«МЕТОДЫ КОДИРОВАНИЯ ДАННЫХ КОМБИНАТОРНЫМИ МНОГОЗНАЧНЫМИ КОДАМИ»

Секция «Інформатика»

СУМЫ 2012

ПЛАН

ВВЕДЕНИЕ

Любое отражение материального мира, которое может быть зафиксировано живым существом или прибором, несет в себе информацию.

Отражение результатов человеческой деятельности или понимания окружающего мира может быть представлено в формализованном виде, например в виде наборов букв или цифр. Такие формализованные наборы обычно называют данными. Данные, полученные от источника информации, называют сообщением. Данные становятся информацией в момент их использования. Не всяким данным суждено стать информацией. Информацией становятся те сообщения, которые снимают неопределенность, существовавшую до их поступления.

Теория информации занимается изучением количества информации в сообщениях безотносительно конкретного их содержания, так как процесс формализации и механизации передачи информации не предусматривает изменения содержания сообщений. Предметом изучения теории информации являются вероятностные характеристики исследуемых объектов и явлений, так как вероятность есть наиболее удобная численная мера неопределенности, с уменьшением которой и связан процесс получения информации. Неопределенность появления того или иного явления, неопределенность нахождения в том или ином состоянии некоторой физической системы или ее отдельных элементов, неопределенность появления той или иной буквы в текстовом сообщении и т. д. можно представить при помощи вероятностных характеристик символов некоторого абстрактного алфавита и изучать его информационные характеристики безотносительно того физического содержания, которое кроется за тем или иным символом. С помощью такого абстрактного алфавита в теории информации моделируются все источники информации.

ОСНОВНЫЕ ЗАДАЧИ ТЕОРИИ ИНФОРМАЦИИ И КОДИРОВАНИЯ

Теория передачи информации является частью теории информации. Предметом изучения теории передачи информации является получение оптимальных методов передачи сообщений.

Сообщения передаются при помощи сигналов, обладающих определенными физическими свойствами. В общем случае сигналом может быть любое изменение начального состояния объекта, которое может вызвать реакцию человека или прибора. Различают сигналы: зрительные (телевизионное изображение), звуковые (звонок), электрические (положительные и отрицательные импульсы), радиосигналы и т. д. Одни сигналы могут вызывать другие. Так, электрический сигнал может вызвать звуковой (в электрическом звонке), световой сигнал — электрический (в фотоэлементе). Сигналы могут быть взаимосвязаны в пространстве и во времени (звуковое кино).

Как правило, число однозначно различимых сигналов, предназначенных для передачи сообщений, значительно меньше количества символов алфавита, описывающего источник сообщений. Во всех случаях, когда число символов исходного алфавита $т_1$ больше числа однозначно различимых качественных признаков щ, являющихся непосредственным переносчиком сообщений, для однозначного представления сообщений необходим процесс кодирования [5].

Код — универсальный способ отображения информации при ее хранении, передаче и обработке в виде системы соответствий между элементами сообщений и сигналами, при помощи которых эти элементы можно зафиксировать. Добавим к этому, что кодирование всегда может быть сведено к однозначному преобразованию символов одного алфавита в символы другого. При этом код есть правило, закон, алгоритм, по которому осуществляется это преобразование.

Назовем исходный, кодируемый алфавит — первичным алфавитом. Число качественных признаков первичного алфавита обозначим —$т_х$.

Алфавит, при помощи которого символы первичного алфавита преобразуются в код, назовем вторичным алфавитом.

Основной задачей теории информации и кодирования как самостоятельной дисциплины является оптимальное использование информационных характеристик источников сообщений и каналов связи для построения кодов, обеспечивающих заданную достоверность передаваемой информации с максимально возможной скоростью и минимально возможной стоимостью передачи сообщений. Частными задачами при этом являются: проблемы измерения количества информации, изучение свойств информации, исследование взаимодействия систем и элементов систем методами теории информации, решение задач прикладного характера[1].

Сообщения передаются от объекта к адресату при помощи совокупности технических средств, которые образуют систему передачи информации. Сколько существует методов отображения информации, столько можно создать и способов ее передачи. Поэтому, говоря в дальнейшем о модели системы передачи информации, будем иметь в виду ее наиболее общий вид (рис. 1). К системам передачи информации относится и почта, и телевидение, и сигнализация при помощи костров, которые разжигали в горах древние воины при приближении противника.

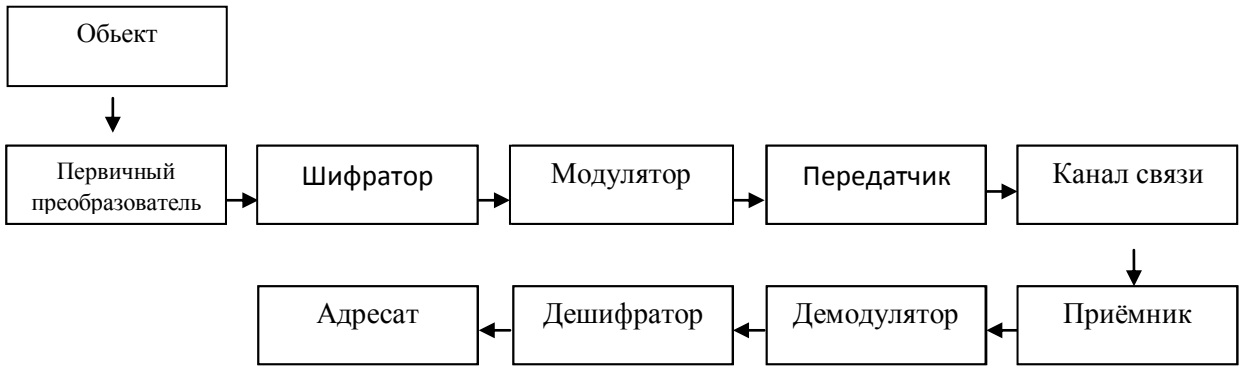

Рис. - 1. Обобщенная модель системы передачи информации.

