本资源来自数缘社区

http://maths.utime.cn:81

欢迎来到数缘社区。本社区是一个高等数学及密码学的技术性论坛，由山东大学数学院研究生创办。在这里您可以尽情的遨游数学的海洋。作为站长，我诚挚的邀请您加入，希望大家能一起支持发展我们的论坛，充实每个版块。把您宝贵的资料与大家一起分享！

数学电子书库

每天都有来源于各类网站的与数学相关的新内容供大家浏览和下载，您既可以点击左键弹出网页在线阅读，又可以点右键选择下载。现在书库中藏书 1000 余本。如果本站没有您急需的电子书，可以发帖说明，我们有专人负责为您寻找您需要的电子书。

密码学论文库

国内首创信息安全专业的密码学论文库，主要收集欧密会（Eurocrypt）、美密会（Crypto）、亚密会（Asiacrypt）等国内外知名论文。现在论文库中收藏论文 4000 余篇（包括论文库版块 700 余篇、论坛顶部菜单“密码学会议论文集”3000 余篇）。如果本站没有您急需的密码学论文，可以发帖说明，我们有专人负责为您寻找您需要的论文。

提示：本站已经收集到 1981－2003 年欧密会、美密会全部论文以及 1997 年－2003 年五大会议全部论文（欧密会、美密会、亚密会、PKC、FSE）。

数学综合讨论区

论坛管理团队及部分会员来源于山东大学数学院七大专业（基础数学、应用数学、运筹学、控制论、计算数学、统计学、信息安全），在数学方面均为思维活跃、成绩优秀的研究生，相信会给您的数学学习带来很大的帮助。

密码学与网络安全

山东大学数学院的信息安全专业师资雄厚，前景广阔，具有密码理论、密码技术与网络安全技术三个研究方向。有一大批博士、硕士及本科生活跃于本论坛。本版块适合从事密码学或网络安全方面学习研究的朋友访问。

网络公式编辑器

数缘社区公式编辑器采用 Latex 语言，适用于任何支持图片格式的论坛或网页。在本论坛编辑好公式后，您可以将自动生成的公式图片的链接直接复制到你要发的帖子里以图片的形式发表。

如果您觉得本站对您的学习和成长有所帮助，请把它添加到您的收藏夹。如果您对本论坛有任何的意见或者建议，请来论坛留下您宝贵的意见。

附录 A：本站电子书库藏书目录

http://maths.utime.cn:81/bbs/dispbbs.asp?boardID=18&ID=2285

附录 B：版权问题

国瑞数码安全系列丛书

信息隐藏技术
—— 隐写术与数字水印

Stefan Katzenbeisser
Fabien A.P. Petitcolas 编

吴秋新 钮心忻
杨义先 罗守山 杨晓兵 译

人民邮电出版社
www.pptph.com.cn

国瑞数码安全系列丛书

信息隐藏技术
——隐写术与数字水印

Stefan Katzenbeisser　Fabien A.P. Petitcolas　编
吴秋新　钮心忻　杨义先　罗守山　杨晓兵　译

人民邮电出版社

图书在版编目(CIP)数据

信息隐藏技术——隐写术与数字水印/(英)卡曾贝塞(Katzenbeisser,S.),(英)佩蒂科勒斯(Petitcolas,F.A.P.)编;吴秋新等译.—北京:人民邮电出版社,2001.9

(国瑞数码安全系列丛书)

ISBN 7-115-09550-7

Ⅰ.信… Ⅱ.①卡…②佩…③吴… Ⅲ.数据通信-安全技术 Ⅳ.TN919

中国版本图书馆 CIP 数据核字(2001)第 049895 号

内 容 提 要

本书详细介绍了涉及数据通信安全的信息隐藏技术,以及用于数字产品知识产权保护的水印技术。除技术本身外,该书还涉及它们的历史、相互的差异、隐蔽通信的破译、水印的删除,以及数字水印和版权问题的法律意义。

本书适合关心网络通信安全和知识产权的读者,他们包括从事网络通信安全和水印制作的工程技术人员、管理人员、法律工作者、学者,也包括从事隐密通信和反盗版的情报人员、技术人员。

国瑞数码安全系列丛书

信息隐藏技术——隐写术与数字水印

◆ 编 Stefan Katzenbeisser Fabien A.P.Petitcolas

译 吴秋新 钮心忻 杨义先 罗守山 杨晓兵

责任编辑 陈万寿

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ pptph.com.cn

网址 http://www.pptph.com.cn

读者热线 010-67129212 010-67129211(传真)

北京汉魂图文设计有限公司制作

北京顺义振华印刷厂印刷

新华书店总店北京发行所经销

◆ 开本:787×1092 1/16

印张:11

字数:254 2001 年 9 月第 1 版

印数:1-4 000 册 2001 年 9 月北京第 1 次印刷

著作权合同登记 图字:01-2001-1103 号

ISBN 7-115-09550-7/TN·1755

定价:20.00 元

版 权 声 明

本书原版书名 Information Hiding Techniques for Steganography and Digital Watermarking

作者 Stefan Katzenbeisser, Fabien A. P. Petitcolas, editors

作 者 简 介

Stefan Katzenbeisser 是维也纳技术大学计算机科学系的一名学生。

Fabien A.P. Petitcolas 从英国剑桥大学获得博士学位，现受聘于微软剑桥研究院。

译 者 的 话

网络信息安全保障迫在眉睫。现在世界上每年因利用计算机网络进行犯罪所造成的直接经济损失令人吃惊。利用计算机通过互联网络窃取机密信息的事例也是屡见不鲜。网络信息安全隐患，将全方位地危及社会的经济、政治、文化等各个方面。

当前，我国社会信息化正以一日千里的速度前进，对网络与信息安全的需求日益增大。与其他领域不同，网络与信息安全问题必须依靠我国自己的力量来解决。引进国外产品或照搬国外先进技术来解决信息安全问题无异于引狼入室。为此，国家已经明确规定："信息安全产品一定要立足国内，自主开发"。

当前国内在网络与信息安全方面的基础（人员和技术）还相当薄弱，急需加强。"在游泳中学游泳"，将国际上最新出版的一些著作以翻译的形式介绍给广大读者，是使国内同行更好更快地了解国际上在信息网络安全方面的最新进展的最佳途径之一。为了取得更好的效果，人民邮电出版社与北京邮电大学信息安全中心和天津市国瑞数码安全系统有限公司共同合作翻译出版了这套"国瑞数码安全系列丛书"。希望本系列丛书能够为促进我国的网络信息安全做出一定的贡献。

天津市国瑞数码安全系统有限公司（http://www.ncs-cyber.com.cn）是一家以网络安全和信息安全为主的高科技企业，公司致力于为我国社会信息化提供全方位的网络安全和信息安全保障。公司拥有一大批国内一流的密码学和信息网络安全专家，硕士、博士学位获得者超过公司员工总数的一半，公司真诚地欢迎更多的有志于我国数码安全的专家加盟（联系电话：010-62383780、022-27237081）。公司已经开发出具有完全自主知识产权的众多信息与网络安全产品。比如：B2B 电子商务安全平台、电子商务加速卡、网站卫士、安全替音电话、WAP 安全解决方案、宽频系统安全结构和多种加密卡和加密算法等。

北京邮电大学信息安全中心（http:\\www.bupt.edu.cn）是一个历史悠久的部级重点实验室，是国务院学位委员会正式批准的全国仅有的三个"密码学"博士点之一，长期致力于网络信息安全的理论和关键技术研究。欢迎有志青年来此攻读硕士、博士和博士后。

本丛书还得到了国家高等学校骨干教师资助计划项目（国家教育部）、国家重点基础研究发展规划项目（编号：G1999035805）、国家杰出青年基金项目（批准号：69425001）和国家自然科学基金项目（批准号：69882002，60073049）的资助。特此致谢。

北京邮电大学信息安全中心的研究生陈亚娟、王自亮、张振涛、张小芬等也参加了本书的部分翻译工作，特此致谢。

序

(Ross J. Anderson)

每隔几年,计算机安全领域必定会有新的发展和突破,新的技术和新的应用也会给网络安全带来新的威胁,从而迫使我们发明新的安全保护机制。当商业领域开始建立网络化的计算机系统时,密码学就立即变得重要起来;当大量 PC 用户要相互交换信息时,计算机病毒就开始流行起来;当因特网发展起来以后,防火墙技术则迅速发展起来。

信息安全研究领域最新热点之一是信息隐藏研究,其驱动力在于信息时代的两大政策问题——版权保护和状态监视。

数字音乐和视频作品极容易被完美复制这一特性使娱乐界非常不安,因为盗版这种数字作品要比模拟家用录音带的盗版频繁得多,而且容易得多。MP3 编码的音乐不断流行起来更加剧了这种恐惧。这些问题的部分解决方案可以是改变音乐和视频作品的销售方式,然而,软件工业很大程度上放弃了版权控制机制,而选择这样一种商业模式,即频繁升级、在线注册以获得技术支持、对大规模盗版进行起诉以及对从商业应用到游戏的每一种软件产品都网络化。但对数字音乐和视频产品,人们希望技术保护机制也将能提供部分解决方案。其中的一种保护机制就是版权标记——将版权标识和序列号隐藏在音频或视频中,并使得盗版者难以消去它们。

因特网的迅猛发展也对国家情报和公安机关产生了巨大影响。他们宣称,随处可得的加密软件可能使搭线监听越来越困难,他们通常的反应是尽力限制加密算法的强度,或者要求拷贝备用的密钥以使他们需要时能即时获得。这激起了人权自由崇尚者的义愤,并谴责这是对个人隐私不可忍受的侵犯。这两种观点从某种程度来讲都过于单纯了一点。绝大多数维护安全的通信侦察并不是与搭线监听有关,而主要是与追踪双方联络的网络有关,并且最典型的犯罪通信工具是预付款的移动电话。在这两种情形中,问题不在于通信的保密性,而是他们的可追踪性。通过使用为版权标识开发的技术,通信也能被隐藏起来,并且它们能帮助犯罪分子逃脱使用"非法"密码系统的法律制裁。

正如对于版权保护和赞同加密者对法律执行辩论的长期解决方案是十分重要的一样,信息隐藏对于隐私也是重要的。从人口普查到医疗记录,大量的个人信息在处理过程中是不应该被别有用心的人识别出来的。有时候,这方面做得很好,但有时候不需要付出太多的努力就可能重新确定数据的主体内容。

随着这些力量的驱动,信息隐藏的研究呈指数级增长,它在最近五年所取得的成就完全可与 1945—1990 年间密码学所取得的成就相提并论,大量的信息隐藏系统被设计出来,其中许多系统已被攻破。关于什么系统能用,什么系统不能用,所感兴趣的研究方向等等,我们现在已有一个公正的见解。

所以我非常高兴在这里看到了第一本在信息隐藏方面很全面的技术书籍,我希望在未来许多年里它将成为该学科的标准参考书。

前　言

本书对隐写术和数字水印作了全面的介绍，这两个研究领域一般都统称为“信息隐藏”。隐写术主要研究如何将秘密信息隐藏在不太容易引起注意的消息之中，从而使得秘密通信不被察觉，而数字水印则源于数字媒体版权保护的需求。

就在几年之前，信息隐藏技术还没有像密码学一样引起研究团体和工业界更多的关注。然而，最近形势迅速发生变化。1996 年，召开了这方面的第一次国际学术会议，其主要驱动力在于对版权保护的关注。由于音频、视频和其它作品都能以数字形式获得，制作完美拷贝变得非常容易，这将会导致大规模非授权的拷贝，而这恰好是音乐、电影、书籍和软件出版业最为担忧的问题。

信息隐藏研究使各种不同背景的研究人员走到一起来，比如：电子工程、信号与图像处理、计算机科学以及密码学等等，这里就不一一列举了。到目前为止，对这个相对较新的研究领域还没有形成一个全面而统一的看法。可以从不计其数的论文和会议文集中获得这方面的相关信息资料。根据一个大型文献目录信息系统统计，1998 年发表了 103 篇专门研究数字水印的论文，而 1992 年只发表了 2 篇，这也表明信息伪装和数字水印越来越重要。本书的目标就是为该研究领域提供一本既可作教科书又可作全面参考手册的书籍。

本书第一章介绍了信息隐藏领域的状况，并对信息隐藏可能的应用作了一个全面的描述。本书第一部分专门讨论隐写术，其中第二章讨论了信息伪装的基本原理。第三章列出了隐写术的各种应用。第四章重点讨论如何攻破伪装通信。

本书的第二部分专门描述了数字水印系统，其中第五章讨论了水印系统的目标和要求。第六章概览了数字水印研究领域中使用的各种方法。第七章的主题是讨论数字水印的关键问题——健壮性。第八章讨论了数字指纹问题。最后一章讨论了因特网上结合水印技术的版权的法律含义。

致　谢

我们非常感谢那些为本书付出艰辛劳动的作者们。尽管他们本身也很忙，但他们还是圆满完成了涉及他们研究课题的有关章节的写作，这需要他们付出相当多的努力，同时也要感谢他们的合作与协助。对我们来讲，能编辑这样一本书以及与他们一起工作是一件十分愉快和荣幸的事情。

我们也要感谢阿尔泰克出版公司（Artech House）的 Viki Williams、Susanna Taggart、Michael Webb 和 Hilary Sardella，是他们帮助我们克服了在本书出版过程中的各种困难。同时，我们也要感谢 Philipp Tomsich，是他帮助我们建立了一个共享的计算机帐户。感谢 Raimund Kirner，是他制作了本书的各种插图。最后，我们还要提及那些不知姓名的校阅者，是他们提供了各种有用的反馈信息，这些反馈信息对我们编写本书有极大的帮助。

Stefan C. Katzenbeisser
Fabien A. P. Petitcolas
1999 年 6 月于维也纳和剑桥

目　录

第二部分　数字水印与版权保护

第一章　信息隐藏入门

(Fabien A.P. Petitcolas)

由于音频、视频和其它作品都能以数字形式获得，制作其完美拷贝变得非常容易，从而可能会导致大规模非授权拷贝，而这极有可能会损害音乐、电影、书籍和软件等出版业的发展。对版权保护的这类关注引发了一个很有意义的研究方向：寻找将版权信息和序列号隐藏到数字媒体中的方法，其目标是：通过序列号来帮助识别版权侵犯者，而版权信息能用来检举和起诉盗版者。

同时，各级政府对普通百姓获取加密服务的限制，也驱使人们研究将私有信息嵌入到表面上看来无关紧要的掩饰信息之中的各种方法。

还有许多其它的应用前景引发人们对信息隐藏这一学科的兴趣。在这一章中，我们将描述它们中的一部分来说明这个研究主题的广阔性。但在此之前，我们将介绍与计算机系统有关的信息隐藏的各主要分支情况，并简要回顾这个引人注目的研究领域。

1.1　信息隐藏学的主要分支

隐蔽信道由 Lampson 在文献[1]中定义为：在多级安全水平的系统环境中(比如，军事计算机系统)，那些既不是专门设计的也不打算用来传输消息的通信路径称为隐蔽信道。这些信道在为某一程序提供服务时，可以被一个不可信赖的程序用来向它们的操纵者泄露信息。为了找到限制这种动机的方法，人们已经详细研究过这些通信信道[2]。在这个主题上我们将不会展开太多，只是介绍以太网上一个隐蔽通信的例子(见 2.7.2 节)，以及关于图像降质的背景和原由(见 3.2.3)。

匿名通信就是寻找各种途径来隐藏通信消息的主体，即消息的发送者和接收者。匿名通信早期的例子包括由 Chaum 在文献[3]所描述的信件匿名重发器和由 Goldschlag、Reed 和 Syverson 在文献[4]中所提出的洋葱路由，其想法是：只要中间参与者不相互串通勾结，通过使用一组邮件重发器或路由器，人们就可以将消息的踪迹隐蔽起来，因此信任是这些工具的基础。根据谁被"匿名"(发送者、接收者，或两者)，匿名通信又分为几种不同的类型。Web 应用强调接收者的匿名性，而电子邮件用户们更关心发送者的匿名性。

隐写术是信息隐藏学的一个重要分支。密码学研究如何保护消息内容，而隐写术专门研究如何隐藏实际存在的信息。隐写术的英文名词 Steganographia 是由 Trithemius(1462—1516)首先构造出来的，一般认为来源于希腊文 στεγανο-ζ、γραφ-ειν，其含义为"被掩盖的笔迹"[5]，该词的现代含义通常理解为将一个信息隐藏在另一个信息之中(图 1.2 显示了 Trithemius 著作的封面)。这方面的例子有：使用不可见墨水给报纸上的某些字母作上标记来向一个间谍发送消息，在一个录音带的某些位置加一些不易察觉的回声等。第二章将介绍一些把数据隐藏到另一些数据中的一般模型，而一些主要的伪装技术将在第三章进行阐述和评论。

与隐写术相反，水印技术需要增加健壮性要求，以对抗各种可能的攻击。在该领域中，术

语"健壮性"的含义一直不是很清楚,它取决于应用的场合,但一个成功的攻击只需使水印标记检测不出来即可。我们将在第七章展示这些攻击方法。健壮性在整个水印系统设计中具有非常重要的份量,这也是我们在本书中将隐写术和数字水印区别对待的原因之一。

水印并不总是需要隐藏起来,正如一些系统需要可见的数字水印[6],但在文献中绝大多数情况下还是强调不可察觉的(或称不可见的、透明的或听不见的,依上下文而定)数字水印,因为它们的应用范围更广泛些。可见的数字水印完全可追溯到13世纪末期出现的标明纸张来源的纸张水印(见5.2.1节)。现代的可见水印可以是放置在数字图像上的各种可见图案(比如,一个公司的标识或版权标记),它们被那些对不可见水印技术不信任的摄影师们广泛使用(见[7])。

从上述简要的介绍,读者也许已经注意到隐写术与水印的另一个根本的不同点,即水印系统所隐藏的信息总是与被保护的数字对象或它的所有者有关,而信息隐写系统可以隐藏任何信息。同时,"健壮性"评判标准也不同,因为隐写术主要关注被隐藏信息的检测,而水印技术则关注被盗版者擦除的可能性。最后,伪装通信通常是点对点的(在发送者与接收者之间),而数字水印技术通常是一点对多点的。

信息隐藏这两个分支的准确术语将在第二章和第五章给出。

1.2 对信息隐藏历史的简要回顾

在这一节里,我们并不想讲述信息隐藏的整个历史全貌,而只介绍一些重要的里程碑式的事件。希望了解更多历史细节的读者请参考文献Kahn[8]和[9,10]。

1.2.1 技术性的隐写术

隐写术最有名的例子可追溯到远古时代。Herodotus(C.486—425 B.C.)在他的著作 *Histories*[11]中讲述到:大约在公元前440年,Histiaus给他最信任的奴隶剃头并将一个消息刺在头上,直到他的头发重新长出后,这条信息才消失,这样做的目的是为了鼓动奴隶们起来反抗波斯人。令人惊讶的是,一些德国间谍在20世纪初期仍然使用这种方法。Herodotus还讲到,在波斯朝廷的一个希腊人Demeratus是如何警告斯巴达将发生一场由波斯国王薛西斯一世发动的入侵的。他先去掉书记板上的蜡,然后将他的消息写在蜡下面的木板上,最后再用蜡覆盖住那个消息。这个书记板看起来完全像一个空白的书记板(它几乎既欺骗了接收方也蒙骗了海关士兵)。有许多隐写技术是由战术家Aeneas发明或记载下来的[13],包括将信函隐藏在信使的鞋底里或妇女的耳饰中,将正文消息写在木板上然后用石灰水把它刷白,以及由信鸽携带便条传送等等。Aeneas也提出了通过改变字母笔画的高度或在掩蔽文体的字母上面或下面挖出非常小的小孔来隐藏正文,后一种方法直到17世纪还在使用,但后来Wilkins(1614—1672)对它进行了改进,他不是挖制小孔而是用无形的墨水制作非常小的斑点[14],并且该方法又被德国间谍在两次世界大战中重新使用起来[8,p.83]。这项技术的现代版本目前仍然在文本安全中使用,并且通过在纸页上打印各种小像素点组成的块来对诸如日期、打印机标识符、用户标识符等信息进行编码。

在1857年,Brewster就已经提出将秘密消息隐藏"在大小不超过一个句号或小墨水点的空间里"的设想[16]。到1860年,制作微小图像的基本难题已经被一个叫Dragon的法国摄影师

解决了，在 1870 至 1871 年弗朗格－普鲁士战争期间，巴黎被围困时，印制在微缩胶片上的消息就是通过信鸽发送出去的[17,18]。在 1905 年的俄日战争期间，把在显微镜下可见的图像隐藏在耳朵、鼻孔以及手指甲里。Brewster 的设想在第一次世界大战期间终于付诸实现，其作法是先将间谍之间要传送的消息经过若干照相缩影步骤后缩小到微粒状，然后粘在无关紧要的杂志等文字材料中的句号或逗号上[12,20]。

不可见墨水已被广泛使用。最初这种墨水是由随处可得的有机物（诸如牛奶或尿）或“将盐溶解于水中”制成的[14,V,pp.37—47]，并且通过加热来显影。化学的进步和第一次世界大战促进人们开发出更加先进的墨水和显影剂的化合物，但随着“万用显影剂”的发明，这项技术就被放弃不用了。“万用显影剂”的原理是根据纤维表面的效果来确定纸张的哪一部分被墨水加湿过[8,pp.523－525]。这导致了印刷品安全领域开发出更加成熟的面向应用的信息隐藏和标记技术[21,22]。纸币中的水印是一项非常古老的防伪技术，最近该项技术有更多的改进，其中包括在印刷旅行支票时使用特殊紫外线荧光墨水。由于在影印机里使用的灯有很高的紫外线成分，它能通过编排这些灯使印出来的影印品用大字母套印上“void”字样。读者若想对这些技术的最新进展有一个全面了解，请参阅 van Renesse 的论文[21,22]。

另一个例子来自建筑学。很早以前，艺术家们就体会出雕塑或绘画作品从不同角度看会显出不同的印象，并且建立了透视画法和夸张技法的准则[23]。纵观 16 和 17 世纪，变形夸张绘画提供了一种理想的手段来伪装危险的政治主张和异教的思想[24]。被隐藏的变形夸张绘画的一部杰作—*Vexierbild*—是 Shō 于 1530 年创作的，他是 Nürnberg 的一位雕刻师 Dürer（1471—1528）的学生。当你从正面直视它时，它像一幅奇怪的山水风景画，但从侧面看过去，则显出一些著名国王的肖像。

1.2.2　语言学中的隐写术

语言学隐写术中广泛使用的方法是藏头诗。最著名的例子可能要算 Giovanni Boccaccio（1313—1375）的诗作 *Amorosa visione*，据说是“世界上最宏伟的藏头诗”作品[25,pp.105—106]。Boccaccio 先创作了三首十四行诗，总共包含大约 1500 个字母，然后创作另一首诗，使连续三行押韵诗句的第一个字母恰好对应十四行诗的各字母。另一首藏头诗的著名例子来自小说 *Hypnerotomachia Poliphili*[26][1]，出版于 1499 年。这本晦涩和高深莫测的书，是由无名氏创作的，它展现了一个修道士和一个女人之间罪恶式的爱情故事。全书 38 章中每章的第一个字母组合在一起恰好拼出“Poliam frater Franciscus Columna peramavit.”[2]。

通过展开藏头诗这种简单的思路，修道士们和其他文化人士找到了更好的把消息隐藏到正文中去的方法。到 16 世纪和 17 世纪，已经出现了大量的关于伪装术的文献，并且其中许多方法依赖于新颖的信息编码手段。Gaspar Schott（1608—1666）在他的 400 页的著作 *Schola Steganographica*[27]中，扩展了由 Trithemius 在 *Polygraphia* 一书中提出的“福哉马利亚（Ave Maria）”编码方法，*Polygraphia* 和 *Steganographia* 是密码学和隐藏学领域所知道的最早的两本专著，见图 1.2。（由于 Trithemius 强烈信仰超自然的力量，他的许多作品，包括 *Steganographia*，都是晦涩难懂的。他写作有关魔力和天使力量方面的故事，将巫婆分为四类，将创世时间定为公

[1] 1592 年在伦敦出版的英文版的书名为 *The Strife of Love in a Dream*。

[2] 被译作：“Brother Francesco Colonna passionately loves Polia”。Colonna 是一个修道士，当该书出版时，仍然健在。

元前 5206 年，并且解释如何借助行星天使和宗教咒语通过心灵感应来传送消息。图 1.2 于 1606 年在德国法兰克福印刷，得到奥地利维也纳国家图书馆的 H. Frodl 许可）。扩展的编码使用 40 个表，其中每个表包含 24 个用四种语言（拉丁语、德语、意大利语和法语）表示的条目（每一条对应于那个时代的字母表中的一个字母）；明文的每个字母用出现在对应表的条目中的词或短语替代，得到的密文看起来像一段祷告、一封简单的信函或一段有魔力的符咒。Schott 也阐明了如何在音乐乐谱中隐藏消息，每一个音符对应一个字母（见图 1.1）。（Gaspar Schott 简单地将字母表中的字母映射到音符。显然，不会有人去演奏这种音乐[27, p.322]。图 1.1 得到英国剑桥 Whipple 科学博物馆版权许可）。Bauer 在文献[28]中还提到另一个方法，它基于巴赫（J. S. Bach）使用的音符出现的次数。John Wilkins 展示了如何进行“两个音乐家既可以通过演奏他们手中的乐器也可以通过话音设备的对话来相互谈论他们的秘密话题”[14, XVIII, pp.143—150]。他也阐述了怎样使用点、线或角来将一个秘密的消息隐藏到一个几何图形中，“点、线的两个端点、图形的角度，它们之中每一个根据不同情形代表一个字母组合”[14, XI, pp.88—96]。

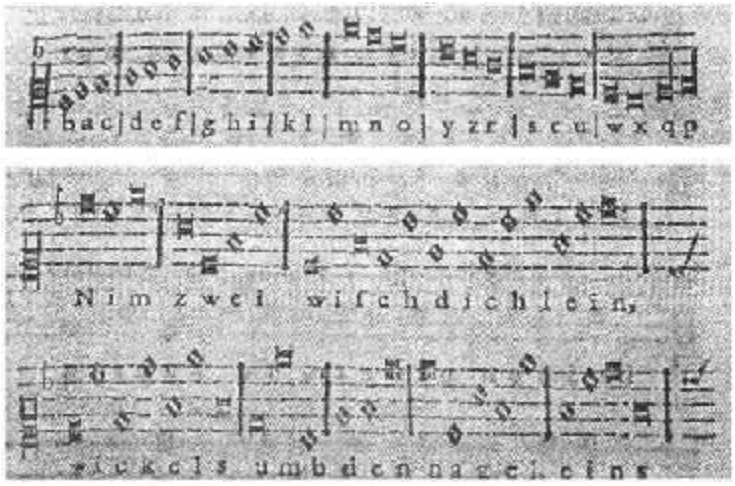

图 1.1　在乐谱中隐藏信息

消息隐藏的一个非常重要的改进就是在掩蔽正文中隐藏消息的位置可随机选择，这是许多现代信息隐藏系统的核心。在古代中国设计和使用的一个安全协议中，消息的发送者和接收者各有一张完全相同的带有许多小孔的掩蔽纸张，而这些小孔的位置是被随机选择并戳穿的，发送者将掩蔽纸张放在一张纸上，将秘密消息写在小孔位置上，移去掩蔽纸张，然后根据纸张上留下的字和空格编写一段掩饰性的文章。接收者只要把掩蔽纸张覆盖在该纸张上就可立即读出秘密消息。在 16 世纪早期，意大利数学家 Cardan（1501—1576）也发明了这种方法，该方法现在被称作卡登格子法。

利用掩蔽材料的预定位置上某些误差和风格特性是另一种选择消息隐藏位置的方法。由 Francis Bacon（1561—1626）在他的 *biliterarie* 字母表中使用的技术就是这方面的一个早期例子[29, pp.266]。目前似乎有一个争论，一些归功于莎士比亚的作品是否是由培根创作的[30]。在这种方法中，每个字母用 5 比特的二进制码来编码，并由这些二进制码来确定打印掩蔽文的字母时是使用标准体还是斜体来把消息隐藏到其中。16 世纪凸版印刷术的可变化性也可担当隐写角色。类似的技术已被用在电子出版的初期项目计划中，版权信息和序列号被隐藏在行间距和文档的其它格式特性之中[31]，据介绍有一种方法是通过对文档的各行提升或降低

1/300 英寸来表示 0 或 1，该方法能抵抗多次影印攻击，并且不会被绝大多数人察觉。

在数学表的世界里也有些这样的例子。在 17 和 18 世纪，对数表和天文历表的出版商们已经使用了在最低有效位中故意引入错误来隐藏信息[32]。今天，数据库和邮件列表提供商通过插入伪造条目来识别那些试图转售他们产品的客户。

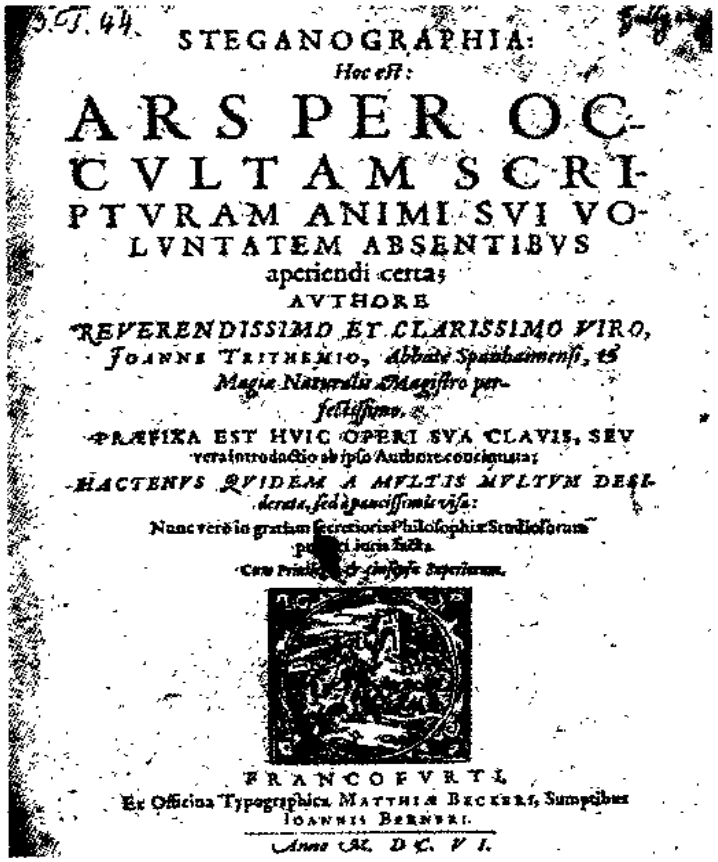
STEGANOGRAPHIA:
Hoc eſt:
ARS PER OC-
CVLTAM SCRI-
PTVRAM ANIMI SVI VO-
LVNTATEM ABSENTIBVS
aperiendi certa;
AVTHORE
REVERENDISSIMO ET CLARISSIMO VIRO,
IOANNE TRITHEMIO, Abbate Spanheimenſi, &
Magiæ Naturalis Magiſtro per-
feciſſimo, &c.
PRÆFIXA EST HVIC OPERI SVA CLAVIS, SEV
veraintroductio ab ipſo Authore conſcripta;
HACTENVS QVIDEM A MVLTIS MVLTVM DESI-
derata, ſed à pauciſſimis viſa:
Nunc verò in gratiam ſecretioris Philoſophiæ Studioſorum
publici juris facta.
FRANCOFVRTI,
Ex Officina Typographica MATTHIÆ BECKERI, Sumptibus
IOANNIS BERNERI.
Anno M. DC. VI.

图 1.2　Trithemius 的著作《Steganographia》的封面。

1.2.3　版权增强

用来对付伪造和保护版权的一个旧式解决方案的实例是 Lorraine 的 Claude Gelle'e (1600—1682)(也被称作 Claude Lorrain)所做的署名图像目录。Lorrain 是一个很有名的风景画家，他的画招致很多模仿和冒充，所以在出现相关的版权法律之前 100 年左右时[3]，他就使用了一种方法来保护他的知识产权。从 1635 年前后到 Lorrain 的临终期 1682 年，他一直保存着一本他称作为 *Liber Veritatis* 的书。这是一本写生形式的素描集，这本书是专为自己制作的，它的页面交替出现，四页蓝色后紧接着四页白色，不断重复着，它大约包含 195 幅素描。

Lorrain 的第二个传记作家 Baldinucci(1624? —1696)写到：创作 *Liber Veritatis* 的目的是为了保护 Lorrain 的画免遭伪造[4]。事实上，只要在素描和油画作品之间进行一些比较就会发现，

[3] 根据 Samuelson[33,p.16]，第一部“版权法”是“圣安妮的法令”，由英国国会于 1710 年制定。

[4] 这本传记的英译本在文献[34]中。

前者专门设计用来作为后者的"核对校验图",并且任何一个细心的观察者根据这本书仔细对照后就能判定一幅给定的油画是不是赝品。

类似的技术现在仍在使用。比如,ImageLock 系统[35]保存一个图像摘要中心数据库,并且周期性地到 Web 网上去搜寻具有相同摘要的图像。基于私钥水印(比如,[36])的跟踪系统也需要中心数据库。不幸的是,除了版权问题的范围扩大了之外(现在是全球性的),其它任何状况都没有改变,因为这种服务仍然不能提供侵犯版权的证据。第 5 章和第 7 章将进一步研究这些问题。

1.2.4 从密码学中获得的启发

虽然信息隐藏学和密码学是两个不同的学科,但我们能从密码学领域中借鉴许多技术和实践经验,以及更深层次的研究准则。1883 年,Auguste Kerckhoffs 阐明了第一个密码系统的设计准则,他在该准则中建议:我们应该假设对手知道加密数据的方法,数据的安全性必须仅依赖于密钥的选择[37][5]。从那时以后,密码学的历史仍重复出现"通过隐匿加密算法来确保安全性"的荒唐事情(假设敌手对正使用的密码系统一无所知),移动电话就是最近发生的例子之一[38]。

运用上述准则,我们可以给一个安全的隐藏系统下一个非正式的定义。一个安全的隐藏系统应该是,任何了解系统但不知道密钥的敌手不能得到任何有关已发生的通信的证据(甚至怀疑的范围)。它将遵守一个核心准则,即被广泛使用的信息隐藏程序步骤应该公开发布,就像商用的密码算法和协议那样。为了使水印技术能用于提供法律证据,更要强有力地遵守 Kerckhoffs 的准则,也就是 Anderson[39,准则 1]所说的:水印系统必须设计成在该假设下能保证经得起一个怀有敌意的专家细致的测试和考验。

所以人们可能期望版权标记系统的设计者会公开发布他们使用的系统机制和原理,并且系统的安全性仅依赖于其使用密钥的保密性。但不幸的是,事实并非如此,许多这种系统的承办商屈从于不泄露协定而不对外公布他们的系统机制,有时候以专利权作为搪塞的理由。

那些靠隐匿算法来确保安全性的系统能工作运转确实是一件很幸运的事情。今天还有许多系统仅仅是把要"隐藏"的数据嵌入到音频或视频文件的最低比特位(见 3.2 节),检测和移去这些隐藏内容对一个能干的对手来说是微不足道的。

1.3 信息隐藏的一些应用

即使所传内容已被加密,军队和情报部门仍需要隐匿通信。在现代战场上,检测到信号就可以马上对之进行攻击。正因为这个原因,军事通信使用诸如扩展频谱调制或流星散射传输的技术使信号很难被敌方检测到或阻塞掉。扩展频谱调制的基本原理将在 6.4.1 节阐述,流星爆发通信是由 Schilling 等首先研究出来的[40]。犯罪团伙也非常重视隐匿通信,他们首选的技术包括预付话费的移动电话、黑掉公用交换机使呼叫可以重选路由(比如,[41])等。作为负面影响,执法部门和反情报局很有兴趣了解这些技术和它们的缺点,这样他们就可以检测和追踪被隐藏的消息。

信息隐藏技术也是攻击军方使用的"多安全级别"系统的基础。一个病毒或其它有敌意的

[5] "Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvenient tomber entre les mains de l'ennemi"[37,p.12]

代码先从"低安全级别"到"高安全级别"传播它自己，然后使用操作系统中的隐蔽信道下载数据，或者将信息直接隐藏到那些不再列入保密范围的数据中去[42]（见2.7.2节）。

信息隐藏技术也可以使用在这样一些场合：需要有在从事某一行为时能隐匿自己身份的能力[6]。"从事某一行为时隐匿自己身份的很明显的动机是通信双方正在从事犯罪活动，并且他们希望避免被抓获"[43]，但更多的是合法的动机，包括公平选举、个人隐私、责任限额（指保险单上规定的最高赔偿金额）。提供这种特性的一种可能的机制是由 Anderson、Needham 和 Shamir 设计和开发的隐藏式的文件系统。其特点是如果一个用户知道文件的名字，他就能找回这个文件；但如果不知道文件的名字，他甚至不能得到文件存在的证据[44]。

匿名通信，包括匿名邮件重发器和 Web 代理[3]，在下列场合是非常需要的：一个合法用户在在线选举中个人投票、提出政治主张、购买性用品、保护在线自由发言以及使用电子现金等。但这些技术也同样会被滥用在诽谤、敲诈勒索以及假冒的商业购买行为上。在信息隐藏的游戏中，游戏者的伦理道德水平并不是很清楚，所以提供此项性能的技术设计需要仔细考虑可能的滥用，而有什么样的滥用可能也不是很清楚。

医疗工业尤其是医学图像系统可以受益于信息隐藏技术。它们使用诸如 DICOM（医用数字图像与通信）之类的标准，该标准中图像数据与诸如患者的姓名、日期和医师等标题说明是相互分离的。有时候患者的文字资料与图像的连接关系会丢失，所以将患者的姓名嵌入到图像数据中可能是一个很有用的安全措施[45,46]。在图像数据中作标记是否将会对病情诊断的精确性有影响仍然是一个悬而未决的问题，但 Cosman 等人最近的研究[47]显示图像的有损压缩对诊断几乎没有什么影响，就让我们相信这种隐藏方法是可行的。另一凸现出来的与医疗工业有关的技术是在 DNA 序列中隐藏信息[48]，这种技术可以用来保护医学、分子生物学、遗传学等领域的知识产权。

在多媒体应用的环境中还提出了许多信息隐藏的其它应用。很多情形下可以使用那些为版权标记所开发的技术，另一些情形下，可以使用改进的技术方案，或者使用发表在技术刊物上的一些思想。这些情形包括：

· Web 网上对授予著作权的资料的自动监控　一个自动的程序搜索 Web 网以寻找带有版权标记的资料，通过这种手段来识别可能的非法使用。另一种使用的技术是从因特网上下载图像，然后计算它们的一个摘要，并将这个摘要与注册在数据库中的摘要进行比较[35,49]。我们将在后面的 7.3.2 节再讲述这些工具。我们发现，它们带来的实际好处并不像广告所说的那么多。

· 无线传输的自动监听　一台计算机监听一座无线基站，并寻找标记信号，该标记用来表征已经广播过的某一段音乐或广告[50,51]。

· 数据附加　为了公众利益附加信息。这些信息可以是关于作品的细节、注解、其它频道[52]，或者是购买信息（最近的商店、价格、生产者等等），这样正在汽车里收听无线电的人只要简单地按一个按钮就可以订购 CD 盘。这些信息也可以是被隐藏的信息，这些信息用来对图像或音乐轨道进行索引，以便提供更有效的从数据库恢复图像和音乐文件的手段（比如，[45,53]）。

[6] 术语"plausible deniability"是由 Roe 在文献[43]中首先引入的，涉及到认可的反问题，也就是具有这样的特性：发送者不应该能错误地否认他发送过一个消息。

· 防篡改　隐藏在数字对象中的信息可以是一段签过名的数字对象的"摘要",这些隐藏的信息能用来防止或检测非授权的修改(比如,[54,55])。

信息隐藏的一些应用和技术将在下一章详细阐述。我们尽量使各章的内容简单明了,使任何计算机专业的研究生理解它们不会有太多的问题。虽然隐写术和数字水印需要一些不同学科(比如密码学、图像处理、信息论和统计学)的知识背景,但详细阐述所有那些信息隐藏技术建立其上的基础理论已经超出了本书的范围。如果你需要了解更多的背景知识,我们建议读者在密码学方面参考 Menezes 的著作[56],在图像处理方面参考 Jain 的著作[57],在信息论方面参考 Cover 的著作[58]。

参考文献

[1] Lampson, B. W., "A Note on the Confinement Problem", *Communications of the ACM*, vol. 16, no. 10, Oct. 1973, pp. 613 – 615.

[2] Gligor, V., "A Guide to Understanding Covert Channel Analysis of Trusted Systems", Technical Report NCSC – TG – 030, National Computer Security Center, Ft. George G. Meade, Maryland, USA, Nov. 1993.

[3] Chaum, D., "Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms", *Communications of the ACM*, vol. 24, no. 2, Feb. 1981, pp. 84 – 88.

[4] Goldschlag, D. M., M. G. Reed and P. F. Syverson, "Hiding routing information", in *Information Hiding: First International Workshop, Proceedings*, vol. 1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp. 137 – 150.

[5] Murray, A. H., and R. W. Burchfield (eds.), *The Oxford English Dictionary: Being a corrected re-issue*, Oxford, England: Clarendon Press, 1933.

[6] Braudaway, G. W., K. A. Magerlein, and F. Mintzer, "Protecting publicly – available images with a visible image watermark", in *Proceedings of the SPIE 2659, Optical Security and Counterfeit Deterrence Techniques*, 1996, pp. 126 – 133.

[7] "The Stockphoto mailing list," < http://stockphoto.joelday.com/ >, 1998. Discussions on photography in general and new technologies in particular.

[8] Kahn, D., *The Codebreakers——The Story of Secret Writing*, New York, New York, USA: Scribner, 1996.

[9] Petitcolas, F. A. P., R. J. Anderson, and M. G. Kuhn, "Information Hiding——A Survey," *Proceedings of the IEEE, vol. 87*, no. 7, Jul. 1999, pp. 1062 – 1078.

[10] Kobayashi, M., "Digital Watermarking: Historical Roots," Technical Report RT0199, IBM Research, Tokyo Research Laboratories, Japan, Apr. 1997.

[11] Herodotus, *The Histories*, London, England: J. M. Dent & Sons, Ltd, 1992.

[12] Newman, B., *Secrets of German Espionage*, London: Robert Hale Ltd, 1940.

[13] Tacticus, A., *How to Survive Under Siege / Aineias the Tactician*, Oxford, England: Clarendon Press, pp. 84 – 90 and 183 – 193, Clarendon ancient history series, 1990.

[14] Wilkins, J., *Mercury: or the Secret and Swift Messenger: Shewing, How a Man May With Privacy and Speed Communicate His Thoughts to a Friend at Any Distance*, London: printed for Rich

Baldwin, near the Oxford - Arms in Warnick - lane, 2nd ed., 1694.

[15] "Aliroo home page," < http://www.aliroo.com/ > . 1997. WitnesSoft and ScarLet security software.

[16] Brewster, D., "Microscope," in *Encyclopedia Britannica or the Dictionary of Arts, Sciences, and General Literature*, vol. XIV, Edinburgh, IX——Application of photography to the microscope, pp. 801 - 802, 8th ed., 1857.

[17] Hayhurst, J., "The Pigeon Post into Paris 1870 - 1871," 1970 < http://www.windowlink.com/jdhayhurst/pigeon/pigeon.html > .

[18] Tissandier, G., *Les merveilles de la photographie*, Boulevard Saint Germain, Paris, France: Librairie Hachette & Cie, VI - Les dépêches microscopiques du siège de Paris, pp. 233 - 248, Bibliothèque des merveilles, 1874.

[19] Stevens, G. W. W., Microphotography——Photography and Photofabrication at Extreme Resolutions, London: Chapman & Hall, 1968.

[20] Hoover, J.E., "The Enemy's Masterpiece of Espionage," *The Reader's Digest*, vol. 48, May 1946, pp. 49 - 53 London edition.

[21] Van Renesse, R.L. (ed.), *Proceeding of the SPIE 2659*, *Optical Security and Counterfeit Deterrence Techniques*, 1996.

[22] Van Renesse, R.L. (ed.), *Proceeding of the SPIE 3314*, *Optical Security and Counterfeit Deterrence Techniques II*, 1998.

[23] Baltrušaitis, J., *Anamorphoses ou thaumaturgus opticus*, Paris, France: Flammarion, pp. 5 and 15 - 19, Les perspectives dépravées, 1984.

[24] Seckel, A., "Your Mind's Eye: Illusions & Paradoxes of the Visual System," Lecture for the National Science Week, University of Cambridge, England, 1998.

[25] Wilkins, E.H., *A History of Italian Literature*, London: Geoffrey Cumberlege, Oxford University Press, 1954.

[26] Anonymous, *Hypnerotomachia Poliphili: the Dream Battles of Polia's Lover*, 1st ed., 1499.

[27] Schott, G., *Schola steganographica*, Jobus Hertz, printer, 1680.

[28] Bauer, F. L., *Decrypted Secrets——Methods and Maxims of Cryptology*, Berlin, Heidelberg, Germany: Spinger - Verlag, 1997.

[29] Bacon, F., *Of the Advancement and Proficiencie of Learning or the Partitions of Sciences*, Leon Lichfiekd, Oxford, for R. Young and E. Forest, vol. VI, pp. 257 - 271, 1640.

[30] Leary, P., *The Second Cryptographic Shakespeare: a Monograph Wherein the Poems and Plays Attributed to William Shakespeare are Proven to Contain the Enciphered Name of the Concealed Author*, *Francis Bacon*, Omaha, Nebraska, USA: Westchester House, 2nd ed., 1990.

[31] Brassil, J., et al., "Electronic Marking and Identification Techniques to Discourage Document Copying," in *Proceedings of INFOCOM'94*, 1994, pp. 1278 - 1287.

[32] Wagner, N. R., "Fingerprinting," in *Symposium on Security and Privacy*, Technical commitee on Security & Privacy, IEEE Computer Society, Oakland, California, USA, 25 - 27 Apr. 1983, pp. 18 - 22.

[33] Samuelson, P., "Copyright and Digital Libraries," *Communications of the ACM*, vol.38, no.4, Apr.1995, pp.15 - 21 and 110.

[34] Röthlisberger, M., *Claude Lorrain: The Paintings*, New York, New York, USA: Hacker Art Books, vol.I: Critical Catalogue, Sources——F. Baldinucci. Translation from Italian of "Notizie de' Proffessori del Disegno," Filippo Baldinucci (1624? - 1696), vol.IV, Florence 1728., pp.53 - 63, 1979.

[35] "ImageLock home page," < http://www.imagelock.com/ >, 1999.

[36] Cox, I.J., et al., "A Secure, Robust Watermark for Multimedia," in *Information Hiding: First International Workshop, Proceedings*, vol. 1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp.183 - 206.

[37] Kerckhoffs, A., "La Cryptographie Militaire," *Journal des Sciences Militaires*, vol.9, Jan.1883, pp.5 - 38.

[38] Piper, F., and M. Walker, "Cryptographic Solutions for Voice Technology and GSM," *Network Security*, Dec.1998, pp.14 - 19.

[39] Anderson, R.J., "Liability and Computer Security: Nine Principles," in *Computer Security - Third European Symposium on Research in Computer Security*, vol.875 of *Lecture Notes in Computer Science*, Springer, 1994, pp.231 - 245.

[40] Schilling, D. (ed.), *Meteor Burst Communications: Theory and Practice*, Wiley Series in telecommunications, New York: J.Wiley and Sons, 1993.

[41] Mulhall, T., "Where Have All The Hackers Gone? A Study in Motivation, Deterrence and Crime Displacement," *Computers and Security*, vol.16, no.4, 1997, pp.277 - 315.

[42] Kurak, C., and J. McHugh, "A Cautionary Note on Image Downgrading," in *Computer Security Applications Conference*, San Antonio, Texas, USA, Dec.1992, pp. 153 - 159.

[43] Roe, M., Cryptography and Evidence, Ph.D. thesis, University of Cambridge, Clare College, 18 Nov. 1997.

[44] Anderson, R.J., R.M. Needham, and A. Shamir, "The Steganographic File System," in *Proceedings of the Second International Workshop on Information Hiding*, vol.1525 of Lecture Notes in Computer Science, Springer, 1998, pp.73 - 82.

[45] Anderson, R.J., and F.A.P. Petitcolas, "On The Limits of Steganography," *IEEE Journal of Selected Areas in Communications*, vol.16, no.4, May 1998, pp.474 - 481.

[46] Hilton, D., "Matching Digital Watermarking Methods to Real Data," Computer Laboratory Seminars, University of Cambridge, 1999.

[47] Cosman, P.C., et al., "Thoracic CT Images: Effect of Lossy Image Compression on Diagnostic Accuracy," *Radiology*, vol.190, no.2, Feb.1994, pp.517 - 524.

[48] Taylor Clelland, C., V. Risca, and C. Bancroft, "Hiding Messages in DNA Microdots," *Nature*, vol.399, 10 Jun.1999, pp.533 - 534.

[49] "Digimarc home page," (http://www.digimarc.com/), 1997.

[50] Blageden, D., and N. Johnson, "Broadcast Monitoring: a Practical Application of Audio Watermarking," Announced for publication in *Proceedings of the SPIE 3657, Security and Watermark-*

ing of Multimedia Contents but Withdrawn. Presented at the conference.

[51] Willard, R., "ICE(Identification Coding, Embedded)", Preprint 3516 (D2 - 3) of the Audio Engineering Society, 1993. Presented at the 74th Convention of the AES, Berlin, 16 - 19 March, 1993.

[52] Gerzon, M.A., and P.G.Graven, "A High Rate Buried - Data Channel for Audio CD," *Journal of the Audio Engineering Sociaty*, vol.43, no.1/2, Jan. - Feb.1995, pp.3 - 22.

[53] Johnson, N.F., "In Search of the Right Image: Recognition and Tracking of Images in Image Databases, Collections, and the Internet," Technical report, George Mason University, Center for Secure Information Systems, Jun.1999.

[54] Friedman, G.L. "The Trustworthy Digital Camera: Restoring Credibity to the Photographic Image," IEEE Transactions on Consumer Electronics, vol.39, no.4, Nov, 1993, pp.905 - 910.

[55] Lin, C. - Y., and S. - F. Chang, "Issues for Authenticating MPEG Video," in *Proceedings of the SPIE 3657, Security and Watermarking of Multimedia Contents*, 1999, pp.54 - 65.

[56] Menezes, A.J., P.C. van Oorschot, and S.A.Vanstone, *Handbook of Applied Cryptography*, Boca Raton, Florida: CRC Press, 1997.

[57] Jain, A.K., *Fundamentals of Digital Image Processing*, Englewood Cliffs: Prentice - Hall, 1989.

[58] Cover, T.M., and J.A.Thomas, *Elements of Information Theory*, New York, Chichester: John Wiley & Sons, 1991.

第一部分　密写与隐写术

第二章　隐写术的基本原理

(Stefan C. Katzenbeisser)

不可视通信的“经典”模型是由 Simmons[1]作为“囚犯问题”首先提出来的。Alice[1] 和 Bob 因为某些罪行被逮捕了,并被关押在两个不同的牢房里。他们想设计一个越狱计划,但不幸的是,他们俩之间的所有通信都要接受名叫 Wendy 的看守人的监督。她不能让他们使用加密通信,并且如果一旦她注意到任何可疑的通信,她将把他们放置在单独的禁闭室里,并禁止他们交换任何信息。所以为了不引起 Wendy 的怀疑,他们双方必须进行不可视的通信,他们得建立一个阈下信道。进行此项工作的一条实用途径就是把有意义的信息隐藏在某些无关紧要的消息里。比如,Bob 可以创作一幅蓝色母牛躺在绿色的草地上的画,并把这幅现代艺术作品发送给 Alice。Wendy 不会想到画中对象的颜色传递着信息。

在本书中我们将假设(实际的监狱里也许是不可能的)Alice 和 Bob 在他们的牢房里能够接入计算机系统并且能以各种格式(比如,文本、数字图像、数字声音等等)交换信息。

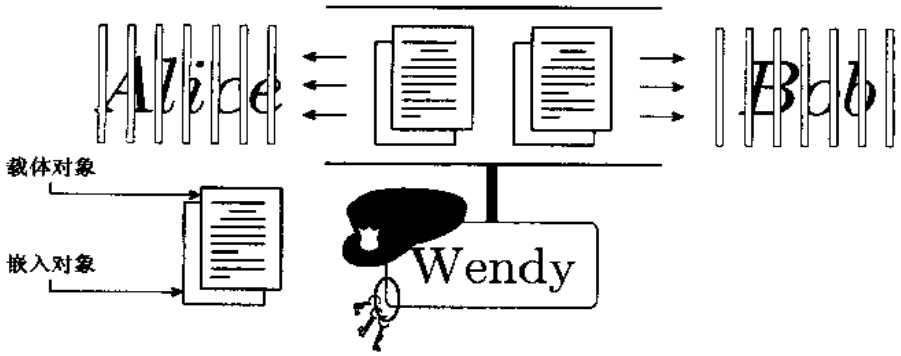

图 2.1　囚犯问题图示

不幸的是,还有另外的问题可能会阻碍 Alice 和 Bob 的逃跑。Wendy 也许会修改 Bob 传给 Alice 的消息。比如,她可能把 Bob 画的母牛的颜色改为红色,因而破坏了隐藏的信息,也就是她扮演了一个主动看守人的角色。甚至更糟糕的是,如果她扮演的是一个恶意角色的话,她还可能伪造消息并且假装自己是其中一个囚犯,通过阈下信道把消息传送给另一个囚犯。

上述模型可以应用到很多需要不可视通信(隐写术)的地方。Alice 和 Bob 代表通信的双方,想以不可见的形式交换秘密信息。监狱看守人 Wendy 代表一个窃听者,她能读取并有可能修改通信双方传送的信息(见图 2.1)。图 2.1 源自[2],得到 Scott Craver 许可。

尽管密码技术试图隐藏一个信息的内容,但信息伪装术更进了一步,它试图隐藏通信存在的事实。任何两个人都可以通过交换含有秘密信息的不保密的消息来进行偷偷摸摸的通信,但各方都要考虑通信的过程中可能存在被动的、主动的甚至恶意的攻击者。

[1] 在密码学领域,通信协议通常要涉及两个名字叫 Alice 和 Bob 的假想人物。标准约定规定在协议中各参与者按字母表顺序命名(比如,在多方协议中,Carol 和 Dave 通常紧随 Alice 和 Bob 之后),或者以这样一种方式命名:名字的第一个字母与角色的第一个字母相匹配(比如,看守人(warden)叫 Wendy)。在本章我们将遵守这个约定。

2.1　秘密通信的构架

信息伪装术的绝大多数应用遵循着一个一般原理，如图 2.2 所示。Alice 打算与 Bob 共享一段秘密信息，她先随机地(从私有的随机消息源 r 中)选取一个无关紧要的消息 c，当把它传送给 Bob 时不会引起什么怀疑，称这个消息为载体对象，然后把秘密信息 m 嵌入到 c 中，也许会用到密钥 k(称作伪装密钥)。这样 Alice 就把载体对象 c 变成一个伪装对象 s。当然，必须以非常小心的方式进行这项工作，使得任何第三方在仅知道表面上无关紧要消息 s 的情况下，不能检测秘密消息 m 的存在性。在一个“完美”的系统里，不管是通过人的感觉还是使用计算机来寻找统计模式，都不能区分正常的载体对象与伪装对象。理论上讲，载体对象可以是任何计算机可读的数据，比如图像文件、数字声音或文档等等。

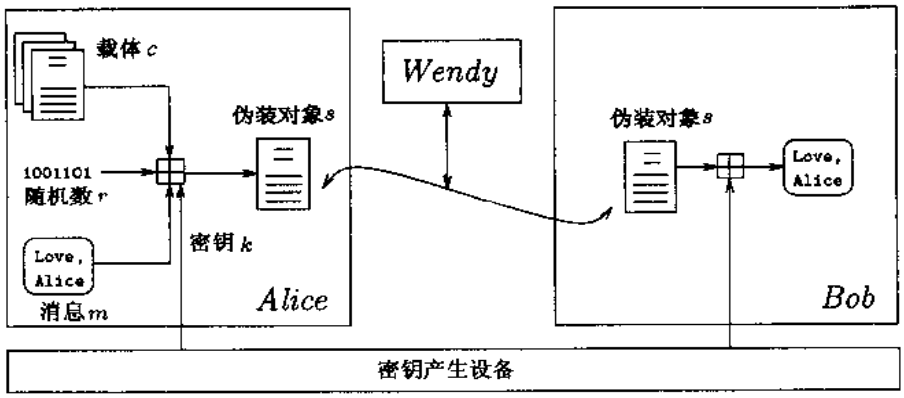

图 2.2　隐写术的原理描述。

然后，Alice 通过一个不安全的信道将 s 传送给 Bob，并且希望 Wendy 不会注意到嵌入的消息。由于 Bob 知道 Alice 使用的嵌入信息方法，也能得到嵌入过程中使用的密钥 k，所以他能重构信息 m。这个提取过程或许有可能不需要原始的载体对象 c。

监视通信过程的第三方，应该无法确定发送者在发送消息时到底是包含了秘密消息还是没有包含任何额外信息。更正式地讲就是，如果一个观察者已经获得了通信双方之间传输的一组载体对象 $\{c_1, c_2, \cdots, c_n\}$，他不可能确定哪一个载体对象 c_i 包含了秘密信息。这样，不可视通信的安全性主要取决于有没有能力将载体对象与伪装对象区分开。

然而在现实中，并不是所有的数据都能用作秘密通信的载体，因为嵌入过程中对载体的修改应该对任何不参与通信过程的一方是不可视的。这就要求载体必须含有足够多的冗余数据，这些冗余数据可以被秘密信息代替。比如，由于测量误差，任何数据(它们是某些物理扫描过程的结果)都包含一个被称作噪声的随机成分。正如我们将在后面几章里所看到的，这个随机成分可被用来作为秘密信息的掩饰。事实上，已得出共识，在绝大多数的信息伪装应用中，随机噪声数据具有很多有用的特性。

显然，一个载体不应该使用两次，因为如果攻击者能够得到一个载体的两个“版本”，那么他就很容易检测到其中的差别并可能重构信息。为了避免偶然的重复使用，发送方和接收方都应该毁掉所有在信息传输中已被使用过的载体。

在学术界，信息伪装协议基本上分为三类：无密钥信息伪装、私钥信息伪装和公钥信息伪

装。后者基于公钥密码学原理。在以下各小节中,将详细讨论这三种类型。

2.1.1 无密钥信息伪装

如果一个信息伪装系统不需要预先交换一些秘密信息(如隐藏用的密钥),我们就称之为无密钥信息伪装。在数学上,嵌入过程可描述为一个映射 $E: C \times M \to C$,这里 C 是所有可能载体的集合,M 是所有可能秘密消息的集合[2]。提取过程也看作一个映射 $D: C \to M$,从载体中提取秘密消息。很显然,必须满足 $|C| \geqslant |M|$。发送和接收双方都必须能够得到嵌入算法和提取算法,但这些算法不能对外公布。

定义 2.1(无密钥信息伪装) 对一个四元组 $\Sigma = \langle C, M, D, E \rangle$,其中 C 是所有可能载体的集合,M 是所有可能秘密消息的集合,且满足 $|C| \geqslant |M|$,$E: C \times M \to C$ 是嵌入函数,$D: C \to M$ 是提取函数,若满足性质:对所有 $m \in M$ 和 $c \in C$,恒有 $D(E(c, m)) = m$,则称该四元组为无密钥信息伪装系统。

在所有实用信息伪装系统中,集合 C 应选择为由一些有意义的但表面上无关紧要的消息所组成(像所有有意义的数字图像的集合,或者像使用第一章讨论过的 Trithemius 字母表所产生的文本等),这样通信双方在交换信息的过程中不致于引起怀疑。嵌入过程定义为这样一种方式:使载体对象和伪装对象在感觉上是相似的。从数学角度说,感觉上的相似性可通过一个相似性函数来定义:

定义 2.2(相似性函数) 设 C 是一个非空集合,一个函数 $sim: C^2 \to (-\infty, 1)$ 称为 C 上的相似性函数,若满足:对 $x, y \in C$

$$sim(x, y) = 1 \Leftrightarrow x = y$$

对 $x \neq y$, $sim(x, y) < 1$

当 C 为数字图像或数字声音的集合时,两个信号之间的互相关系可用来定义为相似性函数。所以,绝大多数实用的信息伪装系统都要努力实现这样的条件,即对所有 $m \in M$ 和 $c \in C$,都有 $sim(c, E(c, m)) \approx 1$。

以前未被使用过的载体对发送者来讲应该是保密的(也就是说,攻击者不能获得用来作秘密通信的载体)。比如,发送者可以通过使用录音或扫描技术来制作载体。对每一次通信过程,载体是随机选择的。比随机选择一个载体更好的方法是,发送者也可以浏览未被使用过的载体数据库,并从中选择一个,使得嵌入过程对它的修改最少。这样的选择过程可以通过相似性函数 sim 来进行。在编码阶段,发送者选择一个载体 c 使之满足性质:

$$c = \max_{x \in C} sim(x, E(x, m)) \tag{2.1}$$

如果载体是一些扫描处理的结果,则原始载体可以反复地数字化。由于硬件里的噪声,每一次处理将产生稍微不同的载体。发送者选择其中最适合通信的一个,这种技术被称作不可视性的选择方法,它在文献[3]中有详细的讨论。

一些研究者提出公共载体数据库的思路。由于一个得到原始载体的攻击者很容易检测到秘密消息,所以发送者从数据库中选择一个元素 c,并对 c 进行一些修改得到一个载体 c',然

[2] 更一般地讲,嵌入过程可被看作集合 $C \times M$ 与集合 C 之间的一个关系(也就是,$E \subset C \times M \times C$),满足:对任意两个元素对 (c_1, m_1) 和 $(c_2, m_2) \in C \times M$, $m_1 \neq m_2$, $E(c_1, m_1) \cap E(c_2, m_2) = \phi$。然而,为记号简明起见,在本章中,我们假设这个关系是一个函数关系。

