; сирена пожарной машины.

По форме сигналы бывают:

- аналоговые;
- дискретные;
- цифровые.

Аналоговый (или непрерывный) сигнал представляет собой физический процесс, информационная характеристика которого изменяется плавно. Пример: звуковой сигнал, естественный световой сигнал. Практически все естественные сигналы аналоговые.

Особенностью аналогового сигнала является размытость границы между двумя соседними его значениями. Общее число значений, которыми можно характеризовать аналоговый сигнал, бесконечно велико.

Дискретный сигнал представляет собой физический процесс, информационная характеристика которого изменяется скачкообразно и может принимать только некоторый ограниченный набор значений. Особенность дискретного сигнала – это чёткое разграничение между двумя разными значениями сигнала. Общее число возможных значений, которые может принимать дискретный сигнал, всегда ограничено.

Например, лампа, включенная в электрическую цепь. Лампа может либо гореть, либо не гореть. Если лампа горит, это служит сигналом о том, что в цепи есть ток. Если не горит – тока нет. Промежуточные значения (с какой яркостью горит лампа) здесь не учитываются – значений только два: либо горит, либо не горит. Другой пример: по телеграфу передаётся некоторое сообщение. Сообщение передаётся с помощью азбуки Морзе, использующей три разных значения: точка, тире и пробел (пауза). Сигнал, который несёт это сообщение, тоже будет иметь только три разных значения: короткий сигнал, длинный сигнал и отсутствие сигнала. Поскольку количество возможных значений сигнала ограничено – это дискретный сигнал.

Дискретные сигналы, как правило, искусственные (создаются человеком или технической системой).

В современных устройствах, относящихся к вычислительной технике, для передачи информации используется цифровой сигнал.

Цифровой сигнал – это частный случай дискретного сигнала, когда информационная характеристика принимает только два возможных значения: либо есть сигнал, либо нет сигнала.

Устройства, использующие для передачи информации цифровые сигналы, называются цифровыми устройствами. Внутри таких устройств передача производится чаще всего с помощью электрического сигнала. Его два возможных значения: либо нет напряжения (когда передаётся 0), либо есть напряжение определенной величины (когда передаётся 1).

Цифровой сигнал чаще всего передаётся не по одной линии, а по нескольким параллельным линиям. Совокупность параллельных проводящих линий, используемых совместно для передачи одного общего цифрового сигнала, называется цифровой шиной.

Цифровая шина характеризуется разрядностью. Разрядность цифровой шины – это количество бит, передаваемых с её помощью за один раз. Если проводящая линия всего одна, то она позволяет передавать за раз один бит. Если проводящих линий восемь, тогда за раз можно передавать восемь бит – это восьмиразрядная шина. В современных компьютерах используются 8-разрядные, 16-разрядные, 32-разрядные и 64-разрядные шины.

Ввиду того, что вычислительные устройства работают с цифровыми сигналами, а все реальные естественные сигналы, как правило, аналоговые, для возможности взаимодействия цифровых устройств с внешним миром необходимы преобразования аналоговых сигналов в цифровые и, наоборот, – цифровых сигналов в аналоговые. Эти преобразования выполняются с помо-

щью специальных микросхем, называемых АЦП (аналого-цифровой преобразователь) и ЦАП (цифро-аналоговый преобразователь). Например, микросхемы АЦП и ЦАП встраиваются в звуковую карту – специальное внутреннее устройство персонального компьютера, обеспечивающее возможность ввода и вывода звуковой информации.

Процесс изменения свойств некоторого материального объекта под воздействием сигнала называется регистрацией сигнала (например, тепловой сигнал заставляет объект нагреваться, радиосигнал вызывает движение электронов в проводнике и т. д.).

Зарегистрированные сигналы представляют собой данные.

Данные – сведения, полученные путем измерения, наблюдения, логических или арифметических операций и представленные в форме, пригодной для постоянного хранения, передачи и обработки. Иначе говоря, данные – это информация, представленная в формализованном виде и предназначенная для обработки ее техническими средствами, например, ПК.

В процессах сбора, обработки и использования данные расчленяются на отдельные элементарные составляющие – элементы данных (или элементарные данные).

Данные необязательно возникают путём регистрации непосредственно исходного сигнала. Чаще всего исходный сигнал сначала преобразуется в сигнал, другой по природе, но такой же по информационным характеристикам. Например, звуковой сигнал, чтобы его можно было записать на магнитофонную ленту, сначала преобразуется в электрический сигнал, а затем в магнитный.

Примеры данных:

– фотоснимок – результат регистрации светового сигнала, излучаемого или отражённого от изучаемых объектов;

– запись на бумаге – результат регистрации мыслей человека (мысли можно рассматривать как множество электрических сигналов, возникающих в нервной системе человека); при этом электрические сигналы нервной системы с помощью мышц руки преобразуются в механическое перемещение карандаша или ручки;

– записанная на магнитной ленте речь человека – результат регистрации звукового сигнала; при этом в качестве средства регистрации сигнала используется магнитофон;

– данные, записанные на дискету, жёсткий диск, лазерный диск или магнитооптический диск, на флэш-память или в оперативную память; и др.

Данные – это пока ещё не информация. Это просто какая-то запись.

Данные могут стать информацией, если к ним применить методы чтения и интерпретации, которые бы позволили вскрыть содержащийся в данных смысл и использовать его для решения той или иной задачи.

Метод чтения должен соответствовать материальному носителю, на котором записаны данные. Например, если данные записаны на бумаге, тогда

нужно включить свет, посмотреть на бумагу, найти буквы и считывать их слева направо, собирая из них слова.

Чтобы получить полную и адекватную информацию из данных, необходимо их не только прочитать, но и правильно интерпретировать (трактовать). Например, считываемые нами слова должны правильно сопоставляться с имеющимися у нас понятиями. Если звуковой файл передать программе, воспроизводящей текст, тогда мы получим неадекватную информацию (вместо музыки несуразный текст из беспорядочно набросанных символов).

