А. И. Куприянов
А. В. Сахаров
В. А. Шевцов

ОСНОВЫ ЗАЩИТЫ ИНФОРМАЦИИ

Высшее профессиональное образование

Учебное пособие

Радиоэлектроника

ACADEMA

ВЫСШЕЕ ПРОФЕССИОНАЛЬНОЕ ОБРАЗОВАНИЕ

А. И. КУПРИЯНОВ, А. В. САХАРОВ, В. А. ШЕВЦОВ

ОСНОВЫ ЗАЩИТЫ ИНФОРМАЦИИ

Допущено
Учебно-методическим объединением
по образованию в области авиации, ракетостроения и космоса
в качестве учебного пособия для студентов, обучающихся
по специальностям «Радиоэлектронные системы», «Средства радиоэлектронной борьбы» и «Информационные системы и технологии»

Москва
ACADEMA
2006

УДК 621.37(075.8)
ББК 32.84я73
К924

Рецензенты:
д-р техн. наук, проф. кафедры «Защита информации»
Московского государственного технического университета им. Н.Э. Баумана
П.Б.Петренко;
д-р техн. наук, проф., заслуженный деятель науки и техники РФ *Е.М. Сухарев*

Куприянов А. И.

К924 Основы защиты информации : учеб. пособие для студ. высш. учеб. заведений / А. И. Куприянов, А. В. Сахаров, В. А. Шевцов. — М. : Издательский центр «Академия», 2006. — 256 с.

ISBN 5-7695-2438-3

Рассмотрены основные проблемы, теоретические положения, потенциальные и технически достижимые характеристики качества, а также технические решения при построении систем защиты важнейшего современного ресурса — информационного — от негативных и деструктивных воздействий, характеризующих конфликт информационных систем с техническими средствами разведки.

Для студентов высших учебных заведений. Может быть полезно специалистам в области защиты информации.

УДК 621.37(075.8)
ББК 32.84я73

ISBN 5-7695-2438-3

Посвящается 75-летию
Московского авиационного института
(Государственного технического университета)
и 60-летию факультета радиоэлектроники
летательных аппаратов МАИ

ПРЕДИСЛОВИЕ

Разными аспектами проблемы защиты информации занимаются юристы, экономисты, связисты, военные, программисты и, разумеется, инженеры. Именно инженерам, точнее — молодым людям, изучающим основы инженерного дела в высших технических учебных заведениях, адресована эта книга.

Труд современного инженера протекает в информационной среде, а информация является основным предметом и продуктом инженерного труда. Поэтому безопасные приемы труда в информационном пространстве также важны для инженера, как выполнение требований техники безопасности в процессе работы по преобразованию вещества и энергии. В силу целого ряда причин, о которых речь пойдет ниже, именно проблемы безопасного обращения с информацией в процессе инженерного труда и творчества приобрели особую актуальность для современного этапа развития нашей технической цивилизации.

Проблема защиты информации возникает там, где есть противоречия, конфликт интересов в информационной среде. Изучая методы и средства защиты, всегда приходится иметь в виду информационные угрозы, их вид, характер и условия проявления. В учебном пособии рассматриваются информационные конфликты и угрозы, характерные для той среды, где функционируют современные технические и организационно-технические системы. Для подобных систем характерны большое разнообразие целей, задач и способов функционирования, значительное структурное разнообразие а также широкий спектр проявлений информационных конфликтов.

Требование широты охвата проблемы защиты информации вступает в противоречие с подробным изучением конкретных методов и средств информационной защиты. Поэтому в название книги внесено уточняющее дополнение «Основы». По той же причине из рассмотрения исключены весьма важные вопросы организации и управления информационной безопасностью предприятия (фирмы), т. е. менеджмент информационной безопасности. Не рассматриваются структура системы законов и подзаконных нормативных актов, регулирующих взаимоотношения в информаци-

онной сфере, а также особенности крайних проявлений информационных конфликтов в форме информационных войн (не путать со скандалами в журналистских тусовках, которые иногда и совершенно неправомерно именуют тем же термином), т.е. вопросы проектирования и применения информационного оружия (как оборонительного, так и наступательного), и некоторые другие вопросы, без которых можно обойтись при первоначальном ознакомлении с проблемой защиты информации.

Учебное пособие написано по материалам лекционных курсов, которые вели авторы на разных факультетах Московского авиационного института (Государственного технического университета).

Глава 1
ПРОБЛЕМА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

1.1. Современное состояние, перспектива и ретроспектива

Очень велико искушение начать ретроспективу проблемы информации и информационной безопасности с истории о том, как сказалась надежность априорных данных на оптимизации стратегии поведения в ходе такой глобальной экологической катастрофы, как Великий Потоп или с того, как и к чему привела информационная незащищенность Адама и Еву. Но, избегая подобных банальностей, все-таки приходится утверждать, что если историю земной цивилизации положить на логарифмическую временну́ю шкалу (рис. 1.1), можно выявить любопытные закономерности.

С давних времен человечество обеспечивало свое существование за счет эксплуатации природных ресурсов, и в этой сфере (собирательство, охота, затем сельское хозяйство) было занято подавляющее большинство населения. На индустриальном этапе развития цивилизации определяющим в жизни человечества было промышленное производство (переработка вещества и энергии). Наконец на современном этапе постиндустриального общества определяющей формой трудовой деятельности стала переработка информации.

В экономически развитых обществах примерно 2 % населения заняты в сельском хозяйстве, 12 % — в промышленности (переработка вещества и энергии), 70 % — в информационной сфере.

Отечественная статистика не дает такой стратификации общества, традиционно разделяя население только по классовым, половым и возрастным признакам. Но есть все основания предполагать, что и для современной России занятость населения в сфере

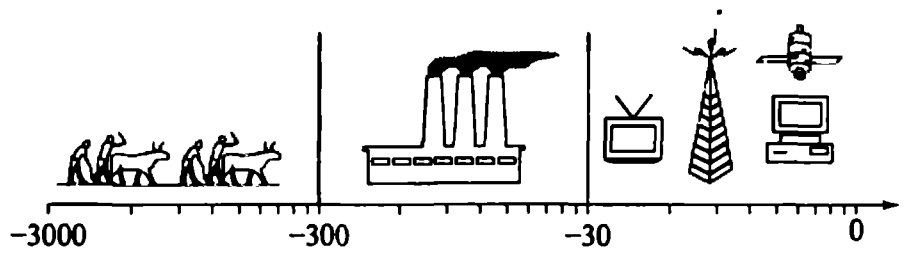


Рис. 1.1. Временна́я шкала смены доминант цивилизации на планете Земля

обслуживания национального информационного ресурса весьма значительна.

Даже исключив из рассмотрения такие высокоинформатизованные области, как управление (государственное, региональное, муниципальное и местное, в военной сфере и экономике). Не рассматривая масс-медиа, образование и науку, функционирующие исключительно в информационном пространстве, можно утверждать, что современные инженеры создают не вещи, а проектную документацию, т. е. информацию. Современные финансисты осуществляют информационное сопровождение финансовых потоков.

Если учесть все сказанное, не покажется большим преувеличением утверждение о том, что XXI в. войдет в историю как век информации и телекоммуникаций, подобно тому, как XX в. был веком электроэнергетики, XIX в. — веком пара, а XVI в. — веком великих географических открытий. Инфокоммуникации уже в самом начале XXI в. достигли высокой степени мобильности (сотовые сети связи). Они приобрели новое системное качество мультимедийности. В XXI в. человечество вступило, имея реальную возможность для создания общедоступной глобальной инфокоммуникационной инфраструктуры.

Такое положение информационной сферы современного постиндустриального информационного общества обусловливает целый ряд его специфических черт.

Во-первых, информация, в процессе получения и распространения которой занято большинство экономически активного населения, стала товаром, причем товаром массового производства и потребления. Но товаром весьма специфическим. На этот товар (информационный продукт) должны распространяться права собственности. Но если традиции и нормы, регулирующие права собственности на вещи, выработаны веками, то с информационными продуктами дело гораздо сложнее. Если некто имеет вещь и передает эту вещь другому, он эту вещь теряет, утрачивая права на нее. По меньшей мере он теряет одно или несколько звеньев триады «владеть — пользоваться — распоряжаться», составляющей основное содержание понятия собственности. Но если он передает некую сумму сведений (информацию, знания), то у него эти сведения тоже остаются. Значит, нужны какие-то иные регуляторы отношений в информационном пространстве.

Во-вторых, массовый характер получения и потребления информации требует разработки методов безопасного обращения с ней. Подобно тому, как массовое участие людей в процессе переработки вещества и энергии требовало массового образования в области безопасных методов труда. Пренебрежение требованиями безопасного обращения с информацией может привести к весьма негативным последствиям. Кто знает, не грозят ли информацион-

ному пространству техногенные катастрофы, подобные экологическим, вызванным нарушениями правил природопользования. Поэтому защита информации, которая должна не только разоблачать, но и предотвращать неправомерное, несанкционированное обращение с ней, приобретает особую актуальность.

В-третьих, в СССР существовала более или менее надежная государственная система мер защиты информации. Она иногда отставала от потребностей жизни, но, в целом, обеспечивала решение поставленных перед ней задач. Преобразования последних лет изменили отношения собственности (в том числе и собственности на информацию). И эти изменения потребовали кардинального пересмотра и значительного совершенствования мер и средств, направленных на обеспечение безопасности информационного ресурса, находящегося в распоряжении государства, отдельных предприятий и организаций, граждан.

В-четвертых, специфические требования к информационной безопасности предъявляются со стороны нынешнего уровня и темпов технического прогресса. За последние 30 лет количество физических процессов и объектов, используемых при подготовке, хранении, распределении и потреблении информации, увеличилось в несколько раз. Появление в информационной сфере каждого нового технического и технологического процесса предъявляет новые специфические требования к обеспечению информационной безопасности.

По современным воззрениям проблема информационной безопасности распадается на две, равноправные и диалектически связанные (рис. 1.2). Это проблема защиты информации (от утраты, искажения, несанкционированного доступа и использования) и защиты от информации (ложной, избыточной).

В мировое информационное пространство могут входить только развитые страны. Государства, не имеющие таких предпосылок всестороннего развития, все дальше отодвигаются на обочину социального и технического прогресса и становятся вечными маргиналами цивилизации. Подобная неравномерность прежде всего и определяет противостояние развитых стран и остального мира, стимулирует углубление противоречий, чревата нестабильностью и угрозами новых войн. Без подключения к мировому информационному пространству страну ожидает экономическое прозяба-

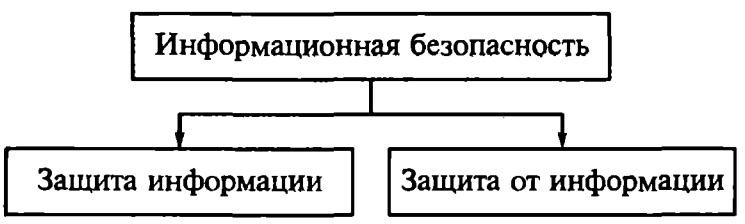


Рис. 1.2. Структура проблемы информационной безопасности

ние. Однако следует отчетливо представлять себе, что участие России в глобальных информационных процессах невозможно без комплексного решения проблем информационной безопасности, предполагающего как рациональное использование мирового информационного ресурса, так и защиту собственного национального информационного пространства от возможности деструктивного и негативного воздействия.

В настоящее время мир озабочен состоянием защиты национальных информационных ресурсов в связи с расширением доступа к ним через открытые информационные сети типа Internet. Кроме того что повсеместно увеличивается число компьютерных преступлений, реальной стала угроза информационных атак на более высоком уровне для достижения политических и экономических целей.

Для предотвращения или нейтрализации последствий таких атак необходимо:

- защищать материально-технические объекты, составляющие физическую основу информационных ресурсов;
- обеспечить нормальное и бесперебойное функционирование баз и банков данных;
- защитить информацию от несанкционированного доступа, искажения, уничтожения;
- сохранить качество информации (своевременности, точности, полноты и необходимой доступности).

Поскольку информационная сфера охватывает все области жизни, постольку информационная безопасность структурируется в

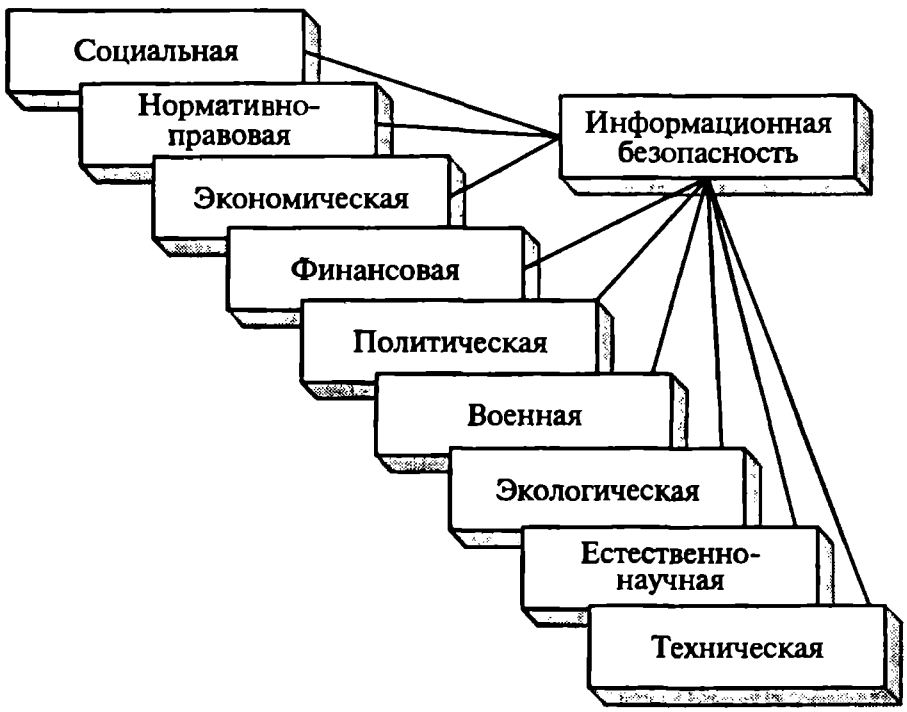


Рис. 1.3. Аспекты проблемы информационной безопасности

совершенно разных, но связанных аспектах (рис. 1.3). Совершенно определенно можно выделить социальные, нормативно-правовые, экономические, финансовые аспекты, информационную безопасность политической и военной сфер, безопасность экологической информации, естественно-научные и технические аспекты информационной безопасности.

Социальный аспект. Показателем цивилизованности общества и уровнем развития демократии является не только свобода доступа к любой информации, но и надежная защита информации ограниченного доступа. Сейчас мы довольно уверенно движемся от всеобщей секретности к информационной культуре. Именно этой идеей проникнута «Концепция информационной безопасности России», утвержденная Президентом России в 2000 г. Общественное мнение продвинуто от тотальных информационных ограничений в сторону информационной культуры гораздо меньше, чем законы РФ и «Концепция информационной безопасности России». Говоря о легком доступе к информации (особенно в Internet), мы прежде всего выделяем его негативные стороны и для борьбы с негативом соглашаемся на применение запретительных методов. Большинство режимных предприятий (кстати, разной формы собственности) не выработали мер безопасности обращения с информацией в условиях применения современных сетевых информационных технологий и пошли по простому, чисто формальному пути. Запретили Internet и e-mail в сфере своей юрисдикции.

Но это частности, а в целом приходится сознавать, что наше общество не вполне готово существовать и нормально функционировать в условиях возможных негативных и деструктивных информационных воздействий, информационно не защищено. Достаточно вспомнить, как население восприняло такое явление, как финансовые пирамиды. Люди всех социальных слоев с удовольствием и даже с азартом бросились исполнять предписания недобросовестных информационных (рекламных) воздействий на массовое сознание. Это наводит на грустные мысли о том, какие беды стране может принести применение информационного оружия массового поражения, если общество не выработает иммунитета к негативным информационным воздействиям, т.е. не научится приемам безопасного обращения с информацией.

Можно доискиваться до причин этого явления, ссылаясь на менталитет, многолетние традиции тотальной пропаганды, некритическое отношение к печатному (и произнесенному по каналам массового воздействия) слову, возможность некого заговора. Сейчас важен факт социальной неустойчивости против деструктивных информационных воздействий.

Нормативно-правовой аспект. Специалисты в области права и информатизации, особенно последние, в настоящее время все чаще

говорят об информационном законодательстве как самостоятельной отрасли права. Эта отрасль должна регулировать общественные отношения по реализации порядка защиты информации как объекта общественных отношений, прав граждан и юридических лиц на владение, использование и распоряжение информационными продуктом и услугами. Однако до сих пор нет достаточно четкого представления о том, что такое информационное законодательство, не определена сфера его правового регулирования, и придание самостоятельности такой отрасли законодательства пока вызывает возражения.

На сегодняшний день законы РФ, прямо или косвенно связанные с информационной сферой, немногочисленны и не образуют целостной системы, обеспечивающей регулирование всего спектра отношений в этой сложной и комплексной сфере.

Проблемы защиты информации регулируются Законами РФ «О государственной тайне», «Об информации, информатизации и информационной безопасности», «О правовой охране программ для электронных вычислительных машин и баз данных», «О правовой охране топологий интегральных микросхем», «Об авторском праве и смежных правах», «О федеральных органах правительственной связи и информации», «Об Архивном фонде Российской Федерации и архивах», «О средствах массовой информации» с последними дополнениями.

В Уголовном кодексе (УК) РФ, вступившем в действие 1 января 1997 г., за преступления в сфере компьютерной информации предусмотрена уголовная ответственность (гл. 28, ст. 272—274). Самая серьезная санкция — лишение свободы на срок от трех до семи лет за создание, использование и распространение вредоносных программ для ЭВМ, повлекших тяжкие последствия. Таким образом, новый УК вводит в употребление новое понятие: «вредоносные программы», под которыми понимаются программы, заведомо приводящие к несанкционированному уничтожению, блокированию, модификации либо копированию информации, а также приводящие к нарушению работы ЭВМ, системы ЭВМ или их сети. В настоящее время такие программы общепринято называть программными закладками и компьютерными вирусами.

В целом, принятие нового УК является шагом вперед в определении понятия «компьютерные преступления» и квалификации отдельных правонарушений. Но пока этот шаг очень робкий и наивный. Юристы считают, что для вхождения России в мировое информационное пространство в качестве полноправного члена нужно:

- разработать национальное законодательство в части правил обращения с информационными ресурсами, регламента прав, обязанностей и ответственности пользователей открытых мировых сетей;

• установить перечень информации, не подлежащей передаче по открытым сетям, и обеспечить систему действенных мер контроля за соблюдением установленного статуса информации;

• активно участвовать в разработке международного законодательства и нормативно-правового обеспечения функционирования мировых открытых сетей.

Создание корректной, полной системы обеспечит сохранение национальных информационных ресурсов России и ее равноправное вхождение в мировое информационное сообщество.

Базовыми для такой системы могут стать федеральные законы, регламентирующие вопросы:

формирование информационных ресурсов Российской Федерации;

перемещение информационных ресурсов и информационных технологий через таможенную границу Российской Федерации;

ответственность за неправомерное использование информационных ресурсов, составляющих национальное достояние или ущемляющих конституционные права граждан Российской Федерации.

В развитие этих законов подзаконными правовыми актами еще только предстоит определить:

порядок проведения официальной регистрации, экспертизы, сертификации и оценки информационных ресурсов, созданных технологическим путем;

права и обязанности субъектов, ответственных за сбор, обработку и предоставление государственных информационных ресурсов;

порядок регламентации доступа к государственным информационным ресурсам;

перечень информационных ресурсов, предоставляемых бесплатно или за плату, не компенсирующую расходы;

создание нормативно-правовой базы для развития системы страхования информационных рисков, направленных на гарантированное обеспечение страховой защиты имущественных интересов хозяйственных субъектов различных форм собственности в виде полного или частичного возмещения ущерба, организацию системы обязательного страхования информационных систем федеральных органов государственной власти, органов исполнительной власти субъектов Российской Федерации и организаций кредитно-финансовой сферы.

Экономический аспект. Он не менее важен, чем социальный и юридический, и в неменьшей степени касается современного специалиста, профессиональная деятельность которого неизбежно протекает в информационном пространстве. Экономические проблемы информационной безопасности можно условно разбить на два подмножества: информационные аспекты безопасности экономики и собственно экономика защиты информации.

Общепризнанно, что современная эффективная рыночная экономика немыслима вне достоверной и надежной информационной среды.

В основе теории экономического равновесия лежит постулат о том, что необходимым условием оптимизации поведения любой фирмы и вообще экономической структуры на рынке и обеспечения максимума прироста общественного продукта, является точное знание, во-первых, своих производственных возможностей, во-вторых, всех условий, которые существуют на рынке, и, в-третьих, всех норм государственного регулирования экономической деятельности.

Если таковых знаний нет, то поведение экономических агентов отклоняется от оптимального и может наносить даже вред экономике общества. Также очевидно, что чем выше информационная обеспеченность деятельности участников социально-экономических отношений, тем выше конкурентоспособность национальной экономики, социальная и политическая стабильность.

Сегодня многие предприятия не имеют возможности получить достоверную информацию о той среде, в которой они работают, тех возможностях, которыми они располагают, о конкурентах и своих конкурентных преимуществах. Во многом поэтому они не в состоянии выдерживать конкуренцию и вынуждены сдавать свои рыночные позиции.

Нынешнее состояние информационного обеспечения хозяйственных отношений неудовлетворительно по целому ряду причин.

Необходимо констатировать, что в информационном пространстве крутится огромное количество недостоверной информации, которая циркулирует сегодня даже по официальным государственным информационным каналам. Достаточно вспомнить те же финансовые пирамиды, когда при помощи средств массовой информации, принадлежащих государству, тиражировалась заведомо ложная, недостоверная информация.

Современная информационная среда характеризуется чрезвычайной неопределенностью. До недавнего времени невозможно было предсказать даже на 3—4 месяца вперед какие будут цены, какая будет ставка процента; на полгода вперед невозможно было точно предсказать, какая будет налоговая система. Столь высокая неопределенность экономической среды ведет к тому, что и экономические субъекты не могут планировать свою хозяйственную деятельность более чем на полгода, на год вперед. Значит, не могут развиваться те секторы экономики, которые требуют долгосрочного планирования.

В экономической среде просто отсутствуют целые сегменты крайне важной для нормального экономического поведения информации. Это касается, прежде всего, информации о распределении прав собственности. Крайне сложно, а иногда и просто не-

возможно получить информацию о том, кто же является собственником тех или иных предприятий, какие права и возможности у этих предприятий и их владельцев.

Серьезную проблему представляет неадекватность информационных показателей, которыми сегодня оперируют участники социально-экономических отношений. Мы очень много сегодня говорим о банкротствах предприятий как необходимом условии оздоровления общества и микроэкономической среды. Но совершенно очевидно, что в условиях информационного хаоса точных и простых критериев признания или не признания того или иного предприятия банкротом просто не существует. И в условиях, когда по формальным показателям неплатежеспособности примерно половина предприятий производственной сферы сегодня могут быть объявлены банкротами, оперирование крайне простыми и примитивными показателями прибыльности, рентабельности не могут дать достоверную оценку состоянию того или иного предприятия.

Это досадное перечисление информационной незащищенности хозяйствующих субъектов можно продолжить. Такая государственная политика в области информационного обеспечения экономической деятельности в системах государственного управления ведет к весьма негативному явлению приватизации информации. Многие информационные потоки, которые должны быть общественным благом и общедоступными, искусственно закрываются и используются в частных интересах.

Другой экономический аспект защиты информации — собственно экономика безопасной информационной деятельности — тоже никак не представлен в нашем образовательном процессе. Даже дипломники, которые решают задачи из области защиты информации или радиоэлектронной борьбы (или из области конфликтных взаимодействий в информационном пространстве), экономические разделы проектов и работ выполняют по типовым заданиям на расчет некой экономической эффективности разработки. Хотя эффективность того, что мешает работать, требует дополнительного и более корректного определения.

Экономика защиты информации требует пристального внимания. И многие аспекты этой проблемы еще не разработаны. Действительно, если рассматривать некий изолированный хозяйствующий субъект (фирму, предприятие и т.п.), можно утверждать, что отсутствие у него адекватной защиты информации (ноу-хау, интеллектуальной собственности) неизбежно приведет к экономическим потерям. Но гипертрофированные меры защиты потребуют такого расхода ресурсов (временных, трудовых, финансовых), при котором упадет эффективность основной хозяйственной деятельности. Значит, между этими крайними условиями где-то должен быть оптимум расходов на обеспечение защиты инфор-

мации. Но методы нахождения этого оптимума пока не известны. В частности, еще и потому, что совершенно не разработана конструктивная теория ценности информации. Отдельные этюды к этой теории, созданные в свое время Р.Л.Стратоновичем, А.А.Харкевичем, В.И.Сифоровым, их последователями и другими учеными, не дают практического методического аппарата измерения семантической ценности информации.

Финансовый аспект. Финансовая составляющая информационной безопасности вплотную примыкает к экономической, но не тождественна ей. Известны истории с чеченскими авизовками, которые нанесли урон финансовой системе страны, соизмеримый с бюджетами регионов. Причина этих коллизий состоит, прежде всего, в довольно тривиальном отсутствии протоколов обеспечения аутентификации информации в финансовых потоках. Сейчас эти задачи решены, разработаны и законодательно внедрены методы использования электронной подписи. Но проблемы здесь еще остаются. Известны факты компьютерных атак на финансовую информацию (хотя банки довольно тщательно скрывают эти факты, как вредящие их коммерческому имиджу). И это далеко не все проблемы, грозящие финансовой сфере, не защищенной со стороны возможной информационной агрессии. Преступные посягательства в финансово-кредитной и банковской сферах за последние годы стали разнообразнее и изощреннее.

Ущерб от различных видов преступных посягательств, связанных с нарушением информационной безопасности в автоматизированных платежных системах, может быть не меньше чем при прямом хищении денег и ценностей. Актуальность этой проблемы возрастает по мере расширения внедрения новых автоматизированных платежных систем. При охвате автоматизированной платежной системой всех регионов страны любая дестабилизация в ее функционировании может нарушить безопасность финансово-платежной системы страны и, как следствие, проявится в сбое всего хозяйственного механизма государства.

Актуальной проблемой организации банковской безопасности является практическое воплощения стратегии и тактики обеспечения информационной безопасности в сфере технологии кредитно-финансовой и банковской деятельности. В этой стержневой проблеме есть целый комплекс вопросов. Одним из важнейших направлений работ по обеспечению информационной безопасности в банковской системе является создание системы защищенных телекоммуникаций, базирующейся на системе спутниковой связи, наземной коммутируемой сети телефонной связи, выделенных каналах передачи данных.

Политический аспект. Анализ его, с точки зрения проблемы информационной безопасности, дает возможность констатировать все большее смещение центра тяжести от силовых факторов к более

скрытым и тонким, базирующимся на информационном воздействии.

Несмотря на прекращение холодной войны ведущие страны мира продолжают модернизировать свои разведывательные службы, совершенствуют техническую разведку, наращивают ее возможности. Внимание к России как объекту разведки усилилось. При этом главными приоритетами иностранных разведок являются процессы становления России как самостоятельного государства в структуре мирового сообщества, ее внутренние и внешние политические ориентиры, военная политика и пути ее практической реализации, происходящие экономические преобразования, направленность научных исследований и технических экспериментов, оценка российского рынка во всех его составляющих. Значительно расширились и облегчились условия ведения разведки на территории России. Фактически договор ОСВ-2, наложив ограничения на развитие и совершенствования средств вооружений, снял ограничения на ведение разведки. Более того, одна из статей этого договора прямо гласит, что «контроль за выполнением соглашений возлагается на национальные средства контроля» (следует читать — «средства разведки»). Россия присоединилась к международному Договору по открытому небу. На очереди Договор по открытому морю. Эти договоры имеют целью контроль за военной деятельностью, но технические средства, используемые в соответствии с этими договорами, конечно, имеют более широкие возможности.

Разведывательная деятельность иностранных государств в настоящее время отличается большим разнообразием используемых сил и средств. Многофункциональные разведывательные космические системы, наземные центры радиотехнической и радиолокационной разведки, стратегические самолеты-разведчики, морские системы и комплексы технической разведки действуют в настоящее время против России непрерывно. При этом расходы на разведывательную деятельность иностранных государств не сокращаются (например, в США они составляют ежегодно около 30 млрд долларов). В сферу интересов технических разведок попадают даже союзники. Достаточно вспомнить обеспокоенность европейских партнеров и союзников США тем, сколь активно внедряется в их политическую, экономическую и, возможно, частную жизнь пресловутая американская система «Эшелон», использующая глобальную сеть космической радио- и компьютерной разведки.

Военный аспект. Информационная безопасность в военном деле — это довольно традиционная область. Военные структуры всегда защищались от средств разведки всеми способами: пассивными и активными.

По мнению отечественных и зарубежных специалистов, боевые действия в современных (и будущих) войнах прежде всего

ведутся не для разгрома сухопутных войсковых группировок противника. Они имеют целью дезорганизацию политического, экономического и военного управления соответствующими структурами противоборствующей стороны. О том, что изменились цель и характер боевых действий, свидетельствует опыт локальных войн последнего времени (после Вьетнама). Сейчас наступает новый этап. Наметилась тенденция перехода от оружия массового уничтожения к высокоточному «информационному оружию». Это не пустые слова. Так, в США создан центр по реализации концепции «Информационная война». Новый орган будет разрабатывать положения по организации и ведению борьбы в новой сфере военного противоборства, решать задачи по подготовке специалистов в данной области, а также определять приоритеты в НИОКР и закупках предназначенных для этих целей вооружений и аппаратуры.

Информационное оружие может существенно изменить характер будущих войн. Иногда утверждается, что будущие войны могут превратиться по существу в «компьютерные войны» с массовым применением компьютеризированных роботов, роботизированного оружия и военной техники. Предполагается, что основу боевой экипировки солдата в будущем образует боевой компьютер. Кроме анализа окружающей обстановки компьютер за счет соответствующего программного обеспечения и сенсоров будет осуществлять медицинский контроль за состоянием солдата и выдавать лечебные рекомендации.

В свою очередь, элементом воздействия на силы и средства противника в будущей войне может стать компьютерное оружие. Оно может быть реализовано, в частности, в виде деструктивных программ, которые могут изменять или уничтожать программы компьютеров, управляющих оружием, военной техникой и войсками.

Опыт военных действий последних лет (на Ближнем Востоке, в Югославии, Ираке) показал, что резко возросшие технические возможности средств разведки сделали неэффективными многие традиционные методы и средства защиты информации. Например, данные космических средств разведки оперативно использовались непосредственно на поле боя, для управления высокоточным оружием, даже для борьбы с иракскими оперативно-тактическими ракетами СКАД. Это значит, что такие традиционные методы скрытия информации о дислокации ракетных комплексов, как пространственное маневрирование в позиционном районе, уже неэффективны.

Возросшие оперативные возможности технических разведок и использование их данных позволило отнести радиоэлектронную борьбу уже не к средствам боевого обеспечения, а к этапу боевых действий. Соответственно возросла роль защиты информации.

Сегодня одним из наиболее существенных объектов безопасности в оборонной сфере являются информационные ресурсы и

информационная структура оборонного потенциала страны (вооруженных сил и военно-промышленного комплекса). Важно, что все современные средства вооружения, военной техники, системы управления войсками и оружием являются системами критических приложений с высоким уровнем компьютеризации. Эти системы могут оказаться весьма уязвимыми с точки зрения воздействия информационного оружия как в военное, так и в мирное время. Последнее может привести к тому, что к угрожаемому периоду оружие сдерживания страны окажется полностью или частично заблокированным за счет скрытого внедрения в программное обеспечение систем управления им программных закладок. О реальности такой ситуации свидетельствует опыт локальных войн последних лет.

Экологический аспект. Проблема экологической безопасности является сегодня одной из важнейших в глобальном масштабе. Она связана с защитой интересов личности, общества и государства от потенциальных и реальных угроз, создаваемых последствиями антропогенного воздействия на среду, а также от природных стихийных бедствий и катастроф.

Экологическая проблема является весьма сложной, многоплановой, комплексной. Она неразрывно связана с экономикой, техникой, правом, военным делом и другими сферами общественной деятельности. Но существенно важны и информационные аспекты проблемы экологической безопасности. Три причины определяют наличие корреляции между экологической и информационной безопасностью:

ощущается недостаточная информированность широких слоев населения об угрозах экологической безопасности, источниках этой угрозы, последствиях экологических бедствий и катастроф т.д. Наиболее характерным примером этого является Чернобыльская катастрофа;

решение большинства экологических проблем и задач связано со сбором и обработкой информации о состоянии окружающей среды (с экологическим мониторингом), моделированием и изучением моделей масштабных глобальных процессов природных явлений. Надежность и безопасность информации в этой сфере экологической деятельности трудно переоценить;

целый ряд систем управления (транспортом, связью, атомной энергетикой, опасными производствами) относится к «критическим». Очевидно, что недооценка вопросов информационной безопасности этих систем может привести к непредсказуемым экологическим последствиям, огромным материальным потерям и человеческим жертвам.

Естественно-научный аспект. Здесь проблемы информационной безопасности легче всего иллюстрировать на примере того, как информационная сфера впитывает и использует новейшие дости-

жения прикладных и фундаментальных научных дисциплин. Так, для реализации информационной агрессии, несанкционированного доступа к охраняемым сведениям и данным могут использоваться все без изъятия физические поля: во всех полях могут существовать процессы переноса вещества и энергии, используемые для передачи и извлечения информации. Не составляют исключения и такие экзотические для использования в приложениях к информационной сфере физические поля, как гравитационное, сейсмическое. Естественно, что использование всех известных и мыслимых физических полей в информационном конфликте предполагает реализацию диалектического баланса мер и контрмер. Необходимо защищать информацию, которая может переноситься сигналами во всех физических полях.

Немаловажно использование достижений информатики и математики в интересах обеспечения информационной безопасности.

Технический аспект. Технический аспект защиты информации тоже приходится рассматривать по-разному. Во-первых, это защита информации, циркулирующей в технических системах, точнее, в организационно-технических, поскольку именно технические средства информационного обмена составляют основное по сложности, стоимости и, возможно, по уязвимости наполнение большинства организационных структур. Во-вторых, это защита информации, основанная на использовании специальных технических средств.

Область технической защиты информации сейчас наиболее продвинута. Уже можно говорить о заложенных основах теории технической защиты информации, т.е. о том, что данная предметная область в своем развитии доросла до некоторых теоретических обобщений, понимания предельных (потенциально достижимых) уровней информационной безопасности и формулировок решаемых задач оптимизации стратегии обеспечения безопасности информации.

Работу всех технических систем сопровождает появление технических каналов утечки информации, т.е. каналов несанкционированного доступа (НСД) к информации (утечка — это очень специфический термин, пришедший из предметной области организационных средств и методов защиты информации, но он все прочнее укореняется и в области информационной безопасности). При этом технические каналы утечки информации могут порождаться вовсе и не информационными системами. Например, спектр излучения факела ракетного двигателя (совсем не информационная система) способен сообщить информацию о том, какой это двигатель (ЖРД или ТРД), о компонентах ракетного топлива, степени отработки, жизненном цикле изделия и совместно с другими разведывательными признаками технической политике в области развития вооружений.

В одной книге невозможно подробно рассмотреть все перечисленные аспекты проблемы информационной безопасности. Поэтому приведенный перечень понадобился лишь для того, чтобы обрисовать круг обсуждаемых далее технических задач защиты информации среди ее комплексных проблем.

1.2. Информационные системы, средства, каналы, сети и среды

Процессы, сопровождающие существование и развитие современного общества, принято объединять под общим названием «информатизация». Информатизация предполагает широкое использование информационных систем, которые обеспечивают доступ к источникам информации (в нетехнических приложениях эти источники часто называют информационными ресурсами), накопление и хранение информации (образование новых информационных ресурсов).

Понятие систем вообще и информационных систем в частности неоднозначно. Разные авторы в разных контекстах могут обозначать этим термином отличающиеся понятия. Изучению систем и системному подходу к исследованию окружающей нас природы, процессам, происходящим в ней, обществе и мышлении человека, посвящено множество работ. Отдельные разделы знаний посвящены исследованию технических и организационных систем. Понятие система применяют в тех случаях, когда пытаются охарактеризовать исследуемый объект как нечто целое, сложное, единое в своем многообразии.

Системность — объяснительный принцип научного познания, требующий исследовать явления в их зависимости от внутренне связанного целого, которое они образуют, приобретая благодаря этому присущие целому новые свойства.

М. Пешель при исследовании принципов построения моделей систем использует философский подход, основанный на единстве и противоположности общего и частного. В одном из тезисов, дающих определение системы, он писал, что «целое больше суммы отдельных частей, однако оно проявляется через отдельные элементы. Для восприятия целого (системы) его необходимо разложить на отдельные элементы; для углубленного восприятия целого необходимо снова собрать отдельные элементы с учетом связей между ними.

За видимой простотой афористичности утверждения о том, что целое больше своих частей, скрыт широкий круг вопросов, как философских, так и конкретно-научных. Ответы на них побуждают выяснить, по каким критериям и на каких основах из множества (и не всегда строго определенного) явлений обособляется

некая категория объектов, приобретающих значение и характер системных.

Внутреннее строение этих объектов описывается в таких понятиях, как элемент, связь, структура, функция, организация, управление, саморегуляция, стабильность, развитие, открытость, активность, среда и др.

Существует несколько десятков определений понятия «система». Так, Л. Берталанфи определял систему как «комплекс взаимодействующих элементов» или как «совокупность взаимодействующих элементов, находящихся в определенных отношениях друг с другом и со средой».

В некоторых определениях осуществляется привязка элементов системы и отношений между ними к целевой функции и временно́му интервалу. В.Н. Сагатовский определил систему как «конечное множество функциональных элементов и отношений между ними, выделенное из среды в соответствии с определенной целью в рамках определенного временно́го интервала».

Американский физиолог У. Кеннон считал синонимом системности принцип гомеостаза как динамического постоянства состава и свойств системы, ее стремление к сохранению стабильного состояния вопреки действию факторов, которые его нарушают. Содержательный смысл этого принципа состоит в том, что, руководствуясь им, исследователь в любом компоненте и отправлении системы усматривает одно из приспособлений, решающих главную задачу — удержание системы в равновесии.

К числу наиболее точных и формальных определений системы можно отнести следующее. Система **S** — это объект, существующий во времени, подвергающийся внутренним и внешним воздействиям (возмущениям), реагирующий на них изменениями своих состояний и обладающий способностью проявить в том или ином виде эти реакции. Таким образом, система **S** определена, если заданы:

множество $\{t\}$ моментов времени t, множество $\{\nu\}$ допустимых воздействий ν, множество$\{g\}$ возможных состояний g, множество $\{r\}$ возможных реакций r;

переходная функция, представленная теми состояниями $g \in \{g\}$, в которых оказывается система **S** в момент $t \in \{t\}$, если в начальный момент времени $t_0 \in \{t\}$ она была в состоянии $g_0 \in \{g\}$ и на нее подействовало возмущение $\nu \in \{\nu\}$;

отношение, связывающее в каждый момент $t \in \{t\}$ реакции $r \in \{r\}$ с состояниями $g \in \{g\}$.

Для нас из этого анализа дефиниций важно, речь идет о некоторой совокупности элементов, противопоставляемой другим совокупностям, именуемым средой функционирования системы или внешней средой. При этом в дальнейшем считается, что совокупность взаимосвязанных элементов, которая для краткости имену-

ется системой, обладает некоторыми собирательными признаками, а именно.

1. Система имеет искусственную, антропогенную природу — она создается людьми. Это сужение понятия позволяет исключить из рассмотрения систему мироздания, системы взглядов, верований и суеверий, социальные и общественно-политические системы и другие очень интересные системные конфигурации.

2. Система обладает целостностью — все ее части работают для достижения единой цели функционирования. Формулировка цели функционирования, определение количественных показателей достижения этой цели (целевая функция) и измеримых характеристик качества функционирования (критериев эффективности) не могут быть заданы изнутри системы. Все эти показатели и характеристики определяются внешней по отношению к системе средой.

3. Совокупность элементов, составляющих систему, обладает разнообразием выполняемых функций, различной сложностью (и стоимостью), т. е. система всегда является большой.

4. Система всегда является сложной в том смысле, что все ее элементы влияют друг на друга и изменение состояния одного из них вызывает изменения состояний других. При этом количественные характеристики взаимного влияния элементов не обязательно обладают свойством линейности. Эти зависимости могут быть и нелинейными, в частности — немонотонными.

5. Практически все системы являются автоматизированными: часть их функций выполняется человеком, а часть (автоматическими) техническими устройствами.

6. Все или почти все воздействия на систему случайны. Поэтому невозможно совершенно точно предсказать конкретное состояние системы, а описание работы системы должно быть вероятностным.

7. Большинство систем, в особенности самые большие и сложные системы, функционируют в конкурентной среде. Поэтому их работа сопровождается конфликтными ситуациями.

Все сказанное о технических системах не изменится, если к названию «система» присоединить эпитет «информационная». В дальнейшем наряду с названиями «система» и «информационная система» придется употреблять и другие термины из словаря системотехники, такие как «подсистема» — часть системы, содержащая все те же признаки, что и система, но являющаяся частью другой системы более высокого иерархического уровня (комплекса). Например, комплекс управления космическим аппаратом состоит из наземного автоматизированного комплекса управления и бортового комплекса, а каждый из них содержит информационные системы для траекторных измерений, передачи командной и телеметрической информации и т. п.

Непременной частью информационных систем являются технические средства — устройства (аппаратура) для добывания, извлечения, передачи, приема, переработки, хранения информации. Рассматривая работу информационных систем, обычно подчеркивают наличие специфических для них подсистем — информационных каналов. Каналы соединяют источники информации с получателями и объединяют такие подсистемы и средства, по которым передаются сигналы, несущие сообщения. Основные каналы соединяют абонентов информационных систем. Работу практически любой информационной системы сопровождает появление нежелательных или даже недопустимых информационных каналов. Это каналы несанкционированного, незаконного доступа к сообщениям, циркулирующим в информационных системах. Иначе такие каналы называются каналами утечки информации. Физические процессы, участвующие в образовании каналов утечки информации, дают им название, например электромагнитные, акустические (гидроакустические, виброакустические), оптические и другие каналы утечки информации.

Если физическая среда функционирования канала утечки информации не важна для описания и исследования его основных характеристик, названием подчеркивают другие основные свойства. Так, говорят о логических каналах несанкционированного доступа к информации, хранимой и перерабатываемой вычислительными средствами, об агентурных каналах доступа к документированной информации.

Следует отметить один очень важный класс информационных систем, специфической чертой и свойством которого является распределенность в пространстве. Это информационные сети. Если сети объединяют передачи информации, говорят о сетях связи и(или) передачи данных. Вычислительные среды, способные в процессе работы обмениваться не только информацией, но и собственными ресурсами, поддерживая тем самым параллельное выполнение общих программ работы, образуют вычислительные сети. Известно довольно много признаков, по которым классифицируют информационные сети. По признаку размещения элементов и средств выделяют наземные, космические, стационарные и мобильные сети. В зависимости от зоны обслуживания говорят о международных, междугородных, региональных, местных и внутренних (локальных) сетях. Классификационным признаком может быть принадлежность информационных сетей: общего пользования, ведомственных, корпоративных учрежденческих и т. п. В зависимости от физической среды функционирования информационных каналов, поддерживающих сеть, говорят о проводных, радиоспутниковых, кабельных, оптоволоконных сетях. Разумеется, всякая классификация условна и подчеркивает не столько различие, сколько общность сетей, наделенных разными признаками.

Одна и та же сеть, например, может быть и региональной, и ведомственной, и оптоволоконной, и ориентированной на передачу данных. Но каждый из признаков классификации определяет специфичность угрозы информационной безопасности сети.

Контрольные вопросы

1. В чем вы видите основные технические проблемы защиты информации?
2. Естественно-научные исследования проблемы защиты информации предусматривают анализ физических причин и материальных предпосылок нарушения информационной безопасности. Перечислите эти причины и предпосылки.
3. Приведите примеры каналов утечки информации, возникающих при работе технических систем.
4. Сформулируйте определения и приведите описание понятий, именуемых терминами «система» и «информационная система».
5. Как вы понимаете термины «канал передачи информации» и «канал утечки информации»? Согласны ли вы считать информационные каналы системами? Почему?

Глава 2
ИНФОРМАЦИЯ

2.1. Количество информации

Разные области и отрасли требуют разных способов измерения такого сложного феномена, как информация. Информационная емкость библиотек измеряется количеством томов, полиграфисты измеряют информацию печатными листами. На радио и телевидении информацию измеряют минутами эфирного времени. Наиболее конструктивна и традиционна мера, используемая в теории и технике передачи информации, — это количество информации по Шеннону. Эта мера связана с условием выбора конкретного сообщения s из ансамбля $s \in S$:

$$I = \log P_{ps}(s) - \log P_{pr}(s), \qquad (2.1)$$

где $P_{ps}(s)$ — апостериорная вероятность того, что принятый сигнал содержит именно сообщение s; $P_{pr}(s)$ — априорная вероятность сообщения s. Если сообщение при передаче не искажено, $P_{ps}(s) \equiv 1$ и $I = -\log P_{pr}(s)$.

Если логарифм берется по основанию 2, информация и энтропия измеряются в двоичных единицах, или битах (BInary digIT). Один бит — это количество информации, содержащееся в неискаженном при передаче сообщении, которое априори может принимать два равновероятных значения $P_{pr}(s_0) = P_{pr}(s_1) = \frac{1}{2}$. Действительно, из (2.1) следует: $I = -\log_2\left(\frac{1}{2}\right)$. Кратные одному биту величины количества информации — 1 байт = 2^3 = 8 бит, 1 Кбайт = = 2^{10} байт = 2^{13} бит и т.д.

Источник информации может формировать разные сообщения, у которых могут отличаться априорные вероятности. Поэтому кроме величины количества информации I (2.1) в сообщении рассматривается еще и среднее количество информации на одно сообщение, которое может быть сформировано источником:

$$H = -\sum_{k=1}^{K} P_{pr}(s_k) \log P_{pr}(s_k), \qquad (2.2)$$

где K — объем полного ансамбля сообщений s_k, $k \in 1:K$, которые может формировать источник.

Как видно из формулы (2.2), энтропия — это математическое ожидание величины $\log P_{pr}(s)$. Чем больше энтропия источника, тем больше, в среднем, степень априорной неопределенности формируемого им сообщения.

После приема сообщения неопределенность уменьшается (во всяком случае, не увеличивается). Поэтому количество информации (2.1) можно трактовать как меру уменьшения неопределенности.

Энтропия, как видно из (2.2), неотрицательна и равна нулю только для вырожденного ансамбля, содержащего только одно сообщение, что $P_{ps}(s_i) = 1$, $P_{ps}(s_j) = 0$; $i, j \in 1:K$.

Энтропия аддитивна: для совокупности источников информации или для ансамбля из нескольких сигналов энтропия равна сумме энтропий каждого:

$$H_N = \sum_{n=1}^{N} H_n = -\sum_{n=1}^{N}\sum_{k=1}^{K} P_{pr}(s_{nk}) \log P_{pr}(s_{nk}). \qquad (2.3)$$

Из (2.2) также видно, что энтропия имеет ту же размерность, что и количество информации и эта размерность зависит от выбора основания логарифма в (2.1) и (2.2). Так, например, для такого источника информации, как запоминающее устройство (ЗУ), объем (энтропия) в 1 Мбайт означает, что ЗУ может формировать сообщения общим суммарным объемом до $2^{20} \cdot 2^3$ бит, или, что то же самое, может запоминать и сохранять информационные массивы общим объемом до $2^{20} \cdot 2^3$ бит. Разумеется, это верхнее, потенциальное значение. Если ЗУ не заполнено (или заполнено данными, не содержащими информации) энтропия такого источника будет меньше.

Другая важная для оценки информативности сообщений величина — взаимная информация. Это информация, которую содержит один ансамбль данных относительно другого. Например, ансамбль принятых сообщений относительно ансамбля сообщений переданных или ансамбль шифровок относительно породившего его ансамбля открытых сообщений. Пусть имеется два ансамбля **A** и **B** дискретных сообщений $a_k \in$ **A** и $b_l \in$ **B**. Совместная вероятность сообщений a_k и b_l — это $P(a_k, b_l)$. Совместная энтропия ансамблей **A** и **B** будет, по определению, математическим ожиданием логарифма совместного распределения:

$$H(\mathbf{A},\mathbf{B}) = -M\{\log(a_k, b_l)\}. \qquad (2.4)$$

Для условного распределения аналогичным образом можно определить условную энтропию:

$$H(\mathbf{A} \mid \mathbf{B}) = -M\{\log(a_k \mid b_l)\} = \sum_k \sum_l P(a_k \mid b_l) \log(a_k \mid b_l). \quad (2.5)$$

Используя теорему Байеса

$$P(a,b) = P(a)P(b \mid a) = P(b)P(a \mid b), \quad (2.6)$$

для (2.5) можно установить, что

$$H(\mathbf{A},\mathbf{B}) = H(\mathbf{A}) + H(\mathbf{B} \mid \mathbf{A}) = H(\mathbf{B}) + H(\mathbf{A} \mid \mathbf{B}). \quad (2.7)$$

Нетрудно также установить, что для условной энтропии справедливо двойное неравенство

$$0 \le H(\mathbf{A} \mid \mathbf{B}) \le H(\mathbf{A}). \quad (2.8)$$

Равенство $P(\mathbf{A} \mid \mathbf{B}) = 0$ выполняется тогда, когда ансамбль **B** содержит всю информацию об ансамбле **A**. В другом случае $H(\mathbf{A} \mid \mathbf{B}) = H(\mathbf{A})$ означает, что ансамбли **A** и **B** содержат независимые сообщения, т.е. **B** не содержит информации об **A**.

Поскольку, как видно из (2.8), условная энтропия $H(\mathbf{A} \mid \mathbf{B})$ никогда не бывает больше $H(\mathbf{A})$, можно утверждать, что знание **B** уменьшает (в среднем) априорную неопределенность относительно **A**. Поэтому разность

$$H(\mathbf{A}) - H(\mathbf{A} \mid \mathbf{B}) = I(\mathbf{A},\mathbf{B}) \quad (2.9)$$

называется количеством взаимной информации в ансамбле **B** относительно **A**.

Используя определения энтропии (2.4), (2.5) и (2.9), можно установить, что взаимная информация может быть определена (измерена) как

$$I(\mathbf{A},\mathbf{B}) = -\mathrm{M}\{\log[P(a_k)]\} + \mathrm{M}\{\log[P(a_k \mid b_l)]\} = M\left\{\log \frac{P(a_k \mid b_l)}{P(a_k)}\right\}. \quad (2.10)$$

Если $P(a_k \mid b_l)$ условная вероятность того, что был передан сигнал a_k при условии, что принят сигнал b_l, а $P(a_k)$ априорная вероятность события $a_k \in \mathbf{A}$, то $I(\mathbf{A},\mathbf{B})$ показывает, сколько в среднем информации об **A** воспроизводится на выходе информационной системы, воспроизводящей сообщение из ансамбля **B**. Иначе говоря, $I(\mathbf{A},\mathbf{B})$ показывает, сколько в среднем получается бит информации о сообщении из ансамбля **A** при наблюдении реализации сообщения из ансамбля **B**. При этом совершенно неважно, формируется ли это сообщение абонентом системы (сети) пере-

дачи информации, радионавигационным маяком или возникает в пространстве вследствие взаимодействия специальным образом организованного электромагнитного поля с радиолокационной целью.

Соотношения (2.1)...(2.10) позволяют сформулировать следующие свойства меры количества информации и энтропии.

• Информация — неотрицательная величина $I(\mathbf{A},\mathbf{B}) \geq 0$, причем равенство соответствует независимым ансамблям сообщений, т.е. случаю, когда информация в **B** об **A** полностью разрушена.

• $I(\mathbf{A},\mathbf{B}) \leq H(\mathbf{A})$ и $I(\mathbf{A},\mathbf{B}) \leq H(\mathbf{B})$, причем равенство соответствует однозначной связи (тождественности) элементов массивов **A** и **B**.

• Поскольку очевидно, что $H(\mathbf{A}|\mathbf{A}) = 0$, из (2.9) следует, что $I(\mathbf{A},\mathbf{A}) = H(\mathbf{A})$, т. е. энтропия источника информации есть ни что иное, как его собственная информация (информация о самом себе).

Перечисленные свойства информации и энтропии подсказывают вывод, важный для объяснения деструктивных свойств помех, сопровождающих работу информационных систем: если **A** — ансамбль исходных сообщений, а **B** — ансамбль сообщений, воспроизводимых информационной системой, то $I(\mathbf{A},\mathbf{B}) = H(\mathbf{A})$ в том и только в том случае, когда преобразование $\mathbf{A} \rightarrow \mathbf{B}$ однозначно и обратимо. Иначе $I(\mathbf{A},\mathbf{B}) < H(\mathbf{A})$ и разность

$$H(\mathbf{A}) - I(\mathbf{A},\mathbf{B}) = H(\mathbf{A}|\mathbf{B}) \qquad (2.11)$$

есть потеря информации об **A** при воспроизведении **B**.

Обозначив величиной T среднее время передачи одного сообщения, можно оценить скорость передачи $R(\mathbf{A},\mathbf{B})$ информации от **A** к **B**:

$$R(\mathbf{A},\mathbf{B}) = \frac{I(\mathbf{A},\mathbf{B})}{T}, \qquad (2.12)$$

где $R(\mathbf{A},\mathbf{B})$ — это тоже удельное количество информации, но не на одно сообщение, а на единицу времени.

С такой скоростью сообщения передаются по информационному каналу или воспроизводится информационной системой, будь то канал передачи или утечки информации. Эта скорость зависит как от свойств источника информации, так и от свойств самого канала. Максимально достижимая скорость при заданном качестве передачи информации $c = \max R(\mathbf{A},\mathbf{B})$ называется пропускной способностью канала передачи и является важнейшей характеристикой информационной системы. Для оценки пропускной способности можно привести следующие рассуждения.

Пусть по дискретному каналу передаются сигналы, содержащие по n символов каждый, и пусть эти символы выбираются из алфавита, который содержит m символов (имеет объем m). Каждый такой сигнал может принимать N разных значений и, соот-

ветственно, переносить информацию об N разных сообщениях. Естественно,

$$\begin{aligned} &\text{при} \quad n=1 \quad N=m; \\ &\text{при} \quad n=2 \quad N=m^2; \\ &\ldots \\ &\text{при} \quad n \quad N=m^n. \end{aligned} \tag{2.13}$$

Если все сигналы в информационной системе независимы (канал без памяти), равновероятны $\left(P_{pr}=\dfrac{1}{N}\right)$ и принимаются без искажений ($P_{ps}=1$), а длительность передачи каждого сигнала $T = n\tau_c$ равна сумме длительности элементов (символов), из которых он составлен, то скорость передачи информации оказывается равной:

$$c = R = \frac{I}{T} = \frac{1}{n\tau_c}\log\frac{P_{ps}}{P_{pr}} = \frac{\log_2 m^n}{n\tau_c} = \frac{1}{\tau_c}\log_2 m. \tag{2.14}$$

Для передачи символа (элементарного сигнала) длительностью τ_c канал передачи данных должен иметь ширину полосы пропускания $w \approx \dfrac{1}{\tau_c}$. Именно такую полосу частот занимает сигнал длительностью τ_c. Поэтому для дискретного канала без памяти соотношение (2.14) позволяет утверждать, что

$$c = w\log_2 m. \tag{2.15}$$

Если условие неискаженной передачи не выполняется и в канале происходят ошибки с вероятностью p, то апостериорная вероятность правильного воспроизведения каждого символа сообщения будет равна уже не единице, а

$$p_{ps} = p\left(b_l \middle| a_k\right) = \begin{cases} 1-p & \text{при } a_k = b_l, \\ \dfrac{p}{m-1} & \text{при } a_k \neq b_l; \end{cases} \tag{2.16}$$

и входящая в (2.14) величина количества информации должна определяться согласно (2.11) с учетом того, что $H(\mathbf{A})$ — энтропия источника сообщений — не зависит от свойств канала, а условная энтропия

$$H\left(\mathbf{B}\middle|\mathbf{A}\right) = -M\left\{\log_2\left(b_k\middle|a_l\right)\right\} = -p\log_2\frac{p}{m-1} + (1-p)\log_2(1-p). \tag{2.17}$$

Используя (2.9), (2.15) и (2.16), можно получить:

$$c = w\left[\log_2 m + p\log_2 \frac{p}{m-1} + (1-p)\log_2(1-p)\right]. \qquad (2.18)$$

Важный частный случай двоичного симметричного канала, когда $m = 2$, а искажения противоположных по значению символов при передаче равновероятны, иллюстрируется рис. 2.1.

Для этого случая пропускная способность выражается соотношением

$$c = w\left[1 + p\log_2 p + (1-p)\log_2(1-p)\right]. \qquad (2.19)$$

Зависимость удельной пропускной способности, нормированной к полосе пропускания информационной системы в двоичном симметричном канале, от вероятности искажения символа представлена на рис. 2.2.

Как видно из (2.17) и графика (см. рис. 2.2), при $p = 0{,}5$ пропускная способность $c = 0$. В содержательных терминах это означает, что при такой вероятности ошибки сообщение можно и не передавать.

Получателю сообщения можно просто выбирать его с равной вероятностью из двух возможных. Иначе говоря, $c = 0$ соответствует обрыву информационного канала или блокировке канала утечки информации.

Скорость передачи информации по каналу максимальна $c = \max$ не только при $p = 0$ (что вполне естественно), но и при $p = 1$. При $p = 1$ все символы при передаче меняются на противоположные, но такой инвертированный, зеркально преобразованный сигнал содержит всю ту же информацию, что и исходный, неискаженный.

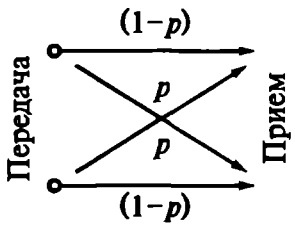


Рис. 2.1. Граф изменений символов в двоичном симметричном канале

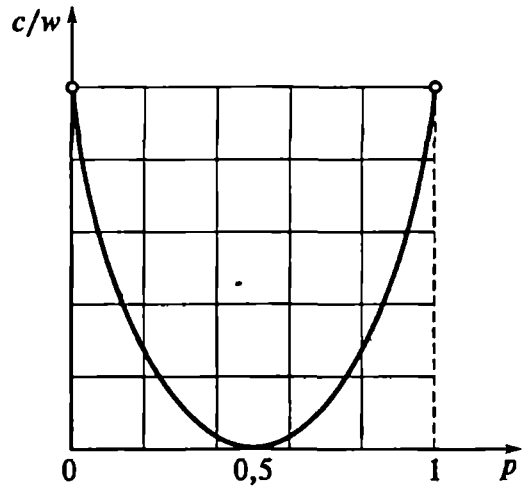


Рис. 2.2. Нормированная пропускная способность двоичного симметричного информационного канала

В этих крайних экстремальных случаях $p = 0$ и $p = 1$ максимум пропускной способности (измеренной в бодах, 1 бод = 1 бит/с) не превосходит численного значения ширины полосы пропускания информационной системы (Гц).

В отличие от дискретных сигналов, состоящих не из конечного набора символов, выбранных из ансамбля (алфавита) конечного размера, непрерывные сигналы $s(t)$ могут принимать значения из континуального множества $s \in [s_{min}; s_{max}]$, где s_{max}, ..., s_{min} — динамический диапазон сигнала, и существовать во всех точках временного континуума $t \in [0; T]$. Формальное применение использованных выше рассуждений для оценки пропускной способности канала передачи непрерывных сообщений может привести к парадоксальным результатам. Непрерывный сигнал, как показано на рис. 2.3, можно уподобить последовательности сколь угодно коротких импульсов (дискретных сигналов), сколь угодно мало отличающихся по амплитуде.

Даже конечный интервал времени существования непрерывного сигнала может содержать бесконечно много таких импульсов. Может показаться, что такая пачка импульсов содержит бесконечно большое количество информации. Кроме того, каждый из этих импульсов имеет амплитуду, равную соответствующему значению сигнала $s(t)$, т.е. принимает значения на сегменте $[s_{min}; s_{max}]$. Если считать, что все множество импульсов различных амплитуд составляет алфавит, то придется признать объем этого алфавита бесконечно большим, а, значит, каждое сообщение, передаваемое символами такого алфавита, содержит бесконечно большой объем информации, т.е. формальное сведение непрерывного сигнала к пачке узких примыкающих друг к другу импульсов приводит к выводу о дурной бесконечности ($\infty \times \infty$) в оценке пропускной способности информационного канала.

Противоречие между ограниченностью динамического диапазона сигнала конечной длительности и кажущейся бесконечностью его информационной емкости разрешается довольно просто. Если канал передачи (или утечки) информации имеет конечную

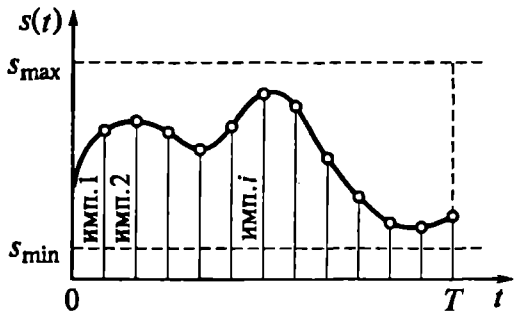


Рис. 2.3. Представление непрерывного сигнала

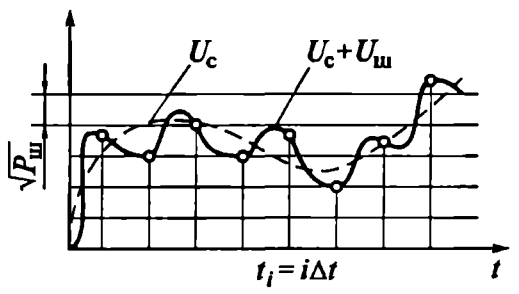


Рис. 2.4. Конечное число различимых градаций сигнала, наблюдаемого на фоне шума

полосу пропускания w (а это условие совершенно естественно и выполняется для всех физически реализуемых систем), то сообщение на приемной стороне канала может быть абсолютно точно восстановлено по своим дискретным значениям, следующим с периодом повторения $\Delta t = \frac{1}{2w}$. Это необходимое и достаточное положение обосновывается фундаментальной для теории и техники информационных систем теоремой В.А. Котельникова. Фактически утверждение о возможности точного восстановления сообщения означает, что выборка дискретных по времени значений сигнала, имеющая объем

$$N = \frac{T}{\Delta t} + 1 = 2Tw + 1 \approx 2Tw, \qquad (2.20)$$

содержит ту же информацию и в том же объеме, что и исходное непрерывное сообщение $s(t)$.

Реально при передаче, приеме и обработке сигнала на него всегда накладываются помехи (шумы). Поэтому, передавая сигнал, бесполезно требовать представления его выборочных значений $s[t = n\Delta t]$ с точностью, превосходящей уровень шумовых флуктуаций: чем больше средний уровень мощности шума $P_{ш}$, сопровождающего прием сигнала мощностью $P_{с}$, тем меньше разных значений сигнала можно различить на приеме (рис. 2.4).

Иначе говоря, в условиях действия помех реально можно различать только m градаций уровня сигнала, причем

$$m = \frac{U_{с} + U_{ш}}{U_{ш}} = \sqrt{\frac{P_{с} + P_{ш}}{P_{ш}}}. \qquad (2.21)$$

Таким образом, непрерывный сигнал длительностью T и шириной спектра w будет содержать такое же количество информации, что и выборка из $N = 2wT + 1$ импульсов, способных принимать любое из $m = \sqrt{1 + \frac{P_{с}}{P_{ш}}} = \sqrt{1 + q}$ значений. Поэтому количество информации, переносимой сигналом $s(t)$, $t \in [0; T]$ по каналу с полосой пропускания w и уровнем помех $\frac{P_{с}}{P_{ш}} = q$, равно

$$I_s = N \log_2 m = 2wT \log_2 \sqrt{1 + \frac{P_{с}}{P_{ш}}} = wT \log_2 (1 + q), \qquad (2.22)$$

а скорость передачи этой информации

$$v = \frac{I_s}{T} = w \log_2 \left(1 + \frac{P_c}{P_ш} \right) = w \log_2 (1 + q). \qquad (2.23)$$

Соотношение (2.23) именуется формулой Шеннона. Оно универсально для любых информационных систем вне зависимости от того, какие физические процессы и поля ими используются для образования каналов передачи либо перехвата информации.

Из (2.23) следует, что для увеличения пропускной способности информационного канала нужно либо увеличивать мощность сигнала (соотношение $q = P_c/P_ш$), либо расширять полосу w. Но эти способы неравноценны. С расширением полосы w пропускная способность растет линейно, а при изменениях q она меняется по логарифмическому закону, т.е. медленнее. Поэтому компенсировать сужение полосы спектра сигнала увеличением его мощности не всегда выгодно.

Если работе информационной системы противодействует нормальный стационарный шум со спектральной плотностью N_0, то $P_ш = wN_0$ и (2.23) преобразуются к виду

$$c = w \log_2 \left(1 + \frac{P_c}{wN_0} \right). \qquad (2.24)$$

Это значит, что при расширении полосы w пропускная способность сначала при малом w быстро растет ($c \sim w$), а затем асимптотически сходится к величине

$$c_\infty = \lim_{w \to \infty} w \log_2 \left(1 + \frac{P_c}{N_0 w} \right) = \lim_{w \to \infty} \log_2 \left(1 + \frac{P_c}{N_0 w} \right)^w = 1{,}443 \frac{P_c}{N_0}. \qquad (2.25)$$

Зависимость (2.24) представлена графиком (рис. 2.5).

Из (2.24), в частности, следует, что для передачи заданного количества информации $I = cT$ с требуемым качеством, отношение энергии сигнала к спектральной плотности аддитивного нормального шума $\frac{P_c T}{N_0}$ должно быть не меньше некоторого порого-

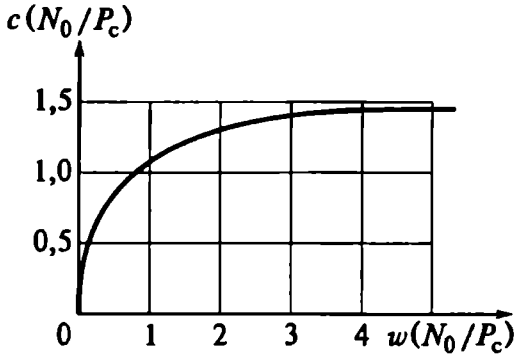


Рис. 2.5. Диаграмма обмена нормированной пропускной способности на полосу канала передачи информации

вого значения. Так, для передачи 1 бит информации нужно, очевидно, чтобы

$$\beta_E = \frac{P_c T}{N_0} > \frac{1}{\log_2 \mathbf{e}} = \ln 2 \approx 0{,}693. \qquad (2.26)$$

Подобно удельным затратам энергетического потенциала $\frac{P_c T}{N_0}$ на 1 бит можно определить удельные затраты полосы пропускания канала на передачу 1 бита информации

$$\beta_w = \frac{w}{R} < \frac{w}{c}. \qquad (2.27)$$

Приняв во внимание, что при $I = 1$ бит, $\max(RT) = cT = 1$ бит, а также учитывая (2.24) и (2.26), можно установить, что

$$\frac{w}{c} \log_2 \left(1 + \frac{P_c T}{N_0} \frac{c}{w} \right) = 1 \qquad (2.28)$$

или

$$\log_2 \left(1 + \frac{\beta_E}{\beta_w} \right)^{\beta_E} = 1. \qquad (2.29)$$

Зависимость (2.27) определяет диаграмму обмена между удельными (на передачу 1 бита информации) затратами энергии и полосы в канале с гауссовскими шумами. Эта зависимость представлена на рис. 2.6. Она называется границей Шеннона и показывает, в какой мере улучшение одного из показателей информационной системы (β_E или β_w) приводит к ухудшению другого. Все оптимальные системы, наилучшим образом использующие для передачи информации энергию сигнала и полосу частот, занимаемую его спектром, располагаются на кривой (см. рис. 2.6). Все реальные системы — выше кривой: они требуют бóльших затрат полосы и энергии. Ниже этой кривой информационные системы работать не могут.

Обычно информационные системы проектируются с расчетом на работу при сравнительно хороших соотношениях сигнала к шуму $\frac{P_c}{P_ш} = q > 1$. Но при этом для (2.24) справедливо приближение

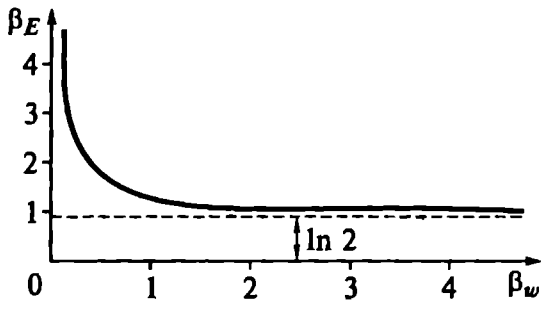


Рис. 2.6. Граница Шеннона для гауссовского канала

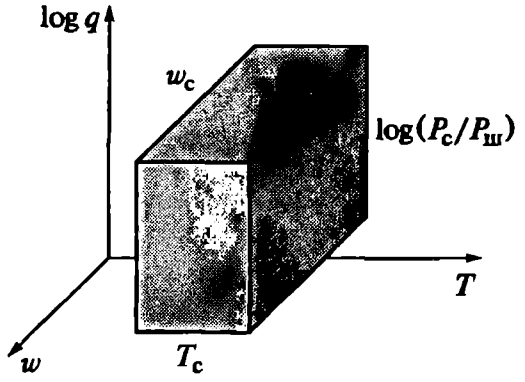


Рис. 2.7. Информационный объем сигнала и информационная емкость канала связи

$$c \approx w \log_2 q. \qquad (2.30)$$

Такая простая зависимость скорости передачи информации от трех переменных w, T и q позволяет поставить в соответствие сигналу некоторый объем в трехмерном пространстве рис. 2.7.

Если рассматриваются спектрально-временные свойства конкретного сигнала, то величина $wT\log q$ называется его (сигнала) информационным объемом. Если в (2.30) фигурируют полоса пропускания w, время наблюдения T и логарифм соотношения сигнала к шуму в канале передачи информации, то говорят, что c — это информационная емкость канала связи.

2.2. Качество информации

Качество функционирования информационных систем связано, прежде всего, с точностью передачи и(или) воспроизведения сообщений. Почти все иные показатели качества некоторым образом связаны с точностью. Так, дальность или зона действия — это расстояние, в пределах которого системы обеспечивают точность. Помехоустойчивость и помехозащищенность — способность сохранять точность в присутствии помех. Надежность — способность обеспечивать точность при сбоях и отказах аппаратуры и т.д.

Если сообщение — дискретная случайная величина, характеристикой точности служит вероятность ошибки его воспроизведения, т.е. вероятность трансформации истинного значения сообщения x_i в некоторое другое x_j, $i \neq j$. Это условная вероятность $P_{ij} = P(x_j | x_i)$. Но полная вероятность ошибки зависит еще и от того, сколько разных сообщений может формироваться и передаваться, а также каково априорное распределение вероятностей на ансамбле возможных сообщений:

$$P_{ош} = 1 - \sum_{j=1}^{m} P(x_j) P(x_j | x_j) = 1 - P_{прав}. \qquad (2.31)$$

Если x — непрерывная случайная величина, принимающая значения на сегменте $x \in (X_{min}, X_{max})$, мерой точности могут служить некоторые неслучайные функции от ошибки ее воспроизведения x^* на выходе информационной системы

$$\Delta = x - x^*, \qquad (2.32)$$

где x^* — оценка сообщения x, т.е. результат воспроизведения сообщения x информационной системой.

Для разных систем в разных ситуациях x может быть функцией времени $x(t)$, или пространственных координат $x(R)$, или и тех и других переменных $x(t, R)$. Соответствующую конструкцию будет иметь и ошибка Δ (2.32).

Полной характеристикой случайной ошибки может служить плотность распределения ее вероятностей $P(x)$. Как правило, от информационных систем требуется высокая точность. Это значит, что плотность распределения ошибок сравнительно узкая, т.е. существенно отличается от нуля в пределах интервала значений x, много меньшего интервала априорных значений. В пределах малой ширины закона распределения ошибок он может быть аппроксимирован квадратичной экспонентой. Иначе говоря, закон распределения вероятностей ошибок можно считать нормальным. При этом для характеристики ошибки достаточно двух величин:

математического ожидания $M\{\Delta\}$, иначе называемого ошибкой смещения;

дисперсии $\sigma_x^2 = M\left\{\left[\Delta - \mathrm{M}(\Delta)\right]^2\right\}$.

Когда ошибка зависит от времени и пространственных координат, для характеристики ошибки может понадобиться ее автокорреляционная функция или эквивалентная величина спектральной плотности.

Системы, осуществляющие несанкционированный доступ к чужому информационному ресурсу, т.е. использующие канал утечки информации, работают в таких условиях, когда модель только нормальных или только аномальных ошибок оказывается недостаточной или неполной. Действительно, работе таких систем сопутствует множество факторов, вызывающих как нормальные, так и аномальные ошибки. Можно считать, что нормальные ошибки вызываются, в основном, естественными причинами (шумами). Аномальные ошибки имеют антропогенную природу: их порождают специально организованные помехи, имеющие целью дезинформировать средства несанкционированного доступа к информации. Разумеется, такое разделение весьма условно: природные, не управляемые людьми факторы могут вызывать сбои и отказы аппаратуры и, как следствие, аномальные ошибки. Специально организованное противодействие может использовать не дезинформирующие, а шумовые помехи, которые не обманывают противника в информационном конфликте, а просто затрудняют его работу, несанкционированный доступ к информации.

Энтропия и количество информации дают возможность сравнивать информационные системы разного функционального назначения и структуры по производительности и информативности, по скорости передачи информации.

Показатель точности хорошо и достаточно описывает качество работы систем передачи и извлечения информации. Но для исследования характеристик информационной безопасности указанные характеристики не очень подходят, поскольку не отражают меры ценности информации.

2.3. Ценность информации

Ценность информации следует определить как максимальную пользу, которую может принести данное количество информации, или как те максимальные потери, к которым приведет утрата данного количества информации. Из этого определения следует, что, во-первых, ценность информации должна быть неубывающей функцией количества информации и, во-вторых, ценность информации может различаться для того субъекта, который эту информацию добывает (желает извлечь пользу из полученных данных), и для того субъекта, кто ее защищает (желает предотвратить потери от утраты или искажения данных).