后他使用这个新载体进行秘密通信。然而，这种方法并不是没有危险性。如果一个攻击者知道了使用的修改技术，他就可以亲自制作"空白"的载体（也就是不含秘密信息的载体），并破坏掉秘密通信。即使他不知道所运用的技术，他也能通过将 c 与伪装对象的比较来创作一个相似的载体。

某些隐藏方法将传统密码学与信息伪装术结合到一起，即发送者在嵌入处理之前对秘密消息进行加密处理。很显然，这种结合方式增加了整个通信过程的安全性，因为攻击者很难检测到嵌入在载体中的密文（密文外表本身具有相当的随机特性）。然而，强健的信息伪装系统并不需要预先加密处理。

2.1.2　私钥信息伪装

无密钥信息伪装不需要什么信息（除了函数 E 和 D）来启动通信过程，这样系统的安全性就完全依赖于它自己的保密性。因为这违反了 Kerckhoffs 的准则（见 1.2.4 小节），在现实中是很不安全的。所以我们必须假定 Wendy 知道 Alice 和 Bob 用于信息传输的算法。理论上讲，她能够从 Alice 和 Bob 之间发送的每一个载体对象中提取秘密信息。一个信息伪装系统的安全性应该仅依赖于由 Alice 和 Bob 协议的某些秘密信息，也就是伪装密钥。不知道这个密钥，任何人都不能从伪装对象中提取秘密信息。

一个私钥伪装系统类似于私钥密码，发送者选择一个载体 c 并使用密钥 k 将秘密信息嵌入到 c 中。如果嵌入过程中使用的密钥对接收者来说是已知的，则他就可逆向操作这个过程并提取出秘密信息，而不知道这个密钥的任何人都不可能得到被隐藏信息的证据。另外，载体 c 和伪装对象之间感觉上是相似的。

定义 2.3（私钥信息伪装）　对一个五元组 $\Sigma=\langle C, M, K, D_K, E_K\rangle$，其中 C 是所有可能载体的集合，M 是所有可能秘密消息的集合，且满足 $|C|\geqslant|M|$，K 是所有可能密钥的集合，E_K：$C\times M\times K\rightarrow C$ 是嵌入函数，D_K：$C\times K\rightarrow M$ 是提取函数，若满足性质：对所有 $m\in M$，$c\in C$ 和 $k\in K$，恒有 $D_K(E_K(c,m,k),k)=m$，则称该五元组为私钥信息伪装系统。

私钥伪装系统需要某些密钥的交换，显然这种额外秘密信息的传输打乱了不可视通信的原始意图。只要涉及到密码学，我们总是假定所有的通信各方都能够通过一个安全的信道来协商密钥。Alice 和 Bob 在入狱前就可以协商好一个伪装密钥。然而，通过利用载体的某些内在特征和一个安全哈希函数 H，完全可以直接从载体计算出一个用于秘密通信的密钥：$k=H(feature)$。如果嵌入处理不改变载体的"内在特征"，接收者就能够重新计算出密钥 k。显然，这样一种内在特征必须具有高度的"载体依赖性"，以达到一个适度的安全水平（然而，这种安全性依赖于 H 的保密性，所以又破坏了 Kerckhoffs 准则）。如果伪装是一幅数字图像，则可以取载体的所有颜色值的最高位比特作为一个"内在特征"。如果将哈希函数 H 依赖于密钥 k'，而 k' 是一个较长时期有效的密钥，则也可以利用上述方法从密钥 k' 中计算出一个秘密会话密钥。

在隐藏信息提取过程中，有些算法还需要额外了解原始载体的某些信息（或者说是从伪装对象中得不到的其它信息）。这种系统的用处很有限，因为使用它们还要求传送原始载体，这是与传统密码学中密钥交换紧密相关的一个问题。这些算法可看作是私钥信息伪装系统的一个特例，其中 $K=C$ 或 $K=C\times K'$，这里 K' 表示密钥的附加集合。

2.1.3 公钥信息伪装

就像公钥密码系统一样，公钥信息伪装系统不依赖于密钥的交换。公钥信息伪装系统需要使用两个密钥：一个私用钥和一个公用钥。公钥存储在一个公开数据库中，并且公钥用于信息嵌入过程，而私钥用于重构秘密信息。

建立公钥信息伪装系统的一种方式是使用公钥密码系统。我们假设 Alice 和 Bob 在入狱前已经交换好某些公钥算法的公钥（这也是一个比较合理的假设）。公钥信息伪装利用这样一个客观事实，即伪装系统里的解码函数 D 能适用于任何载体 c，而不管它是否已经包含秘密信息（前面已经提到 D 是一个作用于整个集合 C 的函数）。在没有隐藏信息的情形下，解码的结果会是秘密消息集合 M 的一个随机元素，我们称之为载体的“自然随机性”。如果这种自然随机性与某些公钥密码系统产生的密文是统计上不可区分的，就可以通过嵌入密文而不是未加密的秘密消息来建立一个安全的信息伪装系统。

一个使用公钥信息伪装的协议已由 Anderson 在[4,5]中提出，它依赖于这样一个事实，即经加密的消息具有足够的随机性以致于可以“躲过明亮的眼睛”。Alice 用 Bob 的公钥加密消息，得到一个“外观随机”的消息，并将它嵌入到 Bob 知道的信道（Wendy 也可能知道）中去，从而替代了每个通信过程所伴随的某些“自然随机性”。我们假设加密算法和嵌入函数是大家都已知的。Bob 事先并不能决定秘密信息是否经由一个特定的载体传输过来，只是猜疑可能有秘密信息到来，并试图用私钥去提取和解密。如果载体确实含有秘密信息，则解密出来的信息就是 Alice 的消息。

由于我们假定 Wendy 已经知道使用的嵌入方法，她也可以试图去提取由 Alice 传给 Bob 的秘密消息。然而，如果加密方法产生外观随机的密文，则 Wendy 将没有证据表明提取出来的信息是否只是一些随机的比特。这样，她就无法确定提取出来的信息是有意义的还是自然随机性的一部分，除非她能攻破这个密码系统。

一个至关重要的方面就是 Bob 必须时刻猜疑伪装技术的使用，并试图对他从 Alice 接收到的每一个载体进行解码（他也许并不怎么了解 Alice 的个性）。如果伪装对象不是专门发给一个特定接收者，而是发送到一个因特网新闻组，则事情会变得更糟糕。虽然这个协议在这种情形也可以工作（只有特定的接收者才能解密出秘密信息，因为只有他才有正确的私钥），但所有可能的接收者都得试图对每一个接收到的对象进行解码。

Craver[2]使用私钥信息伪装和公钥信息伪装对这个协议进行扩展来模拟一个无密钥信息伪装。当攻击者知道嵌入方法时，一个无密钥信息伪装协议不能提供任何安全性，但在绝大多数应用中，无密钥信息伪装仍是首选，这是因为通信双方不需要共享一个伪装密钥。通过使用公钥信息伪装系统执行一个密钥交换协议，Alice 和 Bob 可以共享一个秘钥 k，稍后他们可以在私钥信息伪装系统中使用这个密钥 k。由于不需要事先知道任何伪装密钥（除了他们的公用加密密钥），我们可以将这个通信过程看作是无密钥信息伪装，虽然它并不符合定义 2.1。

在这个协议中，Alice 首先生成一个随机的公用/私有密钥对，同对应公钥密码系统一起使用，然后她把公钥放置在 Bob 知道并看得见的信道上（并且 Wendy 也知道和看得见）。Wendy 和 Bob 都不能确定该信道是否只是包含一些随机比特，还是含有某些信息。然而，Bob 猜疑 Alice 发送的伪装对象可能包含 Alice 的公钥，并且试图提取它。他使用提取出的公钥将一个随机选择的密钥 k 以及一个确认的短消息嵌入到一个载体中，当然密钥 k 和短消息都要事先

用 Alice 的公钥进行加密，并发送给 Alice。另外，Wendy 也可能试图提取由 Bob 发送的秘密信息，但她能够看到的只是随机模样的密文。Alice 猜疑从 Bob 来的消息可能隐藏着信息，于是她从中提取秘密信息并用她的私钥对之解密。现在 Alice 和 Bob 共享了一个隐藏密钥 k。该协议如图 2.3 所示。图 2.3 源自[2]，得到 Scott Craver 许可。

然而，这个协议（在第一步）对中间插入的攻击是很脆弱的。如果 Wendy 是主动的，她能够捕获第一个从 Alice 发给 Bob 的伪装对象，并用她自己的公钥去替代 Alice 的公钥。于是 Bob 使用 Wendy 的公钥而不是 Alice 的公钥对随机密钥 k 加密。现在 Wendy 知道了 Bob 选择的密钥 k，并接着将 k 传送给 Alice：她用 Alice 的公钥对它加密，将它嵌入一个载体中并把结果发送给 Alice。虽然 Alice 正确地接收到 k，但她没有意识到 Wendy 也已经获得了 k。

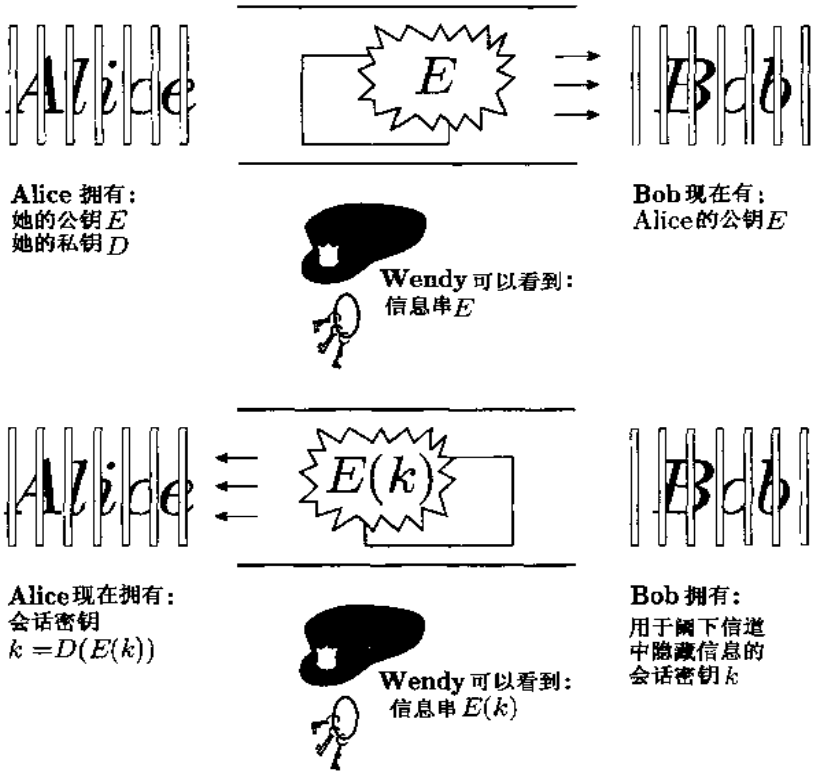

图 2.3　一个信息伪装密钥交换协议的图示

据推测，公钥信息伪装系统和无密钥信息伪装系统都不太可能运用于存在恶意看守者的情形中。Wendy 可以以 Alice 的名义启动上面给出的公钥信息伪装协议或其扩展协议来愚弄 Bob。因为 Bob 无法验证在协议第一步中发送的公钥的有效性，他也不能识别从 Alice 处收到的密钥。这种情形类似于公钥密码系统，需要有公钥证书。在无密钥信息伪装系统里，Bob 不能区分 Alice 和 Wendy 发送的消息。

2.2　隐写系统的安全性

虽然攻破一个信息伪装系统的工作由检测隐藏的信息、提取隐藏的信息和破坏掉隐藏的

信息三部分组成(见第四章),但是,如果一个攻击者能够证明一个秘密信息的存在性,则该系统就已经不安全了。在开发一个正式的信息伪装安全模型中,我们应该假设攻击者具有无限的计算能力,并且能够也乐于对系统进行各种类型的攻击。如果攻击者仍然不能确认他的假设——"一个秘密消息嵌入在一个载体中"——是否正确,则该系统是理论性安全的。

2.2.1 绝对安全性

Cachin 在[6]中,从信息论的角度,给出了信息伪装系统安全性的一个正式定义。其主要思想涉及到载体的选择,而载体被看作是一个具有概率分布为 P_C 的随机变量 C,秘密消息的嵌入过程看作是一个定义在 C 上的函数。设 P_S 是 $E_K(c,m,k)$ 的概率分布,其中 $E_K(c,m,k)$ 是由信息伪装系统产生的所有的伪装对象的集合。

如果一个载体 c 根本不用作伪装对象,则 $P_S(c)=0$。为了计算 P_S,必须给出集合 K 和 M 上的概率分布。利用定义在集合 Q 上的两个分布 P_1 和 P_2 之间的条件熵 $D(P_1 \| P_2)$ 的定义:

$$D(P_1 \| P_2) = \sum_{q \in Q} P_1(q) \log_2 \frac{P_1(q)}{P_2(q)} \tag{2.2}$$

这个条件熵用来度量当真实概率分布为 P_1 而假设概率分布为 P_2 时的无效性,它可以度量嵌入过程对概率分布 P_C 的影响。特别地,我们根据 $D(P_C \| P_S)$ 来定义一个信息伪装系统的安全性。

定义 2.4(绝对安全性) 设 Σ 是一个信息伪装系统,P_S 是通过信道发送的伪装对象的概率分布,P_C 是 C 的概率分布,若有:$D(P_C \| P_S) \leqslant \varepsilon$,则称 Σ 抵御被动攻击是 ε—安全的。若有 $\varepsilon = 0$,则称 Σ 是绝对安全的。

因为当且仅当两个概率分布相等时 $D(P_C \| P_S)$ 等于 0,于是我们可以得出结论:如果一个信息伪装系统嵌入一个秘密消息到载体中去的过程不改变 C 的概率分布,则该系统是(理论上)绝对安全的。绝对安全的系统可以利用以前的填充法来构造。

定理 2.1 存在绝对安全的信息伪装系统。

证明:我们给出一个构造性证明。设 C 是所有长度为 n 的比特串的集合,P_C 是 C 上的均匀分布,e 是秘密消息($e \in C$)。发送者随机选择一个载体 $c \in C$,并计算 $s = c \oplus e$,这里 $\oplus$ 是比特异或运算 XOR。这样产生的伪装对象 s 在 C 上也是均匀分布的,因此 $P_C = P_S$,并且 $D(P_C \| P_S) = 0$。在信息提取过程中,通过计算 $s \oplus c$,就可恢复出秘密消息 e。

上述系统非常简单但没有什么用处,因为没有监狱会让 Alice 和 Bob 去交换随机比特串。

2.2.2 检测秘密消息

一个被动的攻击者(Wendy)不得不判决从 Bob 发往 Alice 的载体 c 是否含有秘密消息。这项任务可公式化为一个统计假设检验问题。所以 Wendy 定义一个检验函数 $f: C \to \{0,1\}$:

$$f(c) = \begin{cases} 1 & c \text{ 中含有秘密消息} \\ 0 & \text{其他} \end{cases} \tag{2.3}$$

当载体在不安全的信道上经过时,Wendy 利用这个函数对它们进行归类。有时候 Wendy 能够准确地将载体归类,有时候她却检测不到隐藏的消息,犯第二类弃真错误。也有可能出现这样的情况,Wendy 从没有包含秘密信息的载体中错误地检测出一个消息,这时候,她犯的是第一

类纳伪错误。实用的信息伪装系统应该尽量使被动攻击者犯第二类错误的概率 β 最大。一个理想的信息伪装系统应该有 $\beta=1$。在下面几段中，我们将指出具有绝对安全性的信息伪装系统(假设攻击者犯第一类错误的概率为 0)具有这样的属性。

对 ε—安全的信息伪装系统，一个被动攻击者所犯的第一类错误的概率 α 和第二类错误的概率 β 通过下列定理联系起来。

定理 2.2(Cachin,1998)

设 $\sum$ 是一个对付被动攻击者为 ε—安全的信息伪装系统，则攻击者检测不到隐藏信息的概率 β 和攻击者错误地检测出一个不是隐藏信息的概率 α 满足关系式：$d(\alpha,\beta)\leqslant\varepsilon$，其中 $d(\alpha,\beta)$ 是按下式定义

$$d(\alpha,\beta)=\alpha\log_2\frac{\alpha}{1-\beta}+(1-\alpha)\log_2\frac{1-\alpha}{\beta}$$

的二元关系熵。

特别地，若 $\alpha=0$，则 $\beta\geqslant 2^{-\varepsilon}$。

为了证明定理 2.2，我们需要条件熵函数的一个特殊性质，即确定型处理不会增加两个概率分布之间的熵。假设 Q_0 和 Q_1 是两个定义在集合 Q 上的随机变量，其概率分布分别为 P_{Q_0} 和 P_{Q_1}，函数 $f\colon Q\to T$，则 $D(P_{T_0}\parallel P_{T_1})\leqslant D(P_{Q_0}\parallel P_{Q_1})$，其中 P_{T_0} 和 P_{T_1} 分别表示 $f(Q_0)$ 和 $f(Q_1)$ 的概率分布。见文献[6]。

证明(定理 2.2)：当载体不包含秘密消息时，所有载体是按 P_C 分布的。我们考虑随机变量 $f(c)$，并计算它的概率分布 π_C。当 $f(c)=1$ 时，攻击者会犯第二类错误，这样 $\pi_C(1)=\alpha$ 和 $\pi_C(0)=1-\alpha$。如果载体中含有秘密消息(注：原文中此处有误)，则载体按 P_S 分布。同样，我们计算 $f(S)$ 的分布 π_S。当 $f(s)=0$，攻击者会犯第一类错误，因为他检测不出隐藏的消息，这样：$\pi_S(0)=\beta$ 和 $\pi_S(1)=1-\beta$。相关熵 $D(\pi_C\parallel\pi_S)$ 可以表示为：

$$\begin{aligned}D(\pi_C\parallel\pi_S)&=\sum_{q\in\{0,1\}}\pi_C(q)\log_2\frac{\pi_C(q)}{\pi_S(q)}\\&=(1-\alpha)\log_2\frac{1-\alpha}{\beta}+\alpha\log_2\frac{\alpha}{1-\beta}\\&=d(\alpha,\beta)\end{aligned}$$

使用上述结论可得：$d(\alpha,\beta)=D(\pi_C\parallel\pi_S)\leqslant D(P_C\parallel P_S)\leqslant\varepsilon$。由于 $\lim_{\alpha\to 0}\alpha\log_2(\alpha/(1-\beta))=0$(利用 De Hospital 法则)，于是，$d(0,\beta)=\log_2(1/\beta)$。因此当 $\alpha=0$ 时，$\beta\geqslant 2^{-\varepsilon}$。

因此，对 $\alpha=0$ 的 ε—安全的信息伪装系统，我们可以得出结论：若 $\varepsilon\to 0$，则概率 $\beta\to 1$；如果 ε 很小，则攻击者不能够以很高的概率检测出隐藏的信息。

2.3　在噪声数据中隐藏信息

正如我们在第 2.1 小节所提到的那样，隐写术是充分利用通信过程中冗余信息的存在性来工作的。图像或数字声音天然地包含噪声形式的冗余。在本节中，不失一般性，我们假设载体 c 可以用二进制数字序列来表示。对于数字声音，这个序列恰好是按时间抽样的序列；对数字图像，这个序列可以通过对图像矢量化来得到(也就是，以从左到右，从上到下的顺序排列像素点的灰度值或颜色值)。设 $l(c)$ 为序列中元素的个数，m 为秘密消息，$l(m)$ 为它的比特长

度。

绝大多数的信息伪装方法所遵循的一般原理是把秘密信息放置在信号的噪声成分中。如果能够对秘密信息进行某种方式的编码，使得它与真正的随机噪声不可区分，则攻击者将没有机会检测到秘密通信。

在二进制序列中隐藏信息的最简单的方法，是将载体 c 中每个元素的最低位比特(LSB)用秘密消息 m 的一个比特去替代。在浮点数算术中，尾数的最低位比特也可用来隐藏信息。由于被隐藏消息的大小一般要比可用来隐藏信息的比特空间数量小得多($l(m) \ll l(c)$)，隐藏完消息后则剩下的 LSB 保持不变。由于改变一个字节(或一个字)的 LSB 仅仅是一个微量的增加或减少，发送者可以假设这种差异完全落在噪声的范围内，所以这种差异一般也就不会引起注意。显然，这种技术不能提供高水平的安全性。攻击者可以简简单单地尝试着对载体"解码"，就像他是一个合法的接收者。另外，即使消息是由真正的随机比特组成，这种算法也会明显地改变载体的统计特性。

这种技术是可以改进的。我们并不把载体的每一个元素用于信息传输，而是可以根据一个密钥以随机方式选择某些元素来传输信息，剩下的元素保持不变。这种选择也可以通过使用伪随机数生成器来进行。文献[7]报道了这样一个系统，随机数生成器的输出结果将消息比特序列扩展到载体中，其扩展的原则是，相邻两个用于消息传输的元素之间的位置距离是由这个随机数决定的。我们将在第三章详细介绍和讨论这些方法。

最后提到的这个方法也非常适合流式载体。所谓流式载体就是发送者在嵌入过程中不能获取整个序列的载体。比如这样一个应用，数字声音文件在录制过程中将秘密消息隐藏其中。另一方面，如果发送者在嵌入过程中可以获取整个载体序列，这样的载体称之为随机获取载体。

Aura[3]引入了一种适用于随机获取载体的灵活性很好的方案，尤其适用于数字图像。他开发了一个基于伪随机置换的私钥信息隐藏系统。根据该方案构造的算法，秘密消息以相当随机的方式分布在整个载体中，更详细的内容请参考第 3.2.2 小节。

2.4 自适应与非自适应算法

上一节描述的所有方法都采用了一个共同的准则，即用秘密消息替代载体的不重要的部分(比如，噪声成分)，虽然这些部分有些特殊的统计特性，而嵌入过程并没有关注它们，所以隐藏消息后会引起载体统计轮廓明显的改变。一个被动的攻击者可以充分利用这个客观事实来攻破该系统，其详细讨论见第四章。正如定义 2.4 所指出的那样，为确保安全性，保持载体统计分布特性不变是至关重要的。

2.4.1 拉普拉斯滤波

大众软件 PGMStealth 通过简单地在每个像素的 LSB 存储秘密信息的一个比特来将秘密信息隐藏在灰度图像中。通过运用离散拉普拉斯算子，有可能检测出隐藏在灰度图像中的秘密的 PGMStealth 消息(一个图像 p 用一个(x, y)—矩阵来表示)：

$$\nabla^2 p(x,y) = p(x+1,y) + p(x-1,y) + p(x,y+1) + p(x,y-1) - 4p(x,y) \tag{2.4}$$

在每一点(x , y)处计算(2.4)式的值,就得到"拉普拉斯滤波"的图像。由于我们可以认为相邻的像素点具有相近的颜色值,所以$\nabla^2 p(x,y)$的柱状图主要聚集在0附近。图2.4(a)显示了印在同一坐标系统里八个拉普拉斯滤波灰度级图像的柱状图,图2.4(b)显示了经软件PGMStealth系统隐藏信息后同一幅图像的柱状图。因为嵌入过程向图像增加了噪声,而这些噪声的统计特性又与真正的随机噪声完全不同,所以新的柱状图与旧的柱状图差别特别大。拉普拉斯滤波虽然不能证明秘密消息的存在性,但它提供了很强的证据证明该幅图像被修改过。

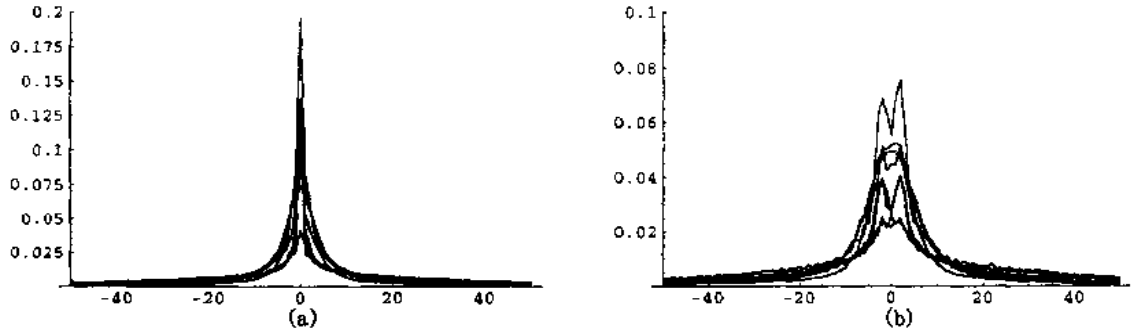

图2.4　(a)八个拉普拉斯滤波后的图像的柱状图;
(b)经PGMStealth系统隐藏信息后同一幅图像经拉普拉斯滤波后的柱状图

2.4.2　使用载体模型

为避免上述的攻击,一些研究者提出了对载体特征模型化,然后产生一个适合于这个模型的(称为适应性)信息伪装算法,但这不是一个容易达到的目标。对秘密信息以某种方式进行编码,使得编码结果与真正随机噪声在统计上不可区分是完全有可能的,然后秘密信息可以放置在图像的高噪声区域。一般来讲,这种方法需要对传输秘密信息的载体的精确模型非常了解。然而,建立这样一种模型并不容易,而且整个策略具有内在的危险性,因为对一个具有很多资源并且愿意在该问题上投入很多时间的攻击者来说,他可以演绎出一个更好的模型,并利用该模型去检测隐藏在载体中的模式,而该模式是发送者使用不怎么复杂的模型建立的。

文献[8]提出了一个依赖噪声模型的嵌入处理算法,它使用数字图像作为载体。在这里我们对该算法进行简单的介绍。设f是噪声成分的概率密度函数(关于y轴对称),η_j是一个服从由f定义的概率分布,且是相互独立的随机变量序列。发送者首先对一个载体的若干个版本求平均,生成一个"零噪声"的图像,然后选择$l(m)$个随机像素用于信息传输,一个像素隐藏一个信息比特。秘密信息的嵌入方式是根据要嵌入的消息比特值,对被选择的像素点的值增加$|\eta_j|$或$-|\eta_j|$,而其余的像素点的值则按η_j进行修改,这些修改应该保持与统计噪声模型f相一致。

使用这种方法,嵌入过程变成参数化的过程,因此在大多数情况下,提取过程也是参数化的过程,并且需要原始载体的知识。

2.5　主动与恶意的攻击者

在设计信息伪装系统时,要特别关注那些主动的和恶意的攻击者。主动攻击者可以在通

信过程中改变一个载体，Wendy 可以捕获一个从 Alice 发往 Bob 的伪装对象，然后修改它并将修改结果发送给 Bob。一般假设一个主动攻击者不能够完全改变载体及其语法含义，但可以对它们进行较小的修改，以使得原来的伪装对象与修改过的伪装对象保持感觉上或语义上的相似性。如果一个攻击者能够伪造消息或者以一个通信方的名义启动和运行信息伪装协议，则该攻击者是一个恶意的攻击者。关于对信息伪装系统的攻击请参看第 4 章和第 7 章。

2.5.1 主动攻击者——健壮的信息伪装

信息伪装系统对载体的修改特别敏感，比如，对数字图像的图像处理技术(像平滑、滤波以及图像变换)和对数字声音的滤波，甚至一个有损压缩都可能导致整个隐藏信息丢失。有损压缩技术试图通过去掉觉察不到的信号成分来减少信息数量，因此也会去掉曾嵌入在信号中的秘密信息。

一个主动攻击者虽然不能从载体中提取秘密消息或者证明秘密消息的存在性，但他可以简单地向传输的载体中加入随机噪声，进而企图破坏掉隐藏的信息。对于数字图像，攻击者也可以采用图像处理技术或将图像转换成另一种图像格式。所有这些技术都对秘密通信产生很大的危害。所以信息伪装系统的另一个实际需求就是健壮性。如果一个信息伪装系统满足以下条件，若不对伪装对象作剧烈的改变，嵌入的信息不可能被改变，则这样的系统被称为健壮性系统。

定义 2.5(健壮性) 设 $\sum$ 是一个信息伪装系统，P 是 $C \to C$ 的一类映射，若对所有的 $p \in P$，

(i) 对私钥信息伪装系统，恒有：

$$D_K(p(E_K(c,m,k)),k) = D_K(E_K(c,m,k),k) = m \tag{2.5}$$

(ii) 对无密钥信息伪装系统，恒有：

$$\mathrm{D(p(E(c,m)))} = \mathrm{D(E(c,m))} = \mathrm{m} \tag{2.6}$$

而不管如何选择 $\mathrm{m}\in\mathrm{M}$、$\mathrm{c}\in\mathrm{C}$、$\mathrm{k}\in\mathrm{K}$，则称该系统为 P—健壮性的信息伪装系统。

显然要把握好安全性和健壮性之间的平衡。一个系统抵御载体修改的健壮性越强，则系统的安全性就越低。因为健壮性只有通过冗余信息编码才能获得，而这又会严重地降低载体的质量，并且有可能改变概率分布 P_S。

许多信息伪装系统被设计成对某一类特殊的映射(比如，JPEG 压缩与解压缩[9,10]、滤波、加入白噪声等等)具有健壮性。一个理想的系统应该对所有的“保持 α—相似性”的映射具有健壮性，(也就是，映射 $p: C \to C$ 具有性质：$sim(c, p(c)) \geqslant \alpha$ 且 $\alpha \approx 1$)。然而，这样一种系统在工程实现中相当困难，而且由于编码嵌入的健壮性将导致很低的带宽。另一方面，一个系统称之为 α—弱的，如果对每一个载体都存在一个“保持 α—相似性”的映射，使得隐藏的信息不能按公式(2.5)和(2.6)的方式恢复出来。

一般地，有两种方法可以使得信息伪装系统具有健壮性。第一种方法是预先了解所有可能的载体修改方式，以使嵌入过程本身就具有健壮性，因而使修改不能彻底地破坏秘密信息[11,12]。第二种方法是设法对攻击者在载体上所作的修改进行逆处理，得到原来的伪装对象。Bender 等人在文献[13]提出了“仿射编码”——一种对付仿射图像变换的对策，以恢复嵌入在伪装对象中的秘密信息。他们通过度量某些编码嵌入的参考模式在形状、大小、方向上的改变，来试图估计变换的参数。由于变换是线性的，所以可以通过逆变换来重构伪装载体。

Cox 等人在文献[14]中指出：健壮性算法必须把信息放置在信号的感观上最重要的部件上，因为信息隐藏在噪声部件里，可以不费吹灰之力地就把它去掉。比如，我们知道在伪装信号的某些变换域里运用的嵌入规则，要比在时间域使用的嵌入算法对修改具有更好的健壮性。我们将充分利用这些特性并构造出健壮性的信息伪装系统，该系统将秘密信息隐藏在图像的离散余弦变换的系数中，见 3.3.1 小节。实验研究表明，这些方法对质量下降到约 60% 的 JPEG 压缩都是健壮的。其它一些方法，主要用于数字水印，即使采用质量下降到 5% 的 JPEG 压缩，隐藏的信息也能保存下来。

2.5.2　阈上信道

如果我们假定一个主动攻击者仅仅能对伪装对象作些较小的修改，并且每个载体都含有某种感觉上最重要的信息，如果不对载体作彻底的语义上的改变，这些重要信息是不可能被去掉的。通过对秘密消息进行编码使它构成感观上最重要的部分，那么信息就能以很高的完整性在通信双方之间传输。Craver[2]称这样一个信道为阈上信道："信息隐藏在视觉明晰处，事实上很显然，如果不对传输的对象作总体的修改，就不可能修改隐藏的信息"。

用于秘密通信的载体可以通过一幅伪装地图描述，伪装地图是载体感观上最重要的部件的正式描述。设 S 是所有伪装地图的集合，一个函数 $f: S \to \{0,1\}^N$，称为伪装地图函数。为了将一个比特串 $x \in \{0,1\}^N$ 嵌入到一个阈上信道中，Alice 选择一个元素 $s \in f^{-1}(x)$，并在不安全的信道上发送一个符合伪装地图 s 的载体。Wendy 可能怀疑使用阈下信道并稍稍改变载体以企图移去嵌入在噪声部件中的秘密消息，但不能够改变这个伪装地图。Bob 从他接收到的载体中重构伪装地图 s 并利用函数 f 来恢复 x，见图 2.5。图 2.5 源自[2]，得到 Scott Craver 许可。

为了达到实用的目的，必须可以根据任何一个伪装地图构造出一个载体，使得如果一个主动的监狱看守对这个载体作微小的修改，则嵌入的比特串不会改变，(也就是，Wendy 不能将传送的载体的伪装地图 s 转变为 s' 使得 $f(s) \neq f(s')$)。进而，也必须能从可用于秘密通信的每一个载体中精确地导出伪装地图。函数 f 应该是公开的，f 与 f^{-1} 计算上应该是容易的。不包含秘密信息的载体应该映射到一个随机串，于是，一个伪装对象与一个无关紧要的载体的唯一区别就在于比特串 $f(s)$ 是否具有含义。

为了利用阈上信道进行伪装通信，Alice 不能显式地隐藏一个有明显意义的比特串(像"这是监狱暴动计划。亲爱的，Alice")，因为 Wendy 可以简单地利用公开函数 f 恢复出嵌入的信息。在这种情况下，Alice 和 Bob 将马上面临被单独禁闭。然而，如果嵌入的消息是一个随机的密钥或者表面上看起来很随机的密文，那么 Wendy 将不会有证据来确定传输的信息是一个随机比特串还是有更多的内容(除非她能够攻破该密码系统)，因而不会引起怀疑。

一个阈上信道可以用来执行一个伪装的密钥交换协议。Alice 生成一个公钥 E 和私钥 D，选择一个伪装地图 $s \in f^{-1}(E)$，发送符合 s 的载体给 Bob。Bob 通过计算伪装地图和利用 f 来恢复 E。他选择一个随机密钥 k，使用 E 对它加密，选择一个符合伪装地图 $s' \in f^{-1}(E(k))$ 的载体，并送回给 Alice，她按类似的方式恢复 $E(k)$，然后用她的私钥 D 将 $E(k)$ 解密。无论如何，Wendy 找不到证据表明嵌入在阈上信道的消息是有意义的，同时 Alice 和 Bob 完成了一次密钥交换协议。虽然该协议对中间插入攻击是很脆弱的(就像 2.1.3 小节所描述的那样)，但一个没有恶意的监狱看守不能够从 Alice 和 Bob 之间传送的信息中推断出 k。

除了由于编码的健壮性导致的低带宽外，这个系统的主要不足之处就是它的可行性。并

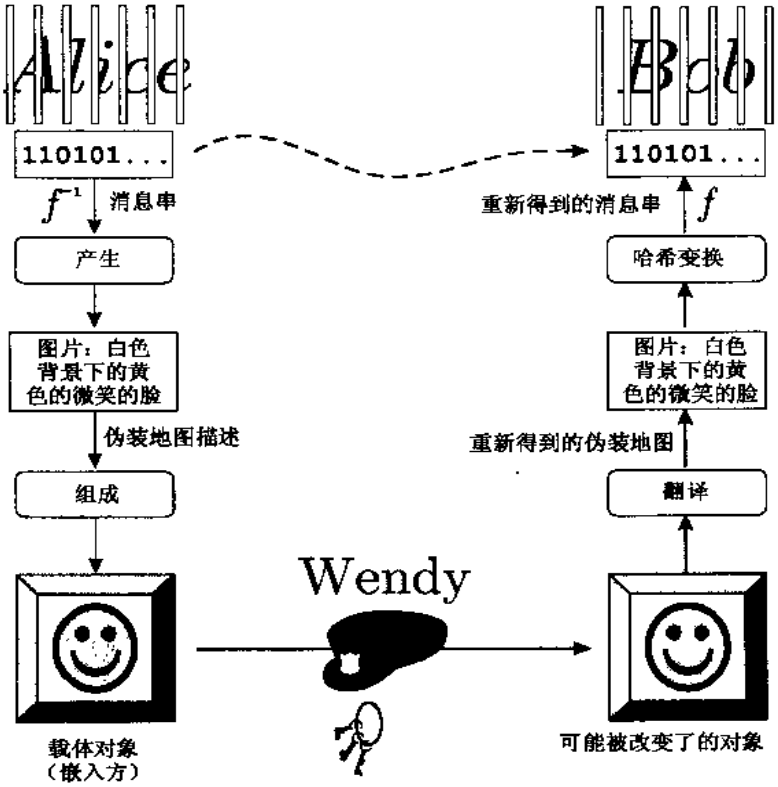

图 2.5 网上信道的示意图

且还不清楚伪装地图函数 f 是否有快速算法。Craver[2]建议使用音频—视频剪辑中“清晰明确易懂的”词作为伪装地图并定义 f 为一个哈希函数。

2.5.3 恶意的攻击者——安全的信息伪装

在有恶意攻击者的场合，光有健壮性是不够的。如果嵌入方法不是依赖于某些发送者与接收者共享的秘密信息（也就是，无密钥信息伪装或公钥信息伪装的情形），那么攻击者可以伪造消息，因为接收者不能够验证发送者身份的正确性。为了避免这种攻击，算法必须既健壮又安全。我们根据下面四个要求来给安全的信息伪装算法下一个定义。

- 使用一个公开算法和一个密钥来隐藏消息，而该密钥必须能唯一地标识发送者。
- 只有正确密钥的拥有者才能检测、提取以及证明隐藏消息的存在性。其他任何人都不应该能找到任何消息存在性的统计证据。
- 即使敌人知道（或者能够选择）一个隐藏消息的内容，他也不应该有机会检测出其他隐藏的消息。
- 检测隐藏的消息在计算上是不可行的。

2.6 在文本中隐藏信息

不像噪声数据，文本数据含有很少的可用来进行秘密通信的冗余信息。隐藏的方法可以

是试图将信息直接编码到文本内容中去(利用语言的自然冗余性),或者将信息直接编码到文本格式中去(比如,调整字间距或行间距)。

已经提出许多方法来将信息直接存储在文本消息中,比如,可以利用偶尔的打字或拼写错误、逗号可以省略、词可以用同义词替代等等。但它们中绝大多数并不是首选的,因为它们会严重降低文本的品质。另外嵌入工作需要用户的介入和交互,所以不能自动完成。

另一方面,可以创作出一篇文本消息仅用作秘密通信的载体。Wayner [15,16]描述了一个最有前景的技术,即 Trithemius 表的自动化版本。他使用上下文自由语法(CFG)来创作载体文本,并根据要传送的秘密消息选择一个创作的载体文本,详细内容请参看 3.7 小节。这样秘密信息不是嵌入在载体中,而是载体本身(实际上是按 CFG 创作载体的方法)就是秘密消息。如果这种语法是明确的话,则接收者就可以运用标准的语法分析提取信息,而这些方法在构造过程中已被充分研究过。根据语法创作出来的每一个词应该形成一个有意义的消息,否则会立刻引起攻击者的怀疑。

如果载体文本以固定格式(像 HTML、LATEX 或 Postscript 文件)的形式传输,则信息可以嵌入到格式中而不是消息内容本身。秘密信息可以存储在行间距或列间距中[17 - 19]。如果两行之间的距离小于某个门限值,就表示隐藏的信息是"0",否则隐藏的信息是"1"。类似的方法也可用于传输 ASCII 文本的信息,偶尔的附加空格字符可以用来构成秘密信息。

许多其它的(更巧妙精细的)方法也可以采用,比如,将信息编码隐藏在字处理系统的断行处。例如,TEX 使用一个非常复杂的算法来计算行与行之间和页与页之间的断点[20],实际上,它将所有断点按下列三个值进行分类:badness、penalty、demerits。粗略地说,一行的 badness 是衡量要插入多少空白空间才能使 TEX 达到"群聚风格"。而 penalty 被分配给每一个可能的断点,它代表在这个特定的位置断开所付出的"美学代价"。每一行的 demerits 参数是基于 badness 和 penalty 来计算的。TEX 试图选择一个断点序列,使得一个段落中所有的 demerits 的总和达到最小。通过调整一些内部参数,可以选择次优的断点,从而储存额外的信息。

在文本消息中能否存在安全和健壮的信息隐藏仍然是一个悬而未决的问题。一个攻击者只需简单地试图重新调整文本的格式,就可以破坏掉所有嵌入在文本格式中的信息。另外,文本消息可以以各种不同的格式进行存储(如 HTML、TEX 的 DVI、Postscript、PDF 或者 RTF),从一种格式转化到另一种格式对嵌入的消息也有很大的损害。

2.7　不可视通信的例子

最后我们想描述一下出现在科学文献中的信息伪装原理的某些应用和成功实现的例子,并以此结束本章。

2.7.1　数字签名方案中的阈下信道

Simmons[21]指出,充分利用数字签名方案的弱点,两个人可以进行不可视通信。ElGamal 型数字签名方案、DSA、ESIGN 以及其它的数字签名方案均可引入阈下信道。作为一个例子,我们想在这里讲述一下如何利用 ElGamal 签名方案[22]来构造一个阈下信道。为了生成 ElGamal 密钥,用户首先选择一个素数 p,选择 Z_p^* 的一个生成元 g 和一个随机数 $x < p$。然后用户计算 $y = g^x \bmod p$,于是公钥为三元组 $\langle y, g, p \rangle$,私钥为 x。为了对消息 M 签名,用户首先

选择一个随机数 k，且使 k 与 $(p-1)$ 互素，计算 $a = g^k \bmod p$，并解 $M \equiv xa + kb \bmod (p-1)$ 方程，求解 b。签名就是 (a, b)。为验证该签名，验证方程 $y^a a^b \equiv g^M \bmod p$。

为了在数字签名中储存附加的秘密信息，接收者必须获得发送者的私钥 x。为了将秘密消息 M' 与某些无关紧要的消息 M 一起发送，M' 在基本的 ElGamal 方案（也就是，发送者计算 $a = g^{M'} \bmod p$ 并从方程 $M \equiv xa + M'b \bmod (p-1)$ 求解 b）中扮演随机数 k 的角色，签名仍然是 (a, b) 并像上述那样进行验证。如果接收者已获得 x，则他可以利用扩展的欧里几得算法重构 M'（给予更强的条件）。（事实上，还有其它弱点，见[23]）。对这种数字签名中的阈下信道的检测，特别强调需要阈下免疫的数字签名方案[24]，即被证明对阈下信道具有免疫的签名方案。

2.7.2 操作系统中的隐蔽信道

如果通信双方都接入同一个计算机系统（或者如果通信双方实际上是运行在同一个主机上的两个进程），可以有许多巧妙的方法来进行不可视通信。几十年来操作系统的设计者们一直担心在高安全性的操作系统里是否可能存在安全漏洞。他们关心的是，恶意的程序也许可以利用隐蔽信道来将敏感信息从高保护级的系统区域传送到低保护级的系统区域。按照 Lampson[25] 的定义，如果一个通信信道既不是有意设计的，也根本不用来传输信息，那么称这种信道为隐蔽信道，它使用一些通常不视为数据对象的实体，从一个客体向另一个客体传输信息[26]。

当运行在一个特定安全级别的系统的一部分（也就是，一个共享资源）能够向另一个系统部分（可能不同安全级别）提供服务时，隐蔽信道就可能会出现。考虑下面的例子，在一个操作系统里，运行在高安全级别的进程 A 能够向一个磁盘写数据，而运行在低安全级别的另一个进程 B 能够访问其文件表（即，由前一个进程创建的所有文件的名字和大小），虽然它没有访问数据本身。这种情形可以导致有一个隐蔽信道，进程 A 通过选择合适的文件名和大小来向 B 发送信息。隐蔽信道的分类和更详细的介绍请参考[27]。

Handel 等人[28]仔细研究了开放系统互联（OSI）网络模型，以查找所有可能用来传输秘密信息的隐蔽信道。他们找到了许多可能的弱点（比如，数据链路层中数据帧的未使用部分）。甚至有一些更为精妙的方法来共享数据，如一个 IP 包的时间戳可以用来传输一比特数据（偶时间增量发送的包代表逻辑 0，奇时间增量发送的包代表逻辑 1）、以太网物理层的碰撞检测系统可以被修改、因特网控制消息可以被利用等等，参见 3.2.8 和 4.4。

2.7.3 视频通信系统

伪装术可以用于将秘密消息嵌入到由视频会议系统录制的视频流中。Westfeld 等人[29]报告了一个系统，该系统将消息隐藏到一个基于 DCT 的有损压缩的视频流中。他们说明，在一个综合业务数字网（ISDN）视频会议系统里，可以嵌入一个 GSM 电话对话（带宽高达 8kbit/s）而不会使视频信号严重降质，从而形成了一个秘密通信。然而，这个信息隐藏速率依赖于用作载体的视频图像的特性。

2.7.4 在可执行文件中隐藏数据

可执行文件以这样的方式包含许多冗余信息，如可以安排独立的一串指令，或者选择一个

指令子集解决特定的问题。代码迷乱技术，最初主要用来保护软件产品的不正当再处理，它能用来在可执行文件中存储额外的信息。这种技术试图把一个程序 P 变换成一个功能等价的程序 P'，而 P' 更难以反向编程，在信息伪装应用中，秘密信息隐藏在所用的一系列变换中。

Collberg 等人[30]指出，唯一的要求就是让这个"看得见"的行为未被别人提防(即，这种行为与普通用户经历的一样)。更精确地说，如果 $P \rightarrow P'$ 是源程序 P 到目标程序 P' 的一个变换，并满足两个条件：如果 P 未能终止或以一个错误信息终止，则 P' 可以终止也可以不终止，否则 P' 终止并产生与 P 一样的输出。Collberg 等人[30,31]列出了许多可用来迷乱 Java 代码的技术。它们中有"分支插入"变换和"循环条件插入"变换。第一个变换通过写两个功能等价的代码块引入一个额外的分支，而代码块根据分支条件来选择。第二个变换扩展循环条件使得循环执行总时间不受影响。

2.8 结　论

1997 年 12 月 16 日，伦敦每日电讯援引了一个非官方的欧盟报告，该报告透露，有一个称作 ECHELON 的系统，专门用于审查欧洲通信，即"欧洲范围内所有的电子邮件、电话和传真通信都例行公事地被美国国家安全局截取，并将所有的目标信息从欧洲大陆通过伦敦的战略性 Hub(网络集线器)再经卫星转移到马里兰的 Fort Meade，再通过北约克角的 Menwith 山的关键 Hub 停留在英国"。这个事件和其它事件表明，使用密码或任何其它可确保私有性的方法对保护公民自由权利都是至关重要的。由于越来越多的国家限制强密码的使用，所以可替代的方法就变得越来越重要。

本章内容表明，在计算机时代，不可视通信是有可能的。几乎任何消息都具有作为秘密通信载体的潜力：数字图像或数字声音的噪声成分可以被修改，格式化的字处理器输出结果可以包含秘密，通过 CFG 可以制作消息，数字签名算法的缺点可以被利用，甚至一个操作系统里的两个进程的通信也能用来交换机密信息。其中许多技术(和攻击它们的方法)将在后面的各章节中深入讨论。

参考文献

[1] Simmons, G. J., "The Prisoners' Problem and the Subliminal Channel," *In Advances in Cryptolopy, Proceedings of CRYPTO'83*, Plenum Press, 1984, pp. 51 – 67.

[2] Craver, S., "On Public-Key Steganography in the Presence of an Active Warden," Technical Report RC 20931, IBM, 1997.

[3] Aura, T., "Practical Invisibility in Digital Communication," in *Information Hiding: First International Workshop , Proceedings*, vol. 1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp. 265 – 278.

[4] Anderson, R. J., "Stretching the Limits of Steganography," in *Information Hiding: First International Workshop, Proceedings*, vol. 1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp. 39 – 48.

[5] Anderson, R. J, and F. A. P. Petitcolas, "On The Limits of Steganography." *IEEE Journal of Selected Area in Communications*, vol. 16, no. 4, 1998, pp. 474 – 481.

[6] Cachin, C., "An Information-Theoretic Model for Steganography." in *Proceeding of the Second International Workshop on Information Hiding*, vol. 1525 of *Lecture Notes in Computer Science*, Springer, 1998, pp. 306 – 318.

[7] Möller, S., A. Pfitzmann, and I. Stirand, "Computer Based Steganography: How It Works and Why Therefore Any Restrictions on Cryptography Are Nonsense, At Best." in *Information Hiding: First International Workshop, Proceedings*, vol. 1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp. 7 – 21.

[8] Fridrich, J., "Innovative C4I Technologies (Secure Image Encryption and Hiding)," Technical report, final report for SBIR project No. AF97 – 043(Phase I), AFRL, Rome, New York, 1998.

[9] Pennebaker, W. B., and J. L. Mitchell, *JPEG Still Image Compression Standard*, New York: Van Nostrand Reinhold, 1993.

[10] Wallace, G. K., "The JPEG Still Picture Compression Standard," *Communications of the ACM*, vol. 34, no. 4, 1991, pp30 – 44.

[11] Johnson, N. F., Z. Duric, and S. Jajodia, "A Role for Digital Watermarking in Electronic Commerce," to appear in *ACM Computing Surveys*.

[12] Johnson, N. F., "An Introduction to Watermark Recovery from Images," in *SANS Intrusion Detection and Response Conference, Proceedings*, 1999.

[13] Bender, W., D. Gruhl, and N. Morimoto, "Technique for Data Hiding," *IBM Systems Journal*, vol. 35, no. 3/4, 1996, pp. 313 – 336.

[14] Cox, I. J., et al., "Secure Spread Spectrum Watermarking for Multimedia," Technical report, NEC Institute, 1995.

[15] Wayner, P., "Mimic Functions," *Cryptologia*, vol. XVI/3, 1992, pp. 193 – 214.

[16] Wayner, P., "Strong Theoretical Steganography," *Cryptologia*, vol. XIX/3, 1995, pp. 285 – 299.

[17] Brassil, J., N. F. Maxemchuk, and L. O' Gorman, "Electronic Marking and Identification Techniques to Discourage Document Copying," in *Proceedings of INFORCOM' 94*, 1994, pp. 1278 – 1287.

[18] Low, S. H., N. F. Maxemchuk, and A. M. Lapone, "Document Identification for Copyright Protection Using Centroid Detection," *IEEE Transactions on Communications*, vol. 46, no. 3, 1998, pp. 372 – 383.

[19] Maxemchuk, N. F., "Electronic Document Distribution," *AT&T Technical Journal*, September/October 1994, pp. 73 – 80.

[20] Knuth, D. E., *The TEXbook*, Reading, MA: Addison Wesley, 1984.

[21] Simmons, G. J., "The Subliminal Channel and Digital Signature," in *Advance in Cryptology, Proceedings of EUROCRYPT' 84*, vol. 209 of *Lecture Notes in Computer Science*, Springer, 1985, pp. 364 – 378.

[22] ElGamal, T., "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms," in *Advances in Cryptology, Proceedings of CRYPTO' 84*, vol. 196 of *Lecture Notes in Computer Science*, Springer, 1985, pp. 10 – 18.

[23] Anderson, R. J., et al., "The Newton Channel," in *Information Hiding: First International Work-*

shop, *Proceedings*, vol.1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp.151 – 156.

[24] Desmedt, Y., "Subliminal-Free Authentication and Signature," in Advances *in Cryptology*, *Proceedings of EUROCRYPT'88*, vol.330 of *Lecture Notes in Computer Science*, Berlin, New York: Springer, 1988, pp.22 – 33.

[25] Lampson, B.W., "A Note on the Confinement Problem," *Communications of the ACM*, vol.16, no.10, 1973, pp.613 – 615.

[26] Kemmerer, R.A., "Shared Resource Matrix Methodology: An Approach to Identifying Storage and Timing Channels," *ACM Transactions on Computer Systems*, vol.1, no.3, 1983, pp.256 – 277.

[27] Meadows, C., and I. Moskowitz, "Convert channels——A Context-Based Review," in *Information Hiding: First International Workshop*, *Proceedings*, vol.1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp.73 – 93.

[28] Handel, T.G., and M.T. Sandford, "Data Hiding in the OSI Network Model," in *Information Hiding: First International Workshop*, *Proceedings*, vol.1174 of *Lecture Notes in Computer Science*, Springer, 1996, pp.23 – 38.

[29] Westfeld, A., and G. Wolf, "Steganography in a Video Conferencing System," in *Proceeding of the Second International Workshop on Information Hiding*, vol.1525 of *Lecture Notes in Computer Science*, Springer, 1998, pp.32 – 47.

[30] Collberg, C., C. Thomborson, and D. Low, "Manufacturing Cheap, Resilient and Stealthy Opaque Constructs," in *Proceedings of the ACM Symposium on the Principles of Programming Languages*, 1998, pp.184 – 196.

[31] Collberg, C., C. Thomborson, and D. Low, "A Taxonomy of Obfuscating Transforms," Technical Report 148, Department of Computer Science, The University of Auckland, 1997.

[32] "Spies Like US," *London Daily Telegraph*, 16Th December 1997.