В простейших каналах связи приемник, передатчик и преобразователь мощности могут быть совмещены. Например, телефон (или телеграф): сигналы от микрофона (или телетайпа) передаются непосредственно по проводной линии связи.

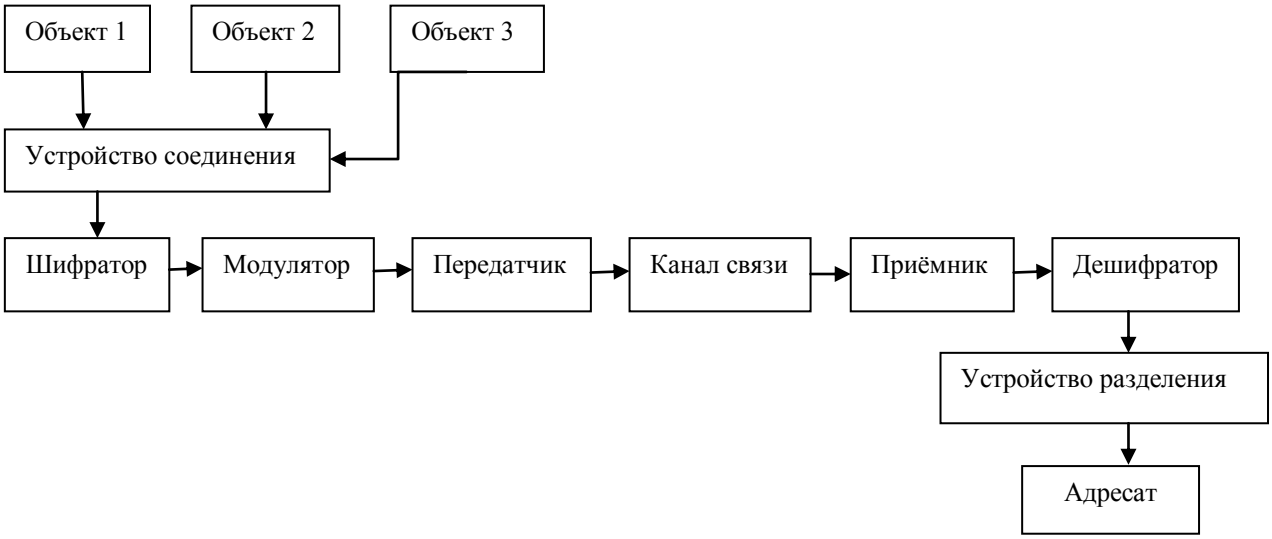

Рис.- 2. Модель многоканальной системы передачи информации.

Для многоканальной системы передачи информации характерны устройства объединения и разделения сигналов (рис. 2). Предположим, требуется передать информацию о состоянии доменных печей. Первичные преобразователи (например, датчики температуры и уровней, газоанализаторы) передают информацию в электронную вычислительную машину, которая ее обрабатывает и затем в определенной последовательности передает на модулятор. В данном случае ЭВМ играет роль устройства объединения и шифратора. Адресатом является световое табло, на котором доменные печи обозначены условными символами. Рядом с символом домны в соответствующих ячейках появляются цифры, отражающие информацию о времени загрузки, проценте содержания кислорода, количестве выплавленного металла и т. д.

Многоканальная система допускает построение кодирующих устройств до устройства объединения, а декодирующих — после устройства

разделения. Однако ее следует пытаться строить так, как показано на рис. 5, что приводит к существенной экономии аппаратуры.

Многоканальная система не обязательно подразумевает передачу информации по нескольким проводам или на нескольких несущих. Не следует путать канал связи и линию связи. Канал связи — совокупность технических средств, предназначенных для передачи информации от объекта к адресату; линия связи — среда, в которой распространяются сигналы, несущие информацию. Для повышения пропускной способности линий связи по ним передают сообщения от нескольких источников одновременно. Такой прием называется уплотнением. В этом случае сообщения от каждого источника передаются по своему каналу связи, хотя линия связи у них общая [2].

Вполне возможно, что у одного объекта может быть несколько адресатов, например, в системах телеуправления, телеизмерения и телесигнализации. В зависимости от структуры связи объекта с адресатами каналы связи могут быть: последовательными — однофидерная линия связи проходит через каждый адресат A_1 — A_5 (рис. 3, а); радиальными — каждый из адресатов A_1 — A_5 соединен с объектом отдельной однофидерной линией; число линии связи больше или равно двум (рис. 3, в); древовидными — однофидерные линии непосредственно не соединяются с объектом, а подключаются к нему через отдельную линию; число линий связи больше или равно трем (рис. 3, г) [1].

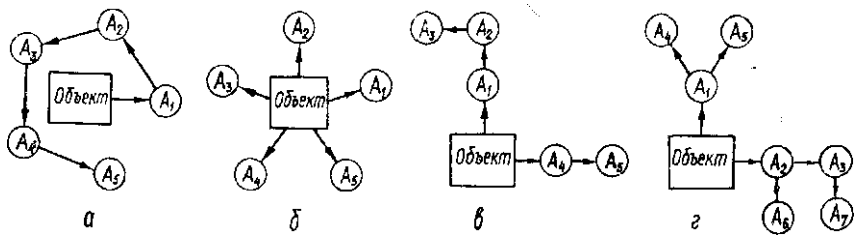

Рис. - 3. Примеры обобщенной структуры канальной связи.

ОСНОВЫ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ

Задача кодера источника – представить подлежащие передаче данные вмаксимально компактной и, по возможности, неискаженной форме. При передаче информации по каналу связи с помехами в принятых данных могут возникать ошибки. Если такие ошибки имеют небольшую величину или возникают достаточно редко, информация может быть использована потребителем. При большом числе ошибок полученной информацией пользоваться нельзя.