Пусть имеются некоторые данные $x_k \in \mathbf{X}$, $k \in 1{:}K$. Например, $\mathbf{X}$ — массив записей на любом физическом носителе. В соответствии со спецификой и назначением этих данных может быть назначена функция риска $r(x_k)$ потери от утраты или искажения конкретного элемента данных x_k. Кроме того, на множестве $\mathbf{X}$ может быть задано распределение вероятностей угроз данным $P(x_k)$, т. е. вероятность утраты или разрушения этих данных. Организуя защиту информации, нужно влиять на распределение $P(x_k)$, используя для этого управляющую переменную $u_k = u(x_k)$, от которой распределение зависит. Для управления защитой информации нужно выделять некоторые ресурсы $u_k \in \mathbf{U}$. Поскольку эти ресурсы всегда ограничены, улучшение защиты одних данных будет проведено в ущерб безопасности других. Усредненный по всему множеству риск равен

$$\bar{R} = \sum_{k=1}^{k} r(x_k) P(x_k, u_k). \tag{2.33}$$

Естественно выбрать такую стратегию защиты (такое распределение ресурса средств защиты), которая минимизирует средний риск.

Для асимптотического непрерывного случая можно принять

$$\bar{R}_{\min} = \min_{u} \int_{\mathbf{U}} r(x) \mathrm{d}P(x, u). \tag{2.34}$$

При заданных функции потерь $r(x_k)$ и распределении вероятных рисков $P(x_k, u)$ вариационная задача (2.34) отыскания такого

управления u или такого распределения $P(x)$, которые минимизируют средний риск, может быть решена методом множителей Лагранжа.

Система, осуществляющая информационную агрессию против массива **X**, будет, по-видимому, так планировать свои действия, чтобы максимизировать приносимую ими пользу. Максимизация пользы может достигаться разными методами, зависящими как от цели агрессивных действий, так и от способа их осуществления, от задействованных ресурсов. Целью может быть несанкционированный доступ к данным для последующего их использования во вред собственнику и владельцу информации (разведка), разрушение данных в массиве записей (информационная диверсия), подмена истинных данных некоторыми ложными (дезинформация). Любые агрессивные действия против чужого информационного ресурса должны преследовать достижение некоторой пользы для одной из сторон информационного конфликта, приобретение этой стороной некоторого качества, усредненная стоимость которого

$$\bar{C} = \sum_{k=1}^{k} c(x_k) Q(x_k, v_k), \tag{2.35}$$

где $c(x_k)$ — функция выигрыша; $Q(x_k)$ — распределение вероятностей получения несанкционированного доступа к информационному массиву **X**; v_k — управляющая переменная.

Из приведенных рассуждений следует, что сторона, осуществляющая незаконное обращение с чужими массивами данных, должна предпринимать такие действия, которые максимизируют стоимость среднего выигрыша

$$\bar{C} = \max_{v} \sum_{k=1}^{k} c(x_k) Q(x_k, v_k), \tag{2.36}$$

влияя на распределение $Q(x_k,v_k)$ посредством выбора значения управляющей переменной v_k.

Решение задачи выбора оптимальной стратегии информационной агрессии можно получить на основе применения методов, аналогичных тем, что минимизируют средний риск системы защиты информации.

Несколько более сложная ситуация будет наблюдаться при выборе оптимальной стратегии действий одной стороны в условиях неопределенности относительно действий противника, т.е. оптимизации поведения в условиях информационного конфликта. Формально неопределенность относительно угроз означает неизвестность входящих в (2.33) и (2.36) распределений $P(x_k,u_k)$ и $Q(x_k,v_k)$.

Удовлетворительной моделью для описания такого конфликта может служить игра двух систем: А — защищающей информацию и В — осуществляющей информационную агрессию.

Поскольку в информационном конфликте несовпадающие интересы противников не точно противоположны, игра относится к классу биматричных неантагонистических бескоалиционных игр с ненулевой суммой [18].

В рассматриваемом частном случае двухстороннего конфликта наблюдается бескоалиционная биматричная игра. Модель бескоалиционной игры представляется в форме [18]:

$$\Gamma = \left(I, \{w_i\}\ i \in I, \{H_i\}\ i \in I\right), \tag{2.37}$$

где I — множество игроков (в частном случае двухстороннего конфликта $I = 2$); H_i — функция выигрыша i-го игрока, заданная на прямом произведении $\mathbf{W} = \prod_{i \in I} w_i$ чистых стратегий игроков.

Поскольку в рассматриваемом случае $I = 2$, такая игра описывается парой матриц размером (2×2): $\mathbf{H_A} = (a_{ij})$ и $\mathbf{H_B} = (b_{ij})$, или, что эквивалентно, матрицей $\mathbf{H_A H_B}$, каждый элемент которой представляет упорядоченную пару $a_{ij} b_{ij}$. Величины a_{ij} и b_{ij} представляют собой выигрыш соответствующего участника конфликта.

Формально разыгрывание сводится к тому, что игроки независимо друг от друга выбирают по элементу из множества чистых стратегий $\mathbf{W} = \{w_1; w_2; \ldots w_I\}$. После этого каждый из игроков получает выигрыш $H_i(w)$, определяемый ситуацией, т.е. совместным выбором игроков.

Участникам бескоалиционной игры ничто не мешает применять смешанные стратегии, т.е. вероятностный выбор некоторой чистой стратегии. Смешанные стратегии предусматривают распределение вероятностей $dp_i = dp(w_i)$ на множестве чистых стратегий, определяющее выбор конкретной чистой стратегии в каждом разыгрывании. Если игроки применяют смешанные стратегии, то математическое ожидание выигрыша i-го игрока оказывается равным

$$\langle H_i \rangle = \int_{\mathbf{W}} H_i(w_1; w_2; \ldots w_I)\, \mathrm{d}p(w_1)\, \mathrm{d}p(w_2) \ldots \mathrm{d}p(w_I), \tag{2.38}$$

где $\mathbf{W}$ — множество ситуаций в чистых стратегиях, т.е. $\mathbf{W} = \prod_{i \in I} w_i$.

Теория бескоалиционных игр изучает поведение разумных игроков. Таким игрокам не выгодно отступать от стратегий, обеспечивающих равновесие в игре. Для ситуации равновесия игры $\Gamma(\cdot)$

$$\mathbf{W}^* = \{w_1^*; w_2^*; \ldots w_I^*\} \tag{2.39}$$

в смешанных стратегиях характерно выполнение условия

$$H_i\left(w^*(w_i)\right) \le H_i\left(w^*\right); \quad w_i \in \mathbf{W}; \quad \forall i \in I. \qquad (2.40)$$

Последнее условие (2.40) как раз означает, что ни одному из игроков не выгодно отступать от ситуации равновесия (2.39), если только другие игроки от нее не отклоняются. А поскольку в бескоалиционных играх никакие стороны не могут вступать в соглашения, ситуация равновесия, приемлемая для каждого игрока, оказывается приемлемой для всех.

Для рассматриваемого случая информационного конфликта в качестве простейшей модели конечной бескоалиционной биматричной игры можно принять следующую.

Для простоты можно предположить, что стороны действуют одним из двух способов. Соответственно этому матрицы $\mathbf{H_A} = (a_{ij})$ и $\mathbf{H_B} = (b_{ij})$ имеют одинаковый ранг, равный 2. В содержательных терминах это означает, что сторона А производит выбор между двумя чистыми стратегиями, предусматривающими или не предусматривающими принятие мер по защите информации. Иначе говоря, она рассматривает некоторое количество информации (массив, файл, документ) как являющееся или не являющееся объектом информационной агрессии. В первом случае применяется один набор мер и средств информационной защиты, во втором — другой. Естественно, что специальные меры по защите информации требуют повышенных затрат и стоят дороже. Нападающая сторона В также выбирает между двумя стратегиями, одна из которых предусматривает информационную агрессию против А, а другая не предусматривает. При этом стоимость мер и средств, обеспечивающих проведение той или иной стратегии, заранее определена как для А, так и для В. Разумеется, практические ситуации гораздо более многообразны. В реальных условиях, предусмотренных законом и соответствующими нормативными актами, защищаемая информация всегда группируется под более чем двумя грифами. Причем каждому из уровней ограничения доступа к информации (каждому грифу) соответствует нормативно устанавливаемый необходимый состав мер и средств обеспечения безопасности. Выполнение требований этих норм определяет как объем затрат на защиту информации, так и стоимость проведения информационной агрессии. Кроме того, защита информации может предусматривать функционирование в условиях активных действий более чем двух противников. Тем не менее, двухальтернативная политика обеспечения информационной безопасности представляет определенный и не только теоретический интерес. Поэтому биматричная игра с функциями выигрыша сторон в виде матрицы размером 2×2 может служить максимально простой, но нетривиальной моделью информационного конфликта.

Интервалы значений расходов ресурсов сторон на реализацию стратегий конфликтного взаимодействия, а также интервалы значений функций выигрыша можно нормировать, как это принято в игровых задачах, к безразмерной единице. Такая нормировка не меняет основных свойств решения, лишь заменяя реальные игры стратегически эквивалентными. В результате такой нормировки смешанными стратегиями сторон А и В в конфликте станут, соответственно, векторы

$$\mathbf{u} = (\xi, 1 - \xi) \text{ и } \mathbf{v} = (\zeta, 1 - \zeta), \tag{2.41}$$

а каждой ситуации разыгрывания однозначно соответствует точка $(\xi;\zeta)$ единичного квадрата $[0;1]\times[0;1]$.

Выигрыши сторон, обозначенные, соответственно, как $H_A(\xi;\zeta)$ и $H_B(\xi;\zeta)$ будут

$$H_A(\xi;\zeta) = \mathbf{uAv}^T = (a_{11} - a_{21} - a_{12} + a_{22})\xi\zeta + (a_{12} - a_{22})\xi_1 + \\ + (a_{21} - a_{22})\zeta + a_{22}; \tag{2.42}$$

$$H_B(\xi;\zeta) = \mathbf{uBv}^T = (b_{11} - b_{12} - b_{21} + b_{22})\xi\zeta + (b_{12} - b_{22})\xi_1 + \\ + (b_{21} - b_{22})\zeta + b_{22}, \tag{2.43}$$

где $\mathbf{v}^T$ — транспонирование соответствующего вектора.

Для того чтобы ситуация $(\xi;\zeta)$ была приемлема для первого игрока, необходимо и достаточно выполнения неравенств на сторонах единичного квадрата:

$$H_A(1;\ \zeta) \le H_A(\xi;\ \zeta); \tag{2.44}$$

$$H_B(0;\ \zeta) \le H_B(\xi;\ \zeta), \tag{2.45}$$

или, с учетом (2.41), неравенств, равносильных (2.42) и (2.43):

$$\alpha\ (1 - \xi) - a(1 - \zeta) \le 0 \tag{2.46}$$

и

$$\alpha\xi\zeta - a\xi \ge 0, \tag{2.47}$$

где обозначено

$$\alpha = a_{11} - a_{21} - a_{12} + a_{22};\ a = a_{22} - a_{12}. \tag{2.48}$$

Множество всех решений для системы неравенств (2.46), (2.47), лежащих в полосе $[0;1]\times(-\infty;\infty)$, состоит:

из всех ситуаций вида $(0;\zeta)$, для которых $\alpha\zeta - a \le 0$;

всех ситуаций вида $(\xi;\zeta)$, для которых $\xi\in(0;1)$, а $\alpha\zeta - a = 0$;

всех ситуаций вида $(1;\zeta)$, для которых $\xi\in(0;1)$, а $\alpha\zeta - a > 0$.

Если $\alpha = a = 0$, то множество решений системы (2.46) и (2.47) есть либо прямая $\xi = 0$, либо прямая $\xi = 1$ (в зависимости от знака перед a в (2.47).

В том случае когда $\alpha \neq 0$, для всех решений системы (2.46) и (2.47), имеющих вид $(0, \zeta)$, должны выполняться условия

$$\begin{aligned} &\zeta \le \frac{a}{\alpha}, \quad \text{если } \alpha > 0; \\ &\zeta \ge \frac{a}{\alpha}, \quad \text{если } \alpha < 0. \end{aligned} \tag{2.49}$$

Множество значений ζ совпадает либо с полупрямой $\left(-\infty; \frac{a}{\alpha}\right]$, либо с полупрямой $\left[\frac{a}{\alpha}; \infty\right)$. Соответственно для решения системы (2.46) и (2.47) вида (2.49) должны выполняться условия

$$\begin{aligned} &\zeta \le 0, \quad \text{если } a > 0; \\ &\zeta \ge 0, \quad \text{если } a < 0. \end{aligned} \tag{2.50}$$

Иначе говоря, ζ должна принадлежать полосе $\zeta \in \left[\frac{a}{\alpha}; \infty\right)$ или $\zeta \in \left(-\infty; \frac{a}{\alpha}\right]$.

Наконец, для решения системы неравенств (2.48) и (2.49) должно быть

$$\zeta = \frac{a}{\alpha}. \tag{2.51}$$

Это значит, что множество таких решений лежит на отрезке между точками $\left(0, \frac{a}{\alpha}\right)$ и $\left(1, \frac{a}{\alpha}\right)$.

Таким образом, множество всех решений биматричной игры $\Gamma(\cdot)$ образует трехзвенную ломаную линию, а множество всех приемлемых для игрока А ситуаций является пересечением этого зигзага с единичным квадратом (рис. 2.8).

При $\frac{a}{\alpha} < 0$ множество приемлемых ситуаций совпадает с одной из вертикальных сторон квадрата. При $\frac{a}{\alpha} = 0$ и $\frac{a}{\alpha} = 1$ это множество состоит из двух сторон квадрата, соответственно нижней

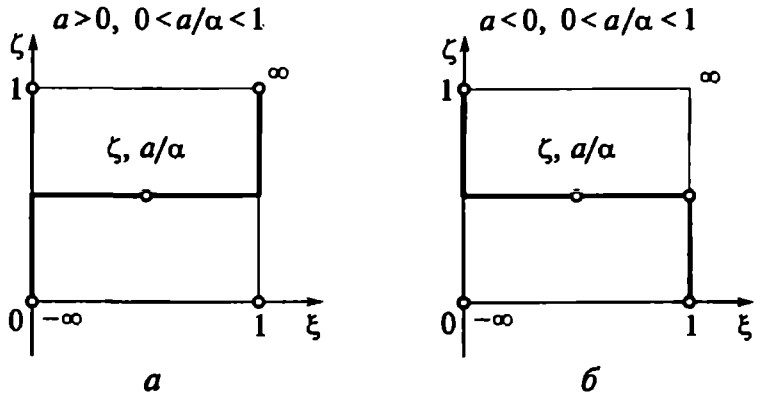


Рис. 2.8. Множество ситуаций, приемлемых для игрока, защищающего информацию

правой и верхней левой. Только при $0 < \frac{a}{\alpha} < 1$ множество приемлемых решений составляет трехзвенный зигзаг (см. рис. 2.8) для $\frac{a}{\alpha} > 0$ и $\frac{a}{\alpha} < 0$.

Аналогичным образом можно построить геометрический образ для интерпретации ситуаций, приемлемых для игрока В.

Введя обозначения, аналогичные (2.48),

$$\beta = b_{11} - b_{12} - b_{21} + b_{22};\ b = b_{22} - b_{21} \qquad (2.52)$$

можно построить геометрические образы ситуаций, приемлемых для игрока В. При $\beta = b = 0$ для игрока В приемлема любая ситуация. Если $\beta = 0$, но $b \neq 0$, множество всех приемлемых ситуаций совпадает либо с нижней, либо с верхней сторонами квадрата $[0;1]\times[0;1]$.

И, наконец, при $\beta \neq 0$ и $0 < \frac{b}{\beta} < 1$ множество приемлемых для игрока В ситуаций образует трехзвенный зигзаг (рис. 2.9).

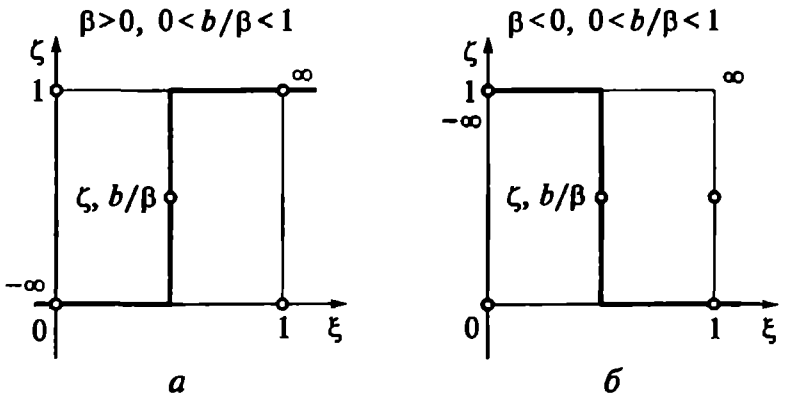


Рис. 2.9. Множество ситуаций, приемлемых для покушения на информационную безопасность

Объединяя множества всех приемлемых ситуаций (см. рис. 2.8. и 2.9) для каждого из рассматриваемой пары игроков, можно получить условия, определяющие ситуацию равновесия в биматричной игре $\Gamma(\cdot)$ (2.36). Если структуры матриц **A** и **B** таковы, что

$$a_{11} - a_{21} - a_{12} + a_{22} \neq 0$$

и (2.53)

$$b_{11} - b_{12} - b_{21} + b_{22} \neq 0,$$

то игра имеет ситуацию равновесия во вполне смешанных (отличных от чистых для каждого игрока) стратегиях, а именно

$$u_1^* = (\xi^*, 1 - \xi^*); \tag{2.54}$$

$$v_2^* = (\zeta^*, 1 - \zeta^*), \tag{2.55}$$

где

$$\xi^* = \frac{b}{\beta} = \frac{b_{22} - b_{21}}{b_{11} - b_{12} - b_{21} + b_{22}}; \tag{2.56}$$

$$\zeta^* = \frac{a}{\alpha} = \frac{a_{22} - a_{21}}{a_{11} - a_{12} - a_{21} + a_{22}}. \tag{2.57}$$

Точка на ограниченном квадрате, соответствующая этим вполне смешанным равновесным стратегиям в информационном конфликте, показана на рис. 2.10.

Анализ полученных соотношений (2.54), (2.55), (2.56) и (2.57) для смешанных стратегий в биматричной игре с нестрого противоположными интересами игроков показывает, что оптимальное поведение игроков, обеспечивающее состояние равновесия, должно быть таким же, как и в антагонистической игре. Но существенно здесь то, что оптимальная стратегия для игрока А оказывается такой же, как в антагонистической игре с матрицей выигрышей $\mathbf{H_B}$, а для игрока В — как в матричной игре с матрицей выигрышей $\mathbf{H_A}$.

При таком игровом подходе к анализу ценности информации за пределами рассмотрения остался вопрос о выборе объема ресурса, который может быть направлен на организацию защиты информации.

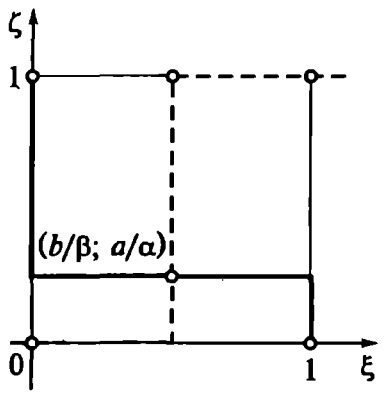


Рис. 2.10. Геометрический образ смешанных равновесных стратегий

Контрольные вопросы

1. По какой причине для количества информации принята логарифмическая мера?

2. Какая величина называется энтропией ансамбля (или источника) сообщений?

3. Перечислите основные свойства энтропии.

4. В каких единицах измеряется энтропия?

5. Как измеряется скорость передачи информации? Чем скорость передачи информации отличается от пропускной способности канала?

6. Полоса пропускания канала передачи информации уменьшилась в два раза, а соотношение сигнала к шуму в канале возросло с 10 до 13 дБ. Как при этом изменилась пропускная способность канала?

7. Один источник информации формирует сообщения x, которые принимают равновероятные значения на сегменте $x \in [-5\ \text{В}; 5\ \text{В}]$, а другой — нормально распределенные с нулевым средним значением $\langle x \rangle$ и со среднеквадратическим значением $\sigma_x = 1$ В. Какой источник обладает большей энтропией?

8. Какова связь между количеством информации и энтропией?

9. Как охарактеризовать качество информации, формируемой измерительной системой?

10. Объясните, как можно измерить ценность информации?

Глава 3
УГРОЗЫ БЕЗОПАСНОСТИ ИНФОРМАЦИИ И ИНФОРМАЦИОННЫЕ АТАКИ

3.1. Информационные угрозы

Угроза — это возможность возникновения такой ситуации (явления, события), следствием которой может стать нарушение безопасности информации. Угрозы информационной безопасности могут возникать на разных этапах жизненного цикла информационных систем и со стороны разных источников. Попытки реализации угроз информации называются информационными атаками на системы или просто атаками. Все известные угрозы информации можно классифицировать по нескольким критериям [6]. Эта классификация иллюстрируется графом рис. 3.1.

Данная классификация (см. рис. 3.1) подчеркивает не столько различие, сколько сходство свойств различных угроз. Так преднамеренные угрозы могут проявляться в случайных и непредсказуемых информационных нарушениях.

Угрозы нанесения ущерба информационным системам и обрабатываемой информации, обусловленные физическими воздействиями стихийных природных явлений, не зависящих от человека, — это случайные угрозы. Причинами случайных воздействий на элементы информационных систем при их эксплуатации могут быть отказы и сбои аппаратуры, непреднамеренные ошибки обслуживающего персонала, помехи в каналах передачи данных, вызванные воздействиями агрессивной внешней среды, ошибки разработчиков аппаратных и программных сегментов информационной системы, аварийные ситуации.

Частота отказов и сбоев аппаратуры, как и вероятность ошибок проектирования, увеличивается при усложнении систем.

Хотя человек как элемент автоматизированной системы обладает по сравнению с техническими средствами рядом преимуществ, прежде всего способностью адаптироваться к возникающим в процессе работы ситуациям, он в то же время имеет ряд недостатков. Основные недостатки — это утомляемость, зависимость психологических параметров от физического и эмоционального состояния, чувствительность к изменениям окружающей среды. Ошибки человека-оператора могут быть логическими (неправильно принятые решения), сенсорными (неправильное восприятие оператором информации) и оперативными, или моторными (непра-

Рис. 3.1. Угрозы информации

вильная реализация решения). Интенсивность ошибок человека может колебаться в пределах нескольких процентов от общего числа операций, выполняемых при обслуживании процесса обработки информации.

К угрозам случайного характера относятся также аварийные ситуации, которые могут возникнуть на объекте размещения автоматизированной системы. Аварийные ситуации — это отказы аппаратуры информационной системы, стихийные бедствия (пожары, наводнения, землетрясения, ураганы, разряды атмосферного электричества и др.). Вероятность подобных событий определяется прежде всего выбором технических решений в процессе проектирования информационной системы, но также организацией процесса ее функционирования.

По сравнению со случайными угрозами, круг искусственных, или преднамеренных, угроз обрабатываемой информации, вызванных человеческой деятельностью, более широк и опасен. Действие преднамеренных угроз направлено практически против всех без исключения элементов и подсистем, в совокупности образующих информационную систему. Чаще других реализуются следующие преднамеренные информационные угрозы:

несанкционированный доступ к информации посторонних лиц, не принадлежащих к числу легальных пользователей, и ознакомление с хранящейся и циркулирующей в информационных системах конфиденциальной информацией;

доступ легальных пользователей информационной системы к информации, на работу с которой они не имеют полномочий;

несанкционированное копирование сведений: программ и данных;

кража физических носителей и оборудования, приводящая к утрате информации;

умышленное уничтожение информации;

несанкционированная модификация документов и баз данных;

фальсификация сообщений, передаваемых по каналам связи;

отказ от авторства сообщения, переданного по каналам связи;

отказ от факта получения информации;

дезинформация, т.е. навязывание ложного сообщения;

разрушение информации деструктивными программными воздействиями, в частности — компьютерными вирусными.

3.2. Информационные атаки

Несанкционированный доступ (НСД) — наиболее распространенный вид информационных атак. Суть НСД состоит в том, что пользователь (нарушитель) получает возможность взаимодействия с информационной системой в нарушение правил разграничения доступа, установленных в соответствии с принятой в организации политикой обеспечения безопасности информации. Для НСД используются ошибки при создании систем защиты, нерациональном выборе средств защиты, их некорректной установке и настройке. НСД может быть осуществлен как с использованием штатных средств информационной системы, так и специально созданными для информационных атак аппаратными и программными средствами.

Источник угроз безопасности информации может находиться как в среде элементов информационной системы, так и располагаться вне этой среды. Источником преднамеренных угроз может выступать как законный пользователь системы, так и постороннее лицо. Если в первом случае за счет жесткого и неукоснительного выполнения всех организационных и режимных мероприятий, связанных с защитой информации, удается практически нейтрализовать действие данного вида угроз информации, то для нейтрализации его действий внешнего квалифицированного злоумышленника необходимо привлечение достаточно большого количества сил и средств.

В последнее время, учитывая широкое распространение информационных систем, интегрированных в глобальные информаци-

онно-вычислительные сети, приходится считаться с опасностью реализации угроз информации со стороны злоумышленника, находящегося вне информационной системы. Реализации подобного рода угроз называются удаленными атаками.

По характеру воздействия удаленные атаки можно разделить на пассивные, не оказывающие непосредственного влияния на работу информационной системы, и активные, наносящие прямой ущерб за счет нарушения конфиденциальности, целостности и доступности информации, а также, возможно, за счет негативного психологического воздействия на потребителя информации и пользователя информационной системы. Очевидной особенностью активного воздействия, по сравнению с пассивным, является принципиальная возможность его обнаружения.

Как отдельную группу можно выделить условно-пассивные информационные атаки, которые имеют целью подготовку к активной информационной атаке и предусматривают ведение компьютерной разведки, взлом системы защиты информации.

Основная цель практически любой атаки — получить несанкционированный доступ к информации. При этом следствием НСД всегда является перехват и(или) искажение информации. Возможность перехвата информации означает получение к ней доступа, но невозможность ее модификации. Иногда это следствие НСД называют утечкой информации. Перехват информации ведет к нарушению ее конфиденциальности.

Возможность искажения информации означает полный или частичный контроль над информационным обменом между подсистемами информационной системы.

Принципиально другой целью атаки является нарушение работоспособности элементов информационной системы. В этом случае не предполагается получение атакующим несанкционированного доступа к информации. Его основная цель — нарушить нормальное функционирование аппаратных и программных средств информационной системы и воспрепятствовать доступу к ресурсам атакованного объекта.

Таким образом, источниками преднамеренных угроз информации с доступом к элементам информационной системы могут быть злоумышленники, обладающие различными сведениями о структуре системы и способе ее функционирования и располагающие соответствующими техническими средствами для информационных атак.

3.3. Технические каналы утечки информации

Физический доступ к элементам информационной системы осуществляется по-разному. Прежде всего, при посредстве сигналов,

сопровождающих работу информационных систем. Каждое устройство хранения, передачи, обработки информации является источником излучения различной природы (электромагнитные, акустические и другие волны). Перехватывая и обрабатывая излучения, сопровождающие работу информационных систем, возможно получать разнообразные сведения о процессах, сопровождающих передачу и обработку данных. Источником подобного рода излучений могут быть различные электронные устройства, элементы системы электропитания (кабели электропитания и заземления), конструкции зданий и сооружений (металлические конструкции, оконные стекла, стены и т.п.), системы жизнеобеспечения (система отопления и вентиляции).

Как видно, потенциальные угрозы безопасности информации отличаются многообразием, сложностью структуры и функций. Их действие направлено практически против всех структурных компонентов современных информационных систем. Поскольку объем материальных ресурсов, выделяемых на защиту информации, обычно ограничен, актуальность приобретает задача их рационального распределения. Естественно, что усилия и материальные средства целесообразно расходовать на нейтрализацию наиболее опасных угроз, наносящих наибольший вред информационным системам. Это требование предусматривает необходимость предварительной оценки возможных угроз информации на всех этапах жизненного цикла информационных систем от замысла до утилизации.

Реализация большинства угроз безопасности информации связана с использованием технических каналов незаконного, несанкционированного доступа к сведениям и данным ограниченного доступа.

Работа радиоэлектронных устройств и систем сопровождается возникновением электромагнитных полей. И эти поля способны переносить сигналы, информативные не только для собственных абонентов, которым предназначаются циркулирующая по каналам систем и средств сообщения. Излучения, сопровождающие работу радиоэлектронных систем (РЭС), информативны для технических средств разведки. Более того, средства разведки могут принимать и использовать такие сигналы, которые РЭС создают в процессе функционирования непреднамеренно.

Большинство современных РЭС являются информационными, т.е. создающими электромагнитные поля и использующими процессы в электронных устройствах для передачи, приема, преобразования, обработки и хранения сообщений. Поэтому работу РЭС можно иллюстрировать предельно обобщенной функциональной схемой (рис. 3.2).

Для РЭС разного функционального назначения (передачи, извлечения информации, радиоуправления) и для источников информации разной природы эта схема должна быть детализирована

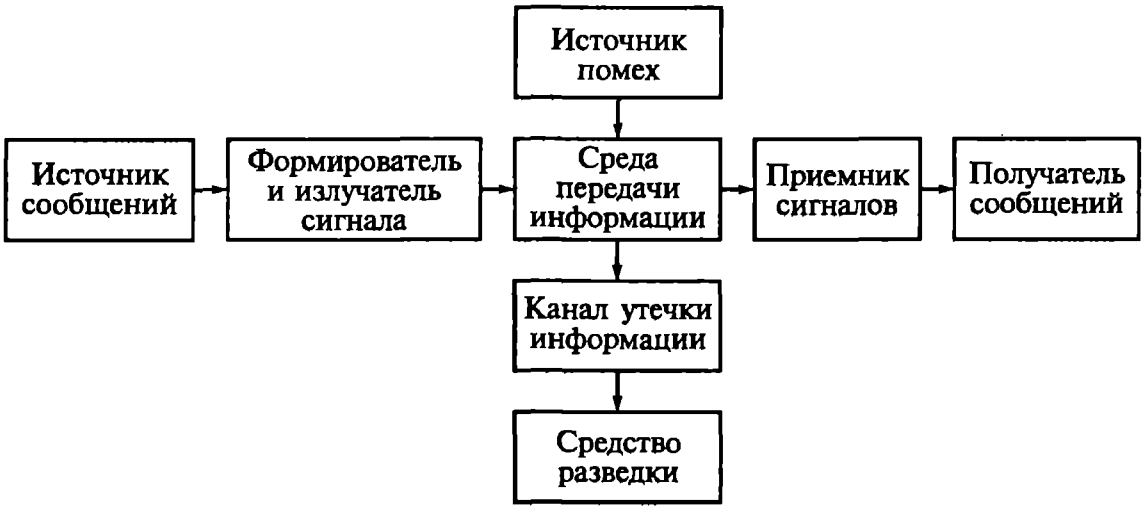


Рис. 3.2. Образование технического канала утечки информации

Рис. 3.3. Классификация технических каналов утечки информации

с тем, чтобы подчеркнуть их особенности. Но существенным остается то, что созданные РЭС электромагнитные поля создают угрозы несанкционированного приема (перехвата) сигналов, информативных для технических средств разведки или, иначе говоря, создают технические каналы утечки информации.

Источники помех, действующих в среде распространения сигналов, создают угрозы разрушения (утраты) и искажения информации. Посредством использования не естественных, природных, а специально созданных помех, возможно деструктивное воздействие на информационные системы. Реализация угрозы деструктивного воздействия может проявляться в виде поражения сигналов РЭС помехами или в виде дезинформации, когда помехи навязывают получателям ложные сообщения.

Тайное или явное получение сведений, циркулирующих по каналам информационных систем и сетей, называется перехватом или утечкой информации. Для перехвата используют технические каналы утечки информации, объединяющие средства несанкционированного доступа к сигналам и среду их (сигналов) распространения. Классификация технических каналов утечки, угрожающих безопасности информации в радиоэлектронных средствах, системах и сетях, иллюстрируется графом на рис. 3.3.

3.3.1. Электромагнитные каналы утечки информации

Электромагнитные каналы утечки информации образуются основными излучениями РЭС. Но не только ими. При работе РЭС непременно происходит побочное излучение в боковых лепестках диаграммы направленности антенных систем. Эти излучения могут обеспечивать довольно высокую плотность потока мощности. Настолько высокую, чтобы оказаться доступными техническим средствам радио- и радиотехнической разведки, располагающимся вне основной трассы распространения сигналов РЭС. Кроме того, нелинейные режимы работы усилителей (прежде всего, оконечных каскадов усиления мощности в передатчиках РЭС) сопровождаются появлением в спектре излучения высших гармоник основной рабочей частоты f_0 высокочастотного сигнала. Это излучения на частотах $f_n = nf_0$. При формировании колебания на основной рабочей частоте обычно используются синтезаторы частот, преобразующие относительно низкую частоту эталонного генератора в требуемое значение f_0. Поэтому в спектре выходного сигнала могут присутствовать колебания на частотах $f_m = f_0/m$.

Каналы утечки информации образуются не только за счет перехвата сигналов, излучаемых антеннами в основных и боковых лепестках диаграммы направленности антенны (ДНА). Различные технические средства, работающие с высокочастотными токами, напряжениями и полями, а также различного рода цепи, распо-

ложенные в непосредственной близости от таких средств, могут обладать антенным эффектом, т. е. непреднамеренно излучать электромагнитные волны. Приемные устройства РЭС при работе используют маломощные гетеродины и другие вспомогательные генераторы, которые тем не менее способны непреднамеренно излучать электромагнитные излучения вполне ощутимой мощности. Все эти колебания могут приниматься средствами разведки и нести информацию о пространственных координатах, параметрах модуляции и других защищаемых сведениях о параметрах и характеристиках РЭС. Кроме того, эти колебания могут быть модулированы сообщениями, циркулирующими по каналам передачи защищаемой информации.

Сосредоточенные антенны образуются неплотными стыками электромагнитных экранов и линий, канализирующих энергию СВЧ, окнами, дверями, вентиляционными отверстиями и другими технологическими проемами в стенах экранированных помещений, отверстиями в металлических кожухах приборов. К распределенным случайным антеннам относятся различного рода кабели, провода систем сигнализации, радиотрансляционные сети, трубы, металлические конструкции и т. п. Кроме основного излучения сигнала на несущей частоте в главном лепестке диаграммы направленности антенны, работу радиоэлектронных систем и средств сопровождают побочные и непреднамеренные электромагнитные излучения (ПЭМИ), которые тем не менее переносят сигналы, информативные для технических средств разведки. ПЭМИ создаются также и системами, не рассчитанными на работу с излучением.

Для защиты информации от утечки по электромагнитным каналам максимально снижают мощность опасного сигнала и принимают меры для уменьшения паразитных и непреднамеренных излучений.

Снижение мощности основного излучения неизбежно уменьшает мощность сигнала и на входе собственных абонентских приемников. Поэтому снижение мощности основного излучения для защиты информации обязательно предусматривает применение всех доступных способов улучшения качества приема слабых сигналов: помехоустойчивое кодирование, оптимальные виды модуляции и оптимальные способы приема.

Для уменьшения мощности паразитных и непреднамеренных излучений, доступных техническим средствам разведки, применяют специальные приемы конструирования РЭС. Создавая антенные системы, добиваются такого амплитудно-фазового распределения на раскрыве, при котором минимизируется уровень излучения по боковым лепесткам диаграммы направленности [1]. Для элементов аппаратуры и устройств, способных создавать непреднамеренные излучения электромагнитных полей, прежде всего предусматривают экранирование [13, 25].

Самовозбуждение усилителей разных типов и назначения, используемых в составе РЭС, возможно за счет образования паразитных положительных обратных связей. В частности, паразитные самовозбуждения наблюдаются в низкочастотных трактах РЭС. Но частоты самовозбуждения таких низкочастотных устройств и подсистем могут быть настолько высокими, что непреднамеренно генерируемые колебания создадут распространяющиеся в пространстве электромагнитные поля. Колебания на частотах самовозбуждения, как правило, оказываются модулированными сигналами, информативными для разведки. Обычно эти сигналы не защищены от расшифровки при перехвате.

Индукционные каналы утечки. Они возникают за счет наводок информационных сигналов через емкостные и индуктивные связи между проводными линиями, по которым распространяются защищаемые информационные сигналы, и другими линиями (телефонной связи, различной сигнализации, силовой сети), выходящими за пределы охраняемых зон и территорий, и потому доступными для контроля техническими средствами разведок.

Для передачи сигналов между различными подсистемами и устройствами РЭС применяют двухпроводные линии, несимметричные и симметричные кабели. Между проводниками, образующими линию передачи сигналов, существует разность потенциалов, которая порождает электрическое поле с напряженностью **E**. По проводникам линии течет электрический ток. Этот ток порождает магнитное поле **H**. Силовые линии электрического поля замыкаются в пространстве между проводами, а силовые линии магнитного поля — вокруг каждого проводника.

На рис. 3.4 показана емкостная связь, возникающая между двумя проводами: сигнальной цепи и цепи перехвата. Напряжение $U_с$ в сигнальной цепи индуцирует в цепи перехвата информации напряжение $U_п$.

По одному из проводов (см. рис. 3.4) передается сигнал, а другой включен в контур цепи перехвата информации. Система совершенно симметрична, поэтому конкретизации того, какой провод в какую цепь включен, не требуется. Напряжение, созданное на проводе цепи перехвата, будет, очевидно, равно

$$U_п = U_с \frac{C_0}{C_0 + C}, \qquad (3.1)$$

где $U_п$ и $U_с$ — соответственно напряжение сигнала в цепи перехвата и напряжение в сигналь-

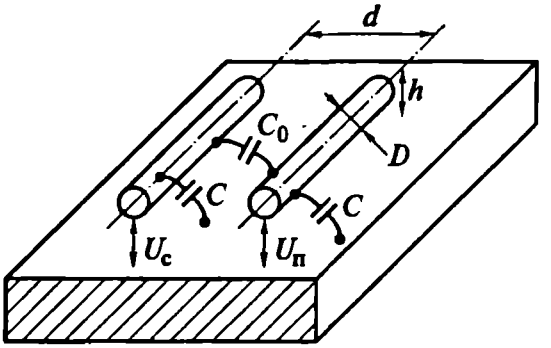


Рис. 3.4. Наводки посредством электрического поля

ном проводе. Емкость C_0 между проводами составляет $C_0 \approx \dfrac{\pi\varepsilon}{\ln\dfrac{2d}{D}}$.

Тем же соотношением определяется и емкость между проводом и заземленной подложкой (с учетом того, что расстояние до подложки h). Поэтому переходное ослабление между сигнальным проводом и цепью индукционного канала перехвата информации составляет

$$\frac{U_{\text{п}}}{U_{\text{с}}} = \frac{\ln\dfrac{2d}{D}}{\ln\dfrac{8dh}{D^2}}. \qquad (3.2)$$

Как видно из (3.2), угроза перехвата тем опаснее, чем больше уровень сигнала в линии и чем меньше расстояние D от сигнальной цепи до датчика сигнала в канале утечки информации. Существенно также влияние емкости C_0 между проводами. Ее нужно всемерно уменьшать, например, применяя экранирование сигнального провода (используя коаксиальную линию).

Магнитное поле, образующееся вокруг токонесущего сигнального провода, индуцирует ток в цепи перехвата информации. Магнитный поток Ф пересекает петлю этой цепи, образованной отрезком провода длиной l и материалом подложки (рис. 3.5). Этот поток создаст на конце провода напряжение

$$U_{\text{п}}(t) = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_d^{d+h} \mu l H \mathrm{d}r = -\mu l \int_d^{d+h} \frac{\mathrm{d}}{\mathrm{d}t} H \mathrm{d}r =$$

$$= -\frac{\mu l}{2\pi}\int_d^{d+h} \frac{\mathrm{d}}{\mathrm{d}t} i_{\text{с}}(t)\frac{\mathrm{d}r}{r} = -\frac{\mu l}{2\pi}\ln\left(\frac{d+H}{d}\right)\frac{\mathrm{d}i_{\text{с}}(t)}{\mathrm{d}t}, \qquad (3.3)$$

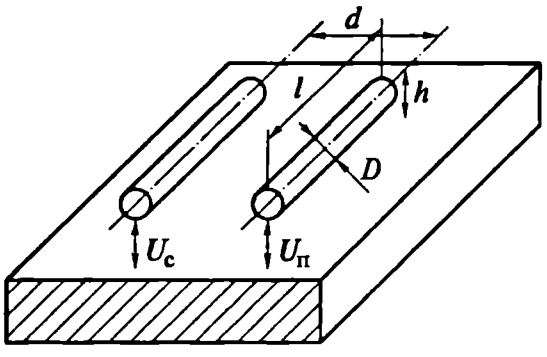


Рис. 3.5. Магнитные наводки

где μ — магнитная проницаемость среды; $\mu H = B$ — индукция; $i_{\text{с}}$ — ток в сигнальном проводе.

При синусоидальном токе $i_{\text{с}} = I_{\text{с}}\cos\omega t$ из (3.3) следует, что амплитуда напряжения на выходе петли перехвата информации с использованием магнитного поля составит

$$U_{\text{п}} = \mu l I_{\text{с}}\omega\ln\left(\frac{d+h}{d}\right). \qquad (3.4)$$

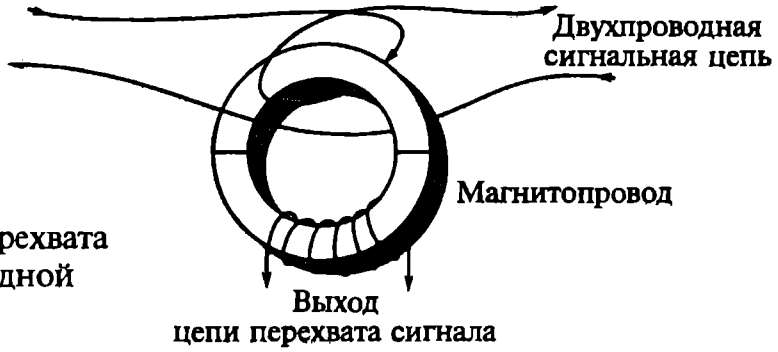


Рис. 3.6. Устройство перехвата сигнала в двухпроводной линии

Как видно, напряжение, наводимое сигнальной цепью в петле перехвата, увеличивается с ростом величины тока в сигнальном проводе и его частоты, с увеличением магнитной проницаемости среды. Поэтому устройства средств перехвата (например, телефонного сигнала) обычно используют трансформаторы на разъемном магнитопроводе с высокой проницаемостью, который охватывает провода линии передачи сигнала изогнутые таким образом, чтобы токи в них создавали магнитные потоки одного направления (токи текут по направлению стрелок (рис. 3.6)), и многовитковой вторичной обмоткой, в которой напряжения на концах последовательно включенных витках суммируются.

Просачивание сигналов. Просачивание сигналов по цепям электропитания происходит потому, что токи в цепях информационных сигналов замыкаются через вторичный источник питания РЭС. Эти токи создают падения напряжения на выходном сопротивлении источника питания и при недостаточной развязке между его выходными и входными цепями могут обнаруживаться во внешней цепи, часто выходящей за границы контролируемой зоны и доступной для контроля с помощью технических средств разведки. Информационный сигнал может проникать в цепи электропитания также в результате того, что ток, потребляемый мощными оконечными каскадами РЭС (выходными каскадами передатчиков, оконечными каскадами звукоусилительной аппаратуры и других технических средств обработки информации), зависит от амплитуды информационного сигнала. Неравномерная нагрузка источника питания при этом приводит к колебаниям потребляемого тока во внешней цепи, т.е. к амплитудной модуляции тока в силовых цепях информационными сигналами.

Аналогичная картина может возникнуть при воздействии информационных сигналов и на другие линии, не предназначенные для передачи информации. В этих линиях может наблюдаться паразитная модуляция информационными сигналами, но могут существовать и наводки, индуцированные токами в сигнальных цепях, если есть условия для индуктивной и (или) емкостной связи с сигнальными линиями. Этот канал утечки возникает чаще всего при параллельной прокладке цепей на участках большой протя-

женности в том случае, если эти цепи выходят за границу контролируемой зоны, где возможен перехват информации.

Для уменьшения опасности перехвата информации за счет просачивающихся сигналов, все линии, выходящие из контролируемой зоны и входящие в нее, снабжаются специальными фильтрами информационных сигналов. Провода и кабели сигнальных цепей экранируют. Кроме того, в пределах контролируемой зоны устраняют все посторонние провода и кабели, а все необходимые, но не задействованные кабели и провода замыкают накоротко и заземляют. Для уменьшения наводок прокладку кабелей сигнальных цепей с другими линиями производят так, чтобы они пересекались под прямым углом. Не допускается гальванический контакт экранирующих оболочек пересекающихся кабелей.

Заземление устройств. Кроме защиты от катастрофических энергетических проявлений (короткое замыкание цепей электропитания на корпус РЭС или удары молний), заземление устройств имеет очень важное значение для снижения уровня угроз перехвата информации. Дело в том, что переходное сопротивление заземления, зависящее от многих факторов: структура и влажность грунта, материала и конструкции заземленного проводника, может быть весьма значительным и колебаться в пределах 2...3 Ом до килоома и даже выше [25]. Токи в цепи заземления, имеющие информативные для средств разведки составляющие, могут создавать на таких сопротивлениях заметные падения напряжений, достаточные для уверенного перехвата информации.

Гальваническую связь с землей, кроме заземлителей, имеют нулевой провод сети электропитания, металлические трубопроводы водопроводной, отопительной и других систем, металлическая арматура железобетонных конструкций, металлические оболочки кабелей связи и т.д. Все эти цепи и конструкции вместе с контуром заземления образуют широкоразветвленную систему заземления.

Различают рабочее, защитное и технологическое заземление.

Рабочее заземление предназначено для подключения к общей шине технических систем с целью использования земли в качестве одного из проводников электрической цепи. Такое заземление применяется во всех однопроводных цепях телефонной и телеграфной связи, сигнализации и некоторых других.

Защитное заземление обеспечивает безопасность обслуживающего персонала, целостность и правильность работы. К защитному заземляющему устройству присоединяются нетоковедущие металлические части РЭС.

Технологическое заземление соединяет с заземлителем экраны оборудования и оболочки кабелей для устранения помех в устройствах с низким уровнем сигнала.

Причинами утечки информации через систему заземления могут быть также электромагнитные поля опасного сигнала в месте

размещения заземляющих проводов и шин заземления; асимметрия линий передачи сигналов, использующих в качестве одного из проводников землю.

Перехват сигнала возможен при съеме токов, наведенных в экранах кабелей, металлических трубах системах водоснабжения и отопления, проходящих в потенциальной зоне контура заземления, а также при регистрации распределения потенциала в грунте в районе контура заземления, при выходе потенциальной зоны заземления за границу контролируемой зоны.

Для парирования угрозы перехвата информации в цепи заземления используются контуры заземления, состоящие из нескольких проводников (заземлителей), соединенных параллельно и заглубленных в грунт на 2...3 м. Иногда, если поверхностные слои грунта очень сухие, применяют глубинные заземлители — электроды, погруженные на большую глубину (10...30 м) во влажные и хорошо проводящие слои.

Заземлители размещаются в контролируемой зоне и исключают возможность гальванического контакта с подземными коммуникациями, выходящими за ее пределы. Если обыкновенные заземлители разместить в пределах контролируемой зоны не удается и их невозможно удалить от подземных коммуникаций, то необходимо применить глубинные заземлители. Длина заземляющих проводов системы внутреннего телевидения не должна превышать 100 м. Величина сопротивления заземления для предотвращения перехвата сигнала не должна превышать 4 Ом.

Параметрический канал. Этот канал утечки информации создается в результате воздействия высокочастотного сигнала на технические информационные системы из-за пределов контролируемой зоны. Такое воздействие осуществляется либо с помощью электромагнитного поля (высокочастотное облучение), либо при гальваническом подключении высокочастотного генератора к электрическим цепям (высокочастотное навязывание).

При взаимодействии внешнего электромагнитного поля с техническими устройствами (с аппаратурой телефонной связи, средствами звукозаписи и звуковоспроизведения) возникает вторичное рассеянное поле.

Параметры этого поля модулируются сообщениями, которые циркулируют в устройствах, рассеивающих внешние высокочастотные поля. Для развязки облучающих колебаний и рассеянных модулированных информационных сигналов используют импульсное излучение.

Для защиты технических систем от утечки информации по параметрическим каналам применяют электромагнитное экранирование технических систем и информационных цепей, шунтируют элементы устройств, способных модулировать высокочастотный облучающий сигнал, конденсаторами, имеющими малое сопро-

тивление для высокочастотных сигналов и не влияющих на низкочастотные информативные для средств разведки.

Эффективный путь парирования угрозы утечки по параметрическому каналу — создание активных помех, маскирующих рассеянное электромагнитное поле или искажающих его информативные параметры. Так, например, для защиты от съема информации посредством лазерного микрофона, который подсвечивает лазерным лучом оконное стекло, вибрирующее в такт изменения звукового давления в помещении, и выделяет акустическую информацию, детектируя фазу отраженного сигнала, применяют шумовую вибрацию стекол. Такая активная модулирующая помеха препятствует перехвату сообщений. Но могут применяться и аддитивные помехи, затрудняющие прием весьма слабых рассеянных сигналов в параметрических каналах утечки информации.

Аппаратные закладки (или закладные устройства). Этим термином обозначаются электронные устройства тайного съема сигналов, циркулирующих по каналам информационных устройств и систем. Перехваченными информационными сигналами модулируется несущее колебание, формируемое передатчиком закладки. Это модулированное колебание излучается или передается по проводам каких-либо неинформационных линий (цепей питания, сигнализации, телефонным линиям).

Обнаружение электронных устройств скрытного съема информации производится по их демаскирующим признакам: тонким проводам неизвестного назначения подключенным к спрятанному и закамуфлированному устройству; наличию в проводах линий неизвестного постоянного напряжения (питания закладки) или информационного сигнала. Весьма продуктивные методы обнаружения закладок используют анализ электромагнитной обстановки (для обнаружения излучений радиозакладок); нелинейную радиолокацию (для обнаружения устройств, содержащих полупроводниковые p—n-переходы); рентгеноскопию для обнаружения устройств, скрытых за непрозрачными для других полей защитными экранами.

3.3.2. Акустические каналы утечки информации

В акустических каналах утечки информации распространяются сигналы, переносимые механическими колебаниями упругой среды. К названию акустический обычно добавляют определения, указывающие и на характер сигнала (например, речевой сигнал), и на среду распространения этого сигнала. Различают воздушные акустические каналы, по которым распространяются колебаниями воздушной среды. В вибрационных каналах акустическая информация переносится механическими колебаниями твердых сред, прежде всего строительных конструкций (панелей стен и перекрытий, труб водоснабжения, отопления, вентиляционных коро-

бов и др.). Жидкие среды могут образовывать гидроакустические каналы утечки информации. Подобно параметрическим электромагнитным каналам, существуют и параметрические акустические каналы, в которых акустические сигналы управляют параметрами некоторых других физических полей.

Приемниками сигналов в воздушных акустических каналах служат разнообразные микрофоны: миниатюрные, высокочувствительные, направленные. Микрофоны объединяются со звукозаписывающими устройствами (диктофонами) или специальными передатчиками для трансляции сигналов по радиоканалам, оптическим и инфракрасным каналам, силовой сети электропитания и другим магистралям. Автономные устройства, конструктивно объединяющие миниатюрные микрофоны с передатчиками, служат для перехвата речевой информации и называются закладными устройствами для перехвата речевой информации, или просто акустическими закладками.

Вибрационный канал также может использоваться для перехвата информации при помощи закладных устройств. Перехваченные виброакустические сигналы усиливаются и транслируются при помощи радиоэлектронных устройств. Поэтому закладные устройства для перехвата виброакустических сигналов называются радиостетоскопами.

Параметрический оптико-электронный канал утечки акустической информации образуется при модуляции лазерного луча, отраженного оконными стеклами. Эти стекла вибрируют в такт колебаниям, создающимися речевыми сигналами в помещениях за стеклами. При вибрации стекол изменяется электрическая длина трассы, по которой распространяется отраженный стеклом оптический луч лазера, и, соответственно, изменяется фаза сигнала в точке приема. Демодуляция принятого оптического сигнала позволяет выделять акустическую информацию.

Приведенный перечень конкретных конфигураций технических каналов перехвата информации не исчерпывает их многообразия, но дает представление о методах реализации угроз информации, доступ к которой ограничивается.

Контрольные вопросы

1. Перечислите основные угрозы информации.
2. Какие вам известны информационные атаки?
3. Чем отличается несанкционированный доступ от перехвата информации?
4. Как возникают электромагнитные каналы утечки информации?
5. Чем отличаются паразитные электромагнитные излучения от непреднамеренных?
6. Для чего применяется экранирование?
7. Дайте определение параметрического канала утечки информации.

Глава 4
ФИЗИЧЕСКИЕ ПОЛЯ, СОЗДАЮЩИЕ КАНАЛЫ УТЕЧКИ ИНФОРМАЦИИ

4.1. Многообразие физических полей

Современные средства несанкционированного добывания информации (технические средства разведки — ТСР) используют для достижения своих целей, без преувеличения, все принципиальные и технологические возможности.

Для этого ТСР формируют каналы утечки информации за счет перехвата сигналов, переносимых всеми мыслимыми физическими полями, которые сознательно или непреднамеренно формируются техническими системами и сопровождают их функционирование.

Физическое поле — это особая форма существования материи, связывающая частицы вещества друг с другом и передающая действие одних объектов на другие с конечной скоростью. Едва ли не самым важным для работы современных информационных систем видом поля, детально изученным физикой и освоенным техникой информационных систем, является электромагнитное поле.

Но для несанкционированного доступа к информации, а следовательно, и для ее защиты весьма важно изучение и использование полей, образованных упругими механическими колебаниями и волнами в воздухе (акустические поля), твердых телах (виброакустические поля), жидких средах (гидроакустические поля).

Такую же физическую природу имеют волновые поля в упругой земной коре. Эти поля информативны для сейсмических разведок. Но волновые процессы, сопровождающие перенос энергии, а следовательно, и сигналов электромагнитными и акустическими полями вовсе не исчерпывают класс физических полей, информативных для технических разведок.

Значительный риск утечки информации создают поля температур, поля концентрации вещества в пространстве.

Эти поля создаются многими техническими системами и комплексами, работа которых сопровождается значительными преобразованиями вещества и энергии (взрывами, радиоактивными излучениями, работой авиационных и ракетных двигателей и т.п.).

4.2. Электромагнитные поля

4.2.1. Электромагнитные поля и волны

Радиоволны, тепловое и ультрафиолетовое излучение, свет, рентгеновское и γ-излучение — это все волны электромагнитной природы, но разной длины. И все эти волны используются техническими средствами разведки. Шкала электромагнитных волн, упорядоченных по частоте f, длине волны λ и названию диапазона, представлена на рис. 4.1.

В соответствии с законом электромагнитной индукции, в контуре, охватывающем изменяющееся магнитное поле, возникает ЭДС, которая возбуждает в этом контуре ток. Проводник здесь не играет существенной роли. Он лишь позволяет обнаружить индуцированный ток. Истинная сущность явления индукции, как установил Дж. К. Максвелл, заключается в том, что в пространстве, где изменяется магнитное поле, возникает изменяющееся во времени электрическое поле. Это изменяющееся во времени электрическое поле Дж. К. Максвелл назвал током электрического смещения.

В отличие от поля неподвижных зарядов, силовые линии изменяющегося во времени электрического поля (тока электрического смещения) могут быть замкнуты так же, как и силовые линии магнитного поля. Поэтому между электрическими и магнитными полями существуют тесная связь и взаимодействие, которые подчиняются следующим законам.

• Переменное во времени электрическое поле в любой точке пространства создает изменяющееся магнитное поле. Силовые линии магнитного поля охватывают силовые линии вызвавшего их переменного электрического поля. В каждой точке рассматриваемого пространства вектор напряженности электрического поля **E** и вектор напряженности магнитного поля **H** ортогональны друг другу.

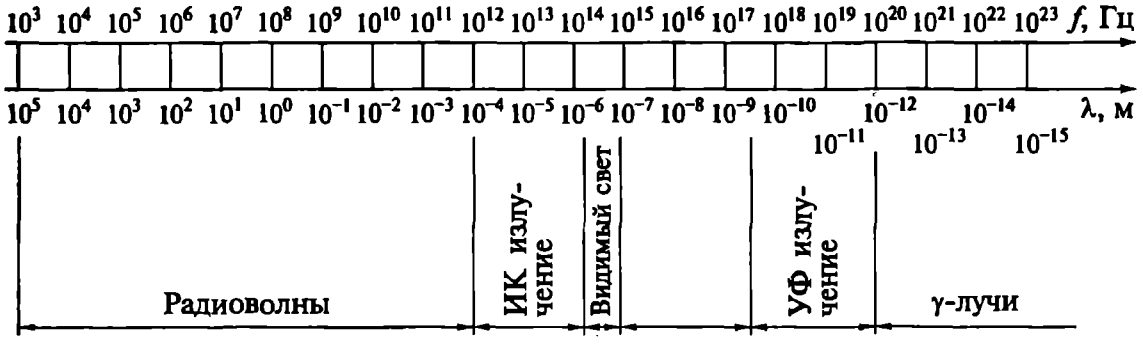


Рис. 4.1. Шкала электромагнитных волн

• Переменное во времени магнитное поле в любой точке пространства создает изменяющееся электрическое поле. Силовые линии электрического поля охватывают силовые линии вызвавшего его переменного магнитного поля. В каждой точке рассматриваемого пространства вектор напряженности магнитного поля **H** и вектор напряженности электрического поля **E** взаимно перпендикулярны.

Совокупность переменного электрического поля **E** и неразрывно связанного с ним переменного магнитного поля **H** образует электромагнитное поле.

Важнейшая особенность электромагнитного поля состоит в том, что оно перемещается в пространстве во все стороны от точки, в которой возникло первоначальное возмущение. Это поле может существовать самостоятельно после того, как источник электромагнитного возмущения перестал действовать. Возникшее в какой-либо точке пространства быстро изменяющееся во времени электрическое поле возбуждает в соседних точках окружающего пространства переменное магнитное поле, которое, в свою очередь, также возбуждает электрическое поле и т. д. Изменяющиеся электрические и магнитные поля, переходя от точки к точке пространства, распространяются в вакууме со скоростью света ($3 \cdot 10^8$ м/с). Перенос волной электромагнитной энергии в пространстве характеризуется вектором, равным векторному произведению напряженностей электрического и магнитного полей, Вт/м^2:

$$\mathbf{П} = \mathbf{E} \times \mathbf{H}. \qquad (4.1)$$

Направление вектора **П** совпадает с направлением распространения волны, а его модуль численно равен количеству энергии, которую волна переносит в единицу времени через единичную площадку, расположенную перпендикулярно направлению распространения волны. Понятие о потоке энергии любого вида было впервые введено Н. А. Умовым в 1874 г. Формула для вектора **П** была получена на основании уравнений электромагнитного поля Дж. Г. Пойнтингом в 1884 г. Поэтому вектор **П** (4.1) именуют вектором Умова — Пойнтинга.

Процесс распространения периодически изменяющегося электромагнитного поля — волновой. Электромагнитные волны излученного поля, встречая на своем пути проводники, возбуждают в них ЭДС той же частоты, что и частота создающего наведенную ЭДС электромагнитного поля. Часть энергии, которую переносят электромагнитные волны, передается токам, возникающим в проводниках.

Расстояние, на которое перемещается фронт волны за время, равное одному периоду электромагнитного колебания, называют *длиной волны*

$$\lambda = cT. \qquad (4.2)$$

Используя поверхности равных фаз, длину электромагнитной волны можно определить как кратчайшее расстояние между двумя поверхностями равных фаз, на которых фазы отличаются на 2π. Поверхность равных фаз — это фронт волны. В зависимости от формы поверхности равных фаз (или волнового фронта) различают плоские, цилиндрические и сферические волны. Все перечисленные типы электромагнитных волн являются поперечными электромагнитными волнами: у них векторы **E** и **H** осциллируют в направлениях, перпендикулярных направлению распространения волны (направлению вектора **П**).

Отношение амплитуд напряженностей взаимосвязанных электрических и магнитных полей, равное для свободного пространства, Ом,

$$Z_0 = \frac{|\mathbf{E}|}{|\mathbf{H}|} = \frac{E}{H} = 120\pi \cong 377, \qquad (4.3)$$

называется *волновым сопротивлением свободного пространства.*

Среднее значение плотности потока энергии за период электромагнитных колебаний связано в соответствии с (4.1) и (4.3) с напряженностью электрического поля соотношением

$$\langle|\mathbf{П}|\rangle = \frac{1}{2}\frac{E^2}{Z_0}.$$

Мощностью излучения источника называется величина, численно равная среднему количеству энергии, которую теряет этот источник электрического поля на образование потока электромагнитной энергии в единицу времени сквозь замкнутую поверхность, охватывающую источник. В том случае, когда гипотетический источник электромагнитного поля находится в центре сферы радиуса R и равномерно (изотропно) излучает во все стороны, мощность его излучения

$$P_\Sigma = 4\pi R^2 \langle|\mathbf{П}|\rangle. \qquad (4.4)$$

4.2.2. Излучение электромагнитных волн радиодиапазона антеннами

Устройства, специально предназначенные для излучения и(или) приема электромагнитных волн, — это антенны. Передающие и приемные антенны обладают свойством взаимности (обратимости), в соответствии с которым одна и та же антенна может как излучать, так и принимать электромагнитные волны. Основные

параметры антенны в режиме излучения, сохраняются и при приеме антенной электромагнитных волн.

Современные антенные устройства весьма разнообразны как с точки зрения выполняемых задач, так и с точки зрения конструкций. Конструкция антенны зависит от рабочего диапазона волн, желаемой направленности излучения, величины излучаемой мощности, места установки и т.д. Принципы действия — физические принципы преобразования антеннами подводимых фидерами электромагнитных колебаний в электромагнитное поле в пространстве и, наоборот, преобразование поля в токи и напряжения на входе приемных устройств также весьма разнообразны.

Мощностью излучения P_Σ называется среднее количество электромагнитной энергии, излучаемой антенной в единицу времени.

Полная мощность P, потребляемая антенной от источника сигнала, складывается из мощности излучения P_Σ и мощности потерь $P_п$. Последняя является следствием конечной проводимости проводников антенны, несовершенства диэлектриков, а также потерь мощности в земле и окружающих предметах. Таким образом, полная мощность в передающей антенне

$$P = P_\Sigma + P_п. \tag{4.5}$$

В тех случаях, когда известна амплитуда тока на входе антенны, в месте соединения антенны с линией передачи (фидерной линией), каждую из мощностей в (4.5) можно представить в виде:

$$P = \frac{R_а I_m^2}{2}; \quad P_\Sigma = \frac{R_\Sigma I_m^2}{2}; \quad P_п = \frac{R_п I_m^2}{2}, \tag{4.6}$$

где $R_а$ — полное сопротивление антенны; R_Σ — сопротивление излучения антенны; $R_п$ — сопротивление потерь; I_m — амплитуда тока на клеммах антенны.

Сопротивление излучения R_Σ равно такому активному сопротивлению, на котором при токе, равном току на входе антенны, рассеивается мощность, равная излучаемой ею мощности. Величина сопротивления излучения антенны зависит от характера распределения тока в антенне, а также от соотношения размеров антенны и длины излучаемой электромагнитной волны. Так, например, все полуволновые вибраторы обладают $R_\Sigma = 73{,}1$ Ом, а все вибраторы длиной в одну волну имеют $R_\Sigma = 210$ Ом. В общем случае сопротивление излучения антенны является комплексной величиной, мнимая часть которой определяет реактивную мощность излучения, локализованную в ближней зоне антенны (зоне индукции).

Антенна преобразует энергию источника электромагнитных колебаний в энергию электромагнитного поля. Коэффициент по-

лезного действия этого преобразователя определяется отношением

$$\eta_a = \frac{P_\Sigma}{P} = \frac{P_\Sigma}{P_\Sigma + P_п} = \frac{1}{1 + P_п/P_\Sigma} \qquad (4.7)$$

и оказывается тем больше, чем больше сопротивление излучения по сравнению с сопротивлением потерь. Величина КПД антенн получается достаточно высокой, например КПД полуволнового вибратора $\eta_a \approx 0{,}9$.

Антенна излучает энергию не изотропно. О энергии, излучаемой антенной в единицу телесного угла в различных направлениях, судят по диаграмме направленности. Диаграммой направленности антенны (ДНА) по полю называется зависимость напряженности электрического поля, создаваемого антенной в равноудаленных точках дальней зоны, от направления излучения. С пространственным представлением этой зависимости работать довольно сложно, поэтому обычно строят не пространственную ДНА, а ее сечение двумя взаимно ортогональными плоскостями, линия пересечения которых совпадает с направлением максимума излучаемой мощности. Обычно одну из этих ортогональных плоскостей совмещают с вектором **E**, а вторую — с вектором **H**.

На диаграмме направленности можно выделить направления максимальной мощности излучения (приема) и меньшего, побочного излучения. Отношение мощностей, излучаемых по главному и побочному направлениям, называется относительным уровнем боковых лепестков и измеряется в децибеллах (дБ). Для примера на рис. 4.2 изображена ДНА, имеющая сравнительно высокую направленность (узкий главный лепесток — глобальный максимум диаграммы — и относительно небольшие значения максимумов излучения по боковым лепесткам).

На практике часто используются антенны с резко выраженными направленными свойствами. При помощи таких антенн определяют направление на объекты, отражающие или излучающие электромагнитные волны (для использования в радиолокации, радионавигации, радиоэлектронной разведке). Направленные антенны увеличивают дальность действия радиоэлектронных устройств за счет концентрации излучаемой энергии в узком секторе пространства, повышают скрытность работы радиоэлектронных систем в условиях радиопротиводействия, уменьшают влияние умышленных помех в условиях ведения радиоэлектронной борьбы.

Как и ширина ДНА, коэффициент направленного действия (КНД) антенны является также числовой характеристикой степени концентрации излучаемой в пространство энергии. КНД

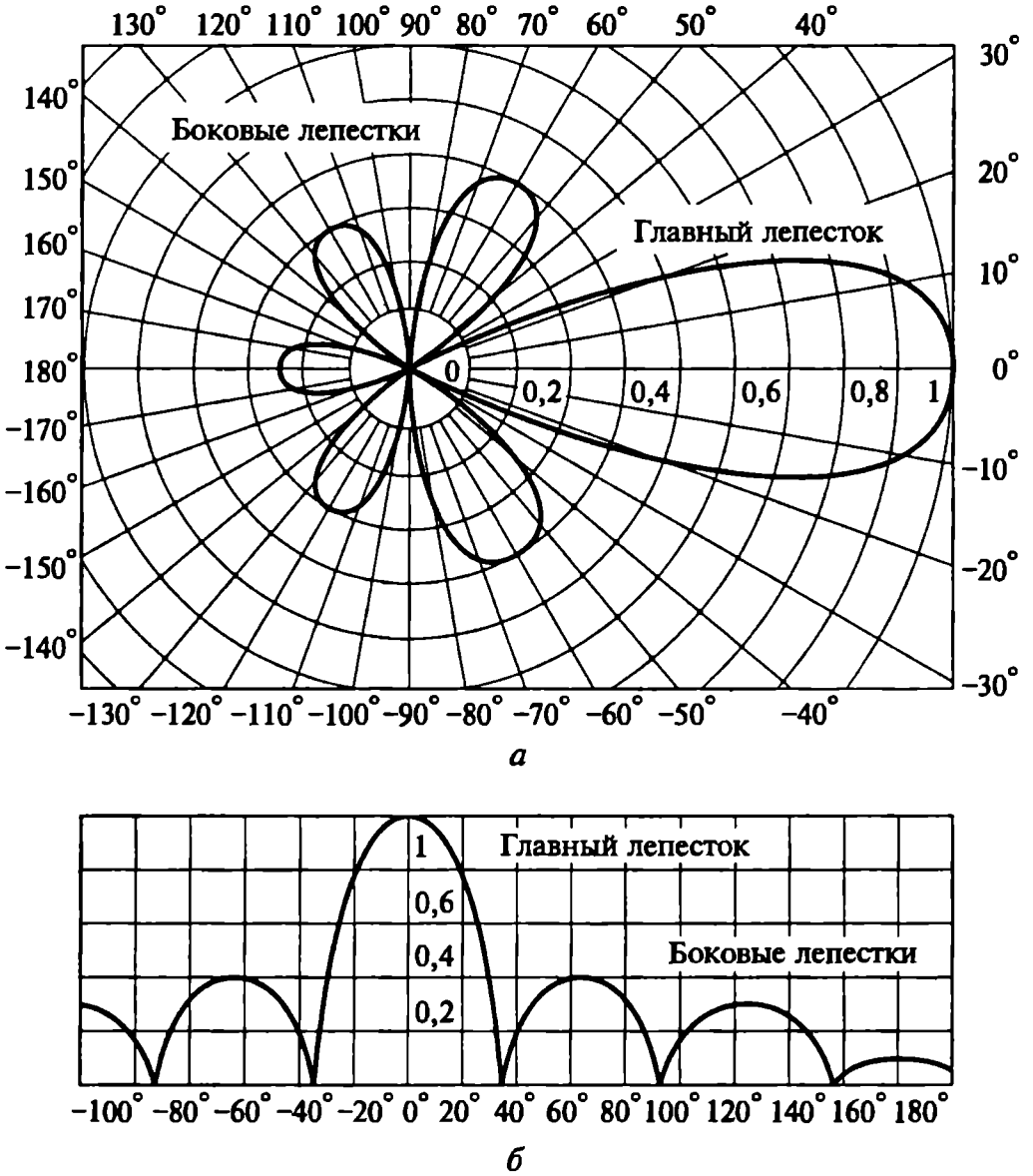


Рис. 4.2. Диаграмма направленности антенны:
а — в полярных координатах; *б* — в прямоугольных координатах

показывает, во сколько раз необходимо увеличить мощность излучения при переходе от направленной антенны к воображаемой ненаправленной (изотропной) антенне при условии, что к ним подводится одинаковая мощность и они имеют одинаковые КПД. Иногда применяют эквивалентное определение КНД как числа, показывающего, во сколько раз мощность излучения антенны, приходящаяся на единицу телесного угла в данном направлении, больше мощности излучения изотропной антенны, отнесенной к единице телесного угла, при равенстве полных мощностей, излучаемых обеими антеннами:

$$D(\theta,\varphi) = 4\pi \frac{P(\theta,\varphi)}{P_\Sigma}, \qquad (4.8)$$

где $P(\theta,\varphi)$ — мощность излучения, приходящаяся на единицу телесного угла в направлении, определяемом углами θ и φ; P_Σ — полная мощность излучения.

Как видно, мощность излучения антенны, приходящаяся на единицу телесного угла, зависит от направления излучения, т.е. от углов θ и φ; и график изменения КНД в пространстве от углов θ и φ отличается от ДНА по мощности лишь постоянным множителем. Максимальная величина КНД колеблется от нескольких единиц у слабонаправленных антенн станций связи до нескольких десятков и даже сотен тысяч у антенн с узкой ДНА (РЛС, космических комплексов и др.).

Две антенны, имеющие одинаковые ДНА и, следовательно, одинаковые КНД, при равной подводимой мощности могут создавать в одинаково расположенных относительно антенн точках приема различные напряженности поля в зависимости от величины потерь энергии в антенне. Для того чтобы учесть влияние потерь энергии в антенне на КНД антенны, вводят понятие коэффициента усиления антенны (КУ), который определяется как произведение КПД на КНД:

$$G = \eta_a D. \tag{4.9}$$

Приемная антенна подводит энергию СВЧ ко входу приемника.

Коэффициент пропорциональности между плотностью потока мощности электромагнитного поля около антенны и мощностью на ее выходе имеет размерность площади и называется эффективной поверхностью приемной антенны S_α:

$$S_\alpha = \frac{P_{\text{вых}}}{|\Pi|}. \tag{4.10}$$

У больших отражательных антенн эффективная поверхность почти равна (несколько меньше) площади геометрического размера раскрыва, нормального направлению на источник излучения. Но и антенны с очень маленькой геометрической поверхностью (проволочные антенны) имеют заметную эффективную поверхность. Для проволочных антенн различных типов чаще вводят такой параметр, как действующая длина (высота) $h_д$. Действующая высота равна отношению напряжения сигнала, создаваемого антенной на входе приемного устройства, к напряженности электрического поля этого сигнала около антенны. Действующая высота всегда меньше геометрической.

Прекрасное описание различных типов антенн, методов определения их параметров и характеристик, методов проектирования и применения можно найти в [1].

4.2.3. Непреднамеренное излучение электромагнитных полей

Каналы утечки информации образуются не только за счет перехвата сигналов, излучаемых антеннами в основных и боковых лепестках ДНА. Различные технические средства, работающие с высокочастотными токами, напряжениями и полями, а также различного рода электрические цепи, расположенные в непосредственной близости от таких средств, могут обладать антенным эффектом, т.е. непреднамеренно излучать электромагнитные волны. Сосредоточенные антенны образуются неплотными стыками электромагнитных экранов и линий, канализирующих энергию СВЧ, окнами, дверями, вентиляционными отверстиями и другими технологическими проемами в стенах экранированных помещений, отверстиями в металлических кожухах приборов. К распределенным случайным антеннам относятся различного рода кабели, провода систем сигнализации, радиотрансляционные сети, трубы, металлические конструкции и т.п. Кроме основного излучения сигнала на несущей частоте в главном лепестке диаграммы направленности антенны, работу радиоэлектронных систем и средств сопровождают побочные и непреднамеренные электромагнитные излучения (ПЭМИ), которые, тем не менее, переносят сигналы, информативные для технических средств разведки. ПЭМИ создаются также и системами, не рассчитанными на работу с излучением: вычислительными системами и различными техническими средствами обработки информации (ТСО).

Составляющие напряженности электрического (поперечная E_θ и продольная E_r), а также магнитного (H_φ) полей, создаваемых источником ПЭМИ, описываются уравнениями [25]:

$$E_\theta = Z_0 k \sin\theta\left[-\left(\frac{\lambda}{2\pi r}\right)^2 \cos\left(\frac{\lambda}{2\pi r} - \omega t\right) - \right.$$

$$\left. -\left(\frac{\lambda}{2\pi r}\right)\sin\left(\frac{\lambda}{2\pi r} - \omega t\right) + \cos\left(\frac{\lambda}{2\pi r} - \omega t\right)\right]; \qquad (4.11)$$

$$E_r = -2Z_0 k \cos\theta\left[-\left(\frac{\lambda}{2\pi r}\right)^2 \cos\left(\frac{\lambda}{2\pi r} - \omega t\right) + \right.$$

$$\left. +\left(\frac{\lambda}{2\pi r}\right) \sin\left(\frac{\lambda}{2\pi r} - \omega t\right)\right]; \qquad (4.12)$$

$$H_\varphi = k \sin\theta\left[-\left(\frac{\lambda}{2\pi r}\right)\cos\left(\frac{\lambda}{2\pi r} - \omega t\right) + \cos\left(\frac{\lambda}{2\pi r} - \omega t\right)\right], \qquad (4.13)$$

где Z_0 — волновое сопротивление свободного пространства; k — коэффициент пропорциональности, $k = \frac{Il}{2r\lambda}$ (I — ток в проводнике длиной l, создающем поле; λ — длина волны, соответствующая частоте $w = 2\pi f$); θ — аргумент радиуса-вектора r, отсчитанный от нормали к направлению тока; r — расстояние от проводника до точки, где определяются **E** и **H**; t — время.