第三章　隐写术综论

(Neil F. Johnson, Stefan C. Katzenbeisser)

在过去几年里，人们提出了许多不同的信息伪装方法，其中大部分可看作是替换系统。这些方法尽量把信号的冗余部分替换成秘密信息(如 2.3 节描述的一样)，它们主要的缺点是对修改伪装载体具有相当的脆弱性。最近，新的健壮性水印技术的研究使得在构造健壮和安全信息伪装系统方面取得了很大进展。因此，这里描述的一些方法是与第六章的数字水印技术紧密相关的。

对信息伪装系统进行分类有许多种方法，可以根据用于秘密通信的伪装载体类型进行分类，也可以根据嵌入过程中对伪装载体的修改方式进行分类。尽管在某些情况下，不可能进行准确的分类，但我们仍采用第二种办法，把信息伪装方法分为如下六类：

- 替换系统：用秘密信息替代伪装载体的冗余部分；
- 变换域技术：在信号的变换域嵌入秘密信息(例如，在频域)；
- 扩展频谱技术：采用了扩频通信的思想；
- 统计方法：通过更改伪装载体的若干统计特性对信息进行编码，并在提取过程中采用假设检验方法；
- 失真技术：通过信号失真来保存信息，在解码时测量与原始载体的偏差；
- 载体生成方法：对信息进行编码以生成用于秘密通信的伪装载体。

下面的章节将分别对这六类伪装技术进行讨论。

3.1 基本定义

在下面的各小节中，我们使用 c 表示嵌入过程中的伪装载体。不失一般性，进一步假定，任何伪装载体能用长为 $l(c)$ 的 c_i 序列(这里，$1 \leqslant i \leqslant l(c)$)表示。例如在数字声音中，它正好是基于时间的取样序列；在数字图像中，通过对图像矢量化也能获得这样的序列(例如，通过按从左向右，从上到下将所有像素排列起来)。对二值图像，c_i 的取值域是 $\{0,1\}$，对量化的图像或声音，c_i 的可能取值是大于 0 并且小于 256 的整数。我们用 s 来表示伪装对象，同样 s 也是一个长为 $l(c)$ 的 s_i 序列。

有时必须对所有的载体元素 c_i 进行索引，我们使用符号 j 作为索引值。如果索引值本身按某种顺序排序，我们就使用记号 j_i 来表示，当提到第 j_i 个载体元素时，即指 c_{j_i}。我们记伪装密钥为 k，k 的结构将在每一种信息伪装应用中分别解释。秘密信息记作 m，m 也是长为 $l(m)$ 的由 m_i 组成的序列，$1 \leqslant i \leqslant l(m)$。除非另外注明，否则我们假定 $m_i \in \{0,1\}$。

在色彩空间(所有可能颜色的集合)中一个颜色元素通常是一个三维向量，参见[1]。最常见的色彩空间是 RGB。由于红、绿和蓝是可加性的三元色，所以任何一种颜色都可以表示成红、绿、蓝成分的加权和。在 RGB 空间中每一个向量描述了这三种成分的精度。另一种常见的色彩空间是 YCbCr，每一个颜色元素表示为一个亮度(Y)成分和两个色度(Cb, Cr)成分，并

且 Y 成分表示一个颜色的亮度，Cb 和 Cr 表示不同的颜色级。使用下面公式，在 RGB 中的一个颜色向量可以转化成一个 $YCbCr$ 向量：

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cb &= 0.5 + (B - Y)/2 \\ Cr &= 0.5 + (R - Y)/1.6 \end{aligned} \tag{3.1}$$

一幅图像 C 就是一个离散函数 $c(x,y)$，它给每个像素点 (x,y) 赋一个（任何色彩空间的）颜色向量。

3.2　替换系统和位平面工具

在各种媒介中有很多方法可以用于隐藏信息，这些方法包括使用 LSB 编码（也常称为位平面或噪音插入工具）、用图像处理或压缩算法对图像的属性（如亮度）进行修改。基本的替换系统，就是试图用秘密信息比特替换掉伪装载体中不重要的部分，以达到对秘密信息进行编码的目的。如果接收者知道秘密信息嵌入的位置，他就能提取出秘密信息。由于在嵌入过程中仅对不重要的部分进行修改，发送者可以假定这种修改不会引起被动攻击者的注意。

3.2.1　最低比特位替换

位平面工具包括应用 LSB 插入和噪音处理之类的方法，这些方法在信息伪装中很常见，而且很容易用于图像和声音[2-6]。伪装载体中能隐藏数量惊人的信息，即使对载体有影响，也几乎察觉不到[5,7,8]。

诸如软件程序 StegoDos[9]、S-Tools[10]、Mandelsteg[11]、EzStego[12]、Hide and Seek[13]、Hide4PGP[14]，White Noise Storm[15]和 Steganos[16]等都使用了位平面工具。在这些伪装方法中主要使用无损图像格式，并且数据能直接处理和恢复。这些程序中，一部分除应用伪装手段外，还采用了压缩和加密技术，以提供更好的隐藏数据的安全性。尽管如此，位平面方法对伪装载体稍微更改的抵抗力仍是相当脆弱的。

嵌入过程包括选择一个载体元素的子集 $\{j_1,\ldots,j_{l(m)}\}$，然后在子集上执行替换操作 $c_{j_i} \leftrightarrow m_i$，即把 c_{j_i} 的 LSB 与 m_i 进行交换（m_i 可以是 1 或 0）。一个替换系统也可以修改载体的多个比特，例如，在一个载体元素的两个最低比特位隐藏两比特信息。在提取过程中，抽出被选择载体元素的 LSB，然后排列起来重构秘密信息。基本方法在算法 3.1 和 3.2 中描述。在这里有一个问题需要解决，即采用什么方法选择 c_{j_i}。

算法 3.1　嵌入过程——最低比特位替换

for $i = 1,\ldots, l$ (c) do

$S_i \leftarrow c_i$

end for

for $i = 1,\ldots, l$ (m) do

compute index j_i where to store ith message bit

$S_{j_i} \leftarrow c_{j_i} \leftrightarrow m_i$

end for

算法 3.2 提取过程——最低比特位替换

for $i = 1, \ldots, l(\mathrm{M})$ do

compute index j_i where the i th message bit is stored

$m_i \leftarrow \mathrm{LSB}(c_{j_i})$

end for

为了能解出秘密信息，接收者必需能获得嵌入过程中使用的索引序列。在最简单的情况下，发送者从第一个元素开始，使用所有的伪装载体元素进行信息传送。通常由于秘密信息比特数比 $l(c)$小，嵌入处理在载体末尾很长一段之前就结束了。这种情况下，剩下的载体元素保持不变。但是这导致了严重的安全问题，载体的第一部分与第二部分，也就是修改的部分和没有修改的部分，具有不同的统计特性。为了解决这个问题，比如共享程序 PGMStealth 中使用了随机序列来延长秘密信息，使得 $l(c) = l(M)$，因而对载体的开始和结尾产生了一致的随机修改。结果是，嵌入过程更改了比传送秘密信息所需要的更多的元素，从而增大了攻击者对秘密通信的怀疑的可能性。

较复杂的方法是，使用伪随机数发生器以相当随机的方式来扩展秘密信息，一个流行的方法是随机间隔法[3]。如果通讯双方使用同一个伪装密钥 k 作随机数发生器的种子，那么他们能生成一个随机序列 $k_1, \ldots, k_{l(m)}$，并且把它们和索引一起按下列方式生成隐藏信息位置来进行信息传送：

$$j_1 = k_1 \qquad (3.2)$$
$$j_i = j_{i-1} + k_i \qquad i \geqslant 2$$

从而，可以伪随机地决定两个嵌入位的距离。由于接收者能获得种子 k 和随机数发生器的信息，因此他能重构 k_i，进一步获得整个元素的索引序列 j_i。这种技术在流载体中尤其有效。见算法 3.3 和 3.4 所示，它们是算法 3.1 和 3.2 的特殊情况。

算法 3.3 嵌入过程——随机间隔法

for $i = 1, \ldots, l$ (c) do

$s_i \leftarrow c_i$

end for

generate random sequence k_i using seed k

$n \leftarrow k_1$

for $i = 1, \ldots, l$ (m) do

$s_n \leftarrow c_n \leftrightarrow m_i$

$n \leftarrow n + k_i$

end for

算法 3.4 提取过程——随机间隔法

generate random sequence k_i using seed k

$n \leftarrow k_1$

for $i = 1, \ldots, l$ (m) do

$m_i \leftarrow \mathrm{LSB}(c_n)$

$n \leftarrow n + k_i$

end for

3.2.2　伪随机置换

如果在嵌入过程中能获得所有的伪装载体比特(即,如果 c 是一个可以任意访问的伪装载体),那么就能把秘密信息比特随机地分散在整个载体中。由于不能保证随后的消息位按某种顺序嵌入,这种技术进一步增加了攻击的复杂度。

Alice 首先尝试(使用一个伪随机数发生器)创建一个索引序列 $j_1, \ldots, j_{l(m)}$,并将第 k 个消息比特隐藏在索引为 j_k 的载体元素中。注意,由于我们对伪随机数发生器的输出不加任何限制,一个索引值在序列中可能出现多次,我们称这种情况为碰撞。如果一个碰撞发生,Alice 将可能在一个载体元素中插入多个消息比特,因而破坏了这些信息。如果与载体元素的个数相比,消息比特较少的话,她可以希望发生碰撞的概率能够忽略,并且被破坏的比特能使用纠错编码进行重构。然而这仅仅适合很短的秘密信息。至少发生一次碰撞的概率 p 能通过下面公式进行估计[1](假定 $l(m) \ll l(c)$:

$$P \approx 1 - \exp\left(-\frac{l(m)[l(m)-1]}{2l(c)}\right)$$

因为 $l(c)$是常数,当 $l(m)$增加时,p 会很快地趋近于 1。例如,如果载体是一个 600×600 像素的图像并且在嵌入过程中选择 200 个像素,p 大约是 5%。另一方面,如果进行信息传输时使用了 600 个像素,p 则增加到 40%左右。我们可以断言只有对非常短的消息,才能忽略碰撞的概率。如果消息长度增加,碰撞必须加以考虑。

为了解决碰撞问题,Alice 可以在一个集合 B 中记录所有已经使用过的载体元素。如果在嵌入过程中,一个载体元素以前没有使用过,她把它的索引加入集合 B,并且使用这个元素。然而,如果载体元素索引已经包含在集合 B 中,那么她就放弃这个元素并伪随机地选择另一个元素。在接收方,Bob 采用相似的技巧。

Aura 在[18]中提出了另一种方法,他使用算法 3.1 和 3.2 的基本替换方案,并且通过使用集合$\{1, \ldots, l(c)\}$的伪随机置换来计算索引 j_i。假设 $l(c)$能表示成两个数字 X 和 Y 的乘积(在数字图像中总是属于这种情况),并且 h_k 是一个任意的依赖于密钥 k 的安全哈希函数。令 k_1、k_2 和 k_3 是三个密钥。如[19,20]所述,算法 3.5 对每一个输入 i $(1 \leqslant i \leqslant XY)$,都输出一个不同的数 j_i(即,若以 $i = 1, \ldots l(c)$为输入对算法求值,它等同于产生集合$\{1, \ldots l(c)\}$的一个伪随机置换)。

Alice 首先把伪装密钥 k 分成三个子密钥 k_1、k_2 和 k_3。在嵌入过程中,她用算法 3.5 计算出的索引 j_i 保存第 i 个消息比特。由于算法 3.5 不生成重复的元素索引,故不会发生碰撞。如果 Bob 已经获得了 k_1、k_2 和 k_3,他能重构 Alice 嵌入秘密消息比特的位置。然而,Aura 的方法需要相当可观的计算时间,因为哈希函数必须计算 $3l(m)$次。

[1] 计算 p 的问题是所谓生日问题的一个实例:一个缸中放 n 个球,并从 1 到 n 编号。假设替换式地从缸中取 m 个球,并列出它们的号码。假定 $m = O(\sqrt{n})$,至少有一个球被取两次的概率是 $P(n,m)$,可由下列公式给出[17]:

$$P(n,m) = 1 - \prod_{i=0}^{m-1}\left(1 - \frac{i}{n}\right) \rightarrow 1 - \exp\left(-\frac{m(m-1)}{2n} + O\left(\frac{1}{\sqrt{n}}\right)\right)$$

算法 3.5　使用伪随机置换计算索引 j_i

$v \leftarrow i \text{ div } X$

$u \leftarrow i \bmod X$

$v \leftarrow (v + h_{k_1}(u)) \bmod Y$

$u \leftarrow (u + h_{k_2}(v)) \bmod X$

$v \leftarrow (v + h_{k_3}(u)) \bmod Y$

$j_i \leftarrow vX + u$

3.2.3　图像降级和隐蔽信道

在 1992 年，Kurak 和 McHugh[5]报道了在高安全级操作系统中的一个安全威胁。这个威胁属于信息伪装技术，它能用于秘密地交换图像，我们称之为图像降级。图像降级是替换系统中的特殊情况，其中图像既是秘密信息又是载体。给定一个同样尺寸的伪装载体和秘密图像，发送者把伪装载体图像灰度(或彩色)值的四个最低比特替换成秘密图像的四个最高比特。接收者从隐藏后的图像中把四个最低比特提取出来，从而获得秘密图像的四个最高比特位。在许多情况下载体的降质视觉上是不易察觉的，并且对传送一个秘密图像的粗略近似而言，四比特足够了。

在多级安全操作系统中，主体(进程、用户)和客体(文件、数据库等)都被指派一个特定的安全级别，参见著名的 Bell-LaPadula[21]模型。主体通常仅允许读取较低安全级别的客体("不能向上读")，同时只能向较高安全级别的客体进行写操作("不能向下写")。第一个限制的原因是明显的，而第二个限制的原因则是试图阻止用户将重要信息变为低安全级别主体可访问的。信息降级，就是通过将机密信息嵌入较低安全级别的客体中，使得机密信息不再机密(信息降级因此得名)，从而破坏了"不能向下写"的原则。第四章将着眼于可能的对策。

3.2.4　载体区域和奇偶校验位

我们称任何一个非空子集 $\{c_1, \ldots, c_{l(c)}\}$ 为一个载体区域。通过把载体分成几个不相接的区域，从而可以在一个载体区域中(而不是单个元素中)贮存一比特信息。一个区域 I 的奇偶校验位能通过下面公式计算出来：

$$p(I) = \sum_{j \in I} LSB(C_j) \bmod 2 \tag{3.3}$$

在嵌入过程中，首先选择 $l(m)$ 个不相接区域 $I_i (1 \leqslant i \leqslant l(m))$，每一个区域在奇偶校验位 $p(I_i)$ 上嵌入一个信息比特 m_i。如果一个载体区域的奇偶验校位与 m_i 不匹配，则将 I_i 中所有值的最低一个比特位进行反转，结果导致 $p(I_i) = m_i$。在译码过程中，计算出所有区域的奇偶校验位，排列起来就可重构消息。另外，使用伪装密钥作为种子，能伪随机地构造载体区域。

尽管这种方法不比简单的比特替换方法健壮，但是它在许多情况下会很有用。首先，发送者可以选择修改载体区域中的某一个元素，从而使得对载体的统计特性改变最小。进一步来说，由 N 个随机选择的元素组成的一个载体区域，奇偶校验位是零的概率 p_0^* 大约是 1/2，几乎与"随机选择一个载体元素的 LSB 是零"的概率 p_0 相独立，因为

$$p_0^* = \sum_{i=0}^{\lceil N/2 \rceil} \binom{N}{2i} (1 - p_0)^{2i} p_0^{N-2i}$$

$$= \frac{p_o^N}{2}\left[\left(1 + \frac{1 - p_0}{p}\right)^N + \left(1 - \frac{1 - p_0}{p_0}\right)^N\right]$$

$$= \frac{1}{2}(1 + (2p_0 - 1)^N) \tag{3.4}$$

公式(3.4)遵循这样一个事实,当且仅当载体区域中有偶数个像素点,并且它们的最低比特位为 1 时,$p(I) = 0$。因为若 $0 < p_0 < 1$,$(2p_0 - 1)^N \to 0$,我们能断定当 N 增加时,无论 p_0 为何值,p_0^* 很快地趋近为 0。这表明嵌入过程对载体的影响能随着 N 的增加而减少。

3.2.5　基于调色板的图像

在基于调色板的图像中,仅用特定色彩空间的一个颜色子集来对图像着色。每一个基于调色板的图像由两部分组成:一部分是调色板,它定义了 N 种颜色索引对(i, c_i)列表,它为一个颜色向量 c_i 指配一个索引 i;另一部分是实际图像数据,它保存每一个像素的调色板索引,而不是保存实际的颜色值。如果整个图像仅使用一小部分颜色值,这种方法大大地减少了文件的尺寸。两种最流行的图像格式是图像交换格式(GIF)和 BMP 位图格式。然而,由于复杂压缩技术的出现,它们的使用也不再那么火了。

一般地,在基于调色板的图像中有两种方法对信息进行编码,或操作调色板,或操作图像数据。正如在上小节描述的替换方法一样,颜色向量的 LSB 也能用于信息传输。另外,因为调色板不需以任何方式排序,在以调色板保存颜色时,可选择对信息进行编码。因为有 $N!$ 个不同方式对调色板进行排序,所以有足够的能力对一个短信息进行编码。然而,所有使用调色板顺序保存信息的方法都不具有健壮性,任何攻击者都能简单地以不同方式排序调色板而毁坏秘密信息(甚至在视觉上他可以不修改图像)。

另外,还可以在图像数据中对信息进行编码。由于调色板上相邻的颜色值在感观上并不接近,这样就不能简单地修改一些图像数据的 LSB。因此,一些信息伪装应用程序(如 ExStego)为使相邻颜色在感观上接近,在开始嵌入处理之前对调色板进行排序。例如,将颜色值根据它们在色彩空间中的欧几里德距离进行保存:

$$d = \sqrt{R^2 + G^2 + B^2} \tag{3.5}$$

由于人类视觉系统对颜色的亮度比较敏感,另一种可行也可能较好的方法是,根据颜色的亮度成份对调色板条目进行排序,具体参见(3.1)。在排序调色板后,可以放心地修改颜色索引的 LSB。

Fridrich [22]提出一种运用轻微差别技巧,不需要对调色板进行排序的方法,即对每一个像素,计算出(在欧几里德范数意义上)最邻近的颜色集。发送者从最近的颜色开始,接着是次接近颜色,并继续下去直到找到一个颜色,它的奇偶位$(R + G + B \bmod 2)$与要编码的秘密比特相匹配为止。一旦找到这种颜色,这个像素就被替换成这个新的颜色。

还有一种伪装应用是使用抖动方法把图片中颜色值的数目减少到$\lfloor N/2 \rfloor$,并且把整个调色板扩倍,轻微修改所有备份的调色板条目。经过这个预处理之后,抖动过的图像的每一个颜色值相应两个调色板条目,再根据秘密信息位选择其中的一个(例如,Mandelsteg [11]、S-Tool [10]、Hide4PGP[14]、和 Hide and Seek[13]都采用了这种方法的变种)。

3.2.6　量化和抖动

数字图像的抖动和量化能用于隐藏秘密信息。Matsui 和 Tanaka [23]基于量化图像操作提

出了两种信息伪装系统，我们在这里简单回顾一下预测编码中的量化。在预测编码中，每一个像素的大小是根据它的邻近区域像素值进行预测的。预测值可能是周围像素值的线性或非线性函数。最简单的情况是计算出相邻像素 x_i 和 x_{i-1} 的差值 e_i，并将它送入量化器 Q，由量化器输出差分信号 $x_i - x_{i-1}$ 的一个离散近似值 Δi（即 $\Delta i = Q(x_i - x_{i-1})$）。结果，在每一步量化中，都引入一个量化误差。对高度互相关信号，我们能预见 Δi 接近于 0，所以，熵编码器具有很高的效率。熵编码器的作用是给定数据的随机模型，试图产生最小冗余的编码。在接收端，对差分信号进行反向量化，并加上前一个信号的取样值来重构序列 x_i 的估计。

也可运用预测编码中的量化误差来达到信息伪装的目的。具体地说，就是调整差分信号 Δ_i 来传送额外信息。在这个方案中，伪装密钥由一个表构成，这个表给每一个可能的 Δ_i 值分配一个比特，例如，可能分配如下：

Δ_i	-4	-3	-2	-1	0	1	2	3	4
	0	1	0	1	1	1	0	0	1

为了在载体信号中保存第 i 个信息比特，计算出量化的差分信号 Δ_i，如果 Δ_i 与要编码的秘密信息比特不匹配（根据秘密表），则 Δ_i 由最接近的 Δ_j 替换，而 Δ_j 的对应比特等于秘密信息比特，最后将得到的 Δ_i 送入编码器。在接收端，信息根据差分信号和伪装密钥进行解码。

在信号的抖动处理过程中也可插入秘密信息，详情参见[23]及 Baharav 和 Shaked 的文章[24]。

3.2.7 在二值图像中的信息隐藏

二值图像（如，数字化的传真数据）以黑白像素分布方式包含冗余。尽管可以实现一个简单的替代系统（例如，某些像素根据某个具体的信息位设置成黑或白），但这些系统很容易受传输错误影响，因而不是很健壮。

Zhao 和 Koch 在[25]提出了一个信息隐藏方案，它使用一个特定图像区域中黑像素的个数来编码秘密信息。把一个二值图像分成矩形图像区域 B_i，分别令 $P_0(B_i)$ 和 $P_1(B_i)$ 为黑白像素在图像块 B_i 中所占的百分比。基本做法是，若某块 $P_1(B_i) > 50\%$，则嵌入一个 1，若 $P_0(B_i) > 50\%$，则嵌入一个 0。在嵌入过程中，为达到希望的像素关系，需要修改一些像素的颜色。修改是在那些邻近像素有相反的颜色的像素中进行的；在具有鲜明对比性的二值图像中，应该对黑白像素的边界进行修改。所有的这些规则都是为了确保不引起察觉。

为了提高整个系统对传输错误和图像修改的健壮性，我们必须调整嵌入处理。如果在传输过程中一些像素改变了颜色，诸如 $P_1(B_i)$ 由 50.6% 下降到 49.5，这种情况就会发生，从而破坏了嵌入信息。因此要引入两个阈值 $R_1 > 50\%$ 和 $R_0 < 50\%$ 以及一个健壮参数 λ，λ 是传输过程中能改变颜色的像素百分比。发送者在嵌入处理中确保 $P_1(B_i) \in [R_1, R_1 + \lambda]$ 或 $P_0(B_i) \in [R_0 - \lambda, R_0]$。如果为达到目标必须修改太多的像素，就把这块标识成无效，即修改 $P_1(B_i)$ 满足下面两个条件中的任何一个：

$$P_1(B_i) < R_0(B_i) - 3\lambda$$

$$P_1(B_i) > R_1(B_i) + 3\lambda$$

然后再为比特 i 伪随机地选择另一个块。在译码过程中，无效的块被跳过，有效的块根据 $P_1(B_i)$ 进行解码。嵌入和提取算法见算法 3.6 和 3.7 所示。

算法 3.6　Zhao 和 Koch 算法——在二进制图像中嵌入数据

for $i = 1, \ldots, l(M)$ do
　do forever
　　pseudorandomly select a new image block B_j
　　/ * Test, if block B_i is valid * /
　　if $P_1(B_j) > R_1 + 3\lambda$　or　$P_1(B_j) < R_0 - 3\lambda$　then continue
　　if($c_i = 1$ and $P_1(B_j) < R_0$) or($c_i = 0$ and $P_1(B_j) > R_1$)then
　　mark block B_j as unusable, i.e. modify block so that
　　either $P_1(B_j) < R_0 - 3\lambda$ or $P_1(B_j) > R_1 + 3\lambda$
　　continue
　　endif
　　break
　enddo
　/ * Embed Secret Message bit in B_j * /
　if $c_i = 1$ then
　　modify B_j so that $P_1(B_j) \geqslant R_1$ and $P_1(B_j) \leqslant R_1 + \lambda$
　else
　　modify B_j so that $P_0(B_j) \leqslant R_0$ and $P_0(B_j) \geqslant R_0 - \lambda$
　end if
end for

算法 3.7　提取过程 (Zhao 和 Koch)

for $i = 1, \ldots, e(M)$ do
　do forever
　　pseudorandomly select image block B_j
　　if $P_1(B_j) > R_1 + 3\lambda$　or　$P_1(B_j) < R_0 - 3\lambda$　then continue
　　break
　　enddo
　　if $P_1(B_j) > 50\%$ then
　　$m_i \leftarrow 1$
　　else
　　$m_i \leftarrow 0$
　　end if
　end for

Matsui 和 Tanaka 在[23]中提出了一个不同的嵌入方案，它在传真文档中使用无损压缩系统来对信息编码。根据以前的 CCITT(现在的国际电信联盟 ITU)[26]建议，传真图像能用游程(RL)编码和哈夫曼编码进行混合编码。RL 技术利用这样一个事实：在二值图像中，连续像素具有同种颜色的概率很高。图 3.1 显示了传真文档中的一个扫描行，我们用 a_i 指出改变颜色

的位置。RL 方法不再显式地对第一个像素颜色进行编码，而是对颜色变化(a_i)的位置和从 a_i 开始的持续同种颜色的像素个数 RL(a_i, a_{i+1})进行编码。我们假定的扫描行如图 3.1 所示，可编码为 < a_0,3 >，< a_1,5 >，< a_2,4 >，< a_3,2 >，< a_4,1 >。从而我们能用一个 RL 元素序列 < a_i, RL(a_i, a_{i+1}) > 来描述一个二值图像。

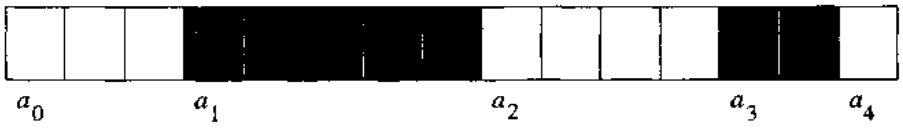

图 3.1 二值图像的一个扫描行

通过修改 RL(a_i, a_{i+1})的最低比特位，可以在一个二值的游程编码图像中嵌入信息。在编码处理中我们修改二值图像的游程长度，若第 i 个秘密消息位 m_i 是 0，我们令 RL(a_i, a_{i+1})为偶数；否则 RL(a_i, a_{i+1})为奇数，就表示 m_i 是 1。例如，可通过下面的方式进行：如果 m_i 是 0，而 RL(a_i, a_{i+1})是奇数，我们就把 a_{i+1}向左移动一个像素。另一方面，如果 $m_i = 1$ 并且 RL(a_i, a_{i+1})是偶数，我们就把 a_{i+1}向右移动一个像素。然而如果游程长度 RL(a_i, a_{i+1})是 1，这种嵌入方法就会出现问题，如果修改游程长度，就可能丢失数据。因此我们必须保证这种情况不发生，所以，所有游程长度为 1 的 RL 元素在嵌入处理前被废弃。

3.2.8 计算机系统中未使用或保留的空间

利用没有使用或保留的空间保存秘密信息，提供了一种隐藏信息的方式，并且伪装载体没有视觉上的降质。操作系统保存文件的方式很容易产生已经分配给文件而并未使用的空间。例如，在 Windows 95 操作系统中，没有压缩地格式化成 FAT16(MS-DOS 兼容)格式的驱动器，典型地使用 32KB 的簇。这意味着分配给文件的最小空间是 32KB，如果一个文件尺寸只有 1KB，那么额外的 31KB 就被"浪费"了。这"额外"空间能用于隐藏信息，并不显示在目录中。在图像或声音的文件头中未使用的空间也能用于隐藏"额外"数据。

另一种在文件系统中隐藏信息的方法是创建一个隐藏分区。如果系统正常起动，则无法看见这些分区。然而，许多情况下，运行一个磁盘配置实用程序(如 DOS 的 FDISK)就能发现隐藏分区。这些观念在一个新颖的伪装文件系统[27,28]中进行了扩展，如果用户知道文件名和密码，就能对文件进行访问；否则，文件系统中不存在此文件的任何痕迹。

OSI 网络协议模型具有能进行信息隐藏的特性[29]。穿过因特网进行信息传输的 TCP/IP 包在包头中有未使用空间。TCP 包头有 6 个未使用(保留)比特，IP 包头有两个保留比特。在每一个通信通道中传送成千上万个包，如果不进行检查，就能提供一个良好的秘密通信信道。大量可获得的伪装工具和其使用的简易性，使得执法部门不得不留意通过正在网上传输的网页图像、声音和其它文件进行非法数据传播。运用信息检测方法和熟悉当前隐藏技术是发现这些行为所必需的(参见第四章)。

3.3 变换域技术

我们已经看到，通过修改 LSB 嵌入信息的方法是比较容易的，但它们对极小的伪装载体修改都具有极大的脆弱性。一个攻击者想完全破坏秘密信息，只需简单地应用信号处理技术。

在许多情况下，即使由于有损压缩的很小变化也能使整个信息丢失。

前面已经提到，在伪装系统的发展中，在信号频域嵌入信息比在时域嵌入信息更具有健壮性。现在所了解的比较健壮的伪装系统实际都是运作在某种频域上。

变换域方法是在载体图像的显著区域隐藏信息，比 LSB 方法能够更好地抵抗攻击，例如压缩、裁剪和一些图像处理。它们不仅能更好地抵抗各种信号处理，而且还保持了对人类感官的不可觉察性。目前有许多变换域的隐藏方法。一种方法是使用离散余弦变换(DCT)[30－33]作为手段在图像中嵌入信息；还有一种使用小波变换[34]。变换可以在整个图像上进行[30]，也可以对整个图像进行分块操作[35,36]，或者是其它的变种。然而，图像中能够隐藏的信息数量和可获得的健壮性之间存在着矛盾[7,37]。许多变换域方法是与图像格式不相关的，并且能承受有损和无损格式转换。

在描述变换域伪装方法前，我们简短地回顾一下能把信号映射到频域中去的傅立叶变换和余弦变换。长 N 序列的离散傅立叶变换定义为

$$S(k) = \mathscr{F}\{s\} = \sum_{n=0}^{N-1} s(n)\exp(-\frac{2in\pi k}{N}) \tag{3.6}$$

这里 $i=\sqrt{-1}$是虚数单位。傅立叶逆变换为

$$s(k) = \mathscr{F}^{-1}\{S\} = \sum_{n=0}^{N-1} S(n)\exp(\frac{2in\pi k}{N}) \tag{3.7}$$

另一个有用的变换是 DCT 变换，公式表示为

$$S(k) = \mathscr{D}\{s\} = \frac{C(k)}{2}\sum_{j=0}^{N} s(j)\cos(\frac{(2j+1)k\pi}{2N})$$

$$s(k) = \mathscr{D}^{-1}\{s\} = \sum_{j=0}^{N} \frac{C(j)}{2}s(j)\cos(\frac{(2j+1)k\pi}{2N}) \tag{3.8}$$

这里如果 $u=0, C(u)=1/2$；否则 $C(u)=1$。DCT 变换最主要的好处是若序列 s 是实数，则 $\mathscr{D}\{s\}$也是实数序列。在数字图像处理中，二维 DCT 变换为

$$S(u,v) = \frac{2}{N}C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} s(x,y)\cos(\frac{\pi u(2x+1)}{2N})\cos(\frac{\pi v(2y+1)}{2N})$$

$$s(x,y) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C(u)C(v)S(u,v)\cos(\frac{\pi u(2x+1)}{2N})\cos(\frac{\pi v(2y+1)}{2N})$$

二维 DCT 变换是目前使用的最著名的有损数字图像压缩系统，JPEG 系统[38,39](参见图 3.2)的核心。JPEG 系统首先将要压缩的图像转换为 YCbCr 颜色空间，并把每一个颜色平面分成 8×8 的像素块。然后，对所有的块进行 DCT 变换。在量化阶段，对所有的 DCT 系数除以一些预定义的量化值(参见表 3.1)，并取整到最接近的整数(根据质量因子，量化值能通过一个

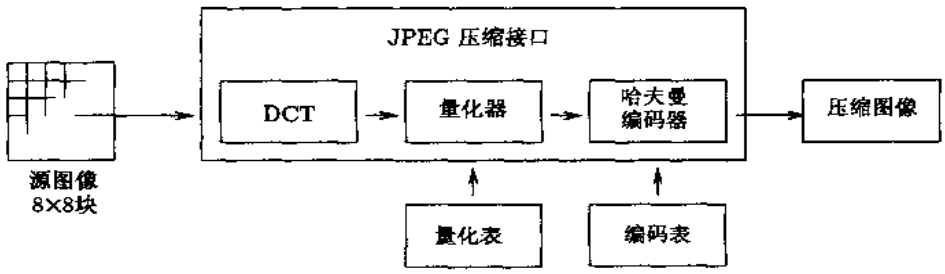

图 3.2　JPEG 图像压缩算法的流程图

常数进行缩放)。这个处理的目的是调整图像中不同频谱成分的影响,尤其是减小了最高频的DCT系数,它们主要是噪声并且不含有图像的细节。最终获得的DCT系数通过熵编码器进行压缩(例如,哈夫曼编码[40]或算术编码)。在JPEG译码时,逆量化所有的DCT系数(也就是乘以在编码阶段中使用的量化值),然后执行逆DCT变换重构数据。恢复后的图像很接近(但不等同)于原始图像。但是如果适当地设置量化值,得到的图像光凭人眼是觉察不到差异的。

表3.1　　在JPEG压缩方案中使用的量化值(亮度成分)

(u,v)	0	1	2	3	4	5	6	7
0	16	11	10	16	24	40	51	61
1	12	12	14	19	26	58	60	55
2	14	13	16	24	40	57	69	56
3	14	17	22	29	51	87	80	62
4	18	22	37	56	68	109	103	77
5	24	35	55	64	81	104	113	92
6	49	64	78	87	103	121	120	101
7	72	92	95	98	112	100	103	99

3.3.1 DCT域中的隐写术

一种在频域中流行的对秘密信息进行编码的方法是在一个图像块中调整两个(或多个)DCT系数的相对大小。我们将描述一个使用数字图像作为载体的系统,并且它与Zhao和Koch[25]提出的技术相似。

在编码处理中,发送者将载体图像分成8×8的像素块,每一块只精确地编码一个秘密信息位。嵌入过程开始时,首先伪随机地选择一个图像块b_i,用它对第i个消息比特进行编码。令$B_i=\mathcal{D}\{b_i\}$为DCT变换后的图像块。

在通信开始前,发送者和接收者必须对嵌入过程中使用的两个DCT系数的位置达成一致,让我们用(u_1,v_1)和(u_2,v_2)来表示这两个索引。这两个系数应该相应于余弦变换的中频,确保信息保存在信号的重要部位(从而使嵌入信息不容易因JPEG压缩而完全丢失)。进一步而言,人们普遍认为中频DCT系数有相似的数量级[41],我们可以假定嵌入过程不会使载体产生严重降质。因为构造的系统要在抵抗JPEG压缩方面是健壮的。我们就选择在JPEG压缩算法中它们的量化值一样的那些DCT系数。根据表3.1,系数(4,1)和(3,2),或者(1,2)和(3,0)是比较好的。

若块$B_i(u_1,v_1)>B_i(u_2,v_2)$就编码为"1",否则编码为"0"。在编码阶段,如果相对大小与要编码的比特不匹配,就相互交换两个系数。由于JPEG压缩(在量化阶段)能影响系数的相对大小,算法应通过在两个系数中加随机值,以确保对某个$x>0$,使得$|B_i(u_1,v_1)-B_i(u_2,v_2)|>x$。$x$值越大,算法抵抗JPEG压缩的能力就越健壮,然而图像的质量就越差。最后,发送者执行逆DCT变换把系数变换回空间域。为了从图像中提取信息,必须对所有图像块进行DCT变换。通过比较每一块中的两个系数,就可以得到隐藏的信息。嵌入和提取算法如算法

3.8 和 3.9 所示。

算法 3.8　DCT-Steg 编码过程

for　$i = 1, \ldots, l\ (M)$ do
　choose one cover-block b_i
　$B_i = \mathscr{D}\{b_i\}$
　if $m_i = 0$ then
　　if $B_i(u_1, v_1) > B_i(u_2, v_2)$ then
　　swap $B_i(u_1, v_1)$and $B_i(u_2, v_2)$
　　end if
　else
　　if $B_i(u_1, v_1) < B_i(u_2, v_2)$ then
　　swap $B_i(u_1, v_1)$and $B_i(u_2, v_2)$
　　end if
　end if
　adjust both values so that $|B_i(u_1, v_1) - B_i(u_2, v_2)| > x$
　$b_i' = D^{-1}\{B_i\}$
　end for
　create stego – image out of all b_i'

算法 3.9　DCT – Steg 解码过程

for $i = 1, \ldots, l\ (M)$ do
　get cover – block b_i associated with bit i
　$B_i = \mathscr{D}\{b_i\}$
　if $B_i(u_1, v_1) \leqslant B_i(u_2, v_2)$then
　$m_i = 0$
　else
　$m_i = 1$
　end if
end for

如果所使用的 DCT 系数的位置和常数 x 选择合适的话，嵌入处理不会对载体产生视觉上的降质。由于在量化处理中两个系数被除以相等的量化值，我们能预见这种方法对 JPEG 压缩是健壮的。因此，它们的相对大小仅受取整的影响。

上面提到的系统最大的缺点可能是算法 3.8 不能废弃某些图像块，在那些图像块里若让 DCT 系数满足所需要的关系，会严重地破坏图像数据。

Zhao 和 Koch[25,31]提出了一个相似的系统，它没有这种缺点。他们是对量化后的 DCT 系数进行操作，并使用块中三个 DCT 系数之间的关系来保存信息。发送者对图像块 b_i 进行 DCT 变换，并对其量化得到 B_i^Q。若一个块对比特 1 进行编码时，让 $B_iQ\ (u_1, v_1) > B_i^Q(u_3, v_3) + D$ 和 $B_i^Q(u_2, v_2) > B_i^Q(u_3, v_3) + D$。另一方面，如对 0 进行编码，让 $B_i^Q(u_1, v_1) + D < B_i^Q$

(u_3, v_3)和 $B_i^Q(u_2, v_2) + D < B_i^Q(u_3, v_3)$。参数 D 是描述一个嵌入位所需两个系数的最小矩离，通常 $D = 1$。D 越大，方法相对于图像处理技术就越健壮。再一次强调，应该在中频选择这三个系数。

在编码时，改变这三个系数的关系使得它们能代表一个秘密信息位。若在编码一个秘密信息位时，所需要的修改太大，那么将这块标识为"无效"，不用于信息传输。如果最大和最小的系数差大于某一常数 MD，就属这种情况。MD 越大，就有更多的块可用于通信。考虑到正确译码，需修改无效块的量化 DCT 系数，让它们满足下面条件之一

$$B_i^Q(u_1, v_1) \leqslant B_i^Q(u_3, v_3) \leqslant B_i^Q(u_2, v_2) \tag{3.9}$$

或

$$B_i^Q(u_2, v_2) \leqslant B_i^Q(u_3, v_3) \leqslant B_i^Q(u_1, v_1) \tag{3.10}$$

然后对块进行逆量化和逆 DCT 变换。

接收者通过应用 DCT 变换和块量化恢复信息。如果在块中选择的三个系数满足条件(3.9)或(3.10)，就忽略该块。否则，通过比较 $B_i^Q(u_1, v_1)$、$B_i^Q(u_2, v_2)$和 $B_i^Q(u_3, v_3)$就可以恢复编码的信息。由于所有修改是在有损量化阶段之后进行的，所以作者称这种嵌入方法对 JPEG 压缩(质量因子是 50%时)是健壮的。

3.3.2 在数字声音中隐藏信息——相位编码

数字声音中隐藏信息通常比数字图像中隐藏信息更难。Moore [42]提到，人类的听觉系统是特别敏感的，声音文件中千万分之一的扰乱都能被觉察出来。尽管可感知噪音极限随着载体噪音级的增加而提高，但最大可容许的噪音级一般是很低的。无论怎样，大家都知道人类听觉系统对声音的相位不太敏感，这个事实在大量数字声音压缩系统中被广泛应用。

在相位编码中[2]，通过载体信号相位谱的一个相位转换来代表一个数据。载体信号 c 分成由 N 个短序列 $c_i(n)$构成的序列，其中 $c_i(n)$长为 $l(m)$，然后进行 DFT 变换，并调用下面公式得到傅立叶变换的幅度 $A_i(k)$和相位 $\phi_i(k)$：

$$A_i(k) = \sqrt{\mathrm{Re}[\mathscr{F}\{c_i\}(k)]^2 + \mathrm{Im}[\mathscr{F}\{c_i\}(k)]^2} \tag{3.11}$$

和

$$\phi_i(k) = \arctan \frac{\mathrm{Im}[\mathscr{F}\{c_i\}(k)]}{\mathrm{Re}[\mathscr{F}\{c_i\}(k)]} \tag{3.12}$$

由于两个连续信号片段间的相位变动很容易被检测出来，因此它们的相位差需要在伪装信号中保持不变。结果，嵌入过程仅在第一个信号片断的相位向量中插入一个秘密消息：

$$\tilde{\phi}_0(k) = \begin{cases} \pi/2 & if \quad m_k = 0 \\ -\pi/2 & if \quad m_k = 1 \end{cases} \tag{3.13}$$

并用原来的相位差创建一个新的相位矩阵

$$\begin{aligned} \tilde{\phi}_1(k) &= \tilde{\phi}_0(k) + [\phi_1(k) - \phi_0(k)] \\ &\cdots\cdots \\ \tilde{\phi}_N(k) &= \tilde{\phi}_{N-1}(k) + [\phi_N(k) - \phi_{N-1}(k)] \end{aligned} \tag{3.14}$$

然后，发送者使用新的相位矩阵 $\tilde{\phi}_i(k)$和原来的傅立叶幅度 $A_i(k)$通过逆傅立叶变换来构造伪

装信号。由于修改了 $\phi_0(k)$，因而也就修改了所有随后信号片段的绝对相位，同时保证了它们的相对差不变。在恢复秘密信息之前，必须采用某种同步。假定知道序列长 $l(m)$，接收者就能计算 DFT，并能检测出相位 $\phi_0(k)$。

3.3.3 回声隐藏

回声隐藏[4]试图在离散信号 $f(t)$ 中引入回声 $f(t-\Delta t)$，生成伪装信号 $c(t)$ 来隐藏信息：

$$c(t) = f(t) + af(t-\Delta t) \tag{3.15}$$

在信号中，通过修改信号和回声间的延迟 Δt 来对信息进行编码。在编码阶段，发送者可以选择 Δt 或 $\Delta t'$。选择 Δt，是在信号中编入代码"0"；选择 $\Delta t'$，是在信号中编入代码"1"。延迟时间 Δt 或 $\Delta t'$ 是以人听不到回声信号为准则进行选取的。

基本的回声隐藏方案仅在一个信号中嵌入一个比特，因此连续载体信号在编码处理前先分成 $l(m)$ 块。连续的块之间用一些不用的取样点隔开，并且间隔大小是随机选取的，这样使得检测和提取秘密信息更加困难。在每一块中，根据式(3.15)嵌入一个秘密比特，最后，将所有的信号块连成一串。

从伪装信号中提取秘密信息前，必须采取某种同步措施，接收者必须能重构 $l(m)$ 个信号块，其中每个信号块是发送者用来嵌入一个秘密信息比特的。然后，每一个信号片段通过信号倒频谱的自相关函数进行解码。Gruhl 等[4]表明自相关函数在延迟时间 Δt 上出现峰值。关于回声隐藏的进一步研究请参见 7.5.1 节。

Chang 和 Moskowitz [43]分析了在数字声音中用于信息隐藏的几种方法，其中，有最低比特位编码(LSB)、相位编码、扩展频谱技术(参见 3.4 节)以及回声隐藏等。最低比特位编码技术健壮性较差，但可以传送大量数据。相位编码对载体信号再取样有健壮性，但由于秘密信息仅在第一个信号片段进行编码，数据传输率很低。相反，扩展频谱和回声隐藏在许多方面有更好的表现。

3.3.4 信息隐藏和数据压缩

在有些情况下，数据压缩系统中融合了信息隐藏算法。可以想像，视频会议系统能在播放的视频图像流中隐藏信息。大多数研究工作是针对有损视频和图像压缩系统隐藏信息，但应该提到，无损压缩系统也能用于秘密信息传输。Cachin [44]展示了怎样通过修改 Willems[45]的"循环乘积"压缩算法构造一个渐近最优的信息伪装系统。

目前已经有了大量的压缩视频和图像的信息伪装系统。最简单的技术是应用 Jpeg-Jsteg [46]工具，以 JPEG 压缩系统中 DCT 系数取整方式隐藏信息(参见 3.3 节)。由于 DCT 对于整数输入其输出是非整数序列，JPEG 系统在编码处理前必须量化 DCT 系数，根据秘密信息位对系数向下或向上取整来隐藏信息。尽管这样一个系统不是健壮的，但检测载体的修改仍是困难的。Westfeld 和 Wolf 在[47]中描述了一个类似的技术，他们的系统作用在量化后的并且是 DCT 编码过的视帧块上。在对块进行适合秘密通信和不适合秘密通信的分类后，以传送秘密信息的方式(详情参见[47])修改块 DCT 系数的模 2 和。更复杂的方法结合了视频压缩和扩展频谱。例如，Hartung 和 Girod 在[48,49]中提出了一种信息隐藏方案，使用他们的扩展频谱水印系统(参见 6.4.1 节)对预压缩视频进行处理。

3.4 扩展频谱和信息隐藏

在20世纪50年代，为了实现一种拦截概率小、抗干扰能力强的通信手段，提出了扩展频谱(SS)通信技术。Pickholtz等[50]把扩展频谱技术定义为这样一种传输方式，"信号在大于所需的带宽内进行传输。带宽扩展是通过一个与数据独立的码字完成的，并且在接收端对这个码字的同步接收被用于解扩和随后的数据恢复"。尽管传输信号的能量可以很大，但在每一个频段上的信噪比很小。即使部分信号在几个频段丢失，其它频段仍有足够的信息可以用来恢复信号。因此，检测和(或)删除一个SS信号是很困难的。这种情况与伪装系统很相似，伪装系统就是试图在整个载体中扩展秘密信息，以达到不可觉察的目的。由于扩展信号很难删除，所以基于SS的隐藏方法都具有可观的健壮性。自从Trikel等人的开创性论文[51]发表后，扩展频谱方法在信息隐藏领域越来越受到重视。

在信息隐藏中，通常使用两个特殊的SS变体，即直接序列扩频和跳频扩频方案。在直接序列扩频方案中，秘密信息与一个伪随机序列调制，扩展倍数是一个称为片率的常量，然后叠加在载体上。另一方面，在跳频方案中，载体信号的频率从一个频率向另一个频率进行跳变。SS被广泛应用在水印中，这一点从6.4.1节中可见一斑。一个比较有趣的直接序列水印方法，由Hartung和Girod [48,49]发明，也能用于伪装目的，这将在6.4.1节描述。

由于SS水印和伪装算法的相似性，在这章我们把讨论限制在介绍信息隐藏中扩展频谱技术应用的数学模型上，并讨论一个称为SSIS的系统，作为例子进行研究。

3.4.1 一个扩展频谱模型

Smith和Comiskey[52]提出了一个扩展频谱伪装系统的一般框架。他们的方法最初使用$N \times M$的灰度级图像作为载体，而这种工作很容易扩展到所有载体的集合，这些载体的尺寸可以定义为两个标量数的乘积。假设Alice和Bob共享一组(至少)l (m)个正交的$N \times M$图像ϕ_i，把它们作为伪装密钥。Alice通过计算图像的加权和，首先产生一个伪装信息E(x,y)：

$$E(x,y) = \sum_i m_i \phi_i(x,y) \tag{3.16}$$

图像ϕ_i互相之间正交，

$$\langle \phi_i, \phi_j \rangle = \sum_{x=1}^{N} \sum_{y=1}^{M} \phi_i(x,y)\phi_j(x,y) = G_i \delta_{ij} \tag{3.17}$$

这里$G_i = \sum_{x=1}^{N} \sum_{y=1}^{M} \phi_i^2(x,y)$，$\delta_{ij}$是Kronecker δ函数。Alice然后计算载体C与秘密信息E这两个图像像素值的和，通过这种方法将秘密信息E编码到载体C中去，生成伪装载体S。

$$S(x,y) = C(x,y) + E(x,y) \tag{3.18}$$

理想情况下，C与所有的ϕ_i正交，(所以$\langle C, \phi_i \rangle = 0$)，并且Bob能通过把伪装图像$S$投影到第$i$个基础图像$\phi_i$上提取出第$i$个消息位$m_i$：

$$\begin{aligned} \langle S, \phi_i \rangle &= \langle C, \phi_i \rangle + \langle \sum_j m_j \phi_j, \phi_i \rangle \\ &= \sum_j m_j \langle \phi_j, \phi_i \rangle \\ &= G_i m_i \end{aligned} \tag{3.19}$$

因此，通过计算 $m_i = \langle S, \phi_i \rangle / G_i$ 就能恢复出秘密信息。注意：在解码过程中不需要原始图像 C。然而在实际中，C 并不完全与所有图像 ϕ_i 正交，所以在(3.19)中必须引入一个误差项 $\langle C, \phi_i \rangle = \Delta C_i$：

$$\langle S, \phi_i \rangle = \Delta C_i + G_i M_i \tag{3.20}$$

我们现在证明在合理假设条件下，ΔC_i 的期望值是 0。令 C 和 ϕ_i 是两个独立的 $N \times M$ 维随机变量，如果我们假定使用零均值随机过程创建所有基础图像，并且它们独立于要传送的消息，那么

$$E[\Delta C_i] = \sum_{x=1}^{N} \sum_{y=1}^{M} E[C(x,y)] E[\phi_i(x,y)] = 0 \tag{3.21}$$

因此，在这些假设条件下，(3.20)中误差项的期望值为 0 。

因此解码过程如下：将伪装图像 S 投影到所有基础函数 ϕ_i 上产生一个近似值

$$s_i = \langle S, \phi_i \rangle = \Delta C_i + G_i m_i \tag{3.22}$$

通过 s_i 来重构秘密信息。从上面条件已知，ΔC_i 的期望值是 0，所以 $s_i \approx G_i m_i$，最后一步就是从 s_i 重构 m_i。若我们将秘密信息编码为 1 和 -1 组成的串，而不是简单使用二进制串，那么 m_i 的值就可以利用 sign 函数进行重构，这里 $G_i \gg 0$，

$$m_i = sign(s_i) = \begin{cases} -1 & if \quad s_i < 0 \\ 0 & if \quad s_i = 0 \\ 1 & if \quad s_i > 0 \end{cases} \tag{3.23}$$

若 $m_i = 0$，表明编码信息已经丢失。在一些比较严重的情况下，$|\Delta C_i|$ 可能变得很大(回想我们已经证明的期望值是 0)，以至于不可能正确地恢复出一个比特。然而，这种情况很少发生，并且能通过采用纠错编码的方式来解决。

在伪装术方面采用扩展频谱技术的主要优点是对图像修改具有较好的健壮性。由于编码后信息扩散到整个频段，想不破坏载体完全删除它是很困难的。实际上，修改伪装载体将增大 ΔC_i 的值，只要不让 $|\Delta C_i| > |G_i m_i|$，这些修改都不会损坏嵌入信息。

3.4.2　SSIS——一个实例研究

Marvel 等[53]提出了一个称作 SSIS 的信息伪装系统，在这里我们把它作为一个实例简要地进行讨论。SSIS 使用扩展频谱技术作为嵌入函数，它的机理描述如下：嵌入处理前，使用传统对称密钥方案对秘密信息进行加密，使用的密钥记为 k_1。接下来，将采用低速率纠错编码对加密的秘密信息进行编码(例如 RS 码)。这一步将提高整个伪装系统的健壮性。然后用一个伪随机序列对获得的编码信息进行调制，这个伪随机序列由一个使用 k_2 作为种子的伪随机码发生器产生。处理后的信号(看似随机的)再被送进一个交织器(它使用 k_3 作为种子)，然后附加在载体上。最后一步，得到的伪装图像将进行适当的量化。

在接收方，提取与嵌入过程刚好相反。由于 SSIS 的设计目标之一是提供一种盲信息伪装系统，就是在秘密信息提取处理时不需要原始图像，所以它使用一种图像恢复技术来得到原始图像的一个估计，如采用自适应 Wiener 滤波器。从得到的原始图像估计中减去伪装图像，得到一个调制和扩展的伪装信息的估计。然后对获得的比特进行逆交织和解调(使用 k_3 和 k_2)。由于 Wiener 滤波器的性能较差，重构的秘密信息可能含有错误比特，因此这种盲信息伪

装系统可以看作是在一个噪音信道中的信息传输。无论如何，纠错编码的使用有益于恢复破坏的信息比特。最后，用 k_1 对秘密信息进行解密。

3.5 统计隐写术

"1－比特"伪装方案是在数字载体中嵌入一个比特，统计伪装技术就是以"1－比特"伪装方案为基础的。具体描述如下，若传送是"1"，就对载体的一些统计特性显著地进行修改，否则就对载体原封不动。所以接收者必须能区分哪些修改了和哪些没有修改。

为了用多个"1－比特"系统构造一个 $l(m)$－比特伪装系统，必须把载体分成不相连的 $B_1,\ldots,B_{l(m)}$ 块。若 m_i 等于 1，就在 B_i 中放一个"1"，否则在嵌入过程中不对块进行修改。每一比特的检测是通过一个测试函数，它能区分修改的载体块和未修改的载体块：

$$f(B_i)=\begin{cases}1 & \text{在嵌入过程修改了块 } B_i\\ 0 & \text{否则}\end{cases} \tag{3.24}$$

函数 f 可以看作是一个假设检验函数，我们对零假设"没有修改块 B_i"和 1 假设"修改了块 B_i"进行比较测试。因此，我们称这一大类的伪装系统为统计伪装系统。接收者为了恢复每一比特秘密信息，要对所有块连续地使用 f 函数。

如何构造(3.24)中的函数 f，这个问题仍待解决。如果我们把 f 解释成假设检验函数，我们就能运用数理统计中的假设检验理论。假设我们能找到一个公式 $h(B_i)$，它依赖载体块 B_i 中的一些元素，并且我们知道在未修改块中 $h(B_i)$ 的分布(也就是这种情况下假设成立)。那么我们能使用标准步骤测试 $h(B_i)$ 是否等于或超过某个特定值。如果我们在嵌入过程中若块 B_i 没有被修改，它的期望值为 0，否则它的期望值比较大，以这种方式控制修改 $h(B_i)$，那么我们能在给定 $h(B_i)$ 的分布情况下，检验 $h(B_i)$ 是否等于 0。

然而，统计伪装技术在许多情况下的应用是很困难的。首先，必须找到一个好的检验统计量 $h(B_i)$，它能区别修改的和未修改的块。另外，对"通常"的载体，必须知道 $h(B_i)$ 分布，这是相当困难的。在实际应用中，为了决定分布的相近公式，都要作出许多假设(很多假设值得怀疑)。

作为一个例子，根据 Pitas 的水印系统[54]，我们想构造一个统计伪装算法，这个算法与 Bender 等人在[2]中的拼凑方法很相似。假设每一个载体块 B_i 是像素 $P_{n,m}^{(i)}$ 的一个矩形集合，令 $S=\{s_{n,m}^{(i)}\}$ 是同样尺寸的矩形伪随机二值图案，并且在 S 中 1 的个数与 0 的个数相等。我们还认为发送者和接收者都能获得 S，在该应用中，S 代表伪装密钥。发送者首先把图像块 B_i 分成同样大小的两个集合 C_i 和 D_i(也就是把所有相应密钥比特 $s_{n,m}$ 等于 0 的像素点放入集合 C)：

$$\begin{aligned}C_i&=\{P_{n,m}^{(i)}\in B_i \mid s_{n,m}=1\}\\ D_i&=\{P_{n,m}^{(i)}\in B_i \mid s_{n,m}=0\}\end{aligned} \tag{3.25}$$

然后发送者对子集 C_i 的所有像素加上一个值 $k(k>0)$，而 D_i 中的像素不变。最后，合并 C_i 和 D_i，形成加了标记的图像块 B_i。

为了提取标记，接收者重构集合 C_i 和 D_i。若块中加了标记，在 C_i 中的所有值比在嵌入时的值大，因此，我们能测试集合 C_i 和 D_i 的均值差。如果我们假定在 C_i 和 D_i 中的所有像素是

独立同分布的随机变量，具体是什么分布可以任意，检验统计量为：

$$q_i = \frac{\overline{C_i} - \overline{D_i}}{\hat{\sigma}_i} \tag{3.26}$$

其中

$$\hat{\sigma}_i = \sqrt{\frac{Var[C_i] + Var[D_i]}{|S|/2}} \tag{3.27}$$

此处$\overline{C_i}$表示在集合C_i中所有像素的均值，Var[C_i]为C_i中随机变量的估计方差，根据中心极限定理，q_i渐近于服从$N(0,1)$正态分布。如果在一个图像块$\tilde{B}_i$中嵌入了一个标记，q_i的期望值将大于0。接收方因此能通过检验块B_i的统计量q_i在$N(0,1)$分布下是否为0来重构第i个秘密信息位。

3.6 变形技术

与替换系统相比，变形技术在解码时要求已知原始图像信息。Alice为得到一个伪装对象，对载体按某种次序进行修改，这种次序是她根据需要传送的秘密信息而定的。Bob为重构Alice相应于秘密消息采用的修改次序，必须测量与原来载体的差异。

在许多应用中，由于接收者必须要得到原始图像，所以这样的系统并不实用。若Wendy也能获得原来的载体，她就能很容易地检测到载体的修改，并且能获得秘密通信的证据。若嵌入和提取函数是公开的并且没有伪装密钥保护，Wendy也可能完全重构秘密信息。因此在本节中，我们假定原始图像能通过一个安全通道传送。

早期隐藏信息的一种方法是在文本中进行的。大部分基于文本的隐藏方法都属于变形类型（也就是利用单词的排列或文档的布局来隐藏信息）。一种技术是调整行和单词的位置，这将在下一小节详细讨论。另外，在文本中加入空格和“不可见”字符提供了一种传递秘密信息的方法。HTML文件能很好地包含额外的空格、制表符和换行点，Web浏览器忽略这些“额外”的空格和行，除非获得Web页的源代码才能发现它们。

3.6.1 在格式化文本中嵌入信息

把格式化文本解释成二值图像，在其中构建嵌入信息的方法，这方面已经做了很多的工作。Maxemchuk等在[55－58]中提出了一些基于文本伪装的方案，它们利用连续行之间或连续单词之间的距离来传送秘密信息。然而，应该提到的是，任何使用文本格式传送信息的伪装系统，都很容易通过重新录入文档而被破坏。

在行间距编码中，行的位置根据秘密信息位进行上移或下移，其它行保持不变，这是为了同步的要求（最初实现时，信息是在每两行中传送）。在移动过的一行中编码一个秘密信息比特，若一行是上移，编码为1，否则是0。当对秘密信息进行译码时，可以使用质心检测法，质心定义为水平轴上一行的中心。我们用Δ_{R+}表示移动行和上一个不动行质心之间的矩离，用Δ_{R-}表示移动行和下一个不动行质心之间的距离，并用Δ_{X+}和Δ_{X-}表示在未修改文档中相应的质心距离。若

$$\frac{\Delta_{R+} + \Delta_{R-}}{\Delta_{R+} - \Delta_{R-}} > \frac{\Delta_{X+} + \Delta_{X-}}{\Delta_{X+} - \Delta_{X-}} \tag{3.28}$$

上一行的距离被增大，类似地，若

$$\frac{\Delta_{R+}+\Delta_{R-}}{\Delta_{R+}-\Delta_{R-}}>\frac{\Delta_{X+}+\Delta_{X-}}{\Delta_{X+}-\Delta_{X-}} \tag{3.29}$$

上一行的距离被减小。注意如果在复制过程时，对页按某个常数因子进行缩放，但(3.28)和(3.29)式中的分数，可以补偿掉这个缩放因子。同样地，垂直打印密度也以近似的方式影响所有的质心。这些特性使得行间距编码技术能够抵御大部分变形攻击。这种嵌入技术的分析参见[57]。

在格式化文本中，另一个可能的嵌入方案是字间距编码，如图 3.3 所示。图中的竖线做为参考，数据嵌在第一个和第三个句子中。根据秘密信息位，改变在载体中选定的两个单词间的水平距离。理论上，在两个单词间任意的间隔都是可行的，唯一的限制是在具体行上所有移动的总和应等于 0，以保持行的正确排序。

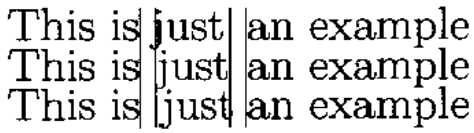

图 3.3 在单词间的空格中嵌入信息

3.6.2 数字图像变形技术

变形技术可以很容易地应用到数字图像上。使用与替换系统类似的方法，发送者首先选择想用于信息传输的 $l(m)$个不同的载体像素。这种选择同样可以通过伪随机数发生器或伪随机置换来实现。发送者若在像素中对 0 进行编码，则保持像素不变；若对 1 编码，则在像素的颜色值中加上一个随机值 Δx。尽管这种方法与替换系统相似，但是却有一个显著的区别，即所选颜色值的 LSB 不必等于秘密信息比特。尤其当对 0 编码时，不需要对载体进行修改。更进一步地，可以适当选择 Δx，使图像可以更好地保持载体的统计特性。接收方用伪装对象中选择的所有 $l(m)$个像素与原始载体中相应像素进行比较，若第 i 个像素不同，则第 i 个消息位是 1，否则是 0。

以上方法在实现上有许多变种，类似 3.2.4 节提到的奇偶校验位方法，某一图像区域的奇偶位，根据信息位是 1 或 0 而进行修改或保持不变。进一步地可以对某个图像区域应用图像处理技术，使得所做修改对观察者不可见。

Sandford 等[59,60]引入了另一种图像变形技术——数据嵌入。与迄今讨论的所有变形技术相比，数据嵌入是试图修改载体中冗余数据出现的顺序，而不是修改它们本身的值。因此嵌入处理保存一个“对表”(也就是相互差别小于一个给定门限值的一对样点的列表)。接收方若能获得“对表”，就能对嵌入过程逆操作。这个表可看成与密码中的密钥相似，接收者通常不能根据载体恢复出这个列表(详情参见[59])。

3.7 载体生成技术

回想以上提到的所有方法，都是通过使用一种嵌入算法，将秘密信息加进一个具体的载体中去。而还有一些伪装技术是生成一个数字对象作为载体进行秘密通信的。

3.7.1 模拟函数

由于信息传输量的爆炸性增长，人类不可能观察世界上所有的通信，这样的任务只能由自

动监督系统完成，因此自动监督系统的重要性日益突出，这一点已在第二章的结论中提到。这些系统通过检查关键字和消息的统计特性来核查通信。例如，由于加密信息和未加密信息具有不同的统计特性，对它们进行区分是可能的。Wayner [61]提出的模拟函数是通过更改消息的统计特性，使它与任何正常文本的特性相匹配，并以此来隐藏一个秘密消息的特征。

众所周知，英语拥有几个统计特性。例如，字母的分布是不均匀的（参见[62]附录中英文两字母组和三字母组的分布频率）。这种事实已在大量数据压缩技术中得到应用（例如哈夫曼编码系统[40]）。给定一个字母表$\sum$和一个概率分布 A，能用哈夫曼编码产生一个最小冗余压缩函数 $f_A:\sum\to\{0,1\}^*$，此处 * 表示 KleeneStar（$\sum^*=\bigcup_{i\geq 0}\{x_1\cdots x_i \mid x_1,\cdots,x_i\in\sum\}$）。我们能使用两个哈夫曼压缩函数构造一个模拟函数 $g:\sum^*\to\sum^*$。

$$g(x)=f_B^{-1}(f_A(x)) \tag{3.30}$$

这个模拟函数能把一个消息（其字母的概率分布为 A）转换成另一个消息（其字母的统计特性为 B）。因此，首先对分布为 A 的文件 x 使用哈夫曼系统压缩，产生一个二进制串文件，这个二进制串又能看成是另一个文件（其分布为 B）经哈夫曼压缩系统后输出的结果。这另一个文件将用作伪装对象，并能通过对二进制串文件运用一个逆哈夫曼压缩函数 f_B^{-1} 进行重构。由于 f_A 和 f_B 是一对一的，所以构造的模拟函数也将是一对一的。Wayner 证明在 f_A 是一个理论最优的哈夫曼压缩函数时，并且在 x 是一个随机比特文件的意义下，这个模拟函数 $g(x)$ 是最优的，并且 $f_A^{-1}(x)$ 也是统计分布 A 最好的近似，$f_A^{-1}(x)$ 也是一对一的。

不使用单个字符分布，而是基于 n 个字符的统计分布频率能构造成一次压缩 n 个字符的哈夫曼编码系统。可是，由哈夫曼系统创建的压缩树随 n 呈指数增长。作为替代，Wayner 又提出通过为每个长 $n-1$ 的字符串 t 创建哈夫曼压缩函数来利用字符之间的依赖性，对文件中可能跟随在 t 后的每个字符的概率进行编码。一个模拟函数能从这些哈夫曼压缩函数的集合中构造出来（详情参见[61]）。

3.7.2　英语文本的自动生成

无论如何，模拟函数仅能用于欺骗机器。由于伪装对象仅根据统计特性创建，完全忽略了语义成分，就人看来，创建的文本完全无意义并且充满了语法和印刷错误。

为解决这个问题，提出了使用自由上下文语法（CFG）。关于 CFG 理论的描述参见[63]。令 $G=<V,\sum,\prod,S>$ 是一个 CFG，这里 V 是一个变量集，$\sum$ 是终止符号集，$\pi\subseteq V\times(\bigcup V\sum)^*$ 是叉积并且 $S\in V$ 是起始符号。叉积可以看成是一个替代规则，它们把一个变量转化成包含结束符号或变量符号的字符串。根据 $\prod$，如果能从起始符 S 开始，通过用终止符号或可变符号序列替换变量连续地生成 s，那么就能由 G 生成一个由终止符号序列定义的字符串 $s\in\sum^*$（正式地：$s\in L(G)$）。例如，令

$$\begin{aligned}\prod=\{&S\to\text{Alice }B,\ S\to\text{Bob }B,\ S\to\text{Eve }B,\ S\to\text{I }A,\\&A\to\text{am working},\ A\to\text{am lazy},\ A\to\text{am tired},\\&B\to\text{is }C,\ B\to\text{can cook},\\&C\to\text{reading},\ C\to\text{sleeping},\ C\to\text{working}\}\end{aligned}$$

我们能从语法$\langle\{S,A,B,C\},\{A,\ldots,Z,a,\ldots,z\},\prod,S\rangle$中派生出句子 I am lazy，Alice is reading 等。若对于每一个字符串 $s\in L(G)$，恰好存在一种从起始符号开始生成 s 的方式，那么

说明语法是明确的。

明确的语法能用作伪装工具。Wayner [61,64]提出一种对模拟函数的扩展。给定一个叉积集合,对变量 V_i 的每一个可能的结果我们指定一个概率。在我们上面例子中,我们能选择

$$\begin{aligned}\Pi = \{&S \rightarrow_{0.5} \text{Alice } B,\ S \rightarrow_{0.3} \text{Bob } B,\ S \rightarrow_{0.1} \text{Eve } B,\ S \rightarrow_{0.1} I\ A,\\ &A \rightarrow_{0.3} \text{am working},\ A \rightarrow_{0.4} \text{am lazy},\ A \rightarrow_{0.3} \text{am tired},\\ &B \rightarrow_{0.5} \text{is } C,\ B \rightarrow_{0.5} \text{can cook},\\ &C \rightarrow_{0.5} \text{reading},\ C \rightarrow_{0.1} \text{sleeping},\ C \rightarrow_{0.4} \text{working}\}\end{aligned}$$

并让 $\Pi_{V_i} = \{\pi_i, 1, \cdots, \pi_{i,n}\}$ 是与变量 V_i 相关的所有叉积的集合。然后发送者对每一个集合 Π_i 构造一个哈夫曼压缩函数 f_{Π_i}。在图 3.4 中显示的是 Π_S 和 Π_A 可能的哈夫曼树。哈夫曼压缩函数能很容易地从这些树中派生出来。例如,结果"Eve B"编码成 110,"A am tired"编成 11 等。

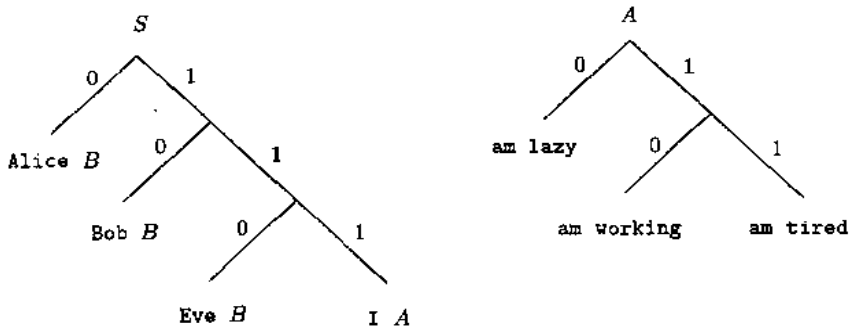

图 3.4 Π_S 和 Π_A 的哈夫曼压缩函数

就伪装目的而言,需要使用逆哈夫曼压缩函数。在编码阶段,发送者根据 CFG 派生出一个具体的字符串,作为伪装对象。从开始符号 S 开始,最左的变量通过一个叉积进行修改,这个叉积由秘密信息和 Π_{V_i} 的哈夫曼压缩函数决定。具体地说,根据秘密信息的下一个比特来回移动哈夫曼树,直到达到一个树节点为止,然后用这个节点上找到的结果替换掉开始符号。重复这种处理(即,最左的变量和一个结果进行交换,这个结果由哈夫曼树和后几个消息比特决定),直到使用了所有的信息比特并且字符串仅由结束字符组成为止。再来看前面的例子,假设秘密信息是 11110。在第一步时,我们来回移动 Π_S 哈夫曼树,并且通过使用前 3 个消息比特最终找到节点"I A"。结果,起始符号 S 替换成"I A"。然后我们来回移动哈夫曼树 Π_A 并且使用另外两个秘密信息比特找到替换位置"am working"。由于派生字符串仅由终止字符组成并且使用了所有秘密消息位,结果描绘 11110 的伪装对象是句子 I am working。

在解码过程中,为了重构在嵌入过程使用的结果,需要对载体进行分解。对于给定的 CFG,通过使用一个分析树能成功地进行译码,参见 [65]。由于结果唯一地决定了秘密信息,并且潜在的语法是明确的,所以接收者能重构秘密信息。

Chapman 和 Davida[66]提出了一个相似的系统。他们的系统由两个函数组成,NICETEXT 和 SCRAMBLE。给定一个按不同类型分类的大字典,和一个描述怎样将不同类型的单词组成有意义句子的类型源。NICETEXT 从字典中选择单词将秘密信息比特转换成句子,并且这个句子符合类型源中给定的句子结构。如果知道使用的字典,SCRAMBLE 就能重构秘密信息。类型源可以从自然语言句子的取样中产生,也可以用 CFG 产生。

3.8 结 论

本章中我们回顾了过去几年在文献中提到的不同伪装方法，以及在有噪声通信信道中的许多灵活、简单的伪装方法。

然而，伪装载体和消息趋向于具有某种独特模式，而且这种模式能被信息伪装分析者利用。通过仔细分析通道噪声的统计特性，能够攻破大多数简单的伪装系统。图像和其它信号容易进行量化、过滤、变换、格式转换等，它们大多数能在数据中留下某种"指印"。设计一个伪装系统时，都必须认真面对这些问题。利用这些特点对秘密信息进行攻击的方法将在下一章介绍。

参考文献

[1] Foley, J., et al., Computer Graphics, Principles and Practice, Reading, MA: Addison Wesley, 1990

[2] Bender, W., D. Gruhl, And N. Morimoto, "Techniques for data hiding," IBM Systems Journal, vol. 35, no. 3/4, 1996, pp. 131 – 336.

