МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение
высшего образования «Хакасский государственный университет им. Н.Ф. Катанова»
Колледж педагогического образования, информатики и права

ПЦК естественнонаучных дисциплин, математики и информатики

РЕФЕРАТ

на тему:
Изучение способов кодирование и шифрования информации

Автор реферата: _____ Емашев М.В.
(подпись) (инициалы, фамилия)

Специальность: 230115 – Программирование в компьютерных системах

Курс: II Группа: И-21
Зачет/незачет: _____
Руководитель: _____ Когумбаева О.П.
(подпись, дата) (инициалы, фамилия)

г. Абакан, 2017г.

Оглавление

Введение

Еще в древние времена люди осознавали, что информация имеет большую ценность. Недаром переписки между императорами были объектом пристального внимания их недругов и друзей. Тогда и возникла задача защиты этих переписок от любопытных глаз.

Одним из методов решения этой задачи была тайнопись – умение составлять сообщения таким образом, чтобы его смысл был недоступен никому, кроме посвященных в тайну. Раньше защиту своей информации могли позволить только очень богатые люди. И лишь несколько десятилетий назад информация приобрела самостоятельную коммерческую ценность и стала широко распространенным, почти обычным товаром.

Информацию производят, хранят, транспортируют, продают и покупают, а значит – воруют и подделывают – и, следовательно, ее необходимо защищать. Современное общество все в большей степени становится информационно–обусловленным, успех любого вида деятельности все сильней зависит от обладания определенными сведениями и от отсутствия их у конкурентов. Возникновение индустрии обработки информации с особой необходимостью привело к возникновению индустрии средств защиты информации.

Среди большого количества методов защиты данных особо выделяют криптографические методы. В отличие от других методов, они опираются лишь на свойства самой информации и не используют свойства ее материальных носителей, особенности узлов ее обработки, передачи и хранения. Образно говоря, криптографические методы строят барьер между защищаемой информацией и реальным или потенциальным злоумышленником.

Цель: Ознакомиться со способами кодирования и шифрования

Задачи: Выявить различия между кодированием и шифрованием информации, а также узнать о их методах и способах.

1.Кодирование

Код – это набор условных обозначений (или сигналов) для записи (или передачи) некоторых заранее определенных понятий.

Кодирование – это процесс формирования определенного представления информации. В более узком смысле под термином «кодирование» часто понимают переход от одной формы представления информации к другой, более удобной для хранения, передачи или обработки. Изменяет форму, но оставляет прежним содержание. Для прочтения нужно знать алгоритм и таблицу кодирования.

На компьютере можно обрабатывать текстовую информацию. При вводе в компьютер каждая буква кодируется определенным числом, а при выводе на внешние устройства (экран или печать) для восприятия человеком по этим числам строятся изображения букв. Соответствие между набором букв и числами называется кодировкой символов.

1.1.Кодирование двоичным кодом

Двоичное кодирование – один из распространенных способов представления информации. В вычислительных машинах, в роботах и станках с числовым программным управлением, как правило, вся информация, с которой имеет дело устройство, кодируется в виде слов двоичного алфавита.

Как правило, вся информация в компьютере представляются с помощью нулей и единиц. Иными словами, компьютеры обычно работают в двоичной системе счисления, поскольку при этом устройстве информация для их обработки получаются значительно более простыми. Ввод данных в компьютер и вывод их для чтения человеком может осуществляться в привычной десятичной форме, а все необходимые преобразования выполняют программы, работающие на компьютере.

Эти программы основаны на представлении данных последовательностью всего двух знаков: 0 и 1. Эти знаки называют двоичными цифрами, по-английски – binary digit или сокращённо bit (бит). Одним битом могут быть выражены два понятия: 0 или 1 (да или нет, чёрное или белое,

истина или ложь и т.п.). Если количество битов увеличить до двух, то уже можно выразить четыре различных понятия. Тремя битами можно закодировать восемь различных значений.