Для уменьшения количества ошибок, возникающих при передаче информации по каналу с помехами, может быть использовано кодирование в канале, или помехоустойчивое кодирование. Возможность использования кодирования для уменьшения числа ошибок в канале была теоретически показана К. Шенноном в 1948 году в его работе "Математическая теория связи". В ней было сделано утверждение, что если скорость создания источником сообщений (производительность источника) не превосходит некоторой величины, называемой пропускной способностью канала, то при соответствующем кодировании и декодировании можно свести вероятность ошибок в канале к нулю. Вскоре, однако, стало ясно, что фактические ограничения на скорость передачи устанавливаются не пропускной способностью канала, а сложностью схем кодирования и декодирования. Поэтому усилия разработчиков и исследователей в последние десятилетия были направлены на поиски эффективных кодов, создание практически реализуемых схем кодирования и декодирования, которые по своим характеристикам приближались бы к предсказанным теоретически [1].

Основные принципы. Типы кодов

Кодирование с исправлением ошибок представляет собой метод обработки сообщений, предназначенный для повышения надежности передачи по цифровым каналам. Хотя различные схемы кодирования очень

непохожи друг на друга и основаны на различных математических теориях, всем им присущи два общих свойства.

Первое − использование избыточности. Закодированные последовательности всегда содержат дополнительные, или избыточные, символы. Количество символов в кодовой последовательности Y всегда больше, чем необходимо для однозначного представления любого сообщения λi из алфавита.

Второе — свойство усреднения, означающее, что избыточные символы зависят от нескольких информационных символов, то есть информация, содержащаяся в кодовой последовательности X, перераспределяется также и на избыточные символы.

Существует два больших класса корректирующих кодов − блочные и сверточные. Определяющее различие между этими кодами состоит в отсутствии или наличии памяти кодера.

Кодер для блочных кодов делит непрерывную информационную последовательность X на блоки-сообщения длиной k символов. Кодер канала преобразует блоки-сообщения X в более длинные двоичные последовательности Y, состоящие из n символов и называемые кодовыми словами. Символы (n-k), добавляемые к каждому блоку-сообщению кодером, называются избыточными. Они не несут никакой дополнительной информации, и их функция состоит в обеспечении возможности обнаруживать (или исправлять) ошибки, возникающие в процессе передачи [2].

Как мы ранее показали, k-разрядным двоичным словом можно представить 2k возможных значений из алфавита источника, им соответствует 2k кодовых слов на выходе кодера. Такое множество 2k кодовых слов называется блочным кодом. Термин "без памяти" означает, что каждый блок из n символов зависит только от соответствующего информационного блока из k символов и не зависит от других блоков.

Кодер для свёрточных кодов работает с информационной последовательностью без разбиения ее на независимые блоки. В

каждый момент времени кодер из небольшого текущего блока информационных символов размером в b символов (блока-сообщения) образует блок, состоящий из v кодовых символов (кодовый блок), причем v > b. При этом кодовый v-символьный блок зависит не только от b-символьного блока- сообщения, присутствующего на входе кодера в настоящий момент, но и от предшествующих m блоков-сообщений. В этом, собственно, и состоит наличие памяти в кодере. Блочное кодирование удобно использовать в тех случаях, когда исходные данные по своей природе уже сгруппированы в какие-либо блоки или массивы [4].

При передаче по радиоканалам чаще используется сверточное кодирование, которое лучше приспособлено к побитовой передаче данных. Кроме этого, при одинаковой избыточности сверточные коды, как правило, обладают лучшей исправляющей способностью.

Линейные блочные коды

Для блочного кода с 2k кодовыми словами длиной в n символов, если он только не обладает специальной структурой, аппарат кодирования и декодирования является очень сложным. Поэтому ограничим свое рассмотрение лишь кодами, которые могут быть реализованы на практике. Одним из условий реализуемости блочных кодов прибольших k является условие их линейности.

Что такое линейный код?

Блочный код длиной n символов, состоящий из 2k кодовых слов, называется линейным (n, k)-кодом при условии, что все его 2k кодовых слов образуют k-мерное подпространство векторного пространства n- последовательностей двоичного поля GF(2).

Если сказать проще, то двоичный код является линейным, если сумма по модулю 2 (mod2) двух кодовых слов также является кодовым словом этого кода.

Работая с двоичными кодами, мы постоянно будем сталкиваться с элементами двоичной арифметики, поэтому определим основные понятия.

Полем называется множество математических объектов, которые можно складывать, вычитать, умножать и делить.

Возьмем простейшее поле, состоящее из двух элементов − нуля - 0 и единицы - 1. Определим для него операции сложения и умножения:

$$0+0=0, \qquad 0\cdot 0=0;$$

$$0+1=1, \qquad 0\cdot 1=0;$$

$$1+0=1, \qquad 1\cdot 0=0;$$

$$1+1=0, \qquad 1\cdot 1=1.$$

Определенные таким образом операции сложения и умножения называются сложением по модулю 2 (mod2) и умножением по модулю 2. Отметим, что из равенства 1+1 = 0 следует, что -1 = 1 и, соответственно, 1+1=1-1, а из равенства $1\cdot 1=1$ − что 1:1=1. Алфавит из двух символов 0 и 1 вместе со сложением и умножением по mod2 называется полем из двух элементов и обозначается как GF(2). К полю GF(2) применимы все методы линейной алгебры, в том числе матричные операции.

Еще раз обратим внимание на то, что все действия над символами в двоичных кодах выполняются по модулю 2.

Желательным качеством линейных блочных кодов является систематичность.

Систематический код имеет формат, изображенный на рис. 4, то есть содержит неизменную информационную часть длиной k символов и избыточную (проверочную) длиной $n - k$ символов.