Коэффициент k в (4.11), (4.12) и (4.13) содержит множитель $\frac{1}{r}$ и, следовательно, уменьшается по мере удаления от источника излучения.

При $r >> \frac{\lambda}{2\pi}$ (в дальней зоне) имеет значение только последнее слагаемое в (4.11) и (4.13), а волновое сопротивление $Z_0 = \frac{E_\theta}{H_\varphi} \approx 377$ Ом. Эта дальняя зона иначе называется зоной излучения или зоной плоской волны. При $r >> \frac{\lambda}{2\pi}$ (в ближней зоне), напротив, в (4.11) и (4.13) следует учитывать только первое слагаемое. Для такого случая оказывается, что $\frac{E_\theta}{H_\varphi} = \frac{Z_0\lambda}{2\pi r} >> Z_0$, что соответствует электрическому полю или полю высокого волнового сопротивления (относительно сопротивления излучения). Если излучатель эквивалентен не короткому проводнику (вибратору) с высоким сопротивлением, а витку (рамке) с низким сопротивлением, то в уравнении (4.11) можно пренебречь первым слагаемым. Тогда волновое сопротивление в ближней зоне оказывается $\frac{E_\theta}{H_\varphi} = \frac{Z_0 2\pi}{\lambda r}$. Этот случай соответствует магнитному полю или полю низкого волнового сопротивления (относительно сопротивления излучения). Условие $\frac{\lambda}{2\pi r} = 1$ определяет границу между дальней и ближней зонами. На рис. 4.3 показаны случаи формирования поля соответственно с высоким (см. рис. 4.3, *а*) и низким (см. рис. 4.3, *б*) импедансом (волновым сопротивлением).

Высокое волновое сопротивление характерно для поля вблизи прямого короткого проводника, по которому течет малый ток. Из-за высокого волнового сопротивления в структуре поля преобладает электрическая составляющая, которая уменьшается по мере удаления от излучателя как $\frac{1}{r^3}$, т. е. быстрее, чем магнитная, про-

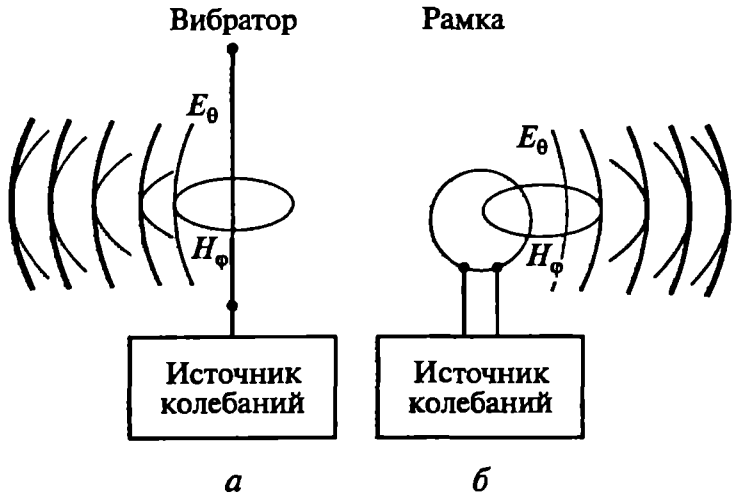


Рис. 4.3. Высоко- (*а*) и низкоимпедансные (*б*) излучатели

порциональная $\frac{1}{r}$. Соответственно этому волновое сопротивление уменьшается, асимптотически приближаясь к Z_0 дальней зоне. Рис. 4.3, *б* соответствует источнику с низкоимпедансным сопротивлением, в силу чего в структуре создаваемом им поля преобладает магнитная составляющая. Но это сопротивление растет по мере удаления от источника, асимптотически стремясь к $Z_0 = 377$ Ом. Изменения волнового сопротивления с расстоянием от источника иллюстрируются графиками (рис. 4.4).

Для передачи высокочастотных сигналов между информационными системами не всегда используется излучение электромагнитных полей антеннами в пространство. Очень часто используются специальные линии передачи (фидерные устройства). Фидерные устройства должны обеспечивать отсутствие излучения электромагнитной энергии при распространении вдоль линии, передачу с минимальными потерями и отвечать ряду специальных требований к конструкции.

Существует довольно обширный набор различных конструкций фидерных линий. Выбор того или иного типа линии зависит от ее назначения, диапазона частот и передаваемой по ней мощности. Прежде всего линии передачи могут быть открытыми или закрытыми.

Простейшим типом открытых линий является симметричная двухпроводная линия (рис. 4.5, *а*). Она слабо излучает электромагнитные волны при условии, что расстояние между проводами L много меньше длины волны λ. Однако уменьшение расстояния между проводами ограничивается передаваемой мощностью. Чем больше передаваемая мощность, тем выше напряжение между проводами. Максимально допустимое напряжение должно быть меньше пробивного, которое определяется расстоянием между проводами.

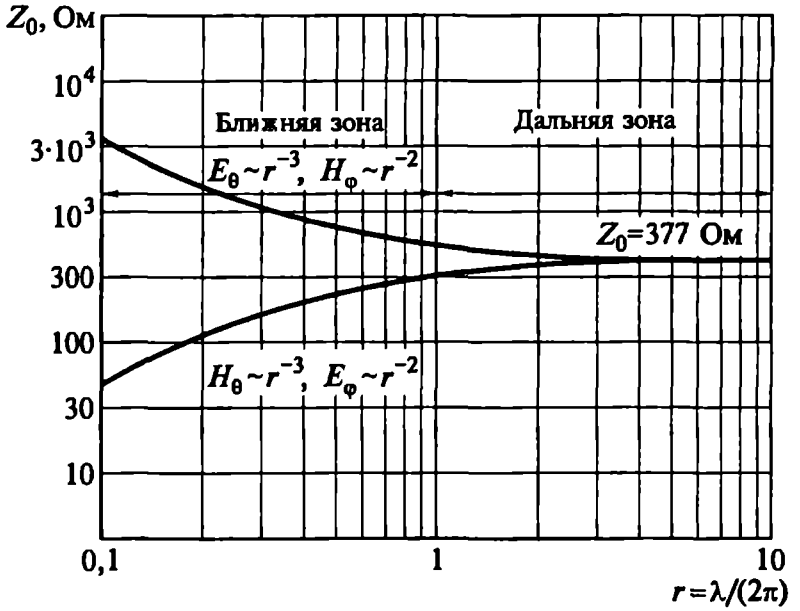


Рис. 4.4. Изменение волнового сопротивления с ростом расстояния до излучателя

Выбор диаметра проводов и расстояния между ними определяется волновым сопротивлением, которое для линий такого вида принимает значения в пределах 300...800 Ом.

В микроминиатюрном радиоэлектронном оборудовании находят широкое применение полосковые линии передачи электромагнитной энергии (рис. 4.5, *б*). Электромагнитное поле в несимметричной полосковой линии (НПЛ) сосредотачивается между разделенными слоем диэлектрика плоским проводником и проводящей подложкой. Волновое сопротивление НПЛ зависит от отношения ширины проводящей полоски к толщине слоя диэлектрика, а также от его диэлектрической проницаемости.

Очень простые и дешевые фидерные линии, обладающие тем не менее сравнительно хорошими эксплуатационными свойствами, получаются при использовании витых пар проводов (рис. 4.5, *в*).

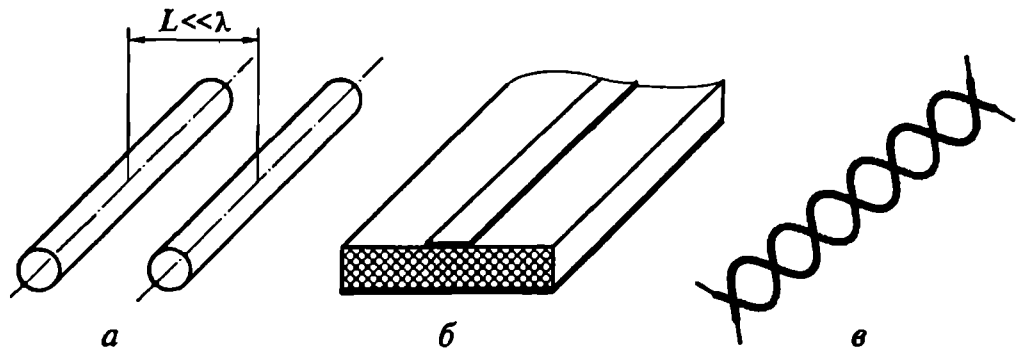
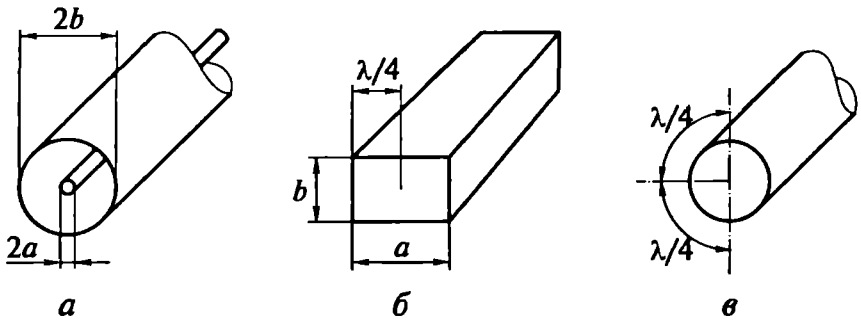


Рис. 4.5. Открытые фидерные линии:
а — двухпроводная линия; *б* — полосковая линия; *в* — витая пара



а *б* *в*

Рис. 4.6. Закрытые фидерные линии:
а — коаксиальный кабель; *б* — прямоугольный волновод; *в* — круглый волновод

Все открытые фидерные линии обладают большими потерями из-за сравнительно плохой экранировки поля. Поскольку потери обусловлены прежде всего излучением электромагнитного поля, применение открытых линий связано с риском утечки информации.

В закрытых линиях передачи электромагнитное поле полностью изолировано от окружающей среды, поэтому в нем теоретически исключаются потери на излучение. Наиболее распространенным среди закрытых линий является коаксиальный кабель (рис. 4.6, *а*). В нем один проводник, покрытый диэлектриком, помещен внутри другого, выполненного, как правило, в виде гибкой металлической оплетки. Наименьшие потери имеют коаксиальные кабели, внутренний провод которых покрыт чешуйчатыми керамическими изоляторами или диэлектрическими шайбами, расположенными на некотором расстоянии друг от друга. Входные колебания от источника подводятся к внешнему и внутреннему проводникам кабеля. Внешний провод кабеля может быть подключен к общей точке электрической схемы устройства (заземлен).

Длина электромагнитной волны в коаксиальном кабеле, заполненном диэлектриком с параметрами ε, μ, равна $\lambda_R = \dfrac{\lambda}{\sqrt{\varepsilon\mu}}$, т. е. не зависит от размеров коаксиальной конструкции, но не больше, чем λ — длина волны в свободном пространстве. Обычно волновое сопротивление коаксиальных кабелей лежит в пределах 30...150 Ом.

С увеличением частоты в двухпроводных линиях возрастают потери как на излучение, так и в изоляторах, в коаксиальном кабеле резко увеличиваются потери в диэлектрике. На волнах десятисантиметрового диапазона и короче потери так велики, что применение коаксиальных линий в ряде случаев становится нецелесообразным.

В сантиметровом и миллиметровом диапазонах волн широко применяются волноводы — полые металлические трубы прямоугольного (рис. 4.6, *б*) или круглого или (рис. 4.6, *в*) сечения.

По волноводу заданных размеров могут распространяться только волны короче определенной длины $\lambda_{в.кр}$, которую именуют критической. В прямоугольном волноводе обычно $\lambda_{в.кр} = 2a$. При приближении длины волны к критической фазовая скорость волны неограниченно возрастает, как и длина волны в волноводе.

Длина волны в волноводе — это расстояние, проходимое волной вдоль оси волновода с фазовой скоростью за время одного периода, т.е. $\lambda_в = v_ф T$. Так как фазовая скорость в волноводе

$$v_ф = \frac{c}{\sqrt{1-\left(\frac{\lambda}{\lambda_{в.кр}}\right)^2}} \tag{4.14}$$

больше скорости света, то и длина волны в волноводе больше длины волны в свободном пространстве для одной и той же частоты электромагнитных колебаний

$$\lambda_в = \frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{\lambda_{в.кр}}\right)^2}}. \tag{4.15}$$

Фазовая скорость характеризует структуру волны, но не определяет скорость переноса энергии. Возрастание фазовой скорости в волноводе при уменьшении частоты электромагнитных колебаний не сопровождается увеличением скорости распространения энергии, переносимой электромагнитной волной вдоль волновода. Наоборот, из-за многократных отражений от стенок скорость переноса энергии уменьшается и оказывается равной

$$u = c\sqrt{1-\left(\frac{\lambda}{\lambda_{в.кр}}\right)^2} \tag{4.16}$$

Иначе говоря, произведение фазовой скорости на скорость распространения энергии (4.16) (или групповую скорость) равна константе — квадрату скорости света в вакууме:

$$v_ф u = c^2. \tag{4.17}$$

Для передачи электромагнитных волн оптического диапазона используются специальные диэлектрические волноводы — световоды. Наиболее перспективный тип световода — гибкий волоконный световод с низкими оптическими потерями, позволяющий передавать свет на большие расстояния. Он представляет собой тонкую нить из оптически прозрачного материала, сердцевина

которой радиуса r_1 имеет показатель преломления n_1, а внешняя оболочка с радиусом r_2 имеет показатель преломления $n_2 < n_1$. Поэтому лучи, распространяющиеся под углами к оси световода, испытывают полное внутреннее отражение на поверхности раздела сердцевины и оболочки и распространяются только по сердцевине. В зависимости от назначения световода его диаметр $2r_1$ составляет от нескольких микрометров до нескольких сотен микрометров, а $2r_2$ — от нескольких десятков до нескольких тысяч микрометров. Величины $2r_1$ и $n_1 - n_2$ определяют число типов волн (мод), которые могут распространяться по световоду при заданной длине волны света. Выбирая $2r_1$ и $n_1 - n_2$ достаточно малыми, можно добиться, чтобы световод работал в одномодовом режиме.

Важнейшими характеристиками световодов являются оптические потери, обусловленные поглощением и рассеянием света в световоде, и информационная полоса пропускания. В последнее время созданы волоконные световоды с очень малыми потерями. Затухание сигнала в них имеет порядок 1 дБ/км в ближней ИК области спектра, а потери на излучение, собственно и образующие канал утечки информации, неизмеримо меньше. Материалом для этих световодов служит кварцевое стекло; различия показателей преломления сердцевины и оболочки достигают легированием стекла (бором, германием, фосфором). Минимально возможные потери в таких световодах составляют ~0,2 дБ/км на волне 1,55 мкм. Полоса пропускания типичных многомодовых волоконных световодов со ступенчатым профилем показателя преломления составляет 20...30 МГц.

Для целей интегральной оптики разработаны тонкопленочные диэлектрические волноводы — световоды, представляющие собой тонкую (порядка длины световой волны) однородную пленку, нанесенную на однородную подложку. Необходимое условие волноводного режима, т. е. существования поверхностных световых волн, заключается в том, что показатель преломления пленки больще показателей преломления подложки и среды над волноводом. Световая волна в таком световоде распространяется в процессе многократных полных отражений от ее стенок. Диэлектрические световоды изготавливаются методом катодного распыления стекла или другого материала на кварцевой подложке.

4.2.4. Собственное излучение электромагнитного поля

Для технических средств разведки информативно собственное излучение разведываемых объектов, если, конечно, это излучение доступно, т.е. если объекты имеют контраст с окружающим фоном.

Излучение объектов разведки может происходить по разным причинам, поэтому источники излучения подразделяются на три

основные группы: тепловые, люминесцентные, смешанные. Для технических средств разведки наибольшее значение имеет тепловое излучение.

Известно, что все тела, температура которых превышает 0 °К, излучают энергию в виде электромагнитных волн. Спектр излучения таких тел является непрерывным и довольно широким. Такие тела способны также поглощать падающее на них внешнее электромагнитное излучение. При определенных условиях может устанавливаться равновесие между излучением и поглощением.

Поверхность макросистемы (тела) характеризуется двумя параметрами: излучательной способностью $R(f, T)$ и поглощательной способностью $A(f, T)$. Величина $R(f, T)$ есть мера количества лучистой энергии, излучаемой за 1 с в единичном интервале частот и внутри единичного телесного угла единицей поверхности тела в направлении нормали к ней при абсолютной температуре T и частоте f. Следовательно, энергия, излучаемая за единицу времени внутри телесного угла $d\theta$ и в интервале частот от f до $f + df$ под углом θ к нормали, будет удовлетворять уравнению

$$\frac{dQ}{dt} = R(f,T)\cos\theta d\Omega df. \qquad (4.18)$$

Косинусоидальное изменение энергии излучения при изменении θ устанавливается законом Ламберта.

Поглощательная способность $A(f, T)$ равна отношению энергии излучения, поглощаемого единицей площади поверхности при определенных f и T, к энергии излучения, падающего на ту же площадь. Таким образом, $A(f, T)$ — безразмерная величина. Если $A(f, T) = 1$, то такое тело, поглощающее всю падающую на него энергию, называется абсолютно черным.

Если падающую на тело энергию обозначить через $K'(f, T)$, то в условиях термодинамического равновесия справедливо равенство

$$K(f, T) = R(f, T) + K(f, T)[1 - A(f, T)],$$

откуда

$$\frac{R(f,T)}{A(f,T)} = K(f,T). \qquad (4.19)$$

Это закон Кирхгофа, согласно которому отношение излучательной способности к поглощательной одинаково для всех веществ, т.е. чем больше энергии поглощает тело, тем больше оно ее и излучает. Для абсолютно черного тела из (4.19) получается $R(f, T) = K(f, T)$, поэтому $K(f, T)$ и есть излучательная способность абсолютно черного тела.

Закон Кирхгофа справедлив не только для интегральных величин, но и для спектральных. Его можно сформулировать еще так: отношение излучательной способности всякого тела к его поглощательной способности равно излучательной способности абсолютно черного тела.

Функция $K(f, T)$ связана со спектральной плотностью излучения $u(f, T)$ соотношением

$$u(f,T) = \frac{4\pi}{c} K(f,T), \tag{4.20}$$

причем спектральную плотность излучения можно определить теоретически на основе следующих рассуждений.

Абсолютно черное тело лучше всего имитируется замкнутым пространством. Стенки, ограничивающие объем пространства, излучают электромагнитные волны, но если установилось равновесие, то в пространстве могут существовать только стоячие волны. Число этих волн в интервале частот от f до $f + df$ оказывается равным

$$n_V = \frac{8\pi f^2 \mathrm{d}f}{c} V, \tag{4.21}$$

где V — объем пространства.

Каждой волне можно поставить в соответствие один гармонический осциллятор. Средняя энергия каждого осциллятора равна KT. Поэтому для классического осциллятора вся энергия составит

$$Q(f)\mathrm{d}n = n_V \langle Q \rangle = \frac{8\pi f^2 \mathrm{d}f}{c} VKT, \tag{4.22}$$

а спектральная плотность излучения будет

$$u(f,T) = \frac{Q(f)}{V} = \frac{8\pi f^2 KT}{c^3}. \tag{4.23}$$

Это закон Релея—Джинса. При низких частотах он согласуется с экспериментальными данными. С увеличением f функция $u(f, T)$, как видно из (4.23), должна неограниченно возрастать вплоть до бесконечности. Этот парадокс получил название ультрафиолетовой катастрофы. Он противоречит экспериментальным данным и здравому смыслу. Суть парадокса в том, что средняя энергия излучения классического осциллятора на всех частотах одинакова. Правильный результат получится, если в (4.23) использовать среднюю энергию не классического, а квантового осциллятора

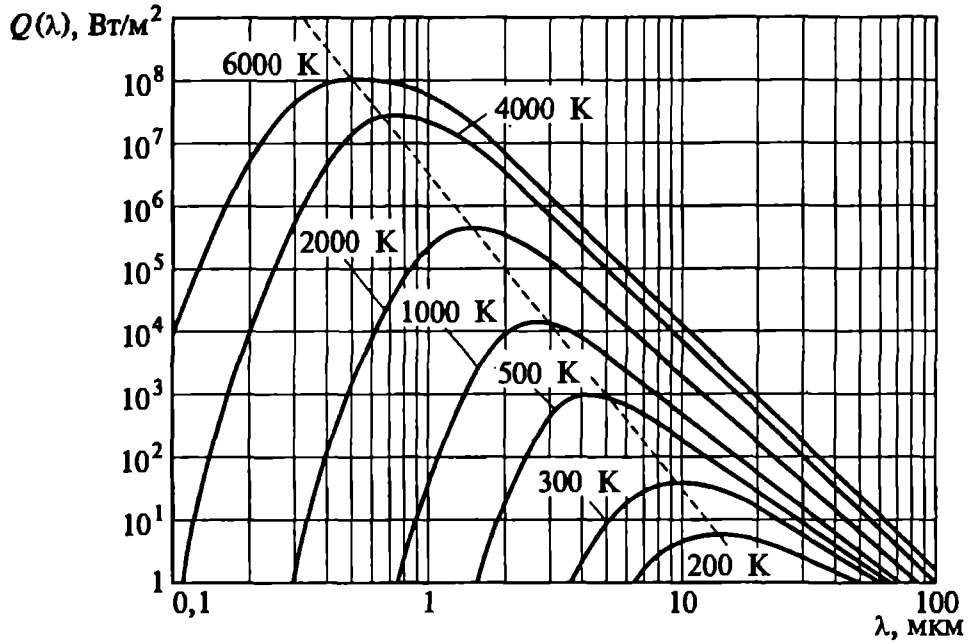


Рис. 4.7. Спектральная плотность излучения абсолютно черного тела (закон излучения Планка)

$$\langle Q(f)\rangle = \frac{hf}{\exp\left(\frac{hf}{KT}\right)-1}. \qquad (4.24)$$

Тогда из (4.22) с учетом (4.24) следует, что

$$u(f,T) = \frac{8\pi f^3 h}{c^3}\frac{1}{\exp\left(\frac{hf}{KT}\right)-1} = \frac{8\pi hc}{\lambda^5}\frac{1}{\exp\left(\frac{hc}{\lambda KT}\right)-1}. \qquad (4.25)$$

Это закон излучения Планка. Кривые, отображающие закон излучения Планка и построенные в соответствии с (4.25), приведены на рис. 4.7.

Если $hf \ll KT$, что возможно при достаточно низкой частоте f, то из (4.25) вытекает закон Релея—Джинса (4.23).

Из формулы закона излучения Планка (4.25) также следуют два других важных закона: Стефана—Больцмана, который получается простым интегрированием (4.25) по всем частотам

$$W = \sigma T^4, \qquad (4.26)$$

где W — полная мощность излучения всего спектра на единицу площади излучающей поверхности; σ — постоянная Стефана—Больцмана, $\sigma = 5{,}67 \cdot 10^{-8}$ Вт/(м$^2 \cdot$ К^4); и закон смещения Вина, который утверждает, что длина волны λ_{max}, на которую приходится максимум мощности в спектре излучения, обратно пропорциональна абсолютной температуре тела:

$$\lambda_{max} T = \text{const} = 2{,}891 \cdot 10^{-3}. \qquad (4.27)$$

4.2.5. Распространение электромагнитных волн

Знание условий распространения электромагнитного поля очень важно для определения опасных расстояний, на которых возможен несанкционированный доступ технических средств разведки к информации, содержащейся в перехватываемых сигналах. Если возможно, пространство, в пределах которого существует опасность перехвата, контролируется, чтобы исключить присутствие технических средств разведки. В иных случаях приходится принимать другие меры для защиты информации, переносимой электромагнитными полями, информативными для разведки.

Условия распространения электромагнитных полей существенно зависят от частоты (длины волны). Распространение радиоволн существенно отличается от распространения ИК излучения, видимого света и более жестких излучений.

Скорость распространения радиоволн в свободном пространстве (в вакууме) равна скорости света. Полная энергия, переносимая радиоволной, остается постоянной, а плотность потока энергии убывает с увеличением расстояния r от источника обратно пропорционально r^2. Распространение радиоволн в других средах происходит с фазовой скоростью, отличающейся от c, и сопровождается поглощением электромагнитной энергии. Оба эффекта объясняются возбуждением колебаний электронов и ионов среды пор действием электрического поля волны. Если напряженность поля $|\mathbf{E}|$ гармонической волны мала по сравнению с напряженностью поля, действующего на заряды в самой среде (например, на электрон в атоме), то колебания происходят также по гармоническому закону с частотой ω пришедшей волны. Колеблющиеся электроны излучают вторичные радиоволны той же частоты, но с другими амплитудами и фазами. В результате сложения вторичных волн с приходящей формируется результирующая волна с новой амплитудой и фазой. Сдвиг фаз между первичной и переизлученными волнами приводит к изменению фазовой скорости. Потери энергии при взаимодействии волны с атомами являются причиной поглощения радиоволн.

Амплитуда электрического (магнитного) поля волны убывает с расстоянием по закону

$$E(r) = \frac{E_0}{r} \exp\left(-\frac{\omega}{c} \kappa r\right), \tag{4.28}$$

где κ — показатель поглощения.

Фаза волны изменяется по закону

$$\Phi(t, r) = \omega t - \frac{\omega}{c} nr, \tag{4.29}$$

где n — показатель преломления.

Показатель преломления n зависит от диэлектрической проницаемости среды ε, ее проводимости σ и частоты волны ω:

$$n = \sqrt{\frac{\varepsilon}{2} + \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + \left(\frac{2\pi\sigma}{\omega}\right)^2}} \qquad (4.30)$$

Среда ведет себя как диэлектрик, если $\left(\frac{4\pi\sigma}{\omega\varepsilon}\right)^2 << 1$, и как проводник при $\left(\frac{4\pi\sigma}{\omega\varepsilon}\right)^2 >> 1$. В первом случае $n \approx \sqrt{\varepsilon}$ — поглощение мало, во втором $n \approx \kappa\sqrt{\frac{2\pi\sigma}{\omega}}$.

В среде с зависящими от частоты ε и σ наблюдается дисперсия волн. Вид частотной зависимости ε и σ определяется структурой среды. Дисперсия радиоволн особенно существенна в тех случаях, когда частота волны близка к характерным собственным частотам среды, например при распространении радиоволн в ионосферной и космической плазме.

При распространении радиоволн в средах, не содержащих свободных электронов (в тропосфере, в толще Земли), происходит смещение связанных электронов в атомах и молекулах среды в сторону, противоположную полю волны **E**, при этом $n > 1$, а фазовая скорость $v_ф < c$ (радиосигнал, несущий энергию, распространяется с групповой скоростью $v_{гр} < c$). В плазме поле волны вызывает смещение свободных электронов в направлении **E**, при этом $n < 1$ и $v_ф < c$.

В однородных средах радиоволны распространяются прямолинейно, подобно световым лучам. Процесс распространения радиоволн в этом случае подчиняется законам геометрической оптики. Учитывая сферичность Земли, дальность прямой видимости можно оценить на основе простых геометрических построений соотношением

$$R = 3{,}57\left(\sqrt{h_{прд}} + \sqrt{h_{прм}}\right), \qquad (4.31)$$

где R — дальность прямой видимости, км; $h_{прд}$ и $h_{прм}$ — высоты расположения передающей и приемной антенн, м.

Однако реальные среды неоднородны. В них n, а следовательно, и $v_ф$ различны в разных участках среды, что приводит к искривлению траектории радиоволны. Происходит рефракция (преломление) радиоволн. С учетом нормальной рефракции радиоволн максимальная дальность определяется более точным, чем (4.31), соотношением

$$R = 4{,}12\left(\sqrt{h_{\text{прд}}} + \sqrt{h_{\text{прм}}}\right). \qquad (4.32)$$

Если n зависит от одной координаты, например высоты h (плоскослоистая среда), то при прохождении волны через каждый плоский слой луч, падающий в неоднородную среду в точке с $n_0 = 1$ под углом φ_0, в пространстве искривляется так, что в произвольной точке среды h выполняется соотношение

$$n(h)\sin\varphi(h) = \sin\varphi_0. \qquad (4.33)$$

Если n убывает при увеличении h, то в результате рефракции луч по мере распространения отклоняется от вертикали и на некоторой высоте $h_{\max}$ становится параллельным горизонтальной плоскости, а затем распространяется вниз. Высота $h_{\max}$, на которую луч может углубиться в неоднородную плоскослоистую среду, зависит от угла падения φ_0. Этот угол можно определить из условия

$$n(h_{\max}) = \sin\varphi_0. \qquad (4.34)$$

В область $h > h_{\max}$ лучи не проникают и, согласно приближению геометрической оптики, волновое поле в этой области должно быть равно 0. В действительности вблизи плоскости $h = h_{\max}$ волновое поле возрастает, а при $h > h_{\max}$ убывает экспоненциально. Нарушение законов геометрической оптики при распространении радиоволн связано с дифракцией волн, вследствие которой радиоволны могут проникать в область геометрической тени. На границе области геометрической тени образуется сложное распределение волновых полей. Дифракция радиоволн возникает при наличии на их пути препятствий (непрозрачных или полупрозрачных тел). Дифракция особенно существенна в тех случаях, когда размеры препятствий сравнимы с длиной волны.

Если распространение радиоволн происходит вблизи резкой границы (в масштабе λ) между двумя средами с различными электрическими свойствами (например, атмосфера — поверхность Земли или тропосфера — нижняя граница ионосферы для достаточно длинных волн), то при падении радиоволн на резкую границу образуются отраженная и преломленная (прошедшая) радиоволны.

В неоднородных средах возможно волноводное распространение радиоволн, при котором происходит локализация потока энергии между некоторыми поверхностями (слоями), за счет чего волновые поля между ними убывают с расстоянием медленнее, чем в однородной среде. Так образуются атмосферные волноводы.

В среде, содержащей случайные локальные неоднородности, вторичные волны излучаются беспорядочно в различных направ-

лениях. Рассеянные волны частично уносят энергию исходной волны, что приводит к ее ослаблению. При рассеянии на неоднородностях размером $l \ll \lambda$ рассеянные волны распространяются почти изотропно. В случае рассеяния на крупномасштабных прозрачных неоднородностях рассеянные волны распространяются в направлениях, близких к направлению исходной волны. При $\lambda \approx l$ возникает сильное резонансное рассеяние.

Влияние поверхности Земли на распространение радиоволн. Оно зависит от относительного расположения передатчика и приемника. Распространение радиоволн — процесс, захватывающий большую область пространства, но наиболее существенную роль в распространении радиоволн играет область, ограниченная поверхностью, имеющей форму эллипсоида рассеяния, в фокусах которого на расстоянии r расположены передатчик и приемник.

Если высоты h_1 и h_2, на которых расположены антенны передатчика и приемника над поверхностью Земли, велики по сравнению с λ, то земная поверхность не влияет на распространение радиоволн. При понижении обеих или одной из конечных точек радиотрассы будет наблюдаться близкое к зеркальному отражение от поверхности Земли. При этом радиоволна в точке приема определяется интерференцией прямой и отраженной волн. Интерференционные максимумы и минимумы обусловливают лепестковую структуру поля в зоне приема. Особенно характерна такая картина для метровых и более коротких радиоволн. Качество радиосвязи в этом случае определяется проводимостью σ почвы.

Почвы, образующие поверхностный слой земной коры, а также воды морей и океанов обладают значительной электропроводностью. Но так как n и κ — функции частоты, то для сантиметровых волн все виды земной поверхности имеют свойства диэлектрика. Для метровых и более длинных волн Земля — проводник, в который волны проникают на глубину $d = \frac{1}{2\pi}\sqrt{\frac{c\lambda_0}{\sigma}}$ (где λ_0 — длина волны в вакууме). Поэтому где для подземной и подводной радиосвязи используются в основном длинные и сверхдлинные волны.

Выпуклость земной поверхности ограничивает расстояние, на котором из точки приема виден передатчик (область прямой видимости). Однако радиоволны могут проникать в область тени на большее расстояние $r_т \approx \sqrt[3]{R_З^2 \lambda}$ (где $R_З$ — радиус Земли), огибая Землю, в результате дифракции. Практически в эту область за счет дифракции могут проникать только километровые и более длинные волны. За горизонтом поле растет с увеличением высоты h_1,

на которую поднят излучатель, и быстро (почти экспоненциально) уменьшается при удалении от него.

Влияние рельефа земной поверхности на распространение радиоволн зависит от высоты неровностей h, их горизонтальной протяженности l, длины волны λ и угла θ падения волны на поверхность. Если неровности достаточно малы и пологи, так что $kh\cos\theta < 1$ (где k — волновое число, $k = \frac{2\pi}{\lambda}$) и выполняется критерий Рэлея $k^2 l^2 \cos\theta < 1$, то они слабо влияют на распространение радиоволн. Влияние неровностей зависит также от поляризации волн. Например, для горизонтально поляризованных волн оно меньше, чем для волн, поляризованных вертикально. Когда неровности не малы и не пологи, энергия радиоволны может рассеиваться (радиоволна от них отражается). Высокие горы и холмы с $h > \lambda$ образуют затененные области. Дифракция радиоволн на горных хребтах иногда приводит к усилению волны из-за интерференции прямых и отраженных волн: вершина горы служит естественным ретранслятором.

Фазовая скорость радиоволн, распространяющихся вдоль земной поверхности (земных волн) вблизи излучателя, зависит от ее электрических свойств. Однако на расстоянии в несколько λ от излучателя $v_ф \approx c$. Если радиоволны распространяются над электрически неоднородной поверхностью, например сначала над сушей, а затем над морем, то при пересечении береговой линии резко изменяются амплитуда и направление распространения радиоволн (наблюдается береговая рефракция).

Распространение радиоволн в тропосфере. Тропосфера — область, в которой температура воздуха обычно убывает с высотой h. Высота тропопаузы на земном шаре не одинакова: она больше над экватором, чем над полюсами, а в средних широтах, где существует система сильных западных ветров, меняется скачкообразно. Тропосфера состоит из смеси газов и паров воды; ее проводимость для радиоволн с λ больше нескольких сантиметров пренебрежимо мала. Тропосфера обладает свойствами, близкими к вакууму, так что у поверхности Земли коэффициент преломления $n = \sqrt{\varepsilon} = 1{,}0003$ и фазовая скорость лишь немного меньше c. С увеличением высоты плотность воздуха падает, поэтому n уменьшается, еще более приближаясь к 1. Это приводит к отклонению траекторий радиолучей к Земле. Такая нормальная тропосферная рефракция способствует распространение радиоволн за пределы прямой видимости, так как за счет рефракции волны могут огибать выпуклость Земли. Практически этот эффект может играть роль только для УКВ. Для более длинных волн преобладает отгибание выпуклости Земли за счет дифракции.

Метеорологические условия могут ослаблять или усиливать рефракцию по сравнению с нормальной, так как плотность воздуха

зависит от давления, температуры и влажности. Обычно в тропосфере давление газов и температура с высотой уменьшаются, а парциальное давление водяного пара увеличивается. Однако при некоторых метеорологических условиях (например, при движении нагретого над сушей воздуха над морем) температура воздуха с высотой увеличивается (температурная инверсия). Особенно велики отклонения летом на высоте 2...3 км, когда часто образуются температурные инверсии и облачные слои. При этом преломление радиоволн в тропосфере может стать столь сильным, что вышедшая под небольшим углом к горизонту радиоволна на некоторой высоте изменит направление и вернется обратно к Земле. В пространстве, ограниченном снизу земной поверхностью, а сверху рефрагирующим слоем тропосферы, волна может распространяться на очень большие расстояния (волноводное распространение). В тропосферных волноводах, как правило, могут распространяться волны с $\lambda < 1$ м.

Поглощение радиоволн в тропосфере пренебрежимо мало для всех радиоволн вплоть до сантиметрового диапазона. Поглощение сантиметровых и более коротких волн резко увеличивается, когда частота колебаний совпадает с одной из собственных частот колебаний молекул атмосферных газов (резонансное поглощение). Молекулы получают от приходящей волны энергию, которая превращается в теплоту и только частично передается вторичным волнам. Известен ряд линий резонансного поглощения в тропосфере: $\lambda = 1{,}35$; 1,5; 0,75 см (поглощение в парах воды) и $\lambda = 0{,}5$; 0,25 см (поглощение в кислороде). Между резонансными линиями лежат области более слабого поглощения (окна прозрачности).

Ослабление радиоволн может быть также вызвано рассеянием на неоднородностях, возникающих при турбулентном движении воздушных масс. Рассеяние резко увеличивается, когда в воздухе присутствуют капельные неоднородности в виде дождя, снега, тумана. Почти изотропное рассеяние Рэлея на мелкомасштабных неоднородностях позволяет радиоволнам распространяться на расстояния, значительно превышающих прямую видимость. Таким образом, тропосфера существенно влияет на распространение УКВ. Для декаметровых и более длинных волн тропосфера практически прозрачна, и на их распространение влияют земная поверхность и более высокие слои атмосферы (ионосфера).

Распространение радиоволн в ионосфере. Ионосферу образуют верхние слои земной атмосферы, в которой газы частично (до 1 %) ионизированы под влиянием ультрафиолетового, рентгеновского и корпускулярного солнечного излучения. Ионосфера электрически нейтральна, она содержит равное количество положительно и отрицательно заряженных частиц, т. е. является плазмой.

Достаточно большая ионизация, оказывающая влияние на распространение радиоволн, начинается на высоте 60 км (слой *D*),

увеличивается до высоты 300...400 км, образуя слои E, F_1, F_2, и затем медленно убывает. В главном максимуме концентрация электронов N достигает 10^2 м$^{-3}$. Зависимость N от высоты меняется со временем суток, года, с солнечной активностью, а также с широтой и долготой.

В зависимости от частоты основную роль в распространении радиоволн играют те или другие виды собственных колебаний. Поэтому электрические свойства различны для разных участков радиодиапазона. При высоких частотах ионы не успевают следовать за изменениями поля, и в распространении радиоволн принимают участие только электроны. Вынужденные колебания свободных электронов ионосферы происходят в противофазе с действующей силой и вызывают поляризацию плазмы в сторону, противоположную электрическому полю волны **E**. Поэтому диэлектрическая проницаемость ионосферы $\varepsilon < 1$. Она уменьшается с уменьшением частоты $\varepsilon = 1 - \dfrac{\omega_0^2}{\omega^2}$. Учет соударений электронов с атомами и ионами дает более точные формулы для диэлектрической проницаемости и проводимости ионосферы:

$$\varepsilon = 1 - \frac{\omega_0^2}{\omega^2 + \nu^2}; \quad \sigma = 1 - \frac{\omega_0^2 \nu}{4\pi\left(\omega^2 + \nu^2\right)}, \tag{4.35}$$

где ν — эффективная частота соударений.

Для декаметровых и более коротких волн в большей части ионосферы $\omega_2 \gg \nu_2$, а показатели преломления n и поглощения κ приближаются к значениям

$$n \approx \sqrt{1 - \frac{\omega_0^2}{\omega^2}}; \quad \kappa \approx \frac{2\pi\sigma}{\omega\sqrt{\varepsilon}}. \tag{4.36}$$

Поскольку для ионосферы $n > 1$, то фазовая скорость распространения радиоволн $v_\text{ф} = \dfrac{c}{n} > c$, а групповая скорость $v_\text{гр} = cn < c$.

Поглощение в ионосфере пропорционально ν, так как чем больше столкновений, тем бóльшая часть энергии, получаемой электроном, переходит в теплоту. Поэтому поглощение больше в нижних областях ионосферы (слой D), где выше плотность газа. С увеличением частоты поглощение уменьшается. Короткие волны испытывают слабое поглощение и могут распространяться на большие расстояния.

Рефракция радиоволн в ионосфере. В ионосфере могут распространяться только радиоволны с частотой $\omega > \omega_0$. При $\omega < \omega_0$ показатель преломления n становится чисто мнимым и электромаг-

нитное поле экспоненциально убывает в глубь плазмы. Радиоволна с частотой ω, падающая на ионосферу вертикально, отражается от уровня, на котором $\omega = \omega_0$ и $n = 0$. В нижней части ионосферы электронная концентрация и ω_0 увеличиваются с высотой, поэтому с увеличением ω излученная с Земли волна все глубже проникает в ионосферу. Максимальная частота радиоволны, которая отражается от слоя ионосферы при вертикальном падении, называется критической частотой слоя:

$$\omega_{кр} = \omega_{max} = \sqrt{\frac{4\pi e^2 N_{max}}{m}}. \qquad (4.37)$$

Критическая частота слоя F_2 (главного максимума) изменяется в течение суток и года в широких пределах (3...10 МГц). Для волн с $\omega > \omega_{кр}$ показатель преломления не обращается в ноль и падающая вертикально волна проходит через ионосферу, не отражаясь.

При наклонном падении волны на ионосферу происходит рефракция, как в тропосфере. В нижней части ионосферы фазовая скорость увеличивается с высотой (вместе с увеличением электронной концентрации N). Поэтому траектория луча отклоняется по направлению к Земле. Радиоволна, падающая на ионосферу под углом φ_0, поворачивает к Земле на высоте h, для которой выполнено условие $\omega = \omega_{кр}$. Максимальная частота волны, отражающейся от ионосферы при падении под углом φ_0, называется максимально применимой частотой $\omega_{max} = \dfrac{\omega_{кр}}{\sin\varphi_0} > \omega_{кр}$. Волны с $\omega < \omega_{max}$, отражаясь от ионосферы, возвращаются на Землю. Этот эффект используется для дальней радиосвязи и загоризонтной радиолокации. Вследствие сферичности Земли величина угла φ_0 ограничена и дальность связи при однократном отражении от ионосферы не превосходит 3500...4000 км. Связь на большие расстояния осуществляется за счет нескольких последовательных отражений от ионосферы и Земли (скачков). Возможны и более сложные волноводные траектории, возникающие за счет горизонтального градиента N или рассеяния на неоднородностях ионосферы при распространении радиоволн с частотой $\omega > \omega_{max}$. В результате рассеяния угол падения луча на слой F_2 оказывается больше, чем при обычном распространении. Луч испытывает ряд последовательных отражений от слоя F_2, пока не попадет в область с таким градиентом N, который вызовет отражение части энергии назад к Земле.

Влияние магнитного поля Земли с напряженностью $\mathbf{H}_0$. Оно сводится к тому, что на электрон, движущийся со скоростью v, действует сила Лоренца $F = \dfrac{e}{c} v \cdot \mathbf{H}_0$, под влиянием которой он вра-

щается по окружности в плоскости, перпендикулярной $\mathbf{H}_0$ с гироскопической частотой ω_H. Траектория каждой заряженной частицы — винтовая линия с осью вдоль $\mathbf{H}_0$. Действие силы Лоренца приводит к изменению характера вынужденных колебаний электронов под действием электрического поля волны, а следовательно, к изменению электрических свойств среды. В результате электрические свойства ионосферы становятся зависимыми от направления распространения радиоволн и описываются не скалярной величиной ε, а тензором диэлектрической проницаемости ε_{ij}. Падающая на такую среду волна испытывает двойное лучепреломление, т.е. расщепляется на две волны, отличающиеся скоростью и направлением распространения, поглощением и поляризацией. Если направление распространения радиоволн перпендикулярно $\mathbf{H}_0$, то падающую волну можно представить себе в виде суммы двух линейно поляризованных волн с $\mathbf{E} \perp \mathbf{H}_0$ и $\mathbf{E} \parallel \mathbf{H}_0$. Для первой «необыкновенной» волны характер вынужденного движения электронов под действием поля волны изменяется (появляется компонента ускорения, перпендикулярная $\mathbf{E}$) и поэтому изменяется n. Для второй «обыкновенной» волны вынужденное движение остается таким же, как и без поля $\mathbf{H}_0$.

Основная часть энергии низкочастотных (НЧ) и очень низкочастотных (ОНЧ) радиоволн практически не проникает в ионосферу. Волны отражаются от ее нижней границы (днем — вследствие сильной рефракции в *D*-слое, ночью — от *E*-слоя, как от границы двух сред с разными электрическими свойствами). Распространение этих волн хорошо описывается моделью, согласно которой однородные и изотропные Земля и ионосфера образуют приземный волновод с резкими сферическими стенками. В этом волноводе и происходит распространение радиоволн. Такая модель объясняет наблюдаемое убывание поля с расстоянием и возрастание амплитуды поля с высотой. Последнее связано со скольжением волн вдоль вогнутой поверхности волновода, приводящим к своеобразной фокусировке поля. Амплитуда радиоволн значительно возрастает в антиподной по отношению к источнику точке Земли. Это объясняется сложением радиоволн, огибающих Землю по всем направлениям и сходящихся на противоположной стороне.

Влияние магнитного поля Земли обусловливает ряд особенностей распространения НЧ волн в ионосфере: сверхдлинные волны могут выходить из приземного волновода за пределы ионосферы, распространяясь вдоль силовых линий геомагнитного поля между сопряженными точками Земли.

Нелинейные эффекты при распространение радиоволн в ионосфере. Такие эффекты проявляются уже для радиоволн сравнительно небольшой интенсивности и связаны с нарушением линейной зависимости поляризации среды от электрического поля

волны. Нагревная нелинейность играет основную роль, когда характерные размеры возмущенной электрическим полем области плазмы во много раз больше длины свободного пробега электронов. Поскольку длина свободного пробега электронов в плазме значительна, электрон успевает получить от поля заметную энергию за время одного пробега. Передача энергии от электрона к ионам, атомам и молекулам при столкновениях затруднена из-за большого различия в их массах. В результате электроны плазмы сильно разогреваются уже в сравнительно слабом электрическом поле, что изменяет эффективную частоту соударений. Поэтому ε и σ плазмы становятся зависящими от напряженности электрического поля **E** волны и распространение радиоволн приобретает нелинейный характер.

Нелинейные эффекты могут проявляться как самовоздействие волны и взаимодействие волн между собой. Самовоздействие мощной волны приводит к изменениям ее поглощения и глубины модуляции. Поглощение мощной радиоволны нелинейно зависит от ее амплитуды. Частота соударений ν с увеличением температуры электронов может как расти (в нижних слоях, где основную роль играют соударения с нейтральными частицами), так и убывать (при соударении с ионами). В первом случае поглощение резко возрастает с увеличением мощности волны (насыщение поля в плазме). Во втором случае поглощение падает (просветление плазмы для мощной радиоволны). Из-за нелинейного изменения поглощения амплитуда волны нелинейно зависит от амплитуды падающего поля, поэтому ее модуляция искажается (автомодуляция и демодуляция волны). Изменение ν в поле мощной волны приводит к искажению траектории луча. При распространении узконаправленных пучков радиоволн это может привести к самофокусировке пучка аналогично самофокусировке света и к образованию волноводного канала в плазме.

Взаимодействие волн в условиях нелинейности приводит к нарушению принципа суперпозиции. В частности, если мощная волна с частотой ω_1 модулирована по амплитуде, то благодаря изменению поглощения эта модуляция может передаться другой волне с частотой ω_2, проходящей в той же области ионосферы. Это явление кросс-модуляции может содействовать перехвату сообщений, переносимых сигналом частоты ω_1.

Распространение радиоволн в космических условиях. Оно имеет особенности за счет того, что из космического пространства к Земле приходит широкий спектр электромагнитных излучений, которые на пути должны пройти через ионосферу и тропосферу. Через атмосферу Земли без заметного затухания распространяются волны двух основных частотных диапазонов: радиоокно соответствует диапазону от ионосферной критической частоты до частот сильного поглощения аэрозолями и газами атмосферы

(10 МГц...20 ГГц); оптическое окно охватывает диапазон видимого и ИК излучения (1...10^3 ТГц). Атмосфера также частично прозрачна в диапазоне низких частот до 300 кГц, где распространяются свистящие атмосферики и магнитогидродинамические волны.

Распространение радиоволн разных диапазонов. Радиоволны очень низких (3...30 кГц) и низких (30...300 кГц) частот огибают земную поверхность вследствие полноводного распространения и дифракции, сравнительно слабо проникают в ионосферу и мало поглощаются ею. Отличаются высокой фазовой стабильностью и способностью равномерно покрывать большие площади, в том числе полярные районы. Это обусловливает возможность их использования для устойчивой дальней и сверхдальней радиосвязи и радионавигации, несмотря на высокий уровень атмосферных помех. Полоса частот 150... 300 кГц используется для радиовещания. Трудности применения этого частотного диапазона связаны с громоздкостью антенных систем, высоким уровнем атмосферных помех, относительной ограниченностью скорости передачи информации.

Средние волны (300...3000 кГц) днем распространяются вдоль поверхности Земли (земная или прямая волна). Отраженная от ионосферы волна практически отсутствует, так как волны сильно поглощаются в слое *D* ионосферы. Ночью из-за отсутствия солнечного излучения слой *D* исчезает, появляется ионосферная волна, отраженная от слоя *E*, и дальность приема возрастает. Сложение прямой и отраженной волн влечет за собой сильную изменчивость поля, поэтому ионосферная волна — источник помех для многих служб, использующих распространение земной волны.

Короткие волны (3...30 МГц) слабо поглощаются слоями *D* и *E* и отражаются от слоя *F*, когда их частоты $\omega < \omega_{max}$. В результате отражения от ионосферы возможна связь как на малых, так и на больших расстояниях при значительно меньшем уровне мощности передатчика и гораздо более простых антеннах, чем в низкочастотных диапазонах. Особенность радиосвязи в этом диапазоне — наличие замираний (федингa) сигнала из-за изменений условий отражения от ионосферы и интерференционных эффектов. Коротковолновые линии связи подвержены влиянию атмосферных помех. Ионосферные бури вызывают прерывание связи.

Для очень высоких частот и УКВ (30...1000 МГц) преобладают распространение радиоволн внутри тропосферы и проникновение сквозь ионосферу. Роль земной волны падает. Поля помех в низкочастотной части этого диапазона все еще могут определяться отражениями от ионосферы, и до частоты 60 МГц ионосферное рассеяние продолжает играть значительную роль. Все виды распространения радиоволн, за исключением тропосферного рассеяния, позволяют передавать сигналы с шириной полосы частот в несколько мегагерц.

Волны УВЧ и СВЧ (1000...10 000 МГц) распространяются в основном в пределах прямой видимости и их прием сопровождается низким уровнем шумов. В этом диапазоне при распространении радиоволн играют роль известные области максимального поглощения и частоты излучения химических элементов (например, линии водорода около 1420 МГц).

Волны СВЧ (свыше 10 ГГц) распространяются только в пределах прямой видимости. Потери в этом диапазоне несколько выше, чем на более низких частотах, причем на их величину сильно влияет количество осадков. Рост потерь на этих частотах частично компенсируется возрастанием эффективности антенных систем. Схема, иллюстрирующая особенности распространения радиоволн различных диапазонов, приведена на рис. 4.8.

Несмотря на то что исторически излучения оптического диапазона волн начали использоваться человечеством гораздо раньше, чем любые другие электромагнитные поля, распространение через атмосферу оптических волн гораздо менее изучено по сравнению с распространением любых волн радиодиапазона. Объясняется это более сложной картиной явлений распространения, а также и тем, что широкое изучение явлений, сопровождающих взаимодействие электромагнитных волн оптического диапазона с атмосферой, началось лишь в последнее время, после изобретения и начала широкого всестороннего применения оптических квантовых генераторов — лазеров.

Три основные явления обусловливают закономерности распространения оптических волн через атмосферу: поглощение, рассеяние и турбулентность. Первые два определяют среднее затухание электромагнитного поля при фиксированных атмосферных условиях и сравнительно медленные изменения поля (медленные замирания) при изменении метеорологических условий. Третье явление — турбулентность — вызывает быстрые изменения

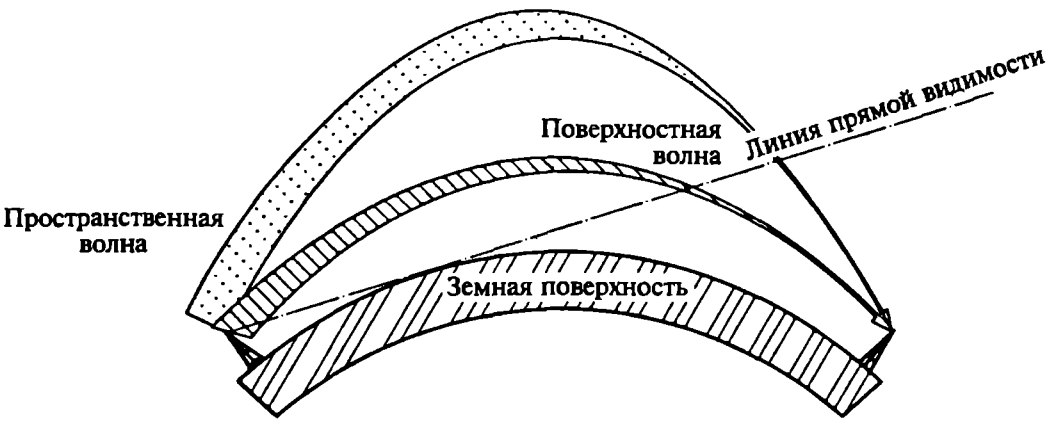


Рис. 4.8. Распространение электромагнитных волн в приземном пространстве

поля (быстрые замирания), наблюдающиеся при любой погоде. Кроме этого, из-за турбулентности наблюдается эффект многолучевости, когда структура пришедшего на прием луча может существенно измениться по сравнению со структурой луча на выходе передающего устройства.

4.3. Акустические поля

Акустические поля распространяются и переносят энергию (и информацию) в упругих средах. Если на каком-либо участке сплошной среды, например в слое воздуха или жидкости, возбудить простое гармоническое движение, то оно будет передаваться соседним участкам, от них в свою очередь к другим участкам и т.д. В результате возмущение от источника будет распространяться в среде с некоторой скоростью v. Результирующее движение будет бегущей волной. Так, плоская тонкая пластинка (мембрана) площадью S колеблется вправо и влево, совершая простое гармоническое колебание с амплитудой x_0 и частотой $\omega = 2\pi f$, возбуждает бегущую волну в окружающем воздухе. Пластинка передает энергию слою воздуха массой dm (рис. 4.9).

Максимальная кинетическая энергия этого слоя воздуха составляет

$$\frac{dmv^2}{2} = \frac{dm\omega^2 x_0^2}{2}; \tag{4.38}$$

$$dQ = \frac{1}{2}(\rho S dx)\omega^2 x_0^2, \tag{4.39}$$

где ρ — плотность воздуха, для нормальных атмосферных условий $\rho = 1{,}225$ кг/м^3.

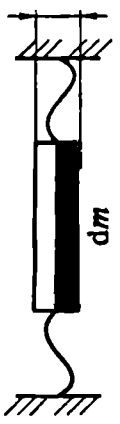


Рис. 4.9. Колебания мембраны и возникновение акустической волны

Поскольку при простом гармоническом движении средняя потенциальная энергия равна средней кинетической энергии, соотношение (4.39) описывает запас энергии в слое воздуха площадью S и толщиной dx. Если колебания начинаются в момент времени $t = 0$, то они распространяются в воздухе (вправо на рис. 4.9) со скоростью $v = dx/dt$, где dx — расстояние, на которое возмущение распространяется за время dt. Разделив (4.39) на dt, можно определить скорость передачи энергии каждому следующему слою толщиной dx:

$$\frac{dQ}{dt} = \frac{1}{2}\rho S \frac{dx}{dt}\omega^2 x_0^2. \tag{4.40}$$

Таким образом, мощность P, излучаемая колеблющейся пластинкой в положительном направлении оси x, можно представить в виде

$$P = \frac{1}{2}\rho S\omega^2 x_0^2 v. \tag{4.41}$$

Плотность потока мощности $П$, переносимой бегущей волной, определяется как мощность, приходящаяся на единицу площади. Из (4.41) плотность потока звуковой волны

$$П = \frac{1}{2}\rho\omega^2 x_0^2 v. \tag{4.42}$$

Скорость распространения звуковых волн в воздухе, как и вообще в газе, определяется соотношением

$$c = \sqrt{\frac{\gamma p_0}{\rho}} = \sqrt{\frac{\gamma RT}{\mu}}, \tag{4.43}$$

где $\gamma = \frac{C_p}{C_V}$ — адиабатическая постоянная, равная отношению теплоемкостей при постоянном объеме и давлении на уровне моря $\gamma = 1{,}41$; p_0 — статическое атмосферное давление; R — универсальная газовая постоянная; μ — молекулярная масса газа.

При уже упомянутых нормальных атмосферных условиях (когда температура $t° = +20°C$, атмосферное давление $p_0 = 10{,}1325 \cdot 10^2$ мбар) скорость звука в воздухе составляет $v = 343$ м/с.

При прохождении звуковой волны элементарные объемы среды совершают колебания около своего положения равновесия. Скорость этих колебаний зависит от звукового давления. В отличие от скорости распространения звука она называется колебательной скоростью v.

Поверхность, на которой расположены частицы, совершающие синфазные колебания, называется фронтом волны. В зависимости от формы этой поверхности различают плоские, цилиндрические и сферические волны. Направление распространения звука перпендикулярно фронту волны, поэтому распространение звука можно описывать с помощью звуковых лучей, которые во всех точках перпендикулярны фронту звуковых волн. Звуковое давление вдоль луча периодически меняется (для чистого тона — по синусоидальному закону).

Расстояние между двумя ближайшими фронтами волны с одинаковой фазой колебаний называется, как и для электромагнитного поля, длиной волны λ. Длина волны обратно пропорцио-

нальна частоте и существенно зависит от свойств звукопроводящей среды: ее плотности и упругости. Длина волны звукового колебания в газах меньше, чем в жидкостях, а в жидкостях, как правило, меньше, чем в твердых телах, для которых справедливо соотношение

$$\lambda = \frac{c}{f} = \frac{1}{f}\sqrt{\frac{E}{\rho}}, \qquad (4.44)$$

где E — модуль упругости (модуль Юнга).

Звуковые волны при распространении в свободной атмосфере благодаря теплопроводности и вязкости воздуха поглощаются тем сильнее, чем выше частота звука и меньше плотность атмосферы. Поэтому резкие вблизи звуки выстрелов или взрывов на больших расстояниях становятся глухими, а колокольный звон — более гулким. Неслышимые звуки очень низких частот (инфразвук) периодами от нескольких секунд до нескольких минут затухают мало, могут распространяться на тысячи километров и даже несколько раз огибать земной шар. Это дает возможность, например, обнаруживать ядерные взрывы, являющиеся мощным источником таких волн.

Температура и плотность атмосферы уменьшаются с увеличением высоты, но на больших высотах температура снова возрастает. На эти регулярные неоднородности накладываются зависящие от метеорологических условий изменения температуры и скорости ветра, а также их случайные турбулентные пульсации различных масштабов. Все перечисленные неоднородности сильно влияют на распространение звука: возникает искривление звукового луча — рефракция, в результате которой наклонный звуковой луч может вернуться к земной поверхности, образуя акустические зоны слышимости и зоны молчания; происходит рассеяние и ослабление звука на турбулентных неоднородностях, сильное поглощение звука на больших высотах и т. д.

Если атмосферные условия благоприятствуют фокусировке ударных волн, возникающих при движении сверхзвуковых реактивных самолетов, у земной поверхности звуковое давление может достичь значений, опасных для сооружений и здоровья людей. Полярные сияния, магнитные бури, землетрясения, ураганы, морские волнения являются источниками звуковых и особенно инфразвуковых волн.

Распространение звуковых волн в водной среде изучает гидроакустика. Особенность подводных звуков — их слабое затухание, вследствие чего под водой звук может распространяться на значительно большие расстояния, чем в воздухе. Так, в диапазоне частот 500...2000 Гц дальность распространения под водой звука средней интенсивности достигает 15...20 км, а в диапазоне ультразву-

ковых частот — 3...5 км. Звук мог бы распространяться и на значительно большие расстояния, однако в естественных условиях, кроме затухания, обусловленного вязкостью воды, ослабление звука происходит за счет рефракции и его рассеяния и поглощения различными неоднородностями среды. Рефракция звука вызывается неоднородностью свойств воды, главным образом по вертикали, вследствие изменения с глубиной гидростатического давления, солености и температуры в результате неодинакового прогрева массы воды солнечными лучами. В результате скорость распространения звука изменяется с глубиной, причем закон изменения зависит от времени года, времени дня, глубины водоема и ряда других причин (например, зимой дальность распространения звука больше, чем летом). Из-за рефракции образуются зоны тени, т.е. области, расположенные недалеко от источника, в которых интенсивность звука очень мала и слышимость отсутствует.

Рефракция может приводить не только к уменьшению, но и увеличению дальности распространения звука, обусловливая явление сверхдальнего распространения звука под водой. На некоторой глубине под поверхностью воды находится слой, в котором звук распространяется с наименьшей скоростью; выше скорость звука увеличивается из-за повышения температуры, а ниже — вследствие увеличения гидростатического давления с глубиной. Этот слой представляет собой своеобразный подводный звуковой канал. Луч, отклонившийся от оси канала вверх или вниз, вследствие рефракции возвращается в него обратно (рис. 4.10).

Если поместить источник и приемник звука в этом слое, то даже звук средней интенсивности (например, звуки взрыва небольших зарядов массой 1...2 кг) может быть зарегистрирован на расстояниях в сотни и тысячи километров.

На распространение звука высокой частоты, в частности ультразвука, у которого длины волн очень малы, оказывают влияние мелкие неоднородности. Такие неоднородности обычно имеются в естественных водоемах. Это микроорганизмы, пузырьки газов и т.д. Они поглощают и рассеивают энергию звуковых волн. В результате с повышением частоты звуковых колебаний дальность их

Рис. 4.10. Распространение звука в водной среде

распространения сокращается. Особенно сильно этот эффект заметен в поверхностном слое воды, где больше всего неоднородностей. Рассеяние звука неоднородностями, а также неровностями поверхности воды и дна вызывает явление подводной реверберации, которая может стать значительной помехой для ряда практического применений гидроакустики. Пределы дальности распространения подводного звука лимитируются также собственными шумами приемников и моря. Шум моря возникает от ударов волн на поверхности воды, морского прибоя, шума перекатываемой гальки, а также создается морской фауной.

Гидроакустика получила широкое практическое применение, в частности в технической разведке, так как никакие виды электромагнитных волн, включая и световые, не распространяются в воде (вследствие ее большой электропроводности) на сколько-нибудь значительные расстояния. Только звук может служить единственным возможным средством получения информации и средством связи под водой. Для этих целей пользуются как звуковыми частотами 300...16 000 Гц, так и ультразвуковыми от 16 кГц и выше. Наиболее широко в гидроакустической разведке применяются эхолоты и гидролокаторы, которыми пользуются для поисковых работ, обнаружения морских и подводных целей. Также используются шумопеленгаторы, определяющие направление на источник акустических колебаний звуковых и инфразвуковых частот.

Для средств акустической разведки информативен прежде всего речевой сигнал. Естественно, что спектральные и энергетические характеристики речевого сигнала весьма индивидуальны и нестационарны. При проектировании технических средств перехвата речевой информации в акустическом канале и при организации работ по защите этой информации используются стандартные усредненные характеристики.

Простейшими являются волны, в которых давление p изменяется по синусоидальному закону:

$$p(t) = p_0 + p_m \sin \omega t, \tag{4.45}$$

где p_0 — статическое давление среды (атмосферное давление); p_m — амплитуда переменной составляющей давления, которая называется звуковым давлением. Эффективное значение звукового давления в случае синусоидальных колебаний меньше амплитудного в $\sqrt{2}$ раз.

Среднее значение потока энергии за один период звукового колебания называется интенсивностью или силой звука J:

$$J = \frac{1}{T}\int_0^T \Pi dt = \frac{p^2}{\rho_c} = \rho_c v^2, \tag{4.46}$$

где p и v — эффективные значения звукового давления и колебательной скорости; ρ_c — волновое, или удельное акустическое сопротивление, $\rho_c = 413 \frac{\text{кгс}}{\text{м}^2}$.

Интенсивность слышимых звуков может меняться в очень широких пределах. Так, например, вблизи самолета с работающими двигателями звуковое давление шума достигает 20 Па и более. В то же время ухо способно различать шепот на расстоянии 0,5 м. При этом звуковое давление составляет всего $2 \cdot 10^{-4}$ Па. Для оценки интенсивности звука широкое применение получило понятие уровня, т.е. логарифмической меры относительной интенсивности:

$$L = 10 \lg \frac{J}{J_0} = 20 \lg \frac{p}{p_0}, \qquad (4.47)$$

где J_0 и p_0 — примерно соответствуют порогу слухового восприятия. Принято, что $J_0 = 10^{-12}$ Вт/м^2; $p_0 = 2 \cdot 10^{-5}$ Па.

Единицей уровня является децибел (дБ). Приращению уровня на 1 дБ соответствует увеличение звукового давления на 12 %, а интенсивности звука — на 26 %. Это приращение уровня находится на пределе различения слухом.

Исследование свойств слуха человека показало, что ощущение громкости зависит как от частоты, так и от интенсивности звука. Наиболее слабый слышимый звук называется порогом слышимости. Если увеличивать интенсивность звука, то при некотором значении наступает ощущение боли в ушах. Соответствующее значение уровня называется порогом болевого ощущения. Ухо способно воспринимать звуки, частота которых лежит в пределах 20...20 000 Гц, и сравнивать по громкости звуки различной частоты. Это позволяет построить так называемые кривые равной громкости (изофоны), приведенные на рис. 4.11. В тех же координатах на рис. 4.11 пунктирной кривой ограничена область, занятая звуками речи, т.е. теми акустическими колебаниями, которые способен создавать голосовой аппарат человека.

В зависимости от частоты звуки равной громкости имеют различный уровень L, поэтому для оценки субъективного ощущения введено понятие уровня громкости $L_г$. Под уровнем громкости понимают уровень звукового давления равногромкого звука частотой 1000 Гц. Для того чтобы отличить уровни громкости от уровней звукового давления, ввели новое наименование единиц уровня громкости — фон. Как видно из кривых (см. рис. 4.11), на низших частотах уровень громкости много ниже уровня звукового давления. Ослабление относительного уровня на низших частотах при общем снижении громкости ведет к искажению звучания.

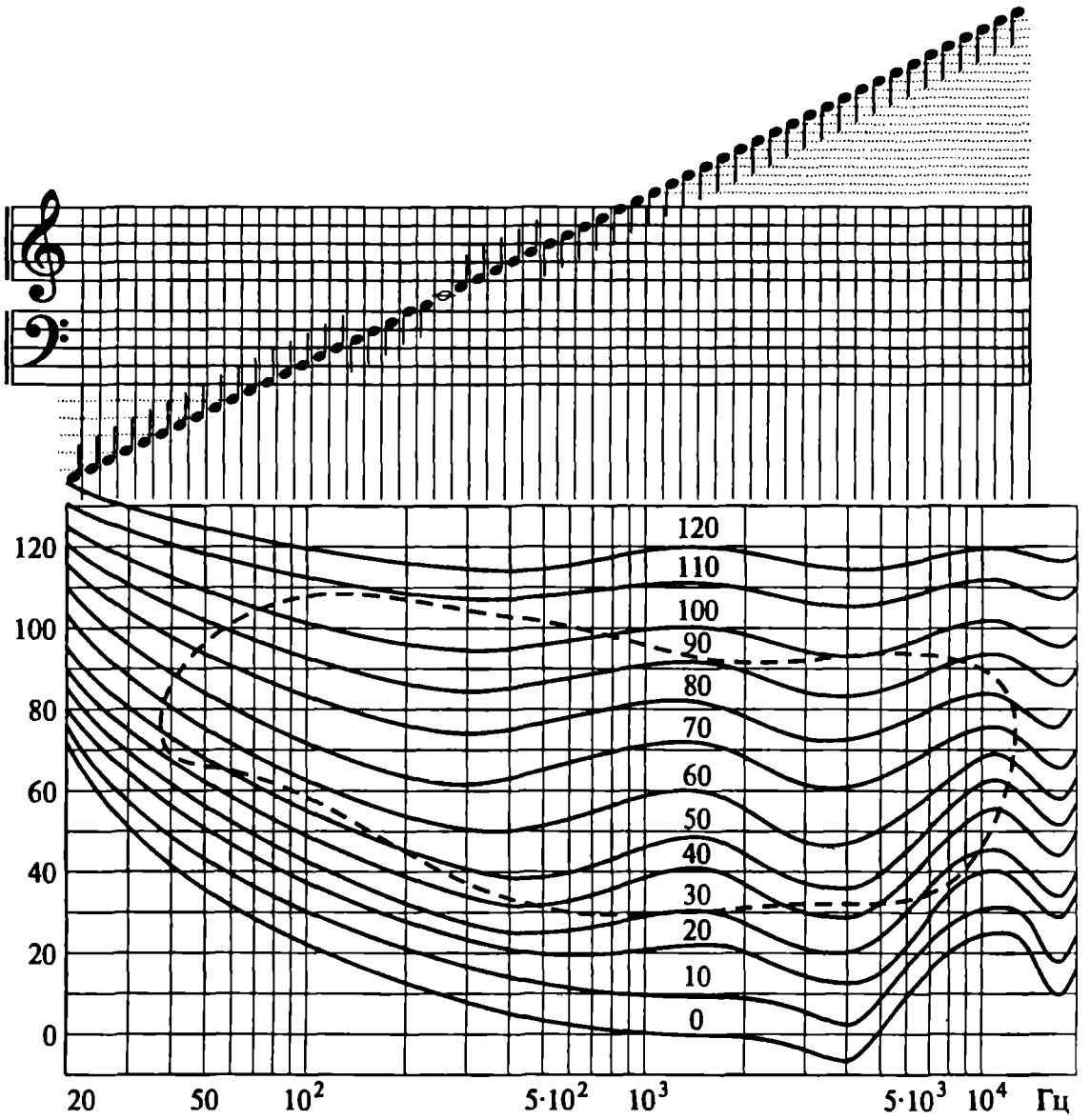


Рис. 4.11. Кривые равной громкости

Ухо человека воспринимает на слух колебания сложной формы как музыкальные звуки, имеющие определенную высоту. Чем больше основная частота звука, тем больше высота ощущаемого звука. Чувствительность уха к небольшим изменениям высоты очень велика. Она максимальна в диапазоне 500...4000 Гц, где человек способен различать разницу по частоте всего на 0,35 %. В то же время при определении высоты тона раздельно звучащих звуков возможны большие погрешности. На частотах выше 3000 Гц ощущение приращения высоты тона намного меньше фактического изменения частоты звука.

Изофоны (см. рис. 4.11) определены для чистых тонов в условиях практически полной тишины. Наличие мешающих звуков приводит к увеличению порога слышимости. Это явление называется маскировкой. Разность между порогом слышимости маскируемого

звука в присутствии мешающего сигнала и в тишине является мерой маскировки. Наибольшее маскирующее действие оказывают звуки, близкие по частоте к маскируемым. При маскировке чистым тоном существенно, лежит ли его частота выше или ниже частоты маскируемого тона. В последнем случае маскирующее действие значительно больше. При маскировке тонального сигнала шумами маскирующее действие оказывают только те составляющие шумов, частоты которых лежат в пределах соответствующей критической полосы слуха. Под критическими полосами слуха понимают интервалы частот, в пределах которых должны находиться составляющие сложного звука, чтобы их интенсивности суммировались.

Явление маскировки широко используется для активной защиты информации от утечки в акустическом канале.

Слуховое восприятие зависит от длительности воздействия звука. Для правильного восприятия высоты тонального сигнала нужно, чтобы его длительность была не менее 20...30 мс. При увеличении длительности воздействия такого сигнала растет ощущение громкости. При длительности воздействия 150...200 мс это ощущение становится максимальным. Дальнейшее увеличение длительности воздействия приводит к постепенному уменьшению ощущения громкости (адаптация).

Важным свойством слуха является бинауральный эффект. В зависимости от угла прихода звуковой волны сигналы, воздействующие на правое и левое ухо, могут в большей или меньшей степени отличаться как по фазе, так и по амплитуде. Слуховой анализатор позволяет человеку определять направление на источник звука. Наибольшая точность локализации получается на средних частотах. Если источник находится впереди слушателей, то точность локализации в горизонтальной плоскости достигает 2...4°. При изменении направления на 180° точность локализации резко падает. В области высших звуковых частот локализации помогает изменение спектра, вызываемого экранирующим действием головы. Поэтому способность локализации сохраняется несмотря на то, что сравнение по фазе становится невозможным.

4.4. Геофизические поля

4.4.1. Сейсмические поля и волны

Технические средства разведки эффективно работают с сейсмическими и гравитационными полями. Поэтому и в таких полях возникает проблема защиты информации.

Для сейсмической разведки информативны волны, распространяющиеся в земной коре. Принимая сигналы, переносимые этими волнами, можно обнаруживать, идентифицировать и пеленго-

вать источник сейсмических колебаний. Для иллюстрации и описания основных закономерностей формирования и распространения волн от сейсмического источника необходимы некоторые элементарные понятия о напряжениях и деформациях в земной коре [14].

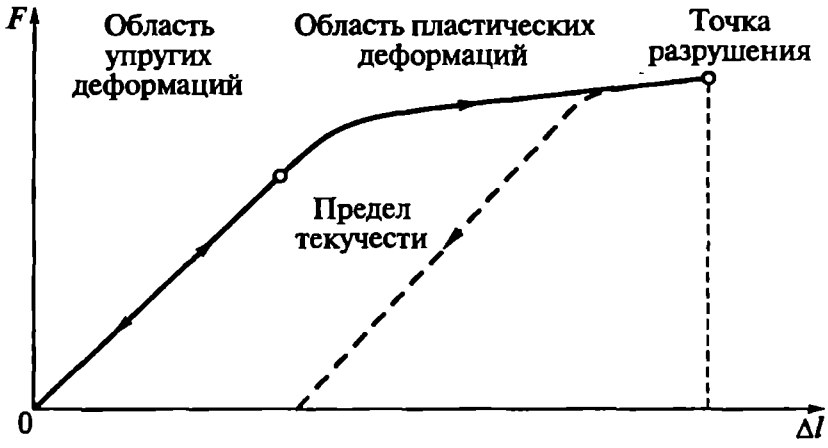
Если на тело действуют внешние силы, то внутри него устанавливается уравновешенная система внутренних сил. Напряжение представляет собой меру интенсивности, с которой действуют эти сбалансированные внутренние силы. Напряжение, действующее на некоторую площадку любой поверхности внутри тела, можно разложить на две компоненты: нормальную, направленную перпендикулярно этой площадке, и сдвиговую (тангенциальную), лежащую в плоскости площадки.