Кодирование – преобразование сообщения в сигнал, т.е. отображение сообщений сигналами в виде определенного сочетания элементарных дискретных символов, называемых кодовыми комбинациями (кодовыми словами).

Код – правило, согласно которому каждому сообщению однозначно ставится в соответствие некоторая кодовая комбинация. Кодер – устройство, осуществляющее кодирование.

Кодер источника (КИ) – кодер, использование которого позволяет путем устранения избыточности существенно уменьшить среднее число символов на букву сообщения (такое кодирование называется оптимальным или эффективным). При отсутствии помех это дает выигрыш во времени передачи или в объеме ЗУ, т. е. повышает эффективность системы передачи данных.

Кодер канала (КК) – позволяет путем внесения избыточности обеспечить достоверность передачи данных при наличии помех (такое кодирование называется помехоустойчивым).

Канал – совокупность средств, предназначенных для передачи сигнала от передатчика к приемнику информации (передатчик, приемник, линия связи и т. д.). Канал связи может быть односторонний (симплексный) и двухсторонний (дуплексный).

Передатчик – служит для преобразования электрического сигнала в сигнал, пригодный для передачи по линии связи.

Модуляцией называется изменение параметров переносчика сигнала в соответствии с функцией, отображающей сообщение. Несущим сигналом может быть ток (телеграфия), гармонические низкочастотные или высокочастотные колебания (телефония и т. д.), высокочастотные импульсы (радиорелейная связь и т. д.). Модулируемые параметры называются информативными и могут быть амплитудой, частотой, фазой и т. д. Модулятор – устройство, осуществляющее модуляцию.

При передаче по каналу связи происходит ослабление и искажение передаваемого сигнала, вносимых каналом и действием помех.

Линейные искажения – определяются частотными и временными характеристиками канала. Нелинейные искажения – определяются нелинейностью звеньев канала и видом модуляции.

Линия связи (ЛС) – это среда, используемая для передачи сигнала от передатчика к приемнику.

Проводные линии связи могут быть воздушными, кабельными, коаксиальными, оптико-волоконными, линиями электропередачи (ЛЭП). Они используются:

– в телефонии – 300–3400 Гц (тональный диапазон);
– в телеграфии – 0–300 Гц (под тональный диапазон);
– в телевидении – 300–3000 мГц;
– ЛЭП – 500–1000 кГц.

Проводные линии связи характеризуются: помехозащищенностью и волновым сопротивлением.

При передаче на большие расстояния необходимо использование промежуточных усилительных пунктов (расстояние зависит от используемых частот и типа ЛС).

Радиолинии могут быть: радиорелейными (РРЛ), коротковолновыми (КВ), тропосферными, ионосферными, космическими и т.д.

Внутриаппаратные тракты – это прежде всего шины информационного обмена в ЭВМ и тракты магнитной записи.

Многоканальные ЛС – обеспечивают несколько каналов, используя различные методы уплотнения и разделения (частотного, временного, кодового).

Помехи – воздействия, искажающие сигнал. Помехи можно классифицировать:

1) Детерминированные (регулярные) – например, фон источника питания и случайные – например, тепловой шум и т.д.

2) Внутренние, возникающие в самой аппаратуре и внешние.

К внешним помехам относятся:

– атмосферные грозовые разряды, космические излучения;
– индустриальные (электротехнические, связанные с коммутацией, сварка, транспорт и т. д.);
– интерферентные (глушители).

3) Аддитивные – которые суммируются с основным сигналом и мультипликативные – которые перемножаются с полезным сигналом.

Приемник осуществляет прием сигнала и его демодуляцию. Демодуляция – отделение полезного сигнала от несущей. Демодулятор – устройство для отделения модулирующего сигнала от несущей.

На выходе приемника получается последовательность кодовых комбинаций, которая вследствие действия помех и наличия искажений может отличаться от переданных комбинаций.

Декодер (декодирующее устройство) преобразует кодовые комбинации в сообщения, поступающие получателю.

Основные проблемы систем передачи информации:

- Обеспечение достоверности передаваемых сообщений (помехоустойчивость).
- Обеспечение высокой эффективности передачи сообщений.

Помехоустойчивость – способность системы противостоять вредному действию помех и искажений. Повышению помехоустойчивости способствует увеличение соотношения сигнал – помеха, выбором метода кодирования (помехоустойчивое кодирование), вида модуляции (искажения), передающей среды и т. д.

Эффективность определяется способностью системы обеспечить передачу заданного количества информации с наименьшими затратами мощности сигнала, времени и полосы частот (т. е. наиболее экономичным способом).

Система передачи данных – совокупность технических средств, обеспечивающих передачу данных.

Первые сети передачи данных появились в начале 50-х годов, когда линии связи соединяли центральные ЭВМ с удаленными терминалами и другими периферийными устройствами. Стремительное увеличение сетей передачи данных было обусловлено созданием вычислительных систем большой производительности с разделением времени.

В связи с широким использованием ЭВМ в различных сферах существенно возрастает потребность в передаче данных. Их применение позволяет использовать:

– удаленный доступ к базам данных (БД) и их обновлению (например: информационные и финансовые службы, продажа авиабилетов и т. д.);
– электронную почту;
– использование удаленных мощных ЭВМ;
– управление объектами в реальном времени и т. д.

Система передачи данных включает:

– аппаратуру передачи данных (АПД);
– модуль управления линией передачи данных (УЛПД);
– модем – модулятор и демодулятор, объединенные в одном устройстве;
– интерфейс – унифицированная система сопряжения.

Аппаратура передачи данных предназначена для преобразования передаваемой дискретной информации (данных) в сигналы, пригодные для передачи по каналам связи, и принятой информации к виду пригодному для обработки на приемной стороне.

Сообщения представляют длинные последовательности битов, которые обычно разбиваются на более короткие последовательности, называемые пакетами. Передающий модуль УЛПД преобразует пакет, представляющий поле данных, в кадр, помещая в него ряд управляющих бит, определяющих начало и конец кадра, адрес источника и приемника, проверочные последовательности (полином 32-го порядка), позволяющие обнаружить ошибки в принятых кадрах, запрос на повторную передачу, если обнаружены ошибки и др.