1.2.Кодирование целых и действительных чисел

Целые числа кодируются двоичным кодом достаточно просто - необходимо взять целое число и делить его пополам до тех пор, пока частное не будет равно единице. Совокупность остатков от каждого деления, записанная справа налево вместе с последним частным, и образует двоичный аналог десятичного числа.

Для кодирования целых чисел от 0 до 255 достаточно иметь 8 разрядов двоичного кода (8 бит). 16 бит позволяют закодировать целые числа от 0 до 65535, а 24 – уже более 16,5 миллионов различных значений.

Для кодирования действительных чисел используют 80-разрядное кодирование. При этом число предварительно преобразовывают в нормализованную форму:

$$\mathbf{3,1414926 = 0,31415926 \cdot 10^{1}}$$

$$\mathbf{300000 = 0,3 \cdot 10^{6}}$$

Первая часть числа называется мантиссой, а вторая – характеристикой. Большую часть из 80 бит отводят для хранения мантиссы (вместе со знаком) и некоторое фиксированное количество разрядов отводят для хранения характеристики.

1.3.Кодирование текстовой информации

Если каждому символу алфавита сопоставить определённое целое число, то с помощью двоичного кода можно кодировать текстовую информацию. Тексты вводятся в память компьютера с помощью клавиатуры. На клавишах написаны привычные нам буквы, цифры, знаки препинания и другие символы. В оперативную память они попадают в двоичном коде. Это значит, что каждый символ представляется 8-разрядным двоичным кодом. Восьми двоичных разрядов достаточно для кодирования 256 различных

символов. Это хватит, чтобы выразить различными комбинациями восьми битов все символы английского и русского языков, как строчные, так и прописные, а также знаки препинания, символы основных арифметических действий и некоторые общепринятые специальные символы.

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Таким образом, человек различает символы по их начертанию, а компьютер — по их коду.

Удобство побайтового кодирования символов очевидно, поскольку байт — наименьшая адресуемая часть памяти и, следовательно, процессор может обратиться к каждому символу отдельно, выполняя обработку текста.

В процессе вывода символа на экран компьютера производится обратный процесс — декодирование, то есть преобразование кода символа в его изображение. Важно, что присвоение символу конкретного кода — это вопрос соглашения, которое фиксируется в кодовой таблице.

Технически это выглядит очень просто, однако всегда существовали достаточно веские организационные сложности. В первые годы развития вычислительной техники они были связаны с отсутствием необходимых стандартов, а в настоящее время вызваны, наоборот, изобилием одновременно действующих и противоречивых стандартов. Для того чтобы весь мир одинаково кодировал текстовые данные, нужны единые таблицы кодирования, а это пока невозможно из-за противоречий между символами национальных алфавитов, а также противоречий корпоративного характера.

Для английского языка, захватившего нишу международного средства общения, противоречия уже сняты. Институт стандартизации США ввёл в действие систему кодирования ASCII (American Standard Code for Information Interchange – стандартный код информационного обмена США). В системе ASCII закреплены две таблицы кодирования базовая и расширенная. Базовая таблица закрепляет значения кодов от 0 до 127, а расширенная относится к символам с номерами от 128 до 255.

Первые 32 кода базовой таблицы, начиная с нулевого, отданы производителям аппаратных средств. В этой области размещаются управляющие коды, которым не соответствуют ни какие символы языков. Начиная с 32 по 127 код размещены коды символов английского алфавита, знаков препинания, арифметических действий и некоторых вспомогательных символов.

Кодировка символов русского языка, известная как кодировка Windows-1251, была введена «извне» - компанией Microsoft, но, учитывая широкое распространение операционных систем и других продуктов этой компании в России, она глубоко закрепилась и нашла широкое распространение.

Другая распространённая кодировка носит название КОИ-8 (код обмена информацией, восьмизначный) – её происхождение относится к временам Действия Совета Экономической Взаимопомощи государств Восточной Европы. Сегодня кодировка КОИ – 8 имеет широкое распространение в компьютерных сетях на территории России и в российском секторе Интернета.

