Титулка

ЗМІСТ

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

СЛОВАРЬ ТЕРМИНОВ

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

Шифротекст, шифрованный (закрытый) текст — данные, полученные после применения криптосистемы.

Шифр, криптосистема — совокупность заранее оговоренных способов преобразования исходного секретного сообщения с целью его защиты.

Символ — это любой знак, в том числе буква, цифра или знак препинания.

Алфавит — конечное множество используемых для кодирования информации символов. Стандартный алфавит может быть изменён или дополнен символами. **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.

Система шифрования (шифрсистема) — это любая система, которую можно использовать для обратимого изменения текста сообщения с целью сделать его непонятным для всех, кроме адресата.

Криптостойкостью — это характеристика шифра, определяющая его стойкость к дешифрованию без знания ключа (т.е. способность противостоять криптоанализу).

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(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.

ВВЕДЕНИЕ

Цель практических работ состоит в изучении основных концепций информационной безопасности, понимание уровней информационной безопасности и целей. Определение угроз на аппаратном и сетевом уровнях.

Первая работа изучить основные понятия и уровни информационной безопасности, составляющие и виды информационных угроз. После чего подготовиться к тестированию по заданным аспектам. На занятии разобрать ситуацию по варианту или предложенную руководителем.

Вторая работа данная работа предполагает настройку устройства, предположительно компьютера. В практическом занятии студент должен провести настройку компьютера, целью является защита от самых распространённых ошибок допускаемыми системными администраторами небольших фирм. После чего протестировать и оформить результаты в отчёт.

Третья работа проведение семинара предполагает ознакомить студента с основными стандартами информационной безопасности. Изучить сервисы и механизмы защиты. Так же предполагает разбор нескольких ситуаций из примеров или предложенные студентами.

Четвёртая работа ознакомиться с одним из методов криптографического преобразования информации, а именно стеганографией. Рассмотреть примеры сокрытия данных в файле, использование шумов и стохастической модуляции. Реализовать преобразование одним из методов.

Пятая работа разделена на две части. Первая, предполагает изучение методов сжатия данных. Изучение алгоритма Хаффмана и Лемпеля-Зива, реализация сжатия больших текстов и оценка актуальности. Вторая, рассматривает резервное сохранение данных, как объект защиты целостности. Ознакомление с возможностями резервного сохранения и различиями реализаций.

Шестая работа является базовой по настройке и работе в сетях, рассматривается вариант «белой», безопасной сети, её подключение и общая настройка прав. Работы выполняются в среде Cisco. Можно использовать другое ПО, если оно предоставляет требуемый функционал.

Седьмая работа – это продолжение шестой работы, где студент должен будет реализовать безопасное подключение всей сети к мировой сети Интернет.

Восьмая работа настроена на исследование анализа больших потоков данных, прослушивание сети. Тут нужно будет дописать

1 ОСНОВНЫЕ ПОНЯТИЯ БЕЗОПАСНОСТИ

Тема: История, основные понятия и уровни информационной безопасности.

Цель: Изучить основные понятия и уровни информационной безопасности, составляющие и виды информационных угроз.

Теоретические ведомости

1 Основные понятия информационной безопасности

Обеспечение защиты информации волновало человечество всегда. В процессе эволюции цивилизации менялись виды информации, для её защиты применялись различные методы и средства.

Процесс развития средств и методов защиты информации можно разделить на три относительно самостоятельных периода:

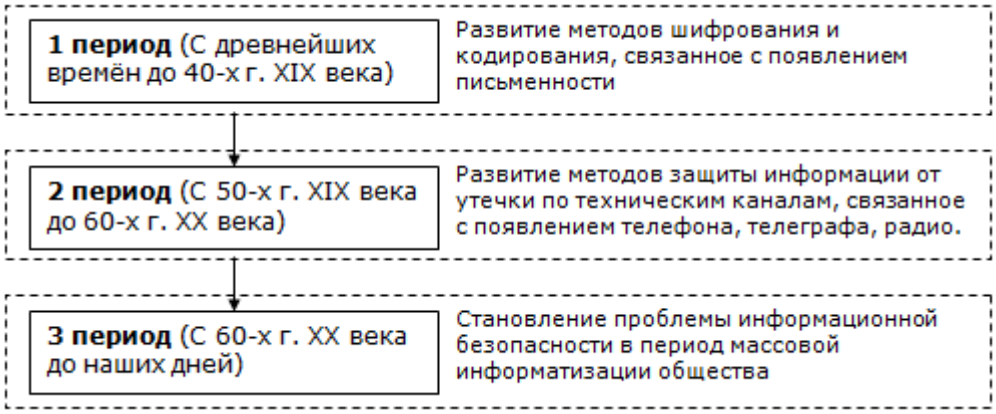

Рисунок 1.1 – Периоды развития

Наблюдаемые в последние годы тенденции в развитии информационных технологий могут уже в недалеком будущем привести к появлению качественно новых (информационных) форм борьбы, в том числе и на межгосударственном уровне, которые могут принимать форму информационной войны, а сама информационная война станет одним из основных инструментов внешней политики, включая защиту государственных интересов и реализацию любых форм агрессии. Это является одной из причин,

почему полезно ознакомиться с основными принципами обеспечения ИБ в ведущих зарубежных странах.

Прежде чем говорить об обеспечении безопасности персональных данных, необходимо определить, что же такое информационная безопасность. Термин "информационная безопасность" может иметь различный смысл и трактовку в зависимости от контекста. В данном пособии под информационной безопасностью мы будем понимать защищённость информации и поддерживающей инфраструктуры от случайных или преднамеренных воздействий естественного или искусственного характера, которые могут нанести неприемлемый ущерб субъектам информационных отношений, в том числе владельцам и пользователям информации и поддерживающей инфраструктуры.[2]

Информационная безопасность – это защищённость информации и поддерживающей её инфраструктуры от случайных или преднамеренных воздействий естественного или искусственного характера, которые могут нанести ущерб владельцам или пользователям информации.

В ряде случаев понятие "информационная безопасность" подменяется термином "компьютерная безопасность". В этом случае информационная безопасность рассматривается очень узко, поскольку компьютеры только одна из составляющих информационных систем. Несмотря на это, в рамках изучаемого курса основное внимание будет уделяться изучению вопросов, связанных с обеспечением режима информационной безопасности применительно к вычислительным системам, в которых информация хранится, обрабатывается и передаётся с помощью компьютеров. Согласно определению, компьютерная безопасность зависит не только от компьютеров, но и от поддерживающей инфраструктуры, к которой можно отнести системы электроснабжения, жизнеобеспечения, вентиляции, средства коммуникаций, а также обслуживающий персонал.

2 Составляющие информационной безопасности

Обеспечение информационной безопасности в большинстве случаев связано с комплексным решением трёх задач:

Конфиденциальность – состояние информации, при котором доступ к ней осуществляют только субъекты, имеющие на него право.

Целостность – состояние информации, при котором отсутствует любое её изменение либо изменение осуществляется только преднамеренно субъектами, имеющими на него право;

Доступность – состояние информации, при котором субъекты, имеющие право доступа, могут реализовывать его беспрепятственно.

Угрозы информационной безопасности – совокупность условий и факторов, создающих потенциальную или реально существующую опасность нарушения безопасности информации.[1, 4]

Атака – это попытка реализации угрозы. Кто предпринимает такую попытку, называется *злоумышленником*. Потенциальные злоумышленники называются *источниками угрозы*.

Угроза является следствием наличия уязвимых мест или уязвимости в информационной системе. Уязвимости могут возникать по разным причинам, например, в результате непреднамеренных ошибок программистов при написании программ.

Угрозы можно классифицировать по нескольким критериям:

- по свойствам информации (доступность, целостность, конфиденциальность), против которых угрозы направлены в первую очередь;
- по компонентам информационных систем, на которые угрозы нацелены (данные, программы, аппаратура, поддерживающая инфраструктура);
- по способу осуществления (случайные/преднамеренные, действия природного/техногенного характера);
- по расположению источника угроз (внутри/вне рассматриваемой ИС).

Обеспечение информационной безопасности является сложной задачей, для решения которой требуется комплексный подход.

3 Уровни защиты информации

3.1 Законодательный уровень

Законодательный уровень является основой для построения системы защиты информации, так как даёт базовые понятия предметной области и определяет меру наказания для потенциальных злоумышленников. Этот уровень играет координирующую и направляющую роли и помогает поддерживать в обществе негативное (и карательное) отношение к людям, нарушающим информационную безопасность.