Принимающий модуль УЛПД на конечном пункте из пакетов собирает сообщения.

Системы передачи данных можно объединять в сети по иерархическому принципу, т. е. в многоуровневую систему. Существуют глобальные и локальные вычислительные сети. Глобальная сеть – сеть, покрывающая площадь, которая больше площади города. Локальная сеть (ЛС) – вычислительная сеть, в которой компьютеры и терминалы расположены в географически ограниченном пространстве, чаще всего в пределах одной организации, учреждения, учебного заведения и т. д.

Характеристиками локальной сети являются: топология (шинная, кольцевая, звездная и древовидная), метод доступа (случайный или детерминированный), физическая передающая среда (витая пара, коаксиальный или оптоволоконный кабель) и др.

Выбор передающей среды определяется необходимой пропускной способностью.

Источником сообщений называют устройство, генерирующее сообщение из ансамбля возможных сообщений. То есть источником сообщений может быть объект, состояние которого определяется некоторым физическим процессом, протекающим во времени или в пространстве по случайному (заранее не известному нам) закону. При передаче и преобразовании информации, как правило, происходит преобразование пространственного распределения во временное.

В зависимости от вида представления сообщения различают дискретные источники, непрерывные источники непрерывного времени и непрерывные источники дискретного времени.

Источник дискретных сообщений формирует дискретные последовательности из ограниченного числа элементарных сообщений.

Для источника непрерывных сообщений характерным является непрерывное изменение во времени или в пространстве физического параметра, значение которого измеряется или передается на расстояние для получения информации о том или ином явлении, факте, процессе и т. п.

Источник сообщений в теории информации полностью определяется статистическими данными о формируемых им сообщениях.

Основными характеристиками любого источника сообщений являются производительность и избыточность.

Под производительностью источника сообщений подразумевают количество информации, вырабатываемое источником в единицу времени. Эту характеристику источника называют также скоростью создания сообщений или потоком входной информации.

При работе источника сообщений на его выходе отдельные символы появляются через некоторые промежутки времени; в этом смысле мы можем говорить о длительности отдельных символов.

Поскольку возможное воздействие помех на источник сообщений принято учитывать эквивалентным изменением характеристик модели канала связи, то производительность источника сообщений равна энтропии источника, приходя-

щейся на единицу времени. Так определяется скорость создания информации при передаче дискретных (по информативному параметру) сигналов.

Избыточность источника информации можно определить через его энтропию. Как мы знаем, энтропия характеризует среднее количество информации, несомое одним символом источника. Она максимальна, когда символы вырабатываются источником с равной вероятностью. Если же некоторые символы появляются чаще других, энтропия уменьшается, а при появлении дополнительных вероятностных связей между символами становится еще меньшей. Чем меньше энтропия источника отличается от максимальной, тем рациональнее он работает, тем большее количество информации несут его символы.

Одной из важнейших характеристик сигнала, вырабатываемого источником, является содержащееся в нем количество информации.

В дискретном случае имеются две причины избыточности: неравновероятность символов и наличие статистической связи между символами. В непрерывном случае – это неэкстремальность распределений (т. е. отклонение от распределений, обладающих максимальной энтропией), что в широком смысле сводится к отклонениям от экстремальности распределения первого порядка и от минимальности связи во времени (от равномерности спектра при его ограниченности).

Хотя с точки зрения наиболее экономичного и эффективного использования информационных систем естественно сводить избыточность сигналов до минимума, не следует думать, что избыточность – явление, играющее лишь отрицательную роль. Наоборот, именно избыточность обеспечивает информационную устойчивость сигналов при воздействии помех. При искажениях, выпадениях и вставках символов именно избыточность позволяет обнаружить и исправить ошибки.

Характер последовательностей, формируемых реальным источником сообщений, зависит от существующих ограничений на выбор знаков. Они выражаются в том, что вероятности реализации знаков различны и между ними существуют корреляционные связи. Эти ограничения приводят к тому, что вероятности формируемых последовательностей существенно различаются.

Существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Например, если монета несимметрична (одна сторона тяжелее другой), то при ее бросании вероятности выпадения «орла» и «решки» будут различаться.

Формулу для вычисления количества информации в случае различных вероятностей событий предложил К. Шеннон в 1948 году. В этом случае количество информации определяется как:

$$I = -\sum_{i=1}^{N} p_i \log_2 p_i, \qquad (1)$$

где I – количество информации;

N – количество возможных событий;

p_i – вероятность i-го события.

Этот подход к определению количества информации называется вероятностным.

Для частного, но широко распространенного и рассмотренного выше случая, когда события равновероятны ($p_i = 1/N$), величину количества информации I можно рассчитать по формуле:

$$I = -\sum_{i=1}^{N} \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N. \qquad (2)$$

Количество информации, которое мы получаем, достигает максимального значения, если события равновероятны.

Понятие условной энтропии связано с понятием энтропии марковского процесса – иначе говоря, изначально рассматривается вероятность появления некоторого сообщения, с учётом того, что известно о появлении некоторого другого сообщения.

Если рассматриваются две статистически зависимые величины, то энтропия объединения двух статистически связанных переменных величин равна сумме энтропии первого ансамбля и энтропии второго относительно первого.

Свойства условной энтропии:

1) Условная энтропия всегда меньше или равна безусловной для той же функции распределения вероятностей, это связано с тем, произведением вероятностей.

2) Если имеется однозначная связь между появлением некоторой последовательности сообщений (условием) и появлением некоторого другого сообщения, то условная энтропия при наличии данного условия становится равной нулю.

3) Если статистические связи между двумя переменными величинами отсутствуют, то условная энтропия их объединения не будет меньше безусловной энтропии их объединения.

Объединение – совокупность двух и более ансамблей дискретных, случайных событий. С объединением связаны понятия условной, безусловной, совместной и взаимной энтропии.