[3] Möller, S., A. Pfitzmann, and I. Stirand, "Computer Based Steganography: How It Works and Why Therefore Any Restrictions on Cryptography Are Nonsense, At Best," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 7 – 21.

[4] Gruhl, D., A. Lu, and W. Bender "Echo Hiding," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 295 – 316.

[5] Kurak, C., and J. McHughes, "A Cautionary Note On Image Downgrading," in IEEE computer Securitn Applications Conference 1992, Proceedings, IEEE Press, 1992, pp. 153 – 159.

[6] Van Schyndel, R. G., A. Tirkel, and C. F. Osborne, "A Digital Watermark," in Proceedings of the IEEE International Conference on Image Processing, vol. 2, 1994, pp. 86 – 90.

[7] Johnson, N. F., and S. Jajodia, "Exploring Steganography: Seeing the Unseen," IEEE Computer, vol. 31, no. 2, 1998, pp. 26 – 34.

[8] Gerzon, M. A., and P. G. Graven, "A High – rate Buried – Data Channel for Audio CD," Journal of the Audio Engineering Society, vol. 43, no. 1/2, 1995, pp. 3 – 22.

[9] "StegoDos – Black Wolf' sPicture Encoder v0.90B," ⟨ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/stegodos.zip⟩, 1993.

[10] Brown, A., "S – Tools for Windows," < ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/s – tools4.zip >, 1996.

[11] Hastur, H., "Mandelsteg," ⟨ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/steg.tar.z⟩, 1994

[12] Machado, R., "EzStego, Stego Online, Stego," ⟨http://www.stego.com⟩, 1997.

[13] Maroney, C., "Hide Seek," "⟨ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/hdsk41b.zip⟩, ⟨http://www.rugeley.demon.co.uk/security/hdsk50.zip⟩, 1994 – 1997.

[14] Repp, H., "Hide4PGP," ⟨http://www.rugeley.demon.co.uk/security/hide4pgp.zip⟩, 1996.

[15] Arachelian, R., "White Noise Storm," ⟨ftp://ftp.csua.berkeley.edu/pub/cypherpunks/

steganography/wns210.zip〉.1994.

[16] Hansmann, F., "Steganos, Deus Ex Machina Communications." < http://www.steganograhpy.com/ > ,1996.

[17] Menezes, A.J., P.C. van Oorschot, and S.A. Vanstone, Handbook of Applied Cryptography, Boca Raton: CRC Press ,1996.

[18] Aura, T., "Practical Invisibility in Digital Communication," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 265 – 278.

[19] Luby, M., and C. Rackoff, "How to Construct Pseudorandom Permutations from Pseudorandom Functions," SIAM Journal on Computation, vol. 17, no. 2, 1988, pp 373 – 386.

[20] Naor, M. and O. Reingold, "On the Construction of Pseudorandom Permutations: LubyRackoff Revisited," Journal of Cryptology, vol. 12, no. 1, 1999, pp. 29 – 66.

[21] Bell, D.E., and L.J. LaPadula, "Secure Computer Systems: Matematical Foundations," Mitre Report ESD – TR – 73 ~ 278(Vol. I – III), Mitre Corporation, Bedford, MA, Apr. 1974.

[22] Fridrich, J., "A New Steganographic Method for Palette – Based Images," in Proceedings of the IS&T PICS conference, Savannah, Georgia, Apr. 1998, pp. 285 – 289.

[23] Matsui, K., and K. Tanaka, "Video – Steganography: How to Secretly Embed a Signature in a Picture," IMA Intellectual Property Project Proceedings, vol. 1, no. 1, 1994, pp. 187 – 205.

[24] Baharav, Z., and D. Shaked, "Watermarking of Dither halftoned Images," in Proceedings of the SPIE 3657, Security and Watermarking of Multimedia Content, 1999, pp. 307 – 316.

[25] Zhao, J., and E. Koch, "Embedding Robust Labels into Images for Copyright Protection," in Proceedings of the International Conference on Intellectual Property Rights for Information, Knowledge and New Techniques, Munchen, Wien: Oldenbourg Verlag, 1995, pp. 242 – 251.

[26] "CCITT Recommendation T6: Facsimile Coding Schemes and Coding Control Functions for Group 4 Facsimile Apparatus for Document Transmission," 1984.

[27] Anderson, R.J., R. Needham, and A. Shamir, "The Steganographic File System," in Proceedings of The Second International Workshop on Information Hiding, V01/1525 of Lecture Notes in Computer Scienc, Springer, 1998, pp. 73 – 82.

[28] "ScramDisk: Free Hard Drive Encryption For Windows 95&98," 〈http://www.scramdisk.clara.net〉, 1998

[29] Handel, T.G., and M.T. Sandford, "Data Hiding in the OSI Network Model," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 23 ~ 38.

[30] Cox, I., et al., "A Secure, Robust Watermark for Multimedia," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 185 – 206.

[31] Koch, E., and J. Zhao, "Towards Robust and Hidden Image Copyright Labeling," in IEEE Workshop on Nonlinear Signal and Image Processing, Jun. 1995, pp. 452 – 455.

[32] Koch, E., J. Rindfrey, and J. Zhao, "Copyright Protection for Multimedia Data," in Proceedings of

the International Conference on Digital Media and Electronic Publishing, Leeds, UK, Dec. 1994.

[33] Ó Runaidh, J. J. K., F. M. Boland, and O. Sinnen, "Watermarking Digital Images for Copyright Protection," in Electronic Imaging and the Visual Arts, Proceedings, Feb. 1996.

[34] Xia, X., C. G. Boncelet, and G. R. Arce, "A Multiresolution Watermark for Digital Images," in Proceedings of the IEEE International Conference on Image Processing(ICIP'97), 1997.

[35] Rhodas, G. B., "Method and Apparatus Responsive to a Code Signal Conveyed Through a Graphic Image," U. SPatent 5,710,834, 1998.

[36] Swanson, M. D., B. Zhu, and A. H. Tewfik, "Transparent Robust Image Watermarking," in Proceedings of the IEEE International Conference on Image processing, vol. 3, 1996, pp. 211 – 214.

[37] Langelaar, G., J. van der Lubbe, and R. Lagendijk, "Robust Labeling Methods for Copy Protection of Images," in Proceedings of the SPIE vol. 3022, Storage and Retrieval for Image and Video Databases V, 1997, pp. 298 – 309.

[38] Pennebaker, W. B., and J. L. Mitchell, JPEG Still Image Compression Standard, New York: Van Nostrand Reinhold, 1993.

[39] Wallace, G. K., "The JPEG Still Picture Compression Standard," Communications of the ACM, vol. 34, no. 4, 1991, pp. 30 – 44.

[40] Huffman, D. A., "A Method for the Construction of Minimum – Redundancy Codes," Proceedings of the IRE, vol. 40, no. 10, 1952, pp. 1098 – 1101.

[41] Smoot, S., and L. A. Rowe, "DCT Coefficient Distributions," in Proceedings of the SPIE 2657, Human Vision and Electronic Imageing, 1996, pp. 403 – 411.

[42] Moore, B. C. J., An Introduction to the Psychology of Hearing, London: Academic Press, 1989.

[43] Chang, L., and I. S. Moskowitz, "Critical Analysis of Security in Voice Hiding Techniques," in Proceedings of the International Conference on Information and Communications Security, vol. 1334 of Lecture Notes in Computer Science, Springer, 1997, pp. 203 – 216.

[44] Cachin, C., "An Information – Theoretic Model for Steganography," in Proceedings of the Second International Workshop on Information hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 306 – 318.

[45] Willems, F. M. "Universal Data Compression and Repetition Times," IEEE Transactions on Information Theory, 1989, pp. 337 – 343.

[46] Upham, D., "Jpeg – Jsteg, modification of the independent JPEG group's JPEG software(release 4) for 1 – bit steganography in JFIF output files," < ftp://ftp.funet.fi/pub/crypt/steganography/ >, 1992 – 1997.

[47] Westfeld, A., and G. Wolf, "Steganography in a Video Conferencing System," in *Proceedings of the Second International Workshop on Insformation Hiding*, vol. 1525 fo *Lecture Notes in Computer Science*, Springer, 1998, pp. 32 – 47.

[48] Hartung, F., and B. Girod, "Copyright Protection in Video Delivery Networks byWatermarking of Pre – Compressed Video." In Multimedia Aplications, Services and Techniques – ECMAST 97, vol. 1242 of Lecture Notes in Computer Science, Springer, 1997, pp. 423 – 436.

[49] Hartung, F., and B. Girod, "Watermarking of Uncompressed and Compressed Video," Signal Pro-

cessing, vol.66, no.3, 1998, pp.283 – 301.

[50] Pickholtz, R.L., D.L. Schilling, and L.B Milstein, "Theory of Spread – Spectrum Communications – A Tutorial," IEEE Transactions on Communications, vol.30, no.5, 1982, pp.855 – 884.

[51] Tirkel, A.Z., G.A. Rankin, and R. van Schyndel, "Electronic Watermark," in Digital Image Computing, Technology and Applications – DICTA 93, Macquarie University, 1993, pp.666 – 673.

[52] Smith, J., and B. Comiskey, "Modulation and Information Hiding in Images," in Information Hiding: First International workshop, Proceedings, vol.1174 of Lecture Notes in Computer Science, Springer, 1996, pp.207 – 227.

[53] Marvel, L.M., C.G. Bonclet, and C.T. Retter, "Reliable Blind Information Hiding for Images," in Proceedings of the Second International Workshop on Information Hiding, vol.1525 of Lecture Notes in Computer Science, Springer, 1998, pp.48 – 61.

[54] Pitas, I., "A Method for Signature Casting on Digital Images," *in International Conference on Image Processing*, vol.3, IEEE Press, 1996, pp.215 – 218.

[55] Maxemchuk, N.F., "Electronic Document Distribution," At&T Technical Journal, Septermber/October 1994, pp.73 – 80.

[56] Low, S.H., et al., "Document Marking and Identification Using Both Line and Word Shifting," in Proceeedings of Infocom'95, 1995, pp.853 – 860.

[57] Low, S.H., N.F. Maxemchuk, and A.M. Lapone, "Document Identification for Copyright Proctection Using Centroid Detection," IEEE Transactions on Communications, vol.46, no.3 1998, pp. 372 – 383.

[58] Low, S.H., and N.F. Maxemchuk, "Performace Comparison of Two Text Marking Methods," IEEE Journal on Selected Areas in Communications, vol.16, no.4, 1998, pp.561 – 572.

[59] Sandford, M.T., J.N. Bradley, and T.G. Handel, "data Embedding Method," in Proceedings of the SPIE 2615, Integration Issues in Large Commercial Media Delivery Systems, 1996, pp.226 – 259.

[60] Sandford, M.T, T.G. Handel, and J.M. Ettinger, "Data Embedding in Degenerate Hosts," Technical Report LA – 95 – 4446UR, Los Alamos National Laboratory, 1996.

[61] Wayner, P., "Mimic Functions," Cryptologia, vol.XVI/3, 1992, pp.193 – 214.

[62] "Basic Cryptanalysis," Headquarters Department of the Army, Field Manual NO 34 – 40 – 2, ⟨ftp://ftp.ox.ac.uk/cryptanalysis/basic-cryptanalysis.ps.tar.gz⟩.

[63] Hopcroft, J.E., and J.D. Ullman, Introduction to Automata Theory, Languages and Computation, Reading, MA: Addison Wesley, 1979.

[64] Wayner, P., "Strong Theoretical Steganography", Cryptologia, vol.XIX/3, 1995, pp.285 – 299.

[65] Aho, A., R. Sethi, and J. Ullman, Compilers: Principles, Techinques and Tools, Reading (MA): Addison Wesley, 1986.

[66] Chapman, M., and G. Davida, "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," in Proceedings of the International Conference on Information and Communications Security, vol.1334 of Lecture Notes in Computer Science, Springer, 1997, pp.335 – 345.

第四章　隐写分析

(Neil F. Johnson)

隐写术(伪装术)主要研究隐匿消息存在性的各种通信方法。如第三章提到的一样,这些消息的数字载体可以是无关紧要的图像、声音、视频、文本或任何其它数字表示的代码或传播载体。隐藏的消息可以是明文、密文、或能看成比特流的任何东西。为了减少因嵌入消息引起的视觉变形,已设计出了许多有创造性的方法。

用电子媒介作为载体,隐藏信息要更改载体的属性,这可能引起某种形式的降质。若施加在图像上,有时人眼可以察觉到降质[1],并可能指出所用的伪装方法和工具的特征。实际上,这些特征已经说明了嵌入消息的存在,结果导致伪装失败。

4.1　隐写分析简介和术语

所有的伪装术和数字水印技术都能描述成下面的简单公式。在一幅图像中存在一个人眼不敏感性测度,即能根据人的感觉特性把图像的信息量分成两部分。一部分信息量记作 t,对这部分信息量处理时不会引起感觉上的降质;另一部分信息量记作 p,对它操作时会引起可感知的降质。那么一个可用于信息隐藏的载体(C)的信息量的等式是:

$$C = p + t \tag{4.1}$$

对伪装系统的用户和想破坏 t 中隐藏信息的攻击者来说,都能已知 t 的大小。只要 t 属于不可觉察区域,那么存在攻击者所使用的某个 t',使得 $C' = p + t'$ 并且 C 和 C' 间没有明显的差别。这种攻击可用于擦除或替换掉区域 t。为伪造数字水印,在文献[2]中就利用了这种攻击的一个变体。如果秘密信息以某种方式隐藏在某种媒体中,致使攻击者不能检测到秘密信息,那么攻击者可以以同样的阈值加入或去掉一些另外的信息,这将覆盖或删掉嵌入的秘密信息。在载体比较敏感的区域中嵌入信息,能更好地抵抗攻击,但是由于存在隐藏信息导致的人为痕迹,反而暴露了信息隐藏。尽管人类感觉系统不容易察觉到某种程度的变形和降质,但它确实存在。这种变形对"正常"载体来说是异常的,在被发现的时候,它就能够说明隐藏信息的存在。不同的隐藏工具使用着不同的隐藏信息的方法,不了解所使用的工具和伪装密钥(如果有的话),进行信息检测是很复杂的。然而,一些隐藏方法具有某些固有的特征,这些特征可用来标识所用的方法和工具。

伪装术的目标是避免传送秘密信息时引起怀疑,从而使秘密信息不可检测。若引起了怀疑,那么就说明隐藏失败了。隐藏分析是发现隐藏的消息并使这些消息无效的一种技术。

对隐藏信息的攻击和分析可能有几种形式,即检测、提取、混淆(攻击者在存在的隐藏信息上进行伪造或覆盖),使隐藏信息无效。这里,我们的目标不是提倡删除或使正确的隐藏信息无效(如版权),而是指出哪些方法是脆弱的和哪些方法可以用来调查非法隐藏信息。

任何载体无论是否存在一个嵌入消息,都能按破坏或使某些隐藏信息失效的方式进行操作。检测隐藏信息的存在性,使我们能在这个阶段仅处理包含隐藏信息的载体,节省了时间。

继续论述之前，我们应了解一下有关攻击和破解伪装方案的新术语。它们与密码学中的术语相似，但仍有一些重要的不同点。正如密码破译者使用密码分析试图破译加密信息一样，伪装分析者应用伪装分析（隐写分析）试图检测隐藏信息的存在。在密码分析中，是对部分明文（也可能没有明文）和部分密文进行分析。在伪装分析中，是在载体对象、伪装对象和可能的部分消息之间进行比较。在密码学上最终结果是密文，而在伪装术上最终结果是伪装对象。在伪装术上隐藏的信息可以加密也可以不加密，若隐藏信息是加密的，那么即使是隐藏信息被提取出来了，为了明白嵌入信息，也还需要应用密码破译技术。

为了定义用于伪装分析的攻击技术，可以参考在密码破译中相对应的技术。对密码破译者来说，可以使用的攻击有唯密文攻击、已知明文攻击、选择明文攻击、和选择密文攻击。在唯密文攻击中，密码破译者知道加密的密文。在已知明文攻击中，密码破译者可能有加密的消息和部分解密的消息。选择明文攻击是对密码破译者最有利的情况，在这种情况下，密码破译者可以任意选择一些明文以及所对应的密文。如果再能获得加密算法和密文，密码破译者就能加密明文，然后在密文中进行匹配。选择密文攻击被用于推测发送者的密钥。密码学的难点不是检测到已加密的信息，而是破译出加密的信息。

对伪装分析专家来说，可定义一些类似的攻击：

- **唯伪装对象攻击** 只可获得伪装对象进行分析。
- **已知载体攻击** 可以获得原始的载体和伪装对象。
- **已知消息攻击** 在某种意义上，攻击者可以获得隐藏的消息。针对系统，分析伪装对象相应隐藏信息的模式有利于将来的攻击。即使已知消息，同样是非常困难，甚至可以认为难度等同与唯伪装对象攻击。
- **选择伪装对象攻击** 知道伪装工具（算法）和伪装对象。
- **选择消息攻击** 伪装分析专家用某个伪装工具或算法对一个选择的消息产生伪装对象。这个攻击的目标是确定伪装对象中相应的模式特征，它可以用来指出具体使用的伪装工具或算法。
- **已知伪装载体和伪装对象攻击** 已知伪装算法（工具），并且可获得原始载体和伪装对象。

即使假定攻击者有最好的攻击条件，提取嵌入的信息仍然是困难的。有时，方法根本不是攻击算法和图像，而是攻击用于加密的口令或选择隐藏信息的比特。“蛮力”攻击对某些伪装工具是成功的，但要达到满意的效果，仍需大量的处理时间[3,4]。

4.2 寻找特征——检测隐藏信息

一些不常规的模式可能会显现出来，它暴露了信息隐藏的事实。在文本中，对粗心的观察者来说，对单词和行间距中一些小的移动要检测出来有些困难[5]。然而，添加的空格和“不可见的”字符用常规的 Word 处理器打开文件时很容易显示出来。在屏幕上显示时，文本看起来是“正常”的，但当用 Word 处理器打开时，空格、制表符和其它字符就会破坏文本的表达。

磁盘上未使用的区域能用于隐藏信息，但有许多有用的磁盘分析工具，它们能报告和过滤存储设备中未使用的簇成分区。由于通过对系统分区信息的分析，一个伪装文件系统也是容易检测的。

过滤器也能用于捕获 TCP/IP 包，TCP/IP 包的包头中可包含隐藏的或非法的信息。因特网防火墙变得越来越复杂，并允许更多的用户定制。正如能设置过滤器来判断数据包是否来自防火墙的域内，以及 SYN 和 ACK 比特是否有效一样，过滤器也能配置成能够捕获那些在未用或保留空间上有信息的数据包。

多媒体为隐藏信息提供了极好的载体。然而，隐藏信息可能造成一些失真，选择合适的伪装工具和载体是成功隐藏信息的关键。甚至隐藏少量的信息，就可能使声音失真和图像总体上降质。“可觉察的噪音”能够泄露隐藏信息的存在。回声和阴影信号能减少听见噪音的可能性，但是它们只需少量的处理就能检测到。

在评估伪装术中使用的许多图像时，就看标示隐藏信息的特征是否变得很明显。这些特征可以是调色板的不常规的排序，也可以是颜色索引中两种颜色间的关系，或者是夸大的“噪音”。一种用于识别这些模式的方法，是对载体图像和伪装图像进行比较，并注意视觉上的差异(已知载体攻击)。当对载体图像和伪装图像进行比较时，很容易注意到细微的变化。例如，在[6]中能找到用于图像的伪装工具所具有的独特特征。

伪装工具主要用于大信息量的隐藏，而水印工具是在一幅图像中隐藏较少信息，但水印是冗余分布在整个图像上的[7]。任何情况下，这些方法都要以保持不可见的方式来嵌入信息和操作图像。然而，对图像的任何操作都会引入一定量的失真，在“原始”图像特性的某个方面引起降质(一些基本嵌入技巧是在头或其它图像浏览器忽略的“未使用”区域内放置信息。这避免了图像的降质，但在比特分析中能够很容易检测到)。

4.2.1　基于调色板的图像

在开始评估一些附加或隐藏了信息的图像之前，需要定义一个“正常”或一般的图像的概念。考虑到数字照片、绘画、素描、图形等各种可能性，要定义一个正常图像确实有些困难。只有在测评许多原始图像和伪装图像关于颜色组成、亮度和像素间关系后，异常的东西才能说明在别的图像中是不“正常”的特征。选择消息和已知载体攻击对检测这些特征是相当有用的。在带调色板或颜色索引的图像中，颜色主要按使用最多到使用最少进行排序，以减少查询时间。颜色值之间可以逐渐改变，但很少以一比特增量方式变化。灰度图像颜色索引是以 1 比特增长的，但所有的 RGB 值是相同的。将类似的方法应用于单色图像，通常 RGB 值中有两个相同，第三个一般有更强的颜色饱和度。一些图像，如，手工素描、分形或裁剪艺术可能在相邻像素颜色值间变化很大。然而，出现单个像素的显著突出，就可能说明存在隐藏信息。

一些图像中加入的内容可看作是夸大的噪音，这是许多采用 8 位图像的位平面工具的一个共有的特征。在许多情况下，使用 8 位图像而又不修改调色板，将会导致颜色随着光栅指针从一个调色板条目改到另一个而发生改变。如果调色板相邻条目的颜色非常接近，那么几乎很少或没有可觉察的变化。然而，如果调色板相邻条目不是颜色相近的，那么 LSB 操作明显地引入了噪音[8]。基于这个原因，伪装软件的许多作者强调使用灰度级图像(具有 256 级灰度)[9]。灰度图像是 8 位图像的特殊情况，并且由于灰度值从一个调色板条目到另一个是逐渐变化的，所以是很好的载体。

在[8]中，作者建议使用相邻调色板条目有很大对比度的图像来挫败伪装软件。在 8 位图像中，像素值的小变化将造成图像颜色很大的变化，同时也在宣告存在隐藏信息。若不修改 8 位图像调色板，在光栅数据中 LSB 的改变可以显示出伪装图像的显著变化。

一些位平面工具通过调色板排序企图减少这种影响[10,11],但即使调色板只含少量的颜色,仅靠排序调色板来防止暴露消息的存在是不够的。一些工具在这方面更进了一步,创建新的调色板[11-13][1]。将8位图像转换为24位图像,可以直接读取需要改变的颜色值,任何变化是几乎无法在视觉上感知的。它的缺点是使图像的尺寸变大许多。一个可能的解决办法是在LSB隐藏信息后,将图像转换回8位图像。即使显著地更改图像调色板的颜色,这种方法仍可以隐藏一个消息存在的事实。

由于在图像调色板中,8位图像限制为256个不同的颜色条目,必须考虑图像使用的不同颜色数。例如,若一个24位图像包括200个独立的颜色并且信息隐藏在LSB中,那么独立的颜色数很容易达到300个,其中可能有两种不同的颜色相差一个比特。再缩减到8位图像时,调色板变成256色,那么将很容易丢失某些新的颜色。

解决这种问题的一个方法是将颜色数减少到某个值,而仍能保持很好的图像质量,并确保颜色数不超过256。这种方法应用在[14,15]中,并且能将颜色数减少到不多于32个。32种颜色通过在调色板中加入相邻颜色能扩展到8个调色板条目,这个调色板与原始颜色非常接近[12]。这种方法生成的伪装图像与原始载体图像也是非常接近的,以至于几乎没有视觉上的差异。无论如何,这种方法也创建了一种有待进一步研究的独特模式。

4.2.2 图像失真和噪音

在伪装图像中检测隐藏信息存在性的一种方法是寻找明显的和重复的模式,它可能指示出伪装工具或隐藏消息的身份或特征。人眼可见的失真或模式是易于检测的。用于验证这些模式的一种方法是将原始载体和伪装图像进行比较,并注意视觉上的差别(已知载体攻击和已知伪装对象攻击)。如果不利用载体图像,这种噪音可能会看作构成图像整体的一部分而不引起注意。其中某些"特征"可自动地用来识别隐藏信息的存在,甚至识别嵌入信息所用的工具。有了这种知识基础,即使无法获得载体图像进行比较,派生的已知特征已足够去提示消息的存在,并能识别出所用的嵌入工具。然而在某些情形下,即使能注意到载体和伪装图像之间的失真,那些可预测的模式也并不是显而易见的。

在[1]中所讨论的图像明显变质是一种失真。在[8]中,通过使相邻调色板条目对比突出,而建立了一套测试图像。一些工具,尤其是那些作用在位平面集上的工具,制作出的伪装图像带有严重失真和噪声[10,16,17]。这些失真是过分的颜色替换造成的,它们宣告了一个隐藏信息的存在性。通过研究像素的"邻近关系",和判断一个不协调的像素是否对该图像是正常的,是否遵循一种模式,或是否类似噪音来检测这种特性。

不是所有的位平面工具都产生这类图像失真。好几种位平面软件和那些变换域中的隐藏工具在嵌入信息后,伪装图像没有可视的失真。尽管这些工具通过了这种测试,但却会出现其它特征模式。

带颜色调色板或索引的8位彩色和灰度图像在视觉上更容易分析。那些在"论文里"提供很好结果的工具也可能有数字特征,这些特征能够使消息的存在性被检测出来[12,13,17-19]。不像[1]中提到或[8]中预测的明显失真,当对屏幕上或印刷的载体和伪装图像进行比较时,一些工具能保持很好的图像完整性并且显示上几乎没有失真[20]。但当研究8位和灰度

[1] Hide4PGP[11]提供命令行参数,让用户更灵活地决定怎样处理调色板。SysCop[13]在灰度图像中更改保存的调色板。

图像的调色板时，就暴露出一些可检测的模式。一种方法是通过在图像中创建一组不同的像素值，然后按亮度进行排序就能成功地检测到一个隐藏消息的存在性（参见 3.1 节），这种方法可以在颜色值间产生一些特殊模式。

一些伪装工具是通过调整图像调色板，来减少颜色表中从一种颜色值换到另一种颜色值所带来的噪音影响[10－14,16,17]。一个例子是将 8 位图像总的不同颜色数减少到小于 256，然后在几个调色板条目之间扩展新的"基本"色。通过按亮度排序调色板，使颜色块看起来没有变化[19]或仅有很少比特的变化[12]。根据整个图像颜色条目的分布，可以画出一个柱状图，该柱状图包含表示颜色值接近的"双峰"[21]（双峰的进一步讨论参见第七章）。这是人类视觉系统极限的一个很好的例子。然而，除一些伪装技术外，改变调色板条目这种方式几乎很少使用。这种类型的模式不会"自然地"出现。灰度和其它单色图像相比，除了所有的 RGB 值相同和相邻条目之间增量相同之外，它们遵循的模式比较相似，当然这种模式和这个例子的模式不一样。还没有发现那些包含大面积颜色相似的图像能产生这种模式。这些图像包含相似颜色，但颜色的变化远远大于由一些伪装工具[11－13,19]生成的伪装图像颜色上的变化。另一种可能发生的模式是调色板中填补颜色值的结构（也就是，根本没有使用的调色板条目的个数）[13]，这些填补条目主要放在颜色表的尾部。记住，颜色表通常按最常使用的颜色到很少使用的颜色进行排序。一个有黑块的照片的 256 色近似图像几乎没有大量的黑调色板条目，通常黑色区域实际是许多接近黑色的不同渐变色组成的。

在因特网上，经常看到的 GIF 图像，通常含有 256 种颜色。如果灰度图像仅使用一部分可获得的灰度级，则颜色表索引仍包含从白色（FF 或 255）到黑色（0）的 256 个灰度级。伪装工具的一个可能特征是减少颜色索引，使其就是实际使用的颜色数。例如，若载体图像是 GIF 灰度图像但仅使用了 9 个灰度级，图像文件仍将有 256 个颜色索引，从文件偏移位置 0x0D 到 0x30C 存放范围从 0 到 255 的值[2]。当用 SysCop[13]处理灰度图像文件时，调色板减少到伪装图像实际使用的颜色数。若在伪装图像中使用了 9 种颜色，那么在伪装图像文件调色板中仅有 9 个不同的 RGB 三元组，而不是期待的 256 种颜色索引。

一个伪装工具，Hide and Seek[17]，将调色板条目的颜色值划分为四组，生成一个特殊的调色板，这是很不寻常的[3]。在处理灰度图像时，将颜色表中 256 个灰度级分成 4 个从 0 到 252 的三元组，其增长步长为 4（也就是，0,4,8,…248,252）。当随意浏览图像时，检测这种情况的关键是注意到图像中"最白的"值是 252 而不是 255。到今天为止，这种特征是 Hide and Seek 所独有的。

在基于 Windows 95 的 Hide and Seek 1.0 版本中，载体图像仍限制为 256 色彩色或灰度的。在 Hide and Seek 的以前的版本中，GIF 图像被用作载体。由于 GIF 图像压缩算法的版权，在 Windows 98 中使用的是 BMP 图像。在版本 4.1 和 5.0 中，伪装图像的调色板不再产生可预见的模式。

4.3　提取隐藏信息

随着伪装系统的开发者和用户对伪装技术了解的增多，隐藏信息的检测变得越来越具有

[2] Offset 是从文件开始偏移的比特数。16 进制值 0x0D 和 0x30C 的十进制表示是 13 和 780。在 GIF 图像中一个典型的灰度级颜色索引是从偏移值 0x0D 到 0x30C，变化的范围是从 0 到 255。

[3] 这种特性在 Hide and Seek 的 4.1 和 5.0 版本存在，Windows 95 版本没有这种特性。

挑战性。如果一个人不能检测到隐藏信息的可能性，那么提取信息也就不大可能。简单地在一幅图像的 LSB 中嵌入信息，若产生了人工痕迹，也就对隐藏的信息提供不了保护。

Hide4PGP[11]对怎样处理 8 比特调色板或对数据应隐藏在哪些比特位上提供了许多可选参数。隐藏信息的缺省贮存区域是 8 位图像的 LSB 和 24 位图像的第四个 LSB(也就是从右数第四位)。BMP 文件有一个 54 字节的头，24 位图像数据紧跟着文件头。而 8 位图像需要一个调色板，文件头后的 1024 字节则用作调色板。由于 Hide4PGP 在位平面中连续地嵌入信息，通过提取伪装图像中的位比特就能获得嵌入文本。如果是隐藏明文并使用 Hide4PGP 中的缺省设置，对于 24 位 BMP 文件，我们就能从第 54 字节开始第四个 LSB 来提取和恢复嵌入信息，对于一个 8 位 BMP 文件，可以从 1078 字节开始提取 LSB 来显示隐藏的明文信息。如果隐藏的信息已被加密，则需要用到密码分析技术去破译加密的内容。如果加密内容有一个可识别的头，攻击者就会有正确的方向进行密码分析或蛮力攻击[1,22]。

选择哪个比特位隐藏信息，有几个选项可以考虑。在许多 8 位图像中，在比特位中隐藏信息能产生可视噪声。为克服这一点，加入了调色板操作选项，这些选项考虑对经常使用的调色板条目进行备份或对相近颜色进行排序。通过调色板排序，Hide4PGP 把相近颜色成对地放在一起，这与[10]中的方法相似。调色板的修改大大地改进了所产生的载体图像的外观效果，但这又加上了这种伪装工具所特有的属性，并提示攻击者可能有隐藏信息。复制条目的数目总是一个偶数，这是一个能用于标示 Hide4PGP 特征的特点(与[23]和[13]相似)。

如果消息加密了，使用这样的一个伪装技术也可能是有问题的。如果加密数据有一个可识别头，那么攻击者就会有正确的方向进行密码分析或蛮力攻击[1,6]。

4.4 破坏隐藏信息

检测隐藏信息的存在性就可以挫败伪装术。按第三章讨论的使用变换域的方法所产生的伪装对象，没有原始图像进行比较，检测隐藏的信息是很困难的。有时候我们需要让伪装对象在通信信道上通过，但是破坏掉所嵌入的信息[6,24]。对每一个隐藏信息方法，嵌入负荷大小(隐藏信息数量)和信息抗攻击的健壮性存在着矛盾。

以添加空格和“不可见”字符形式在文本中隐藏的信息，当用字处理器打开时，很容易暴露。额外的空格和字符也能很快地从文本文档中去掉。

隐藏的信息也可以被覆盖掉。如果信息被加进某些媒体后，加进的信息不能被检测，那么在同样的门限内，存在一定数量的可以被加进或移去的另外的信息，这些信息将覆盖或删除嵌入的秘密信息(参见 4.2 节)。

在文件或文件系统的未使用空间中隐藏信息必须十分小心，文件头和保留空间是寻找“不合适”信息的常见地方。在文件系统中，除非以某种方式保护了伪装区域(如在一个分区)，否则操作系统可以自由地覆盖隐藏的数据，因为这些簇被认为是可用的。这是操作系统特令人讨厌的地方，因为它在运行过程中要做许多缓存和创建临时文件。还有一些软件工具，它们可以“净化”或清除未使用的贮存区域。在清除中，需对簇重写几次以确保删掉了数据。即使在这种极端情况，也有一些工具可以恢复部分被覆盖的信息。

与文件头中未使用或保留空间一样，TCP/IP 包头也能很容易地被检查。正如防火墙过滤器能设置成检验源和目的 IP 地址、SYN 和 ACK 比特的有效性一样，过滤器也能配置成捕获那

些在未使用或保留空间中存在信息的包。如果用 IP 地址更改或欺骗来进行秘密信息传递，在域名服务(DNS)中一次逆向查询就能验证这个地址。如果 IP 地址是错误的，这个包就被丢弃。当 TCP/IP 头在路由处理时可能被覆盖重写时，使用这种技术来隐藏信息就太冒险了。保留的比特可能被覆盖重写，并继续往前传送而不影响包的正常路由。如果通过修改包中的时间戳来传递信息，那么当包通过路由器时，其过滤器也能覆盖时间戳，结果破坏了隐藏信息。

这里讨论破坏隐藏信息的方法，不是有意提倡非法行为去删除或破坏图像中的有效信息，而是评估嵌入系统的思想和研究嵌入方法的健壮性。一些破坏隐藏信息的方法可能需要对伪装图像进行大量修改(参见第七章对数字水印的相关讨论)。由于位平面方法使用了图像的 LSB，它很容易因小的图像压缩处理而改变，因此在使用位平面方法情况下，很容易破坏嵌入信息。破坏采用变换域工具的数据隐藏，需要付出更多的努力。许多变换域方法的目标是使隐藏信息成为图像整体的一部分，唯一可以移去或破坏嵌入信息的方法，就是严重地破坏图像，而破坏后的图像对攻击者来说就没用了。

破坏或删除图像中的隐藏信息来自于各种图像处理技术。用 LSB 方法插入数据，简单使用有损压缩技术，如 JPEG，就足以能破坏嵌入信息。用这种方法作图像压缩，人眼是看不出多大变化的，但已经不再包含隐藏信息了(参见第三章中关于 JPEG 压缩的介绍以及它与伪装术的关系)。对变换域隐藏的信息，破坏嵌入信息的可读性则需要更实质性的处理。许多图像处理技术，如扭曲、裁剪、旋转和模糊化能产生足够大的失真来破坏嵌入信息。这些图像处理技术组合起来使用，可以测试除了 LSB 之外的信息隐藏技术的健壮性[25,26]。这些工具的例子和评价可在文献[6]、[27]和第七章中找到。诸如此类的工具也应该被那些考虑在信息隐藏系统上投资的人使用，因为他们把信息隐藏系统做为工具，为隐藏信息提供安全性。就象系统管理员使用口令破译工具测试用户和系统口令的强度一样，如果口令被破解，管理员应提醒用户，他的口令已经不安全了。

一系列图像处理测试被设计出来，评估位平面和变换域工具的健壮性程度[6]。这些测试对载体进行修改，直到不能找回嵌入信息为止。在一些工具中，这种事实可看成“译码器”的缺点而不是“编码器”的缺点。这些测试的动机是想说明这些技术能经受什么样的攻击，并使一些常规的弱点暴露出来。在这些测试中使用的图像包括数字照片、裁剪艺术和数字艺术。数字照片主要是拥有上千种颜色的 24 位真彩，也可以是 8 位灰度图像。数字照片大部分是 JPEG 图像和 24 位的 BMP 图像。裁剪艺术图像有相对较少的颜色，并且主要是 GIF 图像。数字艺术图像不是照片但却拥有上千种颜色，这些图像可能是 24 位的(BMP 或 JPEG)或 8 位的图像(BMP 或 GIF)。只要需要，图像可以转换为伪装工具或水印工具指定的特定格式。

从每一种图像类型中选择许多图像嵌入已知消息或水印，并且对产生的伪装图像进行消息内容验证。在健壮性测试中，对伪装图像进行许多图像处理操作并检查消息内容，测试包括无损到有损格式的转换，位密度(24 位、8 位、灰度)转换、模糊化、平滑、加入噪声、去除噪声、锐化、边缘增强、掩饰、旋转、缩放、再抽样、扭曲(非对称抽样，与 jitter 攻击[27]相似)、从数字到模拟和从模拟到数字(打印和扫描)、镜像、跳变、加入位平面消息、加入变换域消息等等，以及应用 unZign 和 StirMark 工具来测试水印软件的健壮性。另外也进行一系列测试，以确定对每一种工具能够成功隐藏数据的最小图像。

微小的图像处理或用 JPEG 压缩图像足以破坏位平面隐藏工具，而变换域方法能经受许多图像处理测试。每一种测试，要使用许多图像，可以是 8 位、24 位、无损和有损图像格式。

基于位平面方法的一些工具进行信息隐藏，使用上述任何一种测试都无法恢复信息。变换域嵌入信息能经受其中的许多测试，但无法抵抗这些方法的组合使用。目前已有的工具也可应用于伪装图像来测试其健壮性[25,26]。在破坏隐藏信息使其不可读方面，所注意到的成功的方法是对图像的微小几何变形，然后再抽样和平滑。这种方法组合了轻微模糊化、边缘增强和非对称抽样或弯曲的效果。水印对各种攻击的健壮性的进一步讨论可以在第七章中找到。

声音和视频对这些类似的攻击方法是脆弱的。对信号的操作将改变在噪音中(LSB)嵌入的信息，并足以覆盖或破坏嵌入消息。滤波器可用来滤掉回声或一些微弱信号，但可能没有预期那样成功。对声音中回声隐藏的一种可能的蛮力组合攻击可参见[27](第七章进一步讨论回声隐藏攻击)。

4.5 讨论和结论

这一章对伪装分析给出了一个总的想法和介绍，并且指出了一些伪装方法的弱点和可视性特征。这些工作只是伪装分析方法的一部分。到目前为止，仍然没有设计出一个应用于数字图像伪装的通用的检测技术，并且除了视觉分析之外的各种方法仍在探索之中。大量的图像都是由人工浏览来检查隐藏信息的。我们已经介绍了伪装软件的一些缺点，它们标示了隐藏信息的可能存在性，检测这些"特征"可以包含在检测信息隐藏的自动工具中。检测隐藏信息的工具在伪装分析和验证水印方面是很有应用价值的。

伪装术通过表面上无害的载体来传送秘密信息，以达到隐藏秘密的目的。数字图像伪装术和它的派生产品在应用上越来越广泛。伪装术的商业应用，如数字水印和数字指纹，可以被用于追踪电子媒介的版权和所有者。了解和研究这些应用的缺陷，有助于指导研究者拿出更好的、更健壮的解决办法。在设计更健壮的信息隐藏技术时，必须确保嵌入信息(如版权和许可信息)的抗攻击性，而测试抗攻击性的这类工具，对开发健壮性更强的技术是必不可少的。使用本书介绍的工具和方法，信息隐藏工具使用者能明白，需花费多少(或最小)精力才能使嵌入信息不可读。

伪装工具的易用性和获得上的便利性，使执法部门非常关心非法资料的流通。目前正在探索各种消息检测方法，理解当前技术的检测门限，以揭示这些活动的秘密。因特网伪装术正在开展的工作是研究嵌入、恢复和检测 TCP/IP 包头和其他网络传输中的信息。

成功地伪装秘密信息的关键在于选择合适的机制。而一个看似无害的伪装图像，经过进一步研究分析，可能就会揭示嵌入信息的存在性。

隐蔽通信和伪装术领域的研究将会继续下去。建立更强壮的、能经得起图像操作和攻击的各种方法的研究，也将深入下去。越是将信息放在因特网上共享，这些信息拥有者越需要保护他们自己的信息免于被盗和错误表述。对表面上已损坏的信息进行恢复的技术和伪装分析技术，对计算机法庭和数字交通流量分析[28,29]方面的法律执行权威机构也是很有用的。

参考文献

[1] Kurak, C., and J. McHughes, "A Cautionary Note On Image Downgrading," in IEEE Computer Security Applications Conference 1992, Proceedings, IEEE Press, 1992, pp. 153 – 159.

[2] Craver, S., N. D. Memon, and M. M. yeung, "Resolving Rightful Ownerships with Invisible Watermarking Techniques," Technical Report RC 20755(91985), IBM Research Division, 1997.

[3] Flynn, J., "A Journey Within Steganos," 〈http://www.fravia.org/fly___01.htm〉.

[4] "PhotoShop 4.0/Digimarc: Commercial stupidity-Digimarc downfall," 〈http://www.fravia.org/forgdigi.htm〉, 1997. Original post in Learn Cracking IV on 〈news:tw.bbs.comp.hacker〉.

[5] Brassil, J., et. al., "Document Marking and Identification Using Both Line and Word Shifting," in Proceedings of INFOCOM'95, Apr. 1995, pp. 853 – 860.

[6] Johnson, N. F., and S. Jajodia, "Steganalysis of Images Created Using Current Steganography Software," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 273 – 289.

[7] Koch, E., J. Rindfrey, and J. Zhao, "Copyright Protection for Multimedia Data," in Proceedings of the International Conference on Digital Media and Electronic Publishing, Leeds, UK, Dec. 1994.

[8] Cha, S. D., G. H. Park, and H. K. Lee. "A Solution to the Image Downgrading Problem," in Computer Security Applicatons Conference Proceedings, 1995, pp. 108 – 112.

[9] Aura, T., "Invisible Communication," Technical report, Helsinki University of Technology, Nov. 1995.

[10] Machado, R., "EzStego. Stego Online, Stego," < http://www.stego.com >, 1997.

[11] Repp, H., "Hide4PGP," < http://www.rugeley.demon.co.uk/security/hide4pgp.zip >, 1996.

[12] Brown, A., "S-Tools for Windows," < ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/s-tools4.zip >, 1996.

[13] MediaSec Technologies LLC. "SysCop," < http://www.mediasec.com/ >, 1994 – 1997.

[14] Heckbert, P., "Color Image Quantization for Frame Buffer Display," ACM Computer Graphics, vol. 16, no. 3, Jul. 1982, pp. 297 – 307.

[15] Wayner, P., Disappearing Cryptography, Chestnut Hill, MA: Academic Press, 1996.

[16] Hansmann, F., "Steganos, Deus Ex Machina Communications," < http://www.steganography.com/ >, 1996.

[17] Maroney, C., "Hide and Seek," ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/hdsk41b.zip >, < http://www.rugeley.demon.co.uk/security/hdsk50.zip >, 1994 – 1997.

[18] "stegoDos-Black Wolf's Picture Encode v0.90B," < ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/stegodos.zip >, 1993.

[19] Hastur, H., "Mandelsteg," < ftp://idea.sec.dsi.unimi.it/pub/security/crypt/code/steg.tar.Z >, 1994.

[20] Johnson, n. F., and S. Jajodia, "Exploring Steganograhpy: Seeing the Unseen," IEEE Computer, vol. 31, no. 2, 1998, pp. 26 – 34.

[21] Maes, M., "Twin Peaks: The Histogram Attack on Fixed Depth Image Watermarks," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecutre Notes in Computer Science, Springer, 1998, pp. 290 – 305.

[22] Koch, E., and J. Zhao, "Towards Robust and Hidden Image Copyright Labeling," in IEEE Workshop on Nonlinear Signal and Image Processing, Jun. 1995, pp. 452 – 455.

[23] Arachelian, R., "White Noise Storm," < ftp://ftp.csua.berkeley.edu/pub/cypherpunks/steganography/wns210.zip > ,1994.

[24] Anderson, R.J., and F.A.P. Petitcolas, "On The Limits of Steganography," IEEE Journal of Selected Areas in Communications, vol.16, no.4, 1998, pp.474 – 482.

[25] Petitcolas, F.A.P., and M.G. Kuhn, "StirMark," < http://www.cl.cam.ac.uk/fapp2/watermarking/stirmark/ > ,1997.

[26] "unZign. Watermarking Testing tool," < http://altern.org/watermark/ > ,1997.

[27] Petitcolas, F.A.P., R. Anderson, and M. Kuhn, "Attacks on Copyright Marking Systems," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp.218 – 238.

[28] Johnson, N.F., Z. Duric, and S. Jajodia, "A Role of Digital Watermar King in Electronic Commerce," to appear in in ACM Computing Surveys.

[29] Duric, Z., N.F. Johnson, and S. Jajodia, "Recovering Watermarks from Images," Information & Software Engineering Technical Report: ISE-TR – 99 – 04, to appear in IEEE Transactions on Image Processing, 1999. < http://www.ise.gmu.edu/techrep/1999/ > .

第二部分　数字水印与版权保护

第五章　水印技术简介

(Martin Kutter, Frank Hartung)

本书的第一部分集中介绍了信息伪装，在第二部分，我们将介绍数字水印。一方面水印与信息伪装的联系非常密切，另一方面，相对于信息伪装而言，水印基于不同的原理，有不同的需求和应用，这导致了它们在技术特性方面具有明显的差别。

这一章将对水印及其相关技术作一全面介绍。首先我们要阐明信息伪装与水印的相同点以及它们之间的区别。我们将回顾数字水印的历史概貌，并将它与普通的纸上水印相对比。然后，我们将着眼于一般的水印系统，并讨论其应用、需求和设计问题。

5.1 引　　言

信息伪装和水印都是描述将信息嵌入到伪装载体中的技术，该技术使所传送的信息不可察觉。然而，正如在第三章中介绍的，典型的信息伪装是指在通信双方进行隐蔽的点到点通信，因此，信息伪装对载体数据的修改通常是不太健壮的，或者说只有有限的健壮性，对载体数据传输和存储过程中可能发生的技术性的修改，比如格式转换、压缩、数模变换等，信息伪装对所嵌入信息只能进行有限的保护。

从另外一个方面，为了防止企图移去隐藏的数据，水印提出了更高的要求。因此，当使用载体数据的各方知道隐藏数据的存在，并且想移去它的时候，使用的通常是水印而不是信息伪装。水印的一种比较常见的应用，是通过在数字产品中嵌入版权信息来作为提供产品所有权的证据。很明显，就这种应用而言，嵌入的信息对试图移去它的操作应该是健壮的。其他的应用还有载体数据监控或者说载体数据跟踪，如果用户为了控制正当的版权收入而对所监控数据的发行感兴趣，或者是为了市场和营销的目的，仅仅跟踪数据的发行使其局限于某一范围，那么就可采取载体数据监控或跟踪。此外，数字水印也可用在数字指纹应用场合中来区分所发行的各种数字产品。第八章将讨论数字指纹的问题。

5.2 历史及术语

5.2.1 历史

大约 700 年前，在手工造纸技术中出现了纸张上的水印。在历史文档里发现的最古老的带水印的纸张可追溯至 1292 年，这种纸起源于意大利 Fabriano 的一个城镇，该城镇在造纸工业发展中扮演了重要的角色。在 13 世纪末，大约有 40 家造纸厂共享 Fabriano 的纸业市场，并生产具有不同式样、质量和价格的纸。那时造纸厂生产的生纸表面粗糙，不适合直接使用。这种生纸材料被送给其他工匠，他们利用一种叫做研光机的硬石块将纸表面弄光滑。接着经过处理的纸张被清点、折叠，最后卖给商人。这些商人将它们存放在大仓库里以备转卖，来获取高

额利润。不但 40 家造纸厂之间的竞争很激烈，工匠之间以及商人之间的竞争也很激烈。对任何一方来说，跟踪纸的来源以及对式样和质量进行鉴定都不是一件容易的事。水印的使用是一种完美的解决方法，它能消除任何可能的冲突（见图 5.1）。图 5.1 中的字母组合图 TGE RG（Thomas Goodrich Eliensis——英格兰 Ely 的主教——和 Remy/Remigius Guedon——造纸工匠）是在剑桥地区发现的最早的水印之一（公元 1550）。那时，水印主要用于识别生产纸张的工厂，作为保证纸张质量的一种手段。该图得到英格兰剑桥大学档案室 E. Leedham-Green 的使用许可。其制作技术为 beta X 光照相术，印自[3]。自从水印发明之后，它们迅速在意大利继而在全欧洲传播开来。尽管它们最初是用来表明纸的商标或生产厂家，但是后来却被用作纸的式样、质量和强度的标识，也可用作确定纸的生产日期和鉴别纸的根据。一个很好的能说明水印法律效应的例子是 1887 年发生在法国的一个案例[1]。作为法律证据出示的两封信的水印，说明信的日期已被提前，从而导致了对责任人的起诉、警察长官的免职、内阁的解散以及最终总统 Grévy 的辞职。如果想了解有关纸上的水印、水印历史以及相关法律问题的更多信息，感兴趣的读者可查阅[2]——一个超过 500 个参考书目的大规模列表。

图 5.1　字母组合图 TGE RG

纸上的水印与数字水印之间的相似性是很明显的：在银行票据或邮票上的纸上水印，激发了“水印”[4]这一术语在数字产品环境中的首次使用。最早关注数字图像水印的文献是 Tanaka 等人[5]于 1990 年以及 Caronni[6] 和 Tirkel 等人[4]于 1993 年发表的。

1995 年很显然是讨论这个主题的最适当的时候，并且开始引起越来越多的研究活动。1995 年以来，数字水印得到很多关注，发展得也很迅速（见表 5.1）。尽管还有许多主题有待进一步研究，但实用的方法和系统都已经开发出来了，我们将在第六章中对其中的一些进行介绍。

表 5.1　　据 INSPEC 1999 年 1 月统计，最近几年有关数字水印出版物的数量

经 J.L.Dugelay 允许[7]

年　份	1992	1993	1994	1995	1996	1997	1998
出版物	2	2	4	13	29	64	103

5.2.2　水印术语

现在我们关注的显然是数字的而不是模拟的通信与媒体。正如在模拟媒体中那样，信息伪装和水印方法有一定的使用价值，它们允许信息隐藏或嵌入在其它数据中再进行传输。这些技术使用了多个名字，可是，这些术语经常容易混淆，所以有必要阐明它们的区别。

• **可视水印**　正如名字所说的，是可视的图案，就像插入或覆盖在图像（或视频）上的标识，与可视的纸上的水印很相似。可视水印主要应用于图像，比如用来可视性地标识那些可在图像数据库中得到的，或在 Web 上得到的图像的预览，来防止这些图像被用于商业用途。可

视水印当然也可以应用在视频中，在有些情况下，我们甚至可以考虑在音频中嵌入可听的水印。在IBM数字图书馆工程中，已开发出一个可视水印嵌入的实例[8]。该技术通过修改原始图像的亮度将水印图像与原始图像组合在一起，其亮度的修改程度是水印和一个密钥的函数，其中密钥随机地确定亮度改变的大小，这样做的目的是使攻击者难以移去可视的标识。本书的其它部分我们将着重于不可察觉的水印。

• **数字水印** 与信息伪装相比，数字水印要求有更强的健壮性以抵御攻击。即使知道隐藏信息的存在，不知道密钥（参见1.2.4小节中的Kerckhoffs准则）的攻击者也很难破坏掉嵌入的水印。健壮性的要求实际上暗示了水印方法可以嵌入到载体数据中的信息要比信息伪装方法少得多。因此，信息伪装和水印与其说是相互对立不如说是相互补充。因为在前面的章节中已经涉及伪装方法，所以在这一章的剩余部分，我们将主要着重于水印方法，而不是伪装方法。

• **数字指纹和标签** 是表示水印一些特定应用的术语。在诸如数字产品创作者的信息或接收者的信息被作为水印而嵌入到产品中时，它们就与水印应用联系起来了。如果嵌入信息是标识载体数据的作者或发明人的一个唯一代码，或者是标识数据接收者的一系列代码中一个独特代码，则这样的数字水印就称为数字指纹。有关数字指纹的更多细节将在第八章中介绍。如果嵌入的数据可能包含任何像唯一标识符那样的有用信息，那么这样的数字水印就称为标签。

• **比特流水印** 有时被用作压缩数据的数字水印，例如压缩的视频。

在早期文献中使用的术语是"嵌入签名"而不是"水印"，但现在通常不再使用这个旧名词了，因为它可能会与密码中的数字签名产生混淆。密码中的数字签名是用作认证的，它们被用来检测对签名数据的任何改动，并且确认发送者。而水印只有在特定应用中才用来作认证，通常是被设计成能抵抗改动和修改之类的攻击。

• **脆弱性水印** 是健壮性非常有限的水印，它们是用来检测加水印数据的改动，而不是用来传送不可擦除的信息。脆弱性水印和认证的应用将在5.4.4小节讨论。

5.3 嵌入水印的基本原理

所有嵌入水印的方法都包含这些基本的构造模块，即一个水印嵌入系统和一个水印恢复系统（也称水印提取或水印解码系统）。图5.2展示了一个一般的水印嵌入过程。该系统的输入是水印、载体数据和一个可选的公钥或私钥。水印可以是任何形式的数据，比如数值、文本、图像等等。密钥可用来加强安全性，以避免未授权方恢复和修改水印。所有的实用系统至少使用一个密钥，有的甚至是几个密钥的组合。当水印与私钥或公钥结合时，嵌入水印的技术通常分别称为秘密水印技术和公开水印技术。水印系统的输出称为添加了水印的数据。

对现实世界中健壮的水印系统而言，它们具有一些基本特性：

• **不可感知性** 因嵌入水印而引起的变动应该低于可感知的门限，也就是说一些感知性标准不仅被用来设计水印，也用来对失真进行量化。由于水印所必需的不可感知性，用来嵌入水印的每个样本值（或像素、音素、特征等）只能作微小的修改。

• **冗余性** 尽管在允许的范围内所做的改动很小，为了保证健壮性，通常要将水印信息冗余地分布在载体数据的很多样本（或像素、音素、特征等）中，从而得到整体的健壮性。这意

味着利用一小部分已嵌入水印的数据就可以恢复出水印。很显然，如果在水印恢复过程中有更多的已嵌入水印的数据可用，则水印的恢复将会更有效。健壮性准则将在第七章中介绍。

• **密钥**　一般而言，水印系统使用一个或多个安全的密钥来确保安全，防止修改和擦除水印。一旦水印被人读取到，他就可以很容易地破坏该水印，因为在这种情况下，他不仅知道水印嵌入的策略也知道水印嵌入的位置。

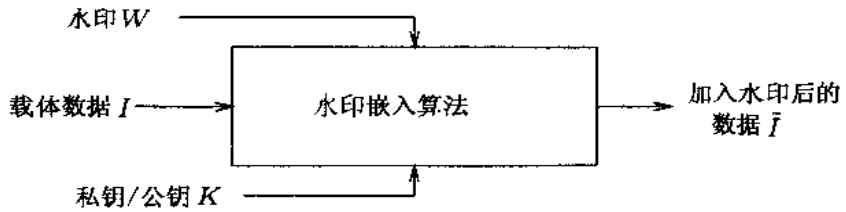

图 5.2　一般的数字水印嵌入方案

这些原则应用于各种类型数据的水印嵌入方案中。加载水印的数据可以是音频、图像、视频、格式化文本、3D 模型、动画模型的参数，也可以是其它类型的数据。

图 5.3 描述了一般的水印恢复过程。该系统的输入是已嵌入水印的数据、私钥或公钥，以及原始数据和(或)原始水印(取决于添加水印的方法)，输出的是水印 W，或者是某种可信度的值，它表明了所考察数据 $\tilde{I}'$ 中存在给定水印的可能性。水印系统可分为三种，它们的区别在于输入输出的种类及其组合(见[9])。

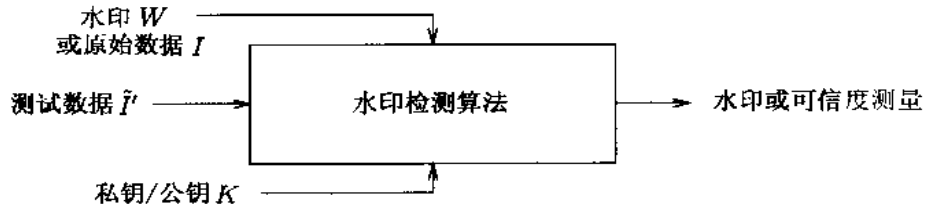

图 5.3　一般的水印恢复方案

• **秘密水印**(也叫非盲化水印)　该类系统至少需要原始的数据。I 型系统从可能失真的数据 $\tilde{I}'$ 中提取水印 W，并使用原始数据作为线索来确定水印在 $\tilde{I}'$ 中的位置。II 型系统也需要所嵌入水印的一个拷贝，得到"$\tilde{I}'$ 中是否含有水印 W"这个问题的"是"或"不是"的答案($\tilde{I}' \times I \times K \times W \rightarrow \{0,1\}$)。由于该系统传输的信息很少，并且需要使用密钥之类的信息，因而它的健壮性比其它方案更好。

• **半秘密水印**(也叫半盲化水印)　该类系统并不使用原始数据来检测($\tilde{I}' \times K \times W \rightarrow \{0,1\}$)，但回答了与上面同样的问题。秘密和半秘密水印的潜在应用包括可作为法庭上证明所有权的证据、DVD 等应用中的版权控制(在该应用中，唱片的使用者需要知道它能不能播放)以及以识别盗版源为目标的数字指纹等等。

• **公开水印**(也叫盲化或健忘水印)　该类系统是目前最具挑战性的问题，因为它既不需要原始的秘密信息 I，也不需要水印 W。实际上，这种系统是从已嵌入水印的数据中提取 n 位信息(水印)：$\tilde{I}' \times K \rightarrow W$。

根据应用，上述水印嵌入和水印恢复方案的输入通常以压缩或解压缩数据的形式存在。很显然，水印技术应该拓宽输入的种类。如果一个用于 MPEG－2 视频的水印方案需要单一的

解压缩视频帧的话，那将没有什么意义，因为这要包含解码和再编码过程，从而使嵌入水印过程的运算代价太高。然而，我们应该注意到，对某些应用来说，设计这样的水印检测方案是有用的，它们不管水印嵌入在什么位置都可检测到。这一点很重要，例如，在需要对代码转换或格式变化有免疫力的应用中。

5.4 水印的应用

水印系统要满足的条件总是建立在应用基础上的。因此，在我们回顾这些条件和最终设计之前，先介绍水印的一些应用。很显然，没有“普遍适用的”水印嵌入方法。虽然总体来说水印嵌入方法应该是健壮的，但是不同的应用对健壮性有不同的要求。

5.4.1 用于版权保护的水印

目前，版权保护可能是水印最主要的应用。其目的是嵌入数据的来源信息以及比较有代表性的版权所有者的信息，从而防止其它团体对该数据宣称拥有版权。这样水印就可以用来公正地解决所有权问题，这种应用要求非常高的健壮性。包含很多图像的 Web 是该应用的推动力量，它所含的这些图像是可随意使用的，但是它们的所有者却要保护它们。除了健壮性以外还要考虑其它的问题，例如，水印必须是清晰的，当其它人嵌入另外的水印时仍然能够确认真正的所有者。因此，除了健壮性之外，还要考虑其它的要求。

5.4.2 用于盗版跟踪的数字指纹

还有其它的一些应用，它们的目的是传输合法接收者的信息而不是数据来源者的信息，主要用来识别数据的单个发行拷贝。这很像软件产品的序列号，对监控和跟踪流通数据的非法拷贝非常有用。这一类应用在发行的每个拷贝中嵌入不同的水印，通常称之为“数字指纹”。对每个拷贝各自嵌入水印的情况，因为它们的发行要面临共谋攻击的危险(见第八章)，所以嵌入的水印应该设计成对共谋攻击而言是安全的[10,11]。同样，对于某些数字指纹应用来说，它们要求水印易于提取，且有很低的复杂度，例如对于 WWW 应用，有专门的 Web 搜索者寻找嵌入了水印的盗版图像。数字指纹应用中的水印也需要很高的健壮性，不仅要能抵抗恶意的攻击，还要能抵抗一些标准数据处理。

5.4.3 用于拷贝保护的水印

在多媒体发行体系中，希望存在这样的一个拷贝保护机制，即它不允许未授权的媒体拷贝。在开放系统中很难实现拷贝保护，然而在封闭或私有系统中，拷贝保护是可行的。在这样的系统中，可用水印来说明数据的拷贝状况。这种情况的一个例子是 DVD 系统，在该系统中，数据中的水印含有拷贝信息。一个符合要求的 DVD 播放器不允许重放或拷贝带有“禁止拷贝”水印的数据，而带有“一次拷贝”水印的数据可以被拷贝，但不允许从该拷贝再进一步制作拷贝[12,13]。

5.4.4 用于图像认证的水印

在鉴定应用中，使用水印的目标是对数据的修改进行检测。这可用所谓的“脆弱性水印”

来实现，它对特定的修改（如压缩）有弱的健壮性，而其它的修改则是破坏性的[14,15]。此外，根据不同的数据类型和应用，相应的健壮性要求也会有所不同。不过，在所有可能的水印应用中，用于鉴定的水印对健壮性的要求最低。应该注意到已经出现了一些新的方法，在这些方法中，诸如块平均或边缘特性之类的数据属性被嵌入到图像中，以用来验证接收到的图像是否还有同样的属性。很明显，如果对修改区域的鉴定感兴趣的话，这种方案需要较高的健壮性。