Законодательно-правовой уровень включает комплекс законодательных и иных правовых актов, устанавливающих правовой статус субъектов информационных отношений, субъектов и объектов защиты, методы, формы и способы защиты, их правовой статус. Кроме того, к этому уровню относятся стандарты и спецификации в области информационной безопасности. Система законодательных актов и разработанных на их базе нормативных и организационно-распорядительных документов должна обеспечивать организацию эффективного надзора за их исполнением со стороны правоохранительных органов и реализацию мер судебной защиты и ответственности субъектов информационных отношений. К этому уровню можно отнести и морально-этические нормы поведения, которые сложились традиционно или складываются по мере распространения вычислительных средств в обществе. Морально-этические нормы могут быть регламентированными в законодательном порядке, т. е. в виде свода правил и предписаний. Наиболее характерным примером таких норм является Кодекс профессионального поведения членов Ассоциации пользователей ЭВМ США. Тем не менее, эти нормы большей частью не являются обязательными, как законодательные меры.

3.2 Административный уровень

Это комплекс мер, предпринимаемых локально руководством организации. Включает комплекс взаимокоординируемых мероприятий и технических мер, реализующих практические механизмы защиты в процессе создания и эксплуатации систем защиты информации. Организационный уровень должен охватывать все структурные элементы систем обработки данных на всех этапах их жизненного цикла: строительство помещений, проектирование системы, монтаж и наладка оборудования, испытания и проверки, эксплуатация.

Разработка политики безопасности - дело тонкое, поскольку у каждой организации есть своя специфика. Здесь бессмысленно переносить практику режимных государственных организаций на коммерческие структуры, учебные заведения или персональные компьютерные системы. В этой области целесообразно предложить, во-первых, основные принципы разработки политики безопасности, а, во-вторых, - готовые шаблоны для наиболее важных разновидностей организаций.

3.3 Процедурный уровень

К процедурному уровню относятся меры безопасности, реализуемые людьми. В отечественных организациях накоплен богатый опыт составления и реализации процедурных (организационных) мер, однако проблема состоит в том, что они пришли из докомпьютерного прошлого, поэтому нуждаются в существенном пересмотре.

Можно выделить следующие группы процедурных мер:

- управление персоналом;
- физическая защита;
- поддержание работоспособности;
- реагирование на нарушения режима безопасности;
- планирование восстановительных работ.

Для каждой группы в каждой организации должен существовать набор регламентов, определяющих действия персонала. В свою очередь, исполнение этих регламентов следует отработать на практике.

3.4 Программно-технический уровень

Согласно современным воззрениям, включает три подуровня: физический, технический (аппаратный) и программный.

Физический подуровень решает задачи с ограничением физического доступа к информации и информационным системам, соответственно к нему относятся технические средства, реализуемые в виде автономных устройств и систем, не связанных с обработкой, хранением и передачей информации: система охранной сигнализации, система наблюдения, средства физического воспрепятствования доступу (замки, ограждения, решётки и т. д.).

Средства защиты аппаратного и программного подуровней непосредственно связаны с системой обработки информации. Эти средства либо встроены в аппаратные средства обработки, либо сопряжены с ними по стандартному интерфейсу.

К аппаратным средствам относятся схемы контроля информации по четности, схемы доступа по ключу и т. д.

К программным средствам защиты, образующим программный подуровень, относятся специальное программное обеспечение, используемое для защиты информации, например антивирусный пакет и т. д. Программы защиты могут быть как отдельные, так и встроенные. Подчеркнём, что формирование режима информационной безопасности является сложной системной задачей, решение которой в разных странах отличается по содержанию и зависит от таких факторов, как научный потенциал страны, степень внедрения средств информатизации в жизнь общества и экономику, развитие производственной базы, общей культуры общества и, наконец, традиций и норм поведения.

4 Виды информационных угроз

Информационные угрозы могут быть обусловлены:

– естественными факторами (пожар, наводнение, и др.);

– человеческими факторами.

Последние, в свою очередь, подразделяются на:

– Угрозы, носящие случайный, неумышленный характер. Это угрозы, связанные с ошибками процесса подготовки, обработки и передачи информации;

– Угрозы, обусловленные умышленными, преднамеренными действиями людей. Это угрозы, связанные с несанкционированным доступом к ресурсам АИС.

Умышленные угрозы преследуют цель нанесения ущерба пользователям АИС и, в свою очередь, подразделяются на активные и пассивные. Угрозы также подразделяются на внутренние, возникающие внутри управляемой организации, и внешние.

Под внутренними угрозами – понимаются угрозы безопасности информации инсайдером (исполнителем) которых является внутренний по отношению к ресурсам организации субъект (инсайдер).