2. Эффективное и помехоустойчивое кодирование информации

Эффективное кодирование решает задачу более компактной записи сообщений, вырабатываемых источником за счет их перекодировки. И применяется практически во всех архиваторах типа *Rar*, *Zip* и др. Особенностью этих архиваторов является то, что они позволяют сжать информацию в относительно небольшое число раз (в 2-3, *max* в 4 раза), но при этом происходит полное восстановление сжатой информации «бит в бит». Если же не требуется восстановление информации «бит в бит», то применяются другие методы перекодировки, позволяющие достичь сжатия в десятки раз. Они основываются на изучении закономерностей создания сообщений источником, изучении свойств самого источника и понимания того, насколько необходимо сохранять начальную информацию для потребителя. Например, при передаче речи можно не передавать ее «бит в бит», а допускать искажения, которые получатель голосового сообщения просто не заметит из-за нечувствительности слухового аппарата человека к этим изменениям. При этом сохранится и разборчивость речи, и узнаваемость голоса, и ее эмоциональная окраска. Частичная потеря этих качеств увеличивает ее сжатие. Еще раз подчеркнем, что эффективное кодирование это сжатие и восстановление информации «бит в бит».

Кодирование – в широком смысле слова – это представление сообщений в форме, удобной для передачи по данному каналу.

Операция, обратная кодированию, называется декодирование.

Поскольку дискретные сообщения складываются из букв, а непрерывные сообщения можно представить последовательностью цифр в каждый момент отсчета, то имеется принципиальная возможность обойтись конечным числом образцовых сигналов, соответствующих отдельным буквам алфавита источника.

При большом объеме алфавита прибегают к представлению букв в другом алфавите с меньшим числом букв, которые будем называть символами.

Поскольку алфавит символов меньше алфавита букв, то каждой букве соответствует некоторая последовательность символов, называемая кодовой комбинацией. Число символов в кодовой комбинации называется ее значностью.

В процессе преобразования букв в символы может преследоваться несколько целей:

1. Первая из них заключается в том, чтобы преобразовать информацию в такую систему символов (код), чтобы он обеспечивал простоту и надежность аппаратурной реализации информационных устройств, т.е.:

- простоту аппаратуры различения отдельных символов;
- минимальное время передачи;
- минимальный объем запоминающего устройства при хранении;
- простоту выполнения в принятой системе арифметических и логических действий.

Статистические свойства источника сообщений и помех в канале связи при этом не принимаются во внимание.

Техническая реализация процесса кодирования в таком простейшем виде при непрерывном входном сигнале осуществляется аналого-кодовыми (цифровыми) преобразователями.

2. Второй целью кодирования является на основании теорем Шеннона – согласование свойств источника сообщений со свойствами канала связи.

Так называемый кодер источника (КИ) имеет целью обеспечить такое кодирование, при котором путем устранения избыточности существенно уменьшается среднее число символов, требующееся на 1 букву сообщения.

При отсутствии помех это непосредственно дает выигрыш во времени передачи или в объеме запоминающего устройства, т.е. повышает эффективность системы. Такое кодирование получило название эффективное кодирование.

При наличии помех в канале эффективное кодирование позволяет преобразовать входную информацию в последовательность символов, наилучшим образом подготовленную для дальнейшего преобразования (максимально сжатую).

Так называемый кодер канала (КК) имеет целью обеспечить заданную достоверность при передаче или хранении информации путем дополнительного внесения избыточности, но уже по простым алгоритмам и с учетом статистических закономерностей помехи в канале связи. Такое кодирование получило название помехоустойчивого.

Целесообразность устранения избыточности сообщения методами эффективного кодирования с последующим перекодированием помехоустойчивым кодом обусловлено тем, что избыточность источника сообщения в большинстве случаев не согласована со статистическими закономерностями помехи в канале связи и потому не может быть полностью использована для повышения достоверности принимаемого сообщения, тогда как можно подобрать подходящий для данной помехи помехоустойчивый код.

Кроме того, избыточность сообщения часто является следствием весьма сложных вероятностных зависимостей и позволяет обнаружить и исправить ошибки только после декодирования всего сообщения, пользуясь сложнейшими алгоритмами и интуицией.

Итак, выбор кодирующих и декодирующих устройств зависит от статистических свойств источника сообщений, а так же уровня и характера помех в канале связи.

Если избыточность источника сообщений и помехи в канале связи практически отсутствуют, то введение как кодера источника, так и кодера канала нецелесообразно.

Когда избыточность источника сообщений высока, а помехи малы, целесообразно введение только кодера источника.

Когда избыточность источника мала, а помехи велики, целесообразно введение кодера канала.

При большой избыточности и высоком уровне помех целесообразно введение обоих дополнительных кодирующих и декодирующих устройств.

Теорема Шеннона для каналов с шумом, утверждающая, что при помощи подходящих кодов можно передавать информацию так, чтобы вероятность ошибки после декодирования была произвольно малой при условии, что скорость передачи не превосходит пропускной способности канала связи, неконструктивна: она не указывает способа построения кода. При конструировании кода решающее значение имеет выбор модели возникновения ошибок в передаваемом слове.

Коды могут быть непрерывные и блочные (рис. 1)..

У блочных кодов каждому элементарному символу алфавита, соответствует строго определенная кодовая комбинация.