Международный стандарт, в котором предусмотрена кодировка символов русского языка, носит названия ISO (International Standard Organization – Международный институт стандартизации). На практике данная кодировка используется редко.

Если проанализировать организационные трудности, связанные с созданием единой системы кодирования текстовых данных, то можно прийти к выводу, что они вызваны ограниченным набором кодов (256). В то же время, очевидно, что если, кодировать символы не восьмиразрядными двоичными числами, а числами с большим разрядом то и диапазон возможных значений кодов станет на много больше. Такая система, основанная на 16-разрядном кодировании символов, получила название универсальной – UNICODE. Шестнадцать разрядов позволяют обеспечить уникальные коды для 65 536 различных символов – этого поля вполне достаточно для размещения в одной таблице символов большинства языков планеты.

Несмотря на тривиальную очевидность такого подхода, простой механический переход на данную систему долгое время сдерживался из-за недостатков ресурсов средств вычислительной техники (в системе кодирования UNICODE все текстовые документы становятся автоматически вдвое длиннее). Во второй половине 90-х годов технические средства достигли необходимого уровня обеспечения ресурсами, и сегодня мы наблюдаем постепенный перевод документов и программных средств на универсальную систему кодирования.

1.4.Кодирование графических данных

Если рассмотреть с помощью увеличительного стекла чёрно-белое графическое изображение, напечатанное в газете или книге, то можно увидеть, что оно состоит из мельчайших точек, образующих характерный узор, называемый растром. Поскольку линейные координаты и индивидуальные свойства каждой точки (яркость) можно выразить с помощью целых чисел, то можно сказать, что растровое кодирование позволяет использовать двоичный код для представления графических данных. Общепринятым на сегодняшний день считается представление чёрно-белых иллюстраций в виде комбинации точек с 256 градациями серого цвета, и, таким образом, для кодирования яркости любой точки обычно достаточно восьмиразрядного двоичного числа.

Для кодирования цветных графических изображений применяется принцип декомпозиции произвольного цвета на основные составляющие. В качестве таких составляющих используют три основные цвета: красный (Red), зеленый (Green) и синий (Blue). На практике считается, что любой цвет, видимый человеческим глазом, можно получить механического смешения этих трёх основных цветов. Такая система кодирования получила названия RGB по первым буквам основных цветов.

Режим представления цветной графики с использованием 24 двоичных разрядов называется полноцветным (True Color).

Каждому из основных цветов можно поставить в соответствие дополнительный цвет, т.е. цвет, дополняющий основной цвет до белого.

Нетрудно заметить, что для любого из основных цветов дополнительным будет цвет, образованный суммой пары остальных основных цветов. Соответственно дополнительными цветами являются: голубой (Cyan), пурпурный (Magenta) и жёлтый (Yellow). Принцип декомпозиции произвольного цвета на составляющие компоненты можно применять не только для основных цветов, но и для дополнительных, т.е. любой цвет можно представить в виде суммы голубой, пурпурной и жёлтой составляющей. Такой метод кодирования цвета принят в полиграфии, но в полиграфии используется ещё и четвёртая краска – чёрная (Black). Поэтому данная система кодирования обозначается четырьмя буквами CMYK (чёрный цвет обозначается буквой К, потому, что буква В уже занята синим цветом), и для представления цветной графики в этой системе надо иметь 32 двоичных разряда. Такой режим также называется полноцветным.

Если уменьшить количество двоичных разрядов, используемых для кодирования цвета каждой точки, то можно сократить объём данных, но при этом диапазон кодируемых цветов заметно сокращается. Кодирование цветной графики 16-разрядными двоичными числами называется режимом High Color.

При кодировании информации о цвете с помощью восьми бит данных можно передать только 256 оттенков. Такой метод кодирования цвета называется индексным.