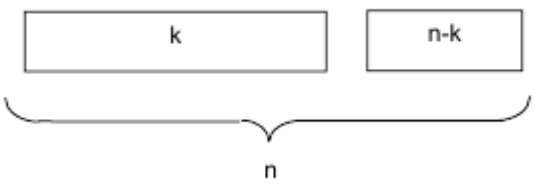

Рис. – 4. Формат систематического кода

Блочный код, обладающий свойствами линейности и систематичности, называется линейным блочным систематическим (n, k)-кодом [2].

КЛАСИФИКАЦИЯ ПОМЕХОУСТОЙЧИВОГО КОДОВ

Работа подавляющего числа современных систем связи основана на передаче сообщений в цифровом виде. Сбой при приеме любого элемента цифровых данных способен вызвать значительное искажение всего сообщения в целом, что, в свою очередь, может привести к полной потере информации, содержащейся в нем. В настоящее время по каналам связи передаются данные со столь высокими требованиями к достоверности передаваемой информации, что удовлетворить эти требования традиционным методами - совершенствованием антенно-фидерных устройств, увеличением излучаемой мощности, снижением собственного шума приемника - оказывается экономически невыгодным или просто невозможным.

Высокоэффективным средством решения данной проблемы является применение помехоустойчивого кодирования, основанного на введении искусственной избыточности в передаваемое сообщение. Теория и техника помехоустойчивого кодирования прошли несколько этапов в своем развитии. Изначально это было просто эмпирическое использование простейших кодов с повторением, с постоянным весом, с одной проверкой на четность и т.д. В дальнейшем теория помехоустойчивого кодирования прошла довольно длинный путь развития, в процессе которого для ее создания использовались основы математической теории – ответвления высшей алгебры и теории чисел с приложением к реальным системам связи.

Теория кодирования возникла в конце 40-х годов с появлением работ Голея, Хэмминга и Шеннона. Выдающиеся два ученые Голей и Хэмминг заложили основу алгебраическим методам кодирования, которые

используются и по сей день, а Шеннон предложил и исследовал понятие случайного кодирования. Отметим, что в современных информационных системах важнейшей задачей является обеспечение информационной безопасности, связанной с методами криптографии и кодирования, теоретические основы которой заложил Шеннон в своих трудах[2].

Помехоустойчивое кодирование передаваемой информации позволяет в приемной части системы обнаруживать и исправлять ошибки. Коды, применяемые при помехоустойчивом кодировании, называются корректирующими кодами. Как правило, корректирующий код может исправлять меньше ошибок, чем обнаруживать. Число ошибок, которые корректирующий код может исправить в определенном интервале последовательности двоичных символов, например, в одной кодовой комбинации, называется исправляющей способностью кода [5].

Физическая среда, по которой передаются данные, не может быть абсолютно надёжной. Более того, уровень шума бывает очень высоким, например, в беспроводных системах связи и телефонных системах. Ошибки при передаче — это реальность, которую надо обязательно учитывать

В разных средах характер помех разный. Ошибки могут быть одиночные, а могут возникать группами, сразу по несколько. В результате помех могут исчезать биты или наоборот — появляться лишние.

Основной характеристикой интенсивности помех в канале является параметр шума — p. Это число от 0 до 1, равное вероятности инвертирования бита, при условии что, он был передан по каналу и получен на другом конце.

Следующий параметр — p2. Это вероятность того же события, но при условии, что предыдущий бит также был инвертирован.

Этими двумя параметрами вполне можно ограничиться при построении теории. Но, в принципе, можно было бы учитывать аналогичные вероятности для исчезновения бита, а также использовать полную информацию о пространственной корреляции ошибок, — то есть

корреляции соседних ошибок, разделённых одним, двумя или более битами.

У групповых ошибок есть свои плюсы и минусы. Плюсы заключаются в следующем. Пусть данные передаются блоками по 1000 бит, а уровень ошибки 0,001 на бит. Если ошибки изолированные и независимые, то 63% блоков будут содержать ошибки. Если же они возникают группами по 100 сразу, то ошибки будут содержать 1% блоков.

Зато, если ошибки не группируются, то в каждом кадре они невелики, и есть возможность их исправить. Групповые ошибки портят кадр безвозвратно. Требуется его повторная пересылка, но в некоторых системах это в принципе невозможно, — например, в телефонных системах, использующие цифровое кодирование, возникает эффект пропадания слов/слогов.

Для надёжной передачи кодов было предложено два основных метода.

Первый — добавить в передаваемый блок данных нескольких «лишних» бит так, чтобы, анализируя полученный блок, можно было бы сказать, есть в переданном блоке ошибки или нет. Это, так называемые, коды с обнаружением ошибок.

Второй — внести избыточность настолько, чтобы, анализируя полученные данные, можно не только замечать ошибки, но и указать, где именно возникли искажения. Это коды, исправляющие ошибки.

Под помехой понимается любое воздействие, накладывающееся на полезный сигнал и затрудняющее его прием [2].

Внешние источники помех вызывают в основном импульсные помехи, а внутренние - флуктуационные. Помехи, накладываясь на видеосигнал, приводят к двум типам искажений: краевые и дробления. Краевые искажения связаны со смещением переднего или заднего фронта импульса. Дробление связано с дроблением единого видеосигнала на некоторое количество более коротких сигналов.

Помехоустойчивые коды делятся на блочные и непрерывные.

Блочными называются коды, в которых информационный поток символов разбивается на отрезки и каждый из них преобразуется в определённую последовательность (блок) кодовых символов. В блочных кодах кодирование при передаче (формирование проверочных элементов) и декодирование при приёме (обнаружение и исправление ошибок) выполняются в пределах каждой кодовой комбинации (блока) в отдельности по соответствующим алгоритмам.

