信息安全国家重点实验室信息隐藏领域丛书
中国科学院大学网络空间安全学院教材

隐写学原理与技术

赵险峰　张　弘　编著

科学出版社

信息安全国家重点实验室信息隐藏领域丛书
中国科学院大学网络空间安全学院教材

隐写学原理与技术

赵险峰　张　弘　编著

科学出版社
北京

内 容 简 介

隐写的主要作用是保护保密通信与保密存储的事实不被发现，而隐写分析的主要作用是发现这类事实。随着网络与多媒体应用的普及，隐写与隐写分析的研究发展很快，它们之间的对抗不断进入更高级的阶段，有必要进行系统的描述与全新的总结。本书将隐写与隐写分析作为一个新学科——隐写学进行了系统阐述，主要内容包括隐写与隐写分析的发展背景、主要性能指标、基本的消息嵌入方法、隐写分布特性保持、矩阵编码、专用隐写分析、湿纸编码、基于±1 的分组隐写编码、通用隐写分析、高维特征通用隐写分析、最优嵌入理论、校验子格编码、自适应隐写、选择信道感知隐写分析与基于深度学习的隐写分析，其中各个子领域的内容也概括了最新的主要研究成果。此外，本书各章的小结与最后一章给出进一步阅读和思考的方向，除最后一章外，每章配有用于巩固知识的思考与实践，附录部分给出了相关的基础知识介绍及实验方案，有助于读者全面学习并形成研究能力。

本书可以作为信息安全相关领域研究人员的参考资料，也可以作为信息安全或相关专业研究生与高年级本科生的教材。

图书在版编目（CIP）数据

隐写学原理与技术 / 赵险峰，张弘编著. — 北京：科学出版社，2018.10

ISBN 978-7-03-059117-3

Ⅰ. ①隐… Ⅱ. ①赵… ②张… Ⅲ. ①密码术-高等学校-教材 Ⅳ. ①TN918.4

中国版本图书馆 CIP 数据核字(2018)第 240124 号

责任编辑：阚 瑞 / 责任校对：郭瑞芝

责任印制：师艳茹 / 封面设计：迷底书装

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

文林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2018 年 10 月第 一 版 开本：720×1 000 1/16

2018 年 10 月第一次印刷 印张：22

字数：430 000

定价：128.00 元

（如有印装质量问题，我社负责调换）

前　言

很多人都了解加密，但对隐写(steganography)的作用不够清楚。一般来说，加密保护保密通信或存储的内容不受非授权的浏览，而隐写保护保密通信或存储的事实不被发现，有学者认为它们提供了机密性的两个方面。在一些应用环境下，保护保密通信或存储的事实非常重要，例如，在不安全或者被监控的环境下，它是进行安全保密通信的前提之一。

隐写是一类信息隐藏技术，它将机密信息隐藏在可公开的数据内容中传输或者保存，使非授权者不但不能浏览保密的内容，而且难以知道保密通信或者机密存储事实的存在，只有授权的接收者才能从隐写后的含密载体中提取隐蔽消息。隐写的主要应用是隐蔽通信与存储，与加密技术不同的是，隐写的主要目的不是对通信或存储内容进行保密，而是要在此基础上隐蔽保密的事实。识别隐写载体的技术称为隐写分析(steganalysis)，它一般通过检测媒体特征的变化判定隐写的存在，因此，隐写的主要设计目标是提高含密载体的隐蔽性，并在此基础上尽量提高数据的隐藏量等性能。

隐写技术历史悠久，在古代人们就已经使用隐写技术，但是，相比密码方法，隐写一直在通信能力上存在较大的劣势，因此长时期没有获得显著的发展。20 世纪末期，随着数字媒体与网络的普及，隐写获得了越来越好的载体源与传输条件，因此，相关的研究非常活跃，也出现了得到广泛关注的应用。当前，人们已经提出了很多以各类数字媒体为载体的隐写方法，互联网上出现了大量的隐写软件，隐写分析方法也日益丰富。