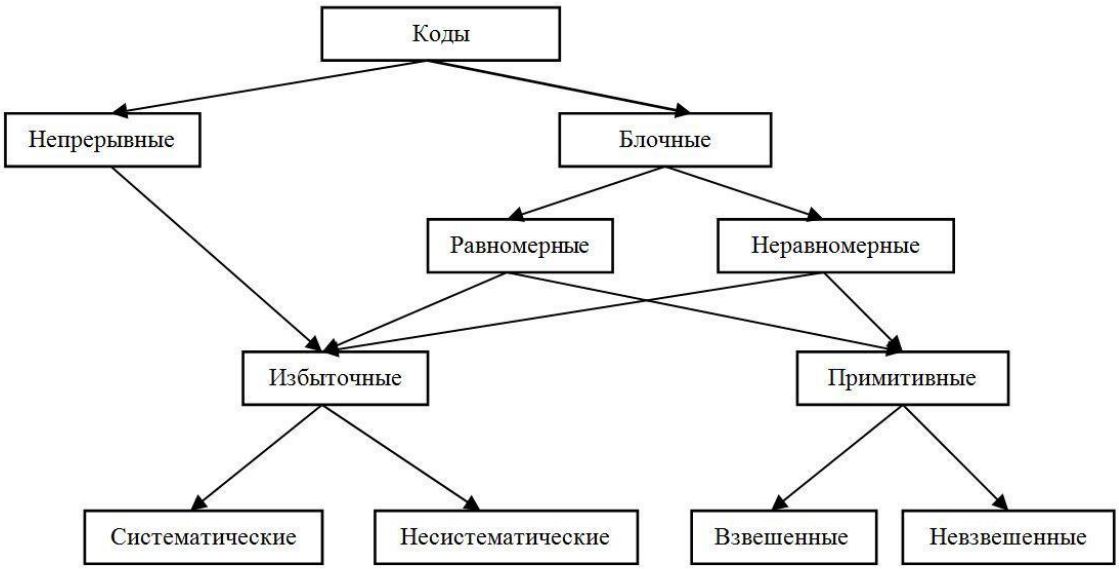

Рисунок 1 – Классификация кодов

Кодовые комбинации неравномерного кода, соответствующие различным символам алфавита источника сообщений, имеют различную длину, а равномерного кода – одинаковую.

Равномерными кодами, например, являются международные телеграфные коды, а неравномерным кодом – код Морзе.

Примитивными кодами являются натуральный двоичный код (НДК) и код Грея, относящиеся к цифровым (импульсным) кодам.

Цифровые коды в зависимости от того, присваивается или нет каждому разряду весовой коэффициент, подразделяют на взвешенные и невзвешенные коды. Например, НДК относится к взвешенным кодам, а код Грея – к невзвешенным.

Из непрерывных кодов наиболее широкое распространение получили рекуррентные коды, иначе называемые цепными кодами, и их разновидность – сверточные коды.

Кодовые символы кодовой последовательности сверточного кода формируются в результате свертки исходной кодовой последовательности с импульсной реакцией кодера сверточного кода.

Один из первых алгоритмов эффективного кодирования информации был предложен Хаффменом в 1952 году. Этот алгоритм стал базой для большого количества программ сжатия информации. Например, кодирование по Хаффмену используется в программах сжатия *ARJ*, *ZIP*, *RAR*, в алгоритме сжатия графических изображений с потерями *JPEG*, а также встроено в современные факс-аппараты.

Эффективное кодирование по Хаффмену состоит в представлении наиболее вероятных (часто встречающихся) букв двоичными кодами наименьшей длины, а менее вероятных – кодами большей длины (если все кодовые слова меньшей длины уже исчерпаны). Это делается таким образом, чтобы средняя длина кода на букву исходного сообщения была минимальной.

До начала кодирования должны быть известны вероятности появления каждой буквы, из которых будет состоять сообщение. На основании этой таблицы вероятностей строится кодовое дерево Хаффмена, с помощью которого производится кодирование букв.

Кодовое дерево (дерево кодирования Хаффмана) – это бинарное дерево, у которого:

– листья помечены символами, для которых разрабатывается кодировка;

– узлы (в том числе корень) помечены суммой вероятностей появления всех символов, соответствующих листьям поддерева, корнем которого является соответствующий узел.

Алгоритм построения дерева Хаффмана:

- Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемый текст.
- Выбираются два свободных узла дерева с наименьшими весами.
- Создается их родитель с весом, равным их суммарному весу.
- Родитель добавляется в список свободных узлов, а двое его детей удаляются из этого списка.
- Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой – бит 0.
- Повторяем шаги, начиная со второго, до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

Существует два подхода к построению кодового дерева: от корня к листьям и от листьев к корню.

Классический алгоритм Хаффмана имеет один существенный недостаток. Для восстановления содержимого сжатого текста при декодировании необходимо знать таблицу частот, которую использовали при кодировании. Сле-

довательно, длина сжатого текста увеличивается на длину таблицы частот, которая должна посылаться впереди данных, что может свести на нет все усилия по сжатию данных. Кроме того, необходимость наличия полной частотной статистики перед началом собственно кодирования требует двух проходов по тексту: одного для построения модели текста (таблицы частот и дерева Хаффмана), другого для собственно кодирования.

Для применения кода на практике желательно, чтобы кодовые слова были как можно короче. Однако чем слова короче, тем их запас меньше. Если попытаться построить префиксный код с очень короткими длинами кодовых слов, то можно потерпеть неудачу – кода с такими длинами слов может не быть.

Пусть $C = \{c_1, c_2, \ldots, c_n\}$ – префиксный двоичный код с длинами кодовых слов $l_1, l_2, \ldots, l_n$. Тогда выполняется неравенство Крафта:

$$\frac{1}{2^{l_1}} + \frac{1}{2^{l_2}} + \cdots + \frac{1}{2^{l_N}} \leq 1 \qquad (3)$$

Основная теорема Шеннона о кодировании для канала без помех:

1. При любой производительности источника сообщений, меньшей пропускной способности канала, существует способ кодирования, позволяющий передавать по каналу все сообщения, вырабатываемые источником.

2. Не существует способа кодирования, обеспечивающего передачу сообщений без их неограниченного накопления, если производительность источника сообщений больше пропускной способности канала.

Кодирование, о котором идет речь в этой формулировке, называется эффективным (безызбыточным) кодированием.

Основная теорема Шеннона о кодировании для канала с помехами:

1. При любой производительности источника сообщений, меньшей, чем пропускная способность канала, существует такой способ кодирования, который позволяет обеспечить передачу всей информации, создаваемой источником, со сколь угодно малой вероятностью ошибки.

2. Не существует способа кодирования, позволяющего вести передачу информации со сколь угодно малой вероятностью ошибки, если производительность источника сообщений больше пропускной способности канала.