Под внешними угрозами – понимаются угрозы безопасности информации инициатором (исполнителем) которых является внешний по отношению к ресурсам организации субъект (удаленный хакер, злоумышленник).

Задания

5 Тестирование

1) В чем заключается проблема информационной безопасности?
2) Дайте определение понятию «информационная безопасность».
3) Что понимается под «компьютерной безопасностью»?

4) Перечислите составляющие информационной безопасности.
5) Приведите определение доступности информации.
6) Приведите определение целостности информации.
7) Приведите определение конфиденциальности информации.
8) Каким образом взаимосвязаны между собой составляющие информационной безопасности? Приведите собственные примеры.
9) Перечислите задачи информационной безопасности общества.
10) Перечислите уровни формирования режима информационной безопасности.
11) Дайте краткую характеристику законодательно-правового уровня.
12) Какие подуровни включает программно-технический уровень?
13) Что включает административный уровень?
14) В чем особенность морально-этического подуровня?

6 Рассмотрение ситуации

Оценив ситуацию соответствующую варианту нужно:

1) Определить источник угрозы.
15) Пострадавшее лицо.
16) Классифицировать вид угрозы.
17) Определить угрозу доступности, целостности, конфиденциальности.
18) Организовать меры по защите.

Так же организовать меры по защите информации в данных обстоятельствах и дальнейшее упреждение данной модели.

Пример выполнения работы

Сотрудница отделения коммерческого банка разместила фото с id своей карты в социальной сети.

В данной ситуации мы можем явно видеть, что сотрудник допустил халатность. В результате чего безопасность компании ставиться под вопрос.

1) Источником угрозы является сотрудница, а так же любые лица пытающиеся проникнуть в административную часть здания с поддельным пропуском на её имя.

2) Пострадавшим лицом является учреждение, в частности отдел по безопасности данного объекта. При бездействии круг пострадавших лиц может сильно увеличиться.

3) Классификация угрозы:

– угроза обусловлена человеческим фактором;

– носящим случайный, неумышленный характер;

– угроза является внутренней.

4) Такие аспекты безопасности как доступность и целостность не нарушены. В данном контексте нарушена только конфиденциальность рабочих пропусков компании.

5) Меры по защите должны включать:

а) Немедленное блокирование пропуска сотрудницы, выдача нового.

б) Усиленная проверка входящих в здание по пропускам в течении недели.

в) Проверка персонала, находящегося в здании.

г) Добавление/удаление пропусков происходят в следящем режиме.

д) Сверка активности сотрудницы.

е) Провести инструктаж на тему "Политика безопасности в организации".

Варианты задания

0) Сотрудница отделения коммерческого банка разместила фото с id своей карты в социальной сети;

1) В СМИ утекли результаты анализов одного из известных деятелей;

2) Ученик, взломав систему оценивания колледжа исправил себе бал по дисциплине;

3) Во время грозы были повреждены электролинии. В связи с этим более 200 клиентов охранной компании остались без наблюдения на 10 часов;

4) Используя брешь в интернет-сети страховой компании, хакер заменил данные нескольких клиентов;

5) Интернет-магазин использует небезопасный канал. Клиент, совершив покупку передал сумму третьему лицу;

6) Подкуплен сотрудник, после чего неизвестный проник в здание отделения полиции.

7) Сотрудник аудиторской компании использовал данные в своих целях;

8) После взлома сервера компании по информационной защите ключи доступа пользователей появились на «чёрном рынке»;

9) Ночью из офиса была украдена печать адвоката, объект находится под охраной;

10) Сотрудник компании по разработке ПО скрыто вставлял мониторинг в продукт;

11) Сбой в работе компании по обеспечению vps серверов;

12) Обнаружен задержка интернет канала биржи. Предположительно злоумышленники, подключившись к каналу получают данные первыми;

13) Сотрудник не соблюдал правила производства. В связи с чем завод потерял несколько партий продукта.

14) Зависание информационной системы на железной дороге привело к столкновению поездов.

15) Офис туристической компании был затоплен во время стихийного бедствия.

2 ЛОКАЛЬНАЯ ЗАЩИТА УСТРОЙСТВ

Тема: Настройка системы от внутренних и внешних угроз.

Цель: Определить цели и методы для настройки устройств, на практике реализовать защиту локального устройства.

Теоретические ведомости

Задания

Ход работы

Вопросы для самоконтроля

3 МИРОВЫЕ СТАНДАРТЫ БЕЗОПАСНОСТИ

Тема: Стандарты информационной безопасности распределённых систем.