5.5　要求和算法设计问题

对应于水印不同的应用和目的，不同的要求导致了各种各样的设计问题。水印的不可感知性是一个共同的要求，并且是独立于应用的。在设计嵌入水印方法的时候还要考虑其它一些要求：

• **水印恢复是否需要原始数据**　对不同的应用而言，水印恢复系统有可能得到原始数据，也有可能得不到原始数据。如果能得到，使用它通常是有帮助的，因为使用原始数据恢复的系统一般都有更好的健壮性。然而，在像数据监控这样的应用中，因为其目的是识别被监控的数据，所以原始数据并没有什么用处。

• **对给定水印的提取或确认**　水印嵌入和恢复有两种方法，它们在本质上是等价的。在第一种方法中，所嵌入的水印是预定义的可容许水印集合中的一种，水印恢复时将嵌入的水印数据与可容许水印集合相比较进行验证。水印恢复的输出是嵌入水印的索引值或者是"没发现水印"的结论。在第二种方法中，嵌入的水印是对水印嵌入系统中给定符号序列的调制。在检测过程中，嵌入的符号通过解调被提取出来。

• **健壮性**　在水印嵌入过程中，健壮性是抵抗修改和（或）恶意攻击的关键要求之一。不过正如前面所说的，相对于其它应用而言，有一些应用对它的重要性要求较低。

• **安全问题和密钥的使用**　对应于不同的应用，密钥的管理状况有很大差别。很明显的例子是公钥水印系统，如 DVD，与此相对的是用于版权保护的私钥水印系统。

5.5.1　不可感知性

对水印系统的一个最重要的要求就是水印可感知的透明度，它是与水印系统的应用和目的无关的。水印嵌入过程引入的人为痕迹不但令人讨厌，并且也不是我们想要的，而且它有可能减少或破坏已嵌入水印数据的商业价值。因此，设计这样的水印方法是重要的，它利用人的视觉和听觉系统的有效性，在不超出可感知门限的约束下，使水印的能量取得最大值。与此相关的有两个问题，第一个问题是对所引入的失真进行可靠的评估，这将在 5.6 节进行介绍。第二个问题出现在对水印数据进行处理的时候，例如，在图像水印中，如果图像按比例缩放，那么水印的可视度将可能会增加。

5.5.2　健壮性

水印嵌入方法最终应该抵抗由标准的或恶意的数据处理所引入的任何类型的失真。迄今为止，还没有提出这样完美的方法，甚至到底存在不存在一种完全安全的水印嵌入方法也未可知。因此，实际系统应在健壮性和诸如可视性、信息速率之类的相互冲突的要求之间进行折衷。根据水印嵌入方法的应用目的，它所要求的健壮性会影响设计的过程。例如在图像水印

中，如果我们需要一种能容忍压缩比很高的 JPEG 压缩的水印嵌入方法，那么使用变换域的方法可能要比使用空间域的方法更有效。同样，如果所用方法要适应一般的几何变换（旋转、不规则缩放、剪切等），那么空间域的方法可能会更合适。考虑水印数据可能要遭受的有意或无意的破坏，可将失真分为两组。第一组包含的是被看作加性噪声的失真，而第二组的失真是由于空间或时间数据的几何修改而引起的，其目的是在水印和用于水印嵌入的密钥间引入不匹配。通常分别称这两种失真为破坏攻击和同步攻击。在[16]中，提出了更好的攻击的分类。

根据应用和水印嵌入的要求，我们认为应考虑的失真和攻击应包括（但并不局限于）以下这些：

- 信号增强（锐化、增加对比度、色彩校正、伽马校正）；
- 加性和乘性噪声（高斯、均衡、斑点、蚊状）；
- 线性滤波（低通、高通、带通滤波）；
- 非线性滤波（中值滤波、形态滤波）；
- 有损压缩（图像：JPEG。视频：H.261、H.263、MPEG－2、MPEG－4。音频：MPEG－2 音频、MP3、MPEG－4 音频、G.723）；
- 局部和全局仿射变换（平移、旋转、缩放、剪切）；
- 数据缩减（剪短、修剪、柱状图修改）；
- 数据合成（标志插入、布景合成）；
- 代码转换（H.263→MPEG－2、GIF→JPEG）；
- D/A 和 A/D 变换（打印—扫描、模拟电视传输）；
- 多重水印；
- 共谋攻击；
- 统计平均；
- 马赛克攻击。

这些攻击中的某些攻击以及其它攻击将在第七章中介绍。基本的原则是设计这样一种水印嵌入方法，它应该足够的健壮，以致于即使攻击成功也只会削弱载体数据的商业价值。

5.5.3 是否需要原始数据的水印恢复

通常在恢复过程中使用原始数据集的水印嵌入方法会提高水印健壮性，这不仅是针对像噪声一样的失真而言，而且也是针对数据中几何失真而言，因为它允许几何失真的检测和倒置。不过，在很多应用中，比如数据监控或跟踪，不可能得到原始数据。在其它一些应用中，比如视频水印应用，由于要处理的数据数量很大，使用原始数据也是行不通的。尽管大多数早期的水印嵌入技术恢复时都要求有原始数据，但一个很明显的趋势是设计不需要原始数据集的水印技术。这是由于这种技术有更为广阔的应用领域。

5.5.4 水印的提取或对给定水印存在性的验证

正如前面所说的，存在两种水印方案：嵌入特定信息或图案，并在随后水印恢复过程中核对这些（已知）信息存在的系统；和向数据中嵌入任意信息的系统。例如，对已知水印存在性进行验证的系统能用于实现版权保护。而在有智能代理的因特网上，图像跟踪不仅要发现图像还要对它们进行分类，这是嵌入任意信息的水印嵌入方案的一个应用实例。嵌入的水印可被

用作图像识别码或数据库访问指针。

应该注意到，这两种方案可以相互交换。允许水印验证的方案可以认为是一种比特水印恢复方案，并且通过要嵌入的任意信息的调制，很容易将其扩展到任意位数。反过来也是如此，假设嵌入信息是已知的话，可将水印恢复方案看作水印验证方案。

5.5.5　水印安全和密钥

在大多数应用中，比如版权保护，要确保嵌入信息的保密性。通常称这种问题及其它相关的问题为“水印安全”。图像数据库的索引是一个不考虑安全问题的应用。如果保密性是一个要求，那么嵌入和提取水印过程中就要使用密钥。可将保密等级分为两级。在保密的最高等级，一个未授权用户既不能读或解码所嵌入水印，也不能检测一个给定的数据集是否包含水印。第二等级允许任何用户检测数据是否嵌入水印，但没有密钥不能读取嵌入的信息。举例来说，这种方案对图像应用中的版权保护很有用，一旦一个有版权保护的图像在一个图像编辑软件中打开，一个短信息将通知用户该图像是被保护的。

这种方案可以包含多重水印，这些水印使用公钥和私钥。也可以是将一个(或几个)公钥与一个私钥组合起来，嵌入一个公开/秘密的联合水印[17]。

如果要设计一个可工作的全面的版权保护系统，诸如密钥产生、分发和管理(可能是由可信任的第三方管理)以及与其它系统的集成等问题，都应仔细考虑。

5.5.6　确定真正的所有者

在包含多个水印的情况下，为了准确地确定真正的所有者，就要能够确定是谁首次给一个数据集嵌入水印。这可以通过加强设计约束来实现，比如水印的不可逆转性[18]，或者使用附加的功能，比如时间戳等。关于这个问题的更多细节将在第七章中给出。

5.6　水印系统的评价和基准

除了设计数字水印方法之外，另外一个重要的问题是建立对水印的评价和基准。这不仅包括对健壮性的评价，还包括由水印处理引入的失真的主观和定量评价。总的来说，在水印健壮性和可察觉性之间有一个权衡。因此，对合理的基准和性能评价来说，我们必须保证所考察的方法是在可比较的条件下进行的[9,19]。

5.6.1　性能评价和表示方式

与应用目的和数据类型无关，水印的健壮性取决于以下几个方面：

• **嵌入信息的数量**　因为它直接影响水印的健壮性，所以是一个重要的参数。要嵌入的信息越多，水印的健壮性就越低。

• **水印嵌入强度**　在水印嵌入强度(对应于水印的健壮性)和水印可感知性之间有一个均衡。高健壮性需要更强的嵌入，这反过来增大了水印的可感知性。

• **数据的大小和种类**　通常，数据的尺寸大小对嵌入水印的健壮性有直接的影响。例如，在图像水印中，太小的图片没有商业价值；不过，一个标记软件程序要能够从它们中恢复出水印，这避免了对它们的“马赛克”攻击[20](见 7.3.2 小节)，并允许 Web 中常用的“覆盖”。对

打印应用，虽然要求高分辨率的图像，但人们也想在它们重采样并在 Web 中使用后，能够保护这些图像。除了数据的大小，数据的种类同样也对水印的健壮性有影响。仍然采用图像水印作为例子，有的方法在扫描自然图像方面具有很高的健壮性，但它们在合成图像（如计算机生成的图像）中的健壮性却很低。

• **秘密信息（如密钥）** 尽管秘密信息的数量对水印的可感知性、健壮性没有直接的影响，但在系统的安全性方面充当了重要的角色。密钥空间，也就是秘密信息所有可能的取值范围要足够大，从而使穷举搜索攻击不可行。读者也应牢记，由于很多系统在设计中没有遵守基本的密码学原理，它们连非常简单的攻击都抵抗不住[20,21]。

考虑这些因素，我们认识到，要想得到合适的基准和性能评估，水印嵌入方法必须要针对不同的数据集进行测试。此外，为了得到统计上的有效结果，必须使用很多不同的密钥和多样化的水印来评估这些方法。嵌入信息的数量通常是确定的，并且取决于具体的应用。可是，如果要对水印嵌入方法进行比较，就要保证所有要考察的方法的嵌入信息数量是一样的。

正如前面所看到的，水印的可感知性和健壮性之间有一个权衡。因此，要进行公平合理的评估和比较，在评估过程中就要考虑水印的可感知性。对水印的可感知性进行评估可以通过主观测试或者通过质量度量来衡量。当使用主观测试时，要遵守一个测试协议，该协议要对测试和评估的过程进行描述。这样的测试通常包括两个步骤。第一步，将失真的数据集按照从最好到最坏的次序排列。第二步，挑选的测试人员对每个数据集进行评定，描述所处理对象的可感知性。举个例子来说，这种评定可基于 ITU-R Rec.500 质量等级级别。表 5.2 列出了等级级别和相应的可感知性以及质量。欧洲的 OCTALIS（通过可信访问链路获得内容）项目中所作的工作表明，经验不同的个体，比如专业的摄影师和研究员，在对水印图像的主观测试中作出的结果相差很大。主观测试对最终的质量评价和测试是有实用价值的，但在研究和开发情况下并不是很有用。

表 5.2　　ITU-R Rec.500 从 1 到 5 范围的质量等级级别

等级级别	损　害	质　量
5	不可察觉	优
4	可察觉，不让人厌烦	良
3	轻微的让人厌烦	中
2	让人厌烦	差
1	非常让人厌烦	极差

在这种情况下，量化失真的度量也就更加有效，并且也使不同方法间的比较趋于合理，因为其结果不依赖于主观评定。表 5.3 列出了用在图像和视频处理中的基于像素的差分失真度量。经过尺度适应后，所列出的大部分度量都可应用于除了图像之外的其它类型的数据中，例如音频数据。现在，在图像和视频的编码压缩领域最流行的失真度量标准是信噪比（SNR），以及峰值信噪比（PSNR）。它们通常以分贝来度量（dB），即，$SNR(dB) = 10\log_{10}(SNR)$。众所周知，这些差分失真度量与人的视觉和听觉系统相互关联得并不是很好。由于复杂的水印嵌入算法以某种方式对这些视（听）觉系统产生影响，因此在数字水印应用中使用差分失真度量可能会

产生问题，使用上面的度量来量化水印处理所引起的失真，可能会造成失真度量的误导。因此，采用一种与人的视觉和听觉系统相适应的失真度量可能会很有用。最近几年，越来越多的研究集中在这种相适应的失真度量上[22－25]，很有可能未来的数字水印系统基准将使用这种质量度量。

表 5.3　　常用的基于像素的视觉失真度量。

差分失真度量	
平均绝对差分	$AD = \frac{1}{XY}\sum_{x,y} \mid p_{x,y} - \tilde{p}_{x,y} \mid$
均方误差	$MSE = \frac{1}{XY}\sum_{x,y} (p_{x,y} - \tilde{p}_{x,y})^2$
L^p—范数	$L^p = \left[\frac{1}{XY}\sum_{x,y} \mid p_{x,y} - \tilde{p}_{x,y} \mid^p\right]^{1/p}$
拉普拉斯均方误差	$LMSE = \sum_{x,y} (\nabla^2 p_{x,y} - \nabla^2 \tilde{p}_{x,y})^2 / \sum_{x,y} (\nabla^2 p_{x,y})^2$
信噪比	$SNR = \sum_{x,y} p_{x,y}^2 / \sum_{x,y} (p_{x,y} - \tilde{p}_{x,y})^2$
峰值信噪比	$PSNR = XY \max_{x,y} p_{x,y}^2 / \sum_{x,y} (p_{x,y} - \tilde{p}_{x,y})^2$
相关失真度量	
归一化互相关	$NC = \sum_{x,y} p_{x,y}\tilde{p}_{x,y} / \sum_{x,y} p_{x,y}^2$
相关质量	$CQ = \sum_{x,y} p_{x,y}\tilde{p}_{x,y} / \sum_{x,y} p_{x,y}$
其它	
全局西格马信噪比	$GSSNR = \sum_b \sigma_b^2 / \sum_b (\sigma_b - \tilde{\sigma}_b)^2$ 其中，$\sigma_b = \sqrt{\frac{1}{n}\sum_{块b} p_{x,y}^2 - \left[\frac{1}{n}\sum_{块b} p_{x,y}\right]^2}$
直方图相似性	$HS = \sum_{c=0}^{255} \mid f_i(c) - f_{\tilde{i}}(c) \mid$ 其中，$f_i(c)$ 是在 256 灰度级图像中灰度级 c 的相对频率。

注：$p_{x,y}$代表一个在原始的未失真的图像中坐标为(x,y)的像素点，$\tilde{p}_{x,y}$代表在嵌入了水印的图像中坐标为(x,y)的像素点。GSSNR 需要将原始图像和嵌入水印的图像分割成 n 个像素点(如，4×4 像素)的块。X 和 Y 分别是行和列的个数。

在确定了参数并选定了失真度量后，下一个要解决的问题就是确定一种有效的健壮性评估策略和视觉表示法。表 5.4 中列出了一些有用的图表，为了对比也列出了可变参数与不可变参数。对所有的评估策略而言，用不同的密钥和多种数据集(如不同的图像)来完成测试是很重要的，然后应该对结果进行平均和划分。如果对作用于一个数据集的两种方法的性能进行直接比较，从而得到该单个数据集的性能评估，那么用不同的密钥重复进行多次测试仍然是非常重要的。在下一段中，我们将简要地解释这四种不同的图表。术语“攻击”是指第七章中

所概要介绍的任何攻击。术语“健壮性”描述水印对这些攻击的抵抗能力。健壮性通常用比特差错率或检测错误来度量，比特差错率定义为错误提取的比特与所有嵌入比特的比率，检测错误定义为 1 减去比特差错的比特数的次方。术语“视觉质量”是指任何一种视觉质量标准，它适合于评价由水印处理所引起的视觉失真。为了说明上述图表的用处，图 5.4 到图 5.7 给出了两种简单图像水印嵌入方法对比的实例[9]。图 5.4 是空间域和多分辨率环境中，扩频隐藏算法的比特差错率对攻击的变化曲线，其视觉质量为 4.5，很明显多分辨率方法的健壮性更高。图 5.5 是空间域和多分辨率环境中，扩频隐藏算法的比特差错率对视觉质量的变化曲线，其攻击为保持 75% 质量的 JPEG 压缩，同样，多分辨率方法表现出更好的性能。图 5.6 是空间域和多分辨率环境中，扩频隐藏算法的攻击对视觉质量的变化曲线，其比特差错率为 0.1，可以清楚地看出，对给定视觉质量，多分辨率方法可以提供更大的压缩比。图 5.7 是空间域和多分辨率环境中，扩频隐藏算法的 ROC 曲线，对应于多分辨率方法的曲线更接近左上角，这表明它的性能更好。这两种方法都运用了扩频调制，但是在不同的域中。一种方法使用空间域，而另一种方法使用多分辨率环境(使用 Daubechies 6 抽头滤波器的三级小波变换)。系统使用一个密钥作为伪随机数生成器的种子，该伪随机数生成器用来生成扩频序列。正如我们用错误比特率来衡量健壮性那样，我们用由 van den Branden lamprecht 和 Farrell 提出的比率来度量视觉质量[22]，并且其中的攻击指的是 JPEG 压缩。所有测试使用的都是 512 × 512、24 位彩色编码的 lena 图。每个测试都重复进行，并且每次都选一个随机的密钥，水印的长度是 100 比特。

表 5.4　　不同图表以及相应的变量和常量

图表类型	参数			
	视觉质量	健壮性	攻击	比特数
健壮性—攻击	固定	变化	变化	固定
健壮性—视觉质量	变化	变化	固定	固定
攻击—视觉质量	变化	固定	变化	固定
ROC	固定	固定	固定/变化	固定

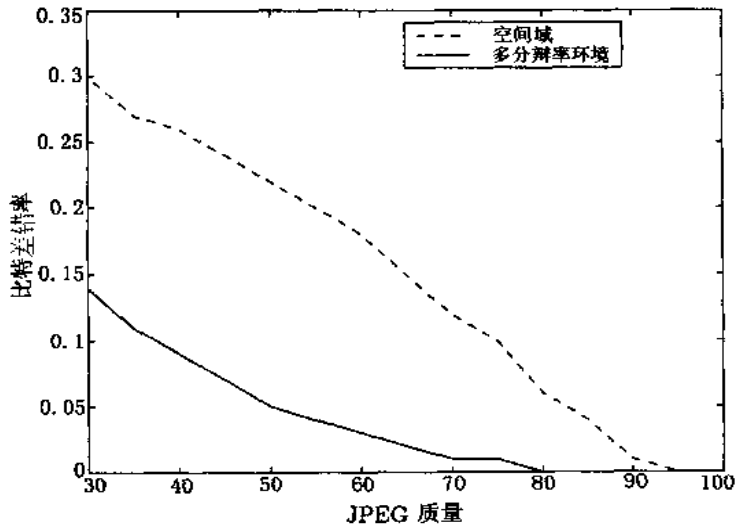

图 5.4　比特差错率对攻击的变化曲线

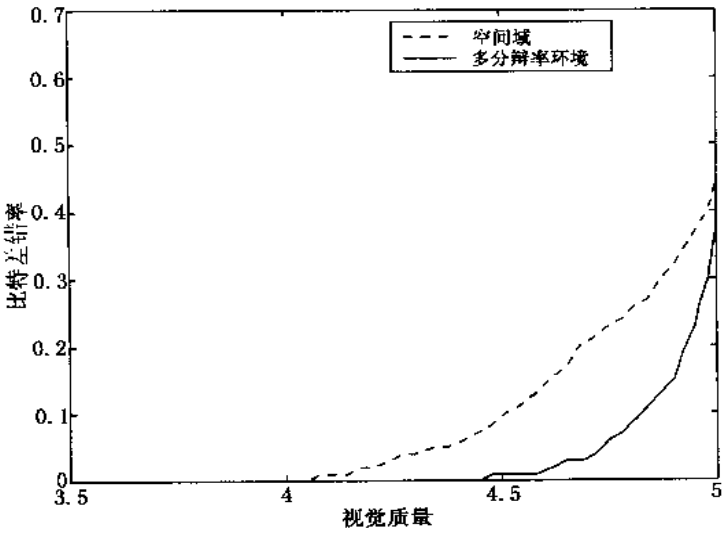

图 5.5　比特差错率对视觉质量的变化曲线

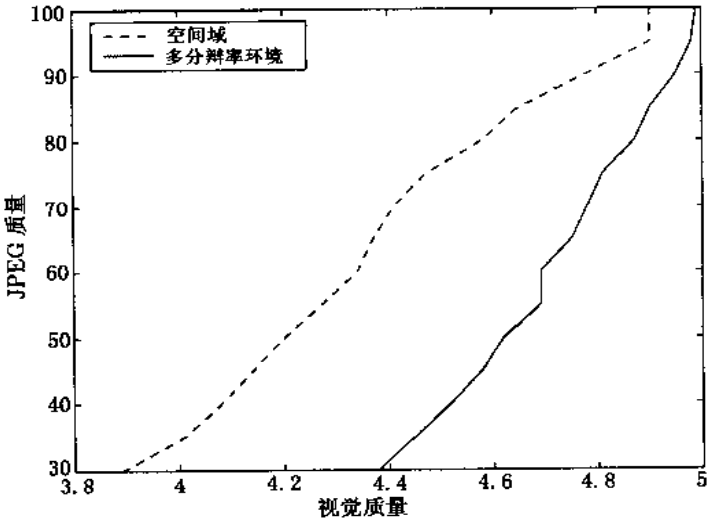

图 5.6　攻击对视觉质量的变化曲线

健壮性—攻击强度图表是将水印健壮性与攻击联系起来的最重要的图表之一。通常这种图表显示了在给定视觉质量的前提下，比特差错（或检测错误）是攻击强度的函数。有几篇论文使用了这种图表，但可惜的是没有明确地给出可视图像质量。这种评价可以对水印健壮性直接进行比较，并给出了该方法针对攻击的所有措施。健壮性—视觉质量图表说明了对某个特定攻击而言，比特差错（或检测错误）与可视图像质量之间的关系。对给定的攻击，这种图表可用来确定在所要求的视觉质量前提下，所能达到的预期比特差错。特别是在一个给定攻击和要求的差错率情况下，确定最低视觉质量时很有用。攻击—视觉质量图表说明了在健壮性给定的情况下，最大可允许的攻击，而它是视觉质量的函数。这种图表给出了在给定视觉质量下，对能承受的水印攻击的直接评估。如果给定视觉质量范围，并且需要估计相应的最大允许失真（比如水印攻击），那么这种图表特别有用。此外，由于在比特差错（或

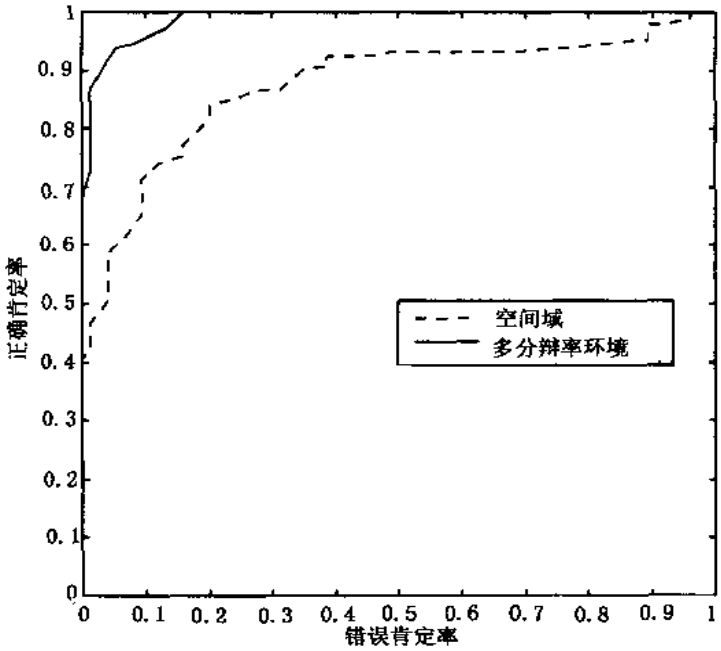

图 5.7 ROC 曲线

检测错误）固定及视觉质量给定的情况下，不同方法健壮性之间的对比，因此这种图表在比较不同水印嵌入方法时也非常有用。对于任何图像，水印检测者都要完成两项任务，即确定图像中是否嵌入了水印和解出嵌入的信息。前者可看作是一个假设检验，在这个检验过程中，检测者要在可能性假设（图像嵌入了水印）和不可能性假设（图像没有嵌入水印）之间作出判断。在此二元假设检验中，可能会出现两种错误，即，在不可能性假设正确时，认可了可能性假设；在可能性假设正确时，认可了不可能性假设。通常称第一种错误为 *I* 型错误或者误肯定，称第二种错误为 *II* 型错误或者误否定。在对所考查水印嵌入方案的所有行为及其可靠性进行评定过程中，接收者操作特性（ROC）图表［26］非常有用。通常，在假设检验中，一个检验的统计量要与一个门限值相比较以确定是这个假设还是另一个假设。用固定的门限值来比较不同的水印嵌入方案可能会导致错误的结果，ROC 图表通过使用多个判决门限从而避免了这个问题。ROC 图表给出了 y 轴的正确肯定部分（*TPF*）与 x 轴的错误肯定部分（*FPF*）之间的关系。正确肯定部分定义为：

$$TPF = \frac{TP}{TP + FN}$$

其中 *TP* 是正确肯定的测试次数，*FN* 是错误否定的测试次数。错误肯定部分定义为：

$$FPF = \frac{FP}{TN + FP}$$

其中 *FP* 是错误肯定的测试总数，*TN* 是正确否定的测试次数。换句话说，ROC 图表给出了由连续变化门限中产生的 TPF-FPF 对。一个理想检测器应具有这样的曲线，从左下角出发到左上角，然后再到右上角。从左下角到右上角的对角线表示一个检测器以相同的可能性随机地选择这一个或另一个假设。因此，检测器的正确度越高，它的曲线就越靠近左上角。通常，用该曲线的积分来度量检测器的性能[26]。为了得到这些图表，要对同样数量的嵌入水印的图

像和未嵌入水印的图像进行测试。如果要评估水印嵌入方法的整体性能，这些测试还应该包括具有不同参数的多种攻击。

5.6.2　水印擦除软件和基准程序

与条件访问和版权保护机制类似，数字水印技术的存在及其潜在的应用可能性促使很多人试图擦除数字水印，很多公开可获得的测试图像水印技术的工具就是实例。Unzign[27]是一个用于 JPEG 格式图像的实用程序。在版本 1.1 中，Unzign 引入了与微小图像变换相结合的像素抖动技术。根据目前研究的水印嵌入技术，该工具好像能有效地擦除或破坏嵌入的水印。然而在擦除水印的同时，Unzign 1.1 版往往会引入不可接受的人为痕迹，现在已发布了改进的 1.2 版本。尽管它减少了人为痕迹，但同时也降低了擦除水印的能力。

StirMark[20,28,29]是一个用于测试图像水印技术健壮性的工具。它给图像提供较小的几何失真，我们将在 7.3.1 小节进行介绍。应用一次 StirMark，仅给图像引入几乎注意不到的质量损失。基于 StirMark 实用程序，对图像水印系统提出了一个一般基准程序。该基准程序使已嵌入水印的图像遭受水印所要抵抗的多种攻击。然后，对这些方法抵抗各种攻击的能力进行平均以用于比较。

5.7　未来和标准化

学术界和工业界对水印技术的关注程度都很高。来自学术界的关注可从数字水印方面的出版物的数量、召开的数字水印和数据隐藏的研讨会的数量上反映出来。来自工业界的关注可从最近几年该领域中成立的公司数量的增加明显地看出来。

除了在大学和工业界的研究活动之外，欧洲共同体资助的几个研究计划以开发实用的数字水印技术为目标来开展研究活动。*TALISMAN*[30](ACTS[1] 的 AC019 计划，即通过标记图像服务和监控访问网络来跟踪作者版权的计划)打算提供欧洲联合服务，它用一个标准的版权机制来保护数字产品，使其避免大规模的商业盗版和非法拷贝。TALISMAN 所期望的成果是得到一个通过数字标记和数字水印来保护视频序列的系统。*OCTALIS*[31](ACTS P119 计划，即通过可信接入链路进行内容提供的计划)是 TALISMAN 和 OKAPI[31](ACTS 051 计划，即用于被保护互操作交互服务的开放内核计划)的后续计划，它的主要目标是得到一个用于公平条件访问和高效版权保护的全球性方法，并通过在因特网和 EBU(欧洲广播联盟)网络上大规模的试验论证它的有效性。

国际标准化组织对水印技术也很感兴趣。比如，新出现的视频压缩标准 MPEG－4(ISO/IEC 14496)提供了一个易于将密码和数字水印结合起来的体制。DVD 工业标准将包含拷贝控制和版权保护机制，该机制使用数字水印来标明多媒体数据的可拷贝状况，比如“一次拷贝”或者“禁止拷贝”标记。

尽管人们都在努力发展和完善数字水印技术，但数字水印技术仍然没有完全成熟和广为人们理解，而且还有很多问题没有得到解决。同时，它的理论基础仍然很弱，大多数系统的设计只是探索性的。

[1] 高级通信技术和服务。

另外一个障碍是在数字水印系统之间进行公正的比较是非常困难的[9]。只要嵌入水印的方法和系统没有以一致的复杂的衡量方法进行评估，就会产生脆弱的、易受攻击的系统，并且有产生事实上的标准的危险，这将导致严重的失败并使整个观念受到怀疑。

因此，对水印的期望要立足于现实。始终要记住这一点，即每一个数字水印系统都在健壮性、水印数据率(有效载荷)和不可察觉性之间有一个权衡。要求每幅图像嵌入10000比特不可见信息并且要它能抵抗所有的攻击，这无论如何都是一个幻想(可行的数目大约要比这低两个数量级)。即使是在可实现的期望下进行设计，水印也只是对非专业人员而言有健壮性，它仍然很容易遭受专家的攻击。

尽管所有权证明是数字水印技术最原始的推动力，但是现在看来，在它作为法律上的证据被接受之前，还要走很长一段路，也有可能这根本就不会发生。在有关版权的应用中，数字水印必须与像密码之类的其它机制相结合才能提供可靠的保护。

尽管如此，在很多应用方面，数字水印仍可为其提供实用的、成功的解决方案。特别是对音频和视频而言，看上去数字水印技术好像要被普遍运用于这些领域中。例如，DVD工业标准的拷贝保护系统将使用数字水印。同样，也有人计划用数字水印来为因特网音频发行提供拷贝保护。使用数字水印的广播监控是另一种应用，很有可能这将在音频和视频中得以广泛地使用。

无论数字水印技术的发展是一个成功的历程还是与此相反，它都值得我们去关注，尽管这个问题至今仍不明朗。数字水印技术仍在发展，但针对水印的攻击技术也在发展。在现实中，切实全面的系统设计对于成功的应用是至关重要的。

参考文献

[1] Emery, O., "Des filigranes du papier," *Bulletin de l' Association technique de l' industrie papetière*, vol.6, 1958, pp.185 – 188.

[2] Weiner, J., and K. Mirkes, Watermarking, no.257 in Bibliographic Series, Appleton, Wisconsin: The Institute of Paper Chemistry, 1972.

[3] Petitcolas, F.A.P.R.J. Anderson, and M.G. Kuhn, "Information Hiding-A Survey," *Proceedings of the IEEE*, vol.87, no.7, 1999, pp.1062 – 1078.

[4] Tirkel, A., et al., "Electronic Water Mark," in *Proceedings DICTA* 1993, dec. 1993, pp.666 – 672.

[5] Tanaka, K., Y. Nakamura, and K. Matsui, "Embedding Secret Information Into a Dithered Multi-level Image," in *Proceedings of the* 1990 *IEEE Military Communications Conference*, 1990, pp.216 – 220.

[6] Caronni, G., "Ermitteln unauthorisierter Verteiler von maschinenlesbaren Daten," Technical report, ETH Zürich, Switzerland, Aug. 1993.

[7] Roche, S., and J.-L. Dugelay, "Image Watermarking Based on the Fractal Transform," in Workshop on Multimedia Signal Processing, IEEE, Los Angels, California, 7 – 9 Dec. 1998, pp.358 – 363.

[8] Braudaway, G.W., K.A. Magerlein, and F. Mintzer, "Color Correct Digital Watermarking of Images," US patent No.5,530,759, 1996.

[9] Kutter, M., and F. A. P. Petitcolas, "Fair Benchmarking for Image Watermarking Systems," in *Proceedings of the SPIE* 3657, *Security and Watermarking of Multimedia Contents*, 1999, pp. 226 – 239.

[10] Boneh, D., and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data," in *Advances in Cryptology, Proceedings of CRYPTO' 95*, vol. 963 of *Lecture Notes in Computer Science*, Springer, 1995, pp. 452 – 465.

[11] Boneh, D., and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data," *IEEE Transactions on Information Theory*, vol. 44, no. 5, 1998, pp. 1897 – 190.

[12] Linnartz, J.-P. M. G., "The 'Ticket' Concept for Copy Control Based on Embedded Signalling," in *Computer Security – 5th European Symposium on Research in Computer Security, vol.* 1485 *of Lecture Notes in Computer Science*, Springer, 1998, pp. 257 – 274.

[13] Bloom, J. A., et al., "Copy Protection for DVD Video," *Proceedings of the IEEE*, vol. 87, no. 7, Jul. 1999.

[14] Schneider, M., and S.-F. Chang, "A Robust Content Based Digital Signature for Image Authentication," in *Proceedings IEEE International Conference on Image Processing* 1996, Lausanne, Switzerland, Sep. 1996.

[15] Xie, L., and G. R. Arce, "A Blind Wavelet Based Digital Signature for Image Authentication," in *Proceedings of the European Signal Processing Conference*, Rhodes, Greece, Sep. 1998.

[16] Hartung, F., J. K. Su, and B. Girod, "Spread Spectrum Watermarking: Malicious Attacks and Counterattacks," in *Proceedings of the SPIE* 3657, *Security and Watermarking of Multimedia Contents*, 1999, pp. 147 – 158.

[17] Hartung, F., and B. Girod, "Fast Public-Key Watermarking of Compressed Video," in *Proceedings IEEE International Conference on Image Processing* 1997, vol. 1, Santa Barbara, California, USA, Oct. 1997, pp. 528 – 531.

[18] Craver, S., et al., "Can invisible watermarks resolve rightful ownerships?" Technical Report RC 20509, IBM research Devision, Jul. 1996.

[19] Petitcolas, F. A. P., and R. J. Anderson, "Evaluation of copyright marking systems," in *IEEE Multimedia Systems*, Florence, Italy, 7 – 11 Jun. 1999.

[20] Petitcolas, F. A. P., R. J. Anderson, and M. G. Kuhn, "Attacks on Copyright Marking Systems," in *Proceedings of the Second International Workshop on Information Hiding*, vol. 1525 of *Lecture Notes in Computer Science*, Springer, 1999, pp. 218 – 238.

[21] Anderson, R. J., "Why Cryptosystems Fail," *Communications of the ACM*, vol. 37, no. 11, Nov. 1994, pp. 32 – 40.

[22] van den Branden Lambrecht, C. J., and J. E. Farrell, "Perceptual Quality Metric for Digitally Coded Color Images," in Proceedings of the European Signal Processing Conference, Trieste, Italy, Sep. 1996, pp. 1175 – 1178.

[23] Westen, S., R. Lagendijk, and J. Biemond, "Perceptual Image Quality Based on a Multiple Channel HVS Model," in *Proceedings of the IEEE International Conference on acoustics, Speech, and Signal Processing*, vol. 4, 1995, pp. 2351 – 2354.

[24] Winkler, S., "A Perceptual Distortion Metric for Digital Color Images," in *Proceedings of the International Conference on Image Processing*, vol. 3, Chicago, IL, Oct. 1998, pp. 399 – 403.

[25] Winkler, S., "A Perceptual Distortion Metric for Digital Color Video," in *Proceedings of the SPIE 3644, Human Vision and Electronic Imaging*, 1999, pp. 175 – 184.

[26] Zweig, M. H., and G. Campbell, "Receiver-Operating Characteristics (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine," *Clinical Chemistry*, vol. 39, no. 4, 1993, pp. 561 – 577.

[27] "UnZign watermark removal software," < http://altern.org/watermark/ >, 1997.

[28] Petitcolas, F. A. P., and M. G. Kuhn, "StirMark 2," < http://www.cl.cam.ac.uk/~fapp2/watermarking/stirmark/ >, 1997.

[29] Petitcolas, F. A. P., and R. J. Anderson, "Weaknesses of Copyright Marking Systems," in *Multimedia and Security-Workshop at ACM Multimedia* '98, 1998, pp. 55 – 61.

[30] "Talisman," < http://www.cordis.lu/esprit/src/talisman.htm >.

[31] "Octalis," < http://www.cordis.lu/esprit/src/octalis.htm >.

第六章　水印技术现状概述

(Jean-Luc Dugelay, Stéphane Roche)

到目前为止,在水印领域里,尽管也有文献考虑音频和视频信号,并且探索其各种安全应用,但大部分出版物所考虑的主要是静止图像的版权问题。本章基于目前最流行的水印应用方面的一些参考文献,讨论目前正在研究的将数据嵌入数字文档的各种技术和方法。

6.1 引　言

在过去的几年里,水印已经成为用来解决数字多媒体(例如,音频、图像和视频)中版权问题和内容认证的主要工具。假设一个所有者想要保护他的图像版权,为了实现这个目标,他在图像里加入一个水印,并且希望不要引起图像视觉上的降质。一旦需要证明他的所有权时,可以通过提取水印来实现,尽管图像可能已被修改过。显然这种情形是根据一种健壮的并且是不可见的签名来实现的(对于其他类型的水印见 5.4 小节)。至少下面三个方面必须考虑:

- **水印中的信息与宿主信号中的信息之间的比率**　这个参数取决于所隐藏的信息特征(例如,身份证、索引值或指示器、明文、标识语等)和宿主信号的特征(例如,文本、图像、视频、声音等),其中前者(所隐藏的信息)部分地取决于所希望的安全服务。
- **由于水印导致的图像降质**　在有失真的信源编码中,这个标准在评价水印系统的性能上是最基本的。见 5.6 节。
- **健壮性**　即对图像"非破坏性"攻击时水印的健壮性。

除此之外,如 5.3 节所述,存在几种提取方式,他们依赖于提取时是否使用原始信息。最后,如 7.4.1 小节所述("死锁"问题),为了设计一种有效的解决方案,必须仔细考虑协议的某些方面。本章主要集中介绍嵌入方法,嵌入方法的选择对于水印特性有决定性的效果,例如:可视性、提取、健壮性和提取方式等。从 1992 年以来,已经提出了许多算法和技术,它们主要来自各种各样的研究领域,如信息伪装术、信源编码和通信等。这里我们没有罗列大量的论文摘要(其中包含很多重复),而是根据目前文献提出的观点和解决方法,对于技术的关键点进行了如下的总体概括(它们的详细叙述将在后面各节进行):

- 隐藏信息的像素点或块的选择;在这部分,给出了私钥、公钥和预测等一些基本概念。
- 隐藏算法操作域的选择;例如空间域或变换域(如,DCT、Mellin-Fourier 变换或小波变换)
- 将隐藏的信息格式编码到宿主信号中的策略(引进冗余和纠错码);
- 将信息和伪装载体融合的方法(调制);基本的思想通常是:在隐藏的信息比特与伪装载体的某些所选择的特征之间构造一个二元关系。
- 水印检测器的优化;这部分主要处理在提取过程中进行的操作,并且它们与嵌入过程中的操作并不是对偶的(因此,它们不可直接从前面的部分推断出)。这个部分还介绍了补偿某些几何攻击和盲模式水印提取的技术。

• 在这章的结论部分，我们将列出从静态图像到视频水印方法的可能的改进与扩展。

我们已经讨论了与目前水印技术有关的所有重要方面，并且列出了许多有关的水印文献。如果读者对于某些特殊的水印方案及其详细描述感兴趣，可以看本章后面的参考文献。

6.2 伪装载体中隐藏位置的选择——密码和心理视觉方面

Kerckhoffs 准则（见 1.2.4 小节）的一个直接应用为：水印算法应该是公开的，并且为了防止水印被偶然地移去，不可以直截了当地得到完整的水印。这可以通过选择水印信息在伪装载体中隐藏的位置来实现。在许多实现方案中，由一个密钥控制的伪随机数生成器决定这些位置（3.2.1 小节阐述了这样一种方法）。只有版权所有者才知道这个密钥，因此他是唯一一个在水印的嵌入和恢复过程中可以获得水印的人。在 6.2.2 小节中阐述了一种更为灵活的方法，它允许恢复过程公开化而又不降低水印的安全性。

除了安全方面外，一个好的水印嵌入位置的选择对于原始图像的视觉失真也起到关键作用。人类视觉系统的精确性随着图像纹理特征的变化而改变。在 6.2.3 小节讨论了这种水印位置的心理视觉问题。

6.2.1 拼凑算法

做为一个例子，我们将讨论由 Bender 等人[1]于 1995 年提出的一种“拼凑”算法。这种算法不是象通常作法那样把一个消息隐藏在伪装载体中，而是简单地回答下面的二元问题：“这个人是否知道在嵌入和恢复一个水印时所使用的密钥?”在拼凑算法中，一个密钥用来初始化一个伪随机数发生器，而这个伪随机数发生器将产生载体中放置水印的位置。

下面概括了拼凑算法的基本思想。在嵌入过程中，版权所有者根据密钥 K_s 伪随机地选择 n 个像素对，然后通过下面的两个公式更改这 n 个像素对的亮度值 (a_i, b_i)：

$$\begin{aligned} \tilde{a}_i &= a_i + 1 \\ \tilde{b}_i &= b_i - 1 \end{aligned} \tag{6.1}$$

这样，版权所有者就对所有的 a_i 加 1 和对所有 b_i 减 1。在提取过程中，也使用密钥 K_s 将在编码过程中赋予水印的 n 个像素对提取出来，并计算这样一个和：

$$S = \sum_{i=1}^{n} \tilde{a}_i - \tilde{b}_i \tag{6.2}$$

如果这个载体确实包含了一个水印，我们可以预计这个和为 $2n$，否则它将近似为零。这种提取方法是基于下面的统计假设的，如果我们在一个图像里随机地选取一些像素对，并且假设它们是独立同分布的，那么有：

$$E[S] = \sum_{i=1}^{n} E[a_i] - E[b_i] = 0 \tag{6.3}$$

因此，只有知道这些修改位置的版权所有者能够得到一个近似值为

$$S \approx 2n \tag{6.4}$$

自从这种算法出现以后，又提出了一些扩展算法[2,3]，目的是为了隐藏多于 1 比特的信息，并且提高这种算法的健壮性（我们很容易验证：一些基本的操作，如图像的一个像素的转换等，将足以骗过版权所有者的）。

6.2.2　公钥密码和公开水印恢复

基于密钥的水印算法有一个明显的缺陷，即它们不允许水印的公开恢复。为了克服这个限制，提出了公钥水印算法，这些系统同时包含公钥和私钥。使用私钥给一个图像加水印，而用公钥验证水印。

Hartung 和 Girod [4] 根据扩频方法，提出了一种公钥水印算法的思想。正如 6.4.1 节所描述的，直接序列扩频技术在扩频和解扩过程中都需要扩频序列 S。不过，由于编码的健壮性，在不知道整个扩频序列的情况下，重建原始信号是有可能的。这样，只有为版权所有者所知的私钥才能够计算整个 S，而公钥只可计算部分 $S(S^{pub})$，因此水印就具有了健壮性。不过，这部分将足以重构水印。

在实际中，公开的扩展序列 S^{pub} 每 N 个比特就有一个比特等于原始序列 S^{orig}，而所有的其他比特以随机的方式选择：

$$S_i^{pub} = \begin{cases} S_i^{orig} & \text{以概率为 } 1/N \\ rand\{-1,1\} & \text{其他情况} \end{cases} \tag{6.5}$$

其他的公钥水印算法见[5,6]。

6.2.3　对于心理视觉水印管理的预测编码

在信源编码中，广泛使用预测模型，它是从一个信号的前一个值预测它的下一个值，前提假设是相邻的样本点或像素点是高度相关的。一般情况下，特别是在一个图像里，相邻像素的值比较接近。因此，粗略地说，在以一种有效的方式对误差进行编码之前，预测编码在预测值和初值之间依然存在计算误差。实际上，可以期望误差的分布以一个很小的方差接近于零，这样可以使用一种有效的二进制方式对误差进行编码(如使用 Huffman 编码)。

在水印应用中，对于心理视觉效应，预测编码是有用的。众所周知，在纹理和边缘区域，人类的视觉系统不太精确，这些地方非常适合嵌入水印[7]。另一方面，人的眼睛对于取值均匀的光滑区域非常敏感，这反而不适合嵌入水印。在预测编码中，误差的分布非常吻合这些特征，所以误差信号可以用来作为水印信号调制的载体。

预测编码由于不需要原始图像[8]，所以对于水印系统也是有益的。这个问题将在 6.6 节进行讨论。

6.3　工作域的选择

正如在拼凑算法中所看到的一样，隐藏操作可以在空间域进行。不过，这种操作经常在变换域里进行。这一节，我们首先介绍文献中所提到的不同工作域，然后阐述采用这种选择的目的。

6.3.1　离散傅立叶变换

离散傅立叶变换(DFT)在信号处理中已被广泛研究，在水印领域里也要考虑它。目的是为了提供控制宿主信号频率的可能性。为了在可视性和健壮性之间获得最好的权衡，选择图像的合适部分来嵌入水印是非常必要的。

给定一个二维信号 f(x,y),DFT 定义为(详细描述见[9]):

$$F(k_1,k_2) = \beta\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} f(n_1,n_2)\exp(-i2\pi n_1k_1/N_1 - i2\pi n_2k_2/N_2) \tag{6.6}$$

这里 $\beta=(N_1N_2)^{-1/2}$, $i=\sqrt{-1}$。而逆 DFT(IDFT)定义为:

$$f(n_1,n_2) = \beta\sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1} F(k_1,k_2)\exp(i2\pi k_1n_1/N_1 + i2\pi k_2n_2/N_2) \tag{6.7}$$

为了在水印和它的载体之间进行相位调制,DFT 对于设计水印是有用的(见 6.5.1 节)。这种变换也应用于将图像分割成多个感觉频段(见 6.3.5 小节)。不过,DFT 更经常以一些派生的方式来使用,例如离散余弦变换或 Mellin-Fourier 变换。下两小节将介绍这两种变换。

6.3.2 离散余弦变换(DCT)

早期,在 JPEG 和 MPEG[10,12]的编码研究中,DCT 已得到广泛的研究,后来 DCT 也考虑了用来在图像[13,14]和视频[7]中嵌入信息。DCT 的概述见 3.3 节。

下面介绍在水印中使用 DCT 的主要论点。DCT 域中的水印嵌入规则对 JPEG 和 MPEG 压缩都具有较强的健壮性,这样水印设计者更容易避免 JPEG/MPEG 压缩攻击。而且以前在信源编码里得出的可视性(即视觉失真)理论能够重新使用,这些研究有助于预测水印对载体图像的视觉影响。最后(但并非最不重要的)一点是为了使计算时间最小,在 DCT 域里水印提供了这种可能性,即,在压缩时直接实现嵌入操作(例如在一个 JPEG 或 MPEG 编码器里)。

下面几节中描述了通过 DCT 进行的几种水印嵌入方法。首先,载体图像的 DCT 系数和水印的 DCT 系数[15,16]可以相加。在更巧妙的算法里,还可以根据水印的比特值构造多个 DCT 系数之间的相关关系[14]。更进一步,还可以根据水印比特去"打乱"图像的量化[7]。

6.3.3 Mellin-Fourier 变换

大多数水印算法在对加了水印的对象进行仿射几何变换后,提取水印时存在许多问题。为了克服这个弱点,Ó Ruanaidh 等人[17]提出了在水印算法中使用 Mellin-Fourier 变换[1]。

Mellin-Fourier 的变换空间是基于傅立叶变换的转换特性,即:

$$f(x_1+a,x_2+b)\leftrightarrow F(k_1,k_2)\exp[-i(ak_1+bk_2)] \tag{6.8}$$

我们很容易证明通过一个平移只有相位被改变。因此,如果这个域(即水印被嵌入的空间)限制在与傅立叶变换的振幅有关的子空间,那么它对于图像的空间坐标平移不敏感。为了对于旋转和缩放不敏感,我们可以考虑对数极坐标变换(LPM),它的定义为:

$$(x,y)\mapsto\begin{cases}x = \exp\rho\cos\theta\\ y = \exp\rho\sin\theta\end{cases}\qquad \rho\in R,\ \theta\in[0,2\pi] \tag{6.9}$$

显然,在笛卡儿坐标系里,任何元素(x,y)的旋转将对应于对数坐标系中的平移。同样,笛卡儿坐标系中的缩放将对应于极坐标系中的平移。见图 6.1。

使用坐标系的适当调整,旋转和缩放都能变成一种平移。这样,这种平移的不变性特性能用于构造一种空间,这种空间对水印图像的任何旋转或缩放操作都不敏感。

[1] 在[17]中,提到了这种变换也可用于 Digimarc 的 PictureMarc

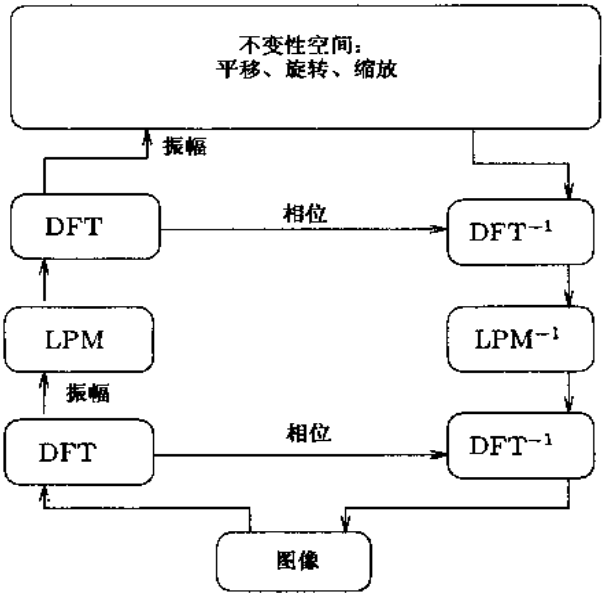

图 6.1　Mellin-Fourier 变换和它的相关不变性

6.3.4　小波域

在即将到来的图像压缩标准 JPEG－2000 中，小波正成为一种关键技术。在目前的几种出版物里，这种技术已经应用于图像水印。现在人们对小波的热情，就像当初在 JPEG 中提倡使用 DCT 一样（即，使用 JPEG－2000 有损压缩可防止擦除水印，并且对图像降质的可视性可以重新使用以前有关信源编码的研究结果。另外，提供了在压缩域嵌入水印的可能性）。除了这些标准，小波的多分辨率思想对根据健壮性和视觉效应设计一个好的信息在载体中的分布方式（即定位）是很有用的。在[18－21]中，详细阐述了有关小波变换和它在图像处理和编码中的应用。

简单地说，小波变换是对图像的一种多尺度、空间—频率分解。图 6.2 表明了 Lena 图像的三级分解，该图属于 M. Antonini 的版权。左上角（LL_3）是最低尺度因子的最低频率段。同样的分辨率下，HL_3块包含有水平方向最高频率和垂直方向最低频率段的信息。同样，LH_3块包含了在最低的尺度因子下，水平方向最低频率和垂直方向最高频率段的信息。以此类推，可以得到中间的和最高的分辨率情形。

构造这些不同分辨率等级的一个方法是将二通道滤波器组与降抽样过程相级联，见图 6.3。二通道滤波器组必须是正交的，并且通过下面的方程来定义：

$$\begin{aligned} H(\omega) &= \sum_k h_k \exp(-jk\omega) \quad \text{高通} \\ G(\omega) &= \sum_k g_k \exp(-jk\omega) \quad \text{低通} \end{aligned} \tag{6.10}$$

信号分解的迭代过程定义为：

$$\begin{aligned} c_{j-1,k} &= \sum_n h_{n-2k} c_{j,n} \\ d_{j-1,k} &= \sum_n g_{n-2k} c_{j,n} \end{aligned} \tag{6.11}$$

信号重构的迭代过程定义为：

$$c_{j,n} = \sum_k h_{n-2k} c_{j-1,k} + \sum_k g_{n-2k} d_{j-1,k} \tag{6.12}$$

在水印研究中，有许多使用小波变换的尝试，我们将简要地讨论一些水印方案。Wang 和 Kuo [22]表明，一种多门限的小波编码方案可以寻找重要的小波系数。他们假设，在一系列信号处理操作之后，这些系数不会有很大的改变。而且，如果这些系数有明显的改变，那么重构出来的图像在视觉上就与原始图像有很大的差别。与预先确定一组系数（例如各种低频系数）的方法相反，这种方法依赖于图像来选择系数。因此，这种方法适合于纹理图像以及光滑图像。

Kundur 等人[23]描述了一种基于小波融合的水印嵌入方法。它的方式是在不同的分辨率

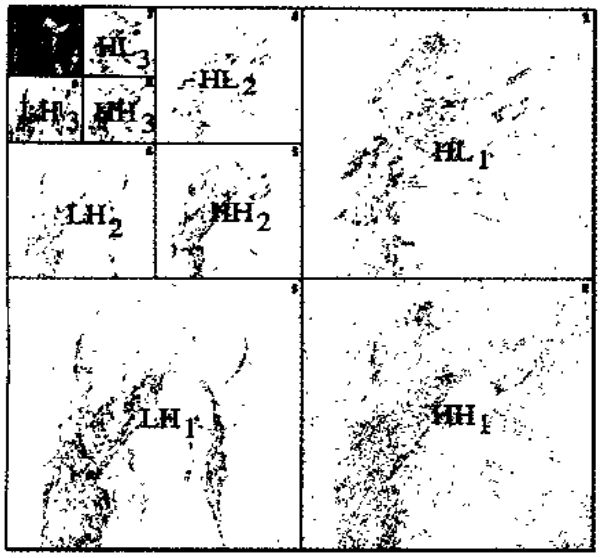

图 6.2 多尺度分解

水平下将水印和图像的小波系数相加。在相加之前，水印的小波系数使用一种人类视觉模型约束（它建立在一种显著性的度量方法上[24]）进行调制。

进一步，Xia 等人[25]提出了一种基于小波变换的分级水印提取方法，其目的在于，如果加水印的图像失真不太严重的话，可以节省计算量。它的基本思想是，用离散小波变换（DWT）将接收到的图像和原始图像（假设是已知的）分解为四个频段（即只做一级分解）。然后，计算所收到的图像和原始图像 HH_1 段的 DWT 系数之间的差值，将它与加在 HH_1 段的水印进行比较，计算它们之间的互相关函数，如果互相关出现一个峰值，那么认为水印被检测到了，否则再考虑同一级的其他频段（也就是 HL_1，LH_1）。如果水印还是没有被检测

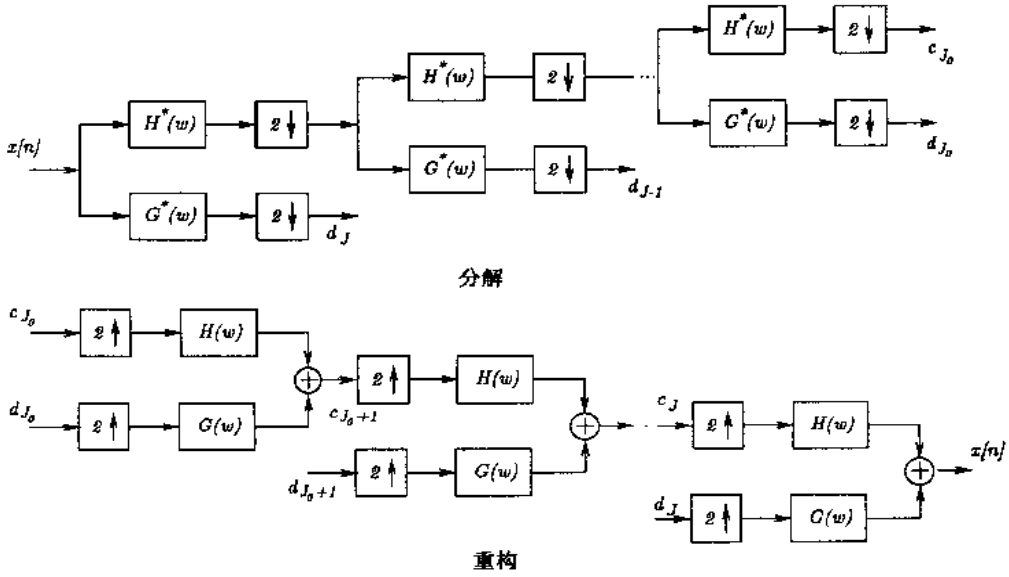

图 6.3 小波变换

到，就计算下一级的 DWT（即二级分解），并且重新尝试检测水印。这个过程一直进行，直到检测到了水印为止或者已经达到了 DWT 的最后一级。

基于 DWT 的其他一些方案也在参考文献[25－28]中提出了。

6.3.5 在感觉频带里分割图像

水印方案的一个关键要求就是水印的不可视性。Delaigle 等人[29]已经开发出了基于人类视觉系统的一种掩蔽模型，该模型使水印对人的视觉影响最小。最终目标就是设计一种迭

代方案，该方案充分考虑感知模型，并将水印保持在视觉门限以下[30]。对于图像和视频，其掩蔽特性还处于研究之中。不过，掩蔽的原理在音频编码领域已经被使用[31]，正如图 6.4（N.Moreau 的版权）所描绘的那样，一个频带的能量可以导致该频带掩蔽其相邻的较低能量的频带。同样，作者假设人类的视觉系统把视觉刺激分割成若干个成分 [32,33]，每一个成分根据三个参数来描述：

- 在视觉域中的位置；
- 空间频率（从傅立叶变换的振幅中计算出来）；
- 方向（从傅立叶变换的相位中计算出来）。

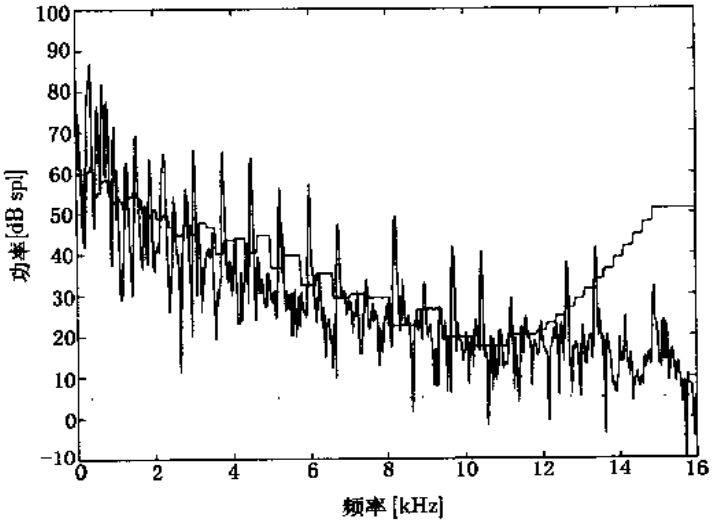

图 6.4　声音掩蔽特性

这些成分通过不同的通道从眼睛传送到大脑皮层。当一个通道成分由于一个能量较高的成分在其邻近通道而变得不可见时，掩蔽效果就产生了。简单地说（更详细的阐述见[29,34]），使用一种图像分析技术能够把图像分割成几个通道（例如 Gabor 滤波），然后计算出每一个通道的局部能量，最后，一个基于通道的频率、方向和位置的对比函数便被计算出来了。这就决定了一种心理视觉掩蔽，使得能量低于这个掩蔽的每一种信号都是不可见的。

6.4　对水印比特进行格式编码

本节将讨论在嵌入水印之前，对水印比特进行格式编码的问题。某些技术可以使任何比特串作为一种水印直接被嵌入，另外一些技术则要求在嵌入之前对水印比特进行变换。

6.4.1　扩展频谱

目前扩展频谱技术用于水印已经引起了人们的许多兴趣[35,36]，其原因非常类似于在伪装术里使用扩频技术的观点。一般来说，用作水印的信息相对于宽带的伪装载体（即图像）而言属于窄带信号。在通过秘密通道（图像）传送信息（水印）之前，扩展频谱技术可以使信息与频带匹配。我们知道，高频区水印嵌入对水印信息的不可视性非常有利，但对健壮性效果极

差，而低频区水印嵌入对健壮性有利，但由于视觉效果无法接受而不可用。扩展频谱技术可以通过将一个低能量信号嵌入到每个频带，来缓解这些矛盾。通过使用密钥来控制伪随机数生成器，扩展频谱技术还提供了保护水印私有性的可能性。

现在我们将介绍两种主要的扩展频谱方法，这些方法在3.4节已经提到过，更详细的信息可以从[37]得到。

直接序列扩频是使用一个宽带伪噪声信号来对原始信号进行时间调制，如图6.5所示。扩展后的信号看上去像伪噪声信号。特别地，即使原始信号是窄带的，但扩展后的信号和伪噪声信号具有相似的频谱。在接收端，通过使用相同的伪噪声信号对接收到的信号进行解调，原始信号就可以被重构出来(见图6.6)。即使在数据传送过程中某些频率丢失了，原始信号仍然可以准确无误地恢复出来，因为原始信息被放置在几个频段里。跳频扩展频谱技术，是使用一个随机过程来对载体的频率进行改变，使其频率值落在一个很宽的范围里，因此，被调制的信号具有很宽的频谱(见图6.7)。