多年来，隐写经常被称为“隐写术”，但是随着隐写与隐写分析方法的发展，有关隐写的理论与技术已逐渐成长为一门较为系统的学科，有必要进行系统的描述与全新的总结。因此，本书采用隐写学作为隐写与隐写分析的总称，并结合最新的研究发展情况对其进行阐述，主要内容包括隐写与隐写分析的发展背景、主要性能指标、基本的消息嵌入方法、隐写分布特性保持、矩阵编码、专用隐写分析、湿纸编码、基于±1 的分组隐写编码、通用隐写分析、高维特征通用隐写分析、最优嵌入理论、校验子格编码、自适应隐写、选择信道感知隐写分析与基于深度学习的隐写分析。为了避免读者陷入各类多媒体编码与应用场景的细节中，在描述上，本书主要通过介绍图像隐写与隐写分析阐述隐写学的基本原理与技术，仅在最后介绍视/音频隐写与隐写分析等子领域的基本发展情况。为了方便读者学习并巩固提高，除最后一章外，书中每章都配有思考与实践，各章的小结与最后一章给出进一步阅读和思考的方向，附录部分回顾所需的基础知识并给出实验方案。

本书是作者在长期从事相关研究与教学工作的基础上完成的，书中主要内容已经在中国科学院大学网络空间安全学院以及计算机与控制学院的信息隐藏课程中讲授。其中，赵险峰教授主要负责撰写正文部分，张弘助理研究员主要负责撰写实验部分。若作为教材使用，全部讲授本书并完成附录实验大概需要 60~80 个学时，建议分为两个学期完成；如果只安排一个学期的基础课程，建议仅主要讲授第 1~6、8、11、12 章的内容，大约需要 40 个学时。本书的主要课件、实验代码与勘误表可以在 http://www.media-security.net 下载。

本书的编著与出版得到了各方面的帮助。本书的出版得到了 NSFC-通用技术基础研究联合基金项目(U1636102)、中国科学院战略性先导科技专项课题(XDA06030600)与信息安全国家重点实验室重点部署项目(2017-ZD-04)的资助；本书的编著得到了信息隐藏领域同行的指导，尤新刚研究员、郭云彪研究员、许舟军高级工程师、张卫明教授、李晓龙教授、朱美能副研究员审阅了本书，提出了大量宝贵意见；作者所在单位的同事与研究生也为本书的编著提供了重要帮助，荆继武教授、薛锐研究员、高丽丽女士、杨林春女士为作者在中国科学院大学开设信息隐藏课程提供了支持，为本书的教学试用创造了条件，夏超、赵增振、马赛、李晔、曹纭、易小伟、尤玮珂、张振宇、刘亚奇、王运韬等提供了部分参考资料、数据和实验程序。对于以上帮助，作者在此一并表示衷心感谢。

作者希望本书能够为隐写学的发展与教学尽一份力量，隐写学科学背景丰厚，涉及多媒体信号处理与编码、统计学、代数学、模式识别、信息论、信源与信道编码、密码学、最优化理论、深度学习等多个领域，知识交叉融合且更新快，作者虽全力以赴，但由于能力与时间有限，总难免仍有不足，敬请读者指正。如发现错误或提供意见和建议，欢迎发送电子邮件至 ih_ucas@163.com。

作　者

中国科学院信息工程研究所，信息安全国家重点实验室

中国科学院大学网络空间安全学院

2018 年 5 月 1 日

目　录

前言

第 1 章　绪言 …… 1
1.1　从密码到信息隐藏与隐写 …… 1
1.1.1　密码方法的一些局限 …… 1
1.1.2　信息隐藏基本概念 …… 2
1.1.3　隐写与隐写分析对抗模型 …… 5
1.2　隐写的应用发展 …… 9
1.3　隐写安全指标 …… 11
1.3.1　基于分布偏差的指标 …… 11
1.3.2　基于抗隐写分析性能的指标 …… 12
1.4　本书内容安排 …… 13
1.5　小结 …… 14
思考与实践 …… 14