В любой точке находящегося в напряженном состоянии тела можно выделить три ортогональные плоскости, на которых напряжения полностью являются нормальными. Вдоль этих плоскостей не действуют сдвиговые напряжения. Пересечение этих плоскостей определяют три ортогональные оси, называемые главными осями напряжений, а нормальные напряжения, действующие в этих направлениях, называются главными напряжениями. Каждое главное напряжение отражает равновесие равных, но противоположно направленных компонент сил. Напряжение считается сжимающим, если силы направлены навстречу друг другу, и растягивающим, если они направлены в противоположные стороны.

Если внутри тела все главные напряжения равны по величине, то режим напряжений называется гидростатическим по аналогии с напряжениями в объеме жидкости, находящейся в покое. В гидростатическом поле напряжений сдвиговых напряжений не существует. Если главные напряжения не равны, сдвиговые напряжения действуют вдоль всех поверхностей внутри напряженного тела, за исключением трех ортогональных плоскостей, пересекающихся по главным осям.

Тело под действием напряжений испытывает изменение формы и (или) размеров, т.е. деформируется. Вплоть до некоторого предельного значения напряжения, называемого пределом текучести материала, величина деформации изменяется пропорционально приложенному напряжению (закон Гука). Упругая деформация обратима: снятие напряжения ведет к снятию деформации. Если напряжение превысит предел текучести, деформация оказывается нелинейной и становится частично необратимой: возникает остаточная (пластическая) деформация. Если напряжение возрастает еще больше, тело разрушается. Характерная кривая зависимости деформации от напряжения показана на рис. 4.12.

Конкретный вид линейной связи между напряжением и деформацией в упругой области определяется для любого вещества характерными для него модулями упругости, каждый из которых



ис. 4.12. Характерная кривая зависимости деформации от напряжения в твердом теле

редставляет собой отношение какого-либо вида напряжения к озникающей в результате его действия деформации. Для стерж-я, первоначальная длина которого l, а площадь поперечного се-ения S, под действием растягивающей силы F, приложенной к орцам, длина увеличится на Δl. Процесс удлинения характеризу-тся модулем Юнга E, определяемым отношением нормирован-ых значений продольного напряжения и продольной деформа-ии

$$E = \frac{F/S}{\Delta l/l}. \tag{4.48}$$

Растяжение такого стержня будет сопровождаться сокращени-м его поперечного размера, т.е. стержень будет испытывать не олько продольную, но и поперечную деформации. Отношение оперечной деформации к продольной называется коэффициен-ом Пуассона σ.

Объемный модуль упругости K выражает отношение напряже-ия к деформации в случае, когда к кубическому образцу веще-тва приложено простое гидростатическое давление P. Возникаю-ая при этом объемная деформация равна величине изменения бъема ΔV, нормированной к первоначальному объему V:

$$K = \frac{P}{\Delta V/V}. \tag{4.49}$$

Аналогичным образом модуль сдвига μ определяется как отно-ение сдвигового напряжения τ к соответствующей сдвиговой де-ормации $\operatorname{tg}\theta$:

$$\mu = \frac{\tau}{\operatorname{tg}\theta}. \tag{4.50}$$

Наконец, модуль продольной деформации ψ определяется как отношение продольного напряжения к продольной деформации при отсутствии поперечной деформации, т.е. когда материал может деформироваться только вдоль одной оси:

$$\psi = \frac{F/S}{\Delta l/l}. \tag{4.51}$$

Сейсмические волны — это импульсы энергии упругой деформации, распространяющиеся во все стороны от источника сейсмических колебаний. За исключением ближайших окрестностей источника, деформации, возникающие в среде при прохождении сейсмического импульса, невелики, и их можно считать упругими. При этом предположении скорости распространения сейсмических импульсов определяются модулями упругости и плотностями веществ, через которые они проходят. Существуют две группы сейсмических волн: объемные и поверхностные волны.

В объеме упругого твердого тела могут распространяться объемные волны двух типов. Продольные волны сжатия вызывают одноосные деформации в направлении распространения. Поэтому движение частицы, связанное с прохождением волны сжатия, — это колебание относительно некоторой фиксированной точки в направлении распространения волны. Поперечные волны сдвига при прохождении создают деформацию в направлении, перпендикулярном направлению распространения волны. Движения отдельных частиц среды в волнах сдвига представляют собой колебания около некоторой фиксированной точки в плоскости, перпендикулярной направлению распространения волны. Если все колебания частиц лежат в одной плоскости, то говорят, что поперечная волна плоскополяризована.

Скорость распространения объемной волны в любом веществе определяется параметрами среды распространения [14]:

$$v = \sqrt{\frac{M}{\rho}}, \tag{4.52}$$

где M — соответствующий модуль упругости вещества; ρ — плотность.

В соответствии с (4.52) скорость v_p продольной объемной волны, создающей одноосную деформацию сжатия, равна

$$v_p = \sqrt{\frac{\psi}{\rho}}, \tag{4.53}$$

или, поскольку $\psi = K + \frac{4}{3}\mu$,

$$v_p = \sqrt{\frac{K + 1,33\mu}{\rho}}. \qquad (4.54)$$

Скорость v_s поперечной объемной волны, создающей деформацию чистого сдвига, равна

$$v_s = \sqrt{\frac{\mu}{\rho}}. \qquad (4.55)$$

Из соотношений (4.52)...(4.55) можно видеть, что продольные волны всегда распространяются быстрее, чем поперечные. Отношение скоростей v_p/v_s в любом веществе определяется исключительно величиной коэффициента Пуассона σ этого вещества:

$$\frac{v_p}{v_s} = \sqrt{\frac{2(1-\sigma)}{1-2\sigma}}, \qquad (4.56)$$

а поскольку коэффициент Пуассона для однородных веществ обычно составляет $\sigma \approx 0,25$, то $v_p \approx 1,7v_s$.

Объемные волны не диспергируют, т.е. все частотные составляющие в волновом импульсе в любом веществе распространяются с одной и той же скоростью, определяемой только модулями упругости и плотностью вещества.

Подавляющее большинство наблюдений приемниками сейсмических разведок базируется на использовании одних лишь продольных волн, и в дальнейшем нашем рассмотрении все внимание будет сконцентрировано именно на этих волнах.

Сейсмический импульс распространяется от источника возмущения со скоростью, определяемой физическими свойствами окружающих пород, слагающих земную кору. Если среда однородная, то он будет иметь одну и ту же скорость во всех направлениях, так что в любой последующий момент времени волновой фронт, представляющий собой геометрическое место точек, которых достигла волновая энергия, будет сферическим. Сейсмические лучи в изотропной среде повсюду перпендикулярны волновым фронтам. Понятие лучей не несет какого-либо физического смысла, но очень помогает при рассмотрении путей перемещения сейсмической энергии в земных недрах.

Следует заметить, что скорость распространения сейсмической волны — это скорость, с которой в среде перемещается сейсмическая энергия. Это не то же самое, что скорость движения частиц среды, смещенных со своих мест в результате прохождения волны. Например, в случае продольных объемных волн их скорость распространения в горных породах обычно равна нескольким тысячам метров в секунду. Возникающие при их прохождении колеба-

тельные движения грунта характеризуются скоростями частиц, зависящими от амплитуды волны. Для слабых сейсмических волн, регистрируемых средствами сейсморазведки, скорости частиц могут составлять лишь 10^{-8} м/с. Им соответствуют смещения всего лишь около 10^{-10} м. Чтобы обнаружить сейсмические волны, приходится измерять эти весьма малые скорости частиц.

По мере распространения сейсмического импульса первоначальная энергия Q, излучаемая наружу от источника, распределяется по сферической оболочке увеличивающегося радиуса. Если радиус этой оболочки r, то энергия, падающая на единичную площадку оболочки, составит $Q/4\pi r^2$. Поэтому вдоль лучевой траектории энергия уменьшается как r^{-2}, а амплитуда волны, в однородной среде пропорциональная квадратному корню из энергии, уменьшается как r^{-1}.

Следующей причиной потерь энергии вдоль луча является то, что реакция земли на прохождение по ней сейсмических волн не является идеально упругой. Упругая энергия постепенно поглощается средой в результате потерь на внутреннее трение, что в конце концов приводит к полному исчезновению сейсмического возмущения. Коэффициент поглощения α определяет долю энергии на расстоянии, равном длине волны. У материалов, из которых обычно состоит поверхность Земли, значения α изменяются 0,25...0,75 дБ/λ.

Обычно предполагается, что в диапазоне частот, используемых в сейсморазведке, коэффициент поглощения не зависит от частоты. Если величина поглощения на единицу длины волны постоянна, то из этого следует, что волны более высоких частот затухают со временем или расстоянием быстрее, чем низкочастотные. Поэтому форма сейсмического импульса с широким частотным спектром в процессе распространения непрерывно изменяется вследствие постепенной потери более высоких частот. В целом эффект поглощения сводится к постепенному удлинению сейсмического импульса по мере распространения.

Источник сейсмических колебаний — это ограниченная по размерам область, внезапное выделение энергии в которой быстро приводит в напряженное состояние окружающую среду. Большинство источников сейсмических колебаний, как уже говорилось, генерирует энергию преимущественно в виде продольных волн, которые в основном и используются в сейсморазведке.

Существует множество разнообразных источников сейсмических колебаний, обладающих различными уровнями энергии и частотными спектрами излучения. В целом сейсмический источник содержит широкий диапазон частотных составляющих в пределах интервала от одного до нескольких сотен герц, хотя часто энергия сконцентрирована в некоторой более узкой полосе частот. Основные источники, информативные для средств сейсмо-

разведки, это различные взрывы. Но информативны также и микросейсмы (слабые источники, например сотрясение земной поверхности от прохождения техники, ударных волн, создаваемых различными летательными аппаратами, от работы промышленных энергетических установок).

В результате работы средств сейсморазведки формируется сейсмограмма — представленная в аналоговой или цифровой форме запись амплитуд смещений и колебаний почвы в функции времени. Для получения сейсмограмм движения грунта преобразуются в электрические сигналы. Эти сигналы усиливаются, фильтруются и регистрируются. В стандартной схеме наблюдений колебания почвы фиксируются в большом числе точек на земной поверхности. Для этой цели обычно применяются многоканальные системы регистрации, у которых число отдельных сейсморегистрирующих каналов иногда достигает нескольких сотен. Во всех системах сейсморегистрации, кроме самых простых, для облегчения последующей цифровой обработки данные записываются в цифровой форме, облегчающей ввод в ЭВМ для последующей вторичной обработки.

В качестве приемников в сейсморазведке применяются электромеханические преобразователи, которые преобразуют механический сигнал на входе (сейсмический импульс) в электрический сигнал на выходе. Иначе сейсмоприемники называются сейсмометрами, или геофонами. При измерениях в воде прохождение сейсмической волны сжатия сопровождается мгновенными изменениями давления, улавливаемыми гидрофонами, которые буксируют за кораблем либо подвешивают на буях в толще воды или (на самых мелких местах) помещают на морском дне. Гидрофоны также используются и при измерениях в условиях сильно насыщенных водой грунтов, которые встречаются на болотах. Приемниками могут служить отдельные геофоны или гидрофоны либо группы этих устройств, последовательное или параллельное соединение которых позволяет получить на выходе суммарный сигнал. Существует несколько разных типов сейсмоприемников, по принципу действия и устройству аналогичных микрофонам. Самый распространенный тип приемника, применяемый сейсмической разведкой на суше, использует электродинамический преобразователь (рис. 4.13). В мягких грунтах сейсмоприемник устанавливается с помощью штыря, а на твердых жестко крепится. Колебания почвы, вызванные прохождением сейсмической волны, передаются сейсмоприемнику.

В идеале форма выходного сигнала сейсмоприемника почти повторяет колебание почвы. Для сохранения формы сейсмического сигнала сейсмоприемники должны иметь в пределах рассматриваемого частотного диапазона плоскую амплитудную характеристику и минимальные фазовые искажения. Поэтому резонансная ча-

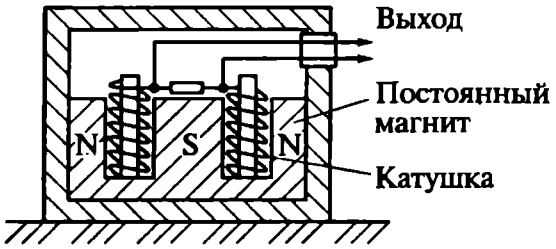


Рис. 4.13. Электродинамический сейсмоприемник

стота сейсмоприемников должна быть гораздо ниже основной частотной полосы сейсмического сигнала, который нужно зарегистрировать. Чаще всего используются сейсмоприемники с резонансной частотой в интервале 4... 15 Гц. Чувствительность сейсмоприемника, измеряемая в вольтах на его выходе, отнесенных к единице скорости движения катушки электродинамического преобразователя, определяется числом витков в катушке и напряженностью магнитного поля. Следовательно, для достижения большей чувствительности требуются приборы больших размеров и более массивной конструкции. Миниатюрные сейсмоприемники, применяемые в сейсморазведке, как правило, имеют чувствительность около 10 В/(м/с).

Электродинамические сейсмоприемники чувствительны только к той составляющей колебания почвы, которая совпадает с осью катушки. Продольные волны, распространяющиеся по вертикали от существующих в земле отражающих горизонтов, вызывают вертикальные колебания грунта, поэтому их легче всего обнаружить с помощью сейсмоприемников, содержащих вертикально стоящую катушку. Для регистрации тех сейсмических колебаний, которые вызывают в основном горизонтальные смещения грунта (как, например, горизонтально поляризованных поперечных волн), требуются сейсмоприемники с горизонтально смонтированной и способной перемещаться только по горизонтали катушкой. Сейсмоприемники обычно размещаются в виде линейных или площадных групп, в которых выходные сигналы нескольких сейсмоприемников суммируются. Такие группы позволяют образовать детекторы, обладающие свойством направленности и позволяющие пеленговать источник сейсмосигнала.

Сейсмические усилители предназначены для усиления сигналов в диапазоне частот от нескольких единиц до нескольких сотен герц (в некоторых системах морской сейсморазведки до нескольких килогерц) и для обеспечения записи сигналов с весьма широким динамическим диапазоном амплитуд.

Амплитуды колебаний почвы вблизи сейсмического источника могут значительно различаться: от первых вступлений интенсивных прямых и поверхностных волн до последующих вступлений очень слабых волн, вернувшихся на земную поверхность после отражения от глубинных горизонтов. Отношение амплитуд, равное 10^6, соответствует динамическому диапазону в 120 дБ; максимальный динамический диапазон сейсмоприемников примерно 140 дБ, а минимальный уровень собственных шумов сейсмических усили-

телей, составляющий около 1 мкВ, практически ограничивает максимальный динамический диапазон сейсморегистрации величиной 120 дБ.

4.4.2. Гравитационные поля

Гравиметрическая разведка исследует вариации гравитационного поля Земли, вызванные различиями в плотности поверхностных и подповерхностных слоев. Основой является понятие аномального тела, которое представляет собой объект, отличающийся по плотности от окружающих его земных пород. Эта область обладает аномальной массой и вызывает локальное искажение гравитационного поля.

Очень многие ситуации приводят к возникновению гравитационных аномалий заметной величины. Интерпретация гравитационных аномалий позволяет получать оценки глубины и формы аномального тела.

Возможность выполнять гравитационные съемки в морских условиях расширяет сферу применения этого метода, так что его можно использовать практически в любых районах мира.

Основой гравиразведки является закон тяготения Ньютона, согласно которому сила притяжения F между двумя массами m_1 и m_2, размеры которых малы по сравнению с расстоянием r между ними, определяется соотношением

$$F = \frac{Gm_1m_2}{r^2}, \tag{4.57}$$

где G — гравитационная постоянная ($G = 6{,}67 \cdot 10^{-11}$ м$^3 \cdot$ кг$^{-1} \cdot$ с$^{-2}$).

Гравитационное притяжение сферической невращающейся однородной Земли с массой M и радиусом r маленькой массы m на ее поверхности составляет

$$F = \frac{GM}{r^2} m = mg, \tag{4.58}$$

где g — ускорение свободного падения; mg — вес тела.

Эллипсоидальная форма Земли, ее вращение, неровности рельефа и неоднородное распределение масс в недрах приводят к тому, что сила тяжести на поверхности меняется.

Гравитационное поле наиболее удобно определять через гравитационный потенциал.

В то время как гравитационное ускорение является векторной величиной, имеющей как амплитуду, так и направление (вертикально вниз), гравитационный потенциал U является скаляром и имеет только амплитуду. Первая производная от U по некоторому направлению дает компоненту силы тяжести по этому направлению.

Среднее значение ускорения свободного падения на поверхности Земли составляет около 9,80 м/с^2. Вариации силы тяжести, вызываемые вариациями плотности подземных масс, обычно имеют порядок 100 мкм/с^2. Единицу 1 мкм/с^2 называют гравитационной единицей (ге). При гравиметрических съемках на суше легко достичь точности в ±0,1 ге, что соответствует приблизительно одной стомиллионной нормального гравитационного поля. На море достижимая точность измерений значительно меньше — примерно ±10 ге. Единицей силы тяжести в системе СГС является миллигал (1 мГал = 10^{-3} Гал = 10^{-3} см/с^2), эквивалентный 10 ге.

В первом поколении приборов гравитационной разведки для относительных измерений значения ускорения силы тяжести использовались небольшие маятники или наблюдались колебания крутильных весов, и хотя эти приборы были портативными, они требовали значительного времени для измерений. Современные приборы, способные измерять силу тяжести значительно быстрее, носят название гравиметров.

Гравиметры в своей основе являются пружинными весами, несущими грузик постоянной массы. Изменения веса этого грузика, вызванные вариациями силы тяжести, приводят к изменению длины пружины и являются мерой изменения силы тяжести. На рис. 4.14 пружина с начальной длиной l удлинилась на величину δl в результате возрастания δg, вызвавшей увеличение веса грузика с массой m. Растяжение пружины пропорционально растягивающей силе (закон Гука), т.е. $m\delta g = k\delta l$ и

$$\delta l = \frac{m}{k}\delta g, \qquad (4.59)$$

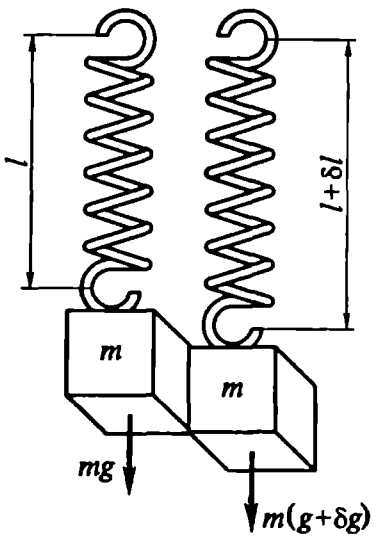


Рис. 4.14. Принцип гравиметрических измерений

где k — постоянная упругости пружины.

В приборах, пригодных для гравитационных измерений на суше, δl должно быть измерено с точностью 10^{-8}. Хотя большая масса и слабая пружина увеличили бы отношение m/k и, следовательно, чувствительность прибора, на практике это сделало бы систему неустойчивой. Следовательно, в реальных приборах требуется усилить удлинение пружины каким-либо оптическим, механическим или электронным способом.

Пружина в гравиметре должна выполнять двойную функцию, а именно: поддерживать грузик и служить измерительным устройством. Эту техническую проблему удалось решить в современных ас-

тазированных приборах, в которых используется дополнительная сила, действующая в том же направлении, что и растяжение (или сжатие) пружины, и, следовательно, усиливающая смещение грузика.

Пример астазированного гравиметра изображен на рис. 4.15. Этот прибор состоит из укрепленного на опоре рычага, несущего грузик и поддерживаемого пружиной, прикрепленной как раз над его точкой опоры. Величина момента силы, действующей со стороны пружины на рычаг, зависит от растяжения пружины и синуса угла θ. При увеличении силы тяжести рычаг опускается и пружина растягивается сильнее. Хотя возвращающая сила пружины возрастает, величина угла θ уменьшается до θ′. Выбирая соответствующим образом взаимное расположение пружины и рычага, можно достичь такого положения, при котором увеличение возвращающего момента с увеличением силы тяжести будет сколь угодно малым. С обычными пружинами рабочий диапазон такого прибора был бы очень мал. Однако применяя пружину нулевой начальной длины (такая пружина предварительно растягивается в процессе ее изготовления так, чтобы возвращающая сила была пропорциональна физической длине пружины, а не ее удлинению), создают приборы с очень высокой чувствительностью в широком диапазоне измеряемых значений силы тяжести. Показания прибора снимают, возвращая рычаг в горизонтальное положение, для чего с помощью микрометрического винта смещают по вертикали точку прикрепления пружины. Температурные эффекты исключаются за счет термостатирования системы. Диапазон величин, измеряемых прибором, составляет 50 000 ге.

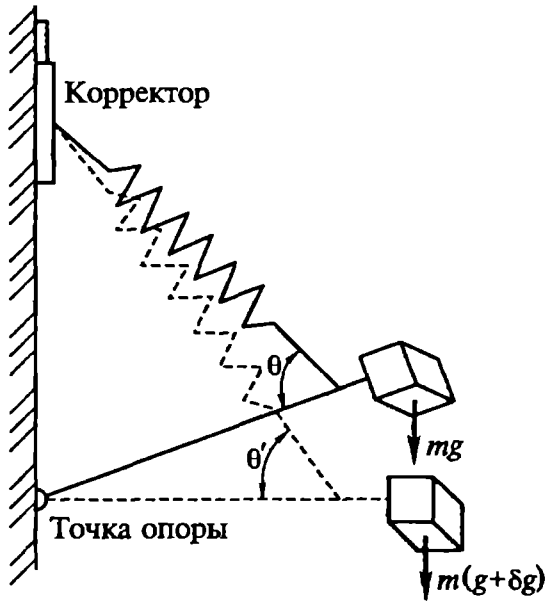


Рис. 4.15. Астазированный гравиметр

Недостатком гравиметров является дрейф — смещение нуля гравиметра, при котором показания прибора постепенно изменяются со временем при измерениях в одной и той же точке. Причиной дрейфа является неидеальная упругость пружин, которые с течением времени испытывают пластические деформации (ползучесть). Дрейф также может являться результатом температурных вариаций, которые, если их не скомпенсировать тем или иным образом, вызывают растяжение или сжатие измерительной системы и в результате порождают вариации измеряемых значений, не

связанные с изменениями силы тяжести. Для исключения влияния дрейфа проводятся повторные измерения в одной и той же точке.

На море силу тяжести можно измерять в отдельных точках с использованием дистанционно управляемого наземного гравиметра, который помещают в водонепроницаемый контейнер и опускают с борта судна на дно. Таким образом могут быть получены результаты, сравнимые по качеству с наземными. Этот метод с успехом используется при относительно небольшой глубине моря. Недостаток его заключается в том, что при каждом измерении прибор необходимо опускать на морское дно, поэтому скорость съемки оказывается очень низкой. Кроме того, при сильных приливных течениях судно, ведущее съемку, нужно ставить на якорь для удержания его над точкой измерения, пока гравиметр находится на дне.

Гравитационные измерения на море могут выполняться непрерывно, если использовать гравиметр, модифицированный для бортового применения. Точность измерений бортовым гравиметром меньше, чем при измерениях на суше вследствие действия значительных горизонтальных и вертикальных ускорений негравитационной природы, действующих на прибор под влиянием морских волн и движения судна. Эти негравитационные ускорения могут вызывать вариации в измеряемом значении силы тяжести вплоть до 10^6 ге. Воздействия горизонтальных ускорений, вызываемых волнами, смещающими судно из стороны в сторону и изменяющими его скорость и курс, можно в значительной степени исключить, если поместить прибор на гиростабилизированную горизонтальную платформу. Стабилизированное измерительное устройство будет реагировать только на вертикальные ускорения. Отклонения платформы от горизонтального положения создают ошибки наклона, которые обычно меньше 10 ге. Внешние вертикальные ускорения, вызываемые движениями волн, не отличимы от силы тяжести, но их влияние можно уменьшить за счет сильного демпфирования подвесной системы и усреднения отсчетов за интервал времени, значительно больший, чем максимальный период волнового движения (около 8 с). При качке судна в вертикальной плоскости, когда оно занимает положение выше и ниже среднего уровня моря, волновые ускорения имеют равные положительные и отрицательные значения и с успехом исключаются осреднением за несколько минут (за несколько периодов качки).

В бортовых гравиметрах с чувствительным элементом на рычаге дополнительная сложность возникает из-за горизонтальных ускорений. Рычаг измерительного устройства осциллирует под действием различных вертикальных ускорений, вызванных движениями судна. Когда рычаг отклоняется от горизонтали, его начинают дальше смещать силы вращения, вызываемые любым горизон-

тальным ускорением. При определенных фазовых соотношениях между горизонтальными и вертикальными компонентами движения судна горизонтальные ускорения могут вызывать такие смещения рычага, которые уже не усредняются во времени.

Примером более сложного возмущающего движения гравиметра является вращение, когда измерительная система под влиянием морских волн описывает в пространстве окружность (рис. 4.16).

В момент времени t_1 (см. рис. 4.16) судно отклоняется вниз, смещая рычаг вверх и вправо и создавая поворот его против часовой стрелки, уменьшающий вертикальное перемещение рычага. Чуть позже, в момент времени t_3, судно отклоняется вверх, смещая рычаг вниз и влево, вызывая поворот его вновь против часовой стрелки, который увеличивает смещение рычага вниз. В таком случае суммарный эффект горизонтальных ускорений должен вызывать систематическую ошибку в положении рычага. Этот эффект известен как эффект кросс-каплинга. Его амплитуда зависит от характеристик демпфирования измерительной системы и амплитудно-фазовых соотношений между горизонтальными и вертикальными движениями. Они приводят к ошибкам в измеряемых значениях силы тяжести, называемым ошибками кросс-каплинга. Эта ошибка мала или незначительна при хороших погодных условиях, но может стать очень большой при сильном волнении. Ошибки кросс-каплинга исправляются с помощью сигналов от двух горизонтальных акселерометров, устанавливаемых на стабилизированной платформе.

Невозможность полностью скомпенсировать посторонние ускорения уменьшает точность набортных измерений в лучшем случае до 10 ге, а действительная величина точности зависит от преобладающего состояния моря.

Измерения силы тяжести с борта самолета в настоящее время не являются удовлетворительными вследствие чрезмерно больших ошибок в применяемых поправках. Поправки могут достигать 16 000 ге при скорости порядка 400 км/ч, а ошибка на 1 % в определении скорости или курса вызывает максимальные ошибки соответственно в 180 и 250 ге. Вертикальные ускорения, связанные с движением самолета, периоды которых превышают время усреднения измерений, исправить нелегко. Подобные погрешности в определенной степени можно исправить использованием автопилотов и автоматических стабилизаторов высоты, но в настоя-

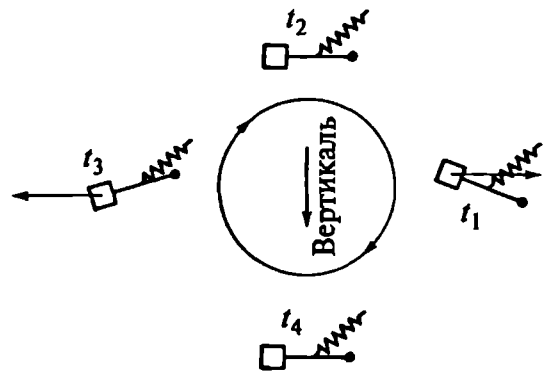


Рис. 4.16. Эффект кросс-каплинга в бортовом гравиметре

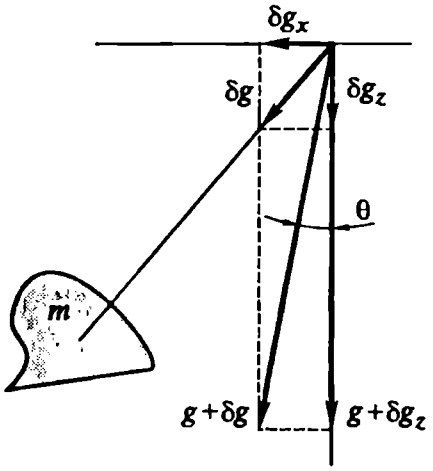


Рис. 4.17. Связь измеряемой напряженности гравитационного поля и компонент аномалии силы тяжести

щее время точность при использовании таких систем достигает только 100 ге.

Цена деления гравиметров может со временем изменяться и должна периодически проверяться. Наиболее распространенной процедурой поверки является снятие показаний на двух или более опорных точках, где абсолютные или относительные значения силы тяжести известны.

Гравиметры эффективно реагируют только на вертикальную составляющую гравитационного притяжения аномальной массы. Гравитационное влияние аномальной массы δg с горизонтальной составляющей δg_x и вертикальной δg_z, на локальное гравитационное поле g иллюстрируется векторной диаграммой (рис. 4.17).

Из геометрических построений (см. рис. 4.17) следует, что в пренебрежении составляющими порядка $0(\delta g^2)$ вертикальная составляющая силы тяжести

$$g + \delta g = \sqrt{(g + \delta g_z)^2 + \delta g_x^2} \approx g + \delta g_z, \tag{4.60}$$

откуда

$$\delta g \approx \delta g_z. \tag{4.61}$$

Следовательно, измеренные возмущения силы тяжести в основном соответствуют вертикальной компоненте притяжения аномального тела. Локальное отклонение от вертикали θ определяется соотношением

$$\theta = \operatorname{arctg}\frac{\delta g_x}{g}, \tag{4.62}$$

а поскольку $\delta g_x \ll g$, то θ обычно незначительно.

Перед тем как результаты гравитационной разведки можно будет интерпретировать, из них необходимо удалить все вариации гравитационного поля Земли, не являющиеся информативными для разведки, поскольку не зависят от различий в плотности земных покровов.

Информативные для технической разведки гравитационные аномалии возникают вследствие различий в плотности, или перепада плотности, между объектом разведки и окружающими его породами. Для тела с плотностью ρ_1, находящегося в среде с плотнос-

тью ρ_2, перепад плотности $\Delta = \rho_1 - \rho_2$. Знак перепада плотности определяет и знак гравитационной аномалии.

Плотности геологических пород относятся к наименее изменчивым из всех геофизических параметров.

Современная технология позволяет производить гравиметры, способные измерять изменения силы тяжести вплоть до 1 мкГал = $= 10^{-8}$ м/с^2. Метод микрогравиметрии — основной метод гравитационной разведки. Именно этот метод применяется для поиска подземных пустот, для обнаружения искусственных подземных сооружений. Другим важным достижением последнего времени в гравиметрической разведке является создание портативного прибора, способного измерять абсолютные значения силы тяжести с высокой точностью.

Контрольные вопросы

1. Перечислите физические поля, используемые техническими средствами разведки.
2. От каких свойств среды зависит скорость распространения электромагнитного поля? Когда эта скорость максимальна? Чему равна максимальная скорость распространения электромагнитного поля?
3. Какая величина называется волновым сопротивлением пространства?
4. Как деформируется диаграмма направленности при увеличении коэффициента усиления антенны?
5. Какая разница в распространении пространственных и поверхностных электромагнитных волн?
6. От каких параметров среды зависит скорость распространения акустической волны?
7. В чем состоит свойство дисперсии волн?
8. Как и в каких единицах измеряется интенсивность звука? Как интенсивность звука связана с мощностью акустических колебаний?
9. Какие цели ставит и какие задачи решает сейсмическая разведка?
10. Чем отличаются цели сейсмической и гравитационной разведок?

Глава 5
ЗАЩИТА ИНФОРМАЦИИ В КАНАЛАХ СВЯЗИ И ПЕРЕДАЧИ ДАННЫХ

5.1. Кодирование для защиты информации от искажения помехами в системах передачи

Для сохранения достоверности и точности при передаче информации по линиям связи в условиях действия помех применяются специальные меры, уменьшающие вероятность появления ошибок. Одной из таких мер, едва ли не самой действенной, является применение помехоустойчивого кодирования. Кодирование дает возможность увеличивать помехоустойчивость передачи информации в обмен на увеличение избыточности и, соответственно, снижение скорости передачи сообщений. Но избыточность при кодировании может вводиться и использоваться по-разному. Во-первых, за счет избыточности можно создавать коды, способные при приеме и декодировании обнаруживать и исправлять (корректировать) ошибки, обусловленные действием помех. Это корректирующие коды. Во-вторых, избыточные символы могут использоваться для создания сигналов, максимально отличающихся друг от друга и потому хорошо противостоящих трансформациям одного сообщения в другое. Такие сигналы предназначаются для приема «в целом». В более сложных случаях информационную избыточность дополняют аппаратурной избыточностью, организуя передачу информации с обратной связью от получателя сообщений к их источнику.

При построении корректирующих кодов из N_0 возможных комбинаций по n символов применяется лишь некоторая часть $N < N_0$. Используемые при передаче N комбинаций символов обычно называются разрешенными кодовыми комбинациями, а остальные $N_0 - N$ — запрещенными. Если под действием помехи передаваемая кодовая комбинация трансформируется в запрещенную, то при некоторых условиях такую ошибку можно обнаружить и даже исправить.

Способность корректирующих кодов обнаруживать и исправлять ошибки можно пояснить следующими рассуждениями. Множество запрещенных кодовых комбинаций всегда можно разбить на N подмножеств N_i, $i \in 1:N$ и каждому подмножеству N_i поставить в соответствие разрешенную кодовую комбинацию B_i. Если искаженная помехами при передаче кодовая комбинация B_i^* принадлежит подмножеству N_i, то принимается решение в пользу

кодовой комбинации B_i. Очевидно, что при таком правиле приема будут исправляться все те ошибки, которые не выводят передаваемую кодовую комбинацию за пределы принадлежащего ей подмножества N_i. Если бы избыточности не было ($N = N_0$), то каждое подмножество N_i содержало бы по одной кодовой комбинации B_i и любые ошибки приема символов неизбежно переводили бы B_i^* в другую разрешенную кодовую комбинацию B_j, $j \neq i$.

При построении корректирующего кода основной задачей является разбиение множества запрещенных кодовых комбинаций на N подмножеств и выработка правила сопоставления их с разрешенными кодовыми комбинациями. Именно по способу такого разбиения различают коды и дают им названия. Для уменьшения вероятности ошибочного декодирования в подмножество N_i включаются те запрещенные кодовые комбинации B_i^*, для которых

$$P(B_i)P(B_k^*|B_i) > P(B_j)P(B_k^*|B_j),\ j \in 1:N,\ j \neq i, \qquad (5.1)$$

где $P(B_i)$ — априорная вероятность передачи кодовой комбинации B_i; $P(B_k^*|B_i)$ — условная вероятность принятия кодовой комбинации B_k^* при передаче кодовой комбинации B_i. Таким образом, в подмножество N_i должны входить кодовые комбинации B_k^*, при приеме которых наиболее вероятной комбинацией является B_i.

При передаче равновероятных сообщений по каналам с независимыми ошибками, когда вероятность появления ошибок уменьшается с увеличением их кратности, для минимизации средней вероятности ошибочного декодирования необходимо в первую очередь исправлять однократные ошибки как наиболее часто встречающиеся, затем двухкратные и т.д. При этом в подмножество N_i следует включить все те кодовые комбинации B_i^*, которые отличаются от B_i меньшим числом символов по сравнению с другими разрешенными кодовыми комбинациями. Соответственно декодер принимает решение о приеме кодовой комбинации B_i, если принятая комбинация B_i^* ближе к B_i, чем любые другие B_j, $j \neq i$. Такое правило принятия решения называется оптимальным по критерию максимума правдоподобия.

Код можно задать таблицей, устанавливающей соответствие между сообщениями и кодовыми комбинациями. Кодирующее устройство (кодер) при этом будет просто запоминающим устройством, в памяти которого хранятся N разрешенных кодовых комбинаций. Соответственно универсальный метод декодирования, пригодный для любого кода, заключается в сличении принятой кодовой комбинации со всеми N разрешенными и нахождении той разрешенной кодовой комбинации, которая отличается от принятой меньшим числом символов. Хотя такие методы кодирования и декодирования и являются универсальными, они не нашли широкого применения из-за большого объема требуемой памяти. В особенно-

сти это ограничение существенно для кодов большой длины. Поэтому к настоящему времени созданы и продолжают разрабатываться коды, не требующие запоминания большого количества комбинаций. Известно много помехоустойчивых кодов, которые классифицируются по различным признакам.

Прежде всего корректирующие коды разделяются на два больших класса: блочные и непрерывные.

При блочном кодировании последовательность элементарных сообщений источника разбивается на отрезки и каждому отрезку ставится в соответствие определенная последовательность (блок) кодовых символов, иначе называемая кодовой комбинацией. Множество всех кодовых комбинаций, разрешенных (возможных) при данном способе кодирования, и есть блочный код.

Длина блока может быть как постоянной, так и переменной. Соответственно? различают равномерные и неравномерные блочные коды. Помехоустойчивые коды являются, как правило, равномерными.

Блочные коды бывают разделимыми и неразделимыми. К разделимым относятся коды, в которых каждый из символов может быть отнесен к одной из двух непересекающихся групп: информационные символы, несущие сообщение, и проверочные, служащие исключительно для обнаружения и исправления ошибок. Такие коды принято обозначать парой чисел (n, k), где n — длина кода; k — число информационных символов. Число разрешенных комбинаций в коде (n, k) не превышает 2^k. К неразделимым относятся коды, у которых нельзя выделить информационные и проверочные символы. Неразделимые коды — это, например, коды с постоянным весом и коды на основе матриц Адамара. Коды с постоянным весом характеризуются тем, что все их кодовые комбинации содержат одинаковое число единиц. Примером такого кода является стандартный телеграфный код, у которого в каждой кодовой комбинации по три единицы и четыре нуля (код «3 из 7» : (7,3)).

Коды с постоянным весом позволяют обнаружить все ошибки кратности $Q = 1, \ldots, n - k$, за исключением случаев, когда число единиц, перешедших в нули, равно числу нулей, перешедших в единицы. В полностью асимметричных каналах, в которых возможны ошибки только одного вида (только трансформации нулей в единицы или единиц в нули), такой код позволяет обнаружить все ошибки. В симметричных каналах вероятность необнаруживаемой ошибки в первом приближении можно определить как вероятность одновременного искажения одной единицы и одного нуля:

$$P_{\text{н.о}} \approx C_3^1 P_{\text{ош}}(1 - P_{\text{ош}})^2 C_4^1 (1 - P_{\text{ош}})^3 = 12 P_{\text{ош}}^2 (1 - P_{\text{ош}})^5, \qquad (5.2)$$

где $P_{\text{ош}}$ — вероятность искажения одного символа.

Среди разделимых кодов выделяются коды линейные и нелинейные. К линейным относятся коды, в которых поразрядная сумма по модулю 2 любых двух разрешенных кодовых слов также является разрешенным кодовым словом. Линейный код называется систематическим, если первые k символов любой его кодовой комбинации являются информационными, а остальные $(n - k)$ символов — проверочными.

Наиболее простой линейный систематический код — это $(n, n-1)$, содержащий один проверочный символ, который равен сумме по модулю 2 всех информационных символов. Такой код называется кодом с проверкой на четность. Он позволяет обнаружить все сочетания ошибок нечетной кратности. Вероятность необнаруженной ошибки в первом приближении можно определить как вероятность искажения двух символов:

$$P_{\text{н.о}} \approx C_n^2 P_{\text{ош}} \left(1 - P_{\text{ош}}\right)^{n-2} \tag{5.3}$$

Подклассом линейных кодов являются циклические коды. У таких кодов все комбинации, образованные циклической перестановкой любой кодовой комбинации, являются также разрешенными кодовыми комбинациями. Это свойство позволяет значительно упростить кодирующее и декодирующее устройства, особенно при обнаружении ошибок и исправлении одиночной ошибки. Примерами циклических кодов могут служить коды Хемминга, коды Боуза—Чоудхури—Хоквингема (БЧХ коды) и некоторые другие.

Примером нелинейного кода является код Бергера, у которого проверочные символы формируются как двоичная запись числа единиц в последовательности информационных символов. Например, таким является код: 00000; 00101; 01001; 01110; 10001; 11010, 11111. Коды Бергера применяются, как правило, в асимметричных каналах. В симметричных каналах они обнаруживают все одиночные ошибки и некоторую часть многократных.

Непрерывные коды не разбиваются на блоки. Операции кодирования и декодирования производятся над непрерывной последовательностью символов. Самые распространенные и удобные для практического применения среди непрерывных — сверточные коды.

К числу основных характеристик кода относятся длина кода n, его основание m, мощность N (число разрешенных кодовых комбинаций), полное число кодовых комбинаций N_0, число информационных символов k, число проверочных символов $r = N - k$, вес кодовой комбинации (число единиц в комбинации), избыточность кода и кодовое расстояние. Избыточность кода определяется выражением

$$\chi = 1 - \frac{\log N}{\log N_0}, \tag{5.4}$$

или для двоичного кода ($m = 2$), когда $N = 2^k$,

$$\chi = 1 - \frac{k}{n} = \frac{r}{n}, \tag{5.5}$$

где $\frac{k}{n}$ — относительная скорость кода.

Для оценки степени сходства разных комбинаций, составляющих код, в пространстве кодовых последовательностей вводится метрика, т.е. определяется правило вычисления расстояния. Наиболее употребительная метрика основана на использовании расстояния Хемминга $d(B_i, B_j)$, которое определяется числом разрядов, где B_i отличается от B_j. Для двоичного кода

$$d(B_i, B_j) = \sum_{k=1}^{n} b_{ik} \oplus b_{jk}, \tag{5.6}$$

где b_{ik} и b_{jk} — символы кодовых комбинаций B_i и B_j соответственно; $\oplus$ — символ операции суммирования по модулю 2.

Наименьшее расстояние Хемминга для данного кода называется кодовым расстоянием d.

При независимых ошибках в канале через кодовое расстояние удобно выражается корректирующая способность кода. Если код имеет $d = 1$, то две кодовые комбинации отличаются минимум в одном символе. Искажение одного символа сразу трансформирует кодовую комбинацию в другую разрешенную, т.е. код с $d = 1$ не способен корректировать ошибки. Чтобы код мог обнаруживать любую одиночную ошибку, необходимо обеспечить кодовое расстояние, равное двум. Рассуждая аналогичным образом, можно получить, что для обнаружения всех ошибок кратности l требуется код с расстоянием

$$d \geq l + 1. \tag{5.7}$$

Для исправления всех ошибок некоторой кратности требуется большее кодовое расстояние, нежели для их обнаружения. Если кратность исправляемых ошибок равна l, то кодовое расстояние должно удовлетворять условию

$$d \geq 2l + 1. \tag{5.8}$$

Помимо режима декодирования с обнаружением и исправлением ошибок используется режим с восстановлением предварительно стертых ненадежных символов. В таких системах решающая схема приемника оперирует с некоторой областью неопределенности. Решение о переданном символе принимается только в случае, если принятый входной сигнал не попадает в указанную область, в противном случае приемник отказывается от принятия

решений и заменяет данный символ специальным символом стирания. Для восстановления стертых символов используются корректирующие коды.

Таким образом, задача построения кода с заданной корректирующей способностью сводится к обеспечению необходимого кодового расстояния путем введения избыточности. При этом желательно, чтобы число используемых проверочных символов было минимальным. К сожалению, задача определения минимального числа проверочных символов, необходимых для обеспечения заданного кодового расстояния, в общем виде не решена. Имеется лишь ряд оценок для максимального кодового расстояния при фиксированных N и k, которые часто используются для выяснения того, насколько код близок к оптимальному, имеющему минимальное кодовое расстояние для заданной корректирующей способности.

Так, для блочного линейного кода (N, k) справедливо неравенство

$$r \geq \log_2 \left(\sum_{i=0}^{\left[\frac{d-1}{2}\right]} C_n^i \right), \tag{5.9}$$

где r — верхняя граница Хемминга; $\left[\frac{d-1}{2}\right]$ — целая часть числа $\frac{d-1}{2}$.

Граница Хемминга (5.9) близка к оптимальной для кодов с большими значениями N/k. Для кодов с малыми значениями N/k более точной является верхняя граница Плоткина:

$$r \geq 2d - 2 - \log_2 d. \tag{5.10}$$

Но существует также блочный линейный код (N,k) с кодовым расстоянием d, для которого справедливо неравенство

$$r \leq \log_2 \sum_{i=0}^{d-2} C_n^i, \tag{5.11}$$

называемое нижней границей Варшамова — Гильберта.

Границы Хемминга (5.9) и Плоткина (5.10) являются необходимыми условиями существования кода с параметрами N, k и d, а граница Варшамова — Гильберта — достаточным условием. Равенство в (5.9) справедливо только для так называемых совершенных кодов. Такие коды исправляют все ошибки кратности $\left[\frac{d-1}{2}\right]$ и

менее и не исправляют ни одной ошибки кратности $l > \left[\frac{d-1}{2}\right]$, где, как и прежде, $\left[\frac{d-1}{2}\right]$ — целая часть числа $\frac{d-1}{2}$. Примером совершенных кодов являются коды Хемминга.

По определению, любой линейный код (N, k) можно получить из k линейно независимых кодовых комбинаций путем их посимвольного суммирования по модулю 2 в различных сочетаниях. Исходные линейно независимые кодовые комбинации называются базисными. Все k базисные комбинации длиной N символов можно расположить по строкам порождающей матрицы

$$\mathbf{G} = ||\, G_{kN} ||. \tag{5.12}$$

С использованием этого обозначения процесс кодирования заключается в выполнении преобразования

$$\mathbf{B} = \mathbf{AG}, \tag{5.13}$$

где **B** — вектор размерностью N, соответствующий кодовой комбинации; **A** — вектор размерности k, соответствующий кодируемому сообщению.

Таким образом, порождающая матрица (5.13) содержит всю необходимую для кодирования информацию, которая должна храниться в памяти кодирующего устройства. Для двоичного кода объем памяти равен kN двоичных символов. При табличном задании кода кодирующее устройство должно запоминать $N2^k$ двоичных символов.

Кодирующее устройство для линейного кода (N, k) (рис. 5.1) состоит из k-разрядного сдвигающего регистра и $r = (N-k)$ блоков сумматоров по модулю 2.

Информационные символы одновременно поступают на вход регистра и на выход кодирующего устройства через коммутатор. С поступлением k-го информационного символа на выходах блоков сумматоров в соответствии с уравнениями формируются проверочные символы, которые затем последовательно поступают

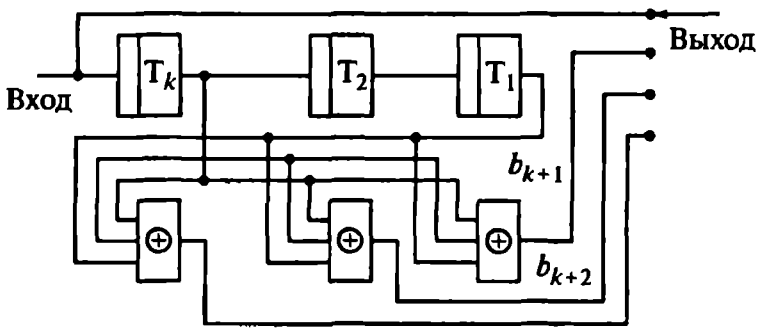


Рис. 5.1. Кодер линейного (N, k) кода

на выход кодера. Процесс декодирования сводится к выполнению операции

$$\mathbf{S} = \mathbf{B}^*\mathbf{H}^{\mathrm{T}}, \tag{5.14}$$

где $\mathbf{S}$ — вектор размерностью $(N-k)$, называемый синдромом; $\mathbf{B}^*$ — вектор принятой кодовой комбинации, возможно, искаженной помехами и поэтому отличающийся от $\mathbf{B}$; $\mathbf{H}$ — проверочная матрица размерности $(r \times N)$ такая, что вектор $\mathbf{B}$ принадлежит коду только в том случае, если $\mathbf{B}\mathbf{H}^{\mathrm{T}} = 0$, где т — символ транспонирования матрицы.

Если принятая кодовая комбинация $\mathbf{B}^*$ совпадает с одной из разрешенных (либо отсутствуют ошибки в принятых символах, либо из-за действия помех одна разрешенная кодовая комбинация трансформировалась в другую), то

$$\mathbf{S} = \mathbf{B}^*\mathbf{H}^{\mathrm{T}} = 0. \tag{5.15}$$

В другом случае $\mathbf{S} \neq 0$ и вид синдрома зависит только от вектора ошибок $\mathbf{e}$, определяемого как

$$\mathbf{B}^* = \mathbf{B} \oplus \mathbf{e}. \tag{5.16}$$

Из определения (5.16) видно, что $\mathbf{e}$ — это такая же последовательность из N символов, как $\mathbf{B}$ и $\mathbf{B}^*$, но имеющая нули на тех позициях, на которых символы $\mathbf{B}^*$ не отличаются от символов $\mathbf{B}$ и единицы на позициях искаженных символов. На основании (5.15) и (5.16) можно утверждать, что

$$\mathbf{S} = \mathbf{B}^*\mathbf{H}^{\mathrm{T}} = (\mathbf{B} \oplus \mathbf{e})\mathbf{H}^{\mathrm{T}} = \mathbf{e}\mathbf{H}^{\mathrm{T}}, \tag{5.17}$$

где $\mathbf{B}^*$ — вектор принятой комбинации с возможными ошибками в некоторых символах; $\mathbf{B}$ — вектор переданной кодовой комбинации.

Из (5.17) следует, что при $\mathbf{S} = 0$ декодер должен принимать решение об отсутствии ошибок, а при $\mathbf{S} \neq 0$ — что ошибки произошли. Число различных синдромов, соответствующих различным сочетаниям ошибок, равно $2^{N-k} - 1$. По конкретному виду синдрома можно в пределах корректирующей способности кода указать на ошибочные символы, а следовательно, и исправить их.

Схема декодера линейного (N, k) кода (рис. 5.2) содержит k-разрядный сдвигающий регистр, $(N-k)$ полусумматоров (сумматоров по модулю 2), схемы сравнения, анализатор и корректор ошибок. На регистре запоминаются информационные символы принятой кодовой последовательности, из которых в блоках сумматоров формируются проверочные символы. В результате сравнения формируемых на приемной стороне проверочных символов с принятыми проверочными символами анализатор ошибок определяет ошибочно принятые символы. Эти решения выносятся на

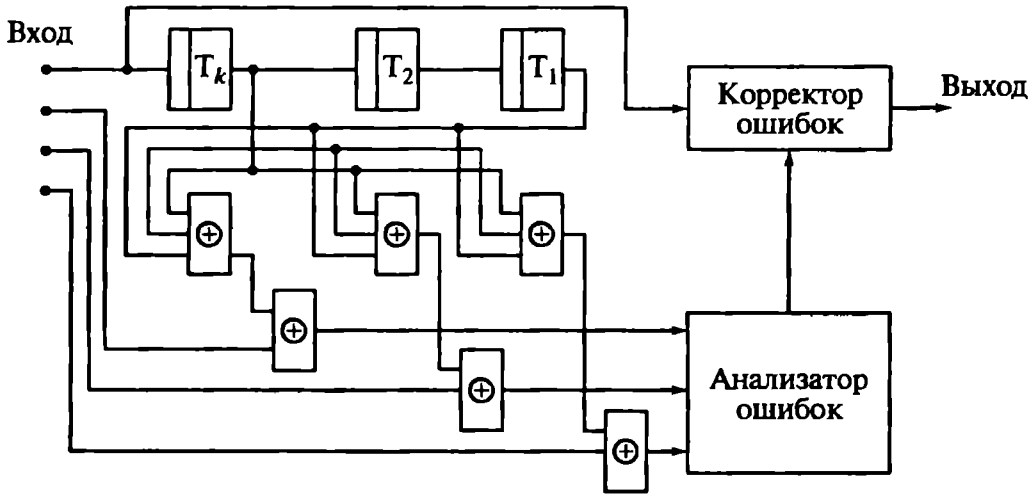


Рис. 5.2. Декодер линейного (N, k) кода

основании анализа синдрома. Исправление информационных символов производится в корректоре.

В общем случае при декодировании линейного кода с исправлением ошибок в памяти декодера нужно хранить таблицу соответствий между синдромами и векторами ошибок. Такая таблица должна содержать 2^{N-k} строк. Для каждой принятой кодовой комбинации декодер должен просматривать всю таблицу. При небольших значениях N эта операция не вызывает затруднений. Но для высокоэффективных кодов длиной $N \gg 10$ разность $(N-k)$ принимает такие значения, что перебор по таблице из 2^{N-k} строк оказывается практически невозможным.

Циклические коды относятся к классу линейных систематических. Поэтому для их построения достаточно знать порождающую матрицу. Но можно указать другой способ построения циклических кодов, основанный на представлении кодовых комбинаций полиномами. Так, всякой кодовой комбинации $\{b_{N-1}, b_{N-2}, \ldots, b_0\}$ может быть поставлено в соответствие число в позиционной двоичной системе, составленное из цифр $b_{N-1}, b_{N-2}, \ldots . b_0$. А значение этого числа определяется полиномом

$$B(x) = b_{N-1}x^{N-1} + b_{N-2}x^{N-2} + \quad + b_0x^0, \qquad (5.18)$$

где x — основание системы счисления; $b \in (0, x)$; суммирование ведется по модулю x.

В частности, комбинации двухосновного кода представляются двоичными числами $b = 0;1$, $x = 2$, и суммирование ведется по модулю 2.

Из эквивалентности кодовых комбинаций полиномам (5.18) следует, что все операции при преобразовании кодированных сообщений могут быть представлены как алгебраические действия над полиномами.

Каждый циклический код (N, k) характеризуется порождающим подиномом. Им может быть любой полином $P(x)$ степени $(N-k)$, который делит без остатка двучлен $x^N \oplus 1$, а также любую разрешенную кодовую комбинацию $B(x)$. Поэтому процесс кодирования сообщения $C(x)$ сводится к отысканию такого полинома $B(x)$, от деления которого без остатка на $P(x)$ получается частное $C(x)$. Иначе говоря, кодовая последовательность должна формироваться по правилу

$$B(x) = C(x)P(x), \tag{5.19}$$

причем $C(x)$ в соответствии с (5.19) представляется многочленом степени не выше $k - 1$.

Однако при кодировании в соответствии с правилом (5.19) формируются только неразделимые коды: информационные и проверочные символы в получаемых кодовых последовательностях оказываются перемешанными. Это свойство затрудняет процесс декодирования, поэтому на практике чаще всего применяется иной метод нахождения полинома $B(x)$.

Если умножить многочлен $C(x)$ на x^{N-k} и полученное произведение разделить на $P(x)$, в остатке будет полином $r(x)$:

$$C(x)x^{N-k} = Q(x)P(x) \oplus r(x). \tag{5.20}$$

Так как операции суммирования и вычитания по модулю 2 совпадают, из (5.20) следует, что полином

$$C(x)x^{N-k} \oplus r(x) = Q(x)P(x) \tag{5.21}$$

делится на порождающий полином $P(x)$ нацело (без остатка). Следовательно, этот полином является разрешенной кодовой последовательностью для кода, заданного порождающим многочленом $P(x)$.

У полинома $C(x)x^{N-k}$ коэффициенты при k старших членах совпадают с коэффициентами $C(x)$, а коэффициенты при $(N - k)$ равны нулю, т.е. совокупность N коэффициентов — это число, равное передаваемому сообщению, увеличенное на $(N-k)$ порядков. Остаток от деления $r(x)$ имеет степень не выше $(N-k)$. Таким образом, коэффициенты при k старших членах полинома $C(x)x^{N-k} \oplus r(x)$ — это информационные символы; совпадающие с символами кодируемого сообщения, а при $(N-k)$ младших — проверочные символы. Эти свойства полиномов подсказывают схемотехнические приемы построения кодеров циклического кода. Для примера на рис. 5.3 приведена схема кодера для кода с порождающим многочленом $P(x) = x^3 \oplus x^2 \oplus 1$.

Триггеры T_1, T_2 и T_3 образуют регистр сдвига. В исходном состоянии ключи K_1 и K_2 находятся в положении 1. Кодируемая последовательность $C(x)$ подается на вход кодера и вместе с этим посту-

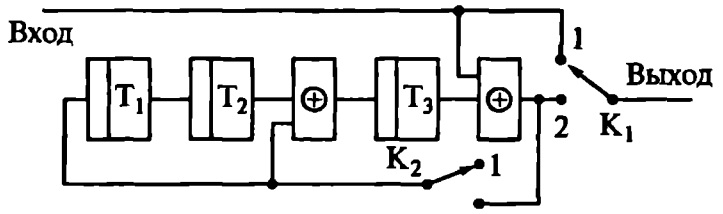


Рис. 5.3. Кодер циклического кода с порождающим полиномом $P(x) = x^3 \oplus x^2 \oplus 1$

пает на выход ячейки T_3 (это соответствует умножению многочлена $C(x)$ на x^3). За четыре такта сдвига происходит деление многочлена $C(x)x^3$ на многочлен $P(x) = x^3 \oplus x^2 \oplus 1$. В результате в регистре записывается остаток, представляющий собой проверочные символы. Ключи K_1 и K_2 перебрасываются в положение 2, и в течение трех следующих тактов содержащиеся в регистре символы поступают на выход кодера.

От порождающего полинома $P(x)$ зависит корректирующая способность кода, поэтому его выбор очень важен. Необходимо помнить, что степень порождающего многочлена должна быть равна числу проверочных символов.

Обнаружение ошибок при использовании циклических кодов сводится к делению многочлена $B^*(x) = B(x) + e(x)$, соответствующего принятой комбинации, на $P(x)$. Если остаток $r(x)$ оказывается равным нулю, то считается, что ошибки нет, в противном случае фиксируется ошибка.

Полином

$$r(x) = [B(x) + e(x)] \bmod P(x) = e(x) \bmod P(x) \qquad (5.22)$$

зависит только от многочлена ошибок $e(x)$ и играет ту же роль, что и вектор-синдром. Поэтому ошибки можно исправлять на основе таблицы соответствий между $e(x)$ и $r(x)$, сохраняемой в памяти декодера, как при линейных нециклических кодах. Однако свойство цикличности позволяет существенно упростить процедуру декодирования.

Один из распространенных алгоритмов исправления ошибок использует следующие свойства синдрома циклического кода. Если имеется циклический код с кодовым расстоянием d, исправляющий все ошибки вплоть до кратности $\left[\frac{d-1}{2}\right]$ включительно (квадратные скобки, как и прежде, обозначают целую часть отношения $\frac{d-1}{2}$), возможны следующие ситуации. Если искажены только проверочные символы, то вес синдрома будет меньше или

равен $\left[\frac{d-1}{2}\right]$, а сам синдром будет совпадать с вектором ошибок; если вектор ошибки искажает хотя бы один информационный символ, то вес синдрома будет больше $\left[\frac{d-1}{2}\right]$; если $r(x)$ — остаток от деления многочлена $b(x)$ на $P(x)$, то остатком от деления полинома $b(x)x^i$ на $P(x)$ является многочлен $r(x)x^i \bmod P(x)$, иначе говоря синдром некоторого циклического сдвига многочлена $b(x)$ является соответствующим циклическим сдвигом синдрома исходного многочлена, взятого по модулю $P(x)$.

Работа алгоритма декодирования иллюстрируется схемой рис. 5.4 для кода с порождающим полиномом $P(x) = x^3 \oplus x^2 \oplus 1$. Такой код имеет кодовое расстояние $d = 3$. Он способен исправлять все однократные ошибки.

Принятая кодовая комбинация одновременно поступает в буферный регистр сдвига, служащий для ее запоминания и циклического сдвига, а также на устройство деления на многочлен $P(x)$ для вычисления синдрома. В исходном состоянии ключ находится в положении 1. После семи тактов принятая кодовая комбинация оказывается полностью загруженной в буферный регистр, а в регистре устройства деления будет вычислен синдром. Если вес синдрома больше единицы, декодер начинает производить циклические сдвиги комбинации в буферном регистре при отсутствии новой комбинации на входе и одновременно вычислять их синдромы $r(x)x^i \bmod P(x)$ в устройстве деления. Если на некотором i-м шаге вес синдрома окажется меньше двух, то ключ переходит в положение 2, обратные связи в регистре деления разрываются. При последующих тактах ошибки исправляются путем подачи содержимого регистра деления на вход сумматора по модулю 2, включенного в буферный регистр. После семи тактов работы декодера в автономном режиме исправленная комбинация в буферном ре-

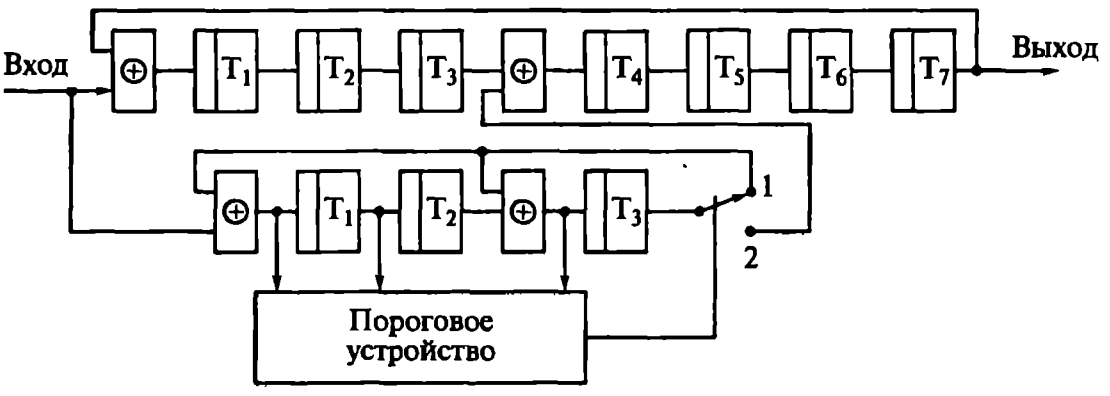


Рис. 5.4. Декодер циклического кода с порождающим полиномом $P(x) = x^3 \oplus x^2 \oplus 1$

гистре возвращается в исходное положение (информационные символы будут записаны в старшие разряды).

К циклическим кодам относятся коды Хемминга, которые являются примерами немногих известных совершенных кодов. Они имеют кодовое расстояние $d=3$ и исправляют все одиночные ошибки. Длина кода выбирается из условия $2^{N-k}=N$, которое имеет простой смысл: число различных ненулевых синдромов равно числу символов в кодовой последовательности. Так, существуют коды Хемминга $(2^r-1, 2^r-r-1)$, в частности коды (7,4), (15,11), (31,26), (63,57) и др. [22]. Ранее использованный в примерах многочлен $P(x)=x^3\oplus x^2\oplus 1$ является порождающим для кода Хемминга (7.4).

Известно, что для любых целых положительных чисел m и $l<N/2$ существует двоичный код БЧХ длины $N=2^m-1$ с кодовым расстоянием $d\geq 2l+1$, причем число проверочных символов $N-k\leq ml$.

Проще реализуется процедура мажоритарного декодирования, применимая для некоторого класса двоичных линейных, в том числе циклических кодов. Основана эта процедура на том свойстве кодов, что у них каждый информационный символ можно несколькими способами выразить через другие символы кодовой комбинации. Если для некоторого символа эти способы проверки дают неодинаковые результаты (одни дают результат 0, а другие 1, что может быть только в случае ошибочного приема), то окончательное решение по каждому из информационных символов принимается по мажоритарному принципу, т.е. по большинству. Декодеры мажоритарных кодов выполняются на регистрах сдвига. Примером кода, допускающего мажоритарное декодирование, является уже выше циклический код (7,3).

Мощные коды (т.е. коды с длинными блоками и большим кодовым расстоянием d) можно строить, объединяя несколько коротких кодов. Так строится, например, итеративный код из двух линейных систематических кодов (N_1,k_1) и (N_2,k_2). Вначале сообщение кодируется кодом первой ступени (N_1,k_1). Кодированная последовательность разбивается на блоки по k_2 символов. Эти символы считаются информационными для кода второй ступени. При кодировании на второй ступени к каждому блоку из k_2 информационных символов приписываются (N_2-k_2) проверочных. В результате получается блок, содержащий N_1N_2 символов, из которых k_1k_2 являются информационными.

Процесс формирования кода можно дополнить третьей итерацией, четвертой и т.д. При декодировании обнаруживают и исправляют ошибки каждого блока: сначала первой ступени, затем — второй. При этом исправляются только те ошибки, которые не были исправлены кодом первой ступени. Минимальное кодовое расстояние для двухмерного итеративного кода равно произведению минимальных кодовых расстояний для кодов первой и второй ступеней, т.е. $d=d_1d_2$.

На итеративный код похож каскадный код, но между ними имеется существенное различие. Первая ступень кодирования в каскадном коде осуществляется так же, как в итеративном. После того как сформированы k_2 блоков кода первой ступени (внутреннего), каждая последовательность из k_1 двоичных (информационных) символов внутреннего кода рассматривается как один символ недвоичного кода второй ступени (внешнего). Основание этого кода $\nu = 2^{k1}$. К этим символам приписываются еще $(N_2 - k_2)$ проверочных символов m-го кода также в виде строк длиной N_1. К каждой из этих строк приписываются двоичные проверочные символы в соответствии с внутренним кодом N_1,k_1.

В процессе приема сначала декодируются (с обнаружением или исправлением ошибок) все блоки внутреннего кода, а затем декодируется блок внешнего m-го кода (N_2,k_2), причем исправляются ошибки, оставшиеся после декодирования внутреннего кода. В качестве внешнего кода используют обычно m-й код Рида — Соломона, обеспечивающий наибольшее возможное значение d при заданных N_2 и k_2, если $N_2 < m$.

Сверточный код — это линейный рекуррентный код. В общем случае он образуется следующим образом. В каждый i-й тактовый момент времени на вход кодирующего устройства поступает k_0 символов сообщения: $c_{i1}c_{i2}...c_{ik0}$. Выходные символы $b_{i1}b_{i2}...b_{ik0}$ формируются по рекуррентному правилу из символов сообщения, поступивших в данный и предшествующие тактовые моменты времени. Величина k называется длиной кодового ограничения. Она показывает, на какое максимальное число выходных символов влияет данный информационный символ. Эта величина играет для сверточного кода ту же роль, что и длина блочного кода. Сверточный код имеет избыточность $\chi = 1 - k_0/N_0$. Обозначение такого кода (k_0/N_0).

Кодер сверточного кода может быть реализован с помощью сдвигающего регистра и сумматоров по модулю 2. Кодирующее устройство, выполненное по схеме (рис. 5.5), на каждый символ сообщения вырабатывает два символа выходной последовательности, которые по очереди подаются на выход через коммутатор.

Выходные символы формируются в результате линейного преобразования входного информационного символа и комбинации,

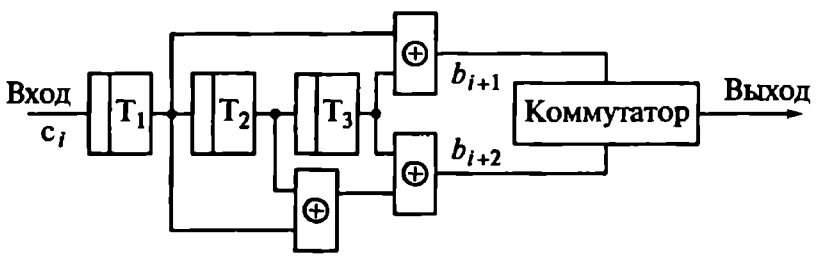


Рис. 5.5. Кодер сверточного кода

записанной в первых двух разрядах регистра. Связь между ячейками сдвигающего регистра и сумматорами по модулю 2 удобно описывать порождающими полиномами $Q_j(x)$, $j \in 1:N$. Для конкретного примера кодера (см. рис. 5.5) $Q_1(x) = x^2 \oplus 1$ описывает связи верхнего сумматора и $Q_2(x) = x^2 \oplus x \oplus 1$ описывает связи нижнего сумматора. Наличие члена x^i, $i = 0,1,2,...$ в порождающем многочлене означает, что $(i+1)$-й разряд регистра сдвига соединен с сумматором. Нумерация разрядов регистра — слева направо.

Сверточный код получается систематическим, если в каждый тактовый момент k_0 выходных символов совпадает с символами сообщения. На практике обычно используются несистематические сверточные коды.

Сверточные коды могут обладать свойством прозрачности. Прозрачные коды оказываются инвариантными по отношению к операции инвертирования сигнала: изменение значений символов на входе декодера на противоположные не влияет на результат декодирования. Это свойство очень удобно и широко используется для борьбы с эффектом обратной работы в радиосистемах передачи информации (РСПИ), использующих сигналы с фазовой модуляцией на 180°.

Корректирующая способность сверточного кода зависит от свободного расстояния $d_{св}$, аналогичного кодовому расстоянию d для блочных кодов.

Декодеры сверточных кодов алгоритмически и схемотехнически довольно сложны. Декодирование с вычислением проверочной последовательности применяется только для систематических кодов. По своей сущности оно ничем не отличается от соответствующего метода декодирования блочных кодов. На приемной стороне из принятых информационных символов формируют проверочные символы по тому же закону, что и на передающей стороне. Затем эти проверочные символы сравнивают с принимаемыми проверочными символами. В результате сравнения образуется проверочная последовательность, которая при отсутствии ошибок состоит из одних нулей. При наличии ошибок на определенных позициях последовательности появляются единичные символы. Закон формирования проверочных символов выбирается так, чтобы по структуре проверочной последовательности можно было определить искаженные символы. Алгоритмы декодирования без вычисления проверочной последовательности используют принцип максимума правдоподобия или последовательное декодирование [23].

За счет введения избыточности можно создавать сигналы, максимально отличающиеся друг от друга. Естественная мера сходства и различия сигналов — коэффициент их взаимной корреляции. Если система передачи информации использует набор сигналов $s_i(t)$, $i \in 1:m$, $t \in [0;T]$ с одинаковой энергией

$$Q = \int_0^T s_i^2(t)\,dt = \text{const}(i), \tag{5.23}$$

то на множестве, содержащем все m сигналов, коэффициент взаимной корреляции определяется соотношением

$$\rho_{ij} = \frac{1}{Q}\int_0^T s_i(t)s_j(t)\,dt, \tag{5.24}$$

где Q — натуральное число.

Сигналы $s_i(t)$ различаются в максимальной степени, если

$$\rho_{ij} = \begin{cases} 1 & \text{при } i = j; \\ \rho_{\min} & \text{при } i \neq j. \end{cases} \tag{5.25}$$

Если $\rho_{\min} = 0$, сигналы $s_i(t)$ называются ортогональными. Теоретически минимальное значение ρ может быть и меньше нуля:

$$\rho_{\min} = \begin{cases} -\dfrac{1}{m-1}, & \text{при } m = 2q; \\ -\dfrac{1}{m}, & \text{при } m = 2q-1. \end{cases} \tag{5.26}$$

Известны системы сигналов, имеющих $\rho_{\min}$ как в (5.26). К ним относятся, например, рассмотренные выше симплексные псевдошумовые сигналы на основе М-последовательностей. Для таких сигналов $m = 2^N - 1$, где N — разрядов регистра сдвига, используемого для генерации М-последовательности.

Из (5.26) следует, что при большом числе сигналов $m \gg 1$ $\rho_{\min} \approx 0$, т. е. оптимальные сигналы очень мало отличаются от ортогональных.

Удобная математическая модель описывает ортогональные сигналы как строки матрицы Адамара размера $m \times m$. Матрица Адамара $\mathbf{H}$ квадратная, состоящая из символов ± 1 и обладающая свойством

$$\mathbf{H}\mathbf{H}^{\mathrm{T}} = m\mathbf{I}, \tag{5.27}$$

где $\mathbf{H}^{\mathrm{T}}$ — транспонированная матрица $\mathbf{H}$; $\mathbf{I}$ — единичная матрица.

Из определения (5.27) матрицы Адамара следует, что любые две ее строки ортогональны. Перестановка строк или столбцов, равно как и умножение ее строк или столбцов на -1, сохраняет ортогональность. Считается, что матрицы Адамара существуют для всех $m = 4Q$, а для всех $m \leq 200$ в настоящее время матрицы Адамара построены. Если $m = 2^q$, то матрицы Адамара образуются

как кронекеровское произведение матриц Адамара меньшего размера. В соответствии с этим правилом

$$\mathbf{H}_{2q} = \begin{pmatrix} \mathbf{H}_{2q-1} & \mathbf{H}_{2q-1} \\ \mathbf{H}_{2q-1} & \bar{\mathbf{H}}_{2q-1} \end{pmatrix}, \tag{5.28}$$

где $\mathbf{H}_i$ — матрица Адамара размера $i \times i$; $\bar{\mathbf{H}}_i$ — матрица Адамара, размера $i \times i$, у которой все элементы заменены на противоположные (1 на −1, и наоборот); $\mathbf{H}_1 = (1)$.

Последовательности символов, составляющих строки получаемых в соответствии с рекуррентным правилом (5.28) матриц Адамара, называются функциями Уолша и обозначаются $\mathrm{wal}(i, t)$. В этом обозначении число i — порядок функции. Оно определяет количество перемен знаков функции на периоде повторения T и называется частостью (секвентностью). Переменная t — это время. Очень удобно использовать безразмерное время $\theta = \frac{t}{T}$ и рассматривать функции Уолша на основном нормированном к единице интервале $\theta \in \left[-\frac{1}{2}; \frac{1}{2}\right]$.

Те функции Уолша, которые на своем периоде оказываются периодическими меандровыми колебаниями, называются функциями Радемахера. Очевидно, что порядок функций Радемахера $i = 2^q - 1$, $Q = 0,1,2....$ Все функции Радемахера генерируются триггерными делителями частоты следования импульсов задающего генератора.

Для функций Уолша справедливо свойство мультипликативности:

$$\mathrm{wal}(i, \theta)\mathrm{wal}(j, \theta) = \mathrm{wal}(i \oplus j, \theta). \tag{5.29}$$

Иначе говоря, порядок функции Уолша, полученной в результате перемножения функций Уолша порядков i и j, равен поразрядной сумме по модулю 2 двоичных значений индексов i и j. Свойство мультипликативности позволяет построить простую логическую схему для генерации всего ансамбля функций Уолша, перемножая функции Радемахера. На рис. 5.6 для примера приведена схема генерации ансамбля из восьми функций Уолша, т.е. всех функций $\mathrm{wal}(i, \theta)$ для $i \in \{0,2....7\}$.

Если ансамбль функций Уолша включает в себя $\mathrm{wal}(0, \theta)$, то такие множества ортогональных сигналов в теории кодирования называются кодами Рида — Мюллера (РМ) первого порядка. Если ко всем комбинациям ортогонального двоичного кода добавить их инверсии, то полученное множество из $2m$ комбинаций будет со-

ставлять биортогональный код. Полученная таким образом система сигналов будет иметь среднее значение коэффициента взаимной корреляции любой пары разных сигналов $(\rho) = -\frac{1}{m-1}$.