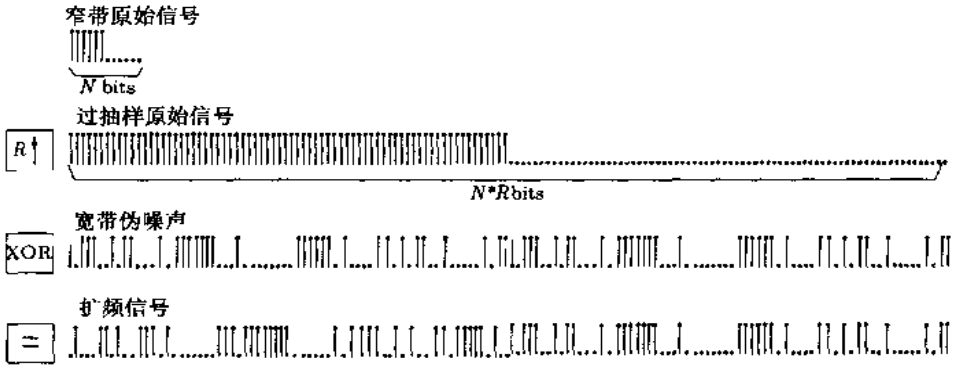

图6.5 直接序列扩频的扩频过程

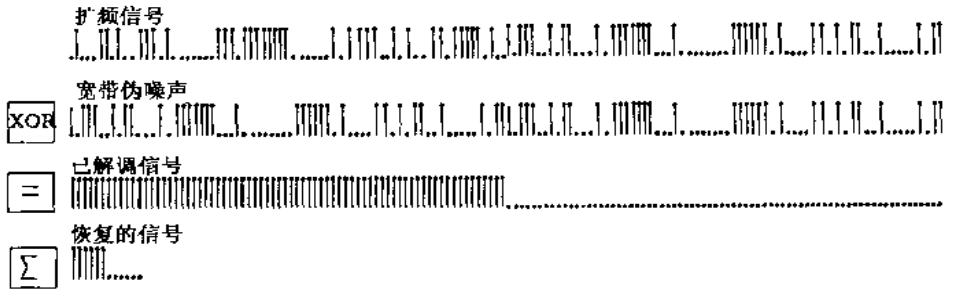

图6.6 直接序列扩频的解扩过程

在这两种方法里，接收到的信号和随机信号之间的再同步是恢复过程的主要问题，它需要了解调制过程中使用的随机过程的特性。特别是对于水印问题，在对图像进行几何攻击后，如何同步的问题就产生了。

现在我们介绍由Hartung等人[36,38]提出的一种实用的水印方案。这种方法是一个视频方案的一部分，不过它也适用于静止图像。令 $a_j(a_j \in \{-1,1\})$ 为一个二进制信号。从 a_i 我们导出它的一个临时扩展信号 b_i

$$b_i = a_j, \qquad j\,cr \leqslant i < (j+1)\,cr \tag{6.13}$$

这里我们记 cr 为片率。为了获得能直接嵌入图像 v_i 中的水印信号，要在 b_i 和一个伪随机信号

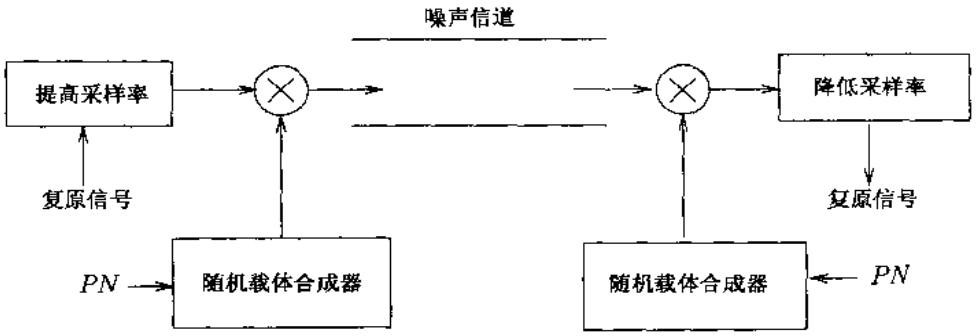

图 6.7　跳频扩频技术

p_i之间进行调制：

$$w_i = \alpha b_i p_i \tag{6.14}$$

其中 α 是控制健壮性和可视性权衡的强度因子。最后的嵌入公式为：

$$\tilde{v}_i = v_i + \alpha b_i p_i \tag{6.15}$$

$\tilde{v}_i$表示加水印后的图像。

水印恢复的基本思想是，对接收到的信号进行解调，并在每一个二进制信息片段里，将每个信号成分相加：

$$\begin{aligned} s_j &= \sum_{i=jcr}^{(j+1)cr-1} p_i \tilde{v}_i \\ &= \sum_{i=jcr}^{(j+1)cr-1} p_i v_i + \sum_{i=jcr}^{(j+1)cr-1} p_i^2 \alpha\, b_i \end{aligned} \tag{6.16}$$

若假设信号 p_i是零均值并且与 v_i统计独立的，我们可以期望 s_j的值为：

$$s_j \approx cr\,\alpha\, a_j \tag{6.17}$$

于是 a_j的值可通过下列公式计算：

$$a_j = sign(s_j) \tag{6.18}$$

6.4.2　低频水印设计

普遍认为，水印必须能够经受住所有的图像操作处理，当然这些操作处理不会破坏图像的可用性。在这些操作中，许多是基于低通滤波的（例如，JPEG 压缩、改变大小等）。这就是为什么许多作者[35,39]提倡设计低通水印的理由，尽管低通水印是在感觉重要区域施以影响从而造成比高通水印更强的人为视觉痕迹。Braudaway[39]使用傅立叶变换创造了一种低通水印，下面描述了这一过程（见图 6.8）。

假定初始水印为 $w_{init}(i,j)$，它是具有原始图像尺寸的有冗余的水印，然后产生缩小尺寸的水印 $w_{redu}(i',j')$并计算它的傅立叶变换，记为 $W_{redu}(u,v)$。然后 $W_{redu}(u,v)$用零来填补，直到达到初始的水印尺寸。最后，计算 $W_{zero-pad}(u',v')$的反傅立叶离散变换，则所得的图像 $w_{low-pass}(i,j)$就是低通水印。这个过程确保在 $w_{low-pass}(i,j)$中，只有低频存在。

6.4.3　纠错码

为了改进基本算法使水印具有很好的健壮性，许多论文[29,40－42]提出使用纠错码技

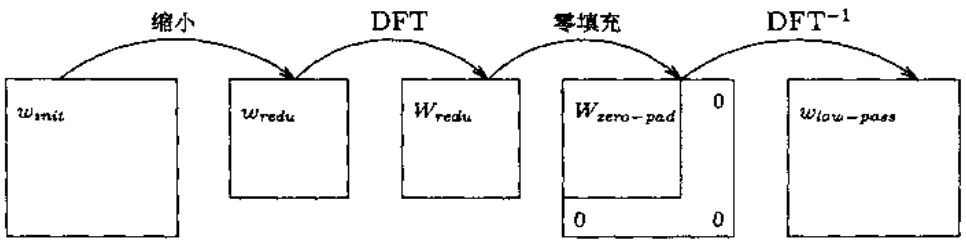

图 6.8 低通水印的产生

术。如果我们将水印问题比作是在一个带噪声的信道中传输信号的话,那么自然就会用到这个方法。这个模型把图像看作是一种信道,把不同的攻击看成是噪声信号。在信道编码中,纠错码被广泛地使用,我们也可以将它用于水印问题中。不幸的是,据我们所知,只有少数一部分论文提到了纠错编码应用于水印系统得到的重要结果。这里部分地解释了纠错编码对水印系统没有产生什么影响的原因。

在传统的信道编码中噪声信号一般可以有效地模型化为高斯噪声,而水印应用正相反,它必须考虑各种攻击(见第 7 章),这些攻击可看作是具有不同特性的噪声,那么高斯假设在这里不再有效了。在这种背景下,设计一种独特的编码来满足不同攻击的要求是很困难的。

由于对水印的支持很有限并且是静态的,特别是如果你考虑一个静止图像,那么为了使用一个纠错码而扩展水印的原始二进制流,有时是根本不可能的。对于音频或视频数据情形就不一样了,如果需要,水印就可以扩展到几个帧上。为了这个目的,水印和图像间的比率是一个关键因素。

对水印使用纠错码依然是个未解决的问题,它要求设计的纠错码非常简洁紧凑并且考虑各种不同的噪声情形。

6.5 水印和载体合并

到目前为止,原始水印数据和原始图像都是分别进行处理来满足水印的各种要求的。这一节将阐述合并水印和载体的技术,以得到加水印的图像。特别是合并的步骤中要根据水印能量和图像能量的关系来处理健壮性和视觉之间的权衡。

6.5.1 相位调制

令 $F(I)$是图像 I 的 DFT,当 I 为实数时,DFT 的特性表明 $F(I)$是复数;如果图像是对称的,那么 $F(I)$为实数。这就得到图像的振幅和相位表示。

Ó Ruanaidh 和 Dowling 在[43]中指出相位调制可能更适合于健壮性水印。两方面的考虑使他们得出这个结论。第一种考虑是在图像的可理解性上,相位成分是非常重要的。根据 Hayes 的文章[44],DFT 的相位成分比振幅成分有更大的心理视觉影响。因此,如果在相位成分中引入带有较高冗余度的水印,那么为了移去水印,恶意的攻击者将会给图像质量带来令人无法接受的破坏。第二种考虑是从通信理论得出的结论,众所周知,相位调制对噪声信号具有较强的健壮性。另外,Ó Ruanaidh 提到相位水印能够经受住图像对比的改变。

为了获得一幅真实的加水印的图像,相位调制水印必须考虑 DFT 变换的图像的反对称特性。因而,如果 δ 是水印大小,频率成分按下列方式调制(我们使用与 3.3 节同样的定义):

$$\phi(k_1,k_2) \leftarrow \phi(k_1,k_2) + \delta \tag{6.19}$$

$$\phi(N_1 - k_1, N_2 - k_2) \leftarrow \phi(N_1 - k_1, N_2 - k_2) - \delta \tag{6.20}$$

如果一个 DFT 系数 $F(k_1,k_2)$有足够高的能量以致于对图像有重要的影响，即如果：

$$A(k_1,k_2)^2 / \sum_{r_1=1}^{N_1-1}\sum_{r_2=1}^{N_2-1} A(r_1,r_2)^2 > \varepsilon \tag{6.21}$$

那么它就适合加入水印。这样就能保证，对水印的攻击将使图像质量损失到令人无法接受的程度。

6.5.2　振幅调制

如前所述，在傅立叶域振幅调制似乎并不适合于水印目的，这是因为傅立叶振幅成分对图像质量的贡献较小。不过，在图像的空间域或它的某些部分，振幅调制能够成功地应用。例如，*Kuntter* 等人[8]提出了一种图像的蓝色成分的振幅调制。令 $I=(R,G,B)$为一幅彩色图像，$p=(i,j)$为图像上的任一位置(见 6.2.1 小节)，并且 $m=\{0,1\}$为水印的二进制数。这个新的加了水印的蓝色成分 $\tilde{B}$ 通过下面的公式由 B 得到：

$$\tilde{B}_{ij} \leftarrow B_{ij} + (2m-1)qY_{ij} \tag{6.22}$$

这里 Y 是亮度成分(见 3.1 节)，q 是用来权衡健壮性和可视性的强度系数。注意到，假设人眼对亮度高的值表现出较低的敏感性，那么根据这个公式，亮度高的值就很适合隐藏较强的水印。

6.5.3　保持亮度均衡的合并

这种方法是基于图像区域块的分类。分类方法是同一个区域块中各像素颜色接近。在[45]中提到的算法概括如下：

- 选择图像块，在这些块中将根据密钥来嵌入水印。
- 在每一个块内部，将像素分类，使其要么是高对比区域的一个成员，要么是低或中对比区域的一个成员。因此，对于每个块，定义了两种区域 R_1和 R_2，它们的平均亮度值能够被计算出来。
- 根据一个预定义的栅格，把 R_1和 R_2分别分割成两个带标号(A 或 B)的区域。因此四个子区域被定义为：$R_{1,A}$，$R_{1,B}$，$R_{2,A}$和 $R_{2,B}$。它们各自包含的像素个数为：$n_{1,A}$，$n_{1,B}$，$n_{2,A}$和 $n_{2,B}$，相应的平均亮度值分别为：$Y_{1,A}$，$Y_{1,B}$，$Y_{2,A}$和 $Y_{2,B}$。
- 令 m 为准备隐藏在一个块中的水印比特。那么嵌入操作定义为：

如果 $m=0$　　$\tilde{Y}_{1,A} - \tilde{Y}_{1,B} = -l$

　　$\tilde{Y}_{2,A} - \tilde{Y}_{2,B} = -l$

如果 $m=1$　　$\tilde{Y}_{1,A} - \tilde{Y}_{1,B} = l$

　　$\tilde{Y}_{2,A} - \tilde{Y}_{2,B} = l$

这里 l 是嵌入水平。由于要维持 R_1和 R_2的平均亮度值。我们可以定义另外两个方程：

$$\frac{n_{1,A}\tilde{Y}_{1,A} + n_{1,B}\tilde{Y}_{1,B}}{n_{1,A} + n_{1,B}} = Y_1 \tag{6.23}$$

$$\frac{n_{2,A}\tilde{Y}_{2,A} + n_{2,B}\tilde{Y}_{2,B}}{n_{2,A} + n_{2,B}} = Y_2 \tag{6.24}$$

根据要被隐藏的信息的比特值 b，可以计算出这些方程中 $\tilde{Y}_{1,A}$，$\tilde{Y}_{1,B}$，$\tilde{Y}_{2,A}$和 $\tilde{Y}_{2,B}$的值。

- 来自于同一个区域的每一个像素按偏移量 $\delta_{i,j}$来修改：

$$\delta_{i,j} = \tilde{Y}_{i,j} - Y_{i,j} \tag{6.25}$$

提取过程的最初三步与嵌入过程的前三步相同。然后，我们计算 $\sigma_1 = \tilde{Y}_{1,A} - \tilde{Y}_{1,B}$和 $\sigma_2 = \tilde{Y}_{2,A} - \tilde{Y}_{2,B}$。从这两个值的符号可以推断出比特值 b。而且，它们的大小也能够为比特提取的精确性提供某些信息，进而提高了所提取比特的可信程度。在版权应用领域，仅仅需要考虑几十个比特。因此，在消息里可以加入一些冗余信息。

6.5.4 基于 DCT 系数量化的合并

正如 6.3.2 小节所提到的那样，DCT 域也适合于嵌入水印。在这一节里，我们将描述一种将图像的 DCT 系数和水印的 DCT 系数组合在一起的方法。这种方法是基于原始 DCT 系数的一个受限的量。由 Koch 和 Zhao[13,14]于 1995 年提出的嵌入算法概括如下。

假定有一个比特序列 m 要隐藏在一幅图像里，发送者根据密钥在这个图像里选择 k 个 8×8 大小的块。如果每一个被选择的块的 DCT 系数 $|a_{i,j}|_{i,j=1,\cdots,8}$没有被计算，那么现在就应该计算它们。然后版权所有者观察一个块的两个 DCT 系数和下一个水印比特之间的关系。如果需要，那么他将对系数进行修改，以使在两个选择好的系数间构成一种特定的关系。更详细地说，对所有的水印比特 $m_k, k=1,\cdots,l(m)$，做如下操作：如果($m_k=1$ 且 $(a_{1,2})_k > (a_{2,1})k$)或者($m_k=0$ 且 $(a_{1,2})_k < (a_{2,1})_k$)，那么所需要的关系已经满足，对系数不作任何改变。否则，就要交换这两个 DCT 系数来满足相应的关系。然后，版权所有者计算所有修改过的图像块的逆 DCT 变换。

由于这些修改，一个图像块恰好嵌入一个水印比特。提取水印时，要检测系数 $(a_{1,2})_k$和 $(a_{2,1})_k$之间的关系。如果 $(a_{1,2})_k < (a_{2,1}{}_k)$，则水印的第 k 个比特认为是 0，否则是 1。

不幸的是，这个算法会产生一些可视的人为痕迹。为了优化这种方法，提出了几种改进措施，包括图像块的选择方式(即，如果版权所有者在不产生可视性改变的情况下，就得不到所期望的关系，那么就不用这些图像块来嵌入水印比特)、被考虑的 DCT 系数的个数(3 个而不应只是 2 个)以及修改 DCT 系数值的方式。

6.5.5 分形编码中基于块替换的合并

Puate 等人[46]提出了基于分形编码的水印嵌入方法。在这里，我们将首先简要回顾分形图像编码。

令 I_{orig}为要压缩的图像，I_0为任意给定的初始图像，I 和 J 为两个普通图像，$d(I,J)$是一个距离测度，它用来度量两个图像之间的差异程度。变换 τ 把一个图像映射到另一个图像上，如果满足 $d(\tau(I),\tau(J)) < \sigma d(I,J)(0<\sigma<1)$，则称这种变换为收缩的，其中 σ 是 τ 的收缩因子。那么随着 n 趋于无穷大，$\tau^n(I_0)$收敛到一个吸引子 I_a，这里 I_a独立于 I_0。而拼贴定理可叙述为：如果存在一种变换 τ，使得 $d(I_{orig},\tau(I_{orig})) < \varepsilon$，$\tau$ 是收缩的并且具有收缩因子 σ，那么 $d(I_{orig},I_a) < \varepsilon/(1-\sigma)$。编码器的任务就是确定一种变换 τ(或称一种“IFS 码”)，使得 $\tau(I_{orig})$尽可能地相似于 I_{orig}，当然是在用来表示 τ 的比特数限制范围内。IFS 码 τ 被传送给译码者，然后他计算吸引子 I_a并作为重构图像。图像的重构错误由拼贴定理来给出其上限。

在 Jacquin 的方法[47]中，编码器从原始图像 I_{orig}中找出变换 τ，而变换 τ 表示为仿射变换

τ_i的总和，其中每个值域块 R_i对应一个仿射变换 τ_i，τ_i将一个特定的定义域块映射到相应的值域块 R_i中。值域块和定义域块在不同的分辨率下被分割，一般值域块的大小为 $B\times B$，而定义域块的大小为 $2B\times 2B$(通常 B 是 8 个像素)。对于每一个值域块，编码器从一个适合变换和选择的定义域块中寻找最好的拼贴匹配。在搜索过程中，候选的定义域块分三步进行变换，即对该块的亮度值进行二次抽样、等距、缩放和平移操作，如图 6.9 所示，该图是一个在值域块和定义域块之间通过旋转和明亮反转变换进行匹配的例子。

为了解码出一个图像，将接收到的 IFS 码 τ 应用于初始的任意一个图像 I_0，以形成图像 I_1。从 I_1到 I_2，从 I_2到 I_3等等重复这个过程，直到得到 I_a。一般来讲，收敛过程大约需要不超过 10 次的迭代。在[47,48]中，可发现更多有关分形图像编码的信息。

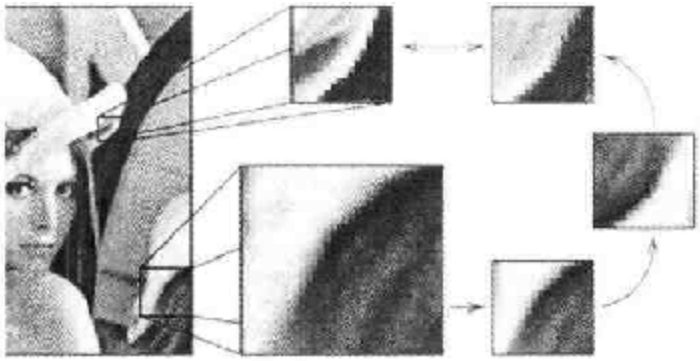

图 6.9　分形编码

用作水印目的时，其关键思想是在匹配时对搜索窗施加额外的限制。例如，不是对整个图像进行扫描，而是定义为两个部分，如图 6.10 所示。根据所隐藏信息的比特值，只考虑它们中的一个。更确切一点，插入过程可概括如下：

- 给定 m 为要隐藏的比特序列，其冗余度为 U；
- 对每个比特 m_k，选择 U—值域块(通过一个密钥随机选择，只有用户知道)：

(1) 如果 $m_k=1$，那么通过在类型为 0 的局部搜索区域搜索它的候选定义域 D_i来对 R_k进行编码(见图 6.10)。

(2) 否则(如 $m_k=0$)，则在类型为 1 的局部搜索区域检索 D_k，来对 R_k编码。

- 正如“经典”的分形图像编码情形，在类型 1 和类型 0 的搜索区域的联合中搜索 D_k，来对剩余的值域块 R_k进行编码。
- 计算吸引子。

提取过程包括两个步骤：

- 从吸引子和每个标识的区域 $\tilde{R}_k$(通过密钥来表示)，寻找相关的定义域块 $\tilde{D}_k$；如果 $\tilde{D}_k$属于类型 1 的区域，那么嵌入的是“1”，否则嵌入的是“0”。
- 对于每个比特 m_k，通过考虑 U 响应集里“0”或“1”的个数多者来作出最后的决定。

这种方法利用了这样一个事实，即一个吸引子是不变的，也就是说，它可以无错误地被编码(即，除了某些块可能存在多个答案外，其他的块使用相同的 IFS 码)。不过，对于大多数攻击，这种特性没有再验证过。

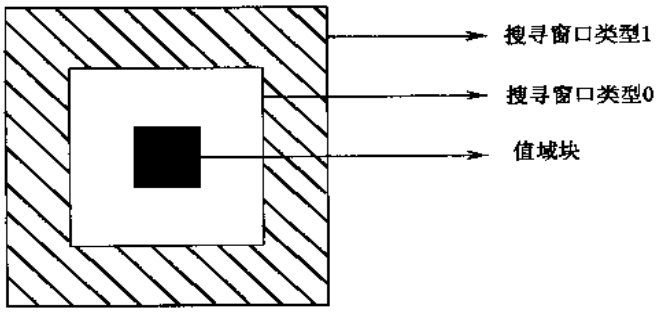

图 6.10 在分形编码中的搜索窗口

6.6 水印检测器的优化

大部分水印检测器的算法可以直接从插入算法导出，不过，有些算法在检测过程中是特定的。这些算法是本节的核心。通过考虑各种可能的亮度值或几何操作，可以改善水印检测的健壮性。

6.6.1 图像预滤波

许多嵌入算法取决于宿主信号的统计假设。为了使得这个假设在提取过程中尽可能可靠，加水印的图像在信息开始提取之前要进行滤波。继续 6.4.1 小节提出的例子，作者[38]提出从滤波后的 $\tilde{v}'$ 而不是 $\tilde{v}$ 中进行提取操作。在振幅调制的情形(见 6.5.2 小节)下，Kutter 表明，从一个均值为 0 方差很小的高斯信号中来进行水印恢复更为适合。而这样一个信号可以从加水印的图像与下面的掩模进行卷积后获得(见[50])。

$$h=\frac{1}{12}\begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 12 & -1 & -1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}$$

6.6.2 重定位和尺寸调整使相位相关性最大

克服一些诸如旋转或缩放之类的几何攻击的简单方法，就是以一定的精确度对所有可能的攻击进行相反的变换，然后，确定哪种反变换可提供最好的结果。原始文档的知识(也就是知道原始图像或嵌入的水印)通常能帮助我们如何去选择作用于加水印的图像的正确的(或至少是适当的)变换。

为了考虑各种可能的加水印图像的再定位和大小调整，有两种策略可供选择。第一种策略(见 6.3.3 小节)是在旋转缩放不变的空间中嵌入水印。第二种策略是对被攻击图像的几何变换做后验估计，然后在提取水印前进行相反的变换。换句话说，第一种方法是预防性的，而第二种方法是治疗性的。相比于预防性方法，后一种方法的主要优点为：通过把水印位置限制

在一个可能很小的不变子空间里而使水印的容量不会降低。不过，这种优点又被完成有效变换估算的困难性冲淡了，而这种变换估算恰恰是治疗性方法所必需的。Braudaway[39]提出了一种方法，他的方法是基于最大相位相关性的，并且要求把原始水印或者原始图像的一部分作为一个参考平面。绘制一个相位相关最大值的一个三维栅格图，栅格的轴有水平尺度因子、垂直尺度因子、相对于加水印图像(可能被操作过)的相关性参考平面的旋转角度。这种三维栅格图确定了相位相关的最大值，这个最大值的坐标给出了水平尺度因子、垂直尺度因子和最终的旋转角度。

6.6.3　自适应门限值改进决策的健壮性

绝大多数处理比特值恢复的决策过程是将提取出来的值与一个门限值进行比较。不幸的是，由于对加水印的图像的攻击，最优的门限值会不同于嵌入和提取阶段的门限值。Kutter[8]建议在水印里引入一些固定的比特，使得在提取过程中可计算出新的最优的门限值。

6.7　从静态图像到视频的扩展

当考虑视频而不是图像时，水印算法所包含的各种权衡(比率、可视性和健壮性)自然要进行显著的调整。对水印信息和宿主信息之间的比率的调整，变得不太重要了。不过，由于时间维数的原因，因水印而导致的视觉失真更加难以控制。而且，攻击的种类在不断地增加，例如，一个视频水印方案必须考虑一些可能的帧率的改变。总之，当需要实时水印时，计算的复杂度问题就变得最基本最重要了。到目前为止，有关水印的大部分研究主要针对静态图像，因此，除了音频和图像，视频水印仍然是一个未解决的问题。诸如“如果加入随时间变化的策略，那么一个视频水印的策略就能执行得更好吗?”的一些基本的问题目前还处在研究之中。不管怎么样，大多数针对静态图像的一些基本思想也可用于运动图像，例如，在[51]中基于宿主信号和水印的 DCT 系数相加的方法，能在 JPEG 流和 MPEG 流中的 I 帧上实现(见[36]，[52])。不过与视频的时间维数有关的一些新问题必须解决。在特定情况下，正如作者们所强调的，存在一种可能的漂移问题。总之，由于 MPEG 视频编码器和译码器中的运动补偿模式，那么在嵌入 I 帧时导致的一些失真也会破坏相邻的 P 帧和 B 帧[53]。

6.7.1　运动矢量量化

与前面所提到的基于图像 DCT 系数修改的方法类似，Jordan[54]提出了根据要隐藏的消息比特对运动矢量的量化进行修改，来实现将一个消息嵌入 MPEG 视频流中的方法。正如前面各小节所看到的，已经有许多种方法在水印信息和图像特性之间构造一种关系。在[7]中，不是使用最近的整数值来量化浮点 DCT 系数的值，作者的作法是，当隐藏的信息比特是“0”时，考虑取最接近的偶数做为量化值，当隐藏的信息比特是“1”时，考虑取最接近的奇数为量化值。在[54]中，对视频，作者建议以同样的方式修改运动参数的量化。

6.8　结　束　语

本章中，基于水印的几个关键问题，我们介绍了涉及范围很广的各种技术。由于应用目标

的多样性(版权、完整性、不可抵赖性)、无限的可能的攻击、要考虑的宿主信号不同的自然属性(音频、图像、视频)、可视性和健壮性之间权衡的复杂性以及水印提取的不同模式,所以水印面临的主要挑战之一就是对文献里所提出的各种方案进行评估和比较的可能性。下一章将讨论水印的这些重要而且又困难的方面。

参考文献

[1] Bender, W., D. Gruhl, and N. Morimoto, "Techniques for Data Hiding," in *Proceedings of the SPIE 2420, Storage and Retrieval for Image and Video Databases III*, 1995, pp. 164 – 173.

[2] Langelaar, G. C., J. C. A. Van der Lubbe, and R. L. Lagendijk, "Robust Labeling Methods for Copy *Protection of Images*," *in Proceedings of SPIE* 3022, *Storage and Retrieval for Image and Video Databases V*, 1997, pp. 298 – 309.

[3] Pitas, I., and T. H. Kaskalis, "Applying Signatures on Digital Images," in *IEEE Workshop on Nonlinear Signal and Image Processing*, Thessaloniki, Greece, Oct. 1995, pp. 460 – 463.

[4] Hartung, F., and B. Girod, "Fast Public-Key Watermarking of Compressed Video," in *International Conference on Image Processing*, Santa Barbara, California, Oct. 1997.

[5] Wong, P. W., "A Public Key Watermark for Image Verification and Authentication," in Proceedings of the International Conference on Image Processing, vol. 1, Chicago, Illinois, Oct. 1998.

[6] Memon, N., and P. W. Wong, "Buyer-seller Watermarking Protocol Based on Amplitude Modulation and the El Gamal Public-Key Cryptosystem," in *Proceedings of the SPIE* 3657, *Security and Watermarking of Multimedia Contents*, 1999, pp. 289 – 294.

[7] Matsui, K., and K. Tanaka, "Video-steganography: How to Secretly Embed a Signature in a Picture," *Journal of the Interactive Multimedia Association Intellectual Property Project*, vol. 1, no. 1, 1994, pp. 187 – 206.

[8] Kutter, M., F. Jordan, and F. Bossen, "Digital Signature of Color Images Using Amplitude Modulation," in *Proceedings of the SPIE* 3022, *Storage and Retrieval for Image and Video Databases V*, 1997, pp. 518 – 526.

[9] Pratt, W. K., *Digital Image Processing*, New York: Wiley, 1991.

[10] Wallace, G. K., "The JPEG Still Picture Compression Standard," *Communications of the ACM*, vol. 34, no. 4, 1991, pp. 40 – 44.

[11] Pennebaker, W. B., and J. L. Mitchell, JPEG Still Image Data Compression Standard, New York: Van Nostrand Reinhold Company, 1992.

[12] Rao, K. R., and P. Yip, Discrete Cosine Transform: *Algorithms*, *Advantages*, *Applications*, New York: Academic Press, 1990.

[13] Koch, E., and J. Zhao, "Towards Robust and Hidden Image Copyright Labeling," in *IEEE Workshop on Nonlinear Signal and Image Processing*, Thessaloniki, Greece, Oct. 1995, pp. 452 – 455.

[14] Zhao, J., "A WWW Service to Embed and Prove Digital Copyright Watermarkings," in *Proceedings of the European Conference on Multimedia Applications, Services and Techniques*, 1996, pp. 695 – 709.

[15] Johnson, N. F., and S. Jajodia, "Exploring Steganography: Seeing the Unseen," *IEEE Computer*,

vol.31, no.2, 1998, pp.26 – 34.

[16] Johnson, N.F., and S. Jajodia, "Steganalysis of Images Created Using Current Steganography Software," in *Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science*, Springer, 1998, pp.273 – 289.

[17] Ó Ruanaidh, J.J.K., and T. Pun, "Rotation, Translation and Scale Invariant Digital Image Watermarking," in *Proceedings of the International Conference on Image Processing*, vol.1, Santa Barbara, California, Oct. 1997, pp.536 – 539.

[18] Mallat, S., A *Wavelet Tour of Signal Processing*, London: Academic Press. 1998.

[19] Vetterli, M., and J. Kovačević, *Wavelets and Subband Coding*, *Englewood Cliffs*: Prentice Hall, 1995.

[20] Antonini, M., et al., "Image Coding Using Wavelet Transform," *IEEE Transactions on Image Processing*, vol.1, no.2, 1992, pp.205 – 220.

[21] Daubechies, I., *Ten Lectures on Wavelets*, SIAM Press, 1992.

[22] Wang. H. – J., and C. – C.J. Kuo, "Image Protection via Watermarking on Perceptually Significant Wavelet Coefficients," in *Proceedings of the IEEE Multimedia Signal Processing Workshop*, Redondo Beach, California, Dec. 1998. pp.278 – 284.

[23] Kundur, D., and D. Hatzinakos, "A Robust Digital Image Watermarking Method Using Wavelet-Based Fusion," in *Proceedings of the International Conference on Image Processing*, vol.1, Santa Barbara, California, Oct. 1997. pp.544 – 547.

[24] Wilson, T.A., S.K. Rogers, and L.R. Myers, "Perceptual-Based Hyperspectral Image Fusion Using Multiresolution Analysis," *Optical Engineering*, vol.34, 1995, pp.3154 – 3164.

[25] Xia, X. – G., C.G. Boncelet, and G.R. Arce, "Wavelet Transform Based Watermark for Digital Images," *Optics Express*, vol.3, no.12, 1998, pp.497 – 511.

[26] Kundur, D., and D. Hatzinakos, "Digital Watermarking Using Multiresolution Wavelet Decomposiiton," in *Proceedings of the IEEE International Conference on Acoustics*, Speech and Signal Processing, vol.6, 1998, pp.2969 – 2972.

[27] Swanson, M.D., B. Zhu, and A. Tewflk, "Multiresolution Scene-Based Video Watermarking Using Perceptual Models," *IEEE Journal on Selected Areas in Communications*, vol.16, . no.4, 1998, pp.540 – 550.

[28] Zeng, W., B. Liu, and S. Lei, "Extraction of Multiresolution Watermark Image for Claiming Rightful Ownership," in *Proceeding of the SPIE 3657, Security and Watermarking of Multimedia Contents*, 1999, pp.404 – 414.

[29] Delaigle, J. – F., C. De Vleeschouwer, and B. Macq. "Watermarking Using a Matching Model Based on the Human Visual System," *École thématique CNRS GDR-PRC ISIS: Information Signal Images*, Marly le Roi, 1997.

[30] Bartolini, F., et al., "Mask Building for Perceptually Hiding Frequency Embedded Watermarks," in *Proceedings of the International Conference on Image Processing*, vol.1, Chicago, Illinois, USA, Oct. 1998, pp.450 – 454.

[31] Moreau, N., Techniques de compression des signaux, Collection technique et scientifique des

télécommunications, Masson, 1995.

[32] Westen, S., R. Lagendijk, and J. Biemond, "Perceptual Image Quality Based on a Multiple Channel HVS Model," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol.4, 1995, pp.2351 – 2354.

[33] Winkler, S., "A Perceptual Distortion Metric for Digital Color Images," in *Proceedings of the International Conference on Image Processing*, vol.1, Chicago, Illinois, USA, Oct.1998, pp.399 – 403.

[34] Goffin, F., et al., "A Low Cost Perceptive Digital Picture Watermarking Method," in *Proceedings of the SPIE* 3022, *Storage and Retrieval for Image and Video Database V*, 1997, pp.264 – 277.

[35] Cox, I., et al., "Secure Spread Spectrum Watermarking for Multimedia," Technical report, NEC Research Institute, 1995.

[36] Hartung, F., and B. Girod, "Digital Watermarking of MPEG – 2 Coded Video in the Bitstream Domain," in *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.4, Munich, Germany, Apr.1997, pp.2621 – 2624.

[37] Pickholtz, R.L., D.L. Schilling, and L.B. Millstein, ."Theory of Spread Spectrum Communications-A Tutorial," *IEEE Transactions on Communications*, vol.30, no.5, 1982, pp.855 – 884.

[38] Hartung, F., and B. Girod," Digital Watermarking of Raw and Compressed Video," in *Proceedings of the European EOS/SPIE Symposium on Advanced Imaging and Network Technologies*, Berlin, Germany, Oct.1996.

[39] Braudaway, G.W., "Protecting Publicly-Available Images with an Invisible Image Watermark," in *Proceedings of the International Conference on Image Processing*, Santa Barbara, California, Oct.1997.

[40] Delaigle, J. – F., et al., "Digital Images Protection Techniques in a Broadcast Framework: Overview," in Proceedings of the European Conference on Multimedia Applications, Services and Techniques, Louvain-la-Neuve, Belgium, May 1996, pp.711 – 728.

[41] Darmstaedter, V., et al., "A Block Based Watermarking Technique for MPEG – 2 Signals: Optimization and Validation on Real Digital TV Distribution Links," in *Proceedings of the European Conference on Multimedia Applications, Services and Techniques*, May 1998.

[42] Hernández, J.R., et al, "The Impact of Channel Coding on the Performance of Spatial Watermarking for Copyright Protection," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol.5, 1998, pp.2973 – 2976.

[43] Ó Ruanaidh, J.J.K., W.J. Dowling, and F.M. Boland, "Phase Watermarking of Digital Images," in *Proceedings of the IEEE International Conference on Image Processing*, vol.3, Sep.1996, pp.239 – 242.

[44] Hayes, M.H., "The Reconstruction of a Multidimensional Sequence," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Apr.1992, pp.140 – 154.

[45] Bruyndonckx, O., J. – J. Quisquater, and B. Macq, "Spatial Method for Copyright Labeling of Digital Images," in *Nonlinear Signal Processing Workshop*, Thessaloniki, Greece 1995, pp.456 – 459.

[46] Puate, J., and F. Jordan, "Using Fractal Compression Scheme to Embed a Digital Signature into an

Image,"in *Proceedings of the SPIE 2915*, *Video Techniques and Software for Full-Service Networks*, 1996, pp.108 – 118.

[47] Jacquin, A.E., "Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations," *IEEE Transactions on Image Processing*, vol. 1, no.1, 1992, pp.18 – 30.

[48] Fisher, Y.(ed.), Fractal Image Compression: *Theory and Application*, New York: Springer-Verlag, 1995.

[49] Depovere, G., T. Kalker, and J. – P.M.G. Linnartz," Improved Watermark Detection Reliability Using Filtering Before Correlation,"in *Proceedings of the International Conference on Image Processing*, vol.1, IEEE Signal Processing Society, Chicago, Illinois, USA, Oct.1998.

[50] Kutter, M., "Watermarking Resisting to Translation, Rotation and Scaling," in *Proceeding of the SPIE 3528*, *Multimedia Systems and Applications*, 1998, pp.423 – 431.

[51] Barni, M., et al., "Robust Watermarking of Still Images for Copyright Protection," in *Proceedings of the 13th International Conference on Digital Signal Processing*, vol.2, Santorini, Greece, Jul.1997, pp.499 – 502.

[52] Swanson, M.D., B.Zhu, and A.H.Tewfik, "Data Hiding for Video-in-Video," in *Proceedings of the International Conference on Image Processing*, vol.1, Santa Barbara, CA, 1997, pp.676 – 679.

[53] Riley, M.J., and I.E.G.Richardson, *Digital Video Communications*, *Artech House*, 1997.

[54] Jordan, F., and T.Vynne, "Motion Vector Watermarking," Laboratoire de Traitement des Signaux École Polytechnique Fédérate de Lausanne, Patent, 1997.

第七章　版权标记系统的健壮性

(Scott Craver, Adrian Perrig, Fabien A.P.Petitcolas)

7.1　健壮性需求

水印抵抗擦除的健壮性是通过对计算时间和图像质量的权衡和综合获得的。为了擦除水印标记,攻击者不得不对带标记内容造成过度的破坏,或者花大量的计算时间来逆转水印嵌入过程。

在这一研究领域,IFPI(国际留声机工业联盟)最近提出声音水印技术的要求时,试图给健壮性下一个定义[1]。它的目的是建立一个水印方案,使它可以产生抵抗盗版活动的证据、跟踪广播机构和其他人员对录音带和唱片的使用情况以及控制拷贝。IFPI 定义的健壮性要求如下:

- 标记机制不能影响唱片的声音品质;
- 标记信息经过以下变换之后必须还能恢复:各种滤波和信号处理操作,包括两个连续的 D/A 和 A/D 转换、稳态压缩或 10%的时间扩展、图像 MPEG 的数据压缩和多频带的非线性振幅压缩、增加加性或乘性噪声、用同一系统加入另外一个标记信号、用于低音和中高音控制系统的高至 15dB 的频率响应失真、群延时失真和陷波滤波等;
- 用任何方法都不能删除或改变嵌入的信息,除非声音质量差到不能用的地步;
- 如果提供 20dB 或更高的信噪比,嵌入标记的数据信道应能在纠错之后还有 20bps 的带宽,且与信号的水平和类型(古典音乐、流行乐、话音)无关。

对静止图片、视频和一般的多媒体对象加入水印标记的要求都是类似的,很少有人尝试去量化它们[2]。

攻击者企图去除(或减弱)内容拥有者的标记的效力,这些标记是为了保护拥有者的所有权和控制所保护的内容而嵌入的。正如计算机安全系统一样,一个系统的安全是与它最薄弱的一环的安全性相同的。如果攻击者破坏了水印生命周期的任何一个阶段,这个攻击就被认为是成功的。所以作品的拥有者和水印软件必须确保任何一个阶段都是非常安全,能抵抗攻击的。

水印的研究者们认识和了解了对水印方案的各种级别的攻击,每一级别攻击都利用了水印过程的不同阶段。为了对已存在攻击的多样性有一个概念,我们可以说,只有第一级别的攻击想删除或减弱水印的存在,其他级别的攻击是利用更狡猾的技巧破坏水印的可用性,而不必破坏水印本身。所以,简单直接的健壮是必要的,但不是完全可以保证安全的。

我们把水印攻击分为四类,如 Craver 等[3]列出的,它们是健壮性攻击、表达攻击、解释攻击和合法攻击。健壮性攻击涉及信号的削弱,是四类攻击中最直观明显的。这些攻击遍布整个范围,从良性的操作如压缩,到用常规的图像处理程序对水印进行表面的攻击,或者用专门设计的程序进行攻击。表达攻击是使水印检测失败,它不是将已加水印的对象进行提取分离,

而是将水印变形到检测器检不出来。解释攻击通常是通过伪造水印而成功的，造成原来水印不能被判断和不再说明任何意义。最后，合法攻击得益于法律问题的漏洞（这些问题中的部分内容在第九章有详细介绍），这大大超出了工程研究的范围。

对这些攻击的彻底研究提出了很多设计问题，我们觉得这些问题对于预防这些攻击起着关键的作用。消费者标准对健壮性有很重要的影响，某种程度上甚至忽视了它们会造成很大的风险："无故损害"介质一定出现对消费者的无故损害，而不仅仅是对职业摄影师和电影制片人造成损害。另一方面，"不可视"水印必须不仅仅对消费者不可视，而且对于拥有复杂的统计武器库的专家也一样。水印工具和水印检测器结构上的缺陷、工程上太容易实现的设计，都会导致水印方案无效；有时候一些在实现中的缺陷、错误，暗示着潜藏在设计中的严重问题。

一旦一个健壮的水印技术变得很实用并投入使用，强大的攻击将会损害曾经信任它的每个人的所有权，而这种损害是用多少技术也无法补救的，因为简单嵌入水印的媒体已经在因特网上以光速被公布和分发了。在使用这些技术之前，我们无法太强调已了解和考虑了这些问题，每一个安全威胁和攻击都将逐渐说明其重要性。

7.2 信号削弱

通过对内容降质而移去水印是所有攻击中最笨和最直接的方法。当水印方案的设计者试图达到工程上的健壮性要求时，他们必须考虑到各类攻击方法。设计出来的方案必须能抵抗常规的攻击，像压缩、裁剪、模糊化、甚至打印和重扫描，或者还希望方案对于预料不到的操作也具有健壮性。但不幸的是，攻击者拥有图像处理操作的庞大武器库来进行攻击，对任何方案的简要分析经常能找到删除水印的一些简单操作方法。

7.2.1 噪声和水印覆盖

加入轻微敏感噪声的一个有趣的方法就是加入另外一个水印。这个操作可以引起微弱的质量下降，攻击者可以方便地用软件实现这个操作以迷惑拥有者。

在检测水印本身的质量降低程度时，我们强调秘密水印和公开水印方案的重要区别，正如5.3节中所定义的那样。公开水印与秘密水印相比，它容易通过被另一个水印覆盖的方法而删除。在信息伪装系统中，秘密水印的秘密信息可以实现可选择信道，在大量的可能位置隐藏水印。这就允许两个水印，每一个按不同的密钥嵌入，他们在不同的位置而不会互相破坏。而公开水印则必须放在每一个检测器都知道的地方，两个水印就会太多占用同一个"空间"。

一些公开水印软件，像 Digimarc 公司的水印软件 PictureMarc 就认识到了这个问题，它拒绝让一个用户在水印已经存在的情况下再加水印。有很多方法可以避免这种情况，从破坏水印到破坏实际的软件产品[4]。你甚至可以重新实现水印算法，还可能将源代码公布给别人用。然而付出大量努力去做这件事，也许对阻止图像盗版事件的发生具有足够的威慑力，尤其是专利法被用于防止人们盗用源代码之后。就连最普通的黑客工具在这种威胁下也不再普遍了，避免了偶然的盗版。一个很有趣的问题，就是找一个简单的、无需努力的方法来战胜这些措施，如果这个对策能被一些应用于标准图像处理程序中的普通操作所击败，攻击就成为可行的甚至是随机的图像偷盗。

这儿有一个用 Adobe Photoshop 分层功能的例子。我们拿一幅在 Photoshop 中用 Digimarc 的

插件程序加入水印的图像，然后发现我们不能加入我们自己的水印了。对它进行连续模糊化直到旧水印提取不出来，但是这需要极端的模糊化以至于不能被认为是成功的攻击。然而我们可以在模糊化后的图像上以最大强度加入我们自己的水印，只是简单的覆盖它，以30%的透明度加在原始图像的上面(30%这个值将随图像和下面的水印强度的变化面变化)。这一般可以在 Adobe Photoshop 中实现，背景图像放入一个层，模糊化和加水印后的图像是另外一层，复合在一起，设置其透明度，并将其平整化。结果看起来好像只是微小的模糊化，这种效果可以通过锐化滤波来修复。图像中现在包括我们加入的非常高强度的新水印，而旧水印则不再能检测得到。

7.2.2 压缩

JPEG 压缩是当前静止图像最为广泛应用的压缩算法。当我们准备将图像发布到网上时，图像要被调整大小、压缩以适应版面设计和带宽的需求。不幸的是，有损压缩会删除一些对可视性影响微小的高频分量，而只保持了低频分量。这就影响了某些数字图像水印方案，这些方案的原理就是将信息嵌入到高频部分以减少失真。

因此，有人建议，水印应该放在图像对感观影响重要的成分中，而不管它所可能引入的失真[5]，但这样可能会留下可视的人为痕迹。

7.2.3 用户质量标准

健壮性是数字水印方案中的主要目的，所以很多东西依赖于这个术语的含义。一个方案在下列情况下通常被认为是健壮的，即如果为成功删除水印而造成对“图像质量”的损害是过度的，或者攻击者为达到目的而付出过多的时间。但是对所有成功水印方案的重要特征—健壮性—的这种定义是建立在一个模糊和不确定的术语的基础之上的。比如我们如何定义质量？什么程度是质量的过度损失？

水印技术的脆弱性之一就是用户可以接受的质量降低的程度，因为这是定义受损内容是否真正有用的质量标准，以及能保证水印必须存在的破坏程度。网页设计师喜欢把图像进行 JPEG 压缩以节省空间和下载时间，用户喜欢看到他们喜欢的电影在录像带上缩减到 250 线，或许能修剪屏幕边缘以适应他们的电视机。这明显不能使内容的受损超出它的可用性和金融价值。

以媒体创造者们的标准来定义质量和健壮性是很容易的。这些人将是真正使用水印软件的人，所以他们对质量的定义已经对加入水印造成的破坏做了限制。水印软件的制造者们必须仔细考虑，把采样率、压缩或颜色量化降低到一定程度就会过度，因为简单地把内容放到印刷品、磁带或电视上是不能被接受的。

还有一点需要考虑，就是由加入水印而引起的质量降低的程度。一些水印软件产品产生了能引起人注意的噪声。如果一个内容制造者乐意接受很大程度的噪声，那么可以公平地说，在内容所涉及的应用领域，更大程度的噪声也可能被用户所接受的。所以，一个水印嵌入程序不能只靠增加嵌入水印的强度来提高健壮性。如果这么做，那它只能用于嵌入引起过度噪声能被接受的领域。在这些领域，水印需要抵抗实质的损害。

7.2.4 平均化

当攻击者可以得到大量的图像时，他可以把它们平均一下产生一个检测不到水印的图像。

这可以很容易地应用于视频应用，同一个水印 W 被加到像 Cox 和 Linnartz 所描述的视频序列的一组图像 $\{I_i\}_{i=1,n}$ 上[6]。如果 f 是特征提取函数(描述水印应加入到哪里)。在这个特征域中，n 个加了水印的帧加起来就是 $nW + \sum_{i=1}^{n} f(I_i)$。当 n 值很大时，它的期望值是 nW。攻击者可以用这个想法对水印进行粗略的估计，把它从各帧中删除。很显然的对策是嵌入多个水印，并让他们在图像中相互独立。读者可以参考 Cox 等[6]的资料了解更详细的解决方法。

平均化可以用于数字指纹，第 8 章将作详细介绍。现在把数字指纹理解为一个数字序列就足够了。介绍中提到的一般方案是这样的，图像的发行商为每一位顾客嵌入一个唯一标识号以跟踪侵权者。在这个事例中，图像是相同的而水印是互不相同的，所以加入水印的图像合计为 $nf(I) + \sum_{i=1}^{n} W_i$，当 n 很大时，它的期望值为 $nf(I)$。各种可能的对策将会在第 8 章中提出。

7.2.5　专门设计的攻击

如果了解水印算法的细节，攻击者可以发明一种攻击，专门设计为删除某种水印。举例来说，通过在傅立叶域改变事先规定好的频率上的值来嵌入水印的方案，对用改变图像像素的方法来删除掉是困难的，但知道这个方案的攻击者可以将图像转变到傅立叶域然后修改相同频率上的值来删除水印。

水印系统中针对扩频的非线性滤波是另一个专门攻击的例子。由于很多水印方案基于扩频技术，Langelaar、Lagendijk 和 Biemond[7]所描述的这种攻击有广泛的意义。通常的想法是把一个已加水印的图像 $\tilde{I}$ 分成两部分，$\tilde{I}$ 和 $\hat{W}$，这样一来原始图像的估计值 $\hat{I}$ 就不再包含水印部分了。Langelaar 等通过实验发现，3×3 中频滤波应用于 Bender et al.[8]、Pitas 和 Kaskalis[9]提出的方法时能提供很好的分离结果，然后把它用于水印的粗略估计，得到 $\tilde{I} - med_{3\times3}(\tilde{I})$。在从 $\tilde{I}$ 中减掉之前，这个估计值还应该提炼一下，因为它仍然包含边缘信息，一些值可能太大。后来 Langelaar 等建议把它通过一个高通滤波器滤波，将输出值调整在 $[-2,2]$ 范围之内。以下给出估计的原始值

$$\begin{cases} \hat{I} = \tilde{I} - \hat{W} \\ \hat{W} = aH_{3\times3}(\tilde{I} - med_{3\times3}(\tilde{I}))\ |_{[-2,2]} \end{cases} \tag{7.1}$$

其中 a 是实验中确定的放大系数。

还有很多专门的攻击。7.2.1 节中描述了如何删除特殊水印，7.5.2 节中描述了一种对固定深度图像水印的攻击。随着水印技术越来越广泛的应用，我们期待更多人公开发表如何破解各种水印的方案，就像分析密码脆弱性的密码分析学一样。

在这里，我们还希望提醒 Kerckhoffs 准则的读者们(见 1.2.4 节)。有很多理由阻止这个理论：现在对二进制代码进行反编译的工具是很复杂的，很多人可以从公司得到源代码，而开发实现一个新的水印算法所需的时间间隔是非常长的。如果我们假想将来水印技术成为普遍的安全技术，那么像一个“借用”备份磁带或工程笔记的清洁工在系统中安装特洛伊木马程序，这种事将成为可能。

正是由于这个原因，水印技术专家仔细研究后公布的算法从长远来看更值得信赖。随着所给水印算法而产生的各种攻击被认为是不可避免的，而且我们不能判断一个方案是否能承受各种攻击，除非它是公开的。

7.3 水印检测的失败

使水印无效，不一定必须删除它，可以通过对内容做处理，使检测器找不到有效的水印。这类攻击叫做表达攻击，最好的例子就是几何变形攻击和马赛克攻击。

7.3.1 变形攻击

虽然很多水印系统能抵抗基本的处理，即那些用标准工具（例如[10]）就能轻松实现的处理，但他们不能应付这些处理的合成和微小随机的几何变形。

StirMark 是一种水印系统的测试工具，它把这样的几何变形应用于图像[11]。如果 A、B、C、D 是图像的四个角，一点 M 可以这样表示：$M = \alpha(\beta A + (1-\beta)D) + (1-\alpha)(\beta B + (1-\beta)C)$，这里 $0 \leqslant \alpha, \beta \leqslant 1$，是 M 相对于四个角的坐标（见图 7.1）。StirMark1 中的随机双线性失真变形是通过用一个很小的随机数从两个方向上改变角度来实现的。新的 M 值可以由上述公式确定，保持(α, β)不变，但用四个新的角计算。这种变换是可逆的。最终结果是，这个软件不能从根本上删除水印，但是它可以阻止一些系统检测到或恢复出他们的水印。

一些还没被注意到的很多变形也可以用到图片中（见图 7.2）。当应用于图像中时，StirMark 引入的失真几乎注意不到。图 7.2 中，“手表”图像经 StirMark 在缺省参数下处理之前的图像(a)和处理之后的图像(b)。为了比较，对一个网格图像进行相同的处理(c)和(d)。图像取自“金链上的怀表”，是 Kevin Odhner 的版权图像（jko@home.com）。除了以前解释过的一般双线性特性，还可以对图片中间的每个像素进行微小偏离，而边缘地区几乎不做处理。实际实现中，这种突起形状就是简单的正弦波。如果(x, y)表示一个像素的坐标，$0 \leqslant x \leqslant X$ 和 $0 \leqslant y \leqslant$，那么变化后的像素坐标表示为(x', y')而 $x' = x + \lambda \sin(\pi y / Y)$，$y' = y + \lambda \sin(\pi x / X)$。在这之上做一个高频位移，即：$\delta = \lambda \sin(\omega_x x)\sin(\omega_y y)(1 + n(x, y))$，$n$ 是随机数，合成为：$x'' = x' + \delta_1$，$y'' = y' + \delta_2$。（注：按译者的理解，这个式子应该是：$x'' = x' + \delta$，$y'' = y' + \delta$）

所有这些变形与适度的 JPEG 压缩结合在一起，也不能从本质上删除水印，但它们能阻止检测器找到水印，能使检测器失去同步。这就说明真正的问题不只是加入水印，还必须能够把它辨别出来。如果保存随机参数，这些微小随机的几何变形还可以用于视频；否则视频播放时会出现抖动。这只是视频的一个方面，更好的攻击应该考虑视频的时间因素。StirMark 还可以实现一系列测试，它们已成为图像水印测试标准[10,12]。

有人也许会试图通过预测盗版者所可能使用的变换来增加水印系统的健壮性，而如果这些变换是难以预测的，这种做法就会很难实现。有人也许会用像嵌入多版本适合可逆变换的水印这类技术。例如，Ó Ruanaidh 和 Pereira 建议使用 Mellin-Fourier 变换（见 6.3.3 节）来解决旋转和缩放问题[13]。你还可以利用几何变形在局部几乎都是线性的这个事实（等价于平移和旋转），用基于块的检测算法。当原始图像可以得到时，它足可以近似地对随机变形进行推断；Davoine 等[14]最近提出，这些可以用三角网格来实现。

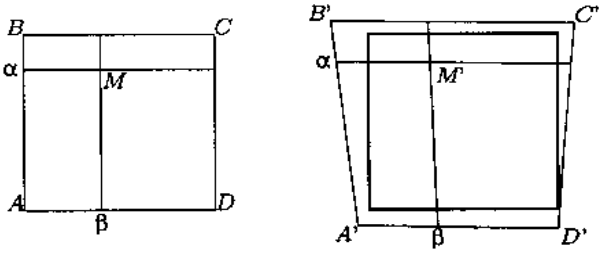

图 7.1 StirMark1 中的随机双线性失真

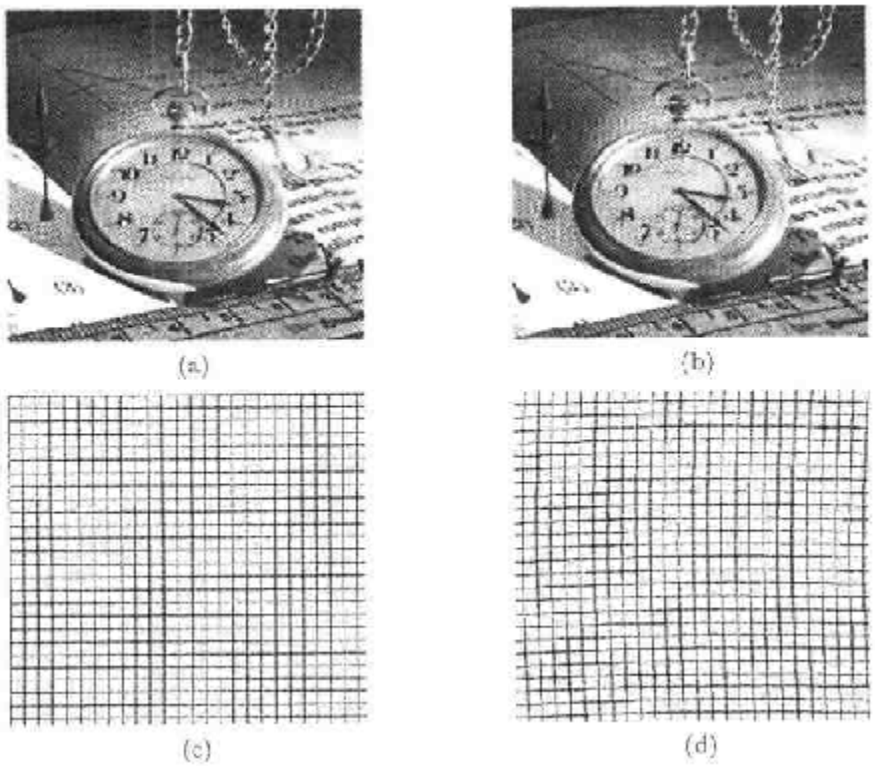

图 7.2　随机几何失真的例子

7.3.2　比特率限制

在 5.6.1 节中我们说明了水印系统比特率和系统对各种攻击的健壮性之间的权衡。很明显，图像越大，越容易隐藏一定数量的比特信息。反过来说也是对的，这就是“马赛克攻击”的基础，即一个图像可以小到不能放下水印。

这一点在表达攻击中非常强调，它有很普遍的适用性。它具有最基本的显著特性，一个已加水印的图像可以去掉水印，还可以把它一个像素一个像素地恢复出来，就像标准浏览器处理水印图像那样。

这类攻击是由版权盗版检测系统所激发的，包括水印方案加上 Web 蜘蛛（或 Web 爬虫），它从网上下载图片并检查它是否含有水印。它还包括把图像分成许多小子图，嵌入网页中合适的地方。一般的网络浏览器把子图重新组合在一起，于是他们看起来与原图一致（见图 7.3）。图 7.3 是经过马赛克攻击后下载图像时浏览器的屏幕显示。这种攻击把水印图像剪切成小块，浏览器修补页面时再粘贴在一起。可以得到这样的软件，读取 JPEG 图片，产生小 JPEG 图像的相应马赛克和自动产生必要的 HTML 代码。在这种情况下，下载马赛克要比下载整幅图像还要快！在这个例子里，我们用 Picture Marc 1.5 版软件对一幅 350 × 280 像素图像加水印。图片为 King's College Chapel，该图得到 J. Thompson 的许可。这种攻击是很普通的，因为所有水印方案都要求加水印的图像有一个最小尺寸（我们不能在一个像素上隐藏一个有意义的水印）。所以通过把图像分成足够小的块，水印检测器将会被迷惑[15]。当然我们希望加水印的图像的最小尺寸能非常小，那么这种攻击方法就不会太有效。

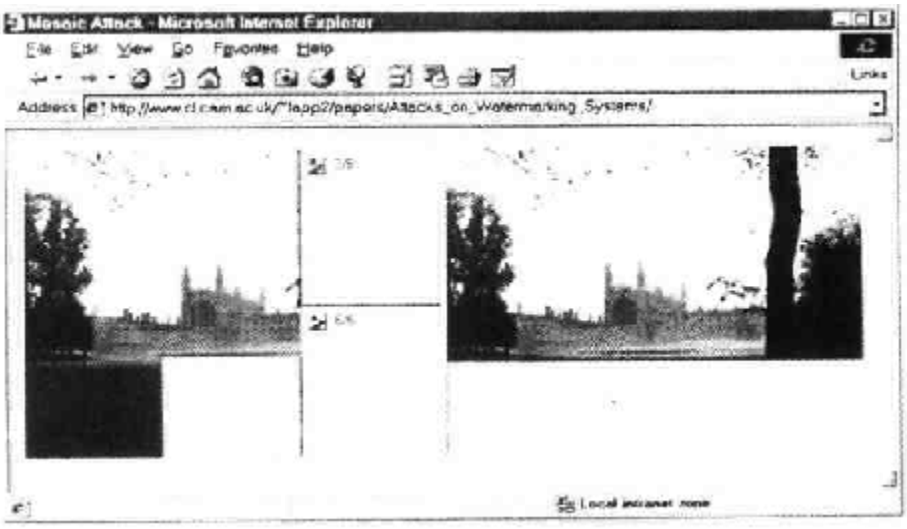

图 7.3 马赛克攻击的例子

7.3.3 意外碰撞和错误警报

当一个水印检测器检测到一个没加入水印的图像中有水印，就会发生错误警报。如果图像还包含了另外一个水印，我们称此为水印碰撞。碰撞经常发生在双方用同一个水印方案的时候。

错误警报和水印碰撞都会在解决一个图像的所有权时产生问题。既然水印检测显示了图像所有权，一个错误警报就可能对无辜图像发布者进行侵权控告。另外，错误警报也可能对合法水印含义产生质疑。在水印碰撞事件中，图像显然包含两个水印，我们无法要求一个不可靠的水印方案提供更有效的版权指示信息。对水印系统错误警报研究感兴趣的读者可以参考 Linnartz 的论文[16]。

这些关注也许看起来很偏执，但是错误警报已经在今天商业应用的水印方案中发现了。用 Holliman[17]开发的水印方案加入水印的图像被发现与一个已经存在的熟知的水印产品相冲突，即使它们是用不同规则嵌入的。当你考虑用自动蜘蛛爬行并寻找加入水印的图像时，这个事实也很令人担忧。

一个有趣的问题是如果有一个任意的图像 I 和一个水印检测函数 D，我们能不能找到一个密钥 k，使得图像看起来是以 k 嵌入水印，即，使得 $D(I, K) = \text{yes}$ 成立？任何商业上用的水印技术都应该使找到这样的 k 值在计算上是不可行的。它还必须使函数 D 在一个图像 I 的个别像素上的敏感性分析在计算上是不可行的(见 7.4.2 节)。否则，我们可以选择一个随机密钥 k，改变像素值直到图像包含所需要的水印。