Непрерывные или рекуррентные коды образуют последовательность символов, не разделяемую на отдельные кодовые комбинации. Кодирование и декодирование непрерывно совершаются над последовательностью элементов без деления их на блоки. Формирование проверочных символов ведётся по рекуррентным (возвратным) правилам, поэтому непрерывные коды часто называют рекуррентными или цепными.

В простейшем цепном коде каждый проверочный элемент формируется путём сложения по модулю 2 соседних или отстоящих друг от друга на определённое число позиций информационных элементов. В канал связи передаётся последовательность импульсов, в которой за каждым информационным следует проверочный.

К непрерывным кодам относятся и свёрточные коды, в которых каждый информационный символ, поступающий на вход кодирующего устройства, вызывает появление на его выходе ряда проверочных элементов, образованных суммированием по модулю 2 данного символа и " k-1 " предыдущих информационных символов. Рекуррентные коды позволяют исправлять групповые ошибки (" пачки ") в каналах связи.

Блочные коды делятся на равномерные и неравномерные. В равномерных кодах, в отличие от неравномерных, все кодовые комбинации содержат одинаковое число n - символов (разрядов) с постоянной длительностью ф0 импульсов символов кода. Равномерные коды в основном и применяются в системах связи, так как это упрощает технику передачи и приёма.

Классическими примерами неравномерного кода являются код Морзе, широко применяемый в телеграфии, и код Хафмена, применяемый для компрессии информации (факсимильная связь, ЭВМ).

Никаких специальных мер по исправлению и обнаружению ошибок в коде Морзе не предусматривается в связи с большой избыточностью самого передаваемого текста. В этом смысле код Морзе не относится к классу корректирующих кодов.

Почти все блочные корректирующие коды принадлежат к разделимым кодам, в которых кодовые комбинации состоят из двух частей: информационной и проверочной. Их символы всегда занимают одни и те же позиции, т.е. располагаются на определённых местах. Как правило, в таких кодах, все кодовые комбинации которых содержат n символов, первые k символов являются информационными, а за ними располагаются (n - k) проверочных символов. В соответствии с этим разделимые коды получили условное обозначение – (n , k) – коды [1].

В неразделимых кодах деление на информационные и проверочные символы отсутствует. К таким кодам относятся, в частности, коды с постоянным весом, так называемые равновесные коды. Например, Международным Консультативным Комитетом по телеграфии и телефонии (МККТТ) рекомендован для использования телеграфный код № 3 - семиразрядный код с постоянным весом, т.е. с числом единиц в каждой кодовой комбинации, равным 3 (W = 3).

Систематические коды образуют наиболее обширную группу (n, k)-разделимых кодов. Особенностью этих кодов является то, что проверочные (корректирующие) символы образуются с помощью линейных операций над информационными. Кроме того, любая разрешённая кодовая комбинация может быть получена в результате линейной операции над набором к линейно независимых кодовых комбинаций. В частности, суммирование по модулю 2 двух и более разрешённых комбинаций также дает разрешённую кодовую комбинацию.

Поскольку теоретической основой получения таких комбинаций является математический аппарат линейной алгебры, то коды и называют линейными, а учитывая, что проверочные символы формируются по определённой системе (правилам), блочные равномерные разделимые линейные коды получили название систематических. Использование аппарата линейной алгебры, в которой важное значение имеет понятие "группа", породило и другое название этих кодов - групповые.

Эти коды получили наибольшее применение в системах передачи дискретной информации.

Несистематические (нелинейные) коды указанными выше свойствами не обладают и применяются значительно реже в специальных случаях. Примером нелинейного кода является уже упоминавшийся неразделимый, равновесный код. Эти коды обычно используются в несимметричных каналах связи, в которых вероятность перехода $1 \rightarrow 0$ значительно больше вероятности перехода $0 \rightarrow 1$ или наоборот. В таких каналах очень маловероятно, чтобы в одном блоке были переходы обоих видов, и поэтому почти все ошибки приводят к изменению веса блока, и, следовательно, обнаруживаются.

Другим примером несистематического кода является код с контрольным суммированием - итеративный код. В этом коде проверочные разряды формируются в результате суммирования значений разрядов как в данной кодовой комбинации, так и одноимённых разрядов в ряде соседних с ней комбинаций, образующих совместный блок. Итеративные коды позволяют получить так называемые мощные коды, т.е. коды с длинными блоками и большим кодовым расстоянием при сравнительно простой процедуре декодирования. Итеративные коды могут строиться как комбинационные посредством произведения двух или более систематических кодов [2].

К комбинационным кодам можно отнести также антифединговые коды, предназначенные для обнаружения и исправления ошибок в каналах с замираниями (федингом) сигналов. Для таких каналов с группированием

ошибок применяют метод перемежения символов или декорелляции ошибок. Он заключается в том, что символы, входящие в одну кодовую комбинацию, передаются не непосредственно друг за другом, а перемежаются символами других кодовых комбинаций исходного систематического или любого другого кода. Если интервал между символами, входящими в одну кодовую комбинацию, сделать длиннее "памяти" (интервала корелляции) канала с замираниями, то в пределах длительности одной исходной кодовой комбинации группирования ошибок не будет. На приёме после обратной "расфасовки" в кодовых комбинациях можно производить декодирование с обнаружением и исправлением ошибок.

В систематических кодах различают два метода формирования проверочной группы символов: поэлементный и в целом.

Наиболее известны среди систематических кодов коды Хемминга, которые исторически были найдены раньше многих других кодов и сыграли большую роль в развитии теории корректирующих кодов. В этих кодах используется принцип проверки на чётность определённого ряда информационных символов.