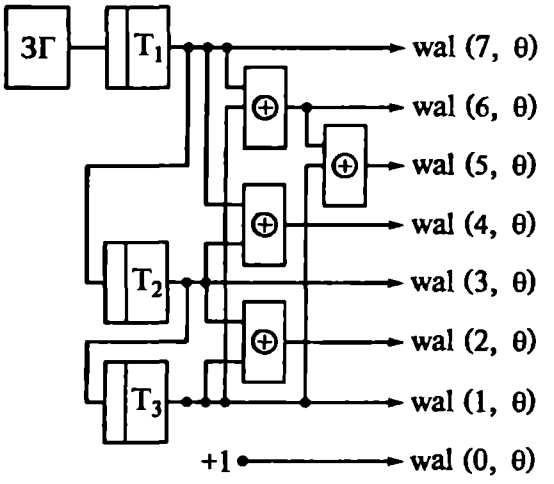


Рис. 5.6. Генератор функций Уолша

Оптимальный приемник для ортогональных и симплексных сигналов (рис. 5.7) содержит, параллельный набор из m корреляторов (последовательно соединенных перемножителей и интеграторов за время длительности сигналов T, которая в m раз превосходит длительность символа $T = m\tau_c$) и устройства выбора максимума, которое выносит решение о том, какому из возможных сигналов наиболее близко принятое колебание. Компаратор на выходе схемы служит для обнаружения сигнала, т.е. принятия решения о том, что выбранное максимальное значение соответствует сигналу на входе приемника, а не шумовому выбросу.

Процедуру, реализуемую при такой обработке сигнала, обычно называют приемом «в целом». Название подчеркивает то обстоятельство, что для вынесения решения о том, какой из возможных сигналов принят, обрабатывается целиком вся наблюдаемая на входе приемника реализация смеси сигнала с помехами.

Таким образом, ортогональные, симплексные и биортогональные сигналы либо оптимальны, либо близки к оптимальным при использовании приема «в целом» в присутствии аддитивного белого гауссова шума. Такие сигналы довольно просто генерировать. Но практическая реализация приема в целом наталкивается на определенные трудности, связанные со сложностью схемотехни-

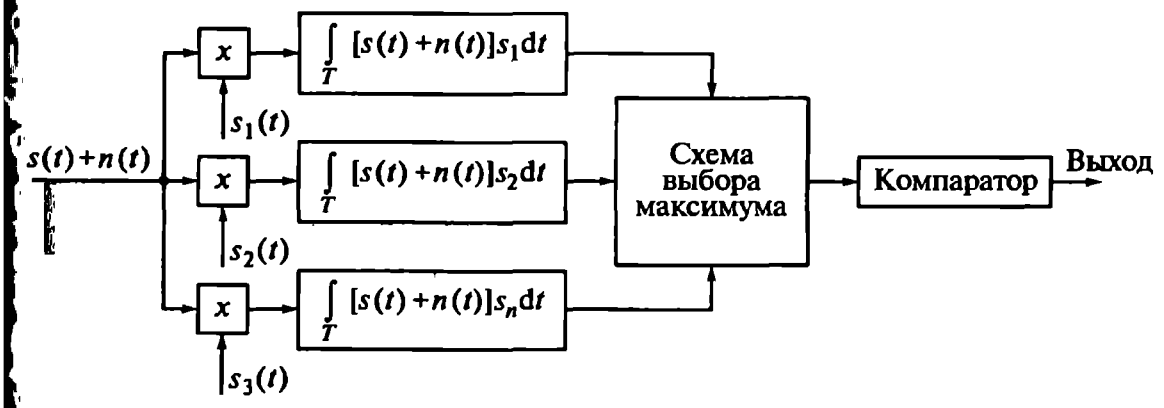


Рис. 5.7. Оптимальный приемник для ортогональных и симплексных сигналов

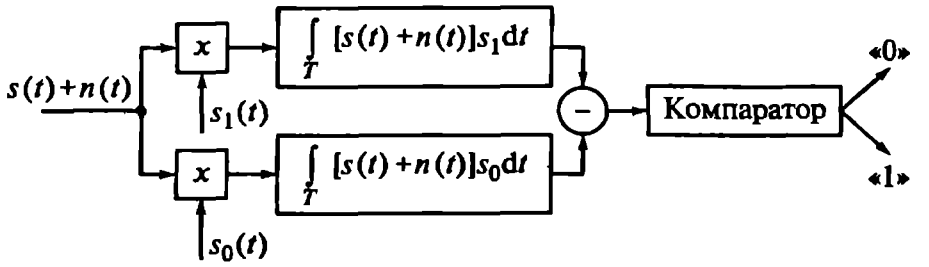


Рис. 5.8. Цифровой согласованный фильтр для приема «в целом»

ческой реализации приемника. Действительно, если блок из k информационных символов, поступающих от источника сообщений, в кодере преобразуются в один из $m = q^k$ сигналов, сложность реализации приемника «в целом», пропорциональная требуемому числу корреляторов, составит

$$\text{Сл} \approx m = Q^k = \exp\{klNQ\} = \mathbf{e}^{\alpha k}, \qquad (5.30)$$

где $\alpha = lNQ > 0$, т.е. экспоненциально растет с увеличением длины блока информационных символов. Для практически интересных значений k такой приемник оказывается технически очень сложным и даже нереализуемым.

Для разрешения проблемы сложности используют регенерацию символов принимаемого сигнала (посимвольный прием), а затем обрабатывают полученную кодовую последовательность двоичных символов, используя цифровые схемы согласованных фильтров. Схема для приема и восстановления символов сигнала представлена на рис. 5.8 [13].

На схеме (см. рис. 5.8) $s_0(t)$ и $s_1(t)$ — это сигналы, которые соответствуют передаче противоположных символов «0» и «1» соответственно. Такая схема оказывается оптимальной для приема и восстановления символов на фоне помехи в виде аддитивного нормального шума.

Разумеется, приемник с двухступенчатой схемой решения, когда на первой ступени восстанавливаются символы кодовой последовательности и лишь на второй ступени эти последовательности обрабатываются в соответствии с процедурой приема «в целом», проигрывает по помехоустойчивости оптимальному приемнику по схеме (см. рис. 5.7). Этот проигрыш служит платой за упрощение практической реализации схемы приема «в целом».

5.2. Обратная связь для адаптации к помеховой обстановке

Реализация любого способа повышения помехозащищенности системы передачи информации связана с введением информационной избыточности. При использовании помехоустойчивых ко-

дов избыточность связана с усложнением структуры кодированных сообщений, которое в конечном счете эквивалентно расширению спектра сигнала или увеличению времени передачи сообщения. При использовании сложных сигналов, предназначенных для приема «в целом», база увеличивается также за счет расширения спектра. Кроме того, повышение помехозащищенности всегда связано с некоторым усложнением систем передачи информации, т.е. увеличением аппаратурной избыточности.

Использование информационной и аппаратурной избыточности путем применения кодов, обнаруживающих и исправляющих ошибки, а также при использовании приема «в целом» сигналов с большой базой — не единственный и, возможно, не самый лучший способ обеспечения помехоустойчивости. Дело в том, что названные методы помехозащиты систем передачи информации оказываются не гибкими. Они проектируются для фиксированных, заранее определенных условий работы (скорее всего, самых тяжелых, наихудших). Но на практике помеховая обстановка в среде, где работают системы, может меняться. Соответственно могут меняться и требования к помехозащите: при меньшей интенсивности помех можно обойтись меньшей избыточностью и обеспечить более высокую скорость передачи информации. Но для такой адаптации скорости передачи информации к изменяющимся помеховым условиям необходимо иметь обратный канал передачи данных от приемника к передатчику. Системы, использующие такой канал, называются системами передачи информации с обратной связью. Обычно используются три основных способа осуществления обратной связи по передаваемой информации.

При первом способе сообщение, принятое и запомненное получателем, ретранслируется источнику информации по обратному каналу. Переданное и ретранслированное сообщения сравниваются. Если ошибки при передаче не случилось и переданное приемником сообщение совпадает с принятым по обратному каналу, передатчик формирует сигнал подтверждения правильности полученных данных. В случае несоответствия сообщения, принятого по каналу обратной связи, тому, которое ранее было передано по прямому каналу, передатчик фиксирует ошибку и формирует специальный сигнал стирания данных в памяти приемного устройства. После стирания передача сообщения повторяется вновь. И так до тех пор, пока не будет зафиксирован факт неискаженной передачи. Поскольку вся передаваемая информация ретранслируется по обратному каналу, подобная обратная связь называется информационной. Функциональная схема РСПИ с информационной обратной связью приведена на рис. 5.9.

Очевидно, что чем больше интенсивность помех в прямом и обратном каналах (см. рис. 5.9) и соответственно вероятность ошиб-

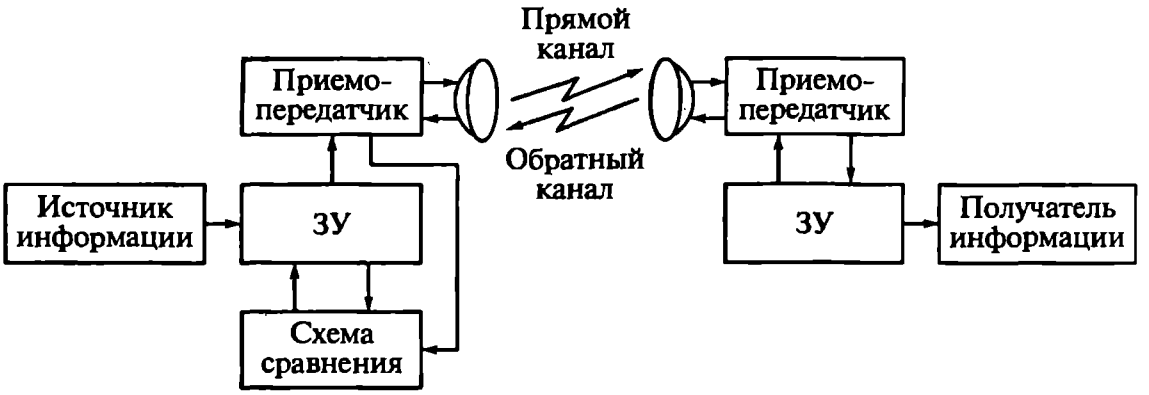


Рис. 5.9. РСПИ с информационной обратной связью

ки при передаче, тем больше следует ожидать повторных передач и тем больше информационная избыточность.

Второй способ использования обратного канала — организация решающей обратной связи. В радиосистемах с решающей обратной связью проверка правильности приема сообщения и принятие решения о необходимости повторной передачи производятся на приемной стороне аппаратурой получателя информации. Функциональная схема такой радиосистемы приведена на рис. 5.10.

Анализ принятой кодовой комбинации выполняется декодирующим устройством приемника. Естественно, что для реализации этой возможности применяется корректирующий код. В случае обнаружения ошибки принятое сообщение считается искаженным и по обратному каналу передается запрос на повторную передачу. Если декодер не обнаруживает ошибок в принятой кодовой комбинации, по обратному каналу передается подтверждение правильности приема (квитанция). Получив квитанцию, удостоверяющую правильность приема, источник сообщений передает следующий блок информации. В противном случае он повторяет передачу предыдущего искаженного блока. Таким образом,

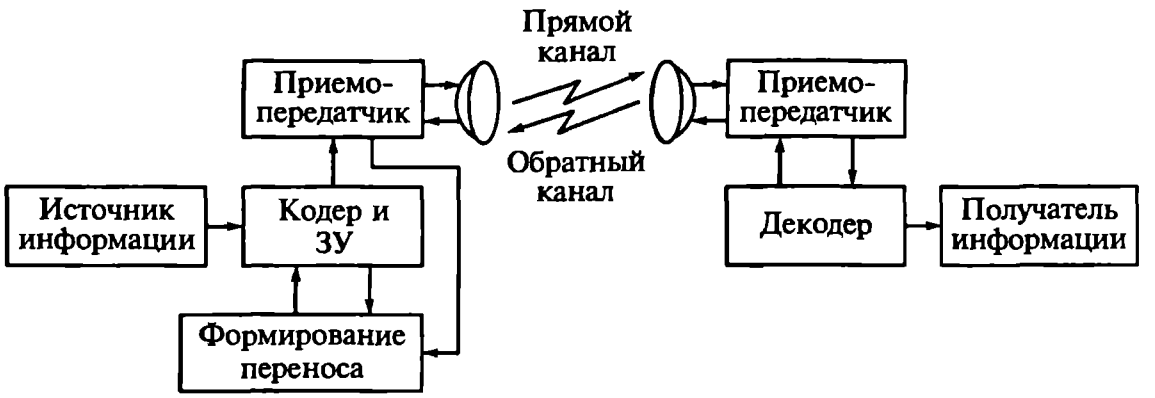


Рис. 5.10. РСПИ с решающей обратной связью

решение о правильности принятого сообщения выносится в точке приема (отсюда название «решающая обратная связь»). Иное название систем с решающей обратной связью — системы с переспросом. Очевидно, что при использовании решающей обратной связи по обратному каналу передается всего одна двоичная единица информации на каждый информационный блок в прямом канале.

Третий способ использует одновременно принципы как информационной, так и решающей обратной связи. Это комбинированная корректирующая обратная связь в системах передачи информации. Например, при решении об ошибки передачи сообщения по обратному каналу посылается квитанция-подтверждение, как при решающей обратной связи. Если приемник выносит решение о правильном приеме, по обратному каналу ретранслируется все принятое сообщение. При этом появляется возможность для устранения трансформации на приеме одной разрешенной кодовой комбинации в другую разрешенную, но, тем не менее, отличающуюся от переданной.

При любом способе осуществления проверочной обратной связи повторная передача сообщения может происходить неограниченное число раз до тех пор, пока не будет принято решение о достоверности принятого сообщения. Но практически максимально возможное число повторений r_{max} всегда ограничивается некоторой величиной, определяемой максимально допустимой задержкой при передаче, т.е. минимально допустимой скоростью передачи информации.

При анализе эффективности цифровых радиосистем передачи информации с проверочной обратной связью вычисляют остаточную вероятность $P_{ост}$ [19], т.е. вероятность тех ошибок, которые не обнаруживаются и не исправляются в результате $r \le r_{max}$ сеансов повторной передачи. Значения $P_{ост}$ и r_{max} зависят от свойств как прямого, так и обратного каналов РСПИ и от характеристик действующих в этих каналах помех.

Процесс передачи сообщения можно представить как последовательность отдельных циклов. Каждый цикл включает в себя передачу блока информации по прямому каналу и передачу соответствующего сообщения по каналу обратной связи. В момент окончания каждого цикла возможны следующие три ситуации:

ошибки в прямом канале отсутствуют, и блок информации принят правильно с вероятностью $P_{прав}$;

случается необнаруживаемая ошибка с вероятностью $P_{н.о}$;

случается ошибка, которая обнаруживается за счет избыточности кода с вероятностью $P_{о.о}$. В последнем случае производится повторная передача сообщения по прямому каналу.

Перечисленные ситуации составляют полную группу случайных событий, поэтому

$$P_{\text{прав}} + P_{\text{н.о}} + P_{\text{о.о}} = 1. \quad (5.31)$$

В результате однократной передачи остаточная (необнаруженная) ошибка будет происходить с вероятностью (5.31)

$$P_{\text{ост1}} = P_{\text{н.о}} = 1 - P_{\text{прав}}. \quad (5.32)$$

Если при первой передаче ошибка обнаруживается (с вероятностью $P_{\text{о.о}}$), цикл повторяется и опять возможны три исхода. Остаточная вероятность ошибки после повтора составит, очевидно,

$$P_{\text{ост2}} = P_{\text{о.о}}(1 - P_{\text{прав}}) = P_{\text{о.о}}(P_{\text{о.о}} + P_{\text{н.о}}) = P_{\text{о.о}}P_{\text{н.о}} + P_{\text{о.о}}^2. \quad (5.33)$$

В результате $r+1$-кратной передачи, когда ошибка обнаруживается r раз, остаточная вероятность ошибки составит $P_{\text{н.о}}P_{\text{о.о}}^r$, где $P_{\text{о.о}}^r$ — это вероятность появления обнаруживаемой ошибки в предыдущих циклах передачи. При неограниченном числе повторений, когда $r \to \infty$,

$$P_{\text{ост}} = P_{\text{н.о}} + P_{\text{о.о}}P_{\text{н.о}} + P_{\text{о.о}}^2P_{\text{н.о}} + \ldots = P_{\text{н.о}}(1 + P_{\text{о.о}} + P_{\text{о.о}}^2 + \ldots). \quad (5.34)$$

В скобках выражения (5.34) заключена сумма бесконечной геометрической прогрессии:

$$P_{\text{ост}} = \frac{P_{\text{н.о}}}{1 - P_{\text{о.о}}}. \quad (5.35)$$

Как видно, вероятность остаточной ошибки зависит не только от вероятности $P_{\text{н.о}}$, но и от вероятности $P_{\text{о.о}}$. При высокой вероятности обнаружения ошибок $P_{\text{о.о}} \to 1$ вероятность остаточной ошибки может существенно превосходить $P_{\text{н.о}}$.

Среднее число передач одного и того же сообщения можно определить соотношением

$$\langle r \rangle = \sum_{r=1}^{\infty} rP(r) = \sum_{r=1}^{\infty} rP_{\text{о.о}}^{r-1}(1 - P_{\text{о.о}}) = \frac{1}{1 - P_{\text{о.о}}}, \quad (5.36)$$

где $P(r)$ — вероятность r-кратной передачи сообщения, вычисляемая в предположении о том, что в каждом из $r-1$ предыдущих циклов передачи обнаруживается ошибка, а в цикле с номером r обнаружения ошибки не происходит, $P(r) = P_{\text{о.о}}^{r-1}(1 - P_{\text{о.о}})$.

Как следует из (5.36), среднее число повторений при передаче сообщений по системе с корректирующей обратной связью зависит от вероятности $P_{\text{о.о}}$, с которой происходит обнаруживаемая ошибка. При уменьшении соотношения сигнал-шум увеличивается вероятность ошибки и, соответственно, монотонно растет $P_{\text{о.о}}$. Но при этом растет и среднее число повторений сообщения, т.е. система с корректирующей обратной связью автоматически умень-

шает скорость передачи информации при ухудшении помеховой обстановки в среде распространения сигнала.

Стойкость цифровой радиолинии с информационной обратной связью к помехам легче всего оценить, предполагая, что для передачи сообщений используется безызбыточный код. Такое предположение совершенно естественно, поскольку достоверность передачи сообщений в радиосистеме с информационной обратной связью определяется не корректирующей способностью кода, а числом повторений. Можно также предположить, что ошибки в прямом и в обратном каналах статистически независимы. Это действительно так: поскольку сообщения в прямом и в обратном каналах не должны влиять друг на друга, постольку независимыми друг от друга будут и помехи в этих каналах. Статистически независимыми предполагаются и искажения отдельных символов передаваемых сообщений (ошибки при передаче не группируются в пакеты). Если даже помехи таковы, что могут воздействовать на группы соседних символов и вызывать пакеты ошибок, то для борьбы с ними приняты специальные меры. Например, символы передаваемого сообщения могут быть перемешаны по известному на приемной стороне закону. При восстановлении на приеме естественного порядка следования символов пакеты ошибок разравниваются по всей длине сообщения.

При безызбыточном кодировании каждое сообщение содержит k информационных символов и искажение любого из них приводит к ошибке и, как следствие, повторной передаче всего блока из k символов. При этом не важно, где конкретно произошла ошибка: в прямом или обратном канале. Необнаруживаемая ошибка соответствует такой комбинации искажений отдельных символов сообщения в прямом и обратном каналах РСПИ, при которых искажения взаимно компенсируются. Пример подобных ошибок — «зеркальные» ошибки, когда при передаче по обратному каналу искажаются те и только те символы, которые были искажены в прямом канале.

Вероятность искажения одного символа в прямом канале $P_{1\rightarrow}$, а в обратном $P_{1\leftarrow}$. Причем эти вероятности достаточно малы, так что $kP_{1\rightarrow} \ll 1$ и $kP_{1\leftarrow} \ll 1$. При обоснованных ранее предположениях о независимости искажений символов помехами ошибка передачи сообщения произойдет в результате одиночной зеркальной ошибки, т.е. тогда, когда в прямом канале исказится один символ, а в обратном — тоже только один и именно тот же самый. Вероятность искажения только одного символа из k информационных символов в прямом канале равна

$$P_{\rightarrow} = kP_{1\rightarrow}(1 - P_{1\rightarrow})^{k-1}. \qquad (5.37)$$

Условная вероятность обратной трансформации символа, который исказился в прямом канале, при ретрансляции сообщения

по обратному каналу (имеется в виду случай, когда трансформация указанного символа не сопровождается другими ошибками в обратном канале) вычисляется по формуле

$$P_{\leftarrow} = P_{1\leftarrow}(1 - P_{1\leftarrow})^{k-1}. \quad (5.38)$$

Основываясь на (5.37) и (5.38), вероятность одиночной зеркальной ошибки можно определить соотношением

$$P_{\text{н.о}} = P_{\rightarrow}P_{\leftarrow} = kP_{1\rightarrow}(1 - P_{1\rightarrow})^{k-1}P_{1\leftarrow}(1 - P_{1\leftarrow})^{k-1} \approx kP_{1\rightarrow}P_{1\leftarrow}. \quad (5.39)$$

Вероятность обнаружения ошибки при использовании информационной обратной связи — это вероятность любой ошибки, кроме зеркальной. Вероятность такого события

$$P_{\text{о.о}} = 1 - P_{\text{прав}} - P_{\text{н.о}} \approx k(P_{1\rightarrow} + P_{1\leftarrow} - P_{1\rightarrow}P_{1\leftarrow}) < 1. \quad (5.40)$$

Вероятность правильного приема команды в одном цикле передачи определяется формулой

$$P_{\text{прав}} = (1 - P_{1\rightarrow})^k(1 - P_{1\leftarrow})^k \approx 1 - k(P_{1\rightarrow} + P_{1\leftarrow} - P_{1\rightarrow}P_{1\leftarrow}). \quad (5.41)$$

Рассматривая предельный случай $r_{max}\rightarrow\infty$, используя соотношения (5.37), (5.38) и учитывая соотношение (5.41), можно получить

$$P_{\text{ост}} \approx P_{\text{н.о}} \approx kP_{1\rightarrow}P_{1\leftarrow}. \quad (5.42)$$

Для прямого канала системы передачи информации вероятность искажения блока из k символов определяется приближенным соотношением

$$P_{\text{иск}} \approx kP_{1\rightarrow}. \quad (5.43)$$

Сравнение (5.42) и (5.43) показывает, что применение системы передачи информации с полной ретрансляцией позволяет существенно уменьшить вероятность ошибки, если обратный канал обладает достаточно высокой помехоустойчивостью ($P_{1\leftarrow} \ll 1$).

При невысоком энергетическом потенциале в обратном канале последнее условие может и не выполняться. Тогда вместо полной ретрансляции применяют другие способы использования обратного канала. При этом скорость передачи информации по обратному каналу выбирается меньшей по сравнению со скоростью в прямом канале РСПИ. Один из таких способов используется при организации уже рассмотренной решающей обратной связи, когда по обратному каналу передается 1 бит информации на каждый блок из k бит информации в прямом канале. За счет уменьшения скорости передачи информации по обратному каналу увеличивается его помехозащищенность. Но использование решающей обратной связи требует применения в прямом канале корректирую-

щих кодов, т. е. передачи кроме k информационных еще и некоторого количества N проверочных символов. Известны способы борьбы с ошибками в обратном канале, приводящими к потере сообщения, основанные на несимметричном кодировании. При этом в обратном канале используются такие коды и правила декодирования, которые обеспечивают вероятность ошибочного приема сигнала переспроса существенно меньшую вероятности ошибки при приеме сигнала подтверждения.

Повторение передачи сообщения при использовании проверочной обратной связи любого типа (информационной, решающей или комбинированной) эквивалентно введению дополнительной избыточной информации. Но количество такой избыточной информации изменяется в зависимости от результатов каждого сеанса приема отдельного сообщения. При благоприятных условиях приема в прямом и обратном каналах искажения сообщений возникают сравнительно редко и, следовательно, среднее число повторных передач оказывается небольшим. Если уровень помех в точке приема сообщений увеличивается, то автоматически увеличивается и количество повторений. Таким образом, при изменении мощности принятого сигнала или мощности помех автоматически регулируется средняя скорость передачи информации по РСПИ. Так работает механизм адаптации РСПИ с обратной связью к помеховой обстановке.

РСПИ с обратной связью применяются для передачи очень важных сообщений. Например, информации при командном радиоуправлении. Очень эффективны адаптивные РСПИ с корректирующей обратной связью при работе в условиях замираний сигнала.

5.3. Искажения кодированных сообщений помехами

Цифровые методы передачи информации, использующие помехоустойчивое кодирование и другие способы внесения избыточности для защиты информации от искажений помехами в линиях связи, применяют сигналы с кодово-импульсной модуляцией (КИМ). Такие сигналы находят применение не только в системах передачи данных и командных радиолиниях, но и в системах связи, для которых ранее использовались сигналы. Поэтому сигналы с КИМ приходится рассматривать как очень важный класс, а качество приема таких сигналов — как важный показатель эффективности защиты информации. В дальнейшем качество защиты информации при использовании для ее передачи цифровых сигналов оценивается вероятностью ошибки приема каждого отдельного элемента (символа). Вопросы синхронизации приемников с передающими устройствами ниже не рассматриваются, хотя

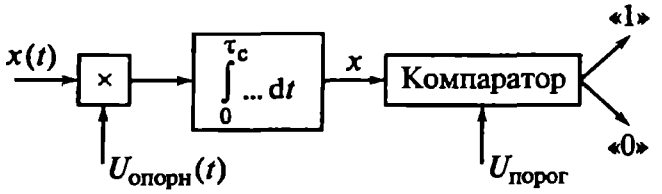


Рис. 5.11. Демодулятор сигнала с КИМ

эти вопросы весьма существенны для оценок точности и достоверности передачи сообщений в многоканальных системах с временным и кодовым разделением каналов.

Для оценки потенциально достижимой вероятности ошибки можно принять следующие предположения и допущения относительно сигнала с кодово-импульсной модуляцией [19].

Сигнал КИМ представляет собой поток из статистически независимых равновероятных двоичных символов $s_0(t)$ и $s_1(t)$ (логические значения символов «0» и «1»); мощности сигнала P_c, длительности символа τ_c, энергии символа $Q_c = P_c\tau_c$.

Сигнал наблюдается приемником средства разведки в аддитивной смеси с нормальным стационарным шумом $n(t)$:

$$x(t) = s(t) + n(t). \tag{5.44}$$

Шум $n(t)$ имеет спектральную плотность N_0.

Сигнал может иметь пассивную паузу (КИМ-АМ), когда передаче символа «0» соответствует пауза в излучении, т.е. $s_0(t) = 0$, или активную паузу (КИМ-ЧМ или КИМ-ФМ), когда $s_0(t) \neq 0$ и $s_1(t) \neq 0$, а энергии сигналов, соответствующих символам «0» и «1» соответственно $s_0(t)$ и $s_1(t)$, одинаковы.

Оптимальный алгоритм работы приемника при сделанных предположениях сводится к вычислению корреляционного интеграла принятого колебания $x(t)$ с опорным напряжением и сравнению значения этого интеграла с пороговым уровнем для принятия решения о сигнале по каждому принятому символу [19]. Работу решающего устройства приемника в соответствии с таким алгоритмом можно иллюстрировать структурной схемой рис. 5.11.

Для сигнала с пассивной паузой

$$\xi = \int_0^{\tau_c} x(t)s_1(t)dt \underset{<}{\overset{>}{}} \frac{Q_c}{2}. \tag{5.45}$$

Если выполняется верхнее неравенство, принимается решение о наличии на входе сигнала $s_1(t)$, если нижнее — $s_0(t)$.

Для сигнала с активной паузой

$$\xi = \int_0^{\tau_c} x(t)\left[s_1(t) - s_0(t)\right]dt \underset{<}{\overset{>}{}} 0. \tag{5.46}$$

Ошибки случаются тогда, когда нормальная случайная величина ξ оказывается выше порога принятия решения при наличии на входе приемника сигнала $s_0(t)$, и тогда, когда ξ меньше порога, а на входе колебание $x(t)$ содержит сигнал $s_1(t)$. Вероятность ошибки, определенная на основе этих соображений, составляет

$$P_{\text{ош}} = \frac{1}{2}\left[1 - \Phi\left(\sqrt{\frac{Q_c}{N_0}(1-\rho_s)}\right)\right], \qquad (5.47)$$

где $\Phi(\cdot)$ — интеграл вероятностей в форме $\Phi(t) = \frac{2}{\sqrt{\pi}}\int_0^t \mathrm{e}^{-t^2}\mathrm{d}t$;

ρ_s — коэффициент взаимной корреляции сигналов $s_1(t)$ и $s_0(t)$, $\rho_s \in [-1; 1]$:

$$\rho_s = \frac{1}{Q_c}\int_0^{\tau_c} s_0(t)s_1(t)\mathrm{d}t. \qquad (5.48)$$

Для сигналов с пассивной паузой и сигналов с КИМ-ЧМ $\rho_s = 0$ (ортогональные сигналы $s_1(t)$ и $s_0(t)$), а для сигналов с КИМ-ФМ $\rho_s = \cos\varphi$, где φ — индекс фазовой модуляции. Таким образом, $\rho_s = -1$ для противоположных сигналов, когда $\varphi = \frac{\pi}{2}$.

В (5.47) нужно учитывать, что при равновероятных символах $s_1(t)$ и $s_0(t)$ средняя мощность сигнала с пассивной паузой в два раза меньше, чем у сигнала с активной паузой.

С учетом сказанного, на основании (5.47) и (5.48) можно получить зависимости вероятностей ошибок оптимального приема символов сигнала с кодово-импульсной модуляцией от отношения сигнала к шуму [19]. Эти зависимости воспроизведены на рис. 5.12.

Разумеется, потенциальные оценки качества приема сигнала дают не больше, чем ориентировочную нижнюю границу вероятности ошибки приема символа, поскольку они определяются для некоторых идеальных моделей сигналов, шумов и способов построения приемника. Реально в приемниках сигналов с кодово-импульсной модуляцией часто применяются некогерентные методы обработки сигналов с КИМ-АМ и КИМ-ЧМ. В приемниках сигналов с КИМ-ФМ всегда приходится применять некоторые разновидности когерентного приема [19].

Способ некогерентного приема сигналов КИМ при амплитудной модуляции (манипуляции) несущего колебания предполагает использование в приемнике детектора огибающей входного сигнала. При этом пороговый уровень различения сигналов $s_1(t)$ и

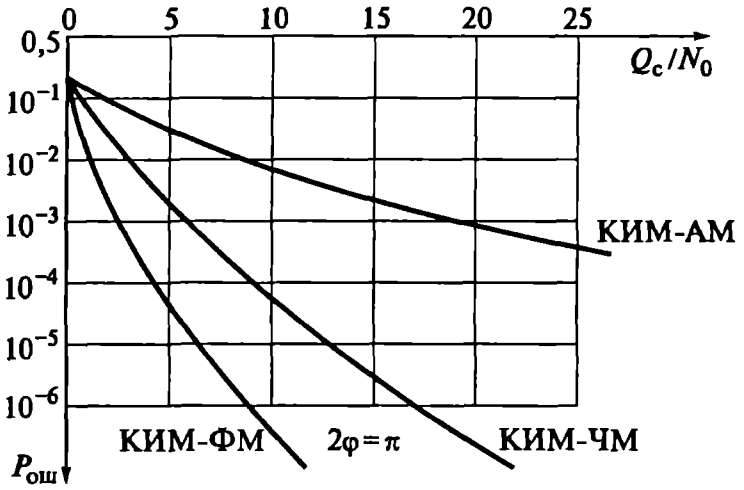


Рис. 5.12. Вероятность ошибки приема символа

$s_0(t)$ зависит от отношения сигнала к шуму $\frac{Q_c}{N_0}$ в полосе $\Delta f \approx \frac{1}{\tau_c}$. Как показано в [19], при оптимально выбранном пороге и отношении сигнала к шуму $\frac{Q_c}{N_0} > 16$ (при этом эффективность противодействия помехам достаточно высока и сигнал может быть принят со сравнительно высокой вероятностью) вероятность ошибки реального некогерентного приемника будет в

$$\left(\frac{P_{ош.нк}}{P_{ош.опт}}\right)_{АМ} \approx \frac{\sqrt{\pi}}{2}\sqrt{\frac{Q_c}{N_0}} \qquad (5.49)$$

раз больше, чем при оптимальном приеме (см. рис. 5.12).

Некогерентный приемник сигналов с КИМ-ЧМ содержит два фильтра, настроенных на частоты сигналов $s_1(t)$ и $s_0(t)$, детекторы огибающей сигналов на выходах этих фильтров и компаратор для сравнения выделенных фильтрами огибающих. Различие в вероятностях ошибок реального и оптимального приемников в этом случае определяется соотношением [19]:

$$\left(\frac{P_{ош.нк}}{P_{ош.опт}}\right)_{ЧМ} \approx 1{,}26\sqrt{\frac{Q_c}{N_0}}, \qquad (5.50)$$

справедливым при $\frac{Q_c}{N_0} > 9$.

При демодуляции сигналов КИМ-ФМ приемник должен использовать фазовый детектор. Независимо от конкретного схемо-

технического решения фазовый детектор перемножает входное колебание $x(t)$ с опорным напряжением $U_{\text{опорн}}(t)$, синхронным и синфазным с несущим (модулируемым) колебанием. Иначе говоря, прием сигналов с ФМ требует в обязательном порядке проведения тех же операций над принимаемым колебанием, выполнение которых предписывается процедурой оптимального когерентного приема. Поэтому следует ожидать, что и характеристики качества приема КИМ-ФМ должны быть такими же, как у оптимального приемника, но с оговорками относительно влияния шумов в канале формирования опорного напряжения фазового детектора. Действительно, когерентное опорное колебание $U_{\text{опорн}}(t)$, обеспечивающее работу фазового детектора при демодуляции КИМ-ФМ, должно формироваться из принятого сигнала. Известно много разных вариантов построения схемы формирования опорного напряжения. Выбор того или иного варианта определяется рядом конкретных условий: индексом фазовой манипуляции, соотношением сигнал-шум, элементной базой, используемой для построения приемника и т.п. Однако в любом случае вместе с опорным колебанием на фазовый детектор будет действовать шум, который не улучшает качества приема и демодуляции сигнала. Поэтому следует считать, что самая нижняя кривая на рис. 5.12, характеризующая вероятность ошибки оптимального приема сигнала с КИМ-ФМ для модуляции на $\pm\frac{\pi}{2}$, это верхняя граница вероятности ошибки в реальном приемнике цифровых сигналов.

5.4. Шифрация для защиты от несанкционированного доступа к информации

Противодействие информационному нападению радиоэлектронных разведок, вскрывающих содержание передаваемых по линиям связи сообщений, осуществляется криптографическими методами. Проблема криптографического обеспечения информационной безопасности составляет основное содержание науки криптологии, которая довольно четко подразделяется на криптографию, изучающую методы создания и применения шифров, и криптоанализ — науку (и искусство) раскрытия шифров. Криптография и криптоанализ неизмеримо старше проблемы информационной безопасности. Легенда, пересказанная римским историком Гаем Светонием, связывает первое применение криптозащиты информации с именем Цезаря, шифровавшего письма Цицерону и другим друзьям в Риме более 2000 лет назад. На современном уровне довольно широкого общественного интереса к криптологии нашлись свидетельства применения криптографических методов защиты информации в еще раньше в древнем Египте, Ки-

тае и других государствах глубокой древности. Научная эра развития криптологии началась именно в наше время и была обусловлена развитием телекоммуникаций на основе применения методов и средств радиоэлектроники, т.е. развитием радиоэлектронных систем связи и передачи данных. Криптология является довольно специфичной и весьма деликатной областью знания и практической деятельности.

Криптографические методы появились и были разработаны для защиты сообщений от перехвата или несанкционированного доступа к информации, передаваемой по каналам связи, и передачи данных. Но применение этих методов оказалось шире. Они вполне подходят и для защиты документов (файлов, записей) на любых носителях. Для описания криптографических методов защиты информации лучше всего подходит терминология, заимствованная из области теории систем передачи информации.

Исходное сообщение, информационную стойкость которого нужно обеспечить, называется открытым текстом. В результате шифрации образуется криптограмма (шифрограмма, шифровка). Для шифрации и расшифровки используется ключ. Этот ключ должен быть известен источнику сообщений (передатчику) и получателю (приемнику), причем известен только им одним. Поэтому в традиционных системах секретной связи ключ передается только по очень надежному каналу, особым образом защищенному от утечки информации (например, перевозится в бронированном автомобиле под охраной, в кейсе, пристегнутом наручником к руке курьера). Хотя последние достижения современной криптологии позволяют создавать системы с облегченными требованиями к защите ключа, но с худшими потенциальными характеристиками информационной стойкости (так называемые криптосистемы с открытым ключом, о которых речь пойдет далее). Процесс образования и передачи криптограммы иллюстрируется блок-схемой (рис. 5.13).

Шифратор преобразует исходный текст С и последовательность символов ключа К по правилу

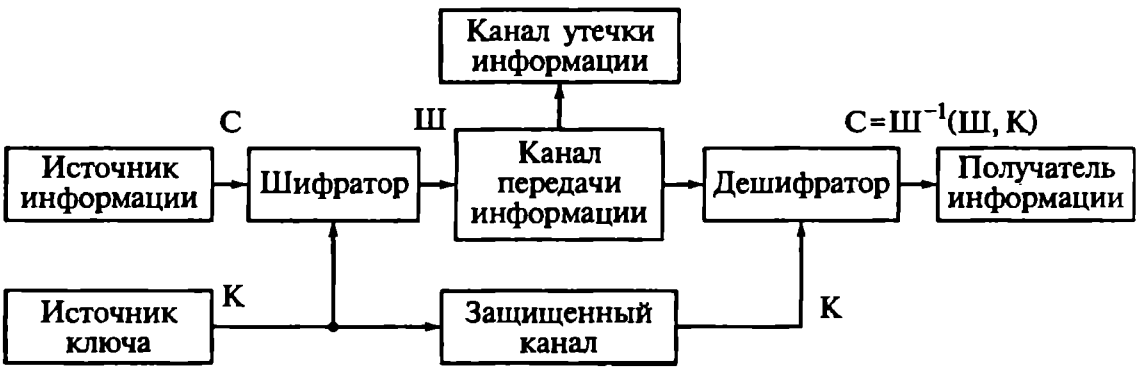


Рис. 5.13. Система передачи информации с секретным ключом

$$Ш = (С + К) \bmod N_С, \qquad (5.51)$$

где $N_С$ — число символов алфавита, которым представляется шифруемый текст.

Чаще всего размеры алфавитов ключа и исходного сообщения одинаковы: $N_С = N_К$. Если системой передачи шифрованной информации используются двоичные сигналы, когда размерность алфавита $N_С = N_К = 2$, правило работы дешифратора выглядит совершенно симметрично (5.51):

$$С = Ш^{-1}(Ш, К) = (Ш + К) \bmod 2 = (Ш \oplus К), \qquad (5.52)$$

где $Ш^{-1}(Ш,К)$ — функция, обратная той, которую реализует шифратор.

Обычно считается, что криптоаналитику известен алгоритм преобразования сообщения в шифраторе, а также полностью доступна криптограмма (это правило Кирхгофа), т.е. считается, что шифрованный сигнал Ш(С,К) достоверно обнаружен, идентифицирован и принят без помех и искажений. Вся неизвестность заключена в исходном открытом тексте С и конкретном выбранном при шифрации ключе. Информацию об открытом тексте можно добыть только на основе знания (статистической) связи перехваченной шифровки и исходного открытого текста. Действия криптоаналитика, осуществляющего несанкционированный доступ к информации, направлены на лучшее использование этой связи. Действия системы защиты информации состоят в таком выборе ключа, чтобы в максимальной степени разрушить связь между шифрограммой и открытым текстом.

Потенциальные, предельно достижимые характеристики доступности смысла и содержания шифрованной информации и, соответственно, характеристики защищенности от этих средств могут определяться на основе положений шенноновской теории связи в секретных системах.

Теоретически достижимую предельную способность шифра обеспечивать защиту информации можно характеризовать условной вероятностью $P_{инф} = P(С|Ш)$, т.е. вероятностью получения открытого текста (сообщения) С при том условии, что была принята криптограмма Ш. У совершенно секретной (по Шеннону) шифросистемы вероятность $P(С|Ш)$ такая же, как и априорная вероятность сообщения С:

$$P(С|Ш) = P(С) \qquad (5.53)$$

для всех возможных криптограмм Ш и сообщений С.

Практически условие (5.53) означает, что шифровка Ш не имеет вероятностной связи с исходным сообщением С и знание шифрограммы не добавляет сведений о сообщении.

Следуя Шеннону, можно в качестве меры неопределенности скрываемого шифром сообщения принять его безусловную энтропию $H(\text{С})$ и условную энтропию $H(\text{С}|\text{Ш})$ при том условии, что криптоаналитик имеет в своем распоряжении результат перехвата шифрованной информации Ш. Естественно, что неопределенность исходного сообщения не уменьшается после получения хоть каких-то сведений, поэтому

$$H(\text{С}) \geq H(\text{С}|\text{Ш}). \qquad (5.54)$$

Если система обеспечивает абсолютную, предельно достижимую информационную стойкость, то $H(\text{С}) = H(\text{С}|\text{Ш})$: прием шифровки не уменьшает неопределенности относительно исходного открытого сообщения, что прямо следует из (5.53).

Шифрограмма формируется при помощи секретного ключа, неизвестного криптоаналитику. По принципу действия именно этот ключ и вносит в шифрограмму неопределенность относительно шифруемого сообщения. Поэтому совместная неопределенность маскируемого сообщения и ключа не меньше, чем неопределенность сообщения

$$H(\text{С}\,|\,\text{Ш}) \geq H(\text{С},\text{К}\,|\,\text{Ш}). \qquad (5.55)$$

Используя определения и известные свойства условной энтропии, следует считать, что

$$H(\text{С},\text{С}\,|\,\text{Ш}) = H(\text{К}\,|\,\text{Ш}) + H(\text{С}\,|\,\text{К},\text{Ш}), \qquad (5.56)$$

но $H(\text{С}\,|\,\text{К},\text{Ш}) = 0$, так как если у криптоаналитика есть и шифровка и ключ к ней, он находится в условиях ничуть не хуже условий законного получателя информации и никакой неопределенности относительно сообщения у него не остается. С другой стороны, знание шифровки не должно добавлять сведений не только о сообщении, но и о ключе, поэтому

$$H(\text{С},\text{К}\,|\,\text{Ш}) = H(\text{К}\,|\,\text{Ш}) \leq H(\text{К}). \qquad (5.57)$$

Неравенство (5.57) отражает уже использованное условие того, что дополнительные данные (наличие перехваченной шифрограммы) не уменьшают неопределенности (энтропии) как открытого сообщения, так и секретного ключа. Или (формально из определения) условная энтропия не может быть больше безусловной. Поэтому, объединяя (5.55) и (5.57), можно получить границу Шеннона для совершенно секретных систем:

$$H(\text{С}) \leq H(\text{К}). \qquad (5.58)$$

В содержательных терминах (5.58) означает, что неопределенность секретного ключа для криптоаналитика должна быть не

меньше неопределенности сообщения, а защищенность информации — предельно достижимой. И если эта граница достигается, вероятность несанкционированного доступа к защищаемой информации оказывается не выше априорной вероятности сообщения.

Ключ — это некоторая последовательность символов. Если k знаков ключа выбираются из алфавита объемом L_k символов, то всего можно сформировать L_k^k разных ключевых последовательностей, обеспечив тем самым

$$H(\mathrm{K}) \le -\sum_k L_k^{-k} \log\left(L_k^{-k}\right) = k \log L_k, \qquad (5.59)$$

причем равенство (5.59) справедливо только для абсолютно случайного выбора ключа, когда вероятность $P(k) = L_k^{-k}$.

Точно также, если шифруемое сообщение представлено m символами алфавита объемом L_c, то

$$H(\mathrm{C}) \le m \log L_c. \qquad (5.60)$$

Соотношения (5.59) и (5.60) совместно устанавливают, что граница Шеннона (5.58) достигается при $k \ge m$, т.е. для достижения потенциальной защищенности информации ключ не должен быть короче шифруемого текста. В частности, из этого условия следует, что ключ нельзя использовать повторно.

Предельные условия, при которых вероятность раскрытия содержания передаваемой информации не превосходит априорной вероятности сообщения, на практике могут и не достигаться. Формально это означает, что условная энтропия ключа уменьшается по мере накопления информации, т.е. по мере увеличения объемов данных перехвата радиоразведкой шифрованных сообщений. Условную энтропию $H(\mathrm{K} \mid \mathrm{Ш}_1, \mathrm{Ш}_2 \ldots \mathrm{Ш}_N)$ можно рассматривать как функцию числа N знаков перехваченных шифровок:

$$H(\mathrm{K} \mid \mathrm{Ш}_1, \mathrm{Ш}_2 \ldots \mathrm{Ш}_N) = H(N). \qquad (5.61)$$

При некотором $N = N_0$ наступают такие условия, при которых $H(\mathrm{K} \mid \mathrm{Ш}_1, \mathrm{Ш}_2 \ldots \mathrm{Ш}_{N0}) = 0$ (точнее, $H(\mathrm{K} \mid \mathrm{Ш}_1, \mathrm{Ш}_2 \ldots \mathrm{Ш}_{N0}) \mid \le \varepsilon$, где ε — оговоренная малая величина). В криптоанализе наименьшее число N_0, для которого выполняются требования малости условной энтропии ключа $H(N_0) \approx 0$, называется расстоянием единственности. Это расстояние показывает, какой длины должна быть перехваченная криптограмма, чтобы по ее анализу можно было бы свести к нулю (приблизительно, но с заданной наперед точностью приближения) неопределенность ключа. В этом смысле N_0 правильнее было бы называть не расстоянием, а длиной единственности.

Неопределенность уменьшается за счет накопления информации, если, конечно, хоть какая-то информация о ключе в криптограмме присутствует, т.е. если $I(\text{К}|\text{Ш}) \neq 0$.

Информации о ключе в шифровке тем больше, чем выше избыточность открытого текста. Действительно, если текст состоит из повторения одного и того же символа (предельно высокая избыточность), то вся криптограмма, в соответствии с (5.51), фактически и есть ключевая последовательность или ее отрезок. Напротив, если открытый текст совершенно случаен и все символы его равновероятны, избыточность равна нулю. В таких условиях, принимая криптограмму, ничего нельзя сказать о ключе.

Естественно, что расстояние единственности должно увеличиваться с увеличением энтропии ключа. В соответствии с принятой Шенноном моделью шифрации, расстояние единственности определяется соотношением

$$N_0 = \frac{H(k)}{\Delta}, \tag{5.62}$$

где $H(k)$ — энтропия ключа; Δ — избыточность открытого текста.

Избыточность открытого текста обусловлена тем, что не все его символы равновероятны, а также тем, что многие символы встречаются в тексте в устойчивых сочетаниях (условные вероятности сочетаний символов открытого текста больше, чем произведения их безусловных вероятностей).

По физическому смыслу и по определению $H(k)$ равно числу знаков в двоичном представлении ключа, а произведение $N_0\Delta$ — числу уравнений, которые можно составить для нахождения каждого неизвестного значения ключа. Для однозначного определения ключа (всех его неизвестных знаков) нужно, чтобы число уравнений было бы не меньше числа неизвестных, т.е. чтобы $N_0\Delta \geq H(\text{К})$, откуда и следует предельное значение N_0 (5.62). Из (5.62) также следует, что для увеличения информационной защищенности сообщений (для усложнения несанкционированной дешифрации) нужно не только увеличивать длину ключа, но и сокращать избыточность открытого текста. Соотношение (5.62) иллюстрирует полезность сжатия данных перед тем как передавать их в шифрованной форме по радиоканалам, защищаемым от перехвата информации средствами радиоразведки. Действительно, избыточность открытого текста количественно определяется как

$$\Delta = 1 - \frac{H(\text{С})}{N \log(L_{\text{С}})}, \tag{5.63}$$

где $H(\text{С})$ — энтропия передаваемого сообщения, составленного из N символов, выбранных из алфавита объемом $L_{\text{С}}$.

Если сообщение С — текст на естественном языке, то для него $\Delta = 0{,}744$ (английский язык) или $\Delta = 0{,}834$ (русский язык). Это значит, что при абсолютно случайном ключе из k символов того же алфавита, в котором представлен открытый текст, для однозначной несанкционированной расшифровки криптоаналитик должен иметь

$$N_0 = \frac{k}{N} = (1{,}19...1{,}11)k \qquad (5.64)$$

символов криптограммы. По такому же количеству символов раскрывается секретный ключ.

Таким образом, хорошие (стойкие к расшифровке) криптосистемы должны устранять избыточность передаваемых сообщений (использовать сжатие данных). Вывод о необходимости сжатия данных за счет устранения избыточности известен еще из донаучной, эвристической криптологии. Идеальных способов сжатия данных нет. Но все применяемые на практике способы используют два основные подхода.

- Из исходного открытого текста удаляются все наиболее часто повторяющиеся символы. Это прежде всего пробелы между словами и другие частые символы. Уже в силу высокой априорной вероятности эти символы малоинформативны: без них нетрудно правильно понять переданное и расшифрованное сообщение. Если иметь в виду шифрованные тексты на естественных языках, самыми избыточными и потому опасными с точки зрения сохранения криптостойкости являются служебные пометки (подписи, даты, адреса, грифы секретности и пр.). Чем длиннее эти пометки, чем больше они содержат символов, тем ниже стойкость криптограммы и, что еще хуже, секретного ключа, которым она зашифрована.
- Увеличивается энтропия шифрованного сообщения. Для этого в исходном открытом тексте разравниваются вероятности различных символов. Иначе говоря, распределение вероятностей символов в шифруемом тексте делается по возможности более близким к равномерному. В текстах на русском языке чаще других попадается буква «О», в английских текстах — «Е». Разравнивание вероятностей достигается за счет рандомизации (когда исходный текст складывается по модулю 2 со специальной не очень длинной последовательностью символов) или за счет применения многоалфавитных подстановок и перестановок.

При многоалфавитных подстановках открытый текст шифруется несколько раз, последовательно. Каждый раз символы шифруемого текста заменяются другими символами, выбранными из того же или другого алфавита. В результате многократного применения таких подстановок относительные частоты появления сим-

волов в криптограмме уже не отражают вероятностей появления символов в исходном тексте на естественном языке. Если распределение вероятностей символов становится точно равномерным, шифрованный текст приобретает максимальную энтропию и, следовательно, минимальную избыточность. В соответствии с (5.63) такая криптосистема будет иметь максимальное расстояние единственности, а значит, и наивысшую при используемом ключе криптостойкость. Практически при шифре с равновероятными символами криптоаналитик не сможет использовать для несанкционированной расшифровки частотный анализ криптограммы.

Перестановки перемешивают символы исходного открытого текста, причем способ перемешивания определяется секретным ключом, известным только законным абонентам системы передачи информации. При перестановках частоты появления отдельных символов в шифровке не изменяются по сравнению с соответствующими частотами в исходном открытом тексте, но статистические связи разрушаются.

Расстояние единственности (5.63) — это теоретическая мера стойкости шифра, исходящая из предположений о том, что криптоаналитик при расшифровке действует некоторым наилучшим для себя образом. Но такая характеристика совершенно не учитывает того, каким ресурсом должен обладать криптоаналитик для успешного раскрытия шифра по криптограммам с заданным расстоянием единственности. Поэтому рабочая характеристика шифра определяется $W(N)$ как средний объем работы (в часах, машинных операциях или других удобных единицах для ЭВМ известного типа и класса), необходимой для криптоанализа и раскрытия криптограммы на основе N знаков шифрованного текста. При этом $W(N)$ определяется для наилучшего криптоаналитического алгоритма.

Наиболее интересна потенциальная оценка рабочей характеристики $W(\infty)$, представляющая средний объем работы по криптоанализу при неограниченном объеме шифрованного текста. Применяя эту оценку, обычно говорят и пишут «шифр требует для раскрытия («взлома») стольких-то лет», а имеют в виду, что при неограниченном количестве знаков перехваченной криптограммы, наилучшим из известных алгоритмов криптоанализа и использовании самой быстродействующей из известных ЭВМ нужно затратить столько-то лет непрерывной работы для раскрытия шифра. Это оценка не доверительной вероятности успеха несанкционированной расшифровки криптограммы $P_{\text{инф}}$, а доверительного интервала времени, по истечении которого раскрытие шифра (ключа и открытого текста) произойдет с вероятностью $P_{\text{инф}} = 1$.

Осознание различия между практической и теоретической стойкостью криптосистем позволило поставить неожиданный и, на первый взгляд, парадоксальный вопрос: раз уж имеет смысл стремиться к обеспечению только практической стойкости шифра,

нельзя ли ее достичь при отказе от сложностей создания и распространения секретного ключа? Положительный ответ на этот вопрос позволяет существенно упростить криптосистему за счет отказа от специального защищенного канала передачи ключа.

Так были созданы двухключевые криптоалгоритмы или, иначе, алгоритмы шифрации с открытым ключом. Особенность таких асимметричных криптосистем состоит в том, что для шифрования они используют один ключ из пары ключей открытого и секретного, а для расшифровки — другой ключ.

Классификация (далеко не полная и приблизительная) алгоритмов шифрования, используемых в настоящее время для защиты информации, иллюстрируется графом на рис. 5.14.

Блочные шифры предусматривают разбиение исходного текста на блоки фиксированной длины и шифрацию каждого блока. При этом возможна шифрация за счет перестановок символов исходного открытого текста по правилу неизвестного для противника ключа или за счет замены символов исходного текста другими символами, выбранными из того же или другого алфавита. Так, уже упомянутый шифр, который применял Цезарь, был шифром замены: каждый символ С исходного текста, представленного символами латинского алфавита, заменялся также латинскими буквами по правилу $(C + 3) \bmod 36$. Модуль 36 — это размерность латинского алфавита (полное число символов). Как шифры замены, так и шифры перестановок в настоящее время в чистом виде не применяются, реальные современные криптоалгоритмы используют их комбинацию.

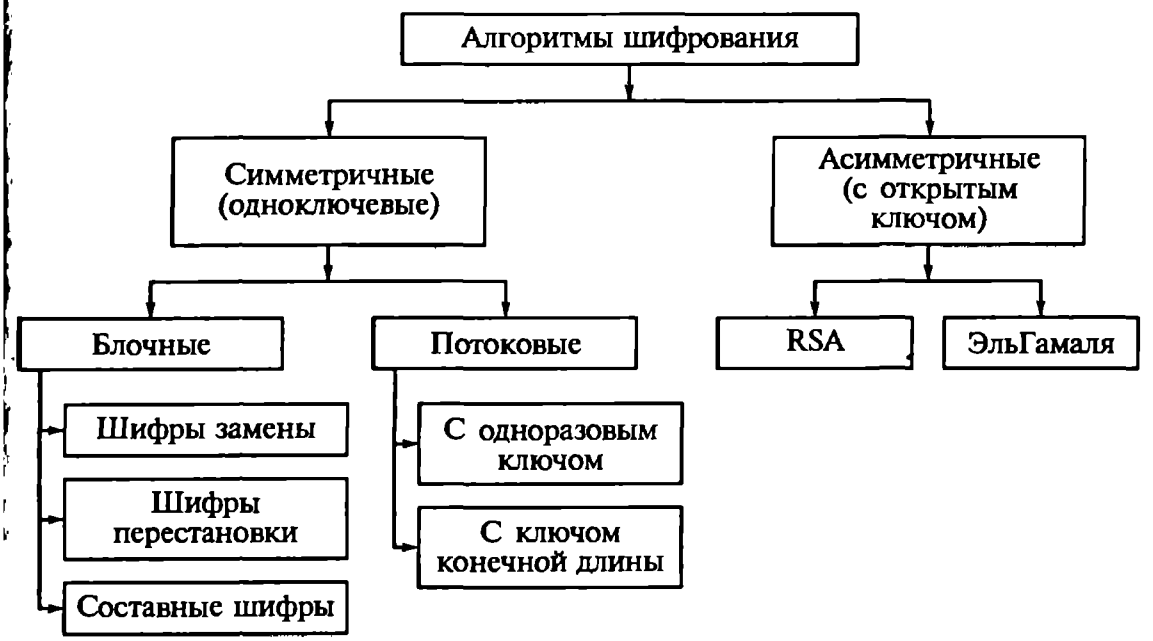


Рис. 5.14. Алгоритмы криптографической защиты информации

Главное свойство блочного шифрования состоит в том, что каждый символ блока текста шифровки является функцией всех или почти всех символов соответствующего блока открытого текста, и никакие два блока открытого текста не могут быть представлены одним и тем же блоком текста шифровки. Основное преимущество простого блочного шифрования состоит в том, что в хорошо сконструированной системе небольшие изменения открытого текста или ключа вызывают большие и непредсказуемые изменения в тексте шифра. Однако при употреблении блочные шифры не свободны от серьезных недостатков. Во-первых, если ко всем блокам применить один и тот же ключ, то даже при сравнительно большой длине блока возможен криптоанализ на основе поиска и обнаружения стандартной текстовой комбинации.

Другой потенциальный недостаток этого шифра связан с размножением ошибок внутри блока. Результатом изменения одного символа в принятом блоке шифровки будет неправильное расшифровывание всего блока. Вследствие отмеченных недостатков простейшие блочные шифры не употребляются для шифрования длинных сообщений. Это недостаток шифра, который проявляется при повторных передачах шифрованных сообщений.

От недостатков блочных шифров свободны поточные шифры, в которых шифрующее преобразование каждого символа исходного сообщения меняется от символа к символу. У большинства поточных шифров секретный ключ К не сам изменяет сообщение в процессе шифрации, а управляет работой генератора ключевого потока. И уже этот генератор формирует последовательность (поток) символов $\{K_1K_2....K_N\}$, взаимодействующих с символами шифруемого сообщения по правилу:

$$Ш_n = C_n \oplus K_n;\ n \in 1:N. \tag{5.65}$$

Так образуется линейный поточный шифр. Поскольку операции сложения и вычитания по модулю 2 совпадают:

$$C_n = (Ш_n)^{-1} = Ш_n \oplus K_n;\ n \in 1:N, \tag{5.66}$$

одинаковыми оказываются схемы шифраторов и дешифраторов.

Длина N генерируемой под управлением ключа последовательности может быть гораздо больше длины ключа. Если величина N очень велика (ключевая последовательность не короче исходного шифруемого сообщения), может показаться, что для такого поточного шифра справедлива граница Шеннона и он оказывается принципиально не раскрываемым, т.е. совершенно секретным. Но это не совсем так. Для совершенной секретности требуется, чтобы длина шифруемого сообщения была не короче длины секретного ключа, а не порождаемой им последовательности ключевого потока.

Для генераторов ключевого потока стоит проблема линейной сложности — проблема определения структуры обратных связей в генераторе на основе максимально короткого регистра сдвига с тем, чтобы получить ключевую последовательность максимальной длины. Большая линейная сложность — это необходимое условие криптостойкости системы с линейным поточным шифром. Разрешение этой проблемы сводится либо к выбору регистра-генератора очень большой длины, либо к применению таких ключевых потоков, в которые нелинейно объединяются последовательности с выходов нескольких независимых регистров-генераторов. Линейная сложность сформированной таким образом поточной ключевой последовательности может быть очень большой только в том случае, если последовательности разных генераторов некоррелированы между собой и незначительна корреляция каждой из них с результирующей последовательностью.

В криптографических приложениях такая последовательность называется гаммой, а метод шифрования — гаммированием.

При однократном использовании абсолютно случайной для незаконного получателя гаммы ключа, превосходящей по длине шифруемое сообщение, шифр обладает потенциальной стойкостью. Если ключ короче открытого сообщения, стойкость шифра далека от потенциально достижимой.

Понятно, что распространение ключей, не менее коротких, чем открытый шифруемый текст, очень затруднено, часто невозможно. Поэтому в качестве ключа передается начальное значение установки генератора псевдослучайной последовательности гаммы. Но тогда длиной секретного ключа оказывается уже не период псевдослучайной последовательности, а длина последовательности символов, определяющих начальное состояние регистра-генератора псевдослучайной последовательности. Для сложных генераторов, объединяющих несколько регистров или создающих нелинейные последовательности, эта длина больше, чем у генераторов линейных рекуррентных псевдослучайных последовательностей, но и у них она короче периода гаммы.

Возможно образование смешанных систем потокового и блочного шифрования с использованием лучших свойств каждого из этих шифров. В таких системах потоковое шифрование комбинируется с перестановками символов в блоках. Открытый текст сначала шифруется, как при обычном потоковом шифровании, затем полученный текст шифровки разбивается на блоки фиксированного размера и в каждом блоке дополнительно производится перестановка под управлением ключа. В результате получается шифр, не размножающий ошибки, но обладающий дополнительным свойством, которого нет у примитивного потокового шифра замены на основе операции суммирования по модулю 2, состоящим в том, что криптоаналитик не знает, какому биту открыто-

го текста соответствует бит текста шифровки. Благодаря этому за шифрованное сообщение становится гораздо более стойким для раскрытия.

5.5. Стандарты симметричных криптосистем

Типичные и стандартизованные алгоритмы симметричного шифрования (одноключевые) — это DES (Data Encryption Standard), служащий стандартом шифрования США, европейский стандарт IDEA (International Data Encryption Algorithm) и российский стандарт ГОСТ 28147—89. Эти блочные симметричные криптоалгоритмы используют такие обратимые преобразования открытых сообщений, при которых значение, вычисленное по одной части шифруемого текста, накладывается на другие части текста. Методика шифрации обеспечивает многократное использование и взаимодействие ключа с символами исходного текста. За счет такого многократного рассеяния и перемешивания исходных данных почти полностью разрушаются корреляционные связи между символами исходного текста, а также символами этого текста и символами шифровки и исходного текста.

Блочные симметричные алгоритмы допускают как аппаратную, так и программную реализацию. Они осуществляют криптографическое преобразование информации для хранения на любых носителях (бумажных, магнитных и других машинных), для передачи данных в сетях ЭВМ. Алгоритм по ГОСТ 28147—89 не накладывает ограничений на степень секретности защищаемой информации.

Алгоритмы DES, IDEA и ГОСТ 28147—89 предусматривают несколько режимов работы, но во всех режимах для шифрации используют секретный ключ, единый для шифрации и расшифровки. Отечественный алгоритм, регламентируемый ГОСТом, работает с ключом длиной 256 бит, образованным конкатенацией (сцеплением) восьми 32-разрядных двоичных чисел K_i:

$$K = K_7 \# K_6 \# K_5 \# K_4 \# K_3 \# K_2 \# K_1 \# K_0. \qquad (5.67)$$

Алгоритм DES использует 64-разрядный ключ, а европейский алгоритм IDEA — 128-разрядный.

Первый режим работы алгоритма шифрации — замена символов открытого текста. Шифруемые данные С, представленные последовательностью двоичных символов, разбиваются на блоки C_n длиной 64 бита так, что каждый блок является конкатенацией двух подблоков: $СЛ_n$, содержащего старшего (левые) 32 бита, и $СП_n$, объединяющего младшие (правые) биты блока C_n, т.е. $C_n = СЛ_n \# СП_n$. Если длина открытого текста С не кратна 64, он до-

полняется нулями так, что шифрации всегда подвергается целое число блоков C_n, $n \in 1:N$ (четное число подблоков).

После разбиения исходного текста на блоки выполняется итеративный процесс шифрования в соответствии со следующим правилом:

$$\begin{cases} \text{СП}_i = F\left\{\left[\text{СП}_{i-1} + \left(\text{K}_j \oplus \text{СЛ}_{i-1}\right)\right] \bmod 2^{32}\right\}; \\ \text{СЛ}_i = \text{СП}_{i-1}, \\ \text{если } i \in 1{:}24,\ j = (i-1) \bmod 8; \end{cases}$$

$$\begin{cases} \text{СП}_i = F\left\{\left[\text{СП}_{i-1} + \left(\text{K}_j \oplus \textit{СЛ}_{i-1}\right)\right] \bmod 2^{32}\right\}; \\ \text{СЛ}_i = \text{СП}_{i-1}, \\ \text{если } i \in 25:31,\ j = 32 - i; \end{cases} \qquad (5.68)$$

$$\begin{cases} \text{СП}_{32} = \text{СП}_{31}; \\ \text{СЛ}_i = F\left\{\left[\text{СП}_{31} + \left(\text{K}_0 \oplus \text{СЛ}_{31}\right)\right] \bmod 2^{32}\right\}, \\ \text{если } i = 32, \end{cases}$$

где i — номер итерации, для ГОСТ 28147—89 $i \in 1:32$, для DES $i \in 1:16$, для IDEA $i \in 1:4$; $F(\cdot)$ — функция шифрования; $\oplus$ — символ суммирования по модулю 2.

Графическая форма представления криптоалгоритма (5.68) представлена схемой на рис. 5.15.

Аргумент функции шифрования $F(\cdot)$ — это сумма по модулю 2 числа СП_{i-1}, полученного на предыдущем шаге итерации шифрования, и 32-разрядного фрагмента ключа K_j. Вычисление функции шифрования предусматривает последовательное выполнение двух операций над полученной 32-разрядной суммой. Первая операция — подстановка. Блок подстановки состоит из восьми узлов замены $Q_1 \ldots Q_8$ с памятью 64 бит каждый. Поступающая на блок подстановки 32-разрядная последовательность двоичных символов разбивается на восемь последовательно идущих четырехразрядных групп, каждая из которых преобразуется в новую четырехразрядную группу соответствующим узлом замены. Каждый узел замены — это таблица из шестнадцати целых чисел в диапазоне 0...15. Входная последовательность определяет адрес строки в таблице. По этому адресу находится число, являющееся выходной последовательностью. Конкатенация трансформированных таким образом в результате выполнения операции подстановки четырехразрядных выходных групп символов представляет собой 32-разрядные группы символов, сдвинутые на 11 дво-

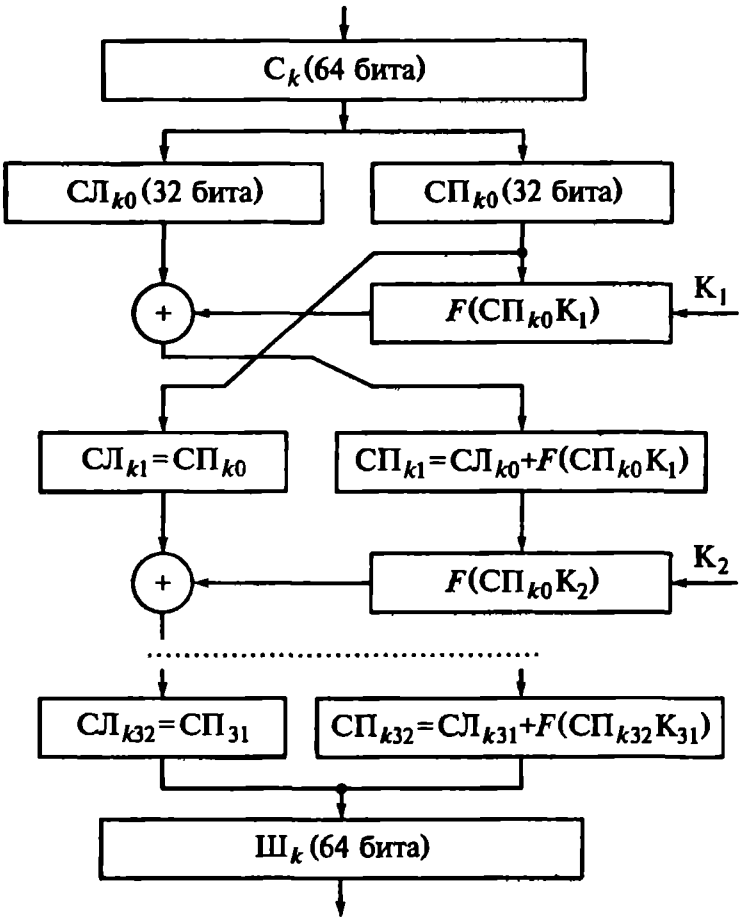


Рис. 5.15. Алгоритм шифрации по ГОСТ 28147—89

ичных разрядов влево. Таблица блока подстановки содержит ключевые элементы, общие для сети шифрующего и расшифровывающего алгоритмов.

Схема вычисления функции шифрования представлена на рис. 5.16. Операция $\Leftarrow 2^{11}$ на схеме означает циклический сдвиг входной последовательности на 11 разрядов влево.

Таким образом, в каждом цикле шифрования используется раундовый ключ. В DESe он содержит 48 бит и вырабатывается по относительно сложному алгоритму, предусматривающему битовые перестановки и замену по таблице. Эти операции легко выполняются аппаратурно, но довольно сложны (требуют значительных затрат времени) при программной реализации. В ГОСТе раундовый ключ — это просто часть (одна из восьми) ключа шифрования.

Вторая операция — циклический сдвиг влево 32-разрядной последовательности, полученной в результате подстановки. Выходной 64-разрядный блок зашифрованных данных формируется в виде

$$Ш_k = СЛ_k(32) \# СП_k(32). \qquad (5.69)$$

Остальные блоки открытых данных в режиме простой замены зашифровываются аналогично.

Следует иметь в виду, что режим простой замены допустим для шифрования данных только в ограниченных случаях. К этим случаям относятся выработка ключа и зашифровывание его с обеспечением имитозащиты для передачи по каналам связи или хранения в памяти ЭВМ.

Следующий режим шифрования — наложение гаммы (гаммирование). Отрытые данные, разбитые на 64-разрядные блоки C_n, $n \in 1:N$, поразрядно складываются по модулю 2 с гаммой шифра $\gamma_ш$, которая вырабатывается блоками по 64 бит, т. е.

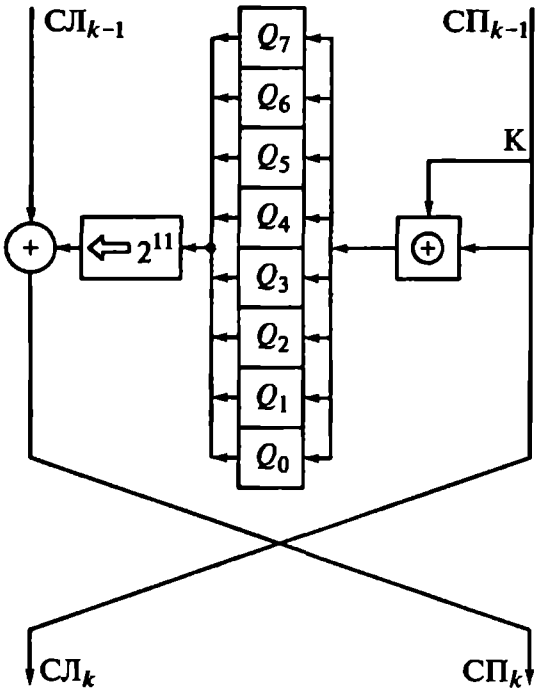


Рис. 5.16. Вычисление функции $F(\cdot)$ в каждом раунде итерационной процедуры шифрования

$$\gamma_ш = \gamma_1 \# \gamma_1 \# \ldots \gamma_1 \# \ldots \gamma_N. \qquad (5.70)$$

Длина блока C_N может быть меньше 64 бит. В этом случае неиспользованная для шифрования часть гаммы шифра из блока $\gamma_ш$ отбрасывается.

Формирование шифрованного текста в режиме гаммирования выполняется итеративно в соответствии с уравнением

$$Ш_i = A\{(Y_{i-1} + D_2) \bmod 32, (Z_{i-1} + D_1 \oplus C_i) \bmod 31\} = C_i \oplus \gamma_i, \qquad (5.71)$$

где A — функция шифрования в режиме простой замены; D_1 и D_2 — константы, заданные в алгоритме ГОСТ 28147—89; Y_i и Z_i — величины, участвующие в итерационной процедуре формирования гаммы:

$$(Y_i, Z_i) = [(Y_{i-1} + D_2) \bmod 2^{32}, (Z_{i-1} + D_1) \bmod 2^{32} - 1];$$
$$[Y_0, Z_0] = A(S), \qquad (5.72)$$

где S — 64-разрядная синхропосылка.

Расшифровка данных возможна только при наличии синхропосылки S, которая не является секретным элементом шифра и может храниться в памяти ЭВМ или передаваться по каналам связи вместе с зашифрованными данными.

В режиме гаммирования с обратной связью, как и в режиме гаммирования, исходные открытые данные, разбитые на 64-разрядные блоки, шифруются путем поразрядного сложения по модулю 2 с гаммой шифра $\gamma_{ш}$, которая вырабатывается блоками по 64 бит. Но аргументом функции шифрования на первом шаге итеративного алгоритма служит синхропосылка:

$$Ш_i = A(S) \oplus C_1 = \gamma_1 \oplus C_i, \; Ш_1 = A(S) \oplus C_1 = \gamma_1 \oplus C_1. \quad (5.73)$$

Алгоритм ГОСТ 28147—89 предусматривает процесс формирования имитовставки. Этот процесс одинаков для любого из режимов шифрования данных. Имитовставка *Ир* — криптографическая контрольная комбинация, предназначенная для защиты шифрограммы от изменений (случайных, вызванных помехами, или преднамеренных, обусловленных несанкционированным вмешательством). *Ир* представляет собой блок из *p* двоичных символов, который вырабатывается либо перед шифрованием всего сообщения, либо параллельно с шифрованием по блокам. Первые блоки открытых данных, которые участвуют в выработке имитовставки, могут содержать служебную информацию (например, адресную часть, время, синхропосылку) и не зашифровываться.

Для получения имитовставки открытые данные, представленные первым 64-разрядным блоком C_1, подвергаются преобразованию, соответствующему первым 16 циклам алгоритма шифрации в режиме простой замены. При этом в качестве ключа для выработки имитовставки используется ключ, по которому шифруются данные.

Полученное после 16 циклов работы 64-разрядное число суммируется по модулю 2 со вторым блоком открытых данных C_2. Результат суммирования снова подвергается преобразованию, соответствующему первым 16 циклам алгоритма зашифрования в режиме простой замены.

Сформированное таким образом 64-разрядное число суммируется по модулю 2 с третьим блоком открытых данных C_3 и т. д. Последний блок C_N, при необходимости дополненный до полного 64-разрядного блока нулями, суммируется по модулю 2 с результатом работы на $N-1$ шаге, после чего зашифровывается в режиме простой замены по первым 16 циклам работы алгоритма. Из полученного 64-разрядного числа выбирается отрезок *Ир* длиной *p* бит.

Имитовставка *Ир* передается по каналу связи или заносится в память ЭВМ после зашифрованных данных. По мере расшифровки данных из полученных блоков открытых данных C_n вырабатывается имитовставка, которая затем сравнивается с имитовставкой *Ир*, переданной или сохраненной вместе с шифровкой. Алгоритм формирования имитовставки при расшифровке тот же, что

и при шифрации, поэтому одинаковые данные обуславливают полное совпадение имитовставок. В случае несовпадения имитовставок все расшифрованные данные считают ложными.

Использование алгоритмом ГОСТ 28147—89 имитовставки повышает его стойкость к подделкам и искажениям.

В отличии от DES и IDEA, алгоритм ГОСТ не предусматривает начальных и конечных битовых перестановок в блоках $С_п$ и $Ш_п$. Это упрощает его реализацию без усложнения стойкости к расшифровке.