我们看到设计一个水印方案需要权衡很多方面，其中之一就是纠错和意外碰撞的权衡。一方面任何算法都需要强的纠错能力来得到很高的健壮性，另一方面，随着纠错能力的增强，水印碰撞的可能性也在增加。有些方案计算水印与水印数据库的相关函数[5,18]。如果我们有大量的用户，这种方法对于错误警报和水印碰撞是脆弱的，因为错误警报或者碰撞的可能性随着用户和图像的数量增大而增大。

7.4 伪造水印

迄今为止，我们介绍了两类水印攻击。信号削弱攻击试图破坏水印但不伤害内容，而表达攻击只是调整内容使得自动检测器不能定位水印。

一个成功的信号削弱攻击意味着，水印方案对某个"合理"操作不健壮。而一个成功的表达攻击不利用水印方案的任何脆弱性。一些表达攻击利用所有水印方案不可避免的脆弱性进行极端破坏，像修改或逆转像素值；但是所有表达攻击所利用的真正的弱点是 Web 蜘蛛和自动检测器所固有的不可救药的无智能性(相对而言，人在网上冲浪就很智能，鼠标点击 java 程序窗口是很平常的事)。

从表达攻击中我们所学到的重要的教训是，在计算机安全中，如果他破坏了水印过程的任何一步，包括从嵌入、发布到检测，这个攻击就是成功的。如果每一个阶段都安全，防守才算成功。最好的例子就是自动水印检测了，世上所有的水印健壮性都被网页上将图像分割到比特这个简单的动作给毁了。

认识到这一点，我们就可以问一问水印过程的其他阶段是否易于被攻击。答案是非常肯定的，我们将用协议攻击的概念来解释。

7.4.1 协议攻击

像前面所描述的那样，很多秘密水印能够抵抗水印的二次嵌入，因为水印嵌入的方式是保密的，或是水印嵌入的位置是保密的，水印所占的"空间"足够大，以至于攻击者为了对每个可能隐藏水印的地方都堆放垃圾，一定会对内容造成过量破坏。也许会有这样的疑问，如果可以加入第二个水印，那么他就可以宣称具有对内容的所有权。这一点之所以不能实现是因为，原始内容的创建者有真正的原始内容，攻击者不知道。它不包含水印，而攻击者所谓的原始内容包含第一次加入的水印，而不包含第二次加入的水印，自然就建立了嵌入次序。

一般而言，多重水印的怀疑可以通过每一方在另一方的正版原图中提取自己的水印的方法来解决。如果每一个原图都是另一个原图的水印版本，那就太奇怪了。如果水印强度非常接近的话，就会阻止任何一方确立所有权、删除水印效果。结果这种情况就可能被聪明的攻击者用到水印方法上。

理论上，原始内容的创建者在图像 I 上加水印 w，产生已加水印图像 $\tilde{I}=I+w$。图像 $\tilde{I}$ 分发到顾客手中，当找到可疑图像 I' 时，计算 $I'-I$，如果 I' 和 $\tilde{I}$ 相同，那么 $I'-I$ 就等于 w，如果 I' 是从 $\tilde{I}$ 衍生出的，$I'-I$ 就接近于 w，这个方案是健壮的。相关函数 $c(w,x)$ 可以用来确定原始水印 w 和所提取的数据 x 之间的相似性。这样，大体上可以从两幅图像的差别中找出水印。很多水印方案基于这个模型，还有一点补充就是，这些方案不需要原始图像[19]。

图 7.4 中，内容的创建者 Alice 把水印 w 加到图像 I 中得到 $\tilde{I}=I+w$，然后发布到网上。Mallory 想把图像偷来自己处理，他是减掉而不是加入第二个水印 x，得到图像 $I'=\tilde{I}-x=I+w-x$。现在，Mallory 声称 I'(而不是 $\tilde{I}$)是他的原始图像，反而把 Alice 拉上法庭告她侵权。比较一下原始图像，Alice 将会发现她的水印 w 现在在 Mallory 的图像 I' 中：

$$I'-I=w-x,\ c(w-x,w)=1 \tag{7.2}$$

既然两个水印可以在同一个图像中存在，那减掉 x 应该对 w 的伤害不是很大。所以，Mallory

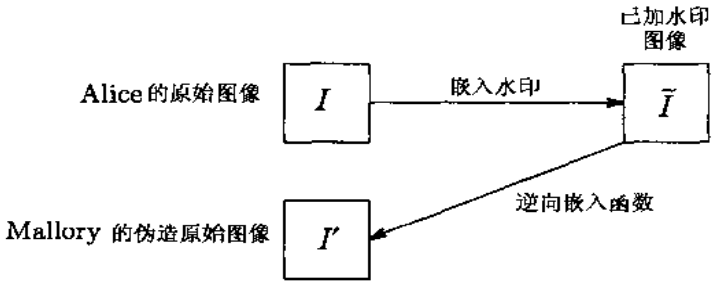

图 7.4 伪造原始图像

没有成功的删除 Alice 的水印。然而，Mallory 也可以表示如下：

$$I - I' = x - w,\ c(x - w, x) = 1 \tag{7.3}$$

换句话说，Mallory 的水印在 Alice 的原始图像中，即使 Alice 没有把它公开。Craver 等[19]收集的 Cox[20]所描述的水印方案的实验数据表明这种攻击是有效的，两个水印的相对强度实质上是相等的，没有真正的证据说明任何一方是图像的原创。

这种攻击是通过减掉一个水印而不是加上一个水印而工作的，所以依赖于水印方案的可逆性。一个好的方法就是把水印嵌入方案做成原始图像的单向函数 h，在这些不可逆方案中，Mallory 就不太可能减去水印 x，因为只有知道 I' 才能算出 $x = h(I')$，而知道 x 才能从 $\tilde{I}$ 中算出 I'。只要 h 难以逆变换，I' 就难以算出。

例如，[19]阐述了[20]中方案的修改版，来自原始图像的比特流用来控制水印加入方法。一个水印是一个实数序列 $w = \{w_0, w_1, \cdots\cdots w_n\}$，原始图像的哈希值用来产生一串比特 $b_0, b_1, \cdots\cdots b_n$。每一个元素 w_i 可以用两种不同的方法嵌入，b_i 的值表明是哪种方法嵌入的。如果有人试图用不同于嵌入时用的比特串来检测水印，检测就会失败。在这个方案中，Mallory 不能简单地从一个图像中减掉一个水印。因为被减掉的那种水印是原始图像的单向函数，Mallory 的水印必须是水印被删除的伪造原始图像的单向函数。换句话说，Mallory 必须猜出一个比特串 $\{b_i\}$ 和一个水印 x，使得当用控制串 $\{b_i\}$ 把 x 从图像中删除时，得到的图像经哈希后得出 $\{b_i\}$。用一个单向哈希函数，这种推测应该是不可能的。因此，这是不可逆水印方案。

然而，不可逆性仅仅是避免这种攻击的一个必要条件。就算 $\tilde{I}'$ 不能计算出来，如果单向哈希函数不正确运用，它也可以被估计出来。我们把前面提到的水印方案作为例子。我们知道没有办法计算出图像 I'，Mallory 就永远也不能执行协议攻击。他只能用一个全零控制串（$b = \{0,0,\cdots\cdots 0\}$）减掉 x 生成图像 I'_0，然后用全 1 控制串再减一次生成图像 I'_1，然后做平均得到最后的图像 $\tilde{I}'$。事实证明，使用任何一个控制串 $b = h(I')$，水印检测器都能找到 x。

这种情况在图 7.5 中做了解释，其中，Mallory 计算图像 I'，使得它与 $\tilde{I} - x$ 尽可能的接近。$\tilde{I}' = I' + x$ 现在与 $\tilde{I}$ 不同。Mallory 能计算出任何一个图像 I'（合理质量的）和水印 x，使得 x 在 $\tilde{I}$ 中存在，即使对图像 I' 加水印 x 得到的图像与 $\tilde{I}$ 不同，伪造攻击也能成功。

7.4.2 Oracle 攻击

在很多应用领域，攻击者可以访问到水印检测器。这检测器可能是配有主要图像处理包的软件，或者是像 DVD 那样的嵌入电器中的电子电路。即使攻击者不知道水印嵌入方法，他仍然可以使用检测器返回的信息来删除水印，对图像做微小变动直到检测器什么都找不出来

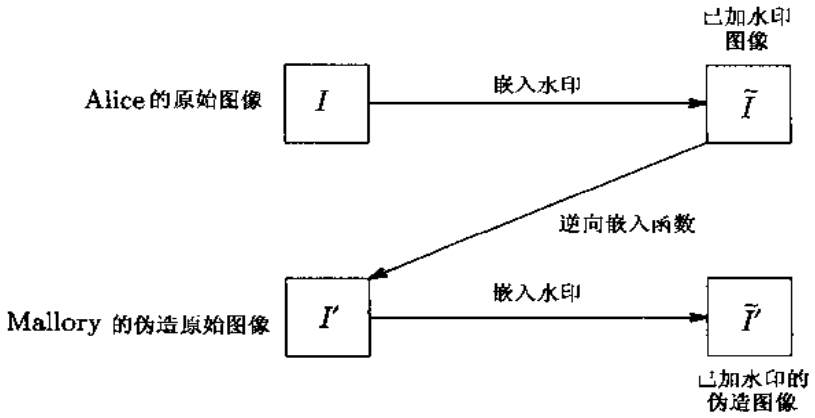

图 7.5　伪造一个水印：四图像事件

[21]。

Linnartz[22]对这种攻击有详细分析。攻击者开始时构造一个图像，使得检测结果与检测器的门限值非常接近，对这个图像进行轻微修改使检测器以接近 0.5 的概率从“水印存在”转换到“水印不存在”状态。构造的图像不需要与原图很相似，这可以通过对图像轻微模糊化（或者替换灰度的像素值）直到检测器找不到水印来实现。

第二步，分析检测器对每一个像素改变的敏感性。增加或减少所给像素的亮度，直到检测器的输出发生改变。对每一个像素重复这个过程。从这些分析中攻击者可以设计像素的组合和修改使得图像失真最小，并且对检测器检测结果改动最大，也就是说水印检测不出来。

一个可能的对策是对检测过程随机化[22]。假设检验不是使用一个门限，而是使用两个门限。在两个门限值之间，检测器得到一个随机答案，在第一个门限值之下表示“存在”，第二个门限值之上表示“不存在”这种随机化决策规则将大大阻止依赖微小改变水印图像的攻击。

也可以把解码过程复杂化。在没有防篡改硬件的情况下，哪一种方法都不是完全满意的。除非有突破性进展。需要公共验证的水印应用（像 DVD）好像注定要在防篡改技术的约束下进行了，或者用一个中心“水印阅读”设备。

最后可能的对策就是依赖于数字水印的意义了。如果水印存在意味着可以通知记录设备有人试图非法拷贝，那么上述设备就有理由不发挥作用或者对拷贝企图发一个悄悄的警告。如果有五次非法拷贝企图之后，攻击者就必须带着 DVD 机去商店重新启动，那么对于管理部门 oracle 攻击就是很受重视和困难的攻击。

7.4.3　特定的 oracle 攻击

这种 oracle 攻击只能对付公开水印，但即使是纯秘密水印，特定的 oracle 攻击也可能成功。这种攻击唯一的需求就是攻击者要有水印嵌入和检测的算法。

攻击者用和版权拥有者相同的方法在图像中一次或多次嵌入自己的水印。攻击者用自己的水印作为秘密水印的随机强度指示器。然后用普通的 oracle 攻击，直到所有新嵌入的水印都被删除掉。因为原始水印随着图像随机改变而变弱，当加入其它水印时，它也会变弱。所以我们可以假设，新加水印的强度为原始水印强度提供一个上限。所以当所有新水印都被删除掉时，原始秘密水印就会以很高的概率被删除掉。

7.5 水印检测

前面提到的攻击都是相当普遍的，并认为嵌入算法是未知的。但是还有很多情况下，攻击者可以得到嵌入和检测的详细细节，隐藏过程相对来说很简单。在这种情况下，被隐藏的信息可以被检测到，最终被删除。

7.5.1 对"回声隐藏"的攻击

对付回声隐藏(3.3.3 节提到的)的"明显"攻击是检测回声，然后通过简单地逆转一下卷积公式来删除它，问题的关键是在不了解原始目标和回声参数的情况下检测回声。显然用于合法系统的技术也可以被攻击者利用，但需要更多一点的工作，回声隐藏系统是基于倒谱分析的，攻击者用相同的检测函数，但是把它与蛮力搜索结合在一起。

以下的倒谱分析想法是由 Bogert 等提出的[23]。假设我们有一个信号 $y(t)$，包含一个简单的单回声(即 $y(t)=x(t)+\alpha x(t-\Delta t)$)。如果我们定义一个 Φ_{xx} 为 x 的功率谱，那么 $\Phi_{yy}(f)=\Phi_{xx}(f)(1+2\alpha\cos(2\pi f\Delta t)+\alpha^2)$，其对数近似于：$\log\Phi_{yy}(f)\approx\log\Phi_{xx}(f)+2\alpha\cos(2\pi f\Delta t)$。这是频率 f 的函数，功率谱增加了"倒频"Δt，就是 $\cos(2\pi\Delta tf)$的频率。后面的函数的自协方差强调峰值出现在"倒频"Δt 上。

如果用倒频的修改版，结果会好些：$Co\Phi o\ln o\Phi$，C 是自协方差函数 $(C(x)=E((x-\bar{x})(x-\bar{x})^*))$，$\Phi$ 是功率谱密度函数，o 是组合运算。对一些像音乐这样的随机信号做实验，结果表明当一个人工回声加到信号上时，这种方法返回相当准确的时延的估计值。在检测函数中只有回声延迟 0.5 与 3 毫秒之间是可以的，低于 0.5 毫秒，函数就不能正确工作，而高于 3 毫秒，回声就能听出来了。顺便提一句，在原始回声隐藏系统中，回声延时选择在 0.5 和 2 毫秒之间，具有最好相关振幅的回声是在 0.8 毫秒附近[24]。

7.5.2 "双峰值"攻击

在一些情况下，要加水印的图像有特定的性质，这些性质帮助恶意的攻击者获取水印本身的信息，例如，一个图片只有少量不同的颜色，像卡通图片，它在颜色矩形图上提供明显的峰值(见图 7.6)。在图 7.6 中，上面两幅图分别是"狒狒"原始图像的颜色矩形图(左)和加了水印图像的颜色矩形图(右)，下面两幅图分别是"君主"原始图像的颜色矩形图(左)和加了水印图像的颜色矩形图(右)。在同一幅图中，画了所有三色元素。双峰值攻击利用这些性质来恢复简单的扩频水印。我们只在灰度图像情况下阐述这种攻击。读者可以参考 Maes 的文章[25]来了解彩色图像情况下这种攻击的详细解释。

在 6.4.1 节中我们已经看到，一个基于扩频的数字水印的简单例子就是对每一个像素随机地加上或减掉一个固定值 d。于是每一个像素值有 50% 的机会增加或减少。设 n_k 是灰度值为 k 的像素的个数，假设对于一个特定的灰度值 k_0，与它相邻的第 d 个临近颜色不会出现，所以 $n_{k_0-d}=n_{k_0+d}=0$。因此，加水印后，期望的像素个数为：$\hat{n}_{k_0-d}=\hat{n}_{k_0+d}=n_{k_0}/2$ 而 $\hat{n}_{k_0}=0$。所以，用一组类似的等式，在某些特定的情况下，可能能恢复出原始矩形图的分布和嵌入水印的值。

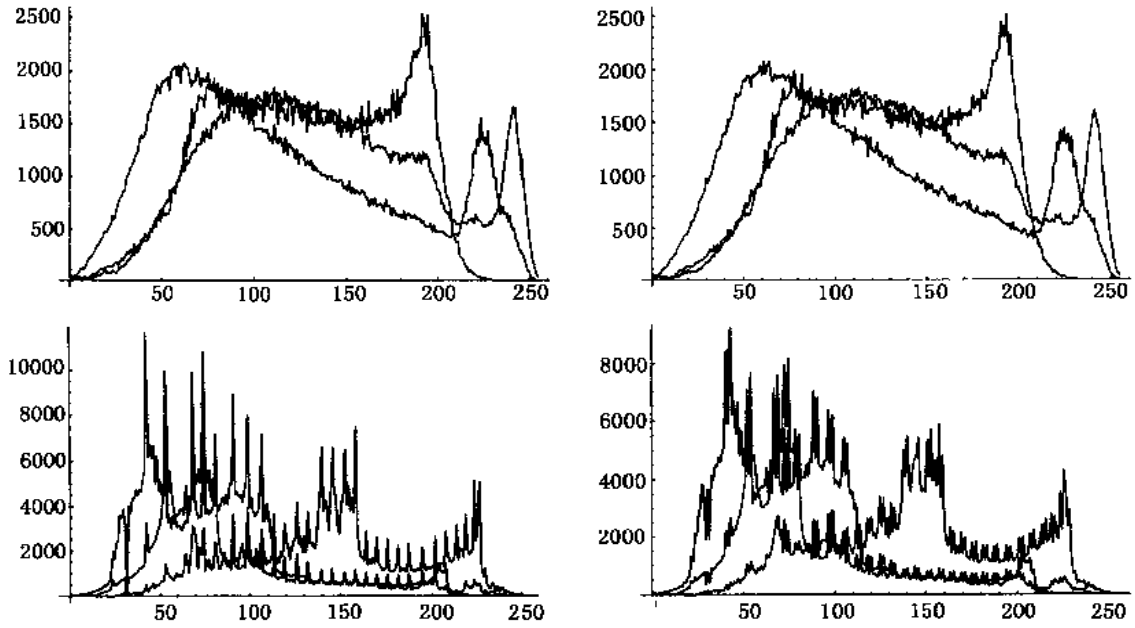

图 7.6　颜色矩形图与"双峰值"攻击

7.6　体系结构问题

我们所考虑的攻击都基于信号处理，但这并不是唯一的问题。在这一节中我们讨论由系统设计失误引起的水印的缺陷，系统设计没有考虑实际问题，如人为因素、用户接口和实现上的缺陷。

7.6.1　人为因素

典型的用户通常对图像水印只是有限的了解，不想花很多的时间去使用图像处理软件中的一个函数。水印应该像"黑箱子"，而用户输入原始图像，通过某种魔力箱子输出加了水印的图像。用户不必要理解具体的细节。

经常是画家和设计者创建的图像需要版权保护。画家不想加入水印后降低画的质量，这使我们相信画家要么加入弱水印[1]要么不加任何东西。很难说服画家们仅仅为了安全性而降低他们作品的图像质量。

7.6.2　用户接口

考虑到上面所描述的人为因素，我们可以看到用户接口(UI)成为安全体系结构的重要组成。Whitten 和 Tygar 演示了一个有缺陷的 PGP 用户接口如何限制了系统的安全性[26]。UI 应该为用户提供一个水印效果的清楚模型，防止用户误用。这需要消除用户对水印技术的错误概念。

[1] 现在的水印方案提出一种在水印强度和图像降质之间的权衡。

因为水印嵌入程序通常是与图像处理应用在一起的，所以用户在处理图像时可能会误删水印。在当前的应用中，比如说 Adobe Photoshop 的 PictureMarc，这个问题就没有解决，用户不清楚水印被减弱是因为改变图像颜色、修改边界或者是什么其它操作导致的。很明显，UI 应该防止用户偶然删掉或削弱水印。一种可能的改进就是在屏幕上放置一个水印强度的标示器，这样用户可以看到水印对图像变换是怎样反应的。理论上水印嵌入应该是图像出版之前的最后一步，因此软件可以将嵌入过程推迟直到用户要保存图像之前。同样在下载图像时，水印可以先提取出来，用户再对没有水印的图像进行操作。水印将变成一个对用户透明的操作。

7.6.3 实现过程的缺陷

大多数对密码系统的攻击来自于机会主义者对偶然发现的漏洞的利用，即使在攻击很脆弱的系统时，也很少使用密码分析学[27]。我们不能期望版权水印系统有什么不同，在因特网上对最广泛应用的图片水印方案的第一次攻击中，这种模式也会被使用。这种攻击利用了实现中的缺陷，而不是水印算法，即使算法是脆弱的（水印可以用 StirMark 变得不能用）（见 7.3.1 节）。

在这个系统中，每个用户有一个 ID 和一个两位口令，这是用户在向水印软件供应商注册时所得到的。在嵌入水印时，检查 ID 和口令是否相符。这种检查想来可以防止攻击者用已知 ID 嵌入水印，但不幸的是，设计上的缺陷允许攻击者用两种不同的方法攻破系统。

第一种攻破系统的方法是：用调试程序攻破软件，使口令检查机制不能正常工作。这种攻击可以从网上得到[4]。第二种方法是与 ID 相匹配的秘密口令只有两个十进制位长，所以只有一百种可能，平均花一百秒就可以找到任何用户 ID 的秘密口令！更别提这个过程自动完成多容易了。我们顺便提一下，如果 ID 公开，那么口令的搜索或分解将使任何一个用户都能被冒充。

对这个程序更深入的分析还可以让罪犯修改 ID、已加水印图像的版权以及应用类型（像成人内容或一般公共内容）。在嵌入水印之前，程序先检查图片中是否有水印，但这个检查可以通过调试程序很轻松地跳过，结果可能用另外一个水印覆盖任何已存在的水印。

对个人号码的穷举搜索可以通过延长搜索时间来防止，但对分解攻击没有显著的解决方法。如果防篡改软件[28]不能给予足够的保护，那么人们可以用在线系统，每个用户和一个信任方共享一个保密的嵌入密钥，用这个密钥嵌入某种数字签名。有两种相互分离的带密钥的操作，即身份认证（可以用签名来实现）和嵌入或隐藏操作。

7.6.4 自动蜘蛛限制

在 7.3.2 节中我们讨论了马赛克攻击，它可以阻止 Web 蜘蛛检测到网上图像中的水印，即使图像外表没被改变。但不幸的是，Web 蜘蛛在网上搜索盗版图像时有更多的问题存在。

第一个问题是带宽，在因特网上爬行需要很高的带宽。这使得个人用户在网上搜索自己的被侵权的图像是不可能的，因此用户需要在公共服务上注册寻找图像。如果他用秘密水印，他将需要用蜘蛛注册私钥，这会引起很多密钥管理和安全问题。

现在让我们看看整个周期，分析一下我们会遇到的问题。设想 Alice 是一个画家，她创造了持有版权的有价值的图像。嵌入秘密水印后，她在 Mallory 上注册检测侵权，并把密钥给了

他[2]。在这个情景中，Mallory 已经偷了 Alice 的图像并自豪地放到了自己的网页上。因为 Mallory 知道 Alice 拥有这些图像的版权，所以他编程让他的 Web 服务器对 Web 蜘蛛不提供所偷图像的服务。假设没有很多公司上网，这个防御措施就很容易实现。为了避免猜疑，当 Web 蜘蛛对被偷图像提出要求时，服务器可以返回一个任意图像。

Web 蜘蛛的另一个缺陷是对于访问控制网站或付费站点，蜘蛛没有身份认证或信息付费就不能访问图像。我们不知道哪个法律能允许蜘蛛或政府代理为了调查而免费搜索访问控制站点。即使有这样的法律，网络服务器也可以检测这样的用户，给他一个假冒的图像。不幸的是，访问控制（付费）站点上的侵权大多直接伤害内容的创建者，因为他们是从创建者作品中挣钱的。如果蜘蛛在检查图像之前不得不用信用卡买下图像，那他就成了网站的目标，提供“有保证的销售”了。这些问题都是摆在 Web 蜘蛛面前要克服的障碍。

Java 应用程序或 ActiveX 控件控制那些能嵌入浏览器显示图片的动态对象，因此他们为蜘蛛又提出了一个大问题，Java 程序甚至可以实时地解码图片。打败这种技术将需要修补网页、检测图片和检测它们是否有水印。

另外，Mallory 甚至可以不真正偷图像而将 Alice 的图像放到自己的网页上。通过 HTML，Mallory 可以不用将图像拷到自己的网上空间，而是在 Alice 的服务器上直接访问图像。然后参观者的浏览器从 Mallory 的服务器上下载网页，从 Alice 的服务器上下载图像。我们相信 Alice 控告侵权将很困难，因为图像只有在她的服务器上才找得到。

我们现在了解了有很多方法可以阻止 Web 蜘蛛找到被盗图像。下面我们将考虑侵权被检测到之后的问题。

7.7 法律攻击

7.7.1 国外服务器

第一种情况是在一个没有加入有关版权保护的伯尔尼协定的国家（见第九章），Mallory 建立自己的网络服务器发布侵权的图像、音乐等。没有办法阻止他做这些非法发布，因为国家不提供起诉的合法法律基础。所有的知识产权形式都有这个问题。过去的事实说明这种情况并不是牵强附会的。事实上，最近的很多事件证明了这种攻击是相当普遍的[3]。随着因特网的延伸，高速连接到对侵权还处于传统意识上的国家，情况变得更糟糕。如果这些国家还不改变他们的法律的话，这个问题不能从技术上解决。

7.7.2 欺骗攻击

在一些实施版权保护的伯尔尼协定的国家，对侵权的执法不是一个简单的任务。在法庭上证明侵权很困难。问题复杂在Alice不能声称“Mallory在他的网络服务器上发布我的版权图

[2] 在这里我们需要说明一点：如果存在类似于公钥密码的水印算法，这种情况就不是问题。每个用户用私钥嵌入水印，每个人能用相关的公钥来验证。不幸的是，这种方法不存在，问题常常是象“水印的圣杯”。

[3] 这种例子包括印度尼西亚的服务器发布高清晰度的侵权音乐。[29]中给了另外一个例子，Warner Bros 在寻找在网上发布 Madonna 即将出版的专集“Ray of Light”的走私犯。

像",我们需要欺骗行为的有声证据。例如,一个公正的目击证人能提供解决方法。但是问题是网络服务器不发出认可的响应,因此证人怎能知道数据来自哪儿呢?有这样的可能,Mallory试图说服 Alice,是 Bob 偷了她的图像。但是 Bob 是个好人,不会偷任何图像。所以 Mallory 需要骗 Alice。一个简单的方法就是用域名服务器(DNS)来欺骗,当 Alice 访问 Mallory 机子上的 <http://www.bob.com/image.gif> 时,Mallory 的 DNS 查找功能返回一个错误的地址为 www.bob.com,然后把"被盗"图像发到 Alice 的浏览器上。

另外一种情况,Mallory 真正偷了 Alice 的图像,Alice 写了起诉书。但是在这期间,Mallory 从服务器上删除了这个图像。如果 Alice 请公证人来看 Mallory 的网站以便证实被偷图像确实在那儿,Mallory 可能拒绝发图像给公证人员。这通常很难办到,但是,对某些域名拒绝数据传输在技术上是可行的。

这些攻击说明法律系统需要扩展以防止这些问题。

7.8 结　论

我们描述了信息隐藏系统的大量攻击手段,揭示了当前水印方案存在的很多缺陷。从这些攻击中,我们知道同步破坏是对付大多数水印技术的有效工具。这使我们提出检测(而不是嵌入)是数字水印的核心问题。我们还说明了软件设计师所不期望的:非常简单的攻击就能欺骗某些水印系统(像马赛克攻击)以及做生意的人们和客户用松散的工具和压缩就造成实施不满意。虽然听起来很耳熟(比如,[27,30]),但确实很令人吃惊的是没有人知道"健壮性"对数字水印的内容意味着什么。

了解这些攻击(以及那些将要出现的攻击)将会帮助版权标记系统的设计者们提出更好的方案。对模板这种新技术的介绍和对抵抗随机几何变形的尝试,已经证明了这个观点的正确性。

水印引人了一种新的模式,在这里信号处理、计算机安全、密码学、法律和商业集中起来保护摄影师、数字画家、歌唱家、作曲家等所有版权拥有者的权益。在加利福尼亚的某个记录标签上,一位高级主管对数字水印的期待是非常意味深长的,他说:"迟早任何密码系统都能被攻破。我们需要水印技术告诉我们是谁干的"[31]。

但是目前发展水平还远没有达到工业上能实现的地步,缺少标准化、互用性问题和对水印系统缺少一系列精确、现实的需求,这些几乎仍阻止着版权保护机制的发展。健壮的数字水印和指纹可能存在(在信号处理灵敏度、感光度中),但在他们被证明之前,本章中提到的攻击方法有助于展示它的反面。

参考文献

[1] "Request for Proposals-Embedded signaling Systems," International Federation of the Phonographic Industry, 54 Regent Street, London W1R 5PJ, 1997.

[2] Kutter, M., and F. A. P. Petitcolas, "A Fair Benchmark for Image Watermarking Systems," in Proceedings of the SPIE 3657, Security and Watermarking of Multimedia Contents, 1999, pp. 226 - 239.

[3] Carver, S., B.-L. Yeo, and M. Yeung, "Technical Trials and Legal Tribulations," Communications

of the ACM, vol.41, no.7, Jul.1998, pp.44 - 54.

[4] Anonymous, "Learn Cracking IV-Another Weakness of PictureMarc," posted by < zguan.bbs@bbs.ntu.edu.tw > on < news:tw.bbs.comp.hacker > , mirrored on < http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/digimarc-crack.html > , 1997. Includes instructions to override any Digimarc watermark using PictureMarc.

[5] Cox, I.J., et al., "Secure Spread Spectrum Watermarking for Images, Audio and Vidio," in International Conference on Image Processing, IEEE, Lausanne, Switzerland, 16 - 19 Sep.1996, pp.243 - 246.

[6] Cox, I.J., and J.-P.M.G. Linnartz, "Some General Methods for Tampering with Watermarks," IEEE Journal of Selected Areas in Communications, vol.16, no.4, May 1998, pp.587 - 593.

[7] Langelaar, G.C., R.L. Lagendijk, and J. Biemond, "Removing Spatial Spread Spectrum Watermarks byNon-linear Filtering," in 9th European Signal Processing Conference, Island of Rhodes, Greece, 8 - 11 Sep.1998, pp.2281 - 2284.

[8] Bender, W., et al., "Techniques for Data Hiding," IBM Systems Joural, vol.35, no.3 and 4, 1996, pp.313 - 336.

[9] Pitas, I., "A Method for Signature Casting on Digital Images," in Internatoinal Conference on Image Processing, vol.3, sep.1996, pp.215 - 218.

[10] Petitcolas, F.A.P., and R.J. Anderson, "Evaluation of Copyright Marking Systems," in IEEE Multimedia Systems, Florence, Italy, 7 - 11 Jun.1999, pp.574 - 579.

[11] Petitcolas, F.A.P., R.J. Anderson, and M.G. Kuhn, "Attacks on Copyright Marking Systems," in Proceedings of the Second International Workshop on Information Hiding, vol.1525 of Lecture Notes in Computer Science, Springer, 1998, pp.218 - 238.

[12] Petitcolas, F.A.P., "Attaques et évaluation des filigranes numériques," in Cinquièmes journées d'études et d'échanges sur la compression et la représentation des signaux audio-visuels(CORESA'99), Centre de recherche et développement de France Télécom(Cnet), EURÉCOM, Conseil Général des Alpes-Maritimes and Télécom Valley, Sophia-Antipolis, France, 14 - 15 Jun.1999.

[13] Ó Ruanaidh, J.J.K., and S. Pereira, "A Secure Robust Digital Image Watermark," in International Symposium on Advanced Imaging and Network Technologies-Conference on Electronic Imaging: Processing, Printing and Publishing in Colour, Europto, International Society for Optical Engineering, European Optical Society, Commission of the European Union, Directorate General XII, Zürich, Switzerland, May 1998.

[14] Davoine, F., et al., "Watermarking et résistance aux déformations géométriques," in Cinquièmes journées d'études et d'échanges sur la compression et la représentation des signaux audiovisuels (CORESA'99), Centre de recherche et développement de France Télécom(Cnet), EURÉCOM, Conseil Général des Alpes-Maritimes and Télécom Valley, Sophia-Antipolis, France, 14 - 15 Jun.1999.

[15] Petitcolas, F.A.P., "2Mosaic," < http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/ > , 1997.

[16] Linnartz, J.-P., T. Kalker, and G. Depovere, "Modeling the False Alarm and Missed Detection

Rate for Electronic Watermarks," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 329 – 343.

[17] Holliman, M., Personal communication, 1999.

[18] Ó Ruanaidh, J. J. K., W. J. Dowling, and F. M. Boland, "Watermarking Digital Images for Copyright Protection," IEE Proceeding on Vision, Signal and Image Processing, vol. 143, no. 4, Aug. 1996, pp. 250 – 256.

[19] Craver, S., et al., "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications," IEEE Joural of Selected Areas in Communications, vol. 16, no. 4, May 1998, pp. 573 – 586.

[20] Cox, I., et al., "Secure Spread Spectrum Communication for Multimedia," Technical report, N. E. C. Research Institute, 1995.

[21] Perrig, A., A Copyright Protection Environment for Digital Images, Diploma dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Feb. 1997.

[22] Linnartz, J.-P. M. G., and M. van Dijk, "Analysis of the Sensitivity Attack Against Electronic Watermarks in Images," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Science, Springer, 1998, pp. 258 – 272.

[23] Bogert, B. P., M. J. R. Healy, and J. W. Tukey, "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking," in Symposium on Time Series Analysis, New York, New York, USA: John Wiley&Sons, Inc., 1963, pp. 209 – 243.

[24] Gruhl, D., W. Bender, and A. Lu, "Echo Hiding," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 295 – 315.

[25] Maes, M., "Twin Peaks: The Histogram Attack on Fixed Depth Image Watermarks," in Proceedings of the Second International Workshop on Information Hiding, vol. 1525 of Lecture Notes in Computer Seience, Springer, 1998, pp. 290 – 305.

[26] Whitten, A., and J. D. Tygar, "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0," in 8th USENIX Security Symposium, August 1999.

[27] Anderson, R. J., "Why Cryptosystems Fail," Communications of the ACM, vol. 37, no. 11, Nov. 1994, pp. 32 – 40.

[28] Aucsmith, D., "Tamper Resistant Software: An Implementation," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer, 1996, pp. 317 – 333.

[29] Wilson, D. L., "Copyright vs. the right to copy," San Jose Mercury News, February 28 1998. < http://www.mercurycenter.com/business/center/copy030198.htm > .

[30] Anderson, R. J., "Liability and Computer Security: Nine Principles," in Computer Security-Third European Symposium on Research in Computer Security, vol. 875 of Lecture Notes in Computer Science, Springer, 1994, pp. 231 – 245.

[31] Yoshida, J., "Digital Watermarking Showdown Between ARIS and BlueSpike," EETimes on-line. 1999. Quotation of a senior executive at a California-based record label.

第八章 数字指纹

(Jong – Hyeon Lee)

8.1 引 言

指纹是指一个客体所具有的模式,它能把自己和其它相似客体区分开。它们有许多的应用,但在本章中,我们只介绍它们怎样用于数据的版权保护。我们感兴趣的技术不是依赖于抵抗篡改,即,不能防止用户拷贝数据,而是它们能使数据所有者去追踪非法散布数据的授权用户。例如,在加密的卫星电视广播中,发送给用户们一组用于解密视频流的密钥,并且电视台能在每个传输包中插入指纹比特来检测非授权的使用。如果一组用户把他们的密钥的子集分给非授权用户,以致于他们也能解密视频流数据,那么一旦捕获到非授权译码者,就至少能追踪到一个密钥提供者[1]。在这方面,指纹通常在叛逆者追踪问题的背景中进行讨论。

指纹还有另一个应用,它们能用作一种高速检索的手段。例如,剑桥大学图书馆使用六个字符的指纹码来检索期刊杂志。如果标题是一个单词,仅需要保留前六个字符。否则,使用杂志标题的第一个单词的前三个字符和标题中其后三个单词的首字符[2]。例如,"Computer & Communications Security Reviews"能使用"comcsr"很容易地找到。在数据库查找中使用相似的思想,但用了包括哈希在内的更复杂的技术。通常加密技术不用于这种目的。在这两种情形中,基本原理是相似的,但外在表现和技术细节是不同的。

指纹操作指对一个客体加入指纹或验证一个客体已固有的指纹的过程。以下几节将讲述指纹的例子、指纹技术的术语和要求、指纹的分类、研究历史和重要指纹方案的简要概况。

8.2 指纹的例子

指纹已经使用了几个世纪,并且我们能找到许多使用它的经典例子。我们列出几个在计算机时代之前的典型指纹技术的例子。这些例子能帮助我们理解指纹是什么和为什么需要它。

• **人的指纹** 大家都知道每一个人的指纹有不同的模式,可借以与其它人相区分。基于调查取证的目的,从入狱者和罪犯中采集人的指纹。由于它是一个比较容易的识别手段,一些国家在居民身份证上也采用它(例如,韩国和法国)。韩国居民身份证上含有其持有人的指纹。法国居民在申请新的身份证时必须提供他的指纹。人的指纹也能用于访问(接入)控制,这可以在几部间谍电影中看到。在第二次世界大战期间,美国战略服务办公室(OSS)用它们代理人的拳头来标识他们[3]。相似的生物测定学手段,如,人的虹膜和声音纹理可以用于同样的目的。一个英国大厦协会已经在自动报话机上运作了大量的人的虹膜扫描系统,并且将在自动报话机上采用这种技术识别用户身份而不是键入密码。

• **射击的子弹** 每种武器有它自己型号的射击子弹,而其型号依赖于制造商和武器类

型。打字机都很相似,但每一个打字机有它自己的排版模式。

• **序列号**:被加工产品的序列号对每一个产品来说是唯一的,并且能用于区分它们自己。

• **爆炸物的编码微粒** 一些爆炸物在加工时被附上一些小的编码微粒,使得爆炸之后能找到这些微粒。通过检查这些微粒,就能识别出生产商、类型、生产时间等。

• **地图** 有时候,绘制地图时故意与真实地图产生适当的轻微的变化,借以识别不同的拷贝。

由于数字数据上的指纹是版权保护较容易和廉价的手段,因此在计算机和通信中对指纹的要求日益增强。我们能找到几个计算机时代的指纹例子。

• **邮件地址的前缀** 在支持 IETF RFC 754[5]的邮件系统中,可以在通常的邮件地址上加一个前缀。这在注册一个在线服务时很有用。例如,Bob(Bob@foo.org)能在 Mallory 站点处注册 Mallory + Bob@foo.org。然后若他在这个地址收到一个未经请求的消息,他能断言 Mallory 把地址送给了某个重要的投递者。

• **PGP 公钥** PGP 是使用最广泛的公钥软件包之一[6],并且对于 PGP 公钥,指纹也被作为最重要的标识方法之一来使用。在 PGP 中,指纹是公钥比特的 MD5[7,pp.436 - 441]哈希值,这些公钥比特包括公开模数和加密指数[8]。指纹几乎是唯一的,并能看作是一个密钥目录中的标识器,例如全球因特网可信注册器[9]。这种注册包括 PGP 指纹和作为公钥标识符的密钥长度。

• **数字声音/视频** 已经建议使用指纹来检查视频数据的私有性。在一个付费电视广播系统中,指纹能用于跟踪非法用户[1].

• **文档** 作为版权保护方法,在文档中使用指纹来阻止拷贝[10,11]。

指纹也能嵌入在计算机程序、多媒体数据、数据比特流等之中。

8.3 术语和要求

标记是客体的一部分并有若干个可能的状态;指纹是标记的集合;发行人是一个授权提供者,他将嵌入指纹的客体提供给用户;授权用户是一个获得授权使用某一嵌入指纹客体的个人;攻击者是非法使用嵌入指纹客体的个人;叛逆者是非法发行嵌入指纹客体的授权用户。

为了帮助理解这些术语,我们想一想图像发行的情形:图像制作者故意在每一个发行拷贝中置入微小的差异。这些差异就是标记,并且这些差异的集合就是指纹。图像制作者是发行人,而购买者就是用户。一组用户可以比较他们的图像并发现故意留下的差异,然后他们可以将这个信息送给或卖给其它人,以至于其它人能制作非法拷贝(这样一个攻击被称为合谋攻击)。在这个例子中,给出差异信息的那组人就是叛逆者,而那些制造非法拷贝的人就是攻击者。

在指纹中一般的威胁模型如下:发行者的目标是识别与攻击者达成妥协的用户,攻击者的目标是防止发行者的识别。

指纹就其本身而言仅提供非法使用的检测而并不能避免非法使用,但检测非法使用的能力可以有助于阻止个人从事这些非法活动。作为拷贝跟踪和拷贝缩影的指纹,其需求包括合谋容忍以及所有对水印的需求:

• **合谋容忍** 即使攻击者获得了一定数量的拷贝(客体),通过比较这些拷贝,不应该能

找到、生成或删除该客体的指纹。特别地，指纹必须有一个共同的交集。

• **客体质量容忍**　加入标记不允许明显地减少客体的用途和质量。

• **客体操作容忍**　如果攻击者篡改客体，除非有太多噪音使客体不可用，否则指纹仍应能存在于客体中。特别地，指纹应能容忍有损数据压缩。

8.4 指纹分类

指纹技术可按下列特征进行分类：加入指纹的客体、检测的灵敏度、嵌入指纹的方法、生成的指纹。这四种分类并不是互相排斥的。下面，我们采用 Wagner[12]的分类法。

8.4.1 基于客体的分类

客体的自然属性是一个最基本的标准，这是因为它能提供一种定制的方法为客体嵌入指纹。基于客体分类时，能分为两种：数字指纹和物理指纹。如果加入指纹的客体是数字格式，使得计算机能处理其指纹，我们称它是数字指纹。

如果一个客体能用它自己的物理特性与其它客体区分开来，我们说这是物理指纹。人的指纹、虹膜模式、声音模式和一些爆炸物的编码微粒都属于这种类型。

8.4.2 基于检测灵敏度的分类

指纹方案对非法使用的灵敏度是另一个分类标准。基于对侵害的检测灵敏度，我们把指纹分为三类：完美指纹、统计指纹和门限指纹。

如果对客体的任何修改使指纹不可识别的同时，也导致了客体不可用，我们称这种指纹为完美指纹。因此指纹生成器总能通过检测一误用客体来识别出攻击者。统计指纹则没有这么严格。假定有足够多误用客体可供检测，指纹生成器能以任意希望的可信度来确认越轨用户。然而，这种识别器不是绝对可靠的。门限指纹是上面两种的混合类型，它允许一定程度上的非法使用，也就是门限，只有达到门限值时，才去识别非法拷贝，这样就允许对一个客体进行拷贝，只要其拷贝数量小于门限即可，并且根本不对这些拷贝作任何检测。当拷贝数量超出门限时，就追踪拷贝者。

8.4.3 基于嵌入指纹方法的分类

基本的指纹处理方法，如识别、删除、添加、修改，也已经被作为另一种分类标准。如果指纹方案由识别和记录那些已成为客体一部分的指纹组成，那么它属于识别类型。例如，人的指纹和虹膜模式。在删除类型指纹中，嵌入指纹时原始客体中的一些合法成分被删除。若在客体中加入一些新的成分来嵌入指纹，那么它就属于添加类型。添加部分可以是敏感的，也可以是无意义的。若修改客体的某部分来嵌入指纹，它就是修改类型，例如变化的地图等。

8.4.4 基于指纹的分类

我们可以区分两种类型的指纹：离散指纹和连续指纹。如果生成的指纹是有限的离散取值，那么就称该指纹为离散的，例如数字文件的哈希值。如果生成的指纹是无限的连续取值，那么我们称该指纹为连续的，大部分物理指纹属于这种类型。

8.5 研究历史

在计算机出现之前，人们只研究和开发了物理指纹技术。随着数字数据重要性的日益增加，人们越来越渴望使用指纹技术保护知识产权，因为指纹只需要轻型的加密性能就达到目的。可加入指纹的数字数据的例子包括：文档、图像、电影、声音、可执行文件等等。

和其它加密技术一样，有一些已知的问题需要解决。当在数字数据中嵌入指纹时，一个必须考虑的问题是合谋问题。假设发行了一个嵌入指纹的数字图像，如果一群获得它的用户比较他们的拷贝，他们能很容易地发现所有的标记。然后用户能去掉这些标记，窜改这些差异，并重售这幅图像而不用担心被追踪。

合谋问题首先由 Blakley 等人[13]提出，并由 Boneh 和 Shaw [14]提出一种对付较多用户合谋的解决方案。Low 和 Maxemchuk [15]给出了一个基于多方密码协议模型的合谋分析。

就像对密码术中密钥的重要性一样，叛逆者追踪对指纹技术非常重要。它由 Chor 等人[1]在广播加密中引入。当数据(例如，付费电视中的电影)以加密形式广播并只卖解密密钥时，不同的密钥卖给不同的付费电视用户。进一步来说，这种加密方案需满足：所有密钥都能用于解密同一个密文。由于每一个用户的解密密钥是不同的，付费电视公司能追踪哪个用户制作了他的密钥的非法拷贝。

最近，几项研究进一步加强了各个方面的指纹方案的功能(如，非对称指纹和匿名指纹)。非对称指纹由 Pfitzmann 和 Schunter[14]提出。与常规的指纹方案不同，这里只有嵌入指纹的客体的买方才知道带指纹的数据。当销售方发现了拷贝，他仍然能识别买方，并向第三方证明：正是这个买方从他那里购买了这个拷贝。Pfitzmann [16,17]也提出一种使用非对称指纹技术进行叛逆者追踪的方案。匿名指纹是由 Pfitzmann 和 Waidner [18]提出的，它类似于盲签名，它使用一个可信的称作注册中心的第三方来识别被怀疑有非法行为的买方。也就是，销售方若没有注册中心的帮助就不能识别他。通过使用注册中心，销售方不再需要发行商保存用户和指纹相对应的详细记录。在下节中，我们将描述这些方案。

8.6 指纹方案

本节中，我们列出指纹技术中的重要成果，并对统计指纹、叛逆者跟踪、合谋安全指纹、非对称指纹和匿名指纹进行概述。

8.6.1 统计指纹

在 1983 年，Wagner [12]提出了基于假设检验的统计指纹。其详细过程如下：假定有 n 个实数 $v_1, v_2, \cdots v_n$ 和 m 个用户，假定我们有足够多的样本数据以致于我们能对统计假设进行检验。为了在统计指纹中适合使用，必须能对每一个 v_j 找到一个值 $\delta_j(\delta_j > 0)$，使得对每一个 $i \neq j$，v_j 的 δ_j 邻域与 v_i 的某个邻域不相交。然后，每个用户在闭区间 $[v_j - \delta_j, v_j + \delta_j]$ 中获得一个数值，它不同于其它用户的数值。大致上，用户获得的数值有一半在区间 $[v_j, v_j + \delta_j]$，另一半在区间 $[v_j - \delta_j, v_j]$ 中。发送给用户 i 的第 j 个数据的版本记作 v_{ij}。

假定数据以某种方式被误用，并且发行商能从他找到的非法拷贝中提取出数值 v_1'，v_2'，

$\cdots v_n'$。对每个在范围 $1 \leq i \leq m$ 中的 i，我们想检验这个假设，即返回数值源自用户 i。为了这样做，对一给定 i，我们检测这种似然统计量

$$L_{ij} := \frac{v'_j - v_{ij}}{\delta_j} \qquad 1 \leq j \leq n \tag{8.1}$$

也就是，$(L_{ij})_{1 \leq j \leq n}$是返回数值和给定用户 i 的数值间的归一化差。

对一给定 i，我们考虑两个不相交子集上$(L_{ij})_{1 \leq j \leq n}$的均值。令 μ_i^h 是这样一些 L_{ij}的均值，此时 v_{ij}是发送给不同用户的 V_j 的两种类型值中的较高的(即在右半区间取值)，令 μ_i^l 是另一些 L_{ij}的均值，此时 V_{ij}是这两个值中较低的(即在左半区间取值)，那么 $\mu_i^h \leqslant 0$ 和 $\mu_i^l \geqslant 0$。令 $\mu_i := \mu_i^l - \mu_i^h$。

假定攻击者在这些值返回为 v_j' 之前不对这些数值进行修改。如果攻击者从用户 i 获得数据，那么 $\mu_i = \mu_i^h = \mu_j^l = 0$。如果他是从别人处获得的，那么我们期望

$$\mu_i^h \approx -0.5 \text{，} \mu_i^l \approx 0.5 \text{，和 } \mu_i^l - \mu_i^h \approx 1 \tag{8.2}$$

因此如果没有作修改，除非 n 非常小，否则应能立刻识别出攻击者。

当攻击者改变了返回值，甚至对较大的 n，$\mu_i^h \approx 0$ 可能也不再能正确识别攻击者，这是因为攻击者可能根据一些分布用非零值修改这些值。然而，可以假定攻击者不能区分两种可能取值中哪个较大和哪一个较小。因此对足够大的 n，如果攻击者的值源自用户 i，可以期望 μ_i 接近于 0。另一方面，如果攻击者的值不是源自用户 i，对大的 n，我们能期望

$$\mu_i = \mu_i^l - \mu_i^h \approx 1 \tag{8.3}$$

因此，人们可以使用下面的算法。对每一个 i，计算出上面两个均值的差值 μ_i。如果对某一个 i，μ_i 接近于 0，并且对所有其它的 $k \neq i$，μ_k 接近于 1，那么这就为误用数据是源自用户 i 的假设提供了证据。通过对所有 i 检查 μ_i 的值，就能识别出哪个用户泄露了信息。

由于这个指纹方案是基于假设检验的，我们可以提高假设检验的可信度，然而，假设毕竟是假设，不能变成确定性事实。

8.6.2　合谋安全指纹

在 1995 年，Boneh 和 Shaw [14,19]引入了 c－安全码来获得合谋安全指纹。他们指出在指纹方案中合谋是最重要的问题，并且提出了针对合谋容忍指纹的一个清晰解决办法。

令 $\sum$是大小为 s 的字符表，表示了 s 个不同的标记状态。$\sum$中的字母可以是从 1 到 s 的整数。给定一个 l－比特的字 $x(x \in \sum^l)$和一个索引集合 $I = \{i_1, \cdots, i_r\} \subseteq \{1, \cdots, l\}$，我们用 $x|_I$ 表示字 $x_{i_1} x_{i_2} \cdots x_{i_r}$，这里 x_i 是 x 的第 i 个字母。我们称 $x|_I$ 为 x 在 I 中位置的约束。一个集合 $\Gamma = \{x^{(1)}, \cdots, x^{(n)}\} \subseteq \sum^l$ 称为一个(l, n)－码。对 $1 \leqslant i \leqslant n$ 中每个 i，将码字 $x^{(i)}$分配给用户 u_i。令 C 是一个用户联合，一个联合就是一个参与制作非法拷贝的用户集合。如果分配给 C 中所有用户的字在第 i 个位置相匹配，这里 i 属于集合$\{1, \cdots, l\}$，我们就说位置 i 对 C 来说是不可检测的；也就是对 $C = \{u_1, \cdots, u_c\}$，$x_i^{(u_1)} = x_i^{(u_2)} = \cdots = x_i^{(u_c)}$。假定第 i 个标记对联合是可检测的，那么该联合能产生一个客体，其中这个标记处于一个不可读的状态，以至于检查人员不能确定不可读标记处于哪一个状态。我们标注一个不可读状态的标记为“?”。令 R 是 C 中不可检测位置集合，那么就 C 中某个用户 u 而言，C 的可行集定义为

$$F(C) = \{x \in (\sum \cup \{?\})^l : x|_R = x^{(u)}|_R\}$$

因此可行集合包含与联合的不可检测比特相匹配的所有字。

我们想用下面这样的特性来构造码字，即没有一个联合能串通起来陷害一个不在联合中的用户。当我们限定 c 个用户组成的联合大小时，我们称这样的码为 c - 防陷害码。形式上，一个 c - 防陷害码码字 Γ 定义成一个满足 $F(W)\cap\Gamma=W$ 的码字，其中 $W\subset\Gamma$，Γ 的大小至多为 c。对二进制字符 $\sum=\{0,1\}$，我们定义 (n,n) - 码 $\Gamma_0(n)$ 为所有恰好含一个 1 的 n 比特二值码字。我们能证明 $\Gamma_0(n)$ 是 n - 防陷害码。

设想一个发行商用一个码 Γ 标识一个客体，并且用户的一个联合 C 合谋产生了一个由某个码 x 标识的非法客体，然后发布这个新客体。当发现一个非法客体时，我们想检测这个联合的一个子集，也就是说，我们想构造一个跟踪算法来检测这个子集。如果有一个跟踪算法 A 满足下面条件，那么我们就说码 Γ 是完全 c - 安全的，即如果一个或至多 c 个用户的联合 C 生成一个码字 x，满足 $A(x)\in C$；也就是，跟踪算法 A 输入 x 时必须输出一个联合 C 的成员。因此一个非法拷贝至少可跟踪到违规联合的一个成员。我们把跟踪算法 A 考虑成一个函数 A：$\{0,1\}^n\to\{1,\ldots,n\}$，这里 n 是用户数。下面将描述这样一个跟踪算法是怎样建立起来的。

假定 c 个用户的一个联合生成客体的一个非法拷贝。能使我们至少以概率 $1-\varepsilon$ 捕获联合 C 的一个成员的指纹方案，我们称之为具有 ε - 误差的 c - 安全码。通常，如果至多 c 个用户的一个联合 C 生成一个码字 x，满足 $Pr[A(x)\in C]>1-\varepsilon$，此处概率被随机比特 r 代替并且由联合作随机选择，若具有这样的算法 A，那么 Γ_r 是具有 ε - 误差的 c - 安全码。在构造合谋安全码字中，我们能获得一个跟踪算法，假定某个联合生成一个码字 x，这个算法能以概率 $1-\varepsilon$ 输出 C 的一个成员。

令 c_m 是一高度为 n 的比特列，前 m 个比特是 1 而其余的是 0。码字 $\sum_0(n,d)$ 由 c_1，$c_2,\ldots,c_{n-1}$ 的所有列组成，每一个重复 d 次。重复次数决定了错误概率 ε。令 $x^{(1)},\ldots,x^{(n)}$ 表示 $\Gamma_0(n,d)$ 的码字。发行商在客体中嵌入 $\sum_0(n,d)$ 的码字之前，他需作下面的随机选择：随机地挑取一个置换 $\pi\in S_l$，这里 S_l 是字母 l 上所有置换的完全对称群。使用码字 $\pi x^{(i)}$ 对用户 u_i 的客体拷贝加入指纹。注意所有的用户都应用同样的置换，并且 π 需要保密。置换对用户保密与用客体中哪个标记编码哪个比特来隐藏信息同样重要。

为派生一个跟踪算法，我们先定义一些概念。令 B_m 是所有具有某种属性的比特位置集，它满足用户在其位置能看到类型为 c_m 的列。也就是说，B_m 是这些比特位置的集合，前 m 个用户看到一个 1 并且其余用户看到一个 0。在 B_m 中元素个数是 d。对于 $2\leqslant s\leqslant n-1$，定义 $R_s=B_{s-1}\cup B_s$。对一个二进制串 x，令 $weight(x)$ 表示 x 中 1 的个数。

假定用户 s 不是生成字 x 的联合 C 的一个成员。保密的置换 π 防止联合知道哪一个标记代表了码字 $\Gamma_0(n,d)$ 中的哪一个比特，联合所知道的唯一信息是它能检测的标记值。观察显示一个不含用户 s 的联合也能确切了解到对所有比特位置 $i\in R_s$ 的相同值。

对一个 $i\in R_s$ 的比特位置，联合 C 不能区分 i 是在 B_s 还是在 B_{s-1} 中。这意味着无论他们使用哪种策略设置 $x|_{R_s}$ 的比特，在 $x|_{R_s}$ 中的 1 以很大概率均匀地分散在 $x|_{B_s}$ 和 $x|_{B_{s-1}}$ 之间。因此如果在 $x|_{R_s}$ 中的 1 不是以很大概率均匀地分布，那么用户 s 是生成 x 的联合的一个成员。通过一些计算，我们能获得在 $\Gamma_0(n,d)$ 中 d 的一种度量，对于 $n\geqslant 3$ 时，$d=2n^2\log(2n/\varepsilon)$，并且跟踪算法能明确表示如下：

算法：给定 $x\in\{0,1\}^l$，找到一个生成 x 的联合的子集。

1. If $weight(x|_{B_1}) > 0$, then output "user 1 is guilty"

2. if $weight(x|_{Bn-1}) < d$, then output "user n is guilty"

3. For all s = 2 to n − 1 do ; let $k = weight(x|_{K_i})$. If

$$weight(x|_{B_{s-1}}) < \frac{k}{2} - \sqrt{\frac{k}{2}\log\frac{2n}{\varepsilon}} \tag{8.4}$$

then output "user s is guilty"

需要澄清一件事，即如果在非法拷贝中找到字 x 包含一些不可读标记，那么在字 x 送给算法前将这些比特置 0，那么算法收到的是 $|0,1|^1$ 中的一个字。

8.6.3　非对称指纹

通常，指纹方案是对称的，也就是说，所有用户都能识别出嵌入指纹的拷贝。在电子商务中，此特性隐含了这样的含义，即发行商无法证明是一个特定用户而不是商人在非法发布客体。为了处理这种情况，Pfitzmann 和 Schunter[20]提出了非对称指纹(见图 8.1)。在这个方案中，只有买方知道具有指纹的数据。如果商人后来在某处发现了它，商人就能识别出它的买方，并能向第三方证明这个事实。

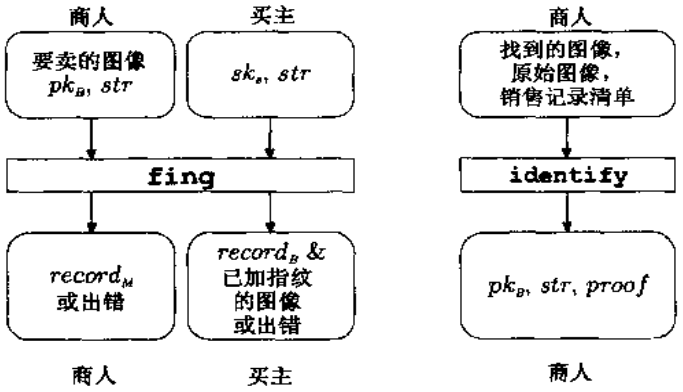

图 8.1　非对称指纹

这个方案由四个协议组成：key − gen、fing、identify、dispute。在密钥生成协议 ken − gen 中，买方生成一个公开钥 pk_B 和一个秘密钥 sk_B，并通过一个认证机构公布公开钥。让我们考虑以数字形式购买一个图像。当对指纹协议 fing 输入参数时，商人输入要卖的图像、用户的身份

pk_B 和一个描述这次购买的字符串 str，str 像一个搜寻关键字用来匹配多用户的输入。而且商人可以输入这个图像以前销售的记录清单。买方输入 str 和他的秘密钥 sk_B。结果输出给买方的是一个有很小误差的图像。买方也可以获得一个记录 $record_B$，它可以保存下来用于以后解决争议。如果协议失败，买方将获得一个"failed"结果。商人获得的结果是一个销售记录 $record_M$ 或"failed"。当商人找到一个拷贝并想识别出原始买主时，将他找到的图像、商人卖出的图像和这个图像的一个销售清单一起送入算法 identify。Identify 的输出是"failed"或是一个买者的身份 pk_B 和一个由买主签名的字符串 proof。争端协议 dispute 是一个在商人、仲裁者和可能的被控买主间的两方或三方协议。商人和仲裁者输入 pk_B 和 str，另外商人还要输入字符串 proof。如果加入被指控的买主，他输入 str，他的秘密钥 sk_B 和记录 $record_B$。协议输出给仲裁者的是一个布尔值，标识是否接受对用户的指控，也就是说，输出结果表示仲裁者是否接受商人已找到的图像就是具有 str 的被控用户买的图像。

8.6.4 叛逆者追踪

如果仅一个人知道秘密，而该秘密出现在晚间新闻上，那么罪犯是显然的。当知道这个秘密的人数量很大时，无论如何，一个更复杂的情况就出现了。如果他们所有人共享完全同样的数据，那么就解决不了谁有罪谁无辜的问题。从秘密共享者中找出叛逆者的一种可能方法是给所有共享者一个稍有差别的秘密。Chor 等人[1]应用这种思想解决盗版问题。关于这个应用，需要识别出是否正在进行盗版，并防止信息传送给盗版用户，但又不损害合法用户。进一步说，应该提供盗版标识的法律证据。在符合上面要求的条件下，描绘了一个应用他们的方案追踪盗版的实例，这个盗版是关于滥用一个广播加密方案的。

Chor – Fiat – Naor 方案[1]的详细操作如下：发行者生成一个有 r 个随机密钥的集合 R，并从 R 中对每个用户分配 m 个密钥，它形成了用户的个人密钥。注意不同的个人密钥可能有非空交集。

一个追踪叛逆者消息由多对（使能块，密码块）组成，参见图 8.2。密码块是在某些秘密随机钥 S 下对实际数据（如几秒钟的视频剪辑）的对称加密，使能块能够允许授权用户获得 S，并且是由发行者在部分或所有的 r 个密钥作用下加密的视频数据组成。每一个用户通过使用他的密钥对加密的视频数据进行解密，从而能计算出 S。也就是说，用户的密钥集和使能块对相应密码块来说成为输入，用来生成解密密钥。

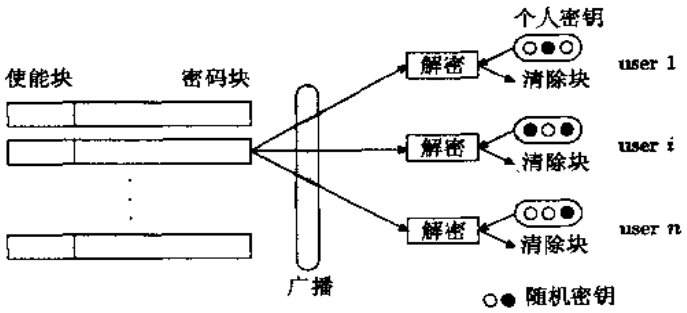

图 8.2 Chor – Fiat – Naor 方案

叛逆者可能合谋，并提供他们密钥的子集给某一非授权用户，以至于非授权用户也能解密密码块。方案的目标是以下面这种方式给用户分配密钥，也就是当俘获一个盗版译码器，并对

它拥有的密钥进行检测时，应可能至少检测出一个叛逆者。

我们期望追踪叛逆者方案对这样的叛逆行为能提供法律的证据。Pfizmann[20]集中考虑了这一点，并指出由于 Chor – Fiat – Naor 方案基于对称指纹，不能提供这种不可否认性。不可否认性具有防止用户否认其行为的特性，通过应用她的非对称指纹方案，Pfitzmann 提出了一种具有不可否认性的非对称追踪叛逆者方案。

8.6.5　匿名指纹技术

Chaum[21]介绍了一种签名方案，它能让一个签名者在数据上签名，而不泄露数据的内容，也就是所谓的盲签名。匿名指纹是盲签名的一个应用。Pfitzmann 和 Waidner[18]介绍了基于一个可信认第三方的匿名非对称指纹和非对称指纹方案[20]。买方能以匿名的形式购买信息，但是如果他们非法重新发行这种信息，仍然能被识别出来，参见图 8.3。

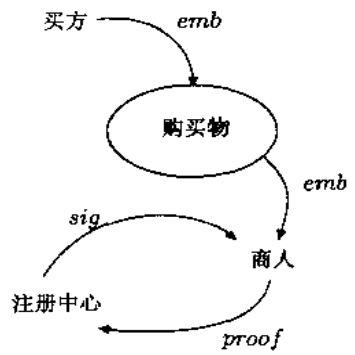

图 8.3　匿名指纹技术

电子市场应该与传统市场一样，提供相似的隐私权，也就是说，在电子购买中应该获得一定级别的匿名性。当人们用现金买商品时，在传统市场上没有人能跟踪这次购买。仅为了使用户用指纹标识他们自己，而破坏所有这样的匿名性是不值得的。

匿名指纹的基本思想如下：买方选择一个假名（即，签名方案中的一个密钥对（sk_B，pk_B）），然后用他的真实身份对其签名，表示他对这个假名负责。他从注册中心获得一个证书 $cert_B$。有了这个证书，注册中心宣布它知道选择这个假名的买方的身份（也就是，注册中心能把一个真人用一个假名代表）。然后当买方进行一次购买时，在不了解商人的情况下用标识这次购买的文本（*text*）计算出一个签名，$sig := sign(sk_B, text)$，然后将信息 $emb := (text, sig, pk_B, cert_B)$ 嵌入购买的数据中。他在一个比特承诺里隐藏这个值，并以零知识方式给商人发送证书和承诺。比特承诺 [22,23]是一种密码技术，它能传递数据并仍能在一段时间内保持秘密，详细信息可见[7, *pp*. 86 – 88]。当需要鉴别时，商人提取出 *emb* 并给注册中心发送 $proof := (text, sig, pk_B)$，并要求验证。作为回答，注册中心向商人发回用户的签名。于是，商人能够使用这个签名来验证所有的值并有证据指控买方。

8.7　结　　论

本章给出了指纹技术的概况和介绍，包括它的例子、分类、研究历史和最近成果及进展。本质上，指纹是区分拷贝的技术。随着对数字版权保护和拷贝控制的日益重视，这种技术可用于数字形式的知识产权的版权保护。

指纹不是设计用来揭示有版权的产品和产品所有者之间的准确关系的，除非他或她违反了它的合法使用。与密码技术相比，这种特性可看成是不完全的或不精确的，但它更可能吸引用户和市场，而高精确度并不一定能解决一切事情。它不仅没有使用加密中那么多的控制，还可提供给用户一些匿名。作为版权保护的应用，指纹技术是一种容易的、轻量级的并且是有效的手段，它也可以作为基于加密的拷贝管理技术的一种补充，并当加密密钥泄露后能提供一种补救保护措施。

参考文献

[1] Chor, B., A. Fiat, and M. Naor, "Tracing Traitors," in Advances in Cryptology, Proceedings of CRYPTO'94, vol. 839 of Lecture Notes in Computer Science, Springer – Verlag, 1994, pp. 257 – 270.