1.5.Кодирование звуковой информации

Приёмы и методы работы со звуковой информацией пришли в вычислительную технику наиболее поздно. К тому же, в отличие от числовых, текстовых и графических данных, у звукозаписей не было столь же длительной и проверенной истории кодирования. В итоге методы кодирования звуковой информации двоичным кодом далеки от стандартизации. Множество отдельных компаний разработали свои корпоративные стандарты, но среди них можно выделить два основных направления.

1.5.1.Метод FM (Frequency Modulation)

Метод FM (Frequency Modulation) основан та том, что теоретически любой сложный звук можно разложить на последовательность простейших гармонических сигналов разных частот, каждый из которых представляет собой правильную синусоиду, а, следовательно, может быть описан числовыми параметрами, т.е. кодом. В природе звуковые сигналы имеют непрерывный спектр, т.е. являются аналоговыми. Их разложение в гармонические ряды и представление в виде дискретных цифровых сигналов выполняют специальный устройства – аналогово-цифровые преобразователи (АЦП). Обратное преобразование для воспроизведения звука, закодированного числовым кодом, выполняют цифро-аналоговые преобразователи (ЦАП). При таких преобразованиях неизбежны потери информации, связанные с методом кодирования, поэтому качество звукозаписи обычно получается не вполне удовлетворительным и соответствует качеству звучания простейших электромузыкальных инструментов с окрасом характерным для электронной музыки. В то же время данный метод копирования обеспечивает весьма компактный код, поэтому он нашёл применение ещё в те годы, когда ресурсы средств вычислительной техники были явно недостаточны.

1.5.2. Метод таблично волнового (Wave-Table) синтеза

Метод таблично волнового (Wave-Table) синтеза лучше соответствует современному уровню развития техники. В заранее подготовленных таблицах хранятся образцы звуков для множества различных музыкальных инструментах. В технике такие образцы называют сэмплами. Числовые коды выражают тип инструмента, номер его модели, высоту тона, продолжительность и интенсивность звука, динамику его изменения, некоторые параметры среды, в которой происходит звучание, а также прочие параметры, характеризующие особенности звучания. Поскольку в качестве образцов исполняются реальные звуки, то его качество получается очень

высоким и приближается к качеству звучания реальных музыкальных инструментов.

2.Шифрование

Шифрование - это такое преобразование информации, которое делает исходные данные нечитаемыми и труднораскрываемыми без знания ключа.

Криптография - наука о защите информации от несанкционированного получения ее посторонними лицами. Сфера ее интересов - разработка и исследование методов шифрования информации.

Сфера интересов криптоанализа противоположная - разработка и исследование методов дешифрования шифрограммы даже без знания ключа.

Под ключом понимается секретная информация, определяющая, какое преобразование из множества возможных преобразований выполняется в данном случае над открытым текстом.

Дешифрование - обратный шифрованию процесс. На основе ключа зашифрованный текст преобразуется в исходный открытый. Процесс получения открытого сообщения из шифрованного без заранее известного ключа называется вскрытием или взломом шифра.

2.1.Виды шифров

Существует несколько классификаций шифров.

По характеру использования ключа алгоритмы шифрования можно разделить на симметричные и несимметричные.

В первом случае в шифраторе отправителя и дешифраторе получателя используется один и тот же ключ.

Во втором случае получатель вначале по открытому каналу передает отправителю открытый ключ, с помощью которого отправитель шифрует информацию. При получении информации получатель дешифрует ее с помощью второго секретного ключа.

При оценке эффективности шифра обычно руководствуются правилом Керкхоффа, согласно которому стойкость шифра определяется только секретностью ключа, т.е. известны все детали алгоритма шифрования и дешифрования, но неизвестен ключ.

Криптостойкостью называется характеристика шифра, определяющая его устойчивость к дешифрованию без знания ключа.

2.2.Шифрование с открытым ключом

Алгоритмы шифрования с открытым ключом используют так называемые необратимые функции, которые обладают следующим свойством. При заданном значении аргумента x относительно просто вычислить значение функции f(x), однако, если известно значение функции f(x), то нет простого пути для вычисления значения аргумента x.