5.6. Двухключевые криптосистемы (криптосистемы с открытым ключом)

В традиционных (одноключевых) криптосистемах одним и тем же секретным ключом и шифруется, и расшифровывается сообщение. Для этого отправитель и получатель сообщения должны располагать идентичными копиями ключа, который передается по особым образом защищенному каналу передачи данных, что значительно усложняет криптосистему. Но осознание различия между теоретической и практической стойкостью криптосистем позволило поставить неожиданный и, на первый взгляд, парадоксальный вопрос: раз уж имеет смысл стремиться к обеспечению только практической стойкости шифра, нельзя ли ее достичь при отказе от сложностей создания и распространения секретного ключа? Поэтому в последние годы были попытки создания систем с открытым распространением ключа. Некоторые из них увенчались успехом.

Для шифрации с открытым ключом применяются криптопреобразования на основе односторонних функций с потайным ходом. Это такие функции F с параметром z, что для данного z можно найти алгоритмы E_z и D_z, позволяющие легко вычислить значение $F_z(x)$ для всех x из области определения $F_z(x)$, а также $F_z^{-1}(x)$ для всех y из области значений. Однако практически для всех значений параметров z и y нахождение $F_z^{-1}(x)$ при неизвестном D_z вычислительно неосуществимо.

Предложено множество односторонних функций. Некоторые из них оказались ненадежными, другие перспективными. Но никому пока не удалось доказать, что какая-то функция является односторонней или односторонней с потайным ходом. Даже стойкость общепризнанной системы RSA основана на недоказанном (хотя и очень правдоподобном) допущении о том, что разложение больших чисел на множители вычислительно неосуществимо.

В криптосистемах с открытым ключом для шифрования и расшифровывания используются разные ключи, и знание одного из

них не дает возможности (практической) определить второй. Поэтому ключ для шифрования может быть сделан общедоступным без потери стойкости шифра, если ключ для расшифровывания сохраняется в секрете, например генерируется и хранится только получателем информации.

В настоящее время два метода шифрования с открытым ключом получили признание и закреплены в стандартах. Национальный институт стандартов и технологий США NIST принял стандарт MD 20899, основанный на алгоритме ЭльГамаля, а на основе алгоритма RSA приняты стандарты ISO/IEC/DIS 9594-8 международной организацией по стандартизации и X.509 международным комитетом по связи.

Криптографическая система с открытым ключом Ривеста — Шамира — Алдемана (RSA). Она основана на трудности разложения очень больших целых чисел на простые сомножители. Алгоритм ее работы предусматривает следующие действия:

1. Источник сообщения выбирает два очень больших простых числа p и q и вычисляет два произведения $n = pq$ и $m = (p-1)(q-1)$, а также некоторое целое число d, взаимно простое с m. На основе этой тройки он вычисляет e, удовлетворяющее условию

$$(de) \bmod m = 1,\ 1 < e < (p-1)(q-1). \tag{5.74}$$

2. Число e сохраняется в секрете, а d и e сообщаются всем абонентам как открытый ключ шифрования (публикуются в справочнике вроде телефонного).

3. Сообщение С, длина которого, определяемого по длине выражаемого им целого числа, находится в интервале $[1; n]$, преобразуется в шифрограмму по правилу

$$Ш = C^d \bmod n. \tag{5.75}$$

4. Получатель сообщения расшифровывает его, возводя шифровку в степень e по модулю n:

$$C = Ш^e \bmod n = C^{de} \bmod n. \tag{5.76}$$

Авторы RSA в примере из своей первой публикации использовали $d = 9007$ и $n = 11438162575788886766923577997614661201021829672124236256256184293570693524573389783059712356395870505898907514 7599290026879543541$. Но для иллюстрации работы алгоритма шифрации RSA в [9] рассматривается более простой пример на малых простых числах $p = 211$ и $q = 223$. В этом случае $n = 47053$ и $m = 46620$. При выборе открытого ключа шифрования $d = 6813$ вычисляется секретный ключ расшифровки $e = 19837$. Теперь, взяв за сообщение название метода RSA, его следует пере-

вести в число. Для этого будем буквам латинского алфавита ставить в соответствие их порядковые номера. Поэтому $R:=18$, $S:=19$, $A:=1$. На двоичное представление каждой буквы отводится по 5 бит, как в коде Бодо. Поэтому сообщению RSA соответствует число $C = ((1 \cdot 32) + 19) \cdot 32 + 18 = 1650$. С помощью открытого ключа получателя это сообщение превращается в шифровку

$$Ш = C^e \bmod n = 1650^{16813} \bmod 47053 = 3071. \qquad (5.77)$$

Получатель шифровки преобразует ее с помощью своего секретного ключа.

Криптостойкость системы RSA основана на том, что m не может быть просто вычислена без знания простых сомножителей p и q, а нахождение этих сомножителей из n считалось трудно разрешимой задачей. Однако недавние работы по разложению больших чисел на сомножители показали, что для этого могут быть использованы разные и даже совершенно неожиданные средства. Сначала авторы RSA предлагали выбрать простые числа p и q случайно, по 50 десятичных знаков каждое. Считалось, что такие большие числа очень трудно разложить на простые сомножители при криптоанализе. Оказалось, что это не так.

Теперь разработчикам криптографических алгоритмов с открытым ключом на базе RSA приходится избегать применения разложимых чисел длиной менее 200 десятичных разрядов. Самые последние публикации предлагают для этого применять числа в 250 и даже 300 десятичных разрядов. А так как для шифрования каждого блока информации приходится соответствующее число возводить в колоссально большую степень по модулю n, то для современных компьютеров это задача на грани возможного. Поэтому для практической реализации шифрования RSA приходится разрабатывать специальные процессоры. Чрезвычайно слабое быстродействие криптографических систем на основе RSA лишь ограничивает область применения, но вовсе не перечеркивает их ценность.

Шифр ЭльГамаля. Он использует схему на основе возведения в степень по модулю большого простого числа. Для этого задается большое простое число p. Сообщения представляются целыми числами C из интервала $(1, p)$. Оригинальный протокол передачи сообщения C выглядит следующим образом.

Отправитель сообщения A и получатель B знают лишь p. A генерирует случайное число X из интервала $(1, p)$. B генерирует случайное число Y из того же интервала.

A шифрует сообщение, формируя криптограмму

$$Ш_A = C^X \bmod p \qquad (5.78)$$

и посылает B.

В шифрует его своим ключом

$$Ш_B = (Ш_A)^Y \bmod p \tag{5.79}$$

и посылает $Ш_B$ абоненту А.

А снимает свой ключ

$$Ш_{AA} = (Ш_B)^{-X} \bmod p \tag{5.80}$$

и возвращает сообщение абоненту В.

Получатель В расшифровывает сообщение:

$$C = (Ш_{AA})^{-Y} \bmod p. \tag{5.81}$$

Криптосистема ЭльГамаля обеспечивает бóльшую степень защиты, чем алгоритм RSA. Этот эффект достигается при том же N, что позволяет почти на порядок увеличить скорость шифрования и расшифровывания. Криптостойкость системы ЭльГамаля основана на том, что можно легко вычислить степень целого числа, т. е. произвести умножение его самого на себя любое число раз так же, как и при операциях с обычными числами. Однако трудно найти показатель степени, в которую нужно возвести заданное число, чтобы получить другое, тоже заданное. В общем случае эта задача дискретного логарифмирования кажется более трудной, чем разложение больших чисел на простые сомножители, на основании чего можно предположить, что сложности вскрытия систем RSA и ЭльГамаля будут сходными.

С точки зрения практической реализации, как программным, так и аппаратным способом ощутимой разницы между этими двумя стандартами нет. Однако в криптостойкости они заметно различаются. Если рассматривать задачу разложения произвольного целого числа длиной в 512 бит на простые множители и задачу логарифмирования целых чисел по 512 бит, вторая задача, по оценкам математиков, несравненно сложнее первой. Однако есть одна особенность. Если в системе, построенной с помощью алгоритма RSA, криптоаналитику удалось разложить открытый ключ n одного из абонентов на два простых числа, то возможность злоупотреблений ограничивается только этим конкретным пользователем. В системе, построенной с помощью алгоритма ЭльГамаля, угрозе раскрытия подвергнутся все абоненты криптографической сети. Кроме того, появились возможности существенно усовершенствовать методы дискретного логарифмирования для отдельных специальных простых чисел.

Преимущества двухключевых систем могли бы привести к полному вытеснению криптосистем с секретными ключами из большинства сетевых приложений, если бы не очень низкая скорость криптопреобразований: они работают на несколько порядков медленнее систем с открытыми ключами. Этот недостаток тем более

существенен, чем длиннее шифруемое сообщение. Противоречие разрешается очень просто, если применять открытое распространение сравнительно коротких секретных ключей, действующих непродолжительное время, например в течение одного сеанса связи. Алгоритм формирования секретного сеансового ключа блочного шифра основывается на использовании односторонней функции дискретного возведения в степень и сводится к следующему:

$$Ш = Ш(С) = a^{C} (\mathrm{mod}\, p). \qquad (5.82)$$

Если a и p известны, то сообщение С нетрудно зашифровать в соответствии с этим алгоритмом (нетрудно получить Ш как результат дискретного возведения в степень (13.17)). Даже при очень больших p Ш(С) вычисляется в результате применения нескольких операций возведения в квадрат и умножения. Например,

$$a^{53} = a^{32+16+4+1} = a \cdot a^2 \cdot (a^2)^2 \cdot (((a^2)^2)^2)^2 \cdot ((((a^2)^2)^2)^2)^2 \qquad (5.83)$$

требует выполнить пять операций умножения типа $a^n \cdot a$ и три операции перемножения полученных величин. Всего для вычисления a^p потребуется примерно $2\log 2p$ (и не более того) операций умножения. Расшифровка для определения С при известном Ш, но неизвестных a и p потребует вычисления обратной функции

$$С = \log_a Ш(С) \ (\mathrm{mod}\, p), \qquad (5.84)$$

т. е. дискретного логарифмирования. Доказано, что если не только p велико, но и $(p - 1)$ имеет большой простой множитель (например, если $0{,}5(p - 1)$ — простое число), вычисление дискретного логарифма потребует примерно $\sqrt{p}$ операций умножения. Разумеется,

$$\sqrt{p} >> 2\log 2p, \qquad (5.85)$$

и функция дискретного возведения в степень при некоторых условиях на a, p и оговорках, сделанных относительно $(p - 1)$, действительно является односторонней функцией. Эти условия сводятся к следующим: число p должно быть простым, а число $a \in [1; p]$ таким, что все его степени (по mod p) принимают значения из множества $[1 : p - 1]$. Иначе говоря, a должно быть примитивным элементом поля Галуа GF(p); такие a всегда существуют. Например, для $p = 7$ и $a = 3$: $a^1 = 3$; $a^2 = 2$; $a^3 = 6$; $a^4 = 4$; $a^5 = 5$; $a^6 = 1$ $(\mathrm{mod}\, p = 7)$.

Работу криптосистемы с открытым ключом можно иллюстрировать на примере обмена шифрованными сообщениями между двумя абонентами. Условно это абоненты А и Б. Предположим, эти абоненты желают передать друг другу конфиденциальные сообщения $C_А$ и $C_Б$ соответственно. Для организации такого обмена

абонент А выбирает случайное число $X_A \in [1; p-1]$ и держит его в секрете, но вычисляет значение дискретной экспоненты:

$$П_A = a^{X}_{A} (\mathrm{mod}\, p). \qquad (5.86)$$

Число $П_A$ сообщается всем, с кем абонент А собирается устанавливать связь. Можно сказать, что в системе связи $П_A$ — это такой же реквизит абонента А, как имя, адрес и номер телефона. Точно также поступает и абонент Б, но, разумеется, выбирая другое число $Х_Б$ и вычисляя другое $П_Б$.

Если А и Б обмениваются конфиденциальными сообщениями, каждый из них вычисляет

$$К_{АБ} = a^{X_А X_Б} = (П_А)^{X_Б} (\mathrm{mod}\, p) = (П_Б)^{X_А} (\mathrm{mod}\, p) \qquad (5.87)$$

и используют его для шифровки и дешифровки сообщений подобно обычному секретному ключу, т. е. абонент А формирует криптограмму $Ш_A$ из сообщения C_A по правилу:

$$Ш_A = (С_A + К_{АБ}) (\mathrm{mod}\, p), \qquad (5.88)$$

а абонент Б, получив $Ш_A$, восстанавливает (расшифровывает) открытый текст с использованием того же вычисленного им ключа $К_{АБ}$, так как

$$С_A = (Ш_A + К_{АБ}) (\mathrm{mod}\, p). \qquad (5.89)$$

Совершенно аналогично происходит передача шифрованных сообщений от Б к А:

$$Ш_Б = (С_A + К_{АБ}) (\mathrm{mod}\, p), \qquad (5.90)$$

поскольку

$$С_Б = (Ш_Б + К_{АБ}) (\mathrm{mod}\, p). \qquad (5.91)$$

Как видно из (13.22), оба абонента могут образовывать идентичные ключи для защиты информации при обмене сообщениями. Причем для каждой пары абонентов сети секретной связи будет формироваться свой ключ, неизвестный и недоступный любой другой паре (даже если в эту пару войдут порознь либо абонент А, либо Б). Работа системы секретной связи с открытым ключом на основе дискретного возведения в степень иллюстрируется блок-схемой (рис. 5.17).

Если некто третий (криптоаналитик, работающий на радиоразведку) попытается перехватить сообщение, ему прежде всего придется по шифровке Ш определить ключ $К_{АБ}$. Но для вычисления значения ключа ему необходимо знать либо X_A, либо $X_Б$. Знание любой из этих величин позволит вычислить $К_{АБ}$, но

$$X_A = \log_a П_Б (\mathrm{mod}\, p) \text{ и } X_Б = \log_a П_A (\mathrm{mod}\, p), \qquad (5.92)$$

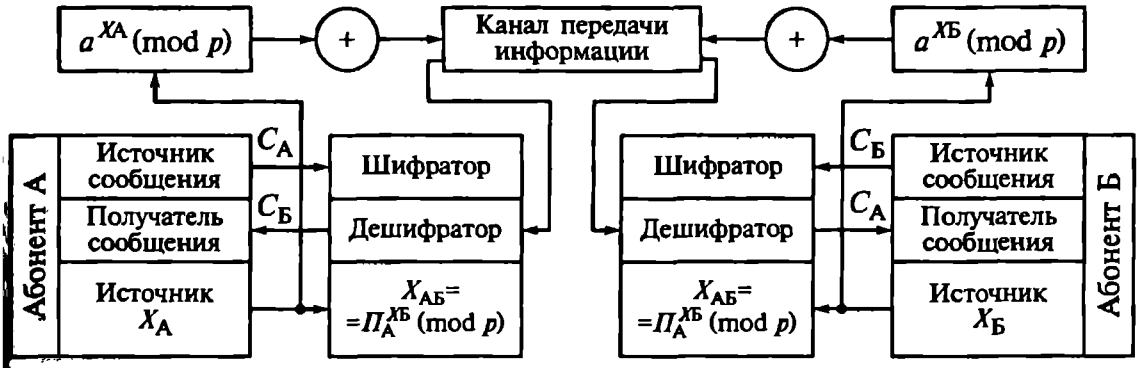


Рис. 5.17. Система связи с открытым ключом

поскольку $П_А$ и $П_Б$ известны. Невозможность (практическая невозможность, т.е. крайняя затруднительность) перехвата сообщения обусловливается трудностью определения секретного ключа X при помощи вычисления дискретного логарифма.

5.7. Стойкость к имитирующим и дезинформирующим помехам (обеспечение подлинности сообщений)

Помехи системам передачи информации могут навязывать получателю ложные сообщения, дезинформировать его. Противодействие такому информационному нападению входит в круг задач радиоэлектронной защиты точно так же, как и противодействие помехам, искажающим сигналы, переносящие эти сообщения. Дезинформируют только те помехи, которые образуют сообщения, подобные истинным, и могут быть приняты как подлинные, созданные собственным источником информации, т.е. дезинформирующие помехи должны имитировать истинные сообщения. Поэтому защита от дезинформирующих помех иначе называется имитозащитой, а способность систем и сообщений противостоять действию дезинформирующих помех — имитостойкостью.

Для обеспечения имитостойкости передаваемых сообщений применяются криптографические методы, в некотором смысле подобные тем, что применяются для обеспечения секретности при передачи информации. Но функции обеспечения секретности (информационной скрытности) и обеспечения подлинности сообщений не тождественны друг другу.

Устойчивость к расшифровке еще не достаточна для обеспечения стойкости сообщений к вредному действию имитирующих помех. Из того факта, что сообщение не может быть расшифровано (может быть расшифровано лишь с достаточно малой вероятностью или по прошествии неприемлемо длительного времени)

еще не следует, что в ходе информационной борьбы противник не может создать ложное, дезинформирующее сообщение. Попытка имитации будет успешной, если система противодействия создаст поддельную шифрограмму $Ш_и$ и эта шифрограмма на приемной стороне будет принята за истинную, посланную законным абонентом системы связи. Вероятность такого события $P_и$.

Подобно потенциальной криптостойкости можно определить предельно достижимый уровень имитостойкости информации как способность системы обеспечивать подлинность передаваемых сообщений. Пусть $N_ш$ — число всех возможных криптограмм, т.е. таких криптограмм, априорная вероятность которых (для системы перехвата) не равна нулю $P(Ш) \neq 0$. Пусть также $N_С$ и $N_К$ — соответственно числа возможных сообщений и ключей, т.е. $P(С) \neq 0$ и $P(К) \neq 0$. Это значит, что для каждой последовательности ключа K существует по крайней мере $N_С$ различных криптограмм и условная вероятность криптограммы для каждого ключа не равна нулю $P(Ш \mid К) \neq 0$. Следовательно, если противник, желающий создать ложное сообщение, выберет совершенно случайно криптограмму из полного числа $N_ш$ (попытается имитировать шифрованное сообщение), вероятность успеха такой имитации будет $P_и = N_С/N_i$. Если же есть какие-либо основания для того, чтобы предпочесть при имитации одни возможные криптограммы другим, вероятность успеха нарушения информационной стойкости будет не меньше. Поэтому

$$P_и \geq \frac{N_С}{N_i}. \tag{5.93}$$

Из (5.93) следует, что для хорошей защиты от имитации, требующей, чтобы $P_и \to 0$, каждое из малого числа $N_С$ возможных сообщений должно при шифрации превращаться в одну из большого числа $N_ш$ криптограмм. Также (5.93) показывает, что нельзя достичь $P_и = 0$, поскольку в этом случае или $N_С = 0$ и ничего нельзя передать, или $N_ш \to \infty$, что столь же нелепо. Иначе говоря, потенциально достижимая защищенность от имитации принципиально не может быть абсолютно совершенной. Предельно достижимый уровень потенциальной защищенности может быть оценен на основе следующих соображений.

Пусть, как и прежде, $P(Ш)$ — вероятность криптограммы Ш(С,К) для системы перехвата, не знающей ключа к шифру; $P_{доп}(Ш)$ — вероятность допустимой криптограммы, возможной при данном секретном ключе К. С этой вероятностью законный получатель сообщения примет криптограмму как возможную (правдоподобную). Условная вероятность $P(Ш \mid К)$ — это вероятность создания криптограммы при известном ключе. Все три величины связаны очевидным неравенством

$$P(\text{Ш}) \le P_{\text{доп}}(\text{Ш}) \le P(\text{Ш}|\text{К}). \qquad (5.94)$$

Поскольку логарифм — монотонная функция своего аргумента, а $P(\text{Ш}) \neq 0$ по определению, будут справедливы и неравенства, равносильные (5.94):

$$\sum_i P(Ш_i)\log P(Ш_i) \le \sum_i P(Ш_i)\log P_{\text{доп}} \le$$
$$\le \sum_i P(Ш_i)\log P_{\text{доп}}(Ш_i|K), \qquad (5.95)$$

где суммирование проводится по всему множеству вероятных криптограмм $i \in 1 : N_{\text{ш}}$. Но

$$-\sum_i P(Ш_i)\log P(Ш_i) + \sum_i P(Ш_i)\log P_{\text{доп}} =$$
$$= H(Ш) - H(Ш|K) = I(Ш, K), \qquad (5.96)$$

т.е. равна разности безусловной энтропии шифрограммы и условной энтропии при условии знания ключа к шифру. По определению эта разность — взаимная информация Ш и К. Она указывает количество информации о ключе К, содержащейся в шифровке Ш.

Входящая в (5.96) величина $\sum_i P(\text{Ш}_i)\log P_{\text{доп}}$ представляет собой среднее значение логарифма вероятности допустимой криптограммы. Но среднее значение некоторой величины не может превосходить ее максимального значения. Поэтому с учетом сделанных обозначений из (5.96) следует, что

$$\log\{\max P_{\text{доп}}(\text{Ш})\} \ge \sum_i P(Ш_i)\log P_{\text{доп}}. \qquad (5.97)$$

Наилучшая, обещающая наибольшую вероятность успеха, попытка имитации шифрованного сообщения состоит в выборе такой конкретной шифровки, которая имеет максимальную вероятность из всех $P_{\text{доп}}(\text{Ш})$:

$$P_{\text{доп}} = \max\{P_{\text{доп}}(\text{Ш}_i)\}, \qquad (5.98)$$

поэтому из (5.97) и (5.98) следует, что

$$\log P_{\text{доп}} \ge I(\text{Ш},\text{К}). \qquad (5.99)$$

Соотношение (5.99) называется нижней границей Симмонса. Равенство в (5.99) достигается тогда, когда $\max\{P_{\text{доп}}(\text{Ш}_i)\}$ равен среднему по i значению вероятности $P_{\text{доп}}(\text{Ш}_i)$, т.е. когда вероятность $P_{\text{доп}}(\text{Ш}_i)$ не зависит от i. При этих условиях оптимальная попытка создания поддельной шифровки сводится к совершенно

случайному выбору подделки из множества возможных (допустимых) криптограмм. Поэтому

$$\log P_{и} \geq -I(Ш,К). \qquad (5.100)$$

Наивысшая достижимая аутентичность, т.е. потенциально достижимая стойкость к подделкам сообщений, соответствует равенству в (5.100). Но из того же соотношения (5.100) следует парадоксальный факт: вероятность обмана (создания поддельного сообщения) тем меньше, чем больше взаимная информация $I(Ш,К)$, т.е. чем больше информации о ключе содержится в шифровке! Таким образом, требование к ключу при обеспечении имитостойкости прямо противоположно требованию к ключу криптозащиты. Парадокс разрешается довольно просто, если учесть, как удостоверяется подлинность (обеспечивается стойкость к обману и подделке) сообщения не в РСПИ, а в обычной житейской и деловой практике. Традиционно для аутентификации документа к нему присоединяют специальное сообщение — подпись и(или) печать. И то и другое сообщение должно быть всем известно и точно указывать на источник, т.е. на того, кто ими обладает и кто их использует для удостоверения подлинности информации. Неразборчивость печати или подписи уменьшает степень доверия к документу (сообщению).

Аналогичная ситуация складывается и в таких широко известных системах аутентификации, как системы опознавания воздушных целей (системы «свой — чужой»). В них сигналы, посылаемые бортом в ответ на запрос подсистемы опознавания целей в составе комплексов ПВО или УВД, должны уверенно идентифицироваться с типом и государственной принадлежностью цели, т.е. они должны быть понятны всем операторам РЛС. Но создавать эти сигналы могут только определенные объекты, и созданные сигналы должны быть надежно защищены от имитации.

Специальное сообщение, удостоверяющее подлинность переданной информации, называется аутентификатором. Такие аутентификаторы, как подпись и печать, присоединенные к сообщению для удостоверения его подлинности, хороши, если сообщение передается на бумажном носителе и не может быть изменено без повреждения этого носителя. При передаче сообщения при помощи сигналов, используемых радиоэлектронными системами вообще и радиосистемами передачи информации в частности, простое присоединение группы символов к основному тексту не может надежно удостоверить его подлинность. Такую группу символов можно перехватить и присоединить к любому ложному сообщению, создав тем самым условия для дезинформации приемника. Для исключения возможности такого обмана необходимо распространить действие аутентификатора на весь текст сообщения, достоверность и подлинность которого требуется подтвердить.

Известны несколько способов формирования и использования такого аутентификатора. Эти способы могут различаются по тому, каково назначение использующих их систем передачи информации и какие требования по имитостойкости предъявляются к системам.

В системах передачи сообщений с повышенной секретностью, когда используется криптозащита информации, аутентификатор присоединяется к исходному шифруемому тексту. После такого сцепления (конкатенации) символов сообщения и аутентификатора производится шифрация полученного расширенного сообщения с использованием секретного ключа, известного только передатчику и приемнику. При шифрации все символы исходного текста обязательно перемежаются и замещаются символами криптограммы. В результате каждый символ криптограммы оказывается зависящим от всех символов исходного текста, символов аутентификатора и символов секретного ключа. Сформированная таким образом криптограмма доставляется получателю, который расшифровывает ее с использованием известного ему ключа и восстанавливает как исходный текст, так и присоединенный к нему аутентификатор. Этот аутентификатор известен только источнику и получателю сообщения. Наличие аутентификатора в полученном и расшифрованном тексте подтверждает подлинность сообщения. Разумеется, тайну аутентификатора нужно охранять не менее строго, чем тайну секретного ключа. Криптографические

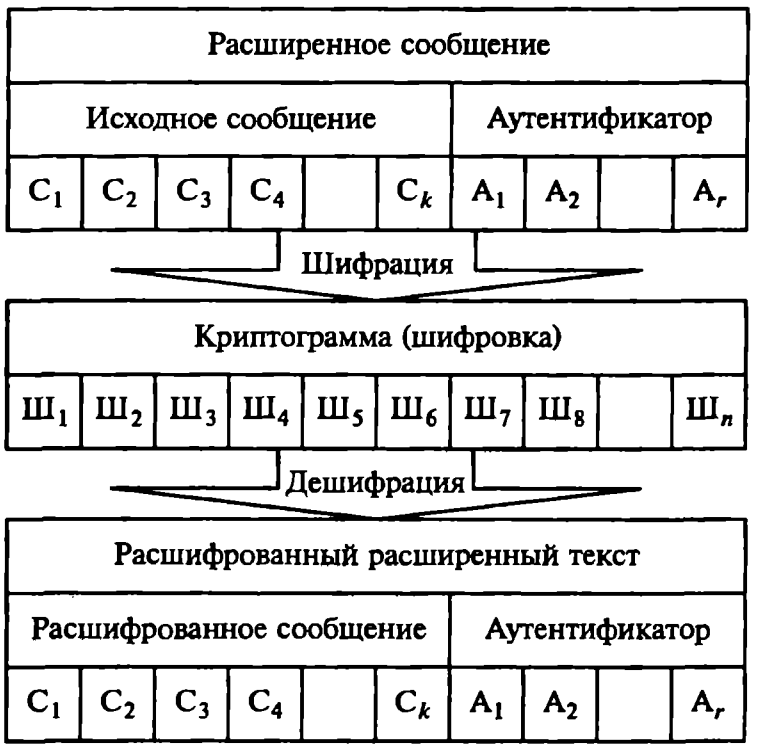


Рис. 5.18. Криптографические методы аутентификации информации в РСПИ

преобразования, совершаемые при передаче имитостойкого сообщения с повышенной секретностью, иллюстрируются на рис. 5.18

Если при шифрации расширенного сообщения используется стойкий криптоалгоритм, то, перехватывая шифровку, противник не может (за приемлемое время) восстановить исходный открытый текст и аутентификатор. В такой ситуации противнику при создании дезинформирующего сообщения не остается ничего иного, как случайным образом сформировать шифротекст в надежде, что он будет воспринят получателем как подлинный. Но если аутентификатор содержит r двоичных символов, то противник при случайной генерации криптограммы сможет угадать неизвестный ему аутентификатор и выдать свое сообщение за подлинное с вероятностью $P_и = 2^{-r}$. Эта вероятность характеризует имитостойкость шифрованного сообщения. Если даже противнику удалось расшифровать криптограмму, это вовсе не значит, что за время вскрытия шифра передатчик и приемник информации по взаимному соглашению не изменили аутентификатор. В случае замены аутентификатора вероятность успеха дезинформации получателя сообщения будет, очевидно, не выше $P_и$.

Возможны случаи, когда шифрация сообщения не нужна или даже нежелательна, как в уже приведенном примере системы опознавания воздушной цели «свой-чужой». Аутентификация таких открытых и общедоступных сообщений совершенно подобна удостоверению их подлинности при помощи подписи. Подчеркивая эту аналогию, способ аутентификации сообщений в линиях связи и на физических носителях, отличных от листов бумаги, называется электронной подписью. Чаще всего используются следующие алгоритмы установления подлинности сообщений при помощи электронной цифровой подписи.

Принцип формирования цифровой подписи на основе схемы ЭльГамаля, положенной в основу отечественного стандарта ГОСТ Р 34.10—94. Все соглашения и требования, касающиеся закрытого и открытого ключей, остаются теми же, что и в случае криптографического закрытия информации.

Пусть имеется большое простое число p, такое, что разложение числа $p - 1$ содержит, по крайней мере, один большой простой множитель, и число a, такое, что $0 < a < p - 1$. Число a нужно выбрать таким, чтобы оно было первообразным корнем в поле вычетов по модулю p. Теория чисел дает несложный тест для проверки этого условия.

Каждым пользователем (абонентом сети распределения аутентифицированной информации) в качестве закрытого ключа принимается случайное число x ($0 < x < p$), а в качестве открытого — совокупность чисел y, a и p. Число y определяется по закрытому ключу, как $y = a^x \bmod p$.

Схема формирования и проверки подписи состоит в следующем. Информационный блок M, на основе которого следует сформировать цифровую подпись, должен иметь длину меньше простого модуля p, т. е. число M должно удовлетворять соотношению: $1 < M < p$. На практике модуль p выбирается таким, что он всегда превышает размер эталонной характеристики сообщения, в качестве которой выступает информационный блок M.

Подписью абонента А, сформированной на основе закрытого ключа y и эталонной характеристики M, служит пара чисел r и s, которые удовлетворяют соотношению:

$$a^M = y^r r^s \bmod p, \tag{5.101}$$

т. е. a^M и $y^r r^s$ дают одинаковый остаток при целочисленном делении на p.

Значения y, a и p в совокупности являются открытым ключом абонента А. Они доступны для всех пользователей, и это позволяет каждому из них убедиться в том, что сообщение действительно подписано абонентом А. Для проверки того факта, что документ подписан абонентом А, проверяется уравнение (5.101).

Система цифровой подписи ЭльГамаля основана на том, что только действительный владелец секретного ключа x может выработать пару чисел (r, s), удовлетворяющую уравнению проверки подписи (5.101). Используя значение x, абонент А вырабатывает подпись по следующему алгоритму:

генерирует случайное число k $(0 < k < p - 1)$, взаимно простое с $(p - 1)$, такое, чтобы наибольший общий делитель чисел k и p был бы равен 1;

вычисляет число $r = a^k \bmod p$;

вычисляет s, решая уравнение $M = (xr + ks) \bmod (p - 1)$;

формирует подпись в соответствии с уравнением (5.101).

Таким образом, подпись из совокупности чисел r и s формируется на основе информационного блока M и закрытого ключа x, а проверяется с помощью открытого ключа (y, a, p) и текущей характеристики M' полученного сообщения. Проверка сводится к установлению истинности равенства (5.101).

Равенство (5.101) истинно тогда и только тогда, когда

$$\left(a^{M'} - y^r r^s \bmod p\right) = 0. \tag{5.102}$$

Нахождение пары чисел (r, s) без знания закрытого ключа вычислительно сложно. Различных подписей, соответствующих одному и тому же документу, может быть чрезвычайно много (k может принимать разные значения). Но выработать правильную подпись может только владелец секретного ключа. Возможные подписи отличаются значением r, но для данного r найти соответствующее

значение s без знания закрытого ключа невозможно (практически невозможно). Для вычисления закрытого ключа по открытому необходимо решить задачу дискретного логарифмирования, которая является вычислительно сложной.

Особенностью схемы цифровой подписи ЭльГамаля является генерация случайного числа k. Не допускается использовать одно и то же значение k для подписи двух разных сообщений, поскольку на основе двух разных подписей, сформированных при одном и том же значении k, имеется возможность вычислить закрытый ключ. Кроме того, при известном значении k нарушитель сможет вычислить и закрытый ключ. Поэтому при формировании цифровой подписи, как и в случае криптографического закрытия информационных блоков, следует избегать повторений чисел k и уничтожать эти числа сразу же после их применения.

В реально используемых системах большое случайное число k генерируется для каждого сообщения и гарантированно уничтожается после применения. При программной реализации обеспечиваемся такая схема шифровывания и формирования подписи, при которой число k появляется только в регистрах микропроцессора и оперативной памяти, а каждое новое число k записывается в ячейки памяти на место предыдущего.

Алгоритм цифровой подписи DSA (Digital Signature Algorithm) является развитием алгоритмов цифровой подписи ЭльГамаля и К.Шнорра. Схема формирования электронной подписи в соответствии с алгоритмом DSA сводится к следующей цепочки действий.

Отправитель и получатель электронного документа используют при вычислении большие целые простые числа: g и p длиной по l бит каждое ($512 < l < 1024$), а также q — большое простое число, делитель числа $(p - 1)$. Числа g, p, q являются открытыми и могут быть общими для всех пользователей сети. Отправитель также выбирает случайное целое число x, $1 < x < q$. Число x является секретным ключом отправителя для формирования электронной цифровой подписи.

Затем отправитель вычисляет значение

$$y = g^x \bmod p. \qquad (5.103)$$

Число y служит открытым ключом для проверки подписи отправителя. Оно передается всем получателям документов.

Для того чтобы подписать документ M, отправитель преобразует его, вычисляя хэш-функцию. Хэш-функция представляет собой одностороннюю криптографическую функцию от сообщения произвольной длины. Значение хэш-функции зависит от каждого бита сообщения и реализуется, как правило, в виде некоторой итерационной процедуры. Значение этой функции $h(M)$ — хэш-код —

для сообщения M произвольной длины имеет фиксированный размер m (обычно 128 или 160 бит). Этот код и является эталонной характеристикой сообщения M. В системах электронной цифровой подписи сообщение M считается подписанным, если подписана его хэш-функция. Поэтому к $h(M)$ предъявляются следующие основные требования:

вычислительно неосуществимо нахождение сообщения M, хэш-функция которого была бы равна заданному значению h;

вычислительно неосуществимо создание двух разных сообщений M_1 и M_2 с равными значениями хэш-функций, т.е. сообщений, удовлетворяющих условию $h(M_1) = h(M_2)$.

Если эти требования не выполняются, то потенциальный злоумышленник может подделать сообщение, подписанное хэш-функцией. Трудоемкость атаки, заключающейся в создании ложного сообщения с тем же значением хэш-функции, что и у данного истинного, в среднем составляет около $2^{m/2}$ вычислений хэш-функций и не зависит от качества криптографических преобразований. Это обстоятельство определяет длину хэш-кода m не менее 128 бит.

Сформировав хэш-функцию

$$m = h(M),\ 1 < m < q, \tag{5.104}$$

отправитель сообщения генерирует случайное целое число k, $1 < k < q$, и определяет число r по правилу:

$$r = \left(g^k \bmod p\right) \bmod q. \tag{5.105}$$

Затем отправитель сообщения вычисляет с помощью секретного ключа x целое число s:

$$s = \frac{m + rx}{k} \bmod p. \tag{5.106}$$

Пара чисел r и s образуют цифровую подпись $S = (r, s)$ под документом M.

Таким образом, подписанное сообщение представляет собой тройку чисел $[M, r, s]$.

Получатель подписанного сообщения $[M, r, s]$ проверяет выполнение условий

$$0 < r < q;\ 0 < s < q \tag{5.107}$$

и отвергает подпись, если хотя бы одно из этих условий не выполнено.

Затем получатель вычисляет значение

$$w = \frac{1}{s} \bmod q, \tag{5.108}$$

значение хэш-функции полученного документа

$$m = h(M) \tag{5.109}$$

и числа

$$u_1 = (m\ w) \bmod q; \quad u_2 = (r\ w) \bmod q. \tag{5.110}$$

После этого получатель с помощью открытого ключа y вычисляет значение

$$v = \left[\left(g^{u_1} y^{u_2}\right) \bmod p\right] \bmod q \tag{5.111}$$

и проверяет выполнение условия $v = r$.

Если условие $v = r$ выполняется, то подпись $S = (r, s)$ под документом M признается получателем подлинной.

Строго доказано, что равенство будет выполняться тогда и только тогда, когда подпись $S = (r, s)$ под документом M получена с помощью именно того секретного ключа x, из которого был получен открытый ключ y. Таким образом, можно надежно удостовериться, что отправитель сообщения владеет именно данным секретным ключом x (не раскрывая при этом значения ключа x) и что отправитель подписал именно данный документ M.

По сравнению с алгоритмом цифровой подписи Эль Гамаля алгоритм DSA имеет ряд преимуществ, а именно:

при любом допустимом уровне стойкости, т. е. при любой паре чисел g и p (512... 1024 бит), числа q, x, r, s имеют длину по 160 бит, сокращая длину подписи до 320 бит.

большинство операций с числами k, r, s, x при вычислении подписи производится по модулю числа q длиной 160 бит, что сокращает время вычисления подписи;

при проверке подписи большинство операций с числами u_1, u_2, u и w также производится по модулю числа q длиной 160 бит, что сокращает объем памяти и время вычисления.

Недостатком алгоритма DSA является то, что при подписывании и проверке подписи приходится выполнять сложные операции деления по модулю q:

$$S = \left(\frac{m + rx}{k}\right) \bmod q; \quad w = \frac{1}{s} \bmod q, \tag{5.112}$$

что не позволяет получать максимальное быстродействие.

Следует отметить, что реальное исполнение алгоритма DSA может быть ускорено с помощью выполнения предварительных вычислений, поскольку значение r не зависит от сообщения M и значения его хэш-функции m. Можно заранее создать строку случайных значений k и затем для каждого из этих значений вы-

числить значения r. Можно также заранее вычислить обратные значения k^{-1} для каждого из значений k и при поступлении сообщения M вычислить значение s для данных значений r и k^{-1}. Эти предварительные вычисления значительно ускоряют работу алгоритма DSA.

Отечественный стандарт цифровой подписи ГОСТ Р34.10—94 предписывает использование следующих параметров:

p — большое простое число длиной 509...512 бит либо 1020...1024 бит;

q — простой сомножитель числа $(p-1)$, имеющий длину 254...256 бит;

a — любое число, меньшее $(p-1)$, но такое, что $a^q \bmod p = 1$;

x — некоторое число, меньшее q;

$y = a^x \bmod p$.

Кроме того, этот алгоритм использует одностороннюю хэш-функцию $h(x)$. Стандарт ГОСТ Р 34.11—94 определяет хэш-функцию, основанную на использовании стандартного алгоритма блочного шифрования ГОСТ 28147—89.

Первые три параметра p, q и a являются открытыми и могут быть общими для всех пользователей сети. Число x является секретным ключом. Число y является открытым ключом. Чтобы подписать некоторое сообщение m, а затем проверить подпись, выполняются следующие действия:

абонент A генерирует случайное число $K < q$ и вычисляет:

$$\begin{aligned} r &= \left(a^k \bmod p\right) \bmod q; \\ s &= \left[xr + kh(m)\right] \bmod q, \end{aligned} \tag{5.113}$$

если $h(m) \bmod q = 0$, то значение $h(m) \bmod q$ принимают равным единице. Если $r = 0$, то выбирают другое значение K и возвращаются к п.1.

Цифровая подпись представляет собой два числа:

$$r \bmod 2^{256};\ s \bmod 2^{256};$$

абонент A отправляет эти числа получателю сообщения B;

абонент B проверяет полученную подпись, вычисляя

$$\begin{aligned} v &= h(m)^{q-2} \bmod q; \\ z_1 &= (sv) \bmod q; \\ z_2 &= \left[(q-2)v\right] \bmod q; \\ u &= \left[\left(a^{z_1} y^{z_2}\right)\right] \bmod q, \end{aligned} \tag{5.114}$$

если $u = r$, то подпись считается верной.

Различие между алгоритмами ГОСТ Р 34.11—94 и DSA заключается в том, что в DSA

$$s = \left(k^{k-1}\left[xr + h(m)\right]\right) \bmod q, \qquad (5.115)$$

и это приводит к другому уравнению проверки подписи. Следует также отметить, что в отечественном стандарте цифровой подписи параметр q имеет длину 256 бит.

Оба рассмотренных алгоритма обеспечения стойкости сообщения к подделкам и искажениям основываются на увеличении избыточности передаваемого сообщения. Разумеется, возможны и иные, отличные от двух приведенных выше, протоколы защиты подлинности сообщений. Но общим для любых протоколов остается то, что аутентификатор присоединяется к исходному тексту. И чем больше внесенная аутентификатором избыточность, тем выше имитостойкость. Присоединенный к сообщению избыточный идентификатор может быть назван электронной подписью. Очевидно, такая подпись подтверждает подлинность сообщения и в том случае, когда оно передается без посредства бумажного носителя.

Способы подтверждения подлинности основаны на внесении избыточности точно так же, как и способы повышения помехоустойчивости. Но для улучшения помехоустойчивости избыточные символы преобразуют сообщения в такие последовательности, которые группируются возможно более близко (в соответствии с принятой метрикой в пространстве сигналов) к неискаженному сигналу. При использовании избыточности для формирования имитостойких сообщений они конструируются иначе: чтобы любые изменения символов в соответствии со стратегией дезинформации распределяли получающиеся кодовые последовательности случайно и равновероятно по всему сигнальному пространству.

Контрольные вопросы

1. Укажите основные пути повышения помехоустойчивости систем передачи информации.
2. Какие коды называются помехоустойчивыми и почему?
3. За счет чего помехоустойчивые коды получают возможность обнаруживать и исправлять ошибки?
4. Какая разница между блоковыми и непрерывными кодами? В чем отличие между разделимыми и неразделимыми помехоустойчивыми кодами?
5. Оцените корректирующую способность кода (15,5). Какой кратности ошибки может обнаруживать и исправлять такой код?
6. Постройте автокорреляционную функцию M последовательности 111100010011010.

7. Покажите, что M последовательность из вопроса 6 генерируется четырехразрядным регистром сдвига с обратной связью $Q_0 = Q_3 \oplus Q_4$. Определите начальное состояние регистра, генерирующее такую последовательность.

8. В чем состоят особенности кодовых последовательностей, предназначенных для приема «в целом»?

9. В чем состоит основное отличие систем с информационной и решающей обратными связями?

10. При каком способе организации обратной связи обеспечивается бóльшая избыточность сигнала в информационных системах? Почему?

11. Зачем нужна синхронизация передающих и приемных подсистем систем передачи информации?

12. В чем разница между блочными и поточными шифрами?

13. Шифровки неосмысленного текста считаются очень устойчивыми к вскрытию. Почему?

14. Чем отличаются криптографические преобразования, используемые для закрытия информации, от преобразований, обеспечивающих аутентификацию?

15. Зачем устраняют избыточность исходного текста перед шифрованием?

16. Перечислите преобразования, используемые в процессе шифрации.

Глава 6
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ СИСТЕМ И СЕТЕЙ СВЯЗИ

6.1. Сети связи

Современные сети связи, обеспечивающие информационный обмен между разными абонентами в разных условиях, могут объединять разные по структуре и характеристикам информационные системы.

Сеть связи — это совокупность пунктов и линий (каналов) связи. Основной функцией сети является доставка сообщений, преобразованных в электрический сигнал, по заданному адресу при условии обеспечения требуемых качественных показателей по времени доставки, верности и надежности. Сеть связи содержит подсистемы двух типов: узлы связи и оконечные пункты. Узлы связи соединяются между собой пучками каналов; оконечные пункты подключаются к узлам связи через соединительные линии или отдельные каналы.

Оконечные пункты оснащаются оборудованием, преобразующим сообщения в электрический сигнал и обратное преобразование, а также обеспечивающим сопряжение с каналами связи. В узлах связи размещается комплекс технических средств, включающий в себя оборудование для организации групповых трактов и каналов передачи информации, технические средства для коммутации и оборудование для управления сетевыми ресурсами.

Сети связи составляют техническую базу Взаимоувязанной сети связи Российской Федерации (ВСС РФ), интегрированную с сетями связи стран СНГ и имеющую выходы на международные сети связи и передачи данных.

Различают два типа сетей — первичные и вторичные. Первичная сеть ВСС состоит из линий передачи и сетевых узлов, организованных на пересечении линий передачи и сетевых станций. На сетевых узлах устанавливается каналообразующая аппаратура систем связи и передачи данных. Сетевые станции, являющиеся оконечными пунктами первичной сети, подключают к линиям передачи системы вторичные сети связи Для образования первичной сети используются линии передачи различного типа: кабельные, спутниковые, оптоволоконные, радиорелейные и т.п. Канал тональной частоты (ТЧ) обеспечивает передачу телефонных сигналов. Стандарт предусматривает полосу пропускания этого

канала в пределах 0,3...3,4 Кгц. Стандартный основной цифровой канал (ОЦК) обеспечивает скорость передачи информации 64 Кбит/с. На базе первичной сети формируются вторичные сети ВСС.

Вторичная сеть ВСС представляет собой комплекс технических средств, содержащий коммутационное и каналообразующее оборудование; оборудование управления сетью и другие подсистемы, размещаемые на узлах связи; оконечное оборудование, размещаемое в оконечных пунктах; каналы, организованные на базе типовых каналов и трактов передачи. Наиболее крупной вторичной сетью является Общегосударственная система автоматизированной телефонной связи (ОГСТфС).

В состав цепи междугородней (и международной) телефонной связи могут входить, кроме кабельных линий (металлических или оптоволоконных), радиорелейные и спутниковые линии связи (рис. 6.1). На этой схеме две наземные станции автоматической междугородней связи (АМТС) соединены с наземными стационарными станциями спутниковой связи кабелем (К) и радиорелейной линией (РРЛ).

Для связи с подвижными абонентами организуются сети мобильной связи, имеющие выход на местные сети. Сети мобильной связи работают в УКВ диапазоне и строятся по радиальной и сотовой схемам. Линии связи этих сетей работают в пределах прямой видимости. В случае реализации радиальной схемы используются довольно мощные базовые и подвижные станции. Для увеличения зоны обслуживания абонентов радиальной сети мобильной связи антенны базовых станций поднимают на максимально возможную для данной местности высоту. Сотовые сети работают с большим количеством базовых станций в зоне обслуживания. Базовые станции оборудуются сравнительно маломощными передатчиками.

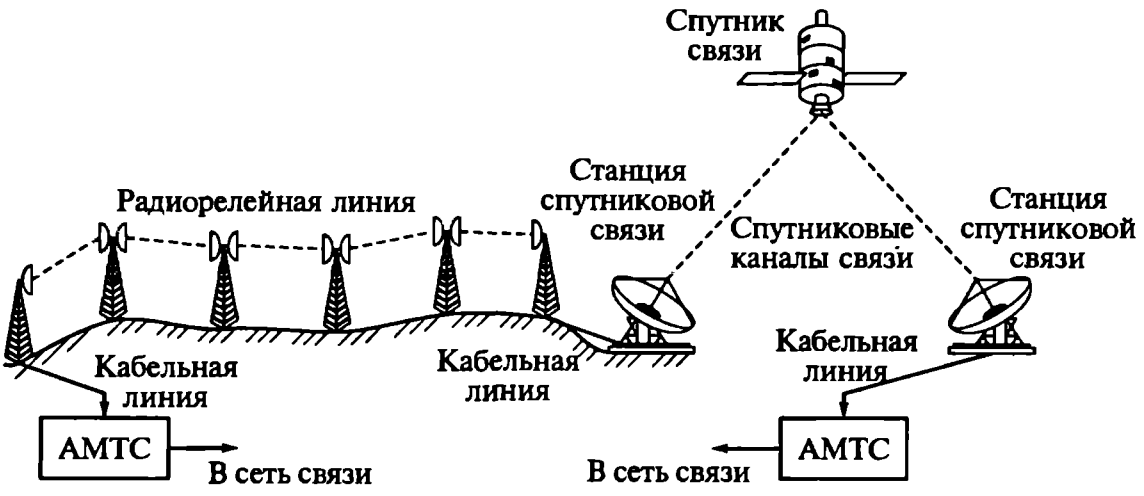


Рис. 6.1. Система телефонной связи

6.2. Основные угрозы безопасности информации и методы защиты информации в кабельных телефонных сетях

Для телефонной сети общего пользования установлены достаточно жесткие стандарты построения и правила функционирования. Однако эти правила не предусматривают обеспечение конфиденциальности связи между абонентами, хотя отдельные части системы телефонной связи сами по себе обладают определенной защищенностью от перехвата циркулирующих в них сообщений. Это обусловлено принимаемыми организационными мерами и некоторыми техническими решениями. Например, ряд кабелей связи прокладывается в коллекторах или тоннелях метро, куда ограничен доступ посторонних лиц; коммутационное оборудование АТС располагается в охраняемых помещениях и т.п.

В то же время имеется ряд участков, где возможен перехват информации из систем кабельной связи. Известно, что происходят взаимные наводки на соседние физические цепи (переходные разговоры), возможны подсадки на АТС (когда из-за неисправностей к цепям, соединяющим двух абонентов, подключается третий) и т.п. При таких нештатных ситуациях сообщения, которыми обмениваются два абонента, становятся доступны случайным или заинтересованным лицам. Персонал, эксплуатирующий сетевое оборудование, имеет возможность прослушивать переговоры при проведении регламентных и ремонтных работ на АТС и в других пунктах сети.

Преднамеренный перехват телефонных переговоров (или, точнее, несанкционированный доступ к информации в телефонных линиях связи) возможен не только при гальваническом подключении к проводам, но также с помощью бесконтактных индукционных или емкостных датчиков, устанавливаемых вблизи разговорных цепей. На всех участках телефонной цепи, где проходят сигналы канала ТЧ, для перехвата необходимо иметь лишь простейшие технические средства. Техника перехвата ушла далеко вперед от простого подслушивания переговоров по медным проводам. В настоящее время имеются возможности перехвата любых каналов связи, инфракрасных систем передачи и даже оптоволоконных линий связи. Единственный радикальный способ предотвращения перехвата телефонных переговоров и раскрытия их содержания посторонними — это шифрование, или скремблирование (перемешивание), речевых сигналов.

Цепь прохождения телефонной информации состоит из нескольких участков: абонентский участок, участок местной сети, участок внутризоновой сети, участок магистральной сети, который является общим для абонентов, обменивающихся телефон-

ными сообщениями. Остальные участки повторяются со стороны каждого абонента. Канал ТЧ, соединяющий абонентов, имеет типовую ширину полосы 0,3...3,4 кГц. При этом на всех участках, кроме абонентского, по физическим цепям передается групповой сигнал, содержащий информацию от разных пар абонентов. Однако на границах участков и внутри участка местной сети там, где происходит транзит по низкой частоте (транзит ТЧ), по физической цепи проходит тот же сигнал, что и на абонентском участке. При этом везде, где проходит канал ТЧ, возможно прямое прослушивание разговоров с помощью минимального набора средств перехвата. В остальных точках цепи для подслушивания необходимо иметь аппаратуру или устройства, выделяющие информативный сигнал ТЧ из группового сигнала.

В цифровых сетях на абонентских участках и в местах транзита, эквивалентных транзитам ТЧ, вместо канала ТЧ используется основной цифровой канал, по которому происходит транзит сигнала с кодово-импульсной модуляцией на скорости 64 Кбит/с. У абонентов цифровых сетей должны быть телефонные аппараты (или абонентские комплекты), оснащенные соответствующими кодирующими устройствами, в том числе аналого-цифровыми (АЦП) и цифроаналоговыми (ЦАП) преобразователями сигналов. Аналогичное оборудование необходимо иметь при перехвате сообщений, который возможен на тех же участках, что и на аналоговой сети.

Почти все технические средства создают технические каналы утечки информации за счет побочных и непреднамеренных излучений. Например, при использовании обычных телефонных аппаратов и при положенной на рычаг микротелефонной трубке (телефон вроде бы отключен) на абонентских проводах, выходящих за пределы помещения, присутствуют электрические сигналы, по которым можно узнать все, о чем говорится в помещении.

При использовании специальных технических средств можно создать дополнительные пути утечки информации. Например, поместить в цифровой телефонный аппарат миниатюрный передатчик и подключить его к микрофонной цепи. С выхода такого передатчика аналоговые речевые сигналы излучаются или передаются по абонентским соединительным линиям на большие расстояния.

Все основные методы защиты от утечки информации можно условно разделить на две группы: организационные, или организационно-технические, и аппаратные, или программно-аппаратные.

К первой группе относятся такие меры, как охрана помещений, где размещается аппаратура связи (коммутационное оборудование, аппаратура уплотнения и т.п.); использование специальных приборов, обнаруживающих подслушивающие устройства при несанкционированном подключении к линиям связи; исполь-

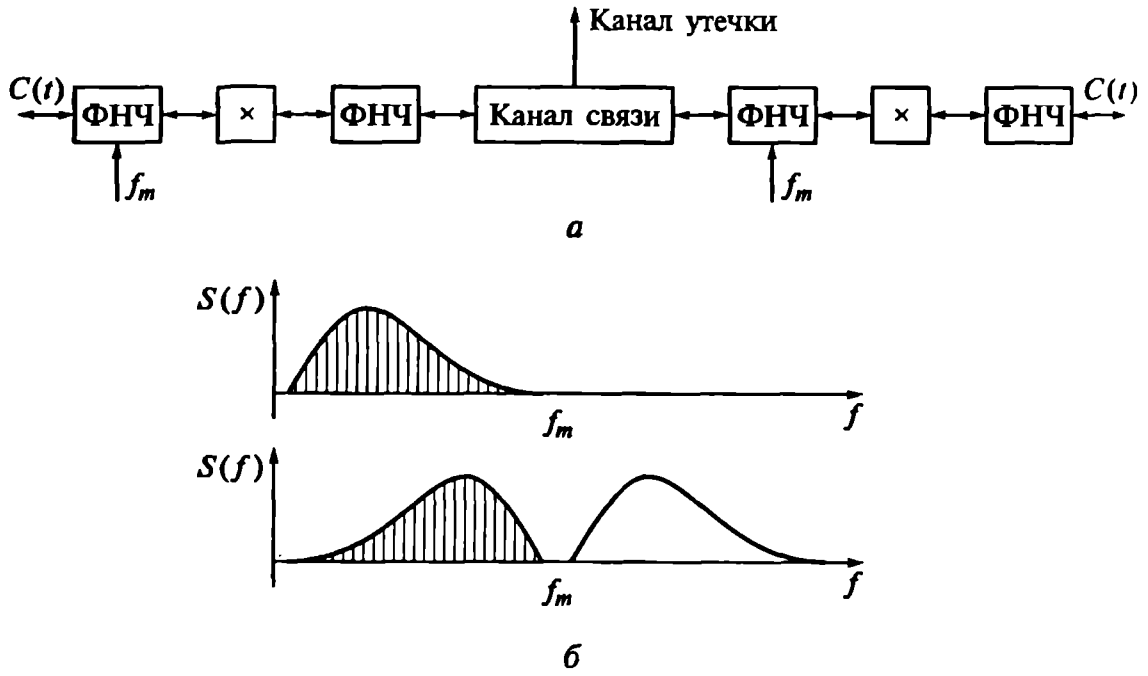


Рис. 6.2. Простейший аналоговый скремблер — инвертор частоты:
а — структурная схема; *б* — преобразование спектра

зование кабелей в герметичной оболочке с контролем разгерметизации (обнаружением утечки газового наполнения) при повреждении этой оболочки; экранирование кабелей; прокладка кабелей в труднодоступных траншеях с защитой и сигнализацией о проникновении.

Организационные, или организационно-технические, методы в ряде случаев оказываются достаточными для защиты конфиденциальной информации. Но если они не обеспечивают требуемого уровня защиты, прибегают к использованию аппаратных, или программно-аппаратных, методов.

Простейшие методы аппаратной защиты используют кодирование речевых сигналов способом, отличающимся от общепринятых и стандартных. Самое известное и простое преобразование сигнала ТЧ — инверсия его частотного спектра. Структурная схема, поясняющая работу инвертора спектра для защиты информации в телефонном канале, изображена на рис. 6.2, *а*; производимые инвертором преобразования спектра передаваемого сигнала иллюстрируются на рис. 6.2, *б*.

Балансный смеситель переносит спектр сигнала на частоту F_m, несколько большую верхней частоты в спектре сообщения $C(t)$, равную 3,4 кГц, а последующий низкочастотный фильтр подавляет верхнюю боковую полосу колебания (на рис. 6.2, *б* эта полоса не заштрихована). В результате спектр преобразованного сигнала на выходе скремблера оказывается симметричным спектру исходного сигнала, имеет с ним одинаковую полосу и одинаковый динамический диапазон. Такое колебание можно передавать по ка-

налу ТЧ, предназначенному для открытой, незащищенной передачи сигнала. На приемной стороне законного получателя информации нужно произвести точно такое же преобразование спектра. В результате восстановится исходный сигнал $C(t)$. Если же принимать преобразованный сигнал с инвертированным спектром, не совершая обратного преобразования, понять сообщение $C(t)$ невозможно. Но и произвести декодирование, совершив зеркальное преобразование спектра, несложно. Поэтому такой способ имеет лишь историческое значение как один из первых способов защиты аналоговых речевых сообщений в телефонных каналах передачи информации. Более сложными и, соответственно, более надежными являются методы защиты, при которых законы кодирования изменяются в процессе передачи информации. Такие методы называются *динамическим кодированием*. В каналах ТЧ это коммутируемая инверсия, частотные перестановки, временны́е перестановки, а также комбинация этих методов. Устройства, изменяющие естественную структуру речевого сигнала для затруднения перехвата, называются *скремблерами* (от *англ.* speech scrambler — перемешиватель речи).

Структурная схема скремблера, расщепляющего полосу спектра, приведена на рис. 6.3.

В соответствии с этой схемой несколько (для линий телефонной связи — пять) октавных фильтров вырезают из спектра исходного сигнала частотные полосы. Фильтрованные таким образом колебания балансными смесителями гетеродинируются на разные частоты в полосе спектра исходного сигнала. Гетеродинированные колебания суммируются, и их сумма передается по каналу связи. Выбором и коммутацией гетеродинных частот можно составить 5! = 120 различных перестановок элементарных полос. Этот выбор как бы является ключом шифрующего преобразования: он известен собственному абоненту системы связи, но держится в секрете от оператора средства перехвата информации.

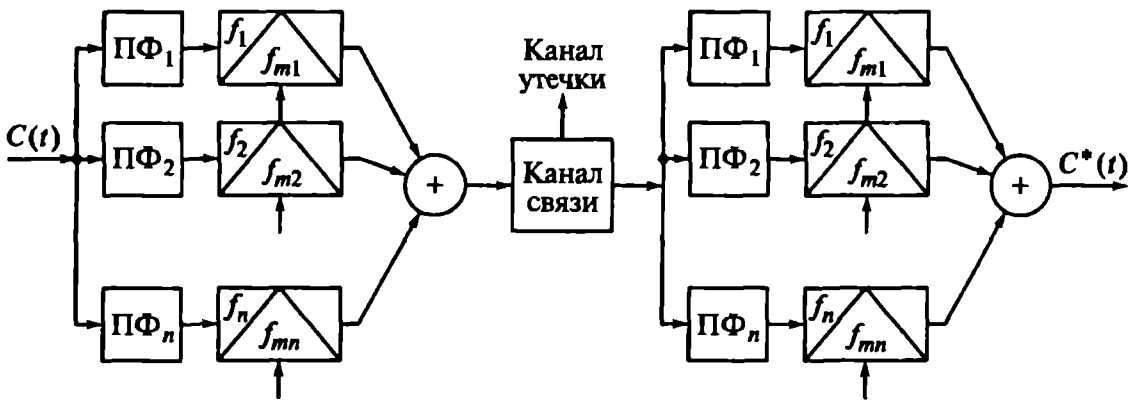


Рис. 6.3. Скремблер, расщепляющий полосу сигнала

Средство перехвата может провести спектральный анализ шифрованного колебания в канале связи и, зная характеристики спектра исходного сигнала $C(t)$, определить ключ шифра. Поэтому выбор гетеродинных частот делают не фиксированным, а изменяющимся во времени по закону, случайному для разведки и точно известному собственному абоненту. Можно сказать, что частоты коммутируются в соответствии с сигналом генератора ключевого потока. А сам генератор управляется секретным ключом, подобным ключу поточного шифра.

Временны́е интервалы постоянства частот гетеродинных колебаний скремблера выбирают из двух условий. Во-первых, они должны быть не больше времени спектрального анализа, чтобы разведка при перехвате информации не смогла восстановить секретный ключ, основываясь на априорном знании характеристики спектра шифруемого сигнала. Во-вторых, переходные процессы в канале, повторяющиеся с частотой коммутации, не должны приводить к ухудшению приема сигнала законным абонентом сети связи.

Совершенно аналогично с частотным можно осуществить временно́е скремблирование, при котором реализация сообщения нарезается временны́ми полосками (отрезками), которые переставляются по случайному для разведки закону, образуя при суммировании групповой сигнал в канале связи. Законный получатель восстанавливает сообщение, зная закон перестановки отрезков сообщения.

Наиболее совершенные системы защиты аналоговых сообщений объединяют частотное и временное скремблирование в канале ТЧ.

Современный уровень развития микроэлектроники позволяет для каналов ТЧ осуществлять динамическое кодирование речевых сигналов в цифровом виде. При этом аналоговый речевой сигнал после микрофона преобразуется в цифровой, затем осуществляются необходимые изменения (фильтрация, перестановки, инверсия и т. п.). После этих преобразований цифровой сигнал снова преобразуется в аналоговый, который передается по каналу ТЧ. На приемном конце декодирование осуществляется аналогичным образом (в обратном порядке).

Все методы скремблирования могут сохранять некоторые признаки исходного защищаемого речевого сигнала. Поэтому при прямом прослушивании скремблированной речи можно извлечь некоторую информацию о говорящем и даже понять отдельные элементы речи: звуки, слоги, слова, фразы. Поэтому говорят, что при скремблировании сохраняется остаточная разборчивость речи.

По степени защиты устройства для конфиденциальной связи условно подразделяют на три группы:

• простейшие — маскираторы, в которых осуществляется простое кодирование;

• средней сложности — скремблеры или перемешиватели, в которых используется не очень сложное динамическое кодирование. При преднамеренном подслушивании с использованием современной вычислительной и другой техники, полезная информация может быть раскрыта если не в реальном времени, то по записи за ограниченное время, значительно меньшее того, в течение которого сохраняется ее актуальность;

• высокой сложности — шифраторы или шифрующие устройства — аппараты засекречивания (ЗАС), в которых используются достаточно сложные алгоритмы преобразований. Стойкость таких систем очень высока. Для несанкционированного извлечения полезной информации потребуется очень большой срок, превышающий время сохранения секретности сообщений.

К устройствам и аппаратам конфиденциальной связи обычно предъявляются специальные требования по уровню побочных и непреднамеренных излучений, по защите от утраты и (или) уничтожения ключей шифропреобразования, некоторые другие.

Сети связи, использующие только абонентские устройства обеспечения конфиденциальности связи, называются *сетями с защитой от абонента до абонента.* Сети связи с применением только канальных или коммутируемых устройств обеспечения конфиденциальной связи — *сетями с защитой от узла до узла.* Сети, в которых одновременно используются и абонентские, и другие типы средств обеспечения конфиденциальности, называются *сетями с комбинированной защитой.* Каждый из этих типов сетей имеет свои преимущества и недостатки [10].

Организуя конфиденциальную передачу информации с защитой от абонента до абонента на сетях общего пользования, абоненты соединяются обычным способом, взаимодействуя в соответствии с ГОСТом на телефонные аппараты. При этом устройства конфиденциальной связи устанавливаются только у абонентов. На узлах коммутации АТС никаких преобразователей нет, и, соответственно, нет открытой информации с выхода телефонных аппаратов. Поэтому не предъявляется каких-либо специальных требований к коммутационному и линейному оборудованию по предотвращению утечки передаваемой информации. Но при таком способе защиты устройствами обеспечения конфиденциальности должны оснащаться все абоненты сети.

В сетях телефонной связи с защитой от узла до узла устройства, обеспечивающие конфиденциальность связи, закрепляются за участком канала связи постоянно или временно, на каждый сеанс обмена информацией. Возможно также использование групповых устройств конфиденциальной связи, которые од-

новременно защищают информацию от нескольких абонентов или весь групповой тракт.

В схемах защиты от узла до узла для обслуживания отдельных, удаленных от узла коммутации, абонентов можно включить скремблеры и другие устройства обеспечения конфиденциальности в абонентскую линию.

Схемы защиты от узла до узла позволяют обеспечить достаточно высокую безопасность связи в выделенных сетях. Однако в них возможно подслушивание чужих разговоров на узлах коммутации, например, при ошибочном подключении третьего лица или несанкционированном подсоединении к цепям связи, по которым конфиденциальная информация передается в незащищенном виде.

При организации сетей с комбинированной защитой используются различные варианты подключения устройств обеспечения конфиденциальности. Условно всех абонентов такой сети можно разделить на две категории: AI — привилегированные абоненты, имеющие абонентские устройства защиты, и AII — обычные абоненты сети коллективного пользования, обслуживаемые канальными устройствами обеспечения конфиденциальности.

Абоненты категории AI между собой соединяются так же, как в сетях с защитой от абонента до абонента. Речевые сигналы или другая конфиденциальная информация от телефонных аппаратов абонентов категории AI проходят через все узлы коммутации в защищенном виде, и на каждом транзитном участке между узлами коммутации происходит дополнительное преобразование (вторичная защита). Сигналы взаимодействия, передаваемые по отдельному каналу сигнализации на участке абонентской линии, передаются в открытом виде, а на транзитных участках они защищаются одновременно с конфиденциальной информации. За счет этого повышается защищенность информации, так как в каналах связи становится невозможным выделение сигналов взаимодействия, адресов абонентов, регламентация времени ведения конфиденциальных переговоров и т. п.

Абоненты категории AII соединяются так же, как в сетях с защитой от узла до узла. Удаленные абоненты этой категории могут подключаться через концентраторы, которые соединяются с узлом коммутации через дополнительные устройства защиты информации.

При организации сетей связи может оказаться необходимым использовать устройства обеспечения конфиденциальности с разными алгоритмами преобразования сигналов. Это обусловлено несколькими причинами.

Во-первых, сроки службы аппаратуры защиты информации составляют десятки лет, за это время появляются новые и совершенствуются используемые алгоритмы преобразований, вводятся новые стандарты.

Во-вторых, для снижения общей стоимости аппаратуры, установленной на сети, и получения наилучшей разборчивости и качества звучания речи целесообразно применять разные типы речепреобразующих устройств для разных каналов связи.

В-третьих, для повышения живучести сети бывает необходимо использовать разные криптографические преобразования в различных зонах сети связи.

6.3. Методы и средства защиты информации в мобильных системах

6.3.1. Защита информации в цифровых системах мобильной связи стандарта GSM

Современные радиосети мобильной связи используют цифровые методы передачи сообщений. Они предоставляют пользователям очень большой набор услуг и хорошо сопрягаются как с цифровыми сетями с интеграцией служб, так и с сетями пакетной передачи данных. Наибольшее распространение получили сети мобильной связи, базирующиеся на стандарте GSM (Group Special Mobile или, как стали расшифровывать это сокращение после широкого мирового распространения стандарта, Global System for Mobil Communication). Весьма перспективным представляется стандарт CDMA (Code Division Multiple Access), основанный на кодовом уплотнении и разделении каналов.

Структура и состав оборудования сетей стандарта GSM иллюстрируется схемой на рис. 6.4.

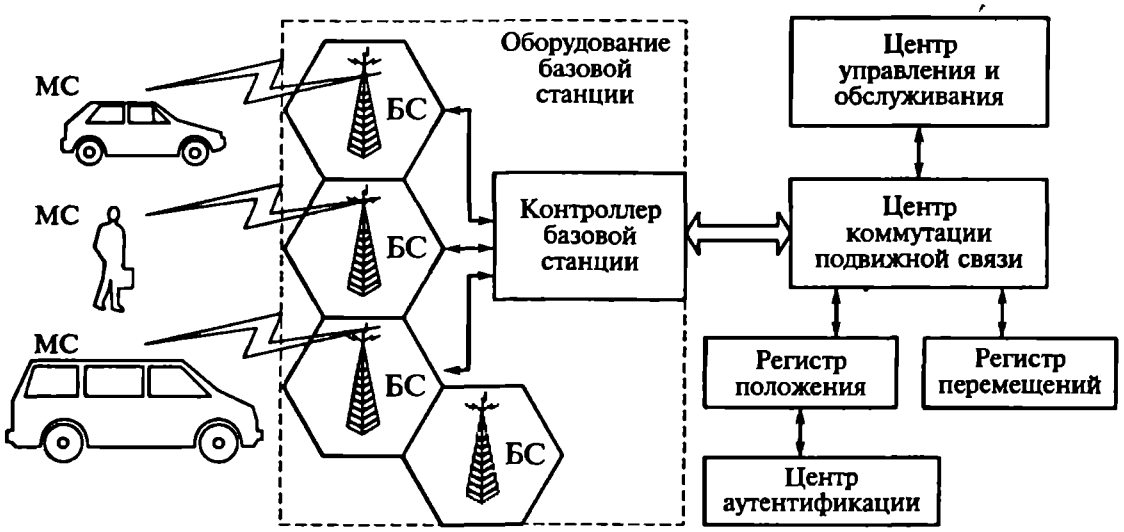


Рис. 6.4. Структурная схема построения сети стандарта GSM

Мобильные станции (МС), транспортные и портативные, оснащены оборудованием для организации доступа абонентов сети GSM к другим сетям связи и передаче данных. Каждая мобильная станция имеет свои международные идентификационные номера (INSI и IMEI), записанные в ее памяти.

Основной идеей и техническим решением, которое позволило резко увеличить емкость радиосети сотовой связи, является повторное использование частот в несмежных сотах. Первым способом организации повторного использования частот, который применялся в аналоговых системах первого поколения, был способ, использующий базовые станции с антеннами круговой направленности. Базовые станции, на которых допускается повторное использование выделенного набора частот, удалены друг от друга на расстояние D, называемое защитным интервалом (рис. 6.5).

Смежные базовые станции, использующие различные частотные каналы, образуют группу из C станций — кластер. Если каждой базовой станции выделяется набор из m каналов с шириной полосы $F_к$ у каждого, то общая ширина полосы $F_с$, занимаемая данной системой сотовой связи, составит $F_с = F_к mC$.

Таким образом, величина C определяет минимально возможное количество каналов в системе, и поэтому ее называют частотным параметром системы, или коэффициентом повторения частот. Коэффициент C не зависит от количества используемых каналов и увеличивается по мере уменьшения радиуса ячейки. Таким образом, при использовании сот меньших размеров можно увеличить повторяемость частот.

Применение шестиугольных сот позволяет минимизировать ширину используемой полосы частот, поскольку такая форма обеспечивает оптимальное соотношение между значениями C и D. Кроме того, шестиугольная форма наилучшим образом вписывается в круговую диаграмму направленности антенны базовой станции, установленной в центре соты.

Размер соты R определяет защитный интервал D между сотами, в которых повторно могут быть использованы одни и те же частоты. Значение защитного интервала D, кроме уже перечисленных факторов, зависит также от допустимого уровня помех и условий распространения радиоволн. Поскольку интенсивность вызовов в пределах всей зоны обслуживания примерно одинакова, соты выбираются одного размера. Размер R определяет также количество абонентов N, спо-

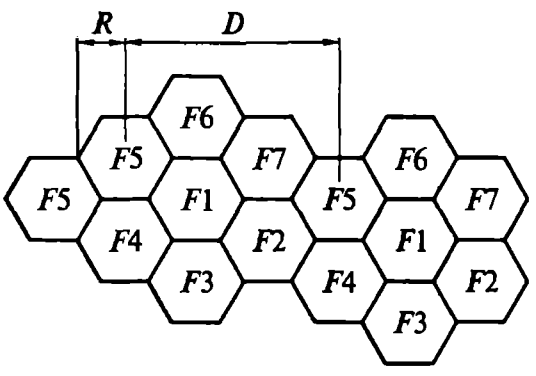


Рис. 6.5. Повторное использование частот в несмежных сотах

собных одновременно вести переговоры на всей территории обслуживания. Следовательно, уменьшение этого размера позволяет не только повысить эффективность использования выделенной полосы частот и увеличить абонентскую емкость системы, но и уменьшить мощность передатчиков и чувствительность приемников базовых и подвижных станций. Это, в свою очередь, улучшает электромагнитную совместимость средств сотовой связи с другими радиоэлектронными средствами и системами.

Эффективным способом снижения уровня помех может быть использование секторных антенн с узкими диаграммами направленности. В секторе такой узконаправленной антенны сигнал излучается преимущественно в одну сторону, а уровень излучения в противоположном направлении сокращается до минимума. Деление сот на секторы позволяет чаще применять частоты в сотах повторно. Весьма эффективный и используемый в настоящее время способ повторного использования частот в организованных таким образом сотах предусматривает применение трехсекторных антенн для каждой базовой станции и трех соседних базовых станций с формированием ими девятью групп частот (рис. 6.6). В этом случае используются антенны с шириной диаграммы направленности 120°

Еще более высокую эффективность использования выделенной полосы частот и, следовательно, наибольшее количество абонентов сети, работающих в этой полосе, обеспечит способ повторного использования частот, при котором шестидесятиградусные диаграммы направленности антенн базовых станций делят каждую ячейку на шесть секторов и каждая рабочая частота ис-

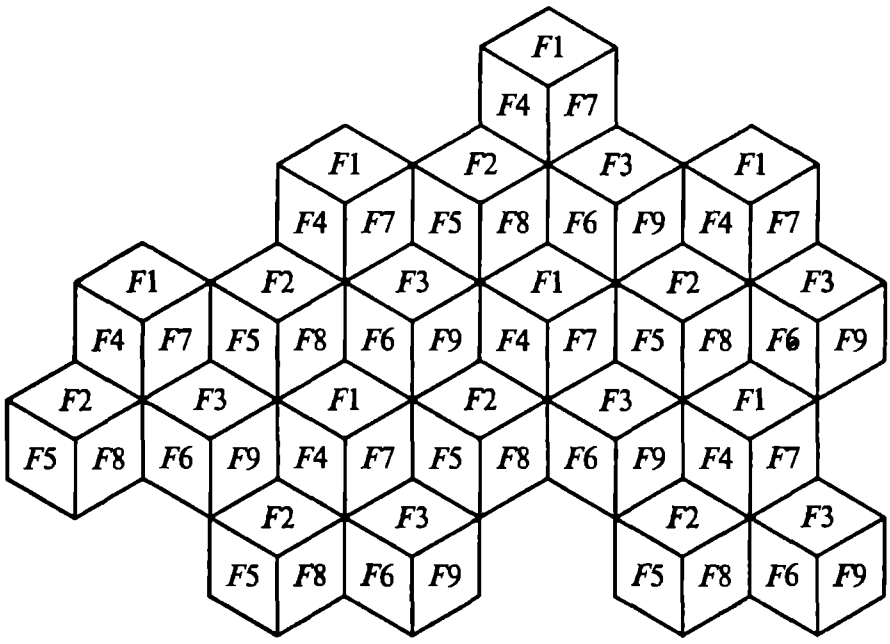


Рис. 6.6. Повторное использование частот в трехсекторных сотах

пользуется дважды в пределах кластера, состоящего из четырех сот [8].