Здесь речь идет о помехоустойчивом кодировании. Помехоустойчивость достигается за счет специальным образом введенной избыточности. Это означает, что количество возможных кодовых комбинаций в коде превышает количество сообщений, которые требуется представить (закодировать) в коде. При этом можно построить коды, обнаруживающие ошибки заданной кратности, исправляющие ошибки заданной кратности, а также и то, и другое вместе. Простейшим кодом, обнаруживающим одиночные ошибки, является код с проверкой на четность (добавляется один дополнительный разряд таким образом, чтобы общее число единиц в каждой кодовой комбинации было четным).

Основным практическим следствием теорем Шеннона является построение системы передачи данных с использованием двух ступеней кодирования-декодирования: для устранения избыточности источника и для повышения помехоустойчивости.

Все методы сжатия информации можно условно разделить на два больших непересекающихся класса:

- сжатие с потерей инфор-мации
- сжатие без потери информации.

1) Сжатие с потерей информации. Сжатие с потерей информации означает, что после распаковки уплотненного архива будет получен документ, который несколько отличается от того, который был в самом начале. Понятно, что чем больше степень сжатия, тем больше величина потери и наоборот.

К алгоритмам сжатия с потерей информации относятся такие известные алгоритмы как *JPEG* и *MPEG*. Алгоритм *JPEG* используется при сжатии фотоизображений. Графические файлы, сжатые этим методом, имеют расширение *JPG*. Алгоритмы *MPEG* используют при сжатии видео и музыки. Эти файлы могут иметь различные расширения, в зависимости от конкретной программы, но наиболее известными являются *MPG* для видео и *MP3* для музыки.

2) Сжатие без потери информации. Эти методы применяют при передаче текстовых документов и программ, при создании резервных копий информации, хранящейся на компьютере.

Методы сжатия этого класса не могут допустить утрату информации, поэтому они основаны только на устранении ее избыточности, а информация имеет избыточность почти всегда. Наличие повторяющихся фрагментов – основание для избыточности. В текстах это встречается редко, но в таблицах и в графике повторение кодов – обычное явление. Так, например, если число 0 повторяется двадцать раз подряд, то нет смысла ставить двадцать нулевых байтов. Вместо них ставят один ноль и коэффициент 20. Такие алгоритмы, основанные на выявлении повторов, называют методами *RLE* (*Run Length Encoding*).

Основные свойства алгоритмов сжатия:

1. У всякого сжатия есть предел. На первый взгляд этот принцип самоочевиден, но из него вытекает не очевидное следствие: уплотнение ранее уплотненного файла в лучшем случае не дает выигрыша, а в худшем случае может привести и к проигрышу в размере результирующего файла. Поэтому прежде чем уплотнять информацию, неплохо знать, не была ли она до этого уплотнена другими средствами.

2. Для всякого метода сжатия можно подобрать файл, применительно к которому данный метод является наилучшим. Справедливо и обратное: для всякого метода сжатия можно подобрать файл, который в результате сжатия не уменьшится, а наоборот увеличится.

Отсюда вывод: все дискуссии о том, что один метод сжатия лучше, чем другой, несостоятельны, поскольку их эффективность зависит от конкретных условий.

В области математики и теории информации линейный код – это важный тип блокового кода, использующийся в схемах определения и коррекции ошибок. Линейные коды, по сравнению с другими кодами, позволяют реализовывать более эффективные алгоритмы кодирования и декодирования информации.

Линейным блоковым кодом называется такой помехоустойчивый код, у которого проверочные символы формируются путем суммирования по модулю два информационных символов, расположенных на определенных позициях, а сумма двух кодовых последовательностей и произведение кодовой последовательности на элемент поля образуют также кодовые последовательности.

Расстояние Хэмминга – число позиций, в которых соответствующие символы двух слов одинаковой длины различны. В более общем случае расстояние Хэмминга применяется для строк одинаковой длины любых q-ичных алфавитов и служит метрикой различия (функцией, определяющей расстояние в метрическом пространстве) объектов одинаковой размерности.

Первоначально метрика была сформулирована Ричардом Хэммингом во время его работы в *Bell Labs* для определения меры различия между кодовыми комбинациями (двоичными векторами) в векторном пространстве кодовых последовательностей.

В некоторых системах счисления, например, в коде Грея целые кодированные числа, различающиеся на 1, имеют расстояние Хэмминга равное 1. Такие числа являются «соседними».

Соседнее кодирование важно при проектировании логических устройств, где необходимо исключить логические гонки.

Недостатком линейных кодов является необходимость выбора некоторых определенных исходных (базисных) разрешенных комбинаций. Необходимость проверки условий формирования разрешенных комбинаций из исходных и запоминания матрицы коэффициентов для формирования проверочных символов, а также необходимость запоминания корректирующих синдромов. Поэтому поиск более простых процедур кодирования и декодирования привел к появлению достаточно эффективного подкласса линейных оптимальных кодов (циклических кодов). Такое название эти коды получили, потому что основной операцией при кодировании и декодировании является цикл и кодовые комбинации формируются на основе циклических перестановок. Сущность циклических перестановок заключается в том, что последний символ кодовой комбинации ставится на место первого, а все остальные символы сдвигаются на одну позицию вправо. Если циклической перестановки подверглась одна из разрешенных комбинаций, то в результате этого появится новая разрешенная комбинация и т. д.

Общим свойством всех разрешенных кодовых комбинаций циклических кодов является их делимость без остатка на некоторый специальный полином (комбинацию), называемым производящим (образующим) порождающим. Синдром ошибки в этом случае проявится в наличии остатка от деления принятой комбинации на производящий полином.

К числу наиболее известных и изученных аналитических функций относятся степенные многочлены – полиномы. Графики полиномов описывают огромное разнообразие кривых на плоскости. Кроме того, возможны рациональные полиномиальные выражения в виде отношения полиномов. Таким образом, круг объектов, которые могут быть представлены полиномами, достаточно обширен, и полиномиальные преобразования широко используются на практике, в частности, для приближенного представления других функций.

Полиномом называют выражение, состоящее из нескольких частей одного вида.

Над полиномами можно выполнять обычные арифметические операции: сложение, вычитание, умножение и деление.