[2] Cambridge University Library, Title Fingerprint Search: Cambridge Union List of Serials, Cambridge, England, 1998.

[3] Kahn, D., The Codebreakers – The Story of Secret Writing, New york, USA: Scribner, 1996.

[4] Card International, Iris – recognition ATMs Planned for Rollout in 1999, Cambridge, England, Dec. 1998.

[5] Postel, J., "Out – of – Net Host Addresses for Mail," IETF RFC 754, The Internet Engineering Task Force, Apr. 1979.

[6] Zimmermann, P. R., The Oficial PGP User's Guide, Cambridge, Massachusetts, USA: MIT Press, 1995.

[7] Schneier, B., Applied Cryptography, New York, USA: John Wiley&Sons, 2nd ed., 1996.

[8] Comp. security. pgp, The comp. security. pgp FAQ, Oct. 1998.

[9] Anderson, R. J., et al., The Global Internet Trust Register 1999, Cambridge, Massachusetts, USA: MIT Press, Apr. 1999.

[10] Brassil, J., et al., "Electronic Marking and Identification Techniques to Discourage Document Copying," in Proceedings of Infocom'94, Jun. 1994, pp. 1278 – 1287.

[11] Heintze, N., "Scalable Document Fingerprinting," in Proceedings of the 2nd USENIX Electronic Commerce Conference, Oakland, California, USA, 1996, pp. 191 – 200.

[12] Wagner, N. R., "Fingerprinting," in Proceedings of the 1983 IEEE Symposium on Security and Privacy, Oakland, California, USA, Apr. 1983, pp. 18 – 22.

[13] Blakley, G. R., C. Meadows, and G. B. Purdy, "Fingerprinting Long Forgiving Messages," in Advances in Cryptology, Proceedings of CRYPTO' 85, vol. 218 of Lecture Notes in Computer Science, Springer – Verlag, 1986, pp. 180 – 189.

[14] Boneh, D., and J. Shaw, "Collusion – secure Fingerprinting for digital Data," in advances in Cryptology, Proceedings of CRYPTO'95, vol. 963 of Lecture Notes in Computer Science, Springer – Verlag, 1995, pp. 452 – 465.

[15] Low, S. H., and N. F. Maxemchuk, "Modeling Cryptographic Protocols and their Collusion Analysis," in Information Hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer – verlag, 1996, pp. 169 – 184.

[16] Pfitzmann, B., "Trials of Traced Traitors," in Information hiding: First International Workshop, Proceedings, vol. 1174 of Lecture Notes in Computer Science, Springer – Verlag, 1996, pp. 49 – 64.

[17] Pfitzmann, B., and M. Waidner, "Asymetric Fingerprinting for Larger Collusions," in Proceedings of the 4th ACM Conference on Computer and Communications Security, 1997, pp. 151 – 160.

[18] Pfitzmann, B., and M. Waidner, "Anonymous Fingerprinting," in Advances in Cryptology, Pro-

ceedings of EUROCRYPT'97, vol.1233 of Lecture Notes in Computer Science, Springer – Verlag, 1997.pp.88 – 102.

[19] Boneh, D., and J. Shaw, "Collusion – secure Fingerprinting for digital Data," IEEE Transactions on Information Theory. vol. IT – 44. no. 5, Sep. 1998, pp. 1897 – 1905.

[20] Pfitzmann, B., and M. Schunter, "Asymmetric Fingerprinting," in Advances in Cryptology, Proceedings of EUROCRYPT'96, vol. 1070 of Lecture Notes in Computer Science, Springr – Verlag, 1996, pp. 84 – 95.

[21] Chaum, D. L., "Blind Signatures for Untraceable Payments," in Advances in Cryptoloty, Proceedings of CRYPTO'82, Plenum Press, 1983, pp. 199 – 203.

[22] Brassard, G., D. L. Chaum, and C Crepeau, "Minimum Disclosure Proofs of Knowledge," Journal of Computer and system Sciences, vol. 37, no2, Oct. 1998. pp. 156 – 189.

[23] Naor, M., "Bit Commitment Using Pseudo – Randomness," in Advances in Cryptology, Proceedings of CRYPTO'89, vol. 435 of Lecture Notes in Computer Science, Springer – Verlag, 1990, pp. 128 – 136.

第九章　因特网版权与水印

(Stanley Lai 和 Fabrizio Marongiu Buonaiuti)

9.1　数字版权和水印[1]

本节主要讨论适合数字技术的版权方面的最新进展，并阐述有关篡改水印系统或反向设计水印系统对版权意味着什么样的问题。水印，作为在数字作品使用过程中跟踪读者的很好的方法，常常被当作是解决多媒体版权保护问题的灵丹妙药。正如在本书其它章节所看到的那样，对这项技术的研究是非常深入的而且硕果累累。然而，作为一个有效的数字水印系统，其标准化、互操作性问题的有效研究、最基本需求集合的定义等等还远没有达到人们的期望值[2]。

水印可以用作数字版权保护的工具。正如前面各章中所看到的那样，各种各样的场合适合于水印发挥其作用。运用可视水印对图像进行版权保护就是一个例子，这样一种标记在表面上就能看得见，但不妨碍该幅图像的正常使用。这种可视性主要用来使图像的任何非法商业意图从表面上就很容易被发现，因而可以协助法律部门进行版权执法。这种水印系统也可用来标识作品的原始所有者。最典型的情形是，水印可用来防止和检测非授权的数字作品的复制和销售(发布)。如果网上在线发布的数据或文档嵌入了水印，那么随后可以利用一个Web爬虫去搜寻那些非授权的发行品(见7.3.2小节)。数字水印是一门较新的学科，并且该领域的研究工作一直在向前推进，人们希望找到一个真正健壮的、盲的、对图像透明的、公开的水印算法，即，能对付第七章所列举的各种攻击的算法。目前水印这项技术仍然有许多局限性，比如，目前仍不清楚对给定大小的一段信息到底能嵌入多少比特的水印。

本节准备在版权方案以及读者私有性之类的更大范围里研究将普通水印系统用作“技术性保护系统”的一些条约。沿着先驱者的足迹来讨论最新的版权技术手段也许是很有帮助的，而这些版权技术影响着对网络在线环境里数字作品发布的管理。

9.1.1　WIPO条约与WIPO的数字议程

1996年，世界知识产权组织(WIPO)召集了一个外交性的国际会议[3]，其目的在于修订用来保护文学和艺术作品的伯尔尼协定，(这是第一份国际性的关于版权的协定，缔结于1886年。同时成立了一个保护文学和艺术作品创作者的所谓的伯尔尼联盟。该协定在后来一系列的外交会议上修订过，最具影响的要算1971年在巴黎会议上的修订工作(也是从那时起称作“伯尔尼协定”))，使其涵盖数字作品的使用与发布。这次会议上通过了两个协定：WIPO版权

[1] 9.1节的一部分作为评论发表在《European Intellectual Property Review》1999年第171期上，9.2节的一部分发表在[1]中，由伦敦CCH出版有限公司出版。

[2] 本节的主要内容来自文献[2]，它涵盖了水印技术的各个层面。

条约(WCT)和 WIPO 履行和解释条约。前者与现在的讨论直接相关,很显然,WCT 条约第一款与伯尔尼协定的关系是互斥的。

WCT 条约第一款第一条规定:"这个条约是一个在保护文学和艺术作品的伯尔尼协定的第二十款的范围内的特殊协议,也被看作是规范由协定所建立的联盟各成员国版权保护行为的合同"。伯尔尼协定的第二十款包含下列规定:"联盟成员国的政府有权在协定框架内签定自己的特殊协议,只要这种协议能确保创作者享有比该协定所赋予的更广泛的权利,或者包含的规定不违背这个协定即可"。WCT 条约第一款第一条规定:"禁止任何可能会导致比伯尔尼协定所准予的保护水平更低的 WCT 条约的解释与阐述"。WCT 条约第一款第一条的第二句涉及了 WCT 条约与伯尔尼协定之外的条约之间的关系,并声明"除了伯尔尼协定,这个条约不应该与任何其它的条约有牵连,也不要对其它条约下的权利与义务产生偏见"。WCT 条约与伯尔尼协定之间的关系显然是互斥的。

WCT 条约中与"数字议程"相关的规定涵盖下列问题:可适用于数字环境中作品的存储与传输的权利[第 8 款第 4 条]、数字环境中对权利的限制与例外[第 10 款第 4 条]、版权保护的技术措施[第 11 款第 4 条]、版权管理信息[第 12 款第 4 条]。虽然 WCT 条约的草案包含了某些规定,这些规定阐述了对存储在电子媒介里数字形式作品的复制权利的申请问题,但它们没有包含到条约的最终版本里去。但是,这个外交性会议还是采纳了如下的一个协议声明:

正如伯尔尼协定第 9 款所宣布的,复制的权利和有依据的例外许可,尤其是数字形式的作品的使用,在数字环境中必须完全申请。由此可推得:一幅受保护的数字形式的作品存储在电子媒介中,那么,在伯尔尼协定第 9 款的意义下就构成一个复制。

早在 1982 年 6 月,一个由管理专家组成的 WIPO/UNESCO 委员会就阐述过:作品存储在电子媒介中是复制,并且 WIPO 正式宣布会毫无顾虑地关注这条原则 [5,p.6],[6,pp.49 - 51]。

9.1.2　技术性的版权保护系统、版权管理信息和它们的欺骗性

在推进 WCT 条约获得通过的准备工作中,大家都认识到,仅仅规定关于数字作品使用的适当的权利是不够的,尤其在因特网上。在这样的环境中,如果没有技术保护手段的支持和版权管理,那么没有任何版权会得到有效的行使。这些问题在 WCT 条约的第 11 和 12 款陈述如下:

第 11 条款

在 WCT 条约的第 11 条款里,要求各签约成员国或组织提供"适当的法律保护和有效的法律修补,来对付创作者在该条约或伯尔尼协定框架下行使他们的权利时所使用的有效的技术手段所造成的欺骗,并且限制那些没有被作者授权的或者法律允许的一些行为"。仔细推敲,"反欺骗"这个条款就术语本身来说,其意义是很宽广的,它的语言暴露了很多当初制定条款时的争论[7]。概要地讲,过去,围绕反欺骗立法的争论主要集中在① 设备的覆盖面,尤其是重用技术;② 禁止软件反向设计这样的行为,在处理版权异议时,要额外合法地发挥它们的作用,比如公平使用条款。从效果来讲,与后者相比,前者关注的是反欺骗条款要在版权法之外制定例外版权[3]。

[3] 关于对欺骗的"行为"和"设施"之间各种禁止的模棱两可和有争议的讨论可见[8,pp.197 - 207]。

WCT 条约的第 11 条款是对过去草案的改进。首先,它应用于"带侵犯目的"的欺骗行为,而不是把法律责任定在设备的生产与销售上,在这些情形里,很少或无法预料一个设备能够用于击败版权保护。其次,对"设备"这个词根本没有参考解释,就像美国立法执行 WCT 条约第 11 条款时所面临的境地[9,S.30]。但对欺骗的"衡量标准"和"法令"却有参考解释,它们可能潜在地将重用技术排除在外[7,p.235][10,第 6(1)款第 30 段]。最后,第 11 条款所涉及的"欺骗"仅适用于那些"版权所有者在法律关注或允许的范围内没有授权"的行为。在这个条文意义下,仍然未涉及到版权例外和免除,虽然可以肯定第 11 条款的适用范围仅相应于版权侵犯[6,p.48]。第 11 条款回避了任何"有效的技术措施"之类的定义,也许考虑到与其早期草案的抵触[11,草案条款 13(3)][6,pp.45 – 47]。

第 12 条款

条款 12(1)要求各签约成员国"提供适当的和有效的法律修补,或者关于有合理基础的人权法修补,以对付任何人故意利用法律知识在这个条约或伯尔尼协定范围内进行引诱、使能、方便地隐匿任何版权的侵犯,包括① 在没有授权情况下,移去或更改任何电子版权管理信息;② 在知道电子版权管理信息无授权地被移去或修改的情况下,没有授权地对公众销售、引进销售、广播或传播数字作品或作品的拷贝"。条款 12(2)定义"版权管理信息"是"当这些信息条目的任何一个附在一个作品的拷贝上或者出现在作品对公众的通信连接上时,用于识别作品、作品的作者、作品中任何权利的所有者的信息,或者是关于作品的使用期限和条件的信息,以及表示这种信息的任何数字或代码。"

外交会议采用了下面有关 WCT 条约的条款 12 的协议声明:

可以理解成有关'在这个条约或伯尔尼协定范围内的任何侵权'还包括专有权利和报酬权利。进一步可以理解成各签约成员国将不依赖这个条款来设计或实现具有强烈形式化影响的版权管理系统,它们在这个条约或伯尔尼协定范围内也是不允许的,在这个条约下,禁止商品的免费流动或者禁止版权的免费分享。

这个协议声明明显限制了为加强任何"比伯尔尼协定更强的"权利和形式而实现的版权管理系统。

在最后正式条约里,条款 12 是以前草案的修改[11,草案条款 14(1)],其中要求某些与侵权目的相关联[4,条款 12(1)]。它不仅要求了解一个禁止行为的性能,而且还要了解这种行为的实际的或构造性知识,即,为什么这种行为使版权侵犯很容易。另外,只有当发布者实际知道版权管理信息被移去的情况下,才追究被修改作品的发布责任。对水印目的特别重要的是"版权管理信息的定义",尤其是涉及到用作所有权和身份鉴别的符号的水印。放宽条款 12(1)的范围到包括控制作品使用的期限和条件、打击篡改在线许可证。同样,条款 12(2)中对"表示这种信息的任何数字或代码"的参考说明涉及到了水印算法。

9.1.3 水印系统的法律保护

从 WCT 条约的条款 11 和 12 可以很清楚地看出,版权管理系统在未来版权执法过程中将扮演一个至关重要的角色。在该系统中,由于水印是构成版权管理信息的一部分(正如 WCT 条约第 12 款第 2 条所定义的),所以它要受到保护,以免遭在数字形式作品的分发过程中任何非授权的删除、修改、扩散和植入。这主要是针对"水印可逆转性"问题的。

如果授权的用户可以将一个水印从文档中去除，则说该水印是可逆转的。在某些情形，这也可能是人们所期待的性质，因为它可以使所有者具有根据其历史来改变一个文档身份的灵活性。反过来，如果又要寻求防篡改性，则水印的可逆转性就很难满足此要求了[2，pp.7－8]。于是，水印的可逆转性越大，则调整 WCT 条约第 12 款的重要性也就越大。

涉及 WCT 条约条款 11 的关于水印的地位问题并不清楚。条约中在没有任何定义的情况下，还不能确定水印是否能成为"有效的技术手段"。欧洲议会和指导委员会关于信息社会中版权与相关权利的某些方面的协调的最新建议的第 6 款第 2 条([7，pp.23]和[10，recital 30，Article 6(1)])提供了下列定义：

术语'技术手段'[...]具体表示为一个处理过程、设备或产品的任何设备、产品或部件，而这些过程、设备或产品是被设计用来防止或抑制任何对版权的侵犯，或者用来防止或抑制任何对版权或与版权相关的任何权利的侵犯[...]。技术手段只有在下列情况下才被认为是有效的：只有通过一个接入码或接入过程，作品或其它主题才可接入用户，包括作品或其它主题的解密、搅乱的还原或其它变换，都要获得版权所有者的授权。

如果水印系统所履行的功能是读者跟踪或作者/所有者的识别，在这个意义下，它不满足上述定义。尽管这样，在第 12 款下所获得的解决任何水印系统的篡改问题的权利还是足够宽阔的。

9.1.4　水印的互操作性

作为计算机软件的问题，互操作性超越了水印的范畴，当版权保护是在各种不同的公开环境中贯彻执行时，不同水印技术之间的兼容性是至关重要的。关于水印的许可报告 [2]已经声明：不同水印系统的兼容性可以在两个基本层面上加以解决，即内在的兼容性和格式上的兼容性。根据该报告，当提到内在的兼容性时，两个不同的水印算法能够在同一环境中并存，当且仅当下列条件满足时：① 他们都允许一个事先确定的最小数量的比特插入数据文件中，使得所有有关的水印都表达相同的信息。② 水印算法必须是同类的，也就是对称钥算法与公钥算法不能一起使用，盲的/非盲的(可读的/可检测的)技术也不能混在一起使用。而格式上的兼容性，简单地说，是指不同的水印格式是否表示同一种信息。

这又提出一个问题，是否允许逆转一个植入水印算法过程(或者通过分解或拆卸)来达到在特定环境中保护它与另一种格式的兼容性的目的。在欧洲范围内，假设有疑问的水印算法被赋予"计算机软件"资格，并满足版权的意图和最低需求条件，则指导委员会，也即指"软件指导委员会"[12]，可以开始发挥作用，一个微代码是否能够保护版权，更准确地说，水印算法是否包含授权信息[13]。在实践中，一个水印算法毫无疑问应该当作软件，因为这种算法包含着编码[2，p.13]，并且可以看作是使一台机器执行特定功能的指令集合。

软件指导委员会的调整机制对软件的可分解性规定了严格的条件[14－18]。至关重要的是，可分解性必须是"为获得必要的信息来实现[...]互操作性所必不可少的"，并履行条款 6(1)(a)－(c)中的条件。可互操作的开发者必须是获得许可的人或者是一个授权的用户[12，条款 6(1)(a)]，不可以反编译他事先知道的接口信息[12，条款 6(1)(b)]，最后，必须禁止他仅对目标程序中那些包含接口信息的部分进行反编译，因为他需要这些接口信息来开发一个独立制作的、可互操作的程序[12，条款 6(1)(c)]。除此之外，还规定了使用反汇编结果的另外三个条件[12，条款 6(2)]。最后一个限制条件是防止任何通过反编译获得的接口信息被用于开

发一个侵犯版权的程序[12,条款6(2)(c)]和其它的事情。

在进行水印互操作性研究的可互操作的开发者们应该记住上述关于逆转水印工程的各种限制条件。这些限制条件也同样适用于那些水印可逆转性的研究者们,他们从事于逆转水印处理的工程[2,p.7]。

9.1.5 对读者隐私的更广阔思考

关于水印的使用与开发方面的版权讨论应该进一步考虑利用水印技术关注的隐私问题以使其更完美。作者们([19,pp.17-21, pp.31-35]、[20,pp.228-238,pp.241]、[21,pp.140-141]、[22,p.34,pp36-37])已经描述了理想的电子版权管理系统的性能应该如下:

· 能检测、防止和审计各种类型的操作,包括:文件打开、打印、输出、拷贝、修改、摘录引用等等;

· 保存和维护各种记录,这些记录显示着哪些操作与权利是确实经授权许可的,以及授权给了谁;

· 捕获某一用户实际上看过、拷贝或打印了些什么的记录;

· 当用户寻求更多的作品接入、或在帐单和支票的最后期限、或每当用户用完存款的时候,把这些有用的记录发送到清算中心。

在管理和控制那些本质上带有侵犯性的技术时,上述属性是必然会出现的,尤其是通过使用诸如水印技术的那些场合。一个系统操作员能够产生一个特定消费群的预测性轮廓描述,这带来一个很大的争议,即读者读、听、看那些匿名选择的材料的自由是应该受美国第一宪法修正案保护的权利([23,p.184]和[24,pp.985-6,1003-30])。为了使管理系统能够保护隐私的唯一可能的办法就是防止个人数据的提取与处理,也就是,为了"保护个人的隐私,有用的数据在它们达到授权的持有者之前就被汇聚或使其匿名"[22,p.36]。人们也已经注意到,在数字时代保护个人隐私的最有效的办法是设计一些技术工具使得它们能够防止或限制个人的识别[25,pp.181-83]或者接受匿名的电子现金[26,pp.415-20,459-70]。

水印系统的开发者应该留意EC(欧共体)通用的隐私法律[27](与美国采用的做法相反,在美国主张隐私法律的努力是非常细琐的。请看例子[28]中专门关注录象带租用记录的隐私),这些法律也许对版权管理系统形成了一个潜在的障碍,而这些系统往往被认为做得太威胁用户的隐私了,比如泄露关于用户太多的信息。欧洲IMPRIMATUR[4]计划最终决定,在开发一个标准化的版权管理模型的过程中,读者隐私应该被看作是一项基本权利,并且在所有的模型中都应该使用"隐私增强"技术[29,pp.86-90]。在最新提出的关于"在信息社会中版权与相关权利"的指导性议案(COM(97) 628最终版)中,委员会在第33条叙述段中承认:各种版权管理信息的技术可以揉合在一起,以处理关于个人使用版权作品的私人数据,也可以用于跟踪在线行为。为了遵照数据保护指导性文件(95/46 EEC),这种技术必须与隐私安全措施结合在一起使用。

同时,为有效地进行版权执法而使用数字水印时,还需要一些负面影响的考虑。争论较多的一点就是:为了公众利益,不要顾虑那些"不太过分的侵犯"行为以及不太重要的隐私争议。应该允许跟踪读者,因而某些版权执法所必需的信息可以为执法目的而进行析取和管理。

[4] "Intellectual Multimedia Property Rights Model and Terminology for Universal Reference"

9.1.6　结论

本节希望在最近的 WIPO 草案里找到证据，说明水印作为一项技术是受版权法保护的。在这个学科里，还有许多方面的内容需要更深入的研究，并且在软件指导委员会里失去反向设计规定的目标的危险性也已经提高了，关注隐私的程度也加强了。

随着时间的流逝，我们将能看到数字水印和数字版权管理是否能够结合在一起，并且结合到什么样的程度。

9.2　因特网版权法之间的相互抵触

这一节的目标是考察专门针对民事侵权法律间的相互抵触所制定的新的法令准则的适用性。1995 年颁布的隐私国际法(内含形形色色的规定)是这样的一部法案，它规定如何改进那些在大不列颠联合王国范围内与法律相抵触的部分条例。它的第三部分是专门针对民事侵权行为的，后来，针对法律行为的一个特定情形，即那些暗示着对因特网上获得的材料进行版权侵犯的行为，“1995 年法案”也从信息技术和知识产权这两个方面作了权衡考虑。大家公认：尽管针对不同国家之间法律标准的协调性问题的传统解决方法，起初来源于针对传统的文学或艺术作品的伯尔尼协定(见 9.1.1 小节)，根据国家之间的条约原则，应该采用作者所在国家的法律还是作品创作地的法律等方面也存在协调性问题。如何解决这些问题将随着情况而变化。不同法律系统之间的模糊越来越多，因而必然要求提出更加清晰明了的解决方法。比如，为保持其作为一个通用法则，作品创作地的法律应该起着决定性的作用。还应考虑对侵犯的适用性，正如，对私有财产的侵犯，涉及到犯罪发生地的地方法律，或者在与民事侵权行为法律相抵触的有关规则下适用的任何其它的法律。

这个问题的考察需要分阶段来进行。从有关 1995 年法案引入后民事侵权行为的法律规则所产生的新抵触的分析开始，尤其是关于它的最吸引人的特性之一到某些例外的灵活条款的规定，这些例外规定，在通用法则下应该运用一个比犯罪发生地的地方法律更“充分合适些”的法律。考虑到这点，经过最近英国学者 Morris 博士的努力，在新法律下被重视的排在后面的选项与一个学术性的研究成果之间作过一个比较分析，也就是大家所共知的“一个民事侵权行为的合适的法律”[30]。

在下一节里我们将评估针对版权侵犯案例的法则的应用价值。这意味着要面对许多原则性问题，侵犯版权可以被看作是一种民事侵权行为吗？或者，相应地，针对民事侵权行为的法律规则的相互抵触同样适用于版权侵犯这种案例吗？从一种特定的角度来决策哪一种版权法将应用于身边这些事件？在努力回答这些问题之后，将要考虑“一个民事侵权行为的合适的法律”方法是否可以适用于因特网上发生的行为，这是否能有效地帮助解决各种问题，而这些问题按诸如犯罪发生地的地方法律这样较严格的标准仍然解决不了。

最后，虽然这篇评述的两步法被正式地切成两部分，一边是法律的相互抵触，另一边是版权保护，但对某些并行的和严格相关的问题的参考是不可以省略的，尤其要参考个人数据的保护以及与加密技术相关的法律问题。

9.2.1 针对英国民事侵权法之间相互抵触的新的准则

1995 年法案的第三部分引入了对英国的民事侵权法之间的相互抵触具有重要意义的两个变化。前者一直存在于一个主题科目的法典编纂条目之中，至今仍游离于通用法律范围之外；而后者处于被最主要的通用法律规则抛弃的地位。而这些通用法律规则不仅控制着法律的选择，而且还控制着英国法院在对民事侵权行为起诉宣判的权限。

• 废除双重起诉准则

双重起诉准则可以看作是一个双重机制，它允许一个在国外犯的所谓的民事侵权行为到英国法院接受审判，并且在犯罪发生地的法律和英国的法律下都要被认定为民事侵权，即在犯罪发生地的地方法律与法案提出的法院所在的国家法律下对民事侵权行为起诉。过去的一个判例，最近在 Boys v. Chaplin [31]有更详细的描述，它是一个与应用于追究民事侵权责任的法律有关的案例。该案例是由一起发生在英国边界的 Malta 的交通事故引出来的。正如后面所解释的，按较严格的案例自然环境，在那里应用了英国法律(这个案例所依据的准则显然导致了对原告机会的限制，因为原告为了满足如此严格的要求，不得不面临着“攀登高峰的努力”[32,p.650])。

更近的红海保险公司与 Bouygues[33 - 35]的案例是沙特阿拉伯的一项建筑工程的保险案例。在此案中，保险公司寻求依照一个起诉的直接理由，该起诉在沙特阿拉伯本地的法律下是允许的，而在香港的法律(假定与英国法律一致)下是不允许的，并且依据该起诉的特殊事实，它被允许作为双重起诉准则的例外，这是缓解双重起诉准则所走的第一步。此案进行到了这种程度，即枢密院的裁决委员会最终同意驳回了双重起诉的要求。显然，这样一个革命性的结论需要一些说明。首先，记住在英国法院枢密院(法律上听讼的最后场合，完全象英国上议院的司法委员会，处理的案例包括从来自英国或英联邦国家的最高法院的案例，到那些放弃其权限的例外，比如最近加拿大和澳大利亚。)的裁决不是原则约束先例的事件，相反地，枢密院本身在裁决时并不受任何最近的权限限制。其次，按 Lord Slynn 的观点，从判决的价值来看，双重起诉准则实际上并不是完全被驳回，而是基于该案例的事实把它作为一个例外。

实际上，到上述案例判决时，英格兰和苏格兰的法律委员会已经考虑了一个有关该主题科目的立法改革的建议，并列举了双重起诉的旧准则将被正式毁除的理由。这项方案导致了隐私国际法律议案的诞生。其后不久，这项法律议案使这个主题科目得到英国上议院特殊公共议案委员会充分的研究[36]。

在 1995 年法案[37]的第三部分的最终文本里，该主题科目新的法令在第 9 节里首先提供了一种预备性描述，紧接着提供了一系列有关第三部分应用意义和范围的说明。事实上，1995 年法案的第 9 节专门解决隐私国际法的某些关键问题，比如，处于民事侵权行为或犯罪嫌疑中的问题的特征化。而这些问题将由争议庭来处理 [38,pp.893 - 894][39,40]。后一个问题在稍后的段落中在这样的意义下被解决，即适用的法律也将适用于是否已经发生起诉民事侵权和犯罪的初步问题，并且，为避免疑虑，第三部分规定被认为同样适用于法庭所在地发生的事件。也就是，在权限内[37,Sec.9,par.7]，它们适用于国外发生的事件[37,Sec.9,par.4,6][41][42,pp.521 - 523] [43]。

然后第 10 节规定：将废除普通法律的强化性准则。而这些强化性准则使双重起诉的要求得到满足，或规定为控制在给定案例中引出来的民事侵权行为或犯罪问题的单一法律的一个

例外，继而，它们不适用于 1995 年法案第三部分实施范围之外的诽谤事件[37,Sec.10]。

最终达成 1995 年法案采用的解决方案，第 11 小节采用的通用准则包括犯罪发生地的地方法律的适用性，通常它们保留在基于一个严格的地域准则的外国立法。但是，经验已经建议我们在采用这个准则的法律系统内附加一系列说明，使其达到这样一种程度，即犯罪发生地的地方法律将作为构成民事侵权或犯罪嫌疑的事件已经发生的所在地方的法律。进一步的规定是关于那些对个人所有权的伤害或损害的起诉，在这些案例中，将应用受害人或所有权正在发生伤害或损害的时刻所在地方的法律，而关于那些仍处于灰色地带的案例，已经设计另一种不同的解决方案，规定使用另一种法律。嫌疑事件的最重要的元素或元素组发生的国家法律的应用，为通用法则的例外铺平了道路，它们在第 12 小节中介绍[37,sec.11],[44]。

• 更充分合适的法律的应用

特别地，1995 年法案的第 12 小节提出要打破两方面因素之间关系的平衡。一方面的因素是将民事侵权或犯罪与通用法则下适用的法律联结在一起。另一方面的因素是将民事侵权或犯罪与特定案例相关的任何其它法律联结在一起。考虑到案例中的效果，即这种平衡向作为更充分合适地管制该案例事件的法律倾斜，然后它将应用于犯罪发生地的地方法律的例外情形。

这个解决方案实际上并不是由 1995 年法案突然提出来的。该解决方案的目标是作为一个适用法律的确定性程度的法令，当然这种确定性程度在普通法律准则里是不会给出的。事实上，以这种条件构思的宽阔的思路在普通法律里已经保留下来了，在大家熟悉的 Boys v. Chaplin[31]引用的案例中也存在这种思路，并且在意味深长的“民事侵权行为的合适的法律”的旗帜下，早些时候在英国法律相互抵触中最有名的诽谤之一的学术性研究里也有这些构思[30,45][46,pp.62～71]。

从稍后开始，独创的和有深刻见解的探索可追踪到五十年代早期。“合适的民事侵权法律”的准许代表着两类相互抵触的法则的汇合，其中一种与合同法相互抵触，另一种与民事侵权法相互抵触。该法律已经作为原则保留在英国或美国的法律系统里。事实上，这个理论的形成首先承认在英国法律范围内存在关于合同的合适法律的定义。反过来，美国法律系统由于缺乏这种定义而遭到很多抱怨。正如 1934 年的 Restatement of Conflict of Laws 法律相互抵触的再声明的第 332 段落显示了在特定领域中美国法律准则的非法律性的强化。

在这个前提下，Morris 博士考虑了在民事侵权领域也引人一个“合适的法律”技术的观点。如果不按照样板策略(审判先例的权威性)，而按照被选作可用法律的弹性和最终合适性，这些考虑会带来好处。在英格兰或苏格兰在这方面的法律到目前为止都缺乏任何权威性的情况下，美国判例法在制定法律参考方面已取得某些突破，不需要受民事侵权法律选择的通用准则的约束，正如法律相互抵触的再声明所表达的那样，是一种使用错误的地方法律的感觉。这里错误的地方意指引出责任的最后事件的发生地[47,Par.377]。

在这方面，虽然美国关于民事侵权法相互抵触的准则缺乏他们在合同法领域所享有的弹性，但 Morris 博士很清楚地指出了许多典型的情形。考虑到转化过来的侵权(比如，不恰当的合适性)，美国法院采用了一个更具弹性的意见，并在一些案例中也采用了。在这些案例中伤害行为发生地的法律与造成损害的地方的法律是不一样的。在这些案例的一个案例中，民事侵权的适当的法律学说已被上议院在 Boys v. Chaplin[31]里引入到英国法律相互抵触中。在那里，由一个发生在涉及英国边界的马耳他的交通事故引发的民事侵权责任的广度要扩大到

英国法律体系中去，即要接受英国法律的审判，以更接近案例本身的事实。

最后 1995 年法案将民事侵权的合适的法律自动纳入英国法律的相互抵触中。这样就将它从双重起诉的要求中解放出来。英国上议院在 Boys v. Chaplin 案例中作了一个示范。在以下小节中，从最近判例法的进展来看，我们将考虑这种法律相互抵触的方法是否或到什么程度可以应用于因特网上版权侵害之类的问题。

9.2.2 信息技术和知识产权方面

我们对因特网的引用需要一些初步的说明。这里的因特网可以被认为仅仅是一种媒介，更精确地说可以被看作是一个传输、存储和交换信息的系统。正是从这个视角来看，因特网已经成为一个知识产权的虚拟战场，这个战场上带来的问题牵扯到了其它的法律问题，特别是法律之间的相互抵触，民事侵权纠纷等问题。这就要求改变当前对知识产权和其它法律单独分析的做法，要在它们之间建立一个不可缺少的进行调和的接口，尽管主要用于国内法律系统的知识产权法是重大的不可抗拒的法律融合进程中的一部分，但是如果没有提供可应用法律的并行系统的支持，它就不会很好地存活下来。因此，无论不同法律的融合和统一进程在向最终不可达到的绝对统一的道路上走多远，知识产权都应该扮演而且将继续扮演不可缺少的角色[48－51]。

• 作为信息交换手段的因特网和法律保护需求

从法律保护的角度来看，因特网上进行的活动应该受到许多种法律的保护。要严格地受知识产权法的保护、也要受用于保护硬件的专利法的保护(而不是通常想象的那样更适合版权保护的保护软件的专利法)[6,52]和用于保护域名的商标法的保护[53－56]、最基本地也要受版权法的保护、还要受处理民事侵权事件的民法的保护、处理不公平竞争的其它国内法律的保护等等。这里还没有提到处理由公共政策或道德因素带来的问题的法律[57,58]。

这些名目繁多的法律问题是对不同国家不同立法的直接的反映。相应地，我们这里只考虑与因特网上信息处理直接相关的法律问题。对由于处理了带有版权色彩的信息而产生民事侵权行为，并进而承担非契约民事补偿责任的案例的引用很有可能会被认为是民事侵权行为，而不论你的行为是多么地符合公共法[59,p.43][60]所定义的特殊特征。

实际上，真正的争论点不在于将版权侵犯看作是对违反相关规则的民事侵权行为，而在于这种行为的侵犯程度。因此，一个恰当的推理途径就是要在提交到(大不列颠)联合王国的所有法案中做出区分，将这些法案区分为与国内知识产权立法相关的法案和与国外知识产权立法相关的法案。在与国外知识产权立法相关的法案中，国际惯例的基本角色就是要统一受保护条目创作地或注册地的法律的应用和犯罪发生地的地方法律的应用之间的地位。而这个过程中，主要是后者要充分地采纳不同的国家处理界定的边界原则。最终，1995 年法案废除了(大不列颠)联合王国中双重起诉的章程，双重起诉是处理境外知识产权侵犯行为的真正障碍。一系列的原因推动了法律之间的通道向更广泛的方面发展。在 1995 年的法案[37,S.9,pp.4－6],[61－64],[65,pp.504－513]中就允许在恰当的时候，英国法院可以应用境外知识产权法。

• 最近判例法中因特网上的版权和法律之间的抵触

到目前为止，我们已经建立了知识产权和法律抵触之间有难度的相互融合的基本框架，考察一下它在判例法中的实际运作情况就显得十分必要。从这个角度来看，以寻求作为信息流

通手段的因特网的使用和无论是在境外还是在境内的版权侵犯之间联系的案例为起点，融合的进程已经有了进步。在本节中，我们首先要考察的就是发生在大西洋彼岸的那些案例。自从 1993 年以来在那里有很多重要的案例产生，其中最著名的就是各种花花公子案例[66－68]。还有 SEGA 有限公司与 Maphia [69]之间的案例、美国与 La Macchia [70]之间的案例、宗教技术中心与 Netcom 在线之间的案例[71]、Frank Music 与 Compuserve[72]之间的案例、还有澳大利亚的 Trumpet 软件公司与 OzEmail[73,74]之间的案例等等。

所有的这些案例都可以一起来考虑，因为它们具有共同的特点，即都具有版权侵犯行为。这些版权侵犯行为通常在于利用 Web 上的资料(杂志的照片、计算机游戏、软件程序或者文本和歌曲来重建自己的“公告牌”、允许普通用户公开访问和下载)，从而侵犯了资料原作者的版权。在所有这些案件的审理过程中，法院最后发现对于版权侵犯来说，尽管在同一情形下，美国的版权法和(大不列颠)联合王国的版权法乃至通用的欧洲法律都不完全一致。在这种意义上，美国的版权法认为只要将计算机系统中的信息以重用的形式上载了，就构成了版权侵犯，而其它的法律系统，特别是日本和澳大利亚的法律则认为随后的信息发布和传输过程才构成版权侵犯，也就是说它将版权侵犯与通信联系起来。美国法律和欧洲法律进一步的差别产生在将信息显示的权利问题上。在花花公子[66]的案例中，美国法院就认为将信息显示出来就是侵犯了版权，而这却完全超出了欧洲公法系统的限制范围。还有就是对像仅仅“浏览”版权保护的资料[75,pp.286－288]，[76,pp.552－556]这样的行为的合法性界定也还不确定。

到现在为止，类似的案例也在欧洲的法庭上出现了，如，在苏格兰的设得兰群岛时报有限公司与 Jonathan Wills 博士之间的庭审案例。该案例的焦点在于是否授权原告(“设得兰群岛时报”的所有者和出版者)终止诉讼进程的权力，以制止被告(“设得兰群岛时报”的新闻报告服务提供商)在他们的大标题新闻中包括“设得兰群岛时报”中出现的类似的条目。下面两个基本的问题是争论的焦点：第一个问题是依据 1988 年版权、设计和专利法案(一个为合并(大不列颠)联合王国内相关的知识产权立法而制定规范的法案的第 7 部分)，被告在他们的网站上发布大标题新闻是否构成“电缆程序服务”的一种形式。这个法案最显著的特点是排除了注册商标，在下面的章节中我们将用“1988 年”法案来指代这一法案。另一个问题是如果大标题新闻不经意中构成了文学作品，而这正触犯受 1988 年法案中第 3 部分保护的版权，那么，按照 1988 年法案的第 20 部分，包含存在问题的这些服务项目是否构成版权侵犯。就这些问题的逻辑关系来说，Lord Hamilton 认为尽管文学特点不是文学作品的必要因素，但是原告已经通过在电缆程序服务中包含恰当的授权禁止信息，给出了初步的版权侵犯说明，然而，被告在辩解中指出的服务的交互特性在起诉阶段看起来也仅仅是偶然的。最终，尽管很不幸的是庭外调解阻碍了最终审判的进一步发展[36,44]，但是，在起诉阶段法庭所采纳的意见却向许多作家认为可行的方向铺平了道路，那就是用为先前开发的通讯方式制定的框架来监控因特网 [76,pp.548－550] [65,pp.495－496] [75,pp.288－293]。

还有一些欧洲案例具有同“设得兰群岛时报”案例相似的情况，其中最著名的就是比利时的比利时职业记者协会与 SCRL Center Station 之间的案例[79]和德国的 Re Copyright in Newspaper Articles Offered On-Line 案例[80]，以及随后在法国发生的两个 Queneau 案例[81－84]。这些问题都带来了类似的在因特网站点上重用文学作品(新闻文章的一部分或一首诗的一部分)的问题。在不同的案例环境中用不同的术语，相应的法庭最终都支持原告关于在受影响的条款中存在版权声明和按照相应的国内法律在 Web 站点上重用文学作品要注明版权的论点，从而

承认了处理由具有文学特点或侵犯特点的重用片段带来保护知识产权的问题，也承认了由于放弃使用私用信息并且像在第二个 Queneau 案例（在这个案例中一个安全泄露使得一个内联网站点可以被公众公开访问）中那样造成意外泄露而产生的知识产权问题的流行做法[84]。

一些关于别的方面的案例可能也值得考虑。例如 Weber 与 Jolly Hotels 之间的案例[85,86]。在这个案例中，中心问题在于在一个被动的 Web 站点上提供信息（也是在这个基础上，共享合同）。这个简单的事实能否引起被告注意原告的论坛，注意到原告的法律应用场合，而在这种场合中被告的法庭在参与者推理论坛通用准则（被告所在国家的法律的权限）下不设置私人权限，从面从应用法律的角度就产生了版权纠纷的后果。面对高级权限的要求（这可能导致在因特网上提供信息的任何人受控于原告起诉他的任何法院的权限的限制），新泽西区法院审慎地坚持权限的要求可能仅仅是由于对因特网的交互式使用。事实上也就是在因特网上进行的商务活动面产生的。

我们在本章的最后阶段迫切地提出的补充观点是要使用不同的计算机安全技术，特别是加密技术。加密技术是对 Web 页面上或本地计算机系统中的信息以特定的经过精心技术处理的格式存放，只有拥有恰当解密密钥的那些用户才可以访问这些数据。这样的系统在电子商务和电子资金传输[87－90]等关键领域中已经普遍存在。像因特网产生地的军用网络一样，网络在向公众和非政府部门发展的过程中产生的公共政策和公共安全问题绝不比它带来的信息共享自由少[91－93]。

Daniel Bernstein 与美国国务院之间的案例正上演了这一幕[94,95]。在这个案例中，一个在加密程序设计领域中的研究员通过传播他自己开发的叫做“Snuffle”的加密模块面表达学术观点的自由受到了特别敏感的美国国家军事安全规定的限制，最终形成了美国国务院限制生产非军事加密产品出口的规定。尽管在这方面美国规章的严格性可能看起来有些过分，但是加密技术无疑将仍然是法律和政策考虑的一个领域，而且也要考虑加密技术带来的关于隐私保护，特别是私有数据保护的一系列问题[95－97]。

9.2.3 结论

本节可能远远没有达到实际的结论。在本节中我们试图描述更多的问题。在现有的两个与法律相关的问题领域（法律之间的抵触和知识产权）之间架起一座桥梁。知识产权传统地被认为是一个国家法律系统内部的事情，它不再仅仅是学术实践中的一个主题了。不可否认，所有这些情况都会随着周围法律框架的发展而发展，而且也会越来越受社会环境、技术环境和经济环境的影响。

在因特网上适应现存法律系统，融合和集成国内、国际法律系统，建立能一致、同质地解决因特网带来的不可避免的法律问题的法律系统是一个新的、很有刺激的挑战。到目前为止，还不可能在国内法律系统内满意地解决这些问题，因为国内法律系统对法律之间的相互抵触采取的规则不必是一致的，而且国内法律系统还有一些相对独立的法令。最终的目标是建立一个我们提到的意义上恰当的法律，这种法律朝着与不同的国内法律高效、和谐地融合而不发生相互抵触的理想境界发展，一直达到可以适应不同特性的程度。

尽管如此，我们必须承认我们在促成不同知识产权法律融合和潜在统一方面，在不同的国际组织内部法律的实现方面都取得了一些成果。欧洲共同体和更广泛的尽管组织分散的世界知识产权组织在最近取得了非凡成绩，特别是在保护私人数据[27,98－102]和版权法的融合

方面取得了非凡成绩。法律的融合和统一通过互相的交融得到了加强和巩固。最终由欧洲共同体提议在它的法律框架内引入一系列与世界知识产权组织制定的条约[6,10,103,104](参看 9.1 节)相一致的规则来实现法律的完美融合和统一。

参考文献

[1] Marogiu Buonaiuti, F., "The Proper Law of a Tort and the Internet," Amicus Curiae, Journal of the Society for Advanced Legal Studies, London, England: CCH Editions Ltd, vol. 13, Jan. 1999, pp. 28 f.

[2] "General Requirements and Interoperability," Imprimatur Report on the Watermarking Technology for Copyright Protection, IMP/14062/A, 1998.

[3] "WIPO Diplomatic Conference on Certain Copyright and Neighbouring Rights Questions," 1996.

[4] "WIPO Copyright Tready," 1996.

[5] "Document prepared by the Internatioal Bureau, WIPO/INT/SIN/98/9," 1998. Presented at the WIPO Seminar for Asia and the Pacific Region on Internet and the Protection of Intellectual Property Rights, Singapore.

[6] Lai, S., "The Impact of the Recent WIPO Copyright Treaty and Other Initatives on Software Copyright in the United Kingdom," Intellecutual Property Quarterly, 1998, pp. 35.

[7] Vinje, T. C., "The New WIPO Copyright Treaty: A Happy Result in Geneva," European Intellectual Property Review, 1997, pp. 230.

[8] Vinje, T., "Copyright Imperiled," European Intellectual Property Review, 1999, pp. 192.

[9] "U. S. Digital Millennium Copyright Act," 1998.

[10] "Proposal for a European Parliament and Council Directive on the harmonization of certain aspects of copyright and related rights in the Information Society OJ C108/6," 1998.

[11] "Partly Consolidated Text of the Copyright Treaty CRNR/DC/55," 1996.

[12] "Council Directive on the Legal Protection of Computer Programs 91/250/EEC of 14 May 1991 (Software directive)," 1991.

[13] "NEC v. Intel," Federal Supplement, vol. 643, 1986, pp. 590. Northern District of California.

[14] Lehmann, M., and C. Tapper, A Handbook of European Software Law, Oxford, England: Clarendon Press, 1993.

[15] Czarnota, B., and R. J. Hart, The Legal Protection of Computer Programs in Europe-A Guide to the EC directive, London: Butterworth, 1991.

[16] Band, J., and M. Katoh, Interfaces on Trial, Boulder: Westview Press, 1995.

[17] Dreier, T., "The Council Directive of 14 May 1991 on the Legal Protection of Computer Programs," European Intellectual Property Review, vol. 9, 1991, pp. 319.

[18] Huet, J., and J. C. Ginsburg, "Computer Programs in Europe: a Comparative Analysis of the 1991 EC Software Directive," Columbia Journal of Transnational Law, vol. 30, 1992, pp. 327.

[19] "Copyright Management and the NII: Report to the Enabling Technologies Committee of the Association of American Publishers," 1996.

[20] Stefik, M., Internet Dreams: Archetypes, Myths and Metaphors, Cambirdge, Massachusetts, USA:

MIT Press, 1996.

[21] Stefik, M., "Shifting the Possible: How Digital Property Rights Challenge Us to Rethink Digital Publishing," Berkeley Technology Law Journal, vol. 12, 1997, pp. 138.

[22] Smith, and Webber, "A New Set of Rules for Information Commerce-Rights-Protection Technolgogies and Personalized Information Commerce Will Affect All Knowledge workers," Commercial Week, 6 Nov. 1995.

[23] Cohen, J., "Some Reflections on Copyright Management Systems and Laws Designed to Protect Them," Berkeley Technology Law Journal, vol. 12, 1997, pp. 161.

[24] Cohen, J., "A Right to Read Anonymously: A Closer Look at Copyright Management in Cyberspace," Connecticut Law Review, vol. 28, 1996, pp. 981.

[25] Glancy, D., "Privacy and Intelligent Transportation Technology," Santa Clara Computer & High Technologies Law Journal, 1995, pp. 151.

[26] Froomkin, M., "Flood Control on the Information Ocean: Living with Anonymity, Digital Cash and Distributed Databases," Journal of Law and Commerce, vol. 15, 1996, pp. 395.

[27] "EC Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data," 94/46/EC, 1995. Reported in OJL. 281/31.

[28] "The Bork Act 18 USCA," par. 2710.

[29] "First Imprimatur Consensus Forum," 1996, < http://www.imprimatur.alcs.co.uk/ > .

[30] Morris, J. H. C., "The Proper Law of a Tort," Harvard Law Review, vol. 64, 1951, pp. 881.

[31] "Boys v. Chaplin," Law Reports, Appeal Cases, 1971, pp. 356.

[32] Rogerson, P., "Choice of Law in Tort: a Missed Opportunity," International and Comparative Law Quarterly, 1995, pp. 650.

[33] "Red Sea Insurance v. Bouygues," Law Reports, Appeal Cases, 1995, pp. 190. Privy Council.

[34] Dickinson, A., "Further Thoughts on Foreigu Torts: Boys v. Chaplin Explained?" Lloyd's Maritime and Commercial Law Quarterly, 1994, pp. 463.

[35] Rodger, B. J., "Bouygues and Scottish Choice of Law Rules in Delict," Scottish Legal Practice Quarterly, vol. 1, 1995, pp. 58.

[36] "House of Lords, Session 1994 - 95, Private International Law (Miscellaneous Provisions) Bill [H. L.]," Proceedings of the Special Public Bill Committee, with Evidence and the Bill (as Amended), London(HMSO), HL Paper 36, 1995.

[37] "Private International Law(Miscellaneous Provisions) Act 1995, Chap. 42," 1995.

[38] Morse, C. G. J., "Torts in Private International Law: A New Statutory Framework," International and Comparative Law Quarterly, vol. 45, 1996, pp. 888.

[39] Collins, L., "International Between Contract and Tort in the Conflict of Laws," International and Comparative Law Quarterly, vol. 16, 1967, pp. 103.

[40] Collins, L., Interaction Between Contract and Tort in the Conflict of Laws, Oxford, England: Clarendon Press, p. 352, International and Comparative Law Quarterly, 1994.

[41] Anton, A. E., "Loi du Royaume-Uni portant diverses dispositions en matiére de droit international privé," Revue Critique de Droit International Privé, vol. 85, 1996, pp. 267.

[42] Briggs, A., "Choice of Law in Tort and Delict," Lloyd's Maritime and Commercial Law Quarterly, 1995, pp.519.

[43] Carter, P.B., "Choice of Law in Tort: the Role of the Lex Fori," Cambridge Law Journal, 1995, pp.38.

[44] Rodger, B.J., "Ascertaining the Statutory Lex Loci Delicti: Certain Difficulties Under the Private international Law (Miscellaneous Provisions) Act 1995," International and Comparative Law Quarterly, vol.47, 1998, pp.205.

[45] Nygh, P.E., "Some Thoughts on the Porper law of a Tort," International and Comparative Law Quarterly, vol.26, 1977, pp.932.

[46] Morse, C.G.J., Torts in Private International Law, Amsterdam, New York, Oxford: North Holland, 1978.

[47] "U.S. Restatement of Conflict of Laws," 1934.

[48] Cohen, B., "A Proposed Regime for Copyright Protection on the Internet," Brooklin Journal of International Law, vol.22, 1996, pp.401.

[49] Delacourt, J.T., "The International Impact of Internet Regulation," Harvard International Law Journal, vol.38, 1997, pp.207.

[50] Dixon, A.N., and L.C. Self, "Copyright Protection for the Information Superhighway," European Intellectual property Review, 1994, pp.465.

[51] Millé, A., "Copyright in the Cyberspace Era," Europena intellectual Property Review, 1997, pp.570.

[52] Longdin, L., "Technological Cross Dressing and Copyright," New Zealand Law Journal, 1998, pp.149.

[53] "Pitman Training Ltd v. Nominet UK," p.797, Fleet Street Reports, 1997, p.797.

[54] "Avnet Inc. V. Isoact Ltd," p.16, Fleet Street Reports, 1998, p.16.

[55] "Prince Plc. v. Prince Sports Groups Inc." p.21, Fleet Street Reports, 1998, p.21.

[56] "Marks &Spencer Plc. v. One In A Million Ltd," p.265, Fleet Street Reports, 1998, p.265.

[57] "The Free Speech Coalition v. Janet Reno," U.S. Law Week, 1997, pp.1125. Judgment C-97-0281, United States District Court for the Northern District of Caifornia.

[58] Taylor, P., "Note, sub Telecommunications-Internet-Free speech challenge to ban on child pornography," Bulletin of Legal Developments, 1997, pp.221.

[59] Cornish, W.R., Intellectual Property, London, England: Sweet &Maxwell, 3rd ed., 1996.

[60] Trooboff, P.D., "Intellectual Property," in, Transnational Tort Litigation, Oxford, England: Clarendon Press, pp.125-154, 1996.

[61] "Gareth Pearce v. Ove Arup," p.641, Fleet Street Reports, 1997, p.641.

[62] "Coin Controls v. Suzo," p.660, Fleet Street Reports, 1997, p.660.

[63] Cornish, W.R., "Note on Intellectual Property and the Conflict of Laws," in Proceedings of the Special Public Bill Committee, pp.64-65, 1996.

[64] Beatson, J., "Note on Intellectual Property and Part III of the Private International Law (Miscellaneous Provisions) Bill," in Proceedings of the Special Public Bill Committee, pp.61-63, 1995.

[65] Dutson, S., "The Internet, the Conflict of Laws, International Litigation and Intellectual Property: the Implications of the International Scope of the Internet in Intellectual Property Infringements," Journal of Business law, 1997, pp. 495.

[66] "Playboy Enterprises Inc. v. George Frena et al." Federal Supplement, vol. 839, 1993, pp. 1552. District of Florida.

[67] "Playboy v. Chuckleberry Publ. Inc." Federal Supplement, vol. 939, 1996, pp. 1032. Southern District of New York.

[68] "Playboy v. Five Senses Production," Intellectual Property Newsletter, vol. 21, no. 6, 1998, pp. 4. U. S. Federal Court.

[69] "Sega Enterprises Ltd v. Maphia," Federal Supplement, vol. 857, 1994, pp. 679. Northern District of California.

[70] "United States v. La Macchia," Federal Supplement. vol. 871, 1994, pp. 535. District of Massachusetts.

[71] "Religious Technology Centre v. Netcom On-line," judgment C – 95 – 20091, 1995. United States District Court for the Northern District of California.

[72] "Frank Music v. Compuserve," Judgment C – 93 – 8153, 1993. United States District Court for the Southern District of New York.

[73] "Trumpet Software v. OzEmail," Australian Federal Court judgment TG21, 1996.

[74] Richardson, M., "Intellectual Property Protection and the Internet," European Intellectual Property Review, 1996, pp. 669.

[75] Dessemontet, F., "Internet, le droit d'auteur et le droit internatioual privé," Schweizerische Juristsche Zeitung, vol. 92, 1996, pp. 285.

[76] Bariatti, S., "Internet: aspect relatives au conflit de lois," Rivista di Diritto Internazionale privato e Processuale, vol. 33, 1997, pp. 545.

[77] "Shetland Times Ltd v. Dr Jonathan Wills," p. 604, Fleet Street Reports, 1997, p. 604.

[78] Campbell, K. J., "Copyright on the Internet: The View from Shetland," European Intellectual Property Review, 1997, pp. 255.

[79] "Association Générale des Journalistes Professionnels de Belgique v. SCRL Central Station," p. 40, European Commercial Cases, 1998, p. 40. Tribunal de Première Instance, Brussels, 16 Oct. 1996.

[80] "Re Copyright in Newspaper Articles Offered On-Line," p. 20, European Commercial Cases, 1998, p. 20. U 127/95, Oberlandsgericht Düsseldorf, 14 May 1996.

[81] "J.-M. Queneau v. Ch. Leroy," p. 22, European Commercial Cases, 1998, p. 22. Tribunal de Grande Instace de Paris, 5 May 1997.

[82] "J.-M. Queneau c/Boue et LAAS TGI Paris," La Semaine Juridique, vol. II, no. 52, 1997, pp. 569.

[83] Galizzi, P., "Internet-Publication on website of material Protected by copyright," Bulletin of Legal Developments, 23 Feb. 1998, pp. 26.

[84] Olivier, F., "Queneau bis ou les faux-semblants de l'usage privé," la Semaine Juridique, vol. 52,

24 Dec. 1997, pp. 571.

[85] "Weber v. Jolly Hotels," Judgment 96 – 2582, 1996. United States District Court for the District of New Jersey. Full text at < http://lw.bna.com/lw/19971014/962582.htm > .

[86] Taylor, P., "Note, sub The Americas-United States, Civil Procedure-Jurisdiction-Internet," Bulletin of Legal Developments, 1997, pp. 249.

[87] "Commission Greem Paper on Electronic Commerce," COM, p. 157, 1997.

[88] Caprioli, E. A., and R. Sorieul, "Le commerce international électronique: vers l'émergence de règles juridiques transnatonales," Journal du droit international, 1997, pp. 323.

[89] Schneider, M. E., and C. Kuner, "Dispute Resolution in International Eletronic Commerce," Journal of International Arbitration, vol. 14, 1997, pp. 5.

[90] Andersen, M. B., "Electronic Commerce: A Challenge to Private Law?" in Saggi Conferenze e Seminari, vol. 32, Centro di Studi e Ricerche di Diritto Comparato e Straniero, Rome, Italy, 1998.

[91] "WIPO Copyright Treaties Implementation Act," U. S. Federal Bill, H. R. 2281.

[92] "Collections of Information Antipiracy Act," U. S. Federal Bill, H. R. 2652. Available at < http://thomas.loc.gov/ > , discussed in the House of Representatives; comments and reactions from either sides at < http://www.dfc.org/ > and < http://www.cic.org/ > .

[93] Marzano, P., "Sistemi anticopiaggio, tatuaggi elettronici e responsabilità on -line: ildiritto d'autore risponde alle sfide di Internet," Il Diritto di Autore, 1998, pp. 149.

[94] "Daniel Bernstein v. United States Department of State", Judgment C – 95 – 0582 of 25 Aug. 1997, 1995. United States District Court for the Northern District of California.

[95] Movius, D. T., "Bernstein v. United States Department of State: Encryption, Justiciability and the First Amendment," Administrative Law Review (American Bar Association), vol. 49, 1997, pp. 1051.

[96] Kuner, C., "Rechtliche Aspekte der Datenverschlüsselung in Internet," Neue Juristische Wochenschrift-ComputerRecht, 1995, pp. 413.

[97] Bréban, Y., and I. Pottier, "Les décrets et arrêtés cryptologie: la mise en oeuvre effective de l'assouplissement des dispositions antérieures," Gazette du Palais, vol. 118, no, 109 – 111, 1998.

[98] "Directive 97/96/EC concerning the processing of personal data and the protection of privacy in the telecommunications sector, O. J. L – 24/1," 1998.

[99] Lai, S., "Database protection in the United Kingdom-The New Deal and its Effects on Software Protection," European Intellectual Property Review, 1998, pp. 32.

[100] Balz, S. D., and O. Hance, "Privacy and the Internet: Intrusion, Surveillance and Personal Data," International Review of Law Computers& Technology, Vol. 10, 1996, pp. 219.

[101] Chalton, S., "The Effect of the E. C. Database Directive on United Kingdom Copyright Law in Relation to Databases: a Comparison of Features," European Intellectual Property Review, 1997, pp. 278.

[102] Smith, S., "Legal Protetion of Factual Compilations and Databases in England-How Will the Database Directive Change the Law in this Area?" Intellectual Property Quarterly, 1997, pp. 450.

[103] Reinbothe, J., M. Martin-Prat, and S. von Lewinski, "The New WIPO Treaties: A First

Résumé,"European Intellectual Property Review,1997,pp.171.

[104] Khlestov,N.,"W.T.O.-WIPO Co-operation:Does It Have a Future?"European Intellectual Property Review,1997,pp.560.

索　引