Каждая из сот обслуживается многоканальным приемопередатчиком — базовой станцией (БС). Она служит интерфейсом между сотовым телефоном и центром коммутации подвижной связи, где роль проводов обычной телефонной сети выполняют радиоволны. Количество каналов базовой станции обычно кратно $2^3 = 8$. Один из каналов является управляющим (control channel). В некоторых ситуациях он может называться также каналом вызова (calling channel). По этому каналу организуются соединения при вызове подвижного абонента сети, а сам разговор начинается только после того, как будет найден свободный в данный момент канал и произойдет переключение на него. Любой из каналов сотовой связи использует при работе пару частот для дуплексной связи (одна частота на передачу, другая — на прием). Поэтому частоты излучения базовой и подвижной станций разнесены.

Контроллер базовых станций может управлять несколькими БС. Он координирует распределение радиоканалов, контролирует соединения и регулирует их очередность, обеспечивает работу с прыгающей частотой, кодирует и декодирует сообщения, выполняет ряд других функций.

Центр коммутации подвижной связи обеспечивает все виды соединений, в которых нуждается в процессе работы подвижная станция. Центр коммутации — это АТС системы сотовой связи. Он имеет интерфейс между фиксированными сетями связи и передачи данных и сетью подвижной станции. Центр коммутации обеспечивает маршрутизацию вызовов, функции управления вызовами, коммутации радиоканалов. Он же поддерживает процедуры обеспечения безопасности, применяемые для управления доступом к радиоканалам.

Центр коммутации осуществляет постоянное слежение за подвижными станциями. Для этого используются регистр положения и регистр перемещений. В регистре положения хранится та часть информации о местоположении подвижной станции, которая позволяет доставлять вызов. Этот регистр содержит международный идентификационный номер мобильного абонента (IMSI) и некоторые другие данные.

Регистр перемещений контролирует перемещение мобильной станции из соты в соту. При каждом таком перемещении в регистр заносится новая информация о номере соты и некоторая другая информация.

Стандарт GSM предусматривает основательные меры по защите информации: обеспечивает аутентификацию сообщений, секретность передаваемых данных, секретность направления вызова.

Для исключения несанкционированного использования ресурсов системы связи в стандарт введены и определены механизмы

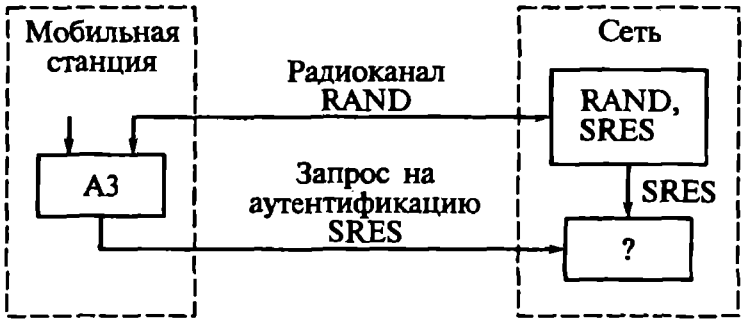


Рис. 6.7. Процедура аутентификации мобильного абонента

аутентификации — удостоверения личности абонента. Каждый абонент на время пользования системой получает стандартный модуль подлинности абонента — SIM-карту, которая содержит международный идентификационный номер подвижного абонента IMSI, индивидуальный ключ аутентификации K_i и алгоритм аутентификации A3. На основе этой информации в результате взаимного обмена данными между подвижной станцией и сетью осуществляется полный цикл аутентификации и разрешается доступ абонента к сети. Процедура проверки сетью подлинности абонента реализуется следующим образом.

Сеть передает случайное число RAND на подвижную станцию. Подвижная станция, используя алгоритм A3, вычисляет значение отклика SRES как функцию RAND и ключа K_i:

$$SRES = A3(K_i, RAND) \qquad (6.1)$$

и посылает вычисленное значение SRES в сеть. Сеть сверяет значение принятого SRES с результатом собственного вычисления. Если оба значения совпадают, подвижная станция получает разрешение пользоваться ресурсами сети: передавать и принимать сообщения. В противном случае связь прерывается, и индикатор подвижной станции должен показать, что опознание не состоялось.

По причине секретности вычисление SRES происходит в рамках SIM. Несекретная информация (такая как ключ K_i) не подвергается обработке в модуле SIM.

Процедура аутентификации иллюстрируется рис. 6.7.

Для обеспечения секретности передаваемой по радиоканалу информации ее шифруют. В стандарте используется алгоритм шифрования с открытым ключом RSA (см. гл. 5). Алгоритм формирования ключей шифрования A8 хранится в памяти SIM-карты. Одновременно с вычислением отклика SRES аппаратура подвижной станции определяет и ключ шифрования K_c по правилу

$$K_c = A8(K_i, RAND). \qquad (6.2)$$

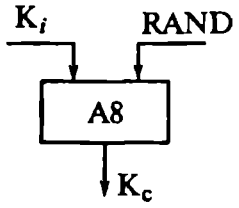


Рис. 6.8. Формирование ключа криптозащиты сообщений в соответствии со стандартом GSM

Для исключения риска утраты ключа он не передается по радиоканалу, а вычисляется и сетью, и абонентским терминалом одновременно с использованием одних и тех же данных и единого алгоритма A8 (рис. 6.8). Для обеспечения секретности вычисление ключа K_c производится в SIM.

Кроме случайного числа RAND сеть посылает подвижной станции идентификационную числовую последовательность. Это число связано с истинным значением K_c и позволяет избежать формирования ложного ключа. Число хранится подвижной станцией и содержится в каждом первом сообщении, передаваемом в сеть. Некоторые сети принимают решение о наличии числовой последовательности действующего ключа шифрования в случае, если необходимо приступить к опознаванию или если выполняется предварительное опознавание, используя правильный ключ шифрования. Но иногда это допущение реально не обеспечивается.

Для установки режима шифрования сеть передает подвижной станции команду CMC (Ciphering Mode Command) на переход в режим шифрования. После получения команды CMC подвижная станция, используя имеющийся у нее ключ, включает режим криптографического преобразования сообщений. Поток передаваемых данных шифруется поточным шифром бит за битом с использованием алгоритма шифрования A5 и ключа шифрования K_c. Процедура установки режима шифрования иллюстрируется на рис. 6.9.

Для исключения идентификации абонента на основе перехвата сообщений, передаваемых по радиоканалу, каждой мобильной станции системы сотовой связи присваивается временный международный идентификационный номер пользователя — TMSI (Time Mobile Subscriber Identity), который действителен только в пределах зоны обслуживания с идентификационным номером LAI

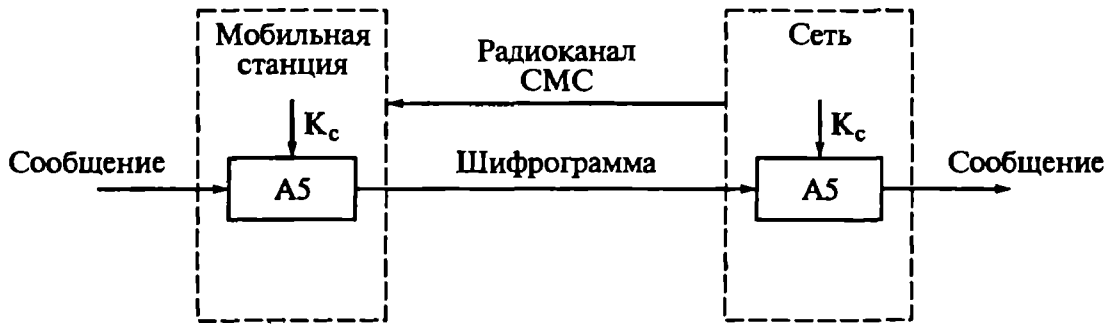


Рис. 6.9. Шифрация сообщений в соответствии со стандартом GSM

(Location Area Identification). В другой зоне обслуживания абоненту присваивается новый TMSI. Если подвижная станция переходит в новую зону обслуживания, то ее TMSI должен передаваться вместе с LAI той зоны, в которой TMSI был присвоен абоненту.

При выполнении процедуры корректировки местоположения по каналам управления между подвижной и базовой станциями происходит двухсторонний обмен служебными сообщениями. Эти сообщения содержат временные номера пользователей TMSI. В этом случае в радиоканале необходимо обеспечить секретность смены TMSI и его принадлежность конкретному абоненту.

В момент эстафетной передачи подвижная станция уже зарегистрирована в регистре перемещения с временным номером TMSI, соответствующим TMSI прежней зоны обслуживания. При входе абонента в новую зону осуществляется процедура опознавания, которая проводится по старому, зашифрованному в радиоканале TMSI, передаваемому одновременно с номером LAI зоны обслуживания. Последний сообщает центрам коммутации и управления информацию о направлении перемещения подвижной станции и позволяет запросить прежнюю зону расположения о статусе абонента, а также его данные, исключив обмен этими служебными сообщениями по радиоканалам управления. При этом по каналу связи сообщение передается как зашифрованный информационный текст с прерыванием сообщения в процессе эстафетной передачи всего на 100...150 мс.

Таким образом, в соответствии с рассмотренными механизмами обеспечения информационной безопасности, действующими в стандарте GSM, секретными считаются следующие данные:

RAND — случайное число, используемое для аутентификации подвижного абонента;

SRES — значение отклика, ответ подвижной станции на полученное случайное число;

K_i — индивидуальный ключ аутентификации пользователя, используемый для вычисления значения отклика и ключа шифрования;

K_c — ключ шифрования, используемый для шифрования-дешифрования сообщений, сигналов управления и данных пользователя в радиоканале;

A3 — алгоритм аутентификации, используемый для вычисления значения отклика из случайного числа с использованием ключа K_i;

A8 — алгоритм формирования ключа шифрования, используемый для вычисления ключа K_c из случайного числа с использованием ключа K_i;

A5 — алгоритм шифрования-дешифрования сообщений, сигналов управления и данных пользователя с использованием ключа K_{ic};

CKSN — номер ключевой последовательности шифрования, который позволяет избежать использование разных ключей на передающей и приемной сторонах, но указывает на действительное число K_c;

TMSI — временный международный идентификационный номер пользователя.

Основным объектом, отвечающим за все аспекты безопасности, является центр аутентификации. Этот центр может быть отдельным объектом или входить в состав какого-либо оборудования, например в регистр местоположения.

Именно центр аутентификации формирует индивидуальные ключи аутентификации пользователей K_i и соответствующие им международные идентификационные номера абонентов IMSI, формирует набор RAND/SRES/K_c для каждого IMSI и раскрытие этих групп для регистра положения при необходимости эстафетной передачи мобильного абонента.

6.3.2. Защита информации в цифровых системах мобильной связи с кодовым разделением каналов

Основная цель разработки сотовых систем подвижной радиосвязи общего пользования с кодовым разделением каналов CDMA (Code Division Multiple Access) состояла в том, чтобы увеличить абонентскую емкость системы и эффективность использования выделенного спектра частот.

Структура подвижной сети стандарта CDMA является иерархической (рис. 6.10). Внутри помещений организуются пикоячейки радиусом до 100 м с очень высокой пропускной способностью, определяемой большой плотностью абонентов. В городах и других населенных пунктах, в пешеходных зонах создаются микроячейки с радиусом обслуживания до 1 км. Сотовые системы, обслуживающие абонентов в автомобилях, оперируют макроячейками радиусом до нескольких десятков километров. Наконец, автомобильный и железнодорожный транспорт, воздушные, морские и речные суда, распределенные по территории с малой плотностью абонентов, могут обслуживаться спутниковым сегментом системы с использованием гиперячеек размерами в сотни и тысячи километров. Для реализации такой системы выделен весьма широкий диапазон рабочих частот, охватывающий полосы 1886...2025 МГц и 2110...2200 МГц, включая в себя 1980...2010 МГц и 2170...2200 МГц для ее спутникового сегмента.

Сети мобильной связи, построенные на базе технологии CDMA с использованием широкополосных шумоподобных сигналов (ШПС), имеют ряд достоинств. Такие системы обладают высокой устойчивостью к действию разного рода сосредоточенных по спектру помех, способностью эффективно функционировать в усло-

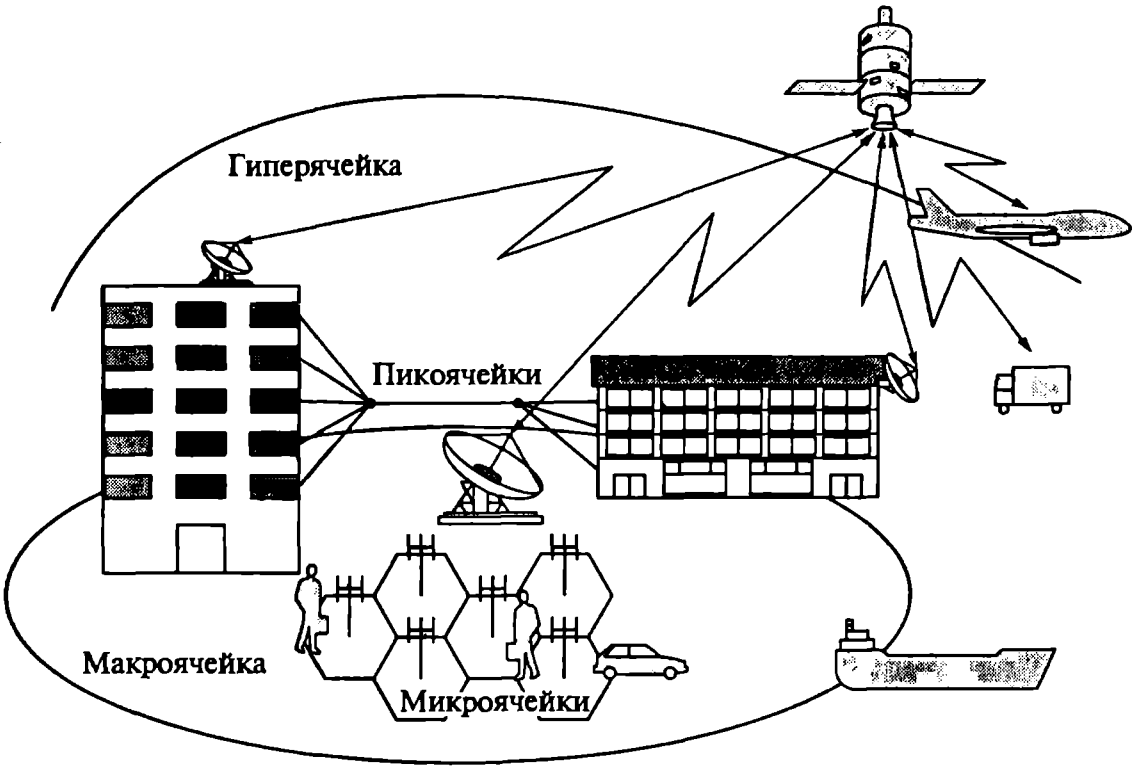


Рис. 6.10. Структура подвижной сети стандарта CDMA

виях многолучевого распространения сигнала и рядом других положительных качеств.

Раздельная обработка сигналов дает также возможность реализовать надежную эстафетную передачу путем так называемой мягкой передачи управления (soft handoff). Во время этой процедуры абонентская станция поддерживает связь одновременно с несколькими (двумя или тремя) базовыми станциями, что облегчает процедуру переключения с одной БС на другую и увеличивает вероятность нормального завершения переключения.

Технология CDMA обеспечивает более высокую эффективность использования частотного спектра по сравнению с той, которая достигается при частотном (FDMA) и временном (TDMA, как в системах стандарта GSM) разделении каналов.

Поскольку все базовые станции работают на одной несущей частоте, в системе не требуется частотное планирование. Это упрощает как начальное развертывание системы CDMA, так и ее последующее развитие. Вместе с тем, при проектировании системы должен соблюдаться баланс мощности сигналов (планирование мощности), чтобы ограничить уровень взаимных помех и улучшить электромагнитную совместимость. Важное свойство CDMA заключается в уменьшении средней излучаемой мощности. Низкая средняя излучаемая мощность системы CDMA позволяет также минимизировать уровень электромагнитных излучений, воздействующих на человека, что снижает ее биологическую опас-

ность. По тем же причинам маловероятно непреднамеренное не гативное воздействие систем CDMA на работу различного рода электронных устройств.

В системе не выделяется заранее определенный частотный (временной) ресурс для организации канала связи между базовой и абонентской станциями, поэтому такая система обладает свойствами эластичности, т.е. возможностью динамического перераспределения ресурсов, что является более простой задачей по сравнению с динамическим перераспределением частотных каналов.

Для систем с CDMA характерна повышенная конфиденциальность обмена сообщениями, поскольку каждому абоненту присваивается свой широкополосный сигнал, имеющий индивидуальную и достаточно сложную структуру.

Современные системы с CDMA используют прямое расширение спектра частот на основе применение 64 видов последовательностей, сформированных по закону функций Уолша. Затем этот сигнал модулирует несущую. Спектр выходного сигнала расширяется псевдошумовой последовательностью.

В приемнике происходит сворачивание спектра на коррелято ре, согласованном с расширяющей спектр псевдошумовой последовательностью. На практике в приемнике мобильной станции имеются несколько корреляторов для приема сигналов с разным временем распространения и одновременной работы с несколькими базовыми станциями.

Стандарт радиоинтерфейса IS-95 CDMA обеспечивает высокую степень защиты передаваемых сообщений и данных об абонентах. Прежде всего он имеет более сложную структуру, чем у стандарта GSM и обеспечивает передачу сообщений кадрами с использованием канального кодирования и перемежения с последующим расширением передаваемых сигналов с помощью составных широкополосных сигналов, сформированных на основе 64 видов последовательностей Уолша и псевдослучайными последовательностями с числом элементов $(2^{15} - 1)$ и $(2^{42} - 1)$.

Безопасность связи обеспечивается также применением процедур аутентификации и шифрования сообщений.

В подвижной станции хранится один ключ *A* и один набор общих секретных данных, которые используются при работе как в режиме с частотным разделением каналов, так и в режиме CDMA. Подвижная станция может передавать электронную цифровую подпись для аутентификации, состоящую из 18 бит. Эта информация передается в начале сообщения (в ответе подвижной станции на запрос сети при поиске станции), добавляется к регистрационному сообщению или пакету данных, передаваемых по каналу доступа. Предусматривается возможность обновления общих секретных данных в подвижной станции.

Шифрование сообщений, передаваемых по каналу связи, осуществляется также с использованием процедур стандарта IS-54B.

В стандарте IS-95 CDMA используется режим «частный характер связи», который обеспечивается наложением на сигнал секретной маски в виде длинного кода (гаммирование).

6.4. Методы представления речевого сигнала

Для повышения качества передачи речевых сообщений по сетям связи и обеспечения информационной безопасности функционирования таких сетей применяются методы преобразования речи.

Наиболее простыми являются методы прямого цифрового преобразования речевых сигналов, при которых каждый отсчет речевого сигнала формируется и преобразуется независимо от других. Поэтому такие способы кодирования иногда называются скалярными. К ним можно, в частности, отнести хорошо известные импульсно-кодовую модуляцию (ИКМ) или дельта-модуляцию (ДМ), а также их многочисленные разновидности.

Другими более сложными и эффективными являются способы представления речевых сигналов, при которых результирующие цифровые значения представлений вычисляются как функции от нескольких значений временной функции или, более точно, являются функционалами от временных функций, взятых на некотором участке анализа. Такие способы кодирования иногда называются векторными. К векторным относятся способы представления, использующие кратковременные амплитудные спектры, частотные (или полосовые) параметры речевых сигналов, коэффициенты линейного предсказания и некоторые другие.

Полное количество информации, содержащейся в непрерывном речевом сигнале, определяется его длительностью, шириной спектра и динамическим диапазоном. Естественно, не вся информация, заключенная в сигнале, полезна. Например, информация, заключенная в фазе речевого сигнала, не содержит сведений о звуках речи и не нужна для ее восприятия.

Динамический диапазон сигнала определяется соотношением пиковой мощности сигнала к минимальной или разностью их уровней

$$D = \log \frac{P_{\max}}{P_{\min}} = \log P_{\max} - \log P_{\min}. \qquad (6.3)$$

Длительность сигнала T, ширина частотного диапазона F и динамический диапазон D — три измерения сигнала. Их произведение

$$V = DFT = FT \log \frac{P_{max}}{P_{min}} \qquad (6.4)$$

есть объем сигнала.

Пропускная способность канала связи также определяется длительностью работы канала, полосой пропускаемых частот и динамическим диапазоном $V_к = T_к F_к D_к$. В этом случае нижняя граница динамического диапазона определяется уровнем помех, а верхняя – перегрузкой канала связи.

Для информационного сжатия (уменьшения объема) сигнала применяется метод компрессии и последующего экспандирования (сокращенно, методы компандирования). Используются непосредственные, параметрические и речеэлементные методы компандирования речевого сигнала. Непосредственные методы делятся на аналоговые и дискретные. Первые основаны на непосредственной компрессии объема речевого сигнала в передающей части тракта путем сжатия любого из входящих в (6.4) трех его измерений или любой из их комбинаций с восстановлением объема сигнала в этих измерениях на приемном конце тракта (в пределах возможностей соответствующего метода). Поэтому в передающей части тракта сигнал в соответствующих измерениях деформируется. При деформации сигнал подвергается искажениям, которые могут рассматриваться как своего рода помехи. Частично эти искажения могут быть скомпенсированы в приемной части тракта, а некоторые из них совсем не поддаются такой компенсации. Для всех методов компандирования характерно сохранение, в основном, микроструктуры речевого сигнала.

В соответствии с тем, по какому измерению объема сигнала происходит компандирование, оно подразделяется на компандирование динамического диапазона, частотного диапазона и временное компандирование.

К методам непосредственной компрессии аналогового речевого сигнала можно отнести и методы ограничения сигнала. При ограничении сигнала по динамическому или частотному диапазону, а также во времени некоторое количество информации исключается за счет выбрасывания отдельных участков частотного диапазона, ограничения громких звуков или исключения среднего (установившегося) временного участка длительных звуков речи. При жестком ограничении сигнала восстановить его на приемном конце, как правило, не удается, т.е. в этом случае имеет место компрессия сигналов, но невозможно его экспандирование.

К методам обработки и непосредственного компандирования речевого сигнала в дискретной форме относятся методы импульсно-кодовой модуляции (ИКМ) и дельта-модуляции (ДМ) в различных модификациях и комбинациях. Так как при двоичной системе кодирования амплитуда сигнала остается неизменной, при

соответствующей компрессии речевого сигнала снижается или скорость, или время передачи сигнала. Тем самым уменьшается его объем с последующим его восстановлением на приемном конце.

Конечно, при этом в передающей части тракта возникают искажения, рассматриваемые как помехи, и сигнал не всегда может быть точно восстановлен. Существенным отличием непосредственных методов компандирования речевого сигнала в дискретной форме от других дискретных методов является то, что сигнал восстанавливается сам по себе, т. е. для восстановления сигнала достаточно иметь только последовательность импульсов без передачи какой-либо вспомогательной информации или управляющих сигналов.

6.4.1. Компрессия аналогового речевого сигнала

Речевой сигнал, с учетом разброса его параметров и индивидуальных особенностей для разных людей, имеет довольно широкий динамический диапазон. Он требует для передачи по каналу связи низкого уровня помех и высокой верхней границы неискаженной передачи. В реальных каналах верхняя граница бывает жестко ограничена требованиями согласования при переходах в другие каналы, перегрузкой усилителей и другими причинами, а уровень помех бывает довольно высоким. Поэтому пропустить речевой сигнал через канал без искажений невозможно из-за перегрузки сильных и маскировки помехами слабых по уровню звуков речи. Выход один — сжать или ограничить динамический диапазон речевого сигнала до величины динамического диапазона канала, повысив тем самым помехозащищенность передачи речи и ее разборчивость на приеме. Компрессия динамического диапазона необходима и для обработки речевого сигнала в тех случаях, когда он должен подвергаться преобразованиям типа вокодерных (от *англ.* voice + coder — собирательное название устройств преобразования речевых сигналов).

Кроме того, при передаче по системе связи желательно сжать частотный диапазон речевого сигнала. Спектр речи занимает полосу частот до 20 кГц, но верхние частоты спектра имеют малую информативность, поэтому сигнал ограничивают по полосе частот. Известно (экспериментально определено и закреплено в отечественных и международных стандартах), что для радиотелефонной передачи речи, особенно в условиях флуктуационных шумов с равномерной плотностью по частоте, ограничение частотного диапазона сверху частотой 3400 Гц и снизу частотой 300 Гц не вызывает заметного снижения разборчивости речи. Дальнейшее сужение частотного диапазона приводит уже к существенному ухудшению разборчивости. В отсутствии шумов и помех достаточная величина разборчивости речи получается даже при передаче

полосы частот шириной 1000 Гц. Соответственно минимальная частота дискретизации для неискаженной передачи речи должна быть $f_д \geq 2f_в \approx 7$ кГц. Но с учетом неидеальности фильтров, восстанавливающих сигнал после цифроаналоговых преобразований на приемной стороне, выбирают $f_д = 8$ кГц. Наиболее простой способ сокращения частотного диапазона речи — это ограничение его сверху и снизу.

Неоднократно предлагался метод сужения частотного диапазона речи путем равномерного деления его на ряд полос и передачи части каждой из них. Этот метод, как не учитывающий распределение речевой информации по частотному диапазону, приводит лишь к тому, что теряется информация и снижается помехозащищенность передачи.

При временном компандировании на передающей стороне линии из речи исключаются некоторые временные интервалы. Вследствие такой дискретизации в передаче образуются паузы, которые могут быть заполнены другой передачей. На приемной стороне интервалы в передаче заполняются предшествующим паузе отрывком речи. Иногда после исключения из речи временных интервалов остающиеся отрывки речи растягиваются на весь интервал до следующего отрывка. Вследствие этого понижаются частоты колебаний и весь частотный диапазон речи соответственно сжимается. Таким образом, временное компандирование превращается в частотное.

Сначала было предложено делить речевой сигнал по времени на равные интервалы и затем исключить их так, чтобы оставались речевые отрывки одинаковой длины и на равных интервалах. Но и этот метод оказался неэффективным, так как длительные звуки сохраняют большую избыточность, а короткие исчезают. В дальнейшем этот метод был усовершенствован с учетом того, что главным признаком распознавания звуков речи являются временные изменения форматных частот. Соответственно этому принципу нет смысла в полной передаче установившихся значений формантных образов, а передавать следует только те временные участки звуков речи, в которых содержатся изменения формантных частот. Поскольку гласные звуки и ряд согласных имеют длительный период, в течение которого формантные изменения не происходят, то такие участки передачи могут быть без ущерба для речи почти исключены из передачи в компрессирующем устройстве. На приеме они могут быть легко восстановлены по предыдущим отрывкам и вставлены в паузы между отрывками звуков речи.

Анализ участков речи, имеющих постоянные формантные частоты, показал, что совершенно свободно, без сколько-нибудь заметного снижения разборчивости и качества звучания речи, может быть исключено примерно 54 % всей длительности передачи. Таким образом, двухкратная временная компрессия речи не дол-

жна вносить искажений при передаче. Была проведена экспериментальная проверка зависимости разборчивости речи от величины временной компрессии и ее периода разделения (период разделения — это промежуток времени, занимаемый оставляемым и выбрасываемым отрывками речи). За коэффициент компрессии принимается отношение паузы между отрывками к периоду разделения. Измеряется слоговая разборчивость для фонетически сбалансированных, но бессмысленных слогов, которые собираются в специальные тестовые таблицы. Коэффициент компрессии изменялся в пределах 0,4...0,9, величина пауз — 0,01...0,24 с.

6.4.2. Дискретные методы передачи и обработки речевого сигнала

Наиболее естественным и, видимо, исторически самым ранним способом перевода речевых сигналов в цифровую форму является импульсно-кодовая модуляция [21], при которой речевой сигнал $s(t)$ подвергается дискретизации по времени и квантованию по уровню. Такие преобразования приводят к возникновению шумов дискретизации и квантования, т.е. снижению отношения сигнал-шум. В соответствии с рекомендациями МККТТ [10] для стандартного телефонного канала принято, что частота дискретизации должна составлять 8 кГц, квантование должно производиться по квазилогарифмическому закону, а кодирование — с восемью двоичными символами. При этом требуется обеспечить скорость передачи оцифрованной речи по каналу связи 64 Кбит/с, тогда как практически допустимые скорости передачи данных по каналу ТЧ обычно не превосходят 2,4 Кбит/с (по выделенным каналам 4,8 или 9,6 Кбит/с).

Другим, хорошо известным способом скалярного кодирования речи является дельта-модуляция (ДМ) со всеми ее разновидностями. При ДМ по каналу связи передаются не сами квантованные значения сигнала, а только их приращения.

Передача речи при помощи дельта-модуляции сопровождается искажениями двух видов. Это частотные ограничения (перегрузка по крутизне) и дробление.

Частотные ограничения обусловлены тем, что приращения преобразуемого сигнала всегда фиксируются с запаздыванием, и это запаздывание тем больше, чем быстрее изменяется сигнал. Уменьшить искажения этого вида можно за счет увеличения частоты дискретизации. Дробления проявляются на тех временных интервалах, где сигнал либо постоянен, либо изменяется очень медленно. Эффект дробления уменьшается при использовании логической обработки и нелинейной шкалы квантования, т.е. адаптации параметров модуляции к характеристикам преобразуемого сигнала.

При использовании всех мер борьбы с искажениями качество передачи речи по каналам с дельта-модуляцией удовлетворяет требованиям МККТТ при частоте дискретизации порядка 40...50 кГц.

Расчеты значений отношения сигнала к шуму для адаптивной дельта-модуляции (АДМ) показали, что при скорости передачи информации больше 40 Кбит/с предпочтительнее использовать ИКМ, а при меньше 40 Кбит/с эффективнее оказывается АДМ.

При использовании способов векторного анализа и синтеза речевых сигналов используется расчленение сигналов на тональные и шумовые интервалы, соответствующие произнесению тональных (вокализованных) и глухих (шумовых) звуков речи. Этот процесс эквивалентен определению сигналов тон—шум $\varphi(t_i)$, которые вычисляют через некоторые временны́е интервалы, длиной 10...25 мсек: $\varphi(t_1)$, $\varphi(t_2)$,.... Здесь $\varphi(t_i) = 1$ или 0, если в окрестности точки t_i был произнесен тональный или шумовой звук.

Одновременно переключательная функция $\varphi(t_i)$ на тональных участках речи определяет значения частоты основного тона $f_0(t_i)$. Это основная частота колебаний голосовых связок. Частоты основного тона $f_0(t_i)$ и сигналы $\varphi(t_i)$ особо важную роль играют при синтезе речи.

Спектрально-полосные методы кодирования речи. На их основе строятся полосные вокодеры. Спектр речевого сигнала на передающем конце разделяется узкополосными фильтрами на частотные полосы (спектральные каналы). В каждом канале путем детектирования и сглаживания фильтрами нижних частот определяются огибающая и средняя интенсивность сигнала. Информация об этих величинах передается в аналоговой или цифровой форме по каналу связи. Кроме того, передаются сигналы тон—шум и значения частоты основного тона.

На приемной стороне $\varphi(t_i)$ управляет подключением генератора шума или генератора импульсов, частота которых определяется частотой $f_0(t_i)$. С помощью этих генераторов создается широкополосный сигнал, который, как и на передающем конце, разделяется на частотные полосы с помощью фильтров. Колебания с выходов фильтров умножаются на значения огибающих канальных сигналов и суммируются друг с другом. Полученный синтезированный сигнал приближенно отображает исходный естественный речевой сигнал, преобразованный на передающей стороне.

Число спектральных каналов обычно варьируется от 7 до 20. Причем с увеличением числа каналов повышаются разборчивость и качество синтезированной речи, но возрастают и требования к пропускной способности канала связи.

Согласно экспериментальным данным, для передачи огибающей сигнала на выходе каждого канала достаточно провести ее дискретизацию с частотой 50 Гц и квантование с помощью трехразрядного двоичного кода, а для частоты основного тона соот-

ветственно 100 Гц и пятиразрядного кода. Следовательно, для 12-канального полосного вокодера потребуется передавать по каналу связи 2300 бит/с, а для 20-канального вокодера — 3500 бит/с.

Основным недостатком полосных вокодеров является техническая сложность и громоздкость реализации, обусловленная большим количеством используемых фильтров. Качество восстанавливаемой речи может снижаться из-за того, что в полосе пропускания фильтра на тональных звуках может оказаться несколько гармоник основного тона и число таких гармоник может меняться во времени. Кроме того, снижение качества обуславливают возможные ошибки при определении $F_0(T_i)$ и $\varphi_0(T_i)$.

Вместо значений частот основного тона можно использовать так называемый основной канал: передавать естественную речь, взятую в частотной полосе, например 250...750 Гц. При этом качество речи, как правило, улучшается, однако скорость передачи увеличивается до 10 Кбит/с. Поскольку передается не только преобразованная речь, в этом случае говорят о полосном полувокодере.

Формантные вокодеры. В них спектральная огибающая речевого сигнала аппроксимируется комбинацией нескольких простых резонансных кривых. Принципы построения формантного вокодера во многом аналогичны принципам естественного речеобразования и восприятия речи. Поскольку речевой тракт представляет собой комплекс резонаторов, резонансные частоты и добротности которых изменяются во времени в соответствии с управляющими сигналами, то и в формантном вокодере происходит выделение из речи управляющих сигналов, которые на приеме воздействуют на резонансные контуры и приближенно воспроизводят передаваемую огибающую спектра. Такими управляющими сигналами могут быть: частоты формант $f_i(t)$, где i — номер форманты, $i \in 1:I$; амплитуды формант $a_i(t)$; ширина их спектра на уровне 3 дБ — $\Delta f_i(t)$.

Существует много способов выделения формантных частот. Некоторые из них зависят даже от различных априорных определений понятия «формантная частота».

Для глухих звуков речи вместо форматных функций используются меняющиеся во времени нулевой, первый и второй моменты частотного спектра ($i = 1, 2, 3$):

$$M_0^{(i)}(t) = \int_{f_i} K(t,f)\,\mathrm{d}f, \text{ или } M_1^{(i)}(t) = \int_{f_i} fK(t,f)\,\mathrm{d}f,$$

$$\text{или } M_0^{(i)}(t) = \int_{f_i} f^2 K(t,f)\,\mathrm{d}f, \qquad (6.5)$$

где интегрирование ведется по первой, второй или третьей формантным областям. При реальных вычислениях, конечно, интег-

рирование заменяется суммированием и $M_s^{(i)}(t)$ вычисляется с некоторым шагом по t.

Момент $M_0^{(i)}(t_0)$ представляет собой интенсивность звука в i-й частотной области для заданного момента времени t_0, $\frac{M_1^{(i)}(t)}{M_0^{(i)}(t_0)}$ — среднюю частоту i-й части спектра, а величина

$$\frac{M_2^{(i)}(t_0)}{M_0^{(i)}(t)} - \left[\frac{M_1^{(i)}(t_0)}{M_0^{(i)}(t_0)}\right]^2 \qquad (6.6)$$

характеризует дисперсию i-й части спектра.

Приведенные функции от моментов заменяют для глухих звуков формантные функции и в дальнейшем формально будут обозначаться, как $a_i(t)$, $f_i(t)$ и $\Delta f_i(t)$.

Таким образом, при использовании формантного вокодера по каналу связи нужно передать дискретизированные по времени и квантованные по амплитуде величины: сигналы тон-шум, значения частоты основного тона и девять функций $a_i(t)$, $f_i(t)$, $\Delta f_i(t)$ $(i = 1, 2, 3)$.

Учитывая плавность изменения во времени выше перечисленных сигнальных параметров параметров, дискретизацию можно проводить с частотой 40 Гц, квантование $a_i(t)$, $f_i(t)$, $\Delta f_i(t)$ в среднем 16 уровнями (4 бита), а частоты основного тона — 128 уровнями (7 бит), что в сумме потребует канала связи с пропускной способностью, равной 1700 бит/с.

На одном из вокодеров формантного типа со скоростью 1200 бит/с получена слоговая разборчивость 80...82 %. Разборчивость речи при использовании формантных вокодеров и передаче со скоростью 2400 бит/с превышает разборчивость обычной телефонной связи.

На приемном конце линии связи при восстановлении (синтезе) речи, применяются управляемые формантные контуры, генератор шума, модуляторы, сумматоры.

Гармонические вокодеры. В гармоническом вокодере спектральная огибающая речевого сигнала $S(t, f)$ в момент времени t_0 приближается с помощью ее представления рядом по ортогональным функциям $\xi_j(f)$ $(j = 0, 1, 2, \ldots, m-1)$ и усечением этого ряда до m членов:

$$K(t_0, f) = \sum_{j=1}^{m} a_j(t_0)\xi_j(f)df;$$
$$a_j(t_0) = \int_{f_1}^{f_2} K(t_0, f)\xi_j(f)df, \qquad (6.7)$$

где $[f_1;f_2]$ — частотный интервал разложения спектральной огибающей речи.

В качестве ортогональных функций обычно выбираются функции:

$$1;\ \frac{\sin 2\pi(f-f_1)}{f_2-f_1}. \tag{6.8}$$

Таким образом, достигается та же цель, что при полосном или формантном вокодере: найти сравнительно небольшое количество числовых параметров, которые бы удовлетворительно приближали спектральную огибающую $S(t,f)$.

Как показали экспериментальные и расчетные данные, в канал связи достаточно передавать значения 5...7 коэффициентов Фурье (6.8) с темпом 50 Гц и квантовать их четырехразрядным двоичным кодом, а также зарезервировать 600 бит/с для передачи основного тона. Иначе говоря, достаточно иметь канал связи с пропускной способностью, не большей 2000 бит/с. Но следует признать, что гармонические вокодеры не нашли практического применения из-за отсутствия существенных преимуществ по сравнению с полосными или формантными вокодерами.

Метод линейного предсказания (ЛПР). Весьма эффективен для представления преобразованной речи. В последнее время он находит самое широкое применение как при анализе речевых сигналов, так и в вокодерных преобразованиях. Текущее значение $S(t_n)$ дискретизированного во времени речевого сигнала аппроксимируется линейной функцией, параметры которой формируются виде линейной комбинации предшествующих значений этого сигнала $S(t_{n-1})$, $S(t_{n-1})$:

$$S(t_n)=\sum_{i=1}^{P} a_i S(t_{n-i}) + e(t_n) = \tilde{S}(t_n) + e(t_n), \tag{6.9}$$

где P — порядок фильтра; $e(t_n)$ — ошибка предсказания; $\tilde{S}(t_n)$ — предсказанный сигнал; a_1, a_2,..., a_p — некоторые (действительные) коэффициенты, которые обеспечивают минимум (6.9).

Коэффициенты a_1, a_2, ..., a_p вычисляются при решении системы из P линейных уравнений, получающихся при обращении в нуль частных производных $\delta(n_0,n_1)$ по a_1, $a_2,\dots,a_p$ по каждому из $a_i\left(i=\overline{1,P}\right)$:

$$\sum_{i=1}^{P} a_i C_{ik}\left(t_{n_0},t_{n_1}\right) = -C_{0k}\left(t_{n_0},t_{n_1}\right),\quad k\in 1:P. \tag{6.10}$$

Линейное предсказание обладает несколькими преимуществами по сравнению с другими способами аппроксимации: оно по-

зволяет выделить периодические составляющие речи. Выбирая порядок P прогнозирующей функции, адекватно отражающей число формант в частотном спектре, можно установить, что пики передаточной функции фильтра часто соответствуют действительным формантам. Это свойство значительно уменьшает трудности, связанные с оценкой положения формант в непрерывной речи.

Самую эффективную компрессию, близкую к предельно достижимой, обеспечивают фонемные вокодеры. При нормальной речи средней интенсивности в секунду произносится 10 фонем из полного числа 66. Это значит, что для передачи каждой фонемы достаточно шести двоичных символов и можно ограничиться пропускной способностью канала 60 бит/с. Правда, в этих условиях не удается передать какие-либо индивидуальные особенности голоса (такая передача речи называется звучащим телеграфом). Этот недостаток наряду с тем, что при определении последовательности произносимых фонем будут допускаться частые ошибки, вовсе не компенсирует некоторую выгоду в уменьшении требований к пропускной способности канала связи.

6.4.3. Критерии оценки систем закрытия речи

Известны четыре основных критерия, по которым оцениваются характеристики качества устройств преобразования и защиты речевых сигналов [15]. Это критерии разборчивости речи, узнаваемости говорящего, степени защиты от перехвата, а также простота реализации, стоимость, ремонтопригодность и другие технические и эксплуатационные характеристики.

Поскольку основным показателем секретности передаваемых речевых сообщений является его неразборчивость при перехвате, сравнение по степеням защиты является определяющим моментом при выборе конкретной системы закрытия речи. В основном распределение по уровням закрытия речевых сообщений соответствует диаграмме, приведенной на рис. 6.11.

Говоря об уровне защиты или степени секретности систем закрытия речи, следует отметить, что эти понятия весьма условные. К настоящему времени не выработано на этот счет четких стандартов или правил. Однако в ряде источников основные уровни защиты определяют как тактический и стратегический, что в некотором смысле перекликается с понятиями практической и теоретической стойкости криптографических систем закрытия данных:

тактический, или низкий, уровень используется для защиты информации от подслушивания посторонними лицами на период времени, измеряемый минутами или днями. Существует большое количество простых методов, способных обеспечивать такой уровень защиты при приемлемой стоимости;

Рис. 6.11. Основные характеристики систем закрытия речи

стратегический, или высокий, уровень защиты информации от перехвата используется в ситуациях, подразумевающих, что высококвалифицированному, технически хорошо оснащенному специалисту потребуется для дешифрования перехваченного сообщения период времени от нескольких месяцев до нескольких лет.

Часто используется и понятие средней степени защиты, занимающее промежуточное положение между тактическим и стратегическим уровнями закрытия.

Как правило, аналоговые скремблеры используются там, где применение цифровых систем закрытия речи затруднено из-за наличия возможных ошибок передачи (наземные линии связи с плохими характеристиками или каналами дальней радиосвязи). Они обеспечивают тактический уровень защиты и хорошо предохраняют переговоры от посторонних «случайных ушей», имеющих ограниченные ресурсы, будь то соседи или сослуживцы. Для таких применений годятся системы со статическим закрытием, т. е. осуществляющие шифрование по фиксированному ключу.

Если же необходимо сохранить конфиденциальность информации в условиях угроз со стороны возможных противников, обладающих достаточным техническим и специальным оснащением, нужно применять аналоговые скремблеры среднего уровня закрытия с динамически меняющимся в процессе разговора ключом. Естественно, что подобные системы будут дороже, чем системы закрытия с фиксированным ключом, однако они настолько осложняют работу по созданию дешифрующего алгоритма, что время, потраченное на это, значительно обесценит добытую информацию из перехваченного сообщения. В таких устройствах закрытия как правило перед началом сообщения передается синхропоследовательность, содержащая дополнительную информацию о ключе.

Если есть основательные предположения о том, что для добы вания ценной информации противник может воспользоваться воз можностями высококвалифицированных специалистов и их тех ническим арсеналом, для гарантированного предотвращения утс чки информации применяют системы закрытия речи, обеспечи вающие стратегическую (самую высокую) степень защиты. С та кой функцией могут справиться лишь устройства дискретизации речи с последующим шифрованием и новый тип аналоговых скрем блеров. Они используют методы преобразования аналогового ре чевого сигнала в цифровую форму, затем применяют методы крип тографического закрытия, аналогичные тем, что используются для закрытия данных, после чего результирующее закрытое сообще ние преобразуется обратно в аналоговый сигнал и подается в ли нию связи. Для раскрытия полученного сигнала на приемном конце производятся обратные преобразования. Эти новейшие гибридные устройства легко адаптируются к существующим коммуникаци онным сетям и предлагают значительно более высокий уровень защиты речевых сообщений, чем традиционные аналоговые скрем блеры, с сохранением всех преимуществ последних в разборчиво сти и узнаваемости восстановленной речи.

6.5. Функциональная схема системы закрытой связи

Традиционно применяются три типа устройств, обеспечиваю щих защиту передаваемой речевой информации: маскираторы, скремблеры и устройства с передачей речи в цифровом виде, на зываемые вокодерами.

Маскираторы защищают речевой сигнал только от прямого про слушивания, совершая над ним некоторые обратимые детерми нированные преобразования. Маскираторы достаточно дешевы и используются, как правило, в бытовых приложениях. Примером засекречивающих преобразований в маскираторах может служить рассмотренная выше инверсия спектра.

Скремблеры — это устройства, реализующие засекречивающие преобразования с применением шифраторов. В зависимости от типа применяемых шифрующих преобразований эти устройства могут обеспечивать достаточно высокую степень защищенности, одна ко чем сложнее преобразования, тем более высокие требования предъявляются к качеству каналов связи для сохранения прием лемого качества дешифрованного сигнала. Вместе с тем даже при использовании достаточно сложных аналоговых преобразований защищенность речевого сигнала при скремблировании нельзя счи тать гарантированной.

Лучшие характеристики скрытия обеспечивают шифраторы речевых сигналов, преобразованных в цифровую форму перед их

передачей. Они способны обеспечить гарантированную защиту передаваемых по каналам связи речевых сообщений. Шифрацию речевых сигналов, представленных в цифровом виде, производят специальные цифровые процессоры.

Общий вид функциональной схемы, иллюстрирующей процессы обработки речи в защищенной системе сотовой связи, приведен на рис. 6.12.

При проектировании систем цифрового закрытия речевых сигналов используется довольно ограниченный набор технических решений.

Во-первых, цифровая последовательность параметров речи с выхода вокодерного устройства подается на вход шифратора, где подвергается преобразованию по одному из криптографических алгоритмов, затем поступает через модем в канал связи. На приемной стороне осуществляются обратные операции по восстановлению речевого сигнала, для которых используются модем и дешифратор. Модем представляет собой отдельное устройство, обеспечивающее передачу данных по одному из протоколов, рекомендованных МККТТ. Шифрующие и дешифрующие функции обеспечиваются либо аппаратно в отдельных специальных устройствах, либо программно, с использованием процессора в составе аппаратуры вокодера.

Во-вторых, шифрующие и дешифрующие функции обеспечиваются самим модемом (это засекречивающий модем). Цифровой поток, несущий информацию о параметрах речи, с выхода вокодера непосредственно поступает на такой модем.

Рис. 6.12. Процессы преобразования и обработки речи в защищенной системе сотовой связи

Контрольные вопросы

1. Системы, сети и каналы связи — в чем различие технических средств, именуемых этими терминами?
2. Перечислите основные информационные угрозы системам и сетям связи.
3. Какие аналоговые методы используются для защиты речевой информации в системах связи?
4. Как обеспечивается секретность в системах связи, использующих стандарт GSM?
5. Как защищают информацию в сотовых системах связи с кодовым разделением?
6. Какие преобразования речевых сигналов используются при вокодерной связи?

Глава 7
ЗАЩИТА ИНФОРМАЦИИ В ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ И СРЕДАХ

7.1. Угрозы информационным ресурсам и информационные атаки на вычислительные системы

Основные угрозы информации, хранимой и перерабатываемой в вычислительных системах, сетях и средах, — это видоизменение и ограничение доступа к данным, нарушение конфиденциальности. В силу нескольких причин угрозы безопасности информации в вычислительных средах особенно опасны. Во-первых, появились новые информационные технологии, увеличившие не только мощность, но и доступность компьютеров. Во-вторых, развитие коммуникаций и универсальность международных стандартов построения информационных систем и протоколов обмена данными открыли возможность дистанционного доступа к информации без физического преодоления систем их охраны и защиты. Особенно остро проблемы защиты данных проявляются при использовании ЭВМ для обработки и хранения информации конфиденциального и секретного характера.

Несанкционированное раскрытие содержания информации встречается наиболее часто. Последствия раскрытия содержания конфиденциальных данных могут быть самыми разными. Так, похищение важных материалов, содержащих государственные или стратегические военные секреты, чревато непредсказуемыми по тяжести последствиями для одной стороны и огромным выигрышем для другой. В большинстве конфликтных взаимодействий (от военных конфликтов до переговоров) знание секретных сведений о возможностях, планах, приоритетах и позициях противника существенно облегчают выбор и оптимизацию стратегии поведения для достижения желаемого результата. Если похищены результаты некоторого исследования или проекта, то для коллектива авторов это существенная потеря, возможно даже катастрофа, имеющая следствием большие материальные и моральные потери. Однако если результаты работы уже опубликованы, факт похищения может даже послужить на пользу работе, сделав ей рекламу. Недаром часто приходится слышать парадоксальные на первый взгляд сообщения об утечке информации, организованной ее владельцами.

Часто конфиденциальная информация, более или менее тщательно оберегаемая от раскрытия, содержит персональные сведения: биографии, анкетные данные, истории болезни, письма, све-

дения о доходах и расходах. Обычно данные о людях наиболее важны для них самих, но, как бы это не муссировалось в детективных романах и фильмах, мало что значат для похитителей. Иногда личные данные могут использоваться для компрометации. Тем не менее персональная информация ценна сама по себе, основной ущерб от ее разглашения — личное несчастье человека.

Другое дело — раскрытие стратегической управляющей информации. Если вскрыт долгосрочный план развития производства или анализ конъюнктуры на рынке, то потери для держателя этой информации будут невелики (владельцы информации все равно останутся при своих сведениях). Но такие сведения очень ценны для конкурентов. Поэтому хотя несанкционированный доступ и раскрытие содержания конфиденциальной информации случаются довольно часто, но редко когда приносят существенный вред.

Большую опасность представляют искажения информации. Во многих организациях важные данные (инвентарные описи, графики работ, контракты) хранятся в компьютерных файлах. Если такие данные будут искажены или стерты, то работа парализуется. Самое опасное в этом то, что шифрация данных не гарантирует их устойчивость против искажения: в примитивных криптографических системах искажения могут быть внесены и без знания ключа. По-видимому, наиболее уязвима для искажения экономическая информация. Потери от ее искажения могут быть чрезвычайно велики.

Незаконное, несанкционированное ограничение доступа к данным не так опасно, как сознательная дезинформация. Действительно, обнаружение информационных ограничений, необоснованно сужающих полномочия и затрудняющих работу человека в автоматизированной системе, всегда рефлексивно навязывают желание преодолеть это препятствие. И в таком конфликте побеждает, как правило, человек. Преодоление ограничений доступа требует затраты некоторого ресурса, прежде всего времени, но и такая потеря оперативности работы вычислительной системы, среды или сети далеко не всегда допустима.

Разумеется, приведенные выше рассуждения носят весьма общий характер и способны помочь в деле защиты информации только в принципе, а не в конкретных жизненных ситуациях (инженеры умеют отличать технические решения, пригодные для воплощения «в принципе», от решений, годящихся для реализации «в кожухе»). Поэтому необходима детализация и конкретизация приведенных общих рассуждений.

Попытки реализации угроз информации в вычислительных системах и средах, а также в сетях сбора и передачи данных (сокращенно информационно-вычислительных системах — ИВС) в дальнейшем именуются общепринятым в настоящее время термином «информационные атаки» или просто называются атаками. Защи-

та циркулирующей и обрабатываемой информации требует осознания и выявления потенциальных угроз на всех этапах жизненного цикла ИВС от замысла, проектирования, создания, эксплуатации и модернизации до утилизации. Каждый их трех перечисленных больших классов информационных атак можно разделить на несколько групп (см. гл. 3).

Причинами случайных воздействий на элементы ИВС могут быть отказы и сбои аппаратуры, помехи в каналах связи, непреднамеренные ошибки обслуживающего персонала, схемные и системотехнические ошибки разработчиков ИВС, структурные, алгоритмические и программные ошибки, аварийные ситуации.

Частота отказов и сбоев аппаратуры обычно увеличивается при усложнении аппаратно-программных комплексов. Человеческий фактор может обусловливать ошибки логические (неправильно принятые решения), сенсорные (неправильное восприятие оператором информации), а также оперативные или моторные (неправильная реализация решения). Интенсивность ошибок человека может колебаться в широких пределах. Хотя человек, как элемент автоматизированной системы, обладает по сравнению с техническими средствами рядом преимуществ (обучаемость, избирательность, способность к работе в конфликтных ситуациях), он имеет и ряд недостатков (зависимость психологических параметров от возраста и психофизиологического состояния, чувствительностью к изменениям окружающей среды и т. п.).

Аварийные ситуации могут возникнуть на объектах размещения элементов и подсистем ИВС вследствие стихийных бедствий и других происшествий, имеющих катастрофические последствия. Вероятность этих событий связана прежде всего с правильным выбором места размещения подсистем ИВК, организацией внутриобъектовых служб и взаимодействия ИВК с внешней средой.

По сравнению со случайными угрозами шире и опаснее круг искусственных или, как их еще называют, преднамеренных угроз информационным системам и обрабатываемой информации. Угрозы преднамеренных атак обусловливаются человеческой деятельностью. Преднамеренные угрозы могут быть направлены практически против всех без исключения элементов ИВС. Они целенаправлены, опасны и регулярны. Большинство наиболее часто реализуемых преднамеренных угроз предусматривает несанкционированный доступ (НСД) посторонних лиц, не принадлежащих к числу законных, легальных пользователей, и ознакомление их с конфиденциальной информацией. Именно НСД открывает возможности для копирования программ и данных, кражи физических носителей, содержащих конфиденциальную информацию; умышленного уничтожения информации; модификации документов и баз данных; фальсификации сообщений, передаваемых по каналам связи; отказа от авторства сообщения; отказа от факта

получения информации; навязывания ложного или ранее переданного сообщения.

Учитывая то, что современные ИВС, построенные на базе сетей ЭВМ, интегрированы в глобальные информационно-вычислительные сети, наиболее опасным является реализация угроз информации, источник которых является внешним по отношению к элементам и подсистемам ИВС. Попытки реализации подобного рода угроз называются удаленными атаками. Можно выделить два вида удаленных атак. Во-первых, удаленные атаки на инфраструктуру и протоколы сетей передачи и обработки данных, входящих в ИВС. Во-вторых, удаленные атаки на телекоммуникационные службы. По характеру воздействия удаленные атаки могут быть пассивными, активными и условно пассивными.

Пассивные атаки не оказывают непосредственного влияния на работу ИВС, но могут нарушать систему обеспечения ее безопасности. Активные атаки имеют целью нанесение прямого ущерба ИВС и предусматривают нарушение конфиденциальности, целостности и доступности информации, а также специальное психологическое воздействие на пользователей ИВС. Очевидной особенностью активного воздействия, по сравнению с пассивным, является принципиальная возможность его обнаружения. Могут случиться и условно-пассивные атаки. Они имеют целью подготовку к активной информационной атаке и включают в себя ведение компьютерной разведки, преодоление (взлом) системы защиты информации.

Удаленные атаки могут преследовать цели нарушения конфиденциальности информации, целостности информации, работоспособности ИВС, блокирования информации. Этот классификационный признак удаленных атак, очевидно, является прямой проекцией трех перечисленных выше основных типов угроз информации. Но все цели достигаются при условии получения несанкционированного доступа к информации.

По характеру начала осуществления воздействия атака может производиться по запросу от атакуемого объекта, наступлению ожидаемого события на атакуемом объекте или быть безусловной. В первом случае атакующий ожидает передачи от потенциальной цели атаки запроса определенного типа. Этот запрос служит условием начала воздействия. Во втором случае атакующий сервер постоянно наблюдает за состоянием операционной системы удаленной цели атаки и при наступлении определенного события в этой системе начинает воздействие. Как и в первом случае, инициатором осуществления начала атаки выступает сам атакуемый объект. Безусловная атака осуществляется немедленно и безотносительно к состоянию системы и атакуемого объекта. В этом случае атакующий сервер является инициатором начала осуществления атаки.

По условию ситуации осуществления воздействия атаки делятся на информационное нападение и ответные (ответно-встречные) воздействия.

Информационным нападением называется внезапное применение информационного оружия для осуществления воздействий на ИВС противостоящей стороны. Информационное нападение эффективно тогда, когда обеспечены его широкомасштабность, долговременность и скрытность. Ответные воздействия осуществляются после установления факта информационного нападения на объекты ИВС и идентификации противника. В случае правильно спланированного противником информационного нападения эффективность ответных воздействий существенно снижается.

Удаленная атака может осуществляться при наличии обратной связи с атакуемым объектом. При этом на некоторые запросы, переданные на атакуемый объект, атакующему требуется получить ответную реакцию. Атакующая сторона при этом приобретает возможность управления удаленной атакой (в идеальном случае — в масштабе реального времени). В то же время прерывание обратной связи может привести к потере управления атакой, снижению ее эффективности и, скорее всего, прекращению атаки.

При атаках без обратной связи на атакуемый объект обычно передаются одиночные запросы, ответы на которые атакующему серверу не нужны.

По расположению субъекта атаки относительно атакуемого объекта можно выделить внутрисегментные и межсегментные удаленные информационные атаки. При осуществлении внутрисегментной атаки субъект и объект атаки находятся в одном сегменте сети ЭВМ. На практике межсегментную атаку осуществлять значительно труднее, чем внутрисегментную, но при этом межсегментная удаленная атака представляет собой большую опасность.

По продолжительности воздействий могут быть разовые и долговременные атаки. Разовые атаки заключаются в ограниченных во времени целенаправленных воздействиях на объекты ИВС. При осуществлении долговременных атак предусматривается проведение продолжительных по срокам многоразовых атак на объекты ИВС, как правило, с использованием различных видов информационного оружия.

По масштабу воздействий рассматриваются локальные и глобальные (широкомасштабные) атаки. Глобальным атакам подвергаются несколько сегментов сети одновременно или последовательно.

Удаленные атаки предусматривают межсетевое взаимодействие информационных систем. Такие взаимодействия описываются иерархической семиуровневой моделью (рис. 7.1) [6]. Каждый конкретный уровень может взаимодействовать только с соседним, а правила взаимодействия полностью описываются соответствую-

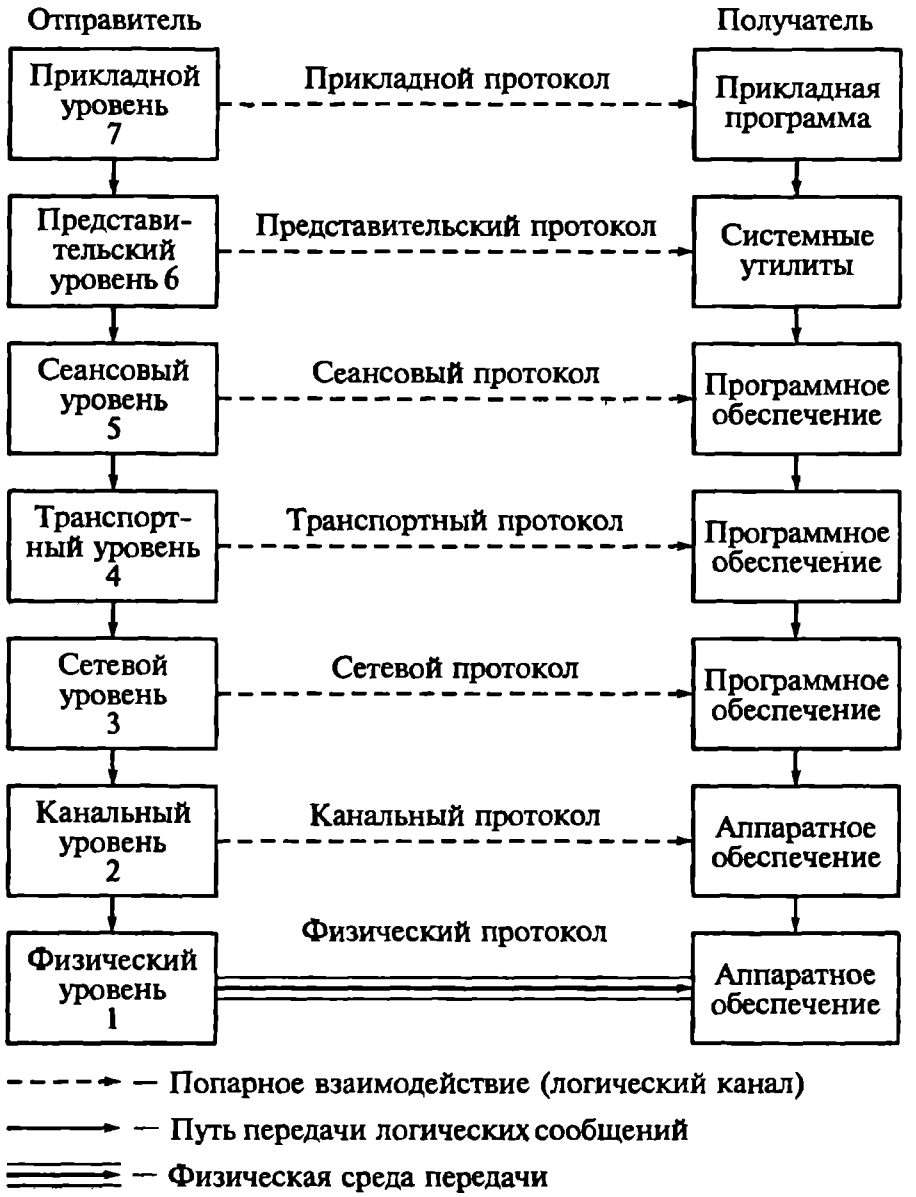


Рис. 7.1. Семиуровневая иерархическая модель сетевого взаимодействия ИВС

щим протоколом. Как видно из рис. 7.1, модель предусматривает декомпозицию процесса межсетевого взаимодействия открытых информационных систем на семь уровней: физический, канальный, сетевой, транспортный, сеансовый, представительный, прикладной.

На основе иерархической модели можно описать и исследовать любой сетевой протокол обмена, как и любую сетевую программу. Удаленная атака реализуется также при использовании специальной сетевой программы. Поэтому модель вполне подходит для оценки возможностей и опасностей информационных атак на ИВС.

Удаленные атаки обычно реализует специалист или группа специалистов очень высокой квалификации (не ниже администрато-

ра системы, а зачастую, и ее разработчика). Одним из таких источников преднамеренных угроз информации (в том числе и удаленных атак) в компьютерных системах являются действия хакеров. Из многих определений и описаний явления хакерства в настоящее время употребительно единственное: хакер — это человек (а иногда группа людей), стремящийся обойти защиты компьютерной системы. Цели хакера могут быть разными, но чаще всего компьютерный взломщик стремится получить дополнительные привилегии и права доступа к ИВС.

Программное обеспечение, наиболее часто подвергаемое атакам со стороны хакеров, — это системы управления базами данных (СУБД), компоненты операционной системы (ОС), сетевое программное обеспечение. Из этой тройки реже всего подвергаются атакам СУБД. Данный факт можно объяснить тем, что СУБД имеют строгую внутреннюю структуру и четко определенный набор операций над данными, хранящимися в базе (базах) данных. Такую структуру легче защищать: доступ к ее элементам легко регламентируется и контролируется. Поэтому взломщики предпочитают получать доступ к файлам базы данных не на уровне СУБД, а средствами ОС. В отличие от СУБД защитить ОС гораздо сложнее, поэтому адекватная организация безопасности является значительно более трудной задачей, чем в случае с СУБД. Реализации того или иного алгоритма хакерской атаки на практике в значительной степени зависят от архитектуры и конфигурации конкретной ОС. Однако есть несколько видов атак, которым может быть подвергнута практически любая ОС. Это кража и (или) подбор пароля, сканирование носителей информации, превышение полномочий.

Кроме выше перечисленных угроз информации, существует достаточно большая группа атак, не связанных с физическим доступом к элементам ИВС. Каждое устройство хранения, передачи, обработки информации является источником излучения различной природы (электромагнитные, акустические волны). Перехватывая и обрабатывая излучения ИВС иногда возможно получать достаточно разнообразные и ценные сведения о процессах, сопровождающих обработку данных в ИВС. Источником подобного рода излучений могут быть различные электронные устройства (системные блоки и мониторы компьютеров, принтеры, сканеры и т. п.). Особенно опасны с точки зрения защищенности информации электромагнитные излучения мониторов и линий передачи данных. Причем для перехвата информации за счет регистрации и обработки вторичных электромагнитных излучений и акустических волн зачастую достаточно использовать простое и, соответственно, недорогое и доступное оборудование. Расстояние, с которого потенциально возможен перехват вторичных излучений, зависит от характера источника и вида создаваемого им из-

лучения. Величина этого опасного расстояния колеблется в пределах до нескольких десятков и сотен метров. Опасность такого вида угроз заключается в том, что злоумышленник получает оперативный доступ к обрабатываемой информации дистанционно, без непосредственного контакта с системой защиты. Поэтому информационные атаки не регистрируется системой защиты или регистрируется слишком поздно. Мероприятия и средства по нейтрализации НСД по каналам перехвата паразитных и непреднамеренных излучений должны разрабатываться и строго контролироваться на всех этапах проектирования и эксплуатации ИВС.

Таким образом, потенциальные угрозы информации в современных ИВС отличаются многообразием, сложностью своей структуры и функций. Их действие направлено практически против всех структурных компонентов современных систем управления, а их источники могут располагаться как в самой ИВС, так быть и вне ее, в том числе удаленными на значительное расстояние. В связи с тем, что объем материальных средств, выделяемых на защиту информации, обычно ограничен, возникает задача рационального их распределения. При этом материальные средства целесообразно расходовать в первую очередь на нейтрализацию угроз, реализация которых может нанести ИВС наибольший вред. Общую задачу оценки угроз можно представить совокупностью следующих составных элементов:

- обоснование структуры и содержания системы показателей, необходимых для исследований и практического решения всех задач, связанных с защитой информации;
- обоснование структуры и содержания тех параметров, которые оказывают существенное влияние на значение показателей уязвимости информации;
- разработка комплексов моделей, отображающих функциональные зависимости показателей от параметров и позволяющих определять значения всех необходимых показателей уязвимости информации во всех представляющих интерес состояниях и условиях жизнедеятельности ИВС;
- разработка моделей для оценки показателей уязвимости при исследованиях и практическом решении различных вопросов защиты. При этом уязвимость информации должна быть оцениваема на всех стадиях и этапах разработки и функционирования ИВС.

Угрозы информационным ресурсам в вычислительных системах и средах реализуются с использованием специальных деструктивных программ. Такие программы, работа которых вызывает нежелательные и даже опасные последствия, обладают целым рядом свойств. И эти свойства можно рассматривать как классификационные признаки деструктивных программ. В частности, такие программы, осуществляющие реализацию угроз информационному ресурсу, должны обладать скрытностью работы (и даже

присутствия в программно-аппаратной среде), возможностью разрушать (искажать) коды программ и данных в памяти ЭВМ, нейтрализовывать работу систем защиты информационных ресурсов.

Деструктивные программы в зависимости от их свойств, целей и способов воздействия на информационные ресурсы подразделяются на компьютерные вирусы, средства несанкционированного доступа к информации и программные закладки.

7.2. Компьютерные вирусы

7.2.1. Общие сведения о компьютерных вирусах

Компьютерные вирусы — один из наиболее распространенных видов деструктивных программ и едва ли не самый серьезный источник угроз информации в современных информационных системах. К настоящему времени не существует четкого определения для вирусных программ. Биологические вирусы, которые представляют собой мельчайшие неклеточные частицы, состоящие из нуклеиновой кислоты и белковой оболочки, не являются живыми существами — они не могут существовать сами по себе, а только в каких-то иных живых клетках. Точно также и компьютерные вирусы могут существовать только внедряясь в программы. Чаще всего компьютерным вирусом называется подпрограмма, которая может заражать другие программы, включая в них свою копию (возможно, модифицированную), и сохранять способность к дальнейшему размножению. При этом она производит несанкционированные и вредные действия.

Можно назвать, по крайней мере, несколько причин широкого распространения компьютерных вирусов:

массовое распространение стандартизованной вычислительной техники, у которой процессоры совместимы на уровне микрокоманд друг с другом, а также с аналогичными подсистемами предыдущих и будущих поколений;

стандартизация программного обеспечения, массовое использование операционных систем, дружественных не только для пользователя, но и для вирусов;

отсутствие информационной культуры у пользователей вычислительной техники и довольно широкая популяризация в печати сведений о создании саморазмножающихся вредоносных программ;

несовершенство законодательства, не предусматривающего ответственности за правонарушения с использованием вычислительной техники.

К настоящему времени компьютерная вирусология накопила определенную историю, которую вкратце можно проиллюстрировать следующими основными эпизодами [11, 12, 29].

Первая известная публикация относится к 1951 г. В ней Джоном фон Нейманом сформулирована и исследована проблема создания саморазмножающихся компьютерных программ.

В 1962 г. Высоцкий и Макилрой (США) создали компьютерную игру, основанную на конфликтном взаимодействии самовоспроизводящихся программных механизмов в памяти компьютера. Победителем считался тот игрок, чьи программы захватывали всю память. Эта игра получила широкое распространение во многих учебных и исследовательских центрах США.

В 1977 г. появляется первый персональный компьютер Apple II. Число компьютеров этой серии за шесть лет перевалило за три миллиона. В это же время произошло бурное развитие сетей передачи информации на базе обычных телефонных каналов (BBS). Одновременно с этим получили широкое распространение программы-вандалы, которые наряду с выполнением некоторой полезной функции разрушали данные в информационной базе персонального компьютера.

В 1980 г. выходит первая европейская публикация о компьютерных вирусах «Самовоспроизводящиеся программы» И. Краузе, содержащая листинги вирусов на языке Ассемблера.

В 1981—1982 гг. появился первый бутовый вирус (Elk Cloner), получивший широкое распространение на компьютерах Apple II. Проявлял он себя очень разнообразно: переворачивал изображение и заставлял мигать экран, выводил на экран разнообразные сообщения.

В 1985—1986 гг. произошли вспышки заражения компьютерными вирусами многих персональных компьютеров. Причиной было бесконтрольное копирование компьютерных файлов.

Первым широко распространенным вирусом для IBM PC стал Пакистанский компьютерный вирус (загрузочный вирус Brain), разработанный братьями Амджатом и Беситом Алби в 1986 г. Вирус заражал компьютеры в отместку за несанкционированное копирование программного обеспечения. Только в США этот вирус заразил около 18 000 компьютеров. Он стал первым вирусом, использовавшим технологию маскировки (СТЕЛС). При попытке чтения зараженного загрузочного сектора вирус подставлял его незараженную копию. В этом же году Р. Бюргер обосновал и показал практически возможность создания файлового вируса.

В 1987 г. появляется вирус Vienna, дизассемблированная копия которого с комментариями была опубликована Р. Бюргером в книге, специально посвященной компьютерным вирусам.

В ноябре 1988 г. отмечена эпидемия сетевого вируса Морриса. Этот вирус вследствие ошибки в программном коде рассылал свои копии по другим компьютерам сети и инфицировал таким образом более 6000 компьютерных систем в США, включая компьютеры NASA Research Institute. Общие убытки от вируса Морриса

были оценены в 96 млн долларов. Для своего размножения он, подбирая пароли из стандартного списка, использовал ошибки в операционной системе Unix.

В 1990 г. появляется первый полиморфный вирус Chameleon, обнаружение которого невозможно по участку постоянного кода (маске или сигнатуре). Это свойство вируса заставило разработчиков антивирусных программ искать другие методы его детектирования.

В начале 1992 г. появляется первый генератор полиморфного кода MtE, на базе которого создается сразу несколько полиморфных вирусов. В конце этого же года отмечается появление первого вируса, заражающего выполняемые модули MS Windows.

1994 г. отмечен появлением первого вируса, заражающего объектные модули.

В августе 1995 г. появляется первый вирус, заражающий документы MS Word (макровирус Concept). Эта вредоносная программа положила начало целой серии макровирусов, поражающих файлы документов (не только текстового редактора Word), имеющие в своем составе макрорасширения на языке высокого уровня (Visual Basic, Word Basic). Этот факт опроверг бытующее мнение, будто заражение вирусом невозможно при открытии файла. В этом же году очень широкое распространение по всему миру получил полиморфный шифрованный вирус DieHard2 (SW4000).

В январе 1996 г. появился первый вирус для Windows 95 (Win95.Boza). В июне того же года выходит в свет первый полноценный вирус (OS2AEP) для OS/2, поражающий exe модули.

В 1997 г. вирусы заняли еще одну нишу: был обнаружен вирус для Linux (Linux.Bliss). В июне этого же года отмечается появление первого самошифрующегося вируса для Windows 95 российского происхождения.

В октябре 1998 г. отмечается появление первого вируса, заражающего файлы HTML.

В марте 1999 г. компьютерный мир потрясла настоящая эпидемия вируса Melissa, который распространялся гораздо быстрее и шире, чем любой другой из появившихся когда-либо ранее. С помощью почтовой программы MS Outlook Express, из адресной книги которого вирус получал адреса электронной почты корреспондентов, Melissa рассылал свои копии по сети. Скорость распространения вируса была действительно впечатляющей: одна из американских организаций сообщила, что вирус сгенерировал около полумиллиона сообщений электронной почты в течение всего трех часов. Относясь к категории сетевых червей, Melissa действительно может рассматриваться как опасность принципиально нового типа.

4 мая 2000 г. началась новая эпидемия, вызванная сетевым червем «I LOVE YOU/Loveletter/LoveBug». Этот вирус, в отличие

от своего предшественника Melissa, рассылающего копии первым 50 абонентам из адресной книги MS Outlook Express, не щадил никого, одинаково уничтожая файлы с расширениями: .vbs, .vbe, js, .jse, .css, .wsh, .set, .hta, .jpg, jpeg. Сообщалось, что в одной из компаний вирусом было уничтожено более 60 Гбайт графики в формате jpeg.

Дополнительно червь пытался скачать из Internet троянского коня, похищающего все пароли Windows. Согласно данным исследовательской фирмы Computer Economics, через день после начала распространения этим вирусом было поражено 45 млн компьютеров. Ущерб, причиненный им, оценивается в несколько миллиардов долларов.

Таким образом, в настоящее время не существует ни одного программного компонента информационных систем, который не был бы подвержен опасности вредоносного воздействия вирусов. Эксперты считают, что сегодня число существующих вирусов превышает 40 тыс., причем ежедневно появляется, по разным оценкам, от 6 до 15 новых. Однако положение вовсе не так трагично, как может показаться. Дело в том, что старые вирусы выходят из обращения. Кроме того, большинство вирусов никогда не покидают резерваций — тщательно охраняемых коллекций в исследовательских лабораториях, а многие настолько плохо сработаны, что просто не могут распространяться дальше машины своего создателя, да еще подпольных Web-страниц и BBS, посвященных проблемам разработки вирусов. Поэтому «диких» (реально циркулирующих) вирусов в настоящее время насчитывается около 400, что, впрочем, тоже немало.