Все используемые в настоящее время криптосистемы с открытым ключом опираются на один из следующих типов необратимых преобразований:

- Разложение больших чисел на простые множители (алгоритм RSA).
- Вычисление логарифма или возведение в степень (алгоритм DH).
- Вычисление корней алгебраических уравнений.

Например, легко в уме найти произведение двух простых чисел 11 и 13 (143). Но попробуйте быстро в уме найти два простых числа, произведение которых равно 437 (19 и 23). Такие же трудности возникают и при использовании вычислительной техники (можно, но долго). Таким образом, в системе кодирования, основанной на разложении на множители, используются два разных ключа. Ключ шифрования основан на произведении двух огромных простых чисел, а ключ дешифрования - на самих простых числах.

Было предложено (в 40-х годах 20-го века) разрабатывать шифр так, чтобы его раскрытие было эквивалентно решению сложной математической задачи. Сложность задачи должна быть такой, чтобы объем необходимых вычислений превосходил бы возможности современных ЭВМ.

В несимметричных системах приходится применять длинные ключи (1024 бит и больше). Это резко увеличивает время шифрования, генерация

ключей становится довольно длительной, зато пересылать ключи можно по открытым каналам связи.

В симметричных алгоритмах используют более короткие ключи, поэтому шифрование и дешифрование происходят быстрее. Но рассылка ключей становится сложной процедурой.

В США для передачи секретных сообщений наибольшее распространение получил стандарт DES. Он предусматривает троекратное шифрование данных разными ключами.

На протяжении всего времени дешифрованию криптограмм помогает частотный анализ появления отдельных символов и их сочетаний. Вероятности появления отдельных букв в тексте сильно различаются. Например, в русском языке буква "о" появляется в 45 раз чаще буквы "ф" и в 30 раз чаще буквы "э". Анализируя достаточно длинный текст, зашифрованный методом замены, можно по частотам появления символом произвести замену и восстановить исходный текст.

По мнению некоторых специалистов, нет не раскрываемых шифров. Рассекретить любую шифрограмму можно либо за большое время, либо за большие деньги (использование нескольких суперкомпьютеров).

Есть и другое мнение. Если длина ключа равна длине сообщения, а ключ генерируется из случайных чисел с равновероятным распределением и меняется с каждым новым сообщением, то шифр невозможно взломать даже теоретически.

2.3.Методы шифрования

Рассмотрим еще одну классификацию шифров. Множество современных методов защитных преобразований можно разделить на четыре группы:

2.3.1.Шифр перестановки

В шифре перестановки все буквы открытого текста остаются без изменений, но перемещаются с их исходных позиций на другие места.

Перестановки получаются в результате записи исходного текста и чтения шифрованного текста по разным путям некоторой геометрической фигуры.

2.3.2.Шифр замены

В шифре замены наоборот, позиции букв в шифровке остаются теми же, что и у открытого текста, но символа заменяются символами другого алфавита.

2.3.3.Аддитивный метод

В аддитивном методе буквы алфавита заменяются числами, к которым затем добавляются числа секретной псевдослучайной последовательности. Ее состав меняется в зависимости от использованного ключа. Этот метод широко используется в военных криптографических системах.

2.3.4.Комбинировынные методы

Комбинированные методы предполагают использование для шифрования сообщения сразу нескольких методов. (например, сначала замена символов, а потом их перестановка).

3.Различие между кодированием и шифрованием

На основании предыдущих пунктов повторим термины кодирования и шифрования на более простом уровне и выявим между ними отличия.

Кодирование – это перевод информации в другую форму, более удобную для хранения, обработки или передачи. Например, любая информация, вводимая в компьютер, автоматически подвергается кодированию в двоичную систему счисления, чтобы компьютер мог ее распознать и в дальнейшем совершать над ней какие-то действия.