第 2 章　图像编码与基本嵌入方法 …… 15
2.1　空域编码图像 …… 15
2.1.1　光栅格式 …… 15
2.1.2　调色板格式 …… 17
2.2　变换域编码图像 …… 17
2.3　基本嵌入方法 …… 19
2.3.1　LSB 替换 …… 20
2.3.2　LSB 匹配与加减 1 …… 21
2.3.3　调色板图像嵌入 …… 21
2.3.4　量化调制 …… 23
2.4　小结 …… 25
思考与实践 …… 25

第 3 章　隐写分布特性保持 …… 26
3.1　分布保持问题 …… 26
3.1.1　LSBR 分布问题与 χ^2 分析 …… 26
3.1.2　分布保持及其困难性 …… 29
3.2　基于调整修改方式的方法 …… 31
3.2.1　F3 隐写 …… 31

3.2.2 F4 隐写……32
3.3 基于预留区的方法……33
3.3.1 预留补偿区的分布恢复……33
3.3.2 预留“死区”的分布保持……38
3.3.3 预留补偿区的二阶分布恢复……41
3.4 基于统计模型的方法……44
3.4.1 HPDM 隐写……44
3.4.2 MB 隐写……45
3.5 小结……49
思考与实践……50
第 4 章 矩阵编码……51
4.1 线性分组纠错码的启发……51
4.2 矩阵编码的一般描述……53
4.2.1 矩阵编码嵌入与提取……53
4.2.2 矩阵编码的一些性质……54
4.3 典型的矩阵编码隐写……56
4.3.1 F5 隐写……56
4.3.2 MME 隐写……58
4.4 小结……60
思考与实践……60
第 5 章 专用隐写分析……61
5.1 对空域隐写的专用分析……61
5.1.1 RS 分析……61
5.1.2 对彩色图像的 RQP 分析……66
5.1.3 SPA 分析……68
5.1.4 直方图特征函数质心分析……70
5.2 对 JPEG 隐写的专用分析……73
5.2.1 对 OutGuess 的块效应分析……73
5.2.2 对 MB 的直方图分析……76
5.2.3 对 F5 隐写的校准分析……78
5.3 对调色板图像隐写的专用分析……82
5.3.1 奇异颜色分析……82
5.3.2 颜色混乱度分析……83
5.4 小结……85
思考与实践……86

第 6 章　湿纸编码……87
6.1　湿点与干点……87
6.2　湿纸编码算法……89
6.2.1　编码原理……89
6.2.2　消息容量分析……90
6.2.3　一个基本算法……91
6.3　典型的湿纸编码隐写……94
6.3.1　量化扰动……94
6.3.2　抗收缩 JPEG 隐写……94
6.3.3　双层隐写……95
6.4　小结……95
思考与实践……96
第 7 章　基于±1 的分组隐写编码……97
7.1　一个特例——LSBM-R……97
7.2　基于和差覆盖集的 GLSBM……99
7.2.1　GLSBM 基本构造方法……99
7.2.2　和差覆盖集的生成……102
7.3　小结……103
思考与实践……104
第 8 章　通用隐写分析……105
8.1　通用隐写分析基本过程……105
8.2　通用空域隐写分析……106
8.2.1　小波高阶统计特征分析……107
8.2.2　SPAM 特征分析……108
8.3　通用 JPEG 隐写分析……110
8.3.1　Markov 特征分析……111
8.3.2　融合校准特征分析……113
8.4　通用盲隐写分析简介……116
8.4.1　隐写分析的多种工作模式……117
8.4.2　算法去失配……118
8.4.3　载体去失配……119
8.5　通用定量隐写分析简介……120
8.6　小结……122
思考与实践……123