3. Криптографическая защита информации

Криптография – наука о методах обеспечения конфиденциальности (невозможности прочтения информации посторонним), целостности данных (невозможности незаметного изменения информации), аутентификации (проверки подлинности авторства или иных свойств объекта), а также невозможности отказа от авторства.

Изначально криптография изучала методы шифрования информации – обратимого преобразования открытого (исходного) текста на основе секретного алгоритма или ключа в шифрованный текст (шифротекст). Традиционная криптография образует раздел симметричных криптосистем, в которых зашифровывание и расшифровывание проводится с использованием одного и того же секретного ключа. Помимо этого раздела современная криптография включает в себя асимметричные криптосистемы, системы электронной цифровой подписи (ЭЦП), хеш-функции, управление ключами, получение скрытой информации, квантовую криптографию.

Криптография не занимается защитой от обмана, подкупа или шантажа законных абонентов, кражи ключей и других угроз информации, возникающих в защищённых системах передачи данных.

Криптография – одна из старейших наук, её история насчитывает несколько тысяч лет.

Открытый (исходный) текст – данные (не обязательно текстовые), передаваемые без использования криптографии.

Шифротекст, шифрованный (закрытый) текст – данные, полученные после применения криптосистемы (обычно – с некоторым указанным ключом).

Шифр, криптосистема – семейство обратимых преобразований открытого текста в шифрованный.

Ключ – параметр шифра, определяющий выбор конкретного преобразования данного текста. В современных шифрах криптографическая стойкость шифра целиком определяется секретностью ключа (принцип Керкгоффса).

Шифрование – процесс нормального применения криптографического преобразования открытого текста на основе алгоритма и ключа, в результате которого возникает шифрованный текст.

Расшифровывание – процесс нормального применения криптографического преобразования шифрованного текста в открытый.

Асимметричный шифр, двухключевой шифр, шифр с открытым ключом – шифр, в котором используются два ключа, шифрующий и расшифровывающий. При этом, зная лишь ключ зашифровывания, нельзя расшифровать сообщение, и наоборот.

Открытый ключ – тот из двух ключей асимметричной системы, который свободно распространяется. Шифрующий для секретной переписки и расшифровывающий – для электронной подписи.

Секретный ключ, закрытый ключ – тот из двух ключей асимметричной системы, который хранится в секрете.

Криптоанализ – наука, изучающая математические методы нарушения конфиденциальности и целостности информации.

Криптоаналитик – учёный, создающий и применяющий методы криптоанализа.

Криптография и криптоанализ составляют криптологию, как единую науку о создании и взломе шифров (такое деление привнесено с запада, до этого в СССР и России не применялось специального деления).

Криптографическая атака – попытка криптоаналитика вызвать отклонения в атакуемой защищённой системе обмена информацией. Успешную криптографическую атаку называют взлом или вскрытие.

Дешифрование (дешифровка) – процесс извлечения открытого текста без знания криптографического ключа на основе известного шифрованного. Термин дешифрование обычно применяют по отношению к процессу криптоанализа шифротекста (криптоанализ сам по себе, вообще говоря, может заключаться и в анализе криптосистемы, а не только зашифрованного ею открытого сообщения).

Криптографическая стойкость – способность криптографического алгоритма противостоять криптоанализу.

Имитозащита – защита от навязывания ложной информации. Другими словами, текст остаётся открытым, но появляется возможность проверить, что его не изменяли ни случайно, ни намеренно. Имитозащита достигается обычно за счет включения в пакет передаваемых данных имитовставки.

Имитовставка – блок информации, применяемый для имитозащиты, зависящий от ключа и данных.

Электронная цифровая подпись, или электронная подпись – асимметричная имитовставка (ключ защиты отличается от ключа проверки). Другими словами, такая имитовставка, которую проверяющий не может подделать.

Центр сертификации – сторона, чья честность неоспорима, а открытый ключ широко известен. Электронная подпись центра сертификации подтверждает подлинность открытого ключа.

Хеш-функция – функция, которая преобразует сообщение произвольной длины в число («свёртку») фиксированной длины. Для криптографической хеш-функции (в отличие от хеш-функции общего назначения) сложно вычислить обратную и даже найти два сообщения с общей хеш-функцией.

Использовавшийся в Древней Греции шифр «скитала», чья современная реконструкция показана на фото, вероятно был первым устройством для шифрования

Роторная шифровальная машина Энигма, разные модификации которой использовались немецкими войсками с конца 1920-х годов до конца Второй мировой войны.

История криптографии насчитывает около 4 тысяч лет. В качестве основного критерия периодизации криптографии возможно использовать технологические характеристики используемых методов шифрования.

Первый период (приблизительно с 3-го тысячелетия до н. э.) характеризуется господством моноалфавитных шифров (основной принцип – замена алфавита исходного текста другим алфавитом через замену букв другими буквами или символами). Второй период (хронологические рамки – с IX века на Ближнем Востоке (Ал-Кинди) и с XV века в Европе (Леон Баттиста Альберти) – до начала XX века) ознаменовался введением в обиход полиалфавитных шифров. Третий период (с начала и до середины XX века) характеризуется внедрением электромеханических устройств в работу шифровальщиков. При этом продолжалось использование полиалфавитных шифров.

Четвёртый период – с середины до 70-х годов XX века – период перехода к математической криптографии. В работе Шеннона появляются строгие математические определения количества информации, передачи данных, энтропии, функций шифрования. Обязательным этапом создания шифра считается изучение его уязвимости к различным известным атакам – линейному и дифференциальному криптоанализам. Однако до 1975 года криптография оставалась «классической», или же, более корректно, криптографией с секретным ключом.

Современный период развития криптографии (с конца 1970-х годов по настоящее время) отличается зарождением и развитием нового направления – криптография с открытым ключом. Её появление знаменуется не только новыми техническими возможностями, но и сравнительно широким распространением криптографии для использования частными лицами (в предыдущие эпохи использование криптографии было исключительной прерогативой государства).

Правовое регулирование использования криптографии частными лицами в разных странах сильно различается – от разрешения до полного запрета.