Цель: Изучить сервисы и механизмы защиты распределённых систем. Разбор планирования систем.

Требования к семинару

1 Студент должен знать

1) Основное содержание стандартов по информационной безопасности распределённых систем;
2) Основные сервисы безопасности в вычислительных сетях;
3) Наиболее эффективные механизмы безопасности;
4) Задачи администрирования средств безопасности.

2 Студен должен уметь

5) Выбирать механизмы безопасности для защиты распределенных систем.

Термины для подготовки

Распределённая информационная система – совокупность аппаратных и программных средств, используемых для накопления, хранения, обработки, передачи информации между территориально удалёнными пользователями.

Сервис (*Сервисная деятельность*) – это вид деятельности, направленный на удовлетворение потребностей социальных субъектов посредством оказания услуг.

Сервис безопасностис – это деятельность государственных и частных организаций, а также отдельных специалистов, направленная на удовлетворение потребностей социальных субъектов в безопасности.

Цель сервиса безопасности – удовлетворение потребностей в безопасности индивидуальных и групповых социальных субъектов. *Сущность сервиса* безопасности состоит в оказании услуг, направленных на обеспечение безопасности.

Услуга безопасности – это деятельность субъекта безопасности, направленная на удовлетворение потребности заказчика в безопасности, а также результат взаимодействия исполнителя и заказчика услуги безопасности, выраженный в виде полезного эффекта.

Темы для обсуждения

1) Механизмы безопасности.
2) Сервисы безопасности в вычислительных сетях.
3) Функций и механизмов безопасности.
4) Администрирование средств безопасности.
5) Международные стандарты.
6) Стандарты ГОСТ и ДСТУ.

Литература для ознакомления

1. Щербаков А. Ю. Введение в теорию и практику компьютерной безопасности. – М.: Издательство Молгачева С. В., 2001.
2. Теория и практика обеспечения информационной безопасности / Под ред. П. Д. Зегжды. – М: Яхтсмен, 1996.
3. Галатенко В. А. Основы информационной безопасности. – М: Интернет-Университет Информационных Технологий – ИНТУИТ. РУ, 2003.
4. Галатенко В. А. Стандарты информационной безопасности. – М: Интернет-Университет Информационных Технологий – ИНТУИТ. РУ, 2004.
5. www.iso.ch – Web-сервер Международной организации по стандартизации.

4 МЕТОДЫ СОКРЫТИЯ ИНФОРМАЦИИ

Тема: Сокрытие информации. Основы стеганографии.

Цель: Ознакомиться с методом стеганографии. Рассмотреть примеры сокрытия информации в файлах, реализовать один из методов.**Теоретические ведомости**

Задания

Ход работы

Вопросы для самоконтроля

5 АРХИВАЦИЯ И РЕЗЕРВНОЕ КОПИРОВАНИЕ ДАННЫХ

Тема: Архивация и резервное копирование данных.

Цель: Изучение методов сжатия данных, алгоритма Хаффмана и Лемпеля-Зива. Рассмотреть резервное копирование данных.

Архивация данных

Архивация (сжатие данных) – есть процесс представления информации в ином виде (перекодирования) с потенциальным уменьшением объёма, требуемого для её хранения. Существует множество классов различных алгоритмов сжатия данных, каждый из которых ориентирован на свою область применения[3].

Основоположником науки о сжатии информации принято считать *Клода Шеннона*. Его теорема об оптимальном кодировании показывает, к чему нужно стремиться при кодировании информации и насколько та или иная информация при этом сожмется. Кроме того, им были проведены опыты по эмпирической оценке, избыточности английского текста. Шеннон предлагал людям угадывать следующую букву и оценивал вероятность правильного угадывания. На основе ряда опытов он пришел к выводу, что количество информации в английском тексте колеблется в пределах $0,6 - 1,3$ бита на символ. Несмотря на то, что результаты исследований Шеннона были по-настоящему востребованы лишь десятилетия спустя, трудно переоценить их значение.

Сжатие данных – это процесс, обеспечивающий уменьшение объёма данных путём сокращения их избыточности. Сжатие данных связано с компактным расположением порций данных стандартного размера. Сжатие данных можно разделить на два основных типа:

Сжатие без потерь (полностью обратимое) – это метод сжатия данных, при котором ранее закодированная порция данных восстанавливается после

их распаковки полностью без внесения изменений. Для каждого типа данных, как правило, существуют свои оптимальные алгоритмы сжатия без потерь.