第 9 章　高维特征通用隐写分析……124
9.1　FLD 集成分类器……124
9.1.1　基本构造……125
9.1.2　参数设置……126
9.2　富模型高维特征隐写分析……127
9.2.1　空域富模型特征分析……127
9.2.2　JPEG 富模型特征分析……133
9.3　随机投影与相位感知分析……138
9.3.1　随机投影特征分析……138
9.3.2　相位感知特征分析……141
9.3.3　相位感知随机投影特征分析……153
9.4　小结……155
思考与实践……156

第 10 章　最优嵌入理论……157
10.1　一般情况……157
10.1.1　PLS 与 DLS 问题……157
10.1.2　最优修改分布的性质……158
10.2　加性模型……161
10.2.1　加性模型下的最优嵌入……161
10.2.2　加性模型最优修改分布求解……164
10.3　最优嵌入模拟……166
10.3.1　基于 Gibbs 抽样的模拟……166
10.3.2　基于子格迭代的模拟……168
10.4　小结……170
思考与实践……170

第 11 章　校验子格编码……171
11.1　STC 基本思想……171
11.2　STC 算法……173
11.3　双层 STC……177
11.3.1　基于三元嵌入分解……178
11.3.2　基于两层嵌入综合……183
11.4　小结……186
思考与实践……186

第 12 章　自适应隐写……187
12.1　限负载自适应隐写……187

12.1.1 基本框架 ……187
12.1.2 图像空域自适应隐写 ……189
12.1.3 JPEG 域自适应隐写 ……198
12.2 限失真自适应隐写 ……201
12.2.1 基本框架 ……201
12.2.2 限平均失真隐写 ……202
12.2.3 限平均统计量失真隐写 ……205
12.3 非加性模型自适应隐写 ……208
12.3.1 子格嵌入与失真修正 ……209
12.3.2 联合失真及其分解 ……214
12.4 小结 ……216
思考与实践 ……217

第 13 章 选择信道感知隐写分析 ……218
13.1 空域图像选择信道感知分析 ……218
13.1.1 基于区域选择的方法 ……219
13.1.2 基于特征权重的方法 ……219
13.2 JPEG 图像选择信道感知分析 ……224
13.3 小结 ……226
思考与实践 ……226

第 14 章 基于深度学习的隐写分析 ……227
14.1 深度卷积神经网络简介 ……227
14.2 针对空域隐写的 CNN 分析 ……231
14.2.1 基本框架的形成 ……231
14.2.2 支持选择信道感知的 CNN 分析 ……236
14.3 针对 JPEG 域隐写的 CNN 分析 ……239
14.3.1 混合深度学习网络 ……239
14.3.2 支持相位感知的 CNN 分析 ……241
14.4 小结 ……244
思考与实践 ……246

第 15 章 其他与后记 ……247
15.1 其他进展 ……247
15.2 部分问题 ……252

参考文献 ……256

附录 A 部分基础知识提要 ……274
A.1 数学与统计学 ……274

A.1.1　群、子群与陪集 ……274
A.1.2　环与域 ……276
A.1.3　线性回归及其误差估计 ……277
A.1.4　Lagrange 乘子法最优化求解 ……279
A.2　信息论与编码 ……280
A.2.1　信息量单位与转换 ……281
A.2.2　Fisher 信息 ……281
A.2.3　KL 散度性质 ……283
A.2.4　Huffman 编码 ……285
A.2.5　线性分组纠错码 ……286
A.3　模式识别 ……290
A.3.1　分类问题与判别函数 ……290
A.3.2　Bayes 分类器 ……291
A.3.3　线性分类器 ……292
A.3.4　支持向量机 ……295
A.3.5　神经网络基础 ……299
A.4　信号处理与检测 ……304
A.4.1　离散 Fourier 变换 ……304
A.4.2　离散余弦变换 ……305
A.4.3　离散小波变换 ……306
A.4.4　最小均方误差直方图修正 ……307
A.4.5　假设检验 ……308
附录 B　实验 ……313
B.1　图像隐写工具的使用 ……313
B.2　图像专用隐写分析 ……316
B.3　JPEG 图像通用隐写分析 ……318
B.4　空域图像自适应隐写 ……320
B.5　JPEG 图像自适应隐写 ……322
B.6　富模型空域图像隐写分析 ……324
B.7　选择信道感知隐写分析 ……326
B.8　空域图像 CNN 隐写分析 ……328
名词索引 ……331