Современная криптография образует отдельное научное направление на стыке математики и информатики – работы в этой области публикуются в научных журналах, организуются регулярные конференции. Практическое применение криптографии стало неотъемлемой частью жизни современного общества – её используют в таких отраслях как электронная коммерция, электронный документооборот (включая цифровые подписи), телекоммуникации и других.

Для современной криптографии характерно использование открытых алгоритмов шифрования, предполагающих использование вычислительных средств. Известно более десятка проверенных алгоритмов шифрования, которые при использовании ключа достаточной длины и корректной реализации алгоритма криптографически стойки.

Распространенные алгоритмы:

- симметричные *DES*, *AES*, ГОСТ 28147-89, *Camellia*, *Twofish*, *Blowfish*, *IDEA*, *RC*4 и др.;
- асимметричные *RSA* и *Elgamal* (Эль-Гамаль);
- хэш-функций *MD*4, *MD*5, *MD*6, *SHA*-1, *SHA*-2, ГОСТ Р 34.11-2012 («Стрибог») .

Криптографические методы стали широко использоваться частными лицами в электронных коммерческих операциях, телекоммуникациях и многих других средах.

В основе построения криптостойких систем лежит многократное использование относительно простых преобразований, так называемых криптографических примитивов. Клод Шеннон известный американский математик и электротехник предложил использовать подстановки и перестановки. Схемы, которые реализуют эти преобразования, называются SP-сетями. Нередко используемыми криптографическими примитивами являются также преобразования типа циклический сдвиг или гаммирование. Ниже приведены основные криптографические примитивы и их использование.

Симметричное шифрование. Заключается в том, что обе стороны-участники обмена данными имеют абсолютно одинаковые ключи для шифрования и расшифровки данных. Данный способ осуществляет преобразование, позволяющее предотвратить просмотр информации третьей стороной.

Асимметричное шифрование. Предполагает использовать в паре два разных ключа – открытый и секретный. В асимметричном шифровании ключи работают в паре – если данные шифруются открытым ключом, то расшифровать их можно только соответствующим секретным ключом и наоборот – если данные шифруются секретным ключом, то расшифровать их можно только соответствующим открытым ключом. Использовать открытый ключ из одной пары и секретный с другой – невозможно. Каждая пара асимметричных клю-

чей связана математическими зависимостями. Данный способ также нацелен на преобразование информации от просмотра третьей стороной.

Цифровые подписи. Цифровые подписи используются для установления подлинности документа, его происхождения и авторства, исключает искажения информации в электронном документе.

Хеширование. Преобразование входного массива данных произвольной длины в выходную битовую строку фиксированной длины. Такие преобразования также называются хеш-функциями или функциями свёртки, а их результаты называют хеш-кодом, контрольной суммой или дайджестом сообщения. Результаты хэширования статистически уникальны. Последовательность, отличающаяся хотя бы одним байтом, не будет преобразована в то же самое значение.

Криптографическим протоколом называется абстрактный или конкретный протокол, включающий набор криптографических алгоритмов. В основе протокола лежит набор правил, регламентирующих использование криптографических преобразований и алгоритмов в информационных процессах. Примеры криптографических протоколов: доказательство с нулевым разглашением, забывчивая передача, протокол конфиденциального вычисления.

Для современных криптографических систем защиты информации сформулированы следующие общепринятые требования:

– зашифрованное сообщение должно поддаваться чтению только при наличии ключа;

– число операций, необходимых для определения использованного ключа шифрования по фрагменту шифрованного сообщения и соответствующего ему открытого текста, должно быть не меньше общего числа возможных ключей;

– число операций, необходимых для расшифровывания информации путем перебора всевозможных ключей должно иметь строгую нижнюю оценку и выходить за пределы возможностей современных компьютеров (с учетом возможности использования сетевых вычислений);

– знание алгоритма шифрования не должно влиять на надежность защиты;

– незначительное изменение ключа должно приводить к существенному изменению вида зашифрованного сообщения даже при использовании одного и того же ключа;

– структурные элементы алгоритма шифрования должны быть неизменными;

– дополнительные биты, вводимые в сообщение в процессе шифрования, должен быть полностью и надежно скрыты в шифрованном тексте;

– длина шифрованного текста должна быть равной длине исходного текста;

– не должно быть простых и легко устанавливаемых зависимостей между ключами, последовательно используемыми в процессе шифрования;

– любой ключ из множества возможных должен обеспечивать надежную защиту информации;

– алгоритм должен допускать как программную, так и аппаратную реализацию, при этом изменение длины ключа не должно вести к качественному ухудшению алгоритма шифрования.

Блочные шифры – одна из фундаментальных составляющих криптографических систем. Существует обширная литература, посвященная блочным шифрам, а сами они являются одной из наиболее хорошо изученных областей криптографии.

Блочный шифр – это функция шифрования, которая применяется к блокам текста фиксированной длины. Текущее поколение блочных шифров работает с блоками текста длиной 128 бит (16 байт). Такой шифр принимает на вход128-битовыйоткрытый текст и выдает128-битовыйшифрованный текст. Блочный шифр является обратимым: существует функция дешифрования, которая принимает на вход128-битовыйшифрованный текст и выдает исходный 128-битовыйоткрытый текст. Открытый и шифрованный текст всегда имеет один и тот же размер, который называется размером блока.

Чтобы зашифровать сообщение, нужен секретный ключ. Невозможно скрыть сообщение, не сохраняя что-нибудь в секрете. Подобно открытому и шифрованному тексту, ключ шифрования также представляет собой строку битов. Наиболее распространены ключи размером 128 и 256 бит.

Заключение

Последние достижения в области теории информации несколько расширили понятия пропускной способности и сжатия данных на случай сетей каналов связи.

Прогресс в теории кодирования позволил создавать модемы для телефонных каналов, вплотную приблизившие скорость передачи информации к их пропускной способности.

Универсальные алгоритмы сжатия данных ныне широко используются для сжатия компьютерных файлов.

Теория информации остается активно развивающейся областью исследований, поставляющей в наш информационный век новые идеи и подходы в сферу проектирования и анализа систем передачи данных и компьютерных систем.