Проверочная группа из r символов формируется поэлементно по соответствующему алгоритму. Коды Хемминга, имеющие dmin = 3, позволяют исправить одну ошибку

Циклические коды также относятся к классу линейных систематических кодов и обладают всеми их свойствами. Коды названы циклическими потому, что циклический сдвиг любой разрешённой кодовой комбинации также является разрешённой комбинацией. Теория построения циклических кодов базируется на разделах высшей алгебры, изучающей свойства двоичных многочленов.

Особую роль в этой теории играют так называемые неприводимые многочлены, т.е. полиномы, которые не могут быть представлены в виде произведения многочленов низших степеней. В связи с этим циклические коды относят к разновидности полиномиальных кодов.

Среди циклических кодов особое место занимает класс кодов, предложенных Боузом и Рой-Чоудхури и независимо от них Хоквингемом . Коды Боуза-Чоудхури-Хоквингема получили сокращённое наименование БЧХ - коды и отличаются специальным выбором порождающего (образующего) циклический код полинома, что приводит к простой процедуре декодирования.

В циклических кодах " r " проверочных символов, добавляемых к исходным " k " информационным, могут быть получены сразу, т.е. в целом, в результате умножения исходной подлежащей передаче кодовой комбинации Q(x) простого кода на одночлен x r и добавлением к этому произведению остатка R(x), полученного в результате деления произведения на порождающий полином P(x).

В процессе кодирования при передаче информации из информационных разрядов в соответствии с определёнными для каждого К. правилами формируются дополнительные символы — проверочные разряды. При декодировании из принятых кодовых слов по тем же правилам вновь формируют проверочные разряды и сравнивают их с принятыми; если они не совпадают, значит при передаче произошла ошибка. Существуют коды, обнаруживающие факт искажения сообщения, и коды, исправляющие ошибки, т. е. такие, с помощью которых можно восстановить первичную информацию.

Ошибки в передаваемых словах могут возникать вследствие либо независимых искажений разрядов (в этом случае применяют, например, коды типа кода Хэмминга), либо искажений группы рядом стоящих разрядов (для таких случаев разработаны коды, исправляющие одиночные пачки ошибок, и коды, исправляющие более одной пачки ошибок); для обнаружения ошибок в процессе вычислений на ЭВМ разработаны так называемые арифметические коды [5].

ИЗБЫТОЧНОСТЬ ИНФОРМАЦИИ

Для нахождения максимальной пропускной способности системы связи необходимо уметь определять максимальное количество информации, которое может быть передано при помощи символов данного алфавита за единицу времени. Мы уже знаем, что максимальное количество информации на символ сообщения H=logm можно получить только в случае равновероятных и независимых символов. Реальные коды редко полностью удовлетворяют этому условию, поэтому информационная нагрузка на каждый их элемент обычно меньше той, которую они могли бы переносить. Энтропия сообщений, представляемых такими кодами, меньше максимальной.

Раз элементы кодов, представляющих сообщения, недогружены, то само сообщение обладает информационной избыточностью. Понятие избыточности в теории информации и кодирования введено для количественного описания информационного резерва кода, из которого составлено сообщение. Сама постановка такой задачи стала возможной именно потому, что информация является измеримой величиной, каков бы ни был частный вид рассматриваемого сообщения [5].

Различают избыточность естественную и искусственную. Естественная избыточность характерна для первичных алфавитов, а искусственная — для вторичных.

Естественная избыточность может быть подразделена на семантическую и статистическую избыточности.

Семантическая избыточность заключается в том, что мысль, высказанная в сообщении, может быть выражена короче. Если сообщение можно сократить без изменения смысла, а затем восстановить содержание, то оно обладает семантической избыточностью. - Для уяснения понятия семантической избыточности рассмотрим следующее сообщение: «Затребованные от нас сводки в положенный срок обработать не можем ввиду того, что на районном вычислительном центре вышла из строя

подстанция». Очевидно, что без особой потери ценности информации это сообщение можно было бы передать так: «Обработка сводок задерживается связи отсутствием электричества». Второе сообщение короче, в нем слова несут гораздо большую информационную нагрузку, чем в первом, т. е. первое сообщение обладает семантической избыточностью по отношению ко второму.

Семантическую избыточность можно устранять различными способами, например, стандартные, часто повторяющиеся сообщения заменять условными обозначениями; сообщения, содержащие различные характеристики одних и тех же элементов, представлять в виде таблиц; применять свертывание информации, аббревиатуры и т. д. Общим при всем при этом остается то, что все преобразования по устранению семантической избыточности производятся в первичном алфавите.

Статистическая избыточность обусловливается неравновероятностным распределением качественных признаков первичного алфавита и их взаимозависимостью.

Например, для английского алфавита, состоящего из 26 букв, максимальное значение энтропии

$$H_{max}=\log_2 m= \log_2 26 = 4{,}7 \text{бит}.$$

Если условно представить частоту появления различных букв в английских текстах, то можно наглядно убедиться в том, что вероятности появления букв английского алфавита далеко не равны, а следовательно, энтропия английского языка меньше, чем 4,7 бит. Действительно, исследования показали, что при учете частоты распределения восьмибуквенных сочетаний, т. е. взаимозависимости между символами, энтропия английского языка уменьшается до 2,35 бит. Если же учитывать статистику следования слов в английских текстах, то энтропия английского языка не превысит 2 бит.

Как видим, избыточность заложена уже в самой природе английского алфавита.

При учете частоты появления букв в текстах, следования букв в различных сочетаниях и слов в различных сообщениях передаваемую информацию можно значительно сжать, сократить. Отношение $H/H_{max}=\mu$ называют коэффициентом сжатия, или относительной энтропией, а величину

$$D = 1 - \frac{H}{H_{\max}} \tag{1}$$

избыточностью. Из выражения (1) очевидно, что избыточность больше у тех сообщений, у которых больше энтропия.