Сжатие с потерями – это метод сжатия данных, при котором для обеспечения максимальной степени сжатия исходного массива данных часть содержащихся в нём данных отбрасывается. Для текстовых, числовых и табличных данных использование программ, реализующих подобные методы сжатия, является неприемлемыми. В основном такие алгоритмы применяются для сжатия аудио и видеоданных, статических изображений.

1 Алгоритмы архивации данных

Алгоритм сжатия данных – это алгоритм, который устраняет избыточность записи данных.

Отношение сжатия – одна из наиболее часто используемых величин для обозначения эффективности метода сжатия.

$$Отношение\ сжатия = \frac{размер\ \ выходного\ \ потока}{размер\ \ входного\ \ потока} \qquad (5.1)$$

Значение 0,6 означает, что данные занимают 60% от первоначального объема. Значения больше 1 означают, что выходной поток больше входного (отрицательное сжатие, или расширение).

Коэффициент сжатия – величина, обратная отношению сжатия.

$$Коэффициент\ \ сжатия = \frac{размер\ входного\ потока}{размер\ выходного\ потока} \qquad (5.2)$$

Значения больше 1 обозначают сжатие, а значения меньше 1 расширение.

Средняя длина кодового слова – это величина, которая вычисляется как взвешенная вероятностями сумма длин всех кодовых слов.

$$L_{cp} = p_1 \cdot L_1 + p_2 \cdot L_2 + \ldots + p_n \cdot L_n, \qquad (5.3)$$

где p_n – вероятности кодовых слов, L_1, L_2, L_3 – длины кодовых слов.

Статистические методы – методы сжатия, присваивающие коды переменной длины символам входного потока, причем более короткие коды присваиваются символам или группам символам, имеющим большую вероятность появления во входном потоке. Лучшие статистические методы применяют кодирование Хаффмана.

Словарное сжатие – это методы сжатия, хранящие фрагменты данных в "словаре" (некоторая структура данных). Если строка новых данных, поступающих на вход, идентична какому-либо фрагменту, уже находящемуся в словаре, в выходной поток помещается указатель на этот фрагмент. Лучшие словарные методы применяют метод Зива-Лемпела.

Рассмотрим несколько известных алгоритмов сжатия данных более подробно.

1.1 Алгоритм Хаффмана

В основе алгоритма Хаффмана лежит идея кодирования битовыми группами. Сначала проводится частотный анализ входной последовательности данных, то есть устанавливается частота вхождения каждого символа, встречащегося в ней. После этого, символы сортируются по уменьшению частоты вхождения.

Основная идея состоит в следующем: чем чаще встречается символ, тем меньшим количеством бит он кодируется. Результат кодирования заносится в словарь, необходимый для декодирования. Рассмотрим простой пример, иллюстрирующий работу алгоритма Хаффмана.

Пусть задан текст «beep boop beer!», рассмотрим таблицу с частотами всех символов:

Таблица 8.1 — Частота символов

Символ	'b'	'e'	'p'	' '	'o'	'r'	'!'
Частота	3	4	2	2	2	1	1

По частоте использования

Символ	'r'	'!'	'p'	'o'	' '	'b'	'e'

После этого создадим элементы бинарного дерева для каждого символа и представим их как очередь с приоритетом, в качестве которого будем использовать частоту.

Возьмём первые два элемента из очереди и создадим третий(рис. 5.1), который будет их родителем. Этот новый элемент поместим в очередь с приоритетом, равным сумме приоритетов двух его потомков. Иначе говоря, равным сумме их частот.

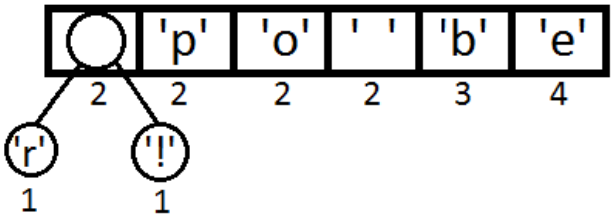

Рисунок 5.1– Пример объединения элементов

Далее будем повторять шаги, аналогичные предыдущему:

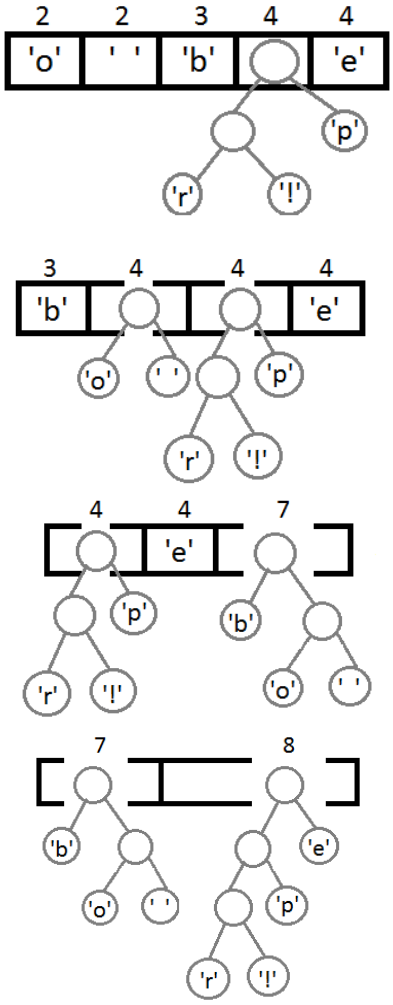

Рисунок 5.2 – Построение дерева

Теперь, после объединения последних двух элементов с помощью их нового родителя, мы получим итоговое бинарное дерево:

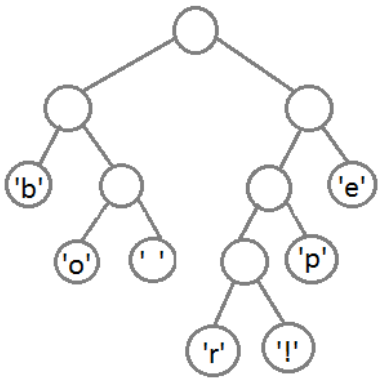

Рисунок 5.3 – Вид бинарного дерева

Осталось присвоить каждому символу его код(рис. 1.1). Для этого запустим обход в глубину и каждый раз, рассматривая правое поддерево, будем записывать в код 1, а рассматривая левое поддерево – 0.

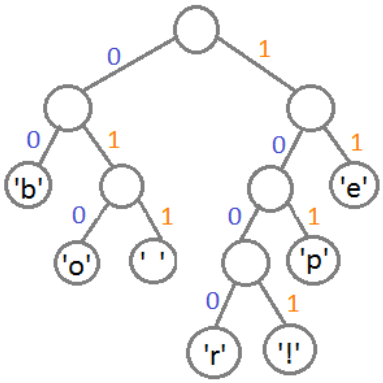

Рисунок 5.4 – Построение кода

В результате соответствие символов кодовым значениям получится следующим:

Таблица 8.2 — Кодовые значения символов

Сивол	'b'	'e'	'p'	' '	'o'	'r'	'!'
Кодовое значение	00	11	101	011	010	1000	1001

Декодирование битов происходит следующим образом: нужно обходить дерево, отбрасывая левое поддерево, если встретилась единица и правое, если встретился 0. Продолжать обход нужно до тех пор, пока не встретим лист, т.е. искомое значение закодированного символа.

Например, закодированной строке «101 11 101 11» и нашему дереву декодирования соответствует строка «pepe».

Входная строка:

```
beep boop beer!
```

Входная строка в двоичном виде:

```
0110 0010 0110 0101 0110 0101 0111 0000 0010 0000 0110
0010 0110 1111 0110 1111 0111 0000 0010 0000 0110 0010
0110 0101 0110 0101 0111 0010 0010 0001
```

Закодированная строка:

```
0011 1110 1011 0001 0010 1010 1100 1111 1000 1001
```

Разница между ASCII-кодировкой строки и её же видом в коде Хаффмана очевидна.

Алгоритм Хаффмана универсальный, его можно применять для сжатия данных любых типов, но он малоэффективен для файлов маленьких размеров (за счет необходимости сохранения словаря). В настоящее время данный метод практически не применяется в чистом виде, обычно используется как один из этапов сжатия в более сложных схемах. Это единственный алгоритм, который не увеличивает размер исходных данных в худшем случае (если не считать необходимости хранить таблицу перекодировки вместе с файлом).