Традиционно вирусные программы подразделяют на семь видов [28]:

1. Программы-вандалы, которые маскируются (по имени) под видом широко используемых программ и выполняют несанкционированные действия (например, стирание информации с диска) при запуске. Способностью к самостоятельному размножению они не обладают. Распространяются при копировании программ пользователями, наиболее часто при этом рассылаются по BBS станциям и электронным конференциям. По сравнению с иными вирусами вандалы не получили широкого распространения по достаточно простым причинам: они либо уничтожают себя вместе с остальными данными на диске, либо демаскируют свое присутствие и уничтожаются пострадавшим пользователем.

2. Бутовые (загрузочные) вирусы. Заражают диски (дискеты, логические диски, жесткие диски).

3. Файловые вирусы. Они заражают выполняемые модули с расширениями .exe, .com .bin, .ovl, .dll.

4. Файлово-бутовые. Редкая разновидность вирусов, заражающая как диски, так и файлы.

5. Макровирусы (макрокомандные вирусы). Заражают файлы документов, содержащие в своем составе программный код макрорасширений (документы Microsoft Word, Exel).

7. Сетевые вирусы (сетевые черви). Редкие и очень сложные по структуре и принципам работы программы, использующие для своего распространения по сети ЭВМ ошибки (дыры) в коде или особенности функционирования сетевого программного обеспечения операционной системы.

Практически все виды вирусов бывают резидентными и нерезидентными. Резидентные вирусы способны оставлять свои копии в оперативной памяти, перехватывать некоторые события (например, обращения к файлам или дискам) и инициировать при этом процедуры заражения обнаруженных объектов (файлов и секторов). Поэтому резидентные вирусы опасны не только во время работы зараженной программы, но и после ее окончания. Резидентные копии таких вирусов остаются жизнеспособными вплоть до очередной перезагрузки, даже если на диске уничтожены все зараженные файлы. Часто от таких вирусов невозможно избавиться восстановлением всех копий файлов с дистрибутивных дисков или backup-копий. Резидентная копия вируса остается активной и заражает вновь создаваемые файлы. Это характерно и для загрузочных вирусов. Форматирование диска при наличии в памяти резидентного вируса не всегда вылечивает диск, поскольку многие резидентные вирусы заражают его повторно после форматирования.

Нерезидентные вирусы, напротив, активны на довольно непродолжительных интервалах времени: только в момент запуска зараженной программы.

По отношению к информации вирусы опасны по-разному.

Безвредные вирусы не производят никаких действий, влияющих на работоспособность системы, кроме размножения собственных копий. Влияние неопасных вирусов на систему ограничивается подачей звуковых сигналов, выдачи визуальных сообщений, не имеющих опасных последствий для системы.

В результате действий опасных вирусов нарушается нормальное функционирование как отдельных программных комплексов, так и всей операционной системы в целом. В алгоритме очень опасных вирусов заложены процедуры и механизмы, результатом выполнения которых могут быть уничтожение информации, разрушение отдельных областей операционной системы. Такие процедуры и механизмы имеют фатальные последствия для программной и информационной среды пользователя.

После появления специализированных антивирусных программных комплексов возникли вирусы, использующие для скрытия своего присутствия в системе специальные приемы (СТЕЛС вирусы). Эти приемы обычно сводятся к перехвату и контролю некоторых системных ресурсов. Процесс лечения системы, зараженной

СТЕЛС вирусами, усложняется еще и тем, что появились вирусы, меняющие свой код, а иногда и способы заражения, в процессе функционирования.

Даже сегодня бороться с такими вирусами достаточно сложно. В антивирусной базе Norton AntiVirus имеется информация о 1016 полиморфных и 942 СТЕЛС вирусах.

7.2.2. Принципы функционирования основных разновидностей вирусов

Упрощенно процесс заражения вирусом программных файлов (exe, jcom, ovl, dll модулей) можно представить следующим образом. Код зараженной программы обычно изменяется таким образом, чтобы вирус получил управление первым, до начала работы программы-носителя. При передаче управления вирусу он находит новую программу и выполняет вставку своей копии в нее. Самый распространенный и часто встречающийся способ заражения файловым вирусом — вставка в конец программы. В этом случае вирус корректирует код заражаемой программы так, чтобы получить управление первым. Для этого обычно первые несколько байтов запоминаются в теле вируса, а на их место вставляется код перехода на начало вируса. В случае com модуля это обычно команда безусловного перехода jmp. В exe модуле вирус изменяет заголовок программы (часто это первые 24 байт). При запуске программы вирус первым получает управление, отрабатывает свой код, после чего восстанавливает скрытые первые байты и передает управление программе-вирусоносителю.

Код вируса может быть вставлен в начало зараженной программы. При вставке в начало файла вирус переписывает первые блоки (или все тело) зараженной программы в ее конец. Поэтому до передачи управления зараженному модулю, вирус должен предварительно переписать перемещенные блоки программы на первоначальное место, переместив часть своего кода таким образом, чтобы она не была затерта.

Вставка вируса в середину файла встречается редко. В этом случае вирус переписывает замещаемые блоки программы в конец модуля. Реже всего такой способ используется для заражения (обычно специализированными вирусами, в том числе и макровирусами) модулей, особенности структуры которых заранее известны (файлы command.com, документы MS Word и т.п.).

При заражении вирусом, вставленным в конец или середину программы, как правило, не производится перенос замещаемых блоков. В этом случае длина зараженной программы не изменяется, а исходный выполняемый модуль безвозвратно разрушается, что маскируется вирусом при его запуске выдачей одного из сообщений операционной системы.

Нерезидентный файловый вирус ищет программы обычно с помощью переменных среды DOS.

Резидентный вирус при инициализации перехватывает и контролирует некоторые прерывания DOS в соответствии с механизмом поиска программ (при запуске, открытии, чтении, записи) обычно перехватываются 21h и 13h прерывания DOS. Заражение *.com и *.exe файлов производится по схемам, различающимся лишь числом байт, изменяемых в начале выполняемого модуля, и связанным с различной структурой файлов данного типа. Распознавание типа модулей вирусом обычно производится по расширению файла и по идентификатору (в *.exe модуле первые 2 байта обычно mz, реже zm).

Для предупреждения повторного заражения файлов вирусы используют специальные приемы. Например, в поле даты и времени создания файла проставляются несуществующие данные. Резидентный вирус вводит дополнительную функцию перехваченного им прерывания, которая возвращает значение, означающее присутствие копии вируса в памяти. Другим наиболее часто используемым способом регистрации наличия резидентного вируса в памяти является запись в редко используемую область памяти некоторой комбинации битов.

Процесс лечения от файловых вирусов обычно заключается в поиске в теле вируса сохраненных байтов, восстановление их в программе носителе и «вырезании» вируса из кода программы.

Разумеется, файловые вирусы используют и другие механизмы взаимодействия с программными модулями. Например, так называемые вирусы-компаньоны (они заражают *.exe модули) не изменяют заражаемого модуля, а используют свойство операционной системы, заключающееся в том, что из нескольких выполняемых модулей с одинаковыми именами DOS в первую очередь выполняет при запуске по имени модуль с расширением com, т. е. вирус записывает в каталог, содержащий заражаемый *.exe файл свой код в модуль с тем же именем, но с расширением com.

Другим примером служат так называемые link-вирусы, представителем которых является печально известный DIR II. Вирус хранит на диске только одну свою копию, а при заражении выполняемых модулей лишь прописывает в соответствующем разделе fat таблицы вместо номера начального кластера файла физический адрес начала вируса на диске.

Бутовый вирус. Он заражает жесткие и гибкие диски (как загрузочные, так и иные). В отличие от файлового вируса он состоит из двух раздельных секций: головы и хвоста. Положение головы вируса всегда одинаково — она расположена в Boot секторе. На жестком диске начальные байты вируса могут располагаться в одном из его двух Boot секторов: главном (MBR по адресу 0/0/1) или Boot секторе логического диска C. Последние байты — в разли-

чных местах, а именно: в кластерах, помеченных в FAT как сбойные, последних физических секторах дискеты или диска, в используемых или неиспользуемых блоках FAT главного каталога или одного из подкаталогов, на дополнительных дорожках дискеты или винчестера. В любом случае хвост вируса должен содержать копию оригинального Boot сектора. Если она не закодирована, то положение хвоста в большинстве случаев может быть определено глобальным контекстным поиском.

Механизм размножения вируса однотипен. При загрузке с зараженной дискеты бутовый вирус, заменяющий в Boot секторе загрузчик DOS, получает управление и сначала копирует себя в старшие адреса памяти. Затем он уменьшает размер памяти, заменяя значения вектора прерываний с Ah для защиты резидентной части вируса и Bh для перехвата обращений к диску. Таким образом, при обращении к диску управление всегда передается вирусу. Вирус запускает стандартный системный загрузчик, и только после этого происходит стандартная загрузка DOS.

Получив управление по прерыванию 13h, вирус анализирует, относится это к дискете или винчестеру. Если это обращение относится к дискете, вирус проверяет, заражена она уже или нет. Для этого считывается Boot сектор и проверяется его содержимое. Если дискета заражена, то вирус приступает к обработке непосредственно прерывания. Если дискета не заражена, то вирус сначала заражает ее.

Бóльшая часть загрузочных вирусов не проверяет системную память на наличие своей уже установленной TSR копии: они либо используют СТЕЛС приемы, при которых повторный запуск кода вируса невозможен, либо ориентируются на то, что код вируса загружается однократно в момент загрузки DOS. После первоначальной загрузки коды загрузочных секторов дисков больше не выполняются ни при каких условиях. Часть вирусов проверяет наличие своей копии. Для этого используются либо специальные вызовы INT 13h с каким-нибудь нестандартным значением, либо помечается заведомо неиспользуемый байт (или слово) в таблице векторов прерываний или в области данных BIOS (0040:00??). Существуют и другие способы обнаружения своей TSR копии.

Важно, что некоторые бутовые вирусы перехватывают прерывания от клавиатуры и поэтому могут сохраниться в памяти при быстрой перезагрузке по команде [Ctrl] + [Alt] + [Del].

Windows-вирусы. Для того чтобы оставить выполняемый код в памяти, они используют три способа [11], которые уже применялись различными вирусами. Самый простой способ — зарегистрировать программу как одно из приложений, работающих в данный момент. Для этого программа регистрирует свою задачу, окно которой может быть свернутым, свой обработчик системных событий и т.д. Второй способ — выделить блок системной памяти

при помощи DPMI вызовов и скопировать в нем свой код (вирус Ph33r). Третий способ — остаться резидентно как V×D драйвер (Windows 3.xx и Windows 95) или как драйвер Windows NT.

Перехват обращений к файлам производится одним из двух способов: либо перехватываются вызовы INT 21h (Hook_V86_Int_Chain, Get/Set_V86_Int_Vector, Get/Set_PM_Int_Vector), либо перехватывается системный вызов API. Затем резидентные Windows вирусы действуют примерно так же, как и DOS вирусы: перехватывают обращения к файлам и заражают их. Для обнаружения уже имеющейся в памяти резидентной копии используются примерно те же способы, которые описаны выше, за исключением V×D вирусов. Известные V×D вирусы загружаются в память при загрузке Windows. Для этого они записывают команду запуска в файл конфигурации Windows system.ini. Если в этом файле уже есть команда запуска вирусного V×D файла, то вирус не производит повторной регистрации своего V×D файла.

Макровирусы. Большинство макровирусов можно считать резидентными, поскольку они присутствуют в области системных макросов в течение всего времени работы редактора. Они, так же как резидентные, загрузочные и файловые вирусы, перехватывают системные события и используют их для своего размножения. К подобным событиям относятся различные системные вызовы, возникающие при работе с документами Word и таблицами Excel (открытие, закрытие, создание, печать и т.д.), вызов пункта меню, нажатие какой-либо клавиши или достижение определенного момента времени. Для перехвата событий макровирусы переопределяют один или несколько системных макросов или функций.

При заражении некоторые макровирусы проверяют наличие своей копии в заражаемом объекте и повторно себя не копируют. Другие макровирусы не делают этого и переписывают свой код при каждом заражении. Если при этом в заражаемом файле или области системных макросов уже определен макрос, имя которого совпадает с макросом вируса, то такой макрос оказывается уничтоженным.

По данным Международной ассоциации компьютерной безопасности (w.icsa.net), доля представителей этого класса в общем числе вирусов, циркулирующих по вычислительным системам и сетям, составляет 2/3, а по данным лаборатории Касперского, эта величина составляет порядка 55 %. Причин тому несколько. Во-первых, это широкое распространение объектов их поражения, т.е. офисных приложений. Сегодня практически нет таких людей, которые бы не использовали в своей повседневной работе текстовый процессор, электронные таблицы, систему обработки базы данных или мастер презентаций. Во-вторых, очень низкий уровень встроенной антивирусной защиты перечислен-

ных приложений. В-третьих, простота создания макровирусов. Для того чтобы написать вирус например, для MS Word, достаточно изучить азы языка программирования VBA. Будучи самым простым и доступным среди всех остальных языков, он предоставляет создателям вирусов все необходимые возможности для того, чтобы уничтожить важную информацию и надолго вывести компьютер из строя. В-четвертых, наиболее популярные офисные приложения (в первую очередь из пакета MS Office), как правило, интегрированы с почтовыми программами (например MS Outlook). Это обстоятельство определяет доступ макровирусов к электронной почте — наиболее удобному и быстрому способу распространения. Поэтому макровирусы имеют неограниченные возможности для молниеносного поражения миллионов компьютеров по всему миру.

Полиморфные вирусы. Обнаружение этих вирусов невозможно (или крайне затруднительно) осуществить при помощи так называемых вирусных масок — участков постоянного кода, специфичных для конкретного вируса [11]. Достигается это двумя основными способами: шифрованием основного кода вируса с непостоянным ключом и случайным набором команд расшифровщика или изменением самого выполняемого кода вируса. Существуют также другие, достаточно экзотические примеры полиморфизма: DOS вирус Bomber, например, не зашифрован, однако последовательность команд, которая передает управление коду вируса, является полностью полиморфной. Полиморфизм различной степени сложности встречается в вирусах всех типов: от загрузочных и файловых DOS вирусов до Windows вирусов и даже макровирусов. Сложность кода служит классификационным признаком полиморфных вирусов.

В настоящее время целью вирусных программ наряду с нанесением максимального ущерба стало получение и контроль доступа к информации (воровство паролей, создание люков в системе защиты вычислительных сетей, получение привилегированного доступа к данным и т.п.). Эти угрозы особенно актуальны для систем, работающих в локальных и глобальных сетях. При этом собственно вирусы занимают менее половины от передаваемых по почтовым каналам вредных и деструктивных данных. Другая половина — это различные программы для несанкционированного доступа (например, троянские программы и Internet черви) [29].

7.2.3. Использование СТЕЛС технологии в вирусных программах

СТЕЛС вирусы обладают способностью скрывать свое присутствие в системе. Известны СТЕЛС вирусы всех типов за исключе-

нием Windows вирусов: загрузочные вирусы, файловые DOS вирусы и даже макровирусы.

Загрузочные СТЕЛС вирусы для скрытия своего кода используют два основных способа. Первый способ заключается в том, что вирус перехватывает команды чтения зараженного сектора (INT 13h) и подставляет вместо него незараженный оригинал. Этот способ делает вирус невидимым для любой программы, включая антивирусы, не способные лечить оперативную память компьютера. Возможен перехват команд чтения секторов на уровне более низком, чем INT 13h.

Второй способ направлен против антивирусов, поддерживающих команды прямого чтения секторов через порты контроллера диска. Такие вирусы при запуске любой программы (включая антивирус) восстанавливают зараженные сектора, а после окончания ее работы снова заражают диск. Поскольку для этого вирусу приходится перехватывать запуск и окончание работы программ, то он должен перехватывать также DOS прерывание INT 21h.

С некоторыми оговорками СТЕЛС вирусами можно назвать вирусы, которые вносят минимальные изменения в заражаемый сектор (например, при заражении MBR правят только активный адрес загрузочного сектора — изменению подлежат только 3 байт) либо маскируются под код стандартного загрузчика.

Большинство файловых СТЕЛС вирусов либо перехватывают DOS вызовы обращения к файлам (INT 21h), либо временно лечат файл при его открытии и заражают при закрытии. Существуют файловые вирусы, использующие для своих СТЕЛС функций перехват прерываний более низкого уровня — вызовы драйверов DOS, INT 25h и даже INT 13h.

Некоторые вирусы используют часть функции полноценного СТЕЛС вируса. Чаще всего они перехватывают функции DOS FindFirst и FindNext (INT 21h, AH 11h, 12h, 4Eh, 4Fh) и уменьшают размер зараженных файлов. Такой вирус невозможно идентифицировать по изменению размеров файлов, если, конечно, он резидентно находится в памяти. Программы, которые не обращаются к увязанным функциям DOS (например, утилиты Norton), а напрямую используют содержимое секторов, хранящих каталог, показывают правильную длину зараженных файлов.

Реализация СТЕЛС алгоритмов в макровирусах является, наверное, наиболее простой задачей — достаточно всего лишь запретить вызов меню File/Templates или Tools/NMacro. Достигается это либо удалением этих пунктов меню из списка, либо их заменой на макросы FileTemplates и ToolsMacro.

Частично СТЕЛС вирусами можно назвать небольшую группу макровирусов, которые хранят свой основной код не в самом макросе, а в других областях документа: в его переменных или в Autotext.

7.3. Программные средства борьбы с вирусами

К настоящему времени разработана довольно широкая и полная система программных средств борьбы с вирусами [11]. Это программы-фаги (сканеры), программы-ревизоры, программы-мониторы, программы-вакцины (иммунизаторы).

Самыми популярными и эффективными антивирусными программами считаются антивирусные фаги (иначе эти программы называются сканерами или полифагами) и ревизоры (CRC сканеры). Часто обе приведенные разновидности объединяются в одну универсальную антивирусную программу, что значительно повышает ее мощность. Реже используют различного типа мониторы (блокировщики) и вакцины (иммунизаторы). Следует, однако, иметь в виду, что, в принципе, нельзя создать универсальное и абсолютно надежное средство борьбы со всеми существующими и будущими вирусами.

Программы-фаги. Принцип работы антивирусных программ-фагов (сканеров) основан на проверке файлов, секторов и системной памяти и поиске в них известных и новых (неизвестных сканеру) вирусов. Для поиска известных вирусов используются маски или, как их еще называют, сигнатуры — некоторая постоянная последовательность кода, специфичная для конкретного вируса. Если вирус не содержит постоянной маски или длина этой маски недостаточно велика, то используются другие методы. Например, перебор всех возможных вариантов кода вирусов. Этот способ эффективно используется для детектирования полиморфных вирусов.

Во многих полифагах используются алгоритмы эвристического сканирования, т. е. анализ последовательности команд в проверяемом объекте, набор некоторой статистики и принятие мягкого решения («возможно, заражен» или «не заражен») для каждого проверяемого объекта.

К достоинствам сканеров относится их универсальность, к недостаткам — низкая скорость сканирования, а также необходимость постоянного обновления антивирусных баз.

Принцип работы типичного алгоритма сканирования сводится к следующему. После загрузки с дискеты, на которой операционная система гарантированно свободна от вируса, программа проверяет дерево каталогов диска, логическое имя которого указывается в виде параметра при запуске. При нахождении *.exe или *.com модуля проверяется его длина. Если длина модуля больше 4 Кбайт, в теле программы ищется сигнатура вируса по соответствующему смещению. Если вирус найден, восстанавливаются скрытые в теле вируса байты начала модуля, после чего длина файла уменьшается на длину вируса и вирус удаляется из зараженного модуля. После этого восстанавливаются исходные время и дата создания файла.

Программы-ревизоры. Они подсчитывают контрольные суммы для присутствующих на диске файлов и системных секторов. Эти суммы сохраняются в базе данных антивируса вместе с некоторой другой информацией: размерами файлов, датами их последней модификации и т. п. При последующем запуске ревизоры сверяют данные, содержащиеся в базе данных, с реально подсчитанными значениями. Если информация о файле, записанная в базе данных, не совпадает с реальными значениями, ревизоры сигнализируют о том, что файл был изменен или заражен вирусом.

Ревизоры, использующие антиСТЕЛС алгоритмы, являются довольно сильным оружием против вирусов: практически 100 % вирусов оказываются обнаруженными почти сразу после их появления в компьютере. Существенным недостатком таких средств борьбы с вирусами является то, что программы-ревизоры распознают наличие вируса в системе уже после его распространения. Кроме того, они не распознают вирусы в новых, только что полученных или записанных файлах, поскольку в их базах данных отсутствует информация об этих файлах. Периодически появляются вирусы, которые используют эту слабость ревизоров, заражая только вновь создаваемые файлы. Такие вирусы остаются невидимыми.

Программы-мониторы. Антивирусные мониторы — это резидентные программы, перехватывающие вирусоопасные ситуации и сообщающие об их возникновении. К вирусоопасным относятся вызовы на открытие для записи в выполняемых файлах, запись в загрузочные секторы дисков, попытки программ остаться резидентно. Иначе говоря, вызовы генерируются вирусами в моменты их размножения.

К достоинствам программ-мониторов относится их способность обнаруживать и блокировать вирус на самой ранней стадии его размножения, что бывает очень полезно в случаях, когда давно известный вирус постоянно «выползает неизвестно откуда». К недостаткам относятся существование путей обхода защиты монитора и большое количество ложных срабатываний. Существуют аппаратные реализации некоторых функций мониторов, в том числе встроенные в BIOS. Однако, как и в случае с программными мониторами, такую защиту легко обойти прямой записью в порты контроллера диска, а запуск DOS утилиты FDISK немедленно вызывает ложное срабатывание защиты.

Программы-вакцины. Антивирусные вакцины (иммунизаторы) подразделяются на два типа: сообщающие о заражении и блокирующие заражение каким-либо типом вируса. Первые обычно записываются в конец файлов (по принципу файлового вируса), и при запуске файла каждый раз проверяют его на предмет обнаружения изменений. Недостаток у таких вакцин один, но он летален: абсолютная неспособность вакцины сообщить о заражении

СТЕЛСвирусом. Поэтому такие иммунизаторы, как и мониторы, в настоящее время практически не используются.

Второй тип вакцин защищает систему от поражения вирусом какого-то определенного вида. Файлы на дисках модифицируются таким образом, что вирус принимает их за уже зараженными. Для защиты от резидентного вируса в память компьютера заносится программа, имитирующая копию вируса. При запуске зараженной программы, вирус распознает вакцину как свою резидентскую копию и не активизируется. Такой тип вакцинации не может быть универсальным, поскольку при его помощи нельзя иммунизировать файлы от всех известных вирусов. Однако несмотря на это подобные программные средства в качестве полумеры могут вполне надежно защитить компьютер от нового неизвестного вируса вплоть до того момента, когда он будет детектироваться антивирусными сканерами.

Антивирусные программные комплексы. В современных условиях лишь они могут обеспечить надежную защиту от вирусных программ, отличающихся большим разнообразием принципов построения и функционирования. Обычно современные антивирусные программные комплексы включают в свой состав монитор, сканер, ревизор и планировщик.

Планировщик используется для координации работы разных компонентов антивирусного пакета и планирования антивирусных мероприятий в вычислительной системе.

Вакцина вследствие своей естественной ограниченности использования низкой универсальности в настоящее время практически не применяется.

Для снижения уровня опасности заражения вычислительной системы вирусными программами нужно выполнять следующие довольно простые правила [11, 29].

- Перед использованием проверять файлы, являющиеся потенциальными носителями вирусов. Такие файлы поступают извне (приносятся на дискетах, поступают из глобальных сетей и т. п.). Для проверки нужно иметь одну или несколько антивирусных программ (программных комплексов), возможности которых хорошо известны. Не запускать непроверенные и неизвестные файлы, в том числе полученные из компьютерной сети. Перед запуском новых программ обязательно проверить их одним или несколькими антивирусами. В случае большого объема документов, поступающих по электронной почте, целесообразно иметь активный антивирус, отслеживающий наличие вирусов. Для этого подходят программы-ревизоры, проверяющие наличие несанкционированных изменений в системных областях и в файловой системе компьютера.
- Круг лиц, работающий на конкретном компьютере, должен быть максимально ограничен, а права и полномочия пользователей строго регламентированы.

• Не следует использовать нелицензионное программное обеспечение. Лучше покупать дистрибутивные копии программного обеспечения у официальных продавцов, чем бесплатно или почти бесплатно копировать их из других источников. Необходимо использовать только хорошо зарекомендовавшие себя источники программ и других файлов. Как следствие из этого правила вытекает необходимость хранения дистрибутивных копий программного обеспечения (в том числе копий операционной системы). Копии желательно хранить на защищенных от записи носителях.

• Для обеспечения целостности и сохранности информации следует периодически сохранять на внешнем носителе файлы, с которыми ведется работа, а также файлы, имеющие наибольшую ценность. Резервные копии позволяют восстановить информацию в случае ее потери (разрушения). Не вредно копировать и хранить все содержимое винчестера.

• Необходимо осознавать и помнить, что не существует такой защиты, которую невозможно было бы обойти. Поэтому не стоит всецело уповать на встроенные системы защиты от вирусов (например, встроенной в BIOS, MS Word и т. п.). Не следует использовать незнакомые и устаревшие антивирусы для обнаружения и лечения компьютеров от новых вирусов. Вероятность обнаружения активного вируса такими программами снижается по мере их устаревания и появления новых вирусов. А пропуск опасных вирусов может способствовать широкому распространению вируса. Так, например, известны случаи, когда полифаг AIDSTEST способствовал заражению практически всех выполняемых модулей нераспознаваемым им вирусом.

• Проблема макровирусов в последнее время перекрывает все остальные проблемы, связанные с прочими вирусами. Существует несколько приемов и встроенных в Word и Excel средств, предотвращающих запуск вируса. Наиболее действенной из них является защита от вирусов, встроенная в Word и Excel (начиная с версий 7.0). Эта защита при открытии файла, содержащего любой макрос, сообщает о его присутствии и предлагает запретить этот макрос. В результате макрос не только не выполняется, но он даже не виден средствами Word и Excel. Такая защита является достаточно надежной, однако абсолютно бесполезна, если пользователь работает с макросами (любыми): она не отличает макросы вируса от не вируса и выводит предупреждающее сообщение при открытии практически любого файла. По этой причине защита в большинстве случаев оказывается отключенной, что дает возможность вирусу проникнуть в систему. Включение защиты от вирусов в уже зараженной системе не во всех случаях помогает: некоторые вирусы, однажды получив управление, при каждом запуске отключают защиту от вирусов и таким образом полностью блокируют ее.

Существуют и другие методы противодействия вирусам.

7.4. Программные закладки

Существует другой, отличный от вирусов, вид деструктивных компьютерных программ, который не обладает способностью к размножению, самодублированию. Такие программы, получая несанкционированный доступ к данным в памяти ЭВМ, перехватывают эти данные. Для перехвата данные несанкционированно копируются и сохраняются в специально созданных разделах памяти или передаются по сети потребителям, не имеющим на то законного права. Такие программы могут, подобно вирусам, искажать или уничтожать данные, но в отличие от вирусов деструктивное действие таких программ селективно направлено на конкретные данные.

Такие программы назвали закладками — по аналогии с тайно помещенными миниатюрными электронными системами перехвата радио-, видео-, и аудиоинформации.

При рассмотрении взаимодействия закладок и программ защиты информации уместны аналогии с взаимодействием вируса и прикладной программы. Вирус может присоединиться к исполняемому файлу, соответствующим образом изменив его, может уничтожить некоторые файлы или встроиться в цепочку драйверов. Закладка отличается более направленным и тонким воздействием. Но и вирус, и закладка должны скрывать свое присутствие в операционной среде компьютерной системы. Особенностью закладок может быть и то, что они фактически становятся неотделимы от прикладных или системных программ, если внедрены в них на стадии разработки программного обеспечения.

Если компьютерная система содержит механизмы защиты от НСД, то несанкционированные действия могут быть вызваны отключением или видоизменением защитных механизмов нелегальным пользователем; входом в систему под именем и с полномочиями реального пользователя.

В первом случае злоумышленник стремится видоизменить защитные механизмы в системе (например, отключить программу запросов паролей пользователей), во втором — каким-либо образом выявить или подделать идентификатор реального пользователя (например, подсмотреть пароль, вводимый с клавиатуры).

И в том и другом случаях НСД можно представить моделью опосредованного доступа — когда проникновение в систему осуществляется на основе некоторого воздействия, произведенного предварительно внедренной в систему одной или несколькими программами.

Например, злоумышленник пользуется информацией, которая извлечена из некоторого массива данных, созданного работой программного средства злоумышленника совместно с системой проверки прав доступа и предоставления этих прав. Предварительно внедренная в систему программа при осуществлении доступа легального пользователя запомнит его пароль и сохранит в заранее

известном доступном злоумышленнику файле, а затем нелегальный пользователь воспользуется данным паролем для входа в систему. Либо злоумышленник изменит часть системы защиты так, чтобы она перестала выполнять свои функции (например, изменит программу шифрования вручную или при помощи другой программы, чтобы она перестала шифровать или изменила алгоритм шифрования на более простой).

Часто используют синонимы термина «закладка»: «логическая бомба», «логический люк», «троянский конь». Обычно в литературе понятие закладки в основном связано с разработкой программного обеспечения, а конкретно — с написанием исходных текстов программ, в которых создаются дополнительные функции. Следовательно, ранее закладка понималась как внутренний объект защищенной системы. Однако закладка может быть и внешним объектом по отношению к защищенной системе.

Программные закладки можно классифицировать по методу и месту их внедрения и применения:

закладки, ассоциированные с программно-аппаратной средой (BIOS);

закладки, ассоциированные с программами первичной загрузки (находящиеся в Master Boot Record или BOOT секторах активных разделов);

закладки, ассоциированные с загрузкой драйверов DOS, командного интерпретатора, сетевых драйверов, т.е. с загрузкой операционной среды;

закладки, ассоциированные с прикладным программным обеспечением общего назначения (встроенные в клавиатурные и экранные драйверы, программы тестирования ПЭВМ, утилиты и оболочки типа NORTON);

исполняемые модули, содержащие только код закладки (как правило, внедряемые в пакетные файлы типа *.bat);

модули-имитаторы, совпадающие с некоторыми программами, требующими ввода конфиденциальной информации;

закладки, маскируемые под программные средства оптимизационного назначения (архиваторы, ускорители и т.д.);

закладки, маскируемые под программные средства игрового и развлекательного назначения (как правило, используются для первичного внедрения закладок).

Как видно, программные закладки имеют много общего с вирусами, особенно в части ассоциирования себя с исполняемым кодом (загрузочные вирусы, вирусы-драйверы, файловые вирусы).

Кроме того, программные закладки, как и многие известные вирусы классического типа, имеют развитые средства борьбы с отладчиками и дисассемблерами.

Для того чтобы закладка смогла выполнить какие-либо функции по отношении к прикладной программе, она должна принять

управление на себя. Иначе говоря, процессор должен начать выполнять инструкции (команды), относящиеся к коду закладки:

закладка должна находиться в оперативной памяти до начала работы программы, которая является целью воздействия закладки, следовательно, она должна быть загружена раньше или одновременно с этой программой;

закладка должна активизироваться по некоторому событию, т.е. при выполнении ряда условий в программно-аппаратной среде управление должно быть передано на программу-закладку.

Разумеется, выполнение перечисленных требований достигается путем анализа и обработки закладкой общих относительно закладки и прикладной программы воздействий (как правило, прерываний). Прерывания должны сопровождать работу прикладной программы или работу всей ЭВМ. В качестве таких прерываний закладками используются прерывания от таймера ПЭВМ; прерывания от внешних устройств; прерывания от клавиатуры; прерывания при работе с диском; прерывания операционной среды (в том числе прерывания при работе с файлами и запуск исполняемых модулей).

Кроме того, возможен случай, когда при запуске программы (в этом случае активизирующим событием является запуск программы) закладка разрушает некоторую часть кода программы, уже загруженной в оперативную память, и, возможно, систему контроля целостности кода или контроля иных событий и на этом заканчивает свою работу.

Таким образом, можно выделить следующие закладки [29].

- Резидентного типа. Такие закладки находятся в памяти постоянно с некоторого момента времени до окончания сеанса работы персонального компьютера (выключения питания или перезагрузки). Закладка может быть загружена в память при начальной загрузке ПЭВМ, загрузке операционной среды или запуске некоторой программы (которая по традиции называется вирусоносителем), а также запущена отдельно.
- Нерезидентного типа. Такие закладки начинают работу по аналогичному событию, но заканчивают ее самостоятельно по истечении некоторого промежутка времени или некоторому событию, при этом целиком выгружая себя из памяти.

Исполнение кода закладки должно сопровождаться операциями несанкционированной записи (для сохранения некоторых фрагментов перехваченной информации) и несанкционированного считывания, которое может происходить отдельно от операций чтения прикладной программы или совместно с ними. При этом под операциями считывания и записи понимаются любые обращения к внешнему устройству (возможно и не связанные с получением информации). Например, считывание параметров устройства или его инициализация; закладка может использоваться для инициирования сбойных ситуаций или переназначения ввода-вывода.

Несанкционированная запись информации закладкой может происходить в массив данных, не совпадающий с массивом пользовательской информации (сохранение информации) или в массив данных, совпадающий с любой частью пользовательского массива (для искажения, уничтожения или навязывания ложной информации закладкой). Поэтому можно рассматривать три основные группы деструктивных функций, которые могут выполняться закладками:

сохранение фрагментов информации, возникающей при работе пользователя, прикладных программ, вводе-выводе данных, во внешней памяти (локальной или удаленной) сети или выделенной ЭВМ;

разрушение функций самоконтроля или изменение алгоритмов функционирования прикладных программ;

навязывание некоторого режима работы (например, при уничтожении информации блокирование записи на диск, при этом информация, естественно, не уничтожается) либо замена записываемой информации навязанной закладкой.

Сохранение фрагментов информации программой-закладкой может происходить при выводе информации на экран видеотерминала, выводе информации в файл или иное внешнее устройство, вводе информации с клавиатуры. Программа выделяет себе в оперативной памяти некоторую область, где помещается информация для обработки (как правило, доступная для непосредственного восприятия: область экрана, клавиатурный буфер). Закладка определяет адрес информативной области программы (иногда этот адрес фиксирован) и анализирует события, связанные с работой прикладной программы или операционной среды. При этом интерес представляют лишь события, результатом которых может стать появление интересующей информации в информативной области. Установив факт интересующего события, закладка переносит всю информативную область либо ее часть в свою область сохранения (непосредственно на диск или в выделенную область оперативной памяти).

При перехвате вывода на экран выделяется область видеобуфера с фиксированным адресом. Видеобуфер, с точки зрения программ, представляет собой область обычной оперативной памяти. Выводимый на экран текст одновременно помещается в видеобуфер, откуда может быть считан и сохранен закладкой. Синхронизирующим событием в этом случае может быть ввод с клавиатуры длинной последовательности символов (обрабатываемого текста), чтение из файла, запуск программ с определенными именами.

Кроме того, возможно периодическое сохранение области экранного буфера, сопряженное с таймерным прерыванием.

Надо отметить, что закладки этого типа имеют малоинформативный результат работы, поскольку на экран помещается небольшое количество информации, а символы пароля, нужные для реализации эффективной информационной атаки, как правило, на экран не выводятся.

Перехват ввода с клавиатуры достаточно опасен, поскольку клавиатура является основным устройством управления и ввода информации. Через клавиатурный ввод можно получить сведения о вводимых конфиденциальных сообщениях (текстах), паролях и тому подобных важных данных. Перехват может происходить двумя основными способами: встраивание в цепочку прерывания INT 9h или анализ содержания клавиатурного порта или буфера по таймерному прерыванию.

Работа закладки основывается на полном сохранением всех фактов нажатий клавиш. Все нажатия сохраняются в специальном скрытом файле. Файл затем изучается, и на основе анализа лицо, пытавшееся получить доступ к зашифрованным файлам, восстанавливает возможные парольные последовательности.

Перехват и обработка файловых операций происходит, когда программное средство защиты информации производит некоторые файловые операции. Для этого открывается файл, часть его или весь файл считывается в буфер оперативной памяти, обрабатывается и затем, возможно, записывается в файл с прежним или новым именем.

Активизирующим событием в данном случае является, как правило, открытие файла (INT 21h, функция 3Dh) либо его закрытие.

В операционной среде закладка, влияющая на файловые операции, порождает в системе новые связи, включая в них свои операции и массивы данных.

Если злоумышленнику известна интересующая его программа с точностью до команд реализации на конкретном процессоре, он может создать модель процесса ее загрузки и выяснить относительные адреса частей программы относительно сегмента оперативной памяти, в который она загружается. Это означает, что возможно произвольное изменение кода программы и, соответственно, отклонение (возможно негативного характера) в работе прикладной программы. Тогда алгоритм действия закладки может быть таким:

закладка загружается в память каким-либо способом;

закладка осуществляет перехват (редактирование цепочки) одного или нескольких прерываний (прерывания DOS «запуск программ и загрузка оверлеев», прерывания BIOS «считать сектор», прерывание таймера);

по одному из трех событий закладка получает управление на себя, и далее выполняются проверка принадлежности запущенной программы или уже работающей (для таймерного прерывания) к интересующим программам, определение сегмента, в ко-

торый загружена программа, запись относительно определенного сегмента загрузки некоторых значений в оперативной памяти так, чтобы отключить схемы контроля и (или) исправить программу нужным образом.

Принципиальная возможность исправления кода следует из того, что вывод о правильности работы программы делается на основе операций сравнения в арифметико-логическом устройстве микропроцессора.

Наконец, возможен случай, когда содержательный код программы защиты вместе со схемой контроля будет удален из памяти полностью, и все последующие операции будут выполнены без влияния программы защиты.

Основным способом активизации разрушающих закладок является запуск ассоциированных с ними программ. При этом закладка получает управление первой и выполняет некоторые действия (изменения адресов прерывания на собственные обработчики, исправление в коде программ защиты и т. п.). В данном случае борьба с воздействием закладок может быть произведена только путем контроля целостности исполняемых файлов непосредственно перед их исполнением. Тогда воздействие закладки можно оценивать количественной вероятностью ее активизации при заданном алгоритме контроля кода запускаемой программы (т. е. вероятностью обнаружения или необнаружения ассоциированного с кодом вируса или закладки).

Особенности применения программно-аппаратных средств защиты состоят в следующем:

собственные программы управления аппаратной частью, как правило, находятся в ПЗУ и, следовательно, не могут быть изменены программным путем;

под управление аппаратным комплексом выделяются средства низкого уровня (операции с портами либо выделенные прерывания);

управление программам аппаратной части передается до загрузки операционной среды, и часть операций, в основном связанных с инициализацией начальных состояний устройства, также происходит до загрузки операционной среды.

Все эти факторы накладывают определенные ограничения на процесс воздействия программных закладок на рассматриваемые программно-аппаратные комплексы. Однако эти ограничения не могут полностью исключить такое воздействие.

Поскольку обращение к аппаратным средствам происходит из прикладных программ и, как правило, через некоторую промежуточную программу управления, воздействие на аппаратное средство может быть сведено к воздействию либо на прикладную программу, либо на программу управления аппаратным комплексом, который находится в ОЗУ.

7.5. Действие вирусов и программных закладок в сетях ЭВМ

Рассматривая использующиеся в настоящее время сети ЭВМ, можно выделить их обобщенную структуру (рис. 7.2).

Абонентские пункты сети могут быть объединены в единую систему с серверами локальных сетей, которые, в свою очередь, соединяются между собой. Таким образом, через линии связи объединяются несколько локальных вычислительных сетей, которые могут быть как территориально близкими, так и разнесенными на значительные удаления.

Серверы локальной сети могут быть подключены к концентратору сообщений, объединяющему потоки информации с нескольких локальных сетей и (или) изолированных абонентских пунктов. Объединенный концентратором сообщений информационный поток поступает на коммутационную машину (коммутатор), которая в свою очередь производит логическую коммутацию передаваемых информационных потоков в другие локальные сети, их объединения, концентраторы сообщений или изолированные абонентские пункты.

Можно выделить следующие функционально законченные элементы сети:

локальные сегменты сети (с различной архитектурой). Их особенностью является возможность использования удаленных ресурсов файловых серверов или других рабочих станций, абонентских пунктов;

коммуникационные сегменты сети, которые производят фрагментирование и объединение пакетов данных, их коммутацию и собственно передачу.

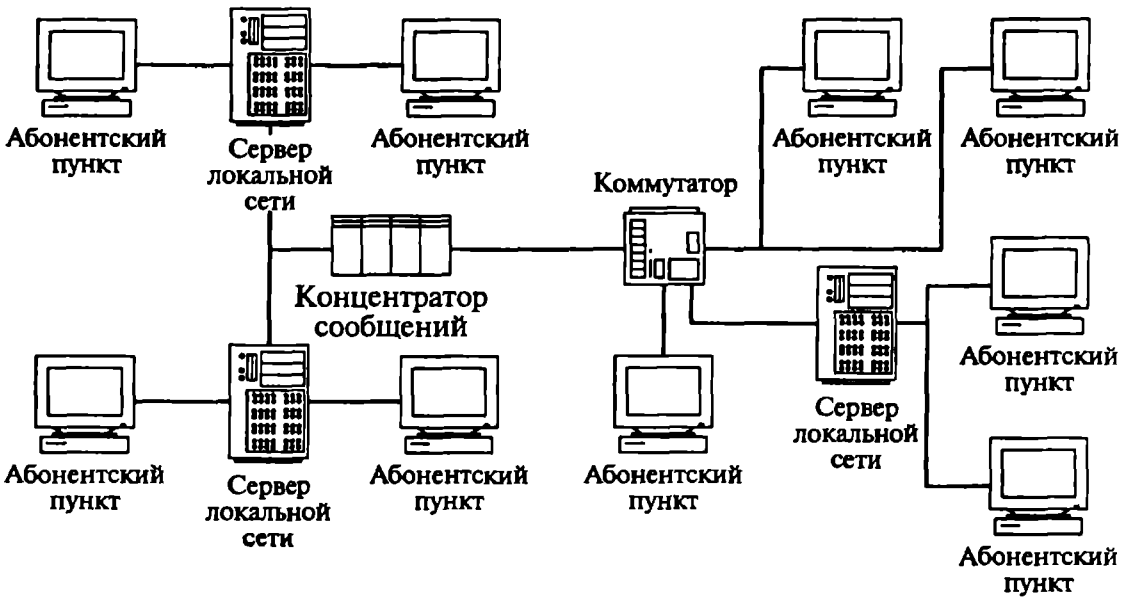


Рис. 7.2. Обобщенная структура сети ЭВМ

Как правило, рабочие станции не могут использоваться для доступа к ресурсам коммуникационных фрагментов вне решения задач передачи сообщений или установления логических соединений.

Для различных сегментов сети можно выделить следующие угрозы безопасности информации:

перехват, искажение, навязывание информации со стороны фрагментов сети;

имитация посылки ложных сообщений на локальные фрагменты сети;

имитация логического канала (удаленный доступ) к ресурсам локальных сегментов сети;

внедрение в циркулирующие по сети данные кодовых блоков, которые могут оказать деструктивное воздействие на программное обеспечение и информацию;

перехват, навязывание, искажение информации при передаче по собственным линиям связи локальных сегментов сети;

внедрение программных закладок в программное обеспечение рабочих станций или в общедоступные ресурсы (во внешней памяти файл-серверов) локальных сегментов сети.

По принципу действий программной закладки на компьютерную сеть можно выделить две основные группы.

1. Существуют закладки вирусного типа, которые способны уничтожать или искажать информацию, нарушать работу программного обеспечения. Закладки такого типа особенно опасны для абонентских пунктов сети и рабочих станций локальных вычислительных сетей. Они могут распространяться от одного абонентского пункта к другому с потоком передаваемых файлов или инфицировать программное обеспечение рабочей станции при использовании удаленных ресурсов (при запуске инфицированных программ в оперативной памяти рабочей станции даже без экспорта выполняемого модуля с файл-сервера).

2. В сетях могут функционировать специально написанные закладки типа «троянский конь» и «компьютерный червь». «Троянский конь» включается и проявляет себя в определенных условиях (по времени, ключевым сообщениям и т. п.). «Троянские кони» могут разрушать и (или) искажать информацию, копировать фрагменты конфиденциальной информации или пароли (ключи), засылать сообщения не по адресу или блокировать прием (отправку) сообщений. Закладки этого типа, как правило, жестко функциональны и учитывают различные особенности и свойства программно-аппаратной среды, в которой работают. Информация для их работы доставляется закладками следующего типа — «компьютерный червь».

«Компьютерные черви» нацелены на проникновение в системы разграничения доступа пользователей к ресурсам сети. Такие

закладки могут приводить к утере (компрометации) матриц установления полномочий пользователей, нарушению работы всей сети в целом и системы разграничения доступа в частности. Примером закладки этого типа является известный репликатор Морриса.

Возможно создание закладок, объединяющих в себе черты и свойства как «троянских коней», так и «компьютерных червей».

Программные закладки представляют опасность как для абонентских пунктов с их программным обеспечением, так и для коммутационной машины и серверов локальных сетей.

Возможны следующие пути проникновения (внедрения) программных закладок в сеть:

заражение программного обеспечения абонентских пунктов вирусами и деструктивными программами типа «троянских коней» и «компьютерных червей» вследствие нерегламентированных действий пользователей (запуска посторонних программ, игр, иных внешне привлекательных программных средств — архиваторов, ускорителей и т. п.);

умышленное внедрение в программное обеспечение абонентских пунктов закладок типа «компьютерных червей» путем их ассоциирования с выполняемыми модулями или программами начальной загрузки;

передача деструктивных программ и вирусов с пересылаемыми файлами на другой абонентский пункт и заражение его в результате пользования зараженными программами;

распространение вирусов внутри совокупности абонентских пунктов, объединенных в локальную сеть общего доступа;

внедрение в программное обеспечение абонентских пунктов вирусов при запуске программ с удаленного терминала;

внедрение вирусов и закладок в пересылаемые файлы на коммутационной машине и (или) на сервере локальной сети.

Телекоммуникационные сети, как правило, имеют неоднородную операционную среду, поэтому передача вирусов по направлению абонентский пункт — коммутатор чрезвычайно затруднена: пользователи с абонентских пунктов не могут получить доступ к программному обеспечению коммутатора, поскольку информация на коммутаторе представляется в фрагментированном виде (в виде пакетов) и не контактирует с программным обеспечением коммутационной машины. В этом случае заражение вирусами может наступать только при пользовании коммутационной машиной как обычной ЭВМ (для игр или выполнения нерегламентированных работ). При этом возможны заражение коммуникационного программного обеспечения и негативное влияние на целостность и достоверность передаваемых пакетов.

Исходя из перечисленных путей проникновения вирусов и возникновения угроз в сети можно детализировать вирусные угрозы.

Для абонентских пунктов:

искажение (разрушение) файлов и системных областей DOS;

уменьшение скорости работы, неадекватная реакция на команды оператора и т.д.;

вмешательство в процесс обмена сообщениями по сети путем непрерывной посылки хаотических сообщений;

блокирование принимаемых или передаваемых сообщений, их искажение;

имитация физических сбоев (потери линии) и т.д.;

имитация пользовательского интерфейса или приглашений для ввода пароля (ключа), с целью запоминания этих паролей (ключей);

накопление обрабатываемой конфиденциальной информации в скрытых областях внешней памяти;

копирование содержания оперативной памяти для выявления ключевых таблиц или фрагментов ценной информации;

искажение программ и данных в оперативной памяти абонентских пунктов.

Для серверов локальных сетей:

искажение проходящей через сервер информации (при обмене между абонентскими пунктами);

сохранение проходящей информации в скрытых областях внешней памяти;

искажение или уничтожение собственной информации сервера (в частности, идентификационных таблиц) и вследствие этого нарушение работы локальной сети;

внедрение вирусов в файлы, пересылаемые внутри локальной сети или на удаленные абонентские терминалы.

Для коммутатора:

разрушение собственного программного обеспечения и вывод из строя коммутационного узла вместе со всеми присоединенными абонентскими терминалами;

засылка пакетов не по адресу, потеря пакетов, неверная сборка пакетов, подмена пакетов;

внедрение вирусов в коммутируемые пакеты;

контроль активности абонентов сети для получения косвенной информации о характере данных, которыми обмениваются абоненты.

7.6. Организационно-технические меры защиты от угроз безопасности сети

Проанализировав возможные вирусные угрозы в сети и их последствия, можно предложить комплекс защитных мер, снижающих вероятность проникновения и распространения деструктивных программ в сети, а также облегчающих локализацию и устранение негативных последствий их воздействия.

Меры защиты нужно предусматривать как на этапе разработки программного обеспечения сети, так и на этапе ее эксплуатации.

Прежде всего на этапе разработки необходимо выявить в исходных текстах программ те фрагменты или подпрограммы, которые могут обеспечить доступ к данным по фиксированным паролям, беспарольный доступ по нажатию некоторых клавиш или их сочетаний, обход регистрации пользователей с фиксированными именами и реализацию тому подобных угроз. Наличие таких фрагментов фактически сведет на нет весь комплекс информационной безопасности сети, поскольку доступ через них возможен как человеком, так и программой-вирусом (закладкой). Присутствие таких фрагментов не всегда является результатом злого умысла. Зачастую подобные фрагменты используется для тестирования программного обеспечения. Для выявления подобных фрагментов может быть произведено сквозное тестирование исходных текстов программного обеспечения независимыми экспертами по стандартным, нормативно утвержденным методикам; тестирование готового программного обеспечения в критических режимах эксплуатации (в период испытаний сети) с фиксацией и устранением выявленных слабостей и отклонений от нормальной работы.

Необходимо также обратить внимание на возможные конфликты прикладного программного обеспечения и средств защиты. Такие конфликты могут возникнуть вследствие конкуренции по ресурсам (захват прерываний, памяти, блокировка клавиатуры и т.д.). Эти моменты, как правило, можно выявить лишь в период испытаний.

Также на этапе разработки должны быть предусмотрены меры защиты от несанкционированного доступа, меры по проверке целостности хранимых на внешних носителях программных средств зашиты, контроль целостности их в оперативной памяти и т.д.

На этапе штатной эксплуатации должны на регулярной основе предприниматься меры защиты и контроля, проводиться разовые эпизодические защитные мероприятия в период повышения опасности информационного нападения, локализационно-восстановительные меры, применяемые в случае проникновения и обнаружения закладок и причинения ими негативных последствий.

Сеть должна иметь общие средства и методы защиты. К ним относятся:

1. Ограничение физического доступа к абонентским терминалам, серверам локальных сетей и коммутационному оборудованию. Для такого ограничения устанавливается соответствующий организационный режим и применяются аппаратные и программные средства ограничения доступа к ЭВМ.

2. При активизации коммуникационного программного обеспечения контролируется его целостность и целостность областей

DOS, BIOS и CMOS. Для такого контроля подсчитываются контрольные суммы и вычисляются хеш-функции, которые потом сравниваются с эталонными значениями для каждой ЭВМ.

3. Максимальное ограничение и контроль за передачей по сети исполняемых файлов с расширениями типа *.exe и *.com, *.sys и *.bin. При этом снижается вероятность распространения по сети файловых вирусов, вирусов типа Driver и загрузочно-файловых вирусов.

4. Организация выборочного и внезапного контроля работы операторов для выявления фактов использования нерегламентированного программного обеспечения.

5. Сохранение архивных копий применяемого программного обеспечения на защищенных от записи магнитных носителях (дискетах), учет и надежное хранение архивных копий.

6. Немедленное уничтожение ценной и ограниченной для распространения информации сразу по истечении потребности в ней (при снижении ее актуальности).

7. Периодическая оптимизация и дефрагментирование внешних носителей (винчестеров) для выявления сбойных или псевдосбойных кластеров и затирания фрагментов конфиденциальной информации при помощи средств типа SPEED DISK.

Проблема защиты от воздействий закладок имеет много общего с проблемой выявления и дезактивации компьютерных вирусов. Она разработана и изучена достаточно подробно [29]. Методы борьбы с закладками сводятся к следующим.

Общие методы защиты программного обеспечения:

контроль целостности системных областей, запускаемых прикладных программ и используемых данных;

контроль цепочек прерываний и фильтрация вызовов, критических для безопасности системы прерываний.

Эти методы действенны лишь тогда, когда контрольные элементы не подвержены воздействию закладок и разрушающее воздействие входит в контролируемый класс. Так, например, система контроля за вызовом прерываний не будет отслеживать обращение на уровне портов. С другой стороны, контроль целостности информации может быть обойден за счет навязывания конечного результата проверок, влияния на процесс считывания информации, изменения хеш-значений, хранящихся в общедоступных файлах.

Включение процесса контроля должно быть выполнено до начала влияния закладки, либо контроль должен осуществляться полностью аппаратными средствами с программами управления, содержащимися в ПЗУ;

создание безопасной и изолированной операционной среды;

предотвращение негативных последствий воздействия вирусов или закладок (например, запись на диск только в зашифрованном

виде на уровне контроллера). В результате этого теряет смысл сохранение информации закладкой, а также запрет записи на диск на аппаратном уровне.

Специальные методы выявления программ с потенциально опасными последствиями:

поиск фрагментов кода по характерным последовательностям (сигнатурам), свойственным закладкам либо, наоборот, разрешение на выполнение или внедрение в цепочку прерываний только программ с известными сигнатурами, заведомо не принадлежащим закладкам;

поиск критических участков кода методом семантического анализа (анализа фрагментов кода на выполняемые ими функции, например выполнение несанкционированной записи, часто сопряженный с дисассемблирование или эмуляцией выполнения).

Как было выяснено, большую опасность представляют программы-закладки, помещенные в ПЗУ (BIOS) и ассоциированные с существенно важными прерываниями. Для их автоматизированного выявления используется следующая методика.

1. Выделяется группа прерываний, существенных с точки зрения обработки информации защищаемой программой. Обычно это прерывания INT 13h, INT 40h (запись и чтение информации на внешние магнитные накопители прямого доступа), INT 14h (обмен с RS232 портом), INT 10h (обслуживание видеотерминала), а также в обязательном порядке прерывания таймера INT 8h, INT ICh и прерывания клавиатуры INT 9h и INT 16h.

2. Для выделенной группы прерываний определяются точки входа (адреса входа) в ПЗУ, используя справочную информацию либо выполняя прерывание в режиме трассировки.

3. Для выделенных адресов создаются цепочки исполняемых команд от точки входа до команды IRET, возврату управления из BIOS.

Запись в сегмент BIOS невозможна, и поэтому закладки в BIOS не могут применять механизм преобразования своего кода во время его исполнения в качестве защиты от изучения [29].

В цепочках исполняемых команд выделяются команды работы с портами, команды передачи управления, команды пересылки данных.

Они используются либо для информативного анализа, либо порождают новые цепочки исполняемых команд. Порождение новых цепочек исполняемых команд происходит тогда, когда управление передается внутри сегмента BIOS.

4. В цепочках анализируются команды, предусматривающие работу с недокументированными портами. Наличие таких команд, как правило, указывает на передачу информации некоторому устройству, подключенному к параллельному интерфейсу (общей шине), например встроенной радиопередающей закладке.

В случае если опасных действий не обнаружено, аппаратно-программная среда ЭВМ без загруженной операционной среды считается безопасной.

Для проверки операционной системы используется аналогичный алгоритм.

1. По таблице прерываний определяются адреса входа для существенно важных прерываний.

2. Эти важные прерывания выполняются покомандно в режиме трассировки с анализом каждой команды по приведенному выше алгоритму. В этом случае команды типа JMP не анализируются, поскольку в режиме покомандного выполнения переходы происходят автоматически. Выполнение происходит до того момента, когда достигается адрес ПЗУ. Для полного анализа необходимо выполнить все используемые программой функции исследуемого прерывания.

7.7. Создание изолированной программной среды

Утечки (потери) информации гарантированно невозможны, если программная среда изолирована:

на ЭВМ с проверенным BIOS установлена проверенная операционная среда;

достоверно установлена неизменность DOS и BIOS;

кроме проверенных программ в данной программно-аппаратной среде не запускалось никаких иных программ;

исключен запуск проверенных программ в какой-либо иной ситуации, т.е. вне проверенной среды.

Выполнение перечисленных условий может быть достигнуто при использовании загрузочной дискеты, без которой невозможен запуск программ. Такая загрузка является надежным методом установления собственной операционной среды, но лишь тогда, когда оператор не допускает ошибок или преднамеренных деструктивных действий (запускает находящиеся на дискете программы без загрузки с дискеты). Подобного рода действия не только нарушают изолированность системы, но и могут привести к внедрению закладок в ранее проверенные программы пользователя или операционную среду.

Концепция ограниченного доверия к программно-аппаратной среде состоит в следующем:

аппаратная среда ЭВМ не содержит закладок и остается неизменной на протяжении всего времени работы. В противном случае работа на данной ЭВМ не ведется;

программная среда (операционная система и прикладные программы) данной ЭВМ может подвергаться воздействию злоумышленника и произвольным образом измениться;

пользователь располагает магнитным носителем, содержащим набор проверенных программ, проверенную DOS и проверенную программу контроля неизменности BIOS и доступных пользователю исполняемых файлов, а также данные для проведения контроля целостности;

использование указанного носителя невозможно иным образом, нежели как после загрузки операционной среды с его помощью.

7.8. Комплексный характер проблемы защиты информации в сетях ЭВМ

Рассмотренные выше примеры воздействия вирусов и закладок на компьютерные системы позволяют сделать вывод об их огромной опасности.

Внедрение закладок способно причинить системе, ее пользователям и абонентам ущерб, соизмеримый и даже превосходящий стоимость самой системы. Результат внедрения способен полностью скомпрометировать системы защиты информации и саму информацию.

На сегодняшний день известно уже несколько случаев проводок фальшивых финансовых документов в системах автоматизированного документооборота банков при помощи внедренных программных закладок блокирования управления технологическими и другими процессами.

В связи с этим разработчикам и менеджерам систем защиты информации рекомендуется обратить на данный аспект серьезное внимание, учесть его в процессе разработки и эксплуатации компьютерных систем и наряду с применением технических и программных мер защиты применять также юридические и организационные меры.

Данные, к которым может быть осуществлен несанкционированный доступ, должны находиться под защитой. Для того чтобы достичь нужного уровня их защиты, следует последовательно пройти четыре препятствия, предпринять реализацию четырех уровней защиты.

Первый уровень — правовой. Этот аспект защиты информации связан с соблюдением этических и юридических норм при передаче и обработке информации. Это важно, хотя законы, защищающие информацию, с которой оперируют компьютеры, еще далеки от совершенства. Может преследоваться незаконное использование секретных данных или информации, составляющей объект авторского права, но никак не копирование чужих файлов. Поэтому этический момент в соблюдении защиты имеет чрезвычайно большое значение.

За многие годы сложились представления, что информация — нематериальный объект, следовательно, цены у него нет. Кроме того, при копировании исходный файл не пропадает, поэтому кража файла, вроде бы, не приносит прямого материального ущерба.

Парадоксальный факт: если некий злоумышленник украдет магнитный диск, на котором записаны файлы с программами или данными, его, злоумышленника, можно привлечь к ответственности. Но только за кражу самого диска. Если злоумышленник скопирует записанные на этом диске файлы, кражу никто фиксировать не станет.

Серьезное препятствие организации правовой борьбы с информационной агрессией и современным хакерством представляет несогласованность правовых норм, принятых в разных странах, хотя компьютерные преступления часто переходит национальные границы. Внутринациональный компьютерный разбой не имеет таких проблем и с юридической точки зрения ничем не отличается от обычной преступности. Но если компьютерное преступление совершено за рубежом или подозреваемые действовали из другой страны, то традиционные концепции суверенитета строго ограничивают применение национального уголовного права и юрисдикции.

Все сказанное означает, что правовые нормы, регулирующие отношения в области защиты информации в информационно-вычислительных системах и сетях, пока что малоэффективны. Тем не менее общественная мораль должна распространить на компьютерные данные те же принципы, которые не позволяют читать чужие письма и рыться в чужих вещах.

Второй уровень — административный. Во всех организационных и организационно-технических системах с учетом правовых норм и производственных требований вводится система уровней компетенции работников. В соответствии с этой системой назначаются приоритеты: кто и какую информацию может собирать и хранить, устанавливаются способы доступа к ней и условия ее распространения, права и обязанности работников, их компетенция и ответственность. Регламентируется процедуры выдачи допусков к данным.

Многие из этих правил определяются внешними факторами: законами и иными нормативными актами. Но большинство проблем решается внутри организации специальными приказами и инструкциями.

Следует иметь в виду, что практическое осуществление административных мер обеспечения информационной безопасности связано с ограничением доступа людей к компьютерам и обрабатываемой ими информацией.

Организационные меры защиты информации по сравнению с этическими кажутся скучными, а по сравнению с программными

и техническими — лишенными конкретности и малоэффективными. Однако они представляют собой мощный барьер на пути незаконного использования информации и основу для других уровней. Никакой другой уровень защиты от НСД не может эффективно работать без соответствующей административной поддержки, но и административные меры не могут работать сами по себе, без опоры на этические и правовые нормы, без технического и программного обеспечения.

Одна из основных причин, по которой трудно проводить в жизнь эффективные административные меры, базируется на стойком общественном мнении, что защита информации — новая и необычная задача. Однако это совсем не так. Защита от НСД применялась во все времена. Просто современные способы представления данных изменились, существенно упростив процедуры НСД. Действительно, во времена преимущественного использования бумажных носителей информации получить доступ к комплекту конструкторской документации на более или менее серьезную техническую систему и скопировать несколько центнеров этой документации было существенно сложнее, чем в наше время переписать дискету.

Другая проблема при введении организационных мер защиты состоит в том, что их реализация почти неизбежно создает неудобства для пользователей. Если хлопот много, эффективность административных мер сведется к нулевой: дверь перестанут запирать, список паролей повесят на стену и т.д. При этом стоит помнить, что любые административные меры защиты вызывают у сотрудников ощущение ограничения их гражданских прав и необходимости выполнять дополнительную работу за ту же зарплату. Поэтому прежде чем вводить административные ограничения следует найти и внедрить рациональные побудительные причины для их исполнения. Нужно четко отдавать себе отчет в том, что большинство организационных мер защиты основано на преимуществе администратора — нарушителя нужно найти и наказать. Считается, что виновных без персональной ответственности не бывает. Поэтому, распределяя ответственность, сразу нужно предусмотреть систему проверок выполнения мер защиты, которые должны быть неожиданными и предельно простыми.

Третий уровень — аппаратно-программный. Он состоит в применении такой процедуры идентификации пользователя, которая открывает доступ к данным и программным средствам. Аппаратная защита может быть выполнена в виде кодовой карточки, ключа и т.п.

Радикальный способ аппаратной защиты — запирать в сейф или сдавать под ответственную охрану съемный жесткий диск. А такая защита, как запирание клавиатуры на ключ или пароли

при загрузке, не выдерживает даже самых простых атак. Во-первых, обычно из 10 ключей, блокирующих клавиатуру ЭВМ, минимум 5 совпадают, и, имея связку из 3 отмычек, можно открыть клавиатуру почти любой ЭВМ с вероятностью единица. Во-вторых, можно загрузить в ЭВМ с гибкого диска свою операционную систему, которая скопирует жесткий диск физически. Иногда даже это действие лишнее: популярный в начале 90-х гг. XX в. администратор диска, запрашивавший пароль при загрузке, «раскалывался», если с дискеты загружали DOS фирмы Digital Research.

Самое слабое место аппаратной защиты — персонал. Люди обычно отказываются от использования любых дополнительных средств защиты, создающих им неудобства в работе. Поэтому применение аппаратных средств защиты требует административной поддержки.

Четвертый уровень — криптографический. Это шифрование данных для скрытия от злоумышленника их смысла. До тех пор пока пользователь не идентифицирован по ключу, смысл данных ему недоступен. Данные в этом случае рассматриваются как сообщения, и для их защиты используется техника из арсенала методов шифрования.

Современным требованиям к криптографическим преобразованиям данных вполне отвечают системы, созданные по стандарту шифрования ГОСТ 28147—89. Так как некоторые данные критичны к искажениям, которые нельзя обнаружить исходя из контекста, приходится использовать лишь такие способы шифрования, которые чувствительны к искажению любого символа. Эти способы криптозащиты гарантируют не только высокую секретность, но и эффективное обнаружение любых искажений или ошибок.

Таким образом, юридические нормы защиты информации от НСД малоэффективны, и применение их в отечественном правовом пространстве затруднено неполнотой и непоследовательностью законодательной базы, хотя новый, действующий ныне Уголовный кодекс и содержит ряд статей, касающихся компьютерных преступлений.

Административные меры надежнее правовых. Они всегда позволяют доказать, что сделано все возможное для защиты данных и предпринято все необходимое для того, чтобы найти и наказать нарушителя правил обращения с информацией.

Меры по защите аппаратуры ЭВМ экзотичны для наших деловых традиций, так что об их эффективности довольно трудно судить ввиду их малой распространенности.

Последняя надежда — криптографическая защита, дает абсолютную защиту данных, если ей пользоваться умело при поддержке необходимых административных мер.

Контрольные вопросы

1. Перечислите основные угрозы информации в вычислительных системах и сетях.
2. Какие вам известны виды деструктивных программ?
3. Какие вам известны виды компьютерных вирусов? В чем различие действия разных вирусов?
4. Как работают компьютерные вирусы?
5. Как работают антивирусные программы?
6. В чем различие компьютерных вирусов и программных закладок?
7. Какие меры борьбы с деструктивными программными воздействиями вам известны?

1. Антенны и устройства СВЧ / под ред. Д. И. Воскресенского. — М. : МАИ, 1999.

2. *Бакулев П. А.* Радиолокационные системы / П. А. Бакулев. — М. : Радиотехника, 2004.

3. *Березин Л. В.* Теория и проектирование радиосистем / Л. В. Березин, В. А. Вейцель. — М. : Сов. радио, 1977.

4. *Вакин С. А.* Основы радиопротиводействия и радиотехнической разведки / С. А. Вакин, Л. Н. Шустов. — М. : Сов. радио, 1968.

5. *Вартанесян В. А.* Радиоэлектронная разведка / В. А. Вартанесян. — М. : Воениздат, 1991.

6. *Герасименко В. А.* Основы защиты информации / В. А. Герасименко, А. А. Малюк. — М. : МИФИ, 1997.

7. *Гоноровский И. С.* Радиотехнические цепи и сигналы / И. С. Гоноровский. — М. : Сов. радио, 1967.

8. *Громаков Ю. А.* Стандарты и системы подвижной радиосвязи / Ю. А. Громаков. — М., 1997.

9. *Жельников В.* Криптография от папируса до компьютера / В. Жельников. — М. : ABF, 1996.

10. *Калинцев Ю. К.* Криптозащита сообщений в системах связи / Ю. К. Калинцев. — М. : МТУСИ, 2000.

11. *Касперский Е. В.* Компьютерные вирусы: что это такое и как с ними бороться / Е. В. Касперский. — М. : СК Пресс, 1998.

12. *Киселев В. Д.* Защита информации в современных системах ее передачи и обработки / В. Д. Киселев, О. В. Есиков, А. С. Кислицин. — М. Солид, 2002.

13. *Куприянов А. И.* Радиоэлектронные системы в информационном конфликте / А. И. Куприянов, А. В. Сахаров. — М. : Вузовская книга, 2003.

14. *Ландау Л. Д.* Теоретическая физика. Т. VII. Теория упругости / Л. Д. Ландау, Е. М. Лифшиц. — М. : Наука, 1965.

15. *Макаров Ю. К.* Методы защиты речевой информации и оценка их эффективности / Ю. К. Макаров, А. А. Хорев // Конфидент. — 2001, № 4.

16. *Меньшаков Ю. К.* Защита объектов информации от технических средств разведки / Ю. К. Меньшаков. — М. : Российский гос. гуманит. ун-т, 2002.

17. Основы радиоуправления : учеб. пособие для вузов / [П. А. Агаджанов, В. А. Вейцель, С. А. Волковский и др.] ; под ред. В. А. Вейцеля. — М. : Радио и связь, 1995.

18. *Оуэн Г.* Теория игр / Г. Оуэн. — М. : Мир, 1971.

19. *Пенин П. И.* Системы передачи цифровой информации / П. И. Пенин. — М. : Сов. радио, 1976.

20. Помехозащищенность радиосистем со сложными сигналами / под ред. Г. И. Тузова. — М. : Радио и связь,1985.

21. *Рабинер Л. Р.* Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер ; под ред. М. В. Назарова, Ю. Н. Прохорова ; пер. с англ. — М. Радио и связь, 1981.

22. Радиотехнические системы передачи информации : учеб. пособие для вузов / [В. А. Борисов, В. В. Калмыков, Я. М. Ковальчук и др.] ; под ред. В. В. Калмыкова. — М. : Радио и связь, 1990.

23. Теория передачи сигналов / [А. Г. Зюко, Д. Д. Кловский, М. В. Назаров и др.] — М. : Радио и связь, 1986.

24. *Тихонов В. И.* Оптимальный прием сигналов / В. И. Тихонов. — М. : Радио и связь, 1983.

25. *Уайт Д.* Электромагнитная совместимость радиоэлектронных средств и непреднамеренные помехи: в 3 вып. / Д. Уайт. — М. : Сов. радио, 1977—1979.

26. *Устинов Г. Н.* Основы информационной безопасности систем и сетей передачи данных / Г. Н. Устинов. — М. : СИГНЕТ, 2000.

27. Цифровые радиоприемные системы : справочник / под ред. М. И. Жодзишского. — М. : Радио и связь, 1990.

28. *Шеннон К. Э.* Теория связи в секретных системах. Работы по теории информации и кибернетике / К. Э. Шеннон. — М. : ИЛ., 1963.

29. *Щербаков А. Ю.* Введение в теорию и практику компьютерной безопасности / А. Ю. Щербаков. — М. : Издатель С. В. Молгачева, 2001.

30. *Янке Е.* Специальные функции / Е. Янке, Ф. Эмде, Ф. Леш. — М. Наука, 1968.

31. *Ярочкин В. И.* Информационная безопасность / В. И. Ярочкин. — М. : Академический проект, 2003.

32. Доктрина информационной безопасности Российской Федерации // www/gov/ru, 2000.

ОГЛАВЛЕНИЕ

Предисловие 3

Глава 1. Проблема информационной безопасности 5

1.1. Современное состояние, перспектива и ретроспектива 5
1.2. Информационные системы, средства, каналы, сети и среды 19

Глава 2. Информация 24

2.1. Количество информации 24
2.2. Качество информации 34
2.3. Ценность информации 36

Глава 3. Угрозы безопасности информации и информационные атаки 45

3.1. Информационные угрозы 45
3.2. Информационные атаки 47
3.3. Технические каналы утечки информации 48
3.3.1. Электромагнитные каналы утечки информации 51
3.3.2. Акустические каналы утечки информации 58

Глава 4. Физические поля, создающие каналы утечки информации 60

4.1. Многообразие физических полей 60
4.2. Электромагнитные поля 61
4.2.1. Электромагнитные поля и волны 61
4.2.2. Излучение электромагнитных волн радиодиапазона антеннами 63
4.2.3. Непреднамеренное излучение электромагнитных полей 68
4.2.4. Собственное излучение электромагнитного поля 74
4.2.5. Распространение электромагнитных волн 78
4.3. Акустические поля 90
4.4. Геофизические поля 97
4.4.1. Сейсмические поля и волны 97
4.4.2. Гравитационные поля 105

Глава 5. Защита информации в каналах связи и передачи данных 112

5.1. Кодирование для защиты информации от искажения помехами в системах передачи 112
5.2. Обратная связь для адаптации к помеховой обстановке 130
5.3. Искажения кодированных сообщений помехами 137
5.4. Шифрация для защиты от несанкционированного доступа к информации 141
5.5. Стандарты симметричных криптосистем 152
5.6. Двухключевые криптосистемы (криптосистемы с открытым ключом) 157

5.7. Стойкость к имитирующим и дезинформирующим помехам (обеспечение подлинности сообщений) 163

Глава 6. Информационная безопасность систем и сетей связи 176

6.1. Сети связи 176
6.2. Основные угрозы безопасности информации и методы защиты информации в кабельных телефонных сетях 178
6.3. Методы и средства защиты информации в мобильных системах 185
6.3.1. Защита информации в цифровых системах мобильной связи стандарта GSM 185
6.3.2. Защита информации в цифровых системах мобильной связи с кодовым разделением каналов 192
6.4. Методы представления речевого сигнала 195
6.4.1. Компрессия аналогового речевого сигнала 197
6.4.2. Дискретные методы передачи и обработки речевого сигнала 199
6.4.3. Критерии оценки систем закрытия речи 204
6.5. Функциональная схема системы закрытой связи 206

Глава 7. Защита информации в вычислительных системах и средах 209

7.1. Угрозы информационным ресурсам и информационные атаки на вычислительные системы 209
7.2. Компьютерные вирусы 217
7.2.1. Общие сведения о компьютерных вирусах 217
7.2.2. Принципы функционирования основных разновидностей вирусов 222
7.2.3. Использование СТЕЛС технологии в вирусных программах 226
7.3. Программные средства борьбы с вирусами 228
7.4. Программные закладки 232
7.5. Действие вирусов и программных закладок в сетях ЭВМ 238
7.6. Организационно-технические меры защиты от угроз безопасности сети 241
7.7. Создание изолированной программной среды 245
7.8. Комплексный характер проблемы защиты информации в сетях ЭВМ 246

Список литературы 251

Учебное издание

Куприянов Александр Ильич
Сахаров Андрей Владимирович
Шевцов Вячеслав Алексеевич

Основы защиты информации

Учебное пособие

Редактор *И. В. Могилевец*
Технический редактор *О. Н. Крайнова*
Компьютерная верстка: *В. А. Крыжко*
Корректоры *М. В. Дьяконова, Т. В. Кузьмина, И. Н. Волкова*

Диапозитивы предоставлены издательством

Изд. № А-1624-I. Подписано в печать 15.10.2005. Формат 60×90/16.
Гарнитура «Таймс». Печать офсетная. Бумага тип. № 2. Усл. печ. л. 16,0.
Тираж 3000 экз. Заказ № 15689.

Издательский центр «Академия». www.academia-moscow.ru
Санитарно-эпидемиологическое заключение № 77.99.02.953.Д.004796.07.04 от 20.07.2004.
117342, Москва, ул. Бутлерова, 17-Б, к. 360. Тел./факс: (095)334-8337, 330-1092.

Отпечатано на Саратовском полиграфическом комбинате.
410004, г. Саратов, ул. Чернышевского, 59.

ОСНОВЫ ЗАЩИТЫ ИНФОРМАЦИИ

ISBN 5-7695-2438-3
9785769524387

Издательский центр «Академия»
www.academia-moscow.ru