Энтропия может быть определена как информационная нагрузка на символ сообщения. Избыточность определяет недогруженность символов. Если $H=H_{max}$, то согласно формуле (1) недогруженности не существует. Поэтому для характеристики степени недогруженности и приняли разность между единицей и μ.

Для английского языка без учета взаимозависимости между словами

$$D = 1 - \frac{2,35}{4,7} = 1 - 0,5 = 0,5$$

Действительно, проведенные эксперименты подтвердили, что удается восстановить содержание английских текстов, составленных из 50% алфавита [1].

Кроме общего понятия статистической избыточности, существуют различные частные понятия, основными из которых являются следующие: избыточность D_s, вызванная статистической связью между символами сообщения, и избыточность D_p, обусловленная неравновероятными распределениями символов в сообщениях.

$$H(A/B) = -\sum_i \sum_j p(a_i) p(b_j / a_i) \log p(b_j / a_i) \tag{2}$$

$$H = -\sum_{i=1}^{m} p_i \log p_i = \sum_{i=1}^{m} p_i \log \frac{1}{p_i} \tag{3}$$

Избыточность D_s определяется выражениями (2),(3) и характеризует информационный резерв сообщений со взаимонезависимыми символами по отношению к сообщениям, в которых наблюдается статистическая связь между символами:

$$D_s = 1 - \frac{H}{H'}$$

где

$$H(A/B) = -\sum_i \sum_j p(a_i) p(b_j / a_i) \log p(b_j / a_i)$$

$$H' = -\sum_i p_i \log p_i$$

Однако выражение для H' само обладает избыточностью за счет неэкстремальности распределения вероятностей отдельных символов (напомним, что максимальная энтропия достигается при равномерном распределении вероятностей $H_{max}=\log_2 m$ для конечного алфавита m).

Избыточность D_p определяется выражениями (3) и:

$$H = H_{\max} = -\sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{m} = \log m$$

и характеризует информационный резерв сообщений с равновероятными символами относительно сообщений, символы которых неравновероятны:

$$D_s = 1 - \frac{H'}{H_{\max}}$$

Полная избыточность

$$D=D_s+D_p-D_sD_p \qquad (4)$$

При малых D_sи D_pполную избыточность вычисляют как сумму частных избыточностей, так как последний член выражения (4) представляет собой произведение дробей, меньших единицы, и с уменьшением D_pи D_sстремится к нулю гораздо быстрее, чем два первых члена.

Устраняется статистическая избыточность путем построения оптимальных неравномерных кодов.При этом статистическая избыточность первичного алфавита устраняется за счет рационального построения сообщений во вторичном алфавите.

Таким образом, естественная избыточность всегда сосредоточена в первичном алфавите, либо в структуре сообщения за исключением того случая, когда избыточность заложена в природе кода и ее нельзя задать одной цифрой на основании статистических испытаний.

Так при передаче десятичных цифр двоичным кодом максимально загруженными бывают только те символы вторичного алфавита, которые передают значения, являющиеся целочисленными степенями двойки. В остальных случаях тем же количеством символов может быть передано большее количество цифр (сообщений). Например, тремя двоичными разрядами мы можем передать и цифру 5, и цифру 8, т. е. для передачи пяти сообщений надо строить код такой же длины, как и для передачи 8 сообщений.

Фактически для передачи цифры 5 в двоичном коде достаточно иметь длину кодовой комбинации

$$L \geq \frac{\log_2 N}{\log_2 m},$$

где N— общее количество передаваемых сообщений.

Lможет быть представлена и так:

$$L \geq \frac{\log_2 m_1}{\log_2 m_2},$$

где m_1 и m_2 соответственно качественные признаки первичного и вторичного алфавитов. Поэтому для нашего примера можно записать:

$$L \geq \frac{\log_2 5}{\log_2 5} = 2,32 \text{ дв. символа}$$

Однако эту цифру необходимо округлить до ближайшего целого числа, так как длина кода не может быть выражена дробным числом. Округление, естественно, производится в большую сторону.

В общем случае, избыточность от округления

$$D_0 = \frac{k - \varphi}{k} \tag{5}$$

где $\varphi = \frac{\log_2 m_1}{\log_2 m_2}$, k— округленное до ближайшего целого значение φ.

Для нашего примера

$$D_0 = \frac{3 - 2,32}{3} = 0,227$$

При передаче одних и тех же сообщений одним и тем же кодом могут наблюдаться разные виды избыточности. Например, при передаче русских текстов в двоичном неравномерном коде избыточность будет как за счет неравномерной статистики появления букв алфавита в текстах, так и за счет избыточности, заложенной в природе двоичного кода, суть которой заключается в том, что вероятность появления символа «0» больше вероятности появления символа «1».

Избыточность, заложенную в природе кода, полностью устранить нельзя. Однако как избыточность от неравновероятного появления символов, так и избыточность от округления автоматически убывает по мере увеличения длины кодового блока [4].

Когда мы говорим о том, что естественная избыточность характерна для первичных алфавитов, а искусственная для вторичных, мы подчеркиваем тот факт, что естественная избыточность существует

всообщении до того, как оно трансформируется в код, чего никогда нельзя сказать об искусственной избыточности.

Искусственная избыточность необходима для повышения помехоустойчивости кодов и ее вводят в виде n_kдобавочных символов. Если в коде всего nразрядов, из них $n_и$ несут информационную нагрузку, то

$$n_k = n - n_и$$

и характеризует абсолютную корректирующую избыточность, а величина

$$D_k = \frac{n - n_и}{n_и}$$

характеризует относительную корректирующую избыточность[1].