1.2 Алгоритм Лемпеля-Зива

Процесс сжатия выглядит следующим образом. Последовательно считываются символы входного потока и происходит проверка, существует ли в созданной таблице строк такая строка. Если такая строка существует, считывается следующий символ, а если строка не существует, в поток заносится код для предыдущей найденной строки, строка заносится в таблицу, а поиск начинается снова. Например, если сжимают байтовые данные (текст),

то строк в таблице окажется 256 (от «0» до «255»). Если используется 10-битный код, то под коды для строк остаются значения в диапазоне от 256 до 1023. Новые строки формируют таблицу последовательно, т. е. можно считать индекс строки ее кодом. Алгоритму декодирования на входе требуется только закодированный текст, поскольку он может воссоздать соответствующую таблицу преобразования непосредственно по закодированному тексту. Алгоритм генерирует однозначно декодируемый код за счет того, что каждый раз, когда генерируется новый код, новая строка добавляется в таблицу строк. LZW постоянно проверяет, является ли строка уже известной, и, если так, выводит существующий код без генерации нового. Таким образом, каждая строка будет храниться в единственном экземпляре и иметь свой уникальный номер. Следовательно, при дешифровании при получении нового кода генерируется новая строка, а при получении уже известного, строка извлекается из словаря.[5]

Кодирование

Пусть мы сжимаем последовательность <<abacabadabacabae>>.

Текущая строка	Текущий символ	Следующий символ	Вывод		Словарь
			Код	Биты	
ab	a	b	0	000	5: ab
ba	b	a	1	001	6: ba
ac	a	c	0	000	7: ac
ca	c	a	2	010	8: ca
ab	a	b	-	-	- -
aba	b	a	5	101	9: aba
ad	a	d	0	000	10: ad
da	d	a	3	011	11: da
ab	a	b	-	-	- -
aba	b	a	-	-	- -
abac	a	c	9	1001	12: abac
ca	c	a	-	-	- -
cab	a	b	8	1000	13: cab
ba	b	a	-	-	- -
bae	a	e	6	0110	14: bae
e	e	-	4	0100	- -

Мы получаем закодированное сообщение:

```
0 1 0 2 5 0 3 9 8 6 4
```

что на 11- бит короче.

```
000 001 000 010 101 000 011 1001 1000 0110 0100
```

Декодирование

Особенность LZW заключается в том, что для декомпрессии нам не надо сохранять таблицу строк в файл для распаковки. Алгоритм построен таким образом, что мы в состоянии восстановить таблицу строк, пользуясь только потоком кодов. Теперь представим, что мы получили закодированное сообщение, приведённое выше, и нам нужно его декодировать. Прежде всего, нам нужно знать начальный словарь, а последующие записи словаря мы

можем реконструировать уже на ходу, поскольку они являются просто конкатенацией предыдущих записей.

Данные		На выходе	Новая запись	
Биты	Код		Полная	Частичная
000	0	a	- -	5: a?
001	1	b	5: ab	6: b?
000	0	a	6: ba	7: a?
010	2	c	7: ac	8: c?
101	5	ab	8: ca	9: ab?
000	0	a	9: aba	10: a?
011	3	d	10: ad	11: d?
1001	9	aba	11: da	12: aba?
1000	8	ca	12: abac	13: ca?
0110	6	ba	13: cab	14: ba?
0100	4	e	14: bae	- -

Достоинства и недостатки

+ Не требует вычисления вероятностей встречаемости символов или кодов.
+ Для декомпрессии не надо сохранять таблицу строк в файл для распаковки. Алгоритм построен таким образом, что мы в состоянии восстановить таблицу строк, пользуясь только потоком кодов.
+ Данный тип компрессии не вносит искажений в исходный графический файл, и подходит для сжатия растровых данных любого типа.
– Алгоритм не проводит анализ входных данных поэтому не оптимален.

Задания

Часть первая

1) Взять данные соответственно варианту из таблицы
2) Удалить лишнюю информацию методом Хаффмана.

3) Провести операцию методом Лемпеля-Зива.

4) Сравнить результаты проведённых операций.

5) Описать актуальность архивации для различных объёмов данных.

6) Сделать выводы по применению методов сжатия в различных криптосистемах.

Часть вторая

Ход работы

Вопросы для самоконтроля

6 CISCO. ЧАСТЬ 1

Тема: Тема.

Цель: Опишите цель.

Теоретические ведомости

Задания

Ход работы

Вопросы для самоконтроля

7 CISCO. ЧАСТЬ 2

Тема: Тема.

Цель: Опишите цель.

Теоретические ведомости

Задания

Ход работы

Вопросы для самоконтроля

8 АНАЛИЗ ДАННЫХ

Тема: Тема.

Цель: Опишите цель.

Теоретические ведомости

Задания

Ход работы

Вопросы для самоконтроля

ДОДАТОК А

ДОДАТОК Б