Шифрование, по сути, тоже является кодированием, но преследует другие цели. Это изменение информации, с целью обеспечения ее конфиденциальности и защиты данных от сторонних лиц.

Отличия кодирования от шифрования:

1. Кодирование информации происходит по стандартным алгоритмам, понятным большинству приборов. Для шифрования используются специальные схемы изменения данных, непременным элементом которых является ключ. С его помощью пользователь или автоматизированная система дешифровки впоследствии сможет получить доступ к информации, выбрав нужный алгоритм дешифрования. В алгоритмах кодирования ключей не предусмотрено.
2. Кодировка — перевод в другой формат, шифрование — способ приведения данных в неудобоваримый для посторонних вид.
3. Кодирование по вполне себе стационарному, стандартному, алгоритму. Шифрование по псевдослучайному, с целью максимально затруднить дешифровку.
4. Шифрование - это способ изменения сообщения или другого документа, обеспечивающее искажение (сокрытие) его содержимого. Кодирование - это преобразование обычного, понятного, текста в код. При этом подразумевается, что существует взаимно однозначное соответствие между символами

текста (данных, чисел, слов) и символьного кода в этом принципиальное отличие кодирования от шифрования.

5. Кодирование — преобразование информации с целью обеспечить удобство ее хранения или передачи. Нет никакого засекречивания. Это просто перевод в другой формат, который по какой-то причине более удобен

   Шифрование — преобразование информации с целью затруднить или сделать невозможным понимание или изменение этой информации неавторизованными лицами в случае перехвата. Здесь есть засекречивание.

Заключение

В данной работе подробно рассмотрены способы кодирования двоичным кодом различной информацией. Также выяснены методы шифрования с ключом и без него.

Были выявлены несколько различий между кодированием и шифрованием информации.

Список литературы

1. Агеев В.М. Теория информации и кодирования: дискретизация и кодирование измерительной информации. — М.: МАИ, 1977.
2. Кузьмин И.В., Кедрус В.А. Основы теории информации и кодирования. — Киев, Вища школа, 1986.
3. Простейшие методы шифрования текста/ Д.М. Златопольский. – М.: Чистые пруды, 2007 – 32 с.
4. Угринович Н.Д. Информатика и информационные технологии. Учебник для 10-11 классов / Н.Д.Угринович. – М.: БИНОМ. Лаборатория знаний, 2003. – 512 с.
5. Кузьмин И.В. Теоретические основы информационной техники. – Х.: ХВКИУ, 1969. – 162 с.
6. Шастова Г.А. Кодирование и помехоустойчивость передачи телемеханической информации. – М.: Энергия, 1966. – 454 с.
7. Кодирование информации. Двоичные коды: Справочник / Под ред. Н. Т. Березнюка. – Х.: Вища шк., Изд-во при Харьк. ун-те, 1978. – 252 с.
8. Заренин Ю. Г. Корректирующие коды для передачи и обработки информации. – К.: Техніка, 1965. – 170 с.
9. Удалов А. П., Супругин Б. А. Избыточное кодирование при передаче информации двоичными кодами. – М.: Связь, 1964. – 270 с.
10. Элиас П. Безошибочное кодирование // Коды с обнаружением и исправлением ошибок / Под ред. А. М. Петровского. - М.. 1956. – С. 59-71.
11. Элиас П. Кодирование и декодирование // Лекции по теории систем связи / Под ред. К. Дж. Багдади. – М., 1964. – С. 289-317.
12. Возенкрафт Дж., Рейфен Б. Последовательное декодирование. – М.: Изд-во иностр. лит., 1963. – 156 с.
13. Оливер Б. Эффективное кодирование // Теория информации и ее приложения. – М., 1959. С. 1-15.

14. Котов П.А. Повышение достоверности передачи цифровой информации. – М.: Связь, 1966. – 184 с.

15. Мартынов Ю.М. Обработка информации в системах передачи данных. – М.: Связь, 1969 – 263 с.