ОЦЕНКА ПОМЕХОУСТОЙЧИВОСТИ НЕРАЗДЕЛИМЫХ КОДОВ

Помехоустойчивость передачи дискретных сообщений определяется совместным действием большого числа взаимосвязанных факторов: избыточностью сообщений, способом кодирования, свойствами сигналов-переносчиков, видом модуляции, характером искажений сигналов и помех в -канале, нарушением синхронизации передаваемых и принимаемых сигналов, способами демодуляции и декодирования принимаемых сигналов и др. [2]

Задачей систем связи обычно является передача наибольшего объема информации за определенный период времени с вероятностью ошибок $P_{ош} < P_{доп}$. Решение этой задачи требует оценки помехоустойчивости кодов. Методы оценки, применяющие, например, минимальное кодовое расстояние, для некоторых, в частности, неразделимых кодов неприемлемы. В качестве такого метода может быть использован

универсальный подход оценки кодов, в соответствии с которым определяется доля обнаруживаемых ошибочных комбинаций

$$D = 1 - \frac{M}{N}; \tag{6}$$

где M - число разрешенных кодовых комбинаций ;

N - общее число кодовых комбинаций ;

Выражение (6) определяет потенциальную помехоустойчивость кода. Она предполагает, что источник информации генерирует кодовые комбинации с равной вероятностью и что вероятности их перехода во все кодовые слова, как разрешенные, так и запрещенные, также равны. Такой подход, однако, ограничен, так как не учитывает реальных свойств канала связи и источника информации, в котором обычно генерирование кодовых слов происходит с разной вероятностью. С различной вероятностью на практике также происходят и переходы разрешенных слов в другие и в самих себя.

Учесть это можно введением вероятности P_i генерирования слов источником информации и вероятностей перехода разрешенного слова в запрещенное $p_i^З$, перехода разрешенного слова в разрешенное $p_i^Н$ и перехода разрешенного слова с самого себя p_i^i. [3]

Очевидно, что

$$\sum_{i=1}^{M} P_i = 1; \tag{7}$$

$$p_i^З + p_i^Н + p_i^i = 1; \tag{8}$$

Доля обнаруживаемых ошибочных комбинаций

$$Z = \sum_{i=1}^{M} \mathrm{P}_i \, p_i^{3} = \sum_{i=1}^{M} \sum_{j=M+1}^{N} \mathrm{P}_i \, p_{ij}^{3}; \qquad (9)$$

где p_{ij}^{3} - вероятность перехода i-го разрешенного слова в j-тое запрещенное.

Доля необнаруживаемых ошибочных комбинаций

$$V = \sum_{i=1}^{M} \mathrm{P}_i \, p_i^{\mathrm{H}} = \sum_{i=1}^{M} \sum_{j=1,\, j\neq i}^{M} \mathrm{P}_i \, p_{ij}^{\acute{I}} \qquad (10)$$

где $\mathrm{p}_{ij}^{\mathrm{H}}$ - вероятность перехода i -го разрешенного слова в j-тое разрешенное.

Доля ошибочных переходов

$$W = Z + V. \qquad (11)$$

Доля правильных переходов

$$\Pi = \sum_{i=1}^{M} \mathrm{P}_i \, p_i^{i} \qquad (12)$$

В частном случае при равновероятностном переходе разрешенной кодовой комбинации в любую ($\mathrm{p}_{ij}^{3} = \mathrm{p}_{ij}^{\mathrm{H}} = \mathrm{p}_{i}^{i} = 1/\mathrm{N}$) получим, что

$$\mathrm{p}_i^{3} = \sum_{j=M+1}^{N} \mathrm{p}_{ij}^{3} = \frac{1}{N}(N - M) = \frac{N - M}{N}; \qquad (13)$$

$$Z = \frac{N - M}{N} \sum_{i=1}^{M} \mathrm{P}_i = 1 - \frac{M}{N}; \qquad (14)$$

$$p_i^{\text{Н}} = \sum_{j=1,\, j\neq i}^{M} p_{ij} = \sum_{j=1,\, j\neq i}^{M} \frac{1}{N} = \frac{M-1}{N}; \tag{15}$$

$$V = \frac{M-1}{N}\sum_{i=1}^{M} P_i = \frac{M-1}{N}\times 1 = \frac{M-1}{N}; \tag{16}$$

$$\Pi = \sum_{i=1}^{M} P_i \frac{1}{N} = \frac{1}{N}. \tag{17}$$

Следовательно, при равновероятных переходах вероятностные характеристики источника информации не влияют на величины Z, V и П.

В случае равновероятного генерирования источником информационных слов, то есть при $P_i = 1/M$ получим:

$$Z = \sum_{i=1}^{M} \frac{1}{M} p_i^{\text{З}} = \frac{1}{M}\sum_{i=1}^{M} p_i^{\text{З}}; \tag{18}$$

$$V = \sum_{i=1}^{M} \frac{1}{M} p_i^{\text{Н}} = \frac{1}{M}\sum_{i=1}^{M} p_i^{\text{Н}}; \tag{19}$$

$$\Pi = \frac{1}{M}\sum_{i=1}^{M} p_i^{i}\ . \tag{20}$$

Это значит, что значения Z, V и П при равной вероятности генерирования слов зависят только от вероятностей переходов $p_{ij}^{\text{З}},\ p_{ij}^{\text{Н}},\ p_i^{i}$ [3].

ЛИТЕРАТУРА

1 Цымбал В.П. Теория информации и кодирование. - К.: Выща шк., 1992.

2 Шульгин В.И. Основы теории передачи информации. В двух частях. - Харьков “ХАИ”., 2003.

3 Борисенко А.А., Онанченко Е.Л. Оценка помехоустойчивости неразделимых кодов. / Вест. Сум. ун-та , 1994, №2. с. 64-68.

4 Игнатов В.А. Теория информации и передачи сигналов: Учебник для вузов. – М.: Сов. радио, 1979. 280 с., ил.

5 Жураковський Ю.П., Полторак В.П. Теорія інформації та кодування: Підручник. – К.:Вища шк, 2001.-255 с.