第1章 绪 言

一般人们可能认为，保密通信是采用密码技术的通信，由发送方将消息加密后发送给消息的接收者，但是这种观点并不完全正确。在一些情况下，保密通信方法不但要保护消息内容不泄露，而且需要保护保密通信的行为不被识别。信息安全中机密性的一个定义是[1]：能够确保敏感或机密数据的传输和存储不遭受未授权的浏览，甚至可以做到不暴露保密通信的事实。密码数据存在伪随机特性，与非密码数据相比在统计特征上显著不同，因此，直接发送密文的保密通信难以掩盖保密通信的行为事实，直接存储密文的保密存储难以掩盖保密存储的行为事实，这对很多保密技术的用户来说是不希望发生的。

满足以上机密性全部要求的通信一般称为隐蔽通信(covert communication)。在现代隐蔽通信中，通常消息仍然被加密，但是加密消息的存储与传输方式一般是隐蔽的，即非授权方难以识别保密存储与保密通信的存在。为了达到这个目的，隐蔽通信或存储需要看起来像普通的日常行为。以下将说明，隐写(steganography)就是隐蔽通信或隐蔽存储的一种重要形式。为了描述简单，以下一般仅用隐蔽通信指代类似的需求。

1.1 从密码到信息隐藏与隐写

隐写是一类信息隐藏(information hiding)方法，在给出隐写的定义之前需要先介绍信息隐藏的概念，而要理解信息隐藏需要深入了解密码技术的局限。在很多情况下，引入信息隐藏的目的是满足密码方法难以满足的信息安全需求。

1.1.1 密码方法的一些局限

密码方法主要解决消息保密传输、数据来源认证与完整性认证等信息安全问题，但是密码方法并不解决以下两方面的问题。

(1)保密通信的行为隐蔽性问题。现代密码方法加密的数据具有伪随机性(pseudo randomness)，这种性质可以用游程、熵与各种自相关值等一系列统计特征来刻画，其特性是自然数据不具有的；Huffman 编码与算术编码等无损压缩数据的随机性很强，但是，通过解压缩可以验证这些数据是未经加密的。因此可以认为，将密码数据直接发送到信道上没有保密通信的行为隐蔽性，不利于在环境不安全的对抗场合进行安全通信。例如，当前已经出现了一系列检测网络加密流量[2,3]与密码协议[4]的方法。

(2)松散环境下的内容保护与内容认证问题。密码体制一般假设密文接收者是可信的，但是在诸如数字内容分发等应用中，合法接收者也存在肆意散布数字内容的版权违规可能，为了加以制约，存在保护解密后数据内容的需求，但这并不是密码方法解决的问题。需指出，在信息安全领域，内容(content)与数据有不同的含义：内容依赖于媒体表达的信息，与具体编码形式没有直接的关系，而数据一般是指信息的具体表现形式。例如，将一个 BMP 图像文件转换为 JPEG 格式图像，内容基本是一致的，而数据形式发生了很大的变化。显然，密码学的数字签名、验证码等方法是面向保护数据的，而在实际应用中，也需要对内容的所有权或者真实性进行认证或者保护，例如，一个视频可以经过多种格式编码或者被裁剪，但是人们可能希望验证其内容的版权所有者、购买用户或者内容的真实性，这显然不是针对数据安全的密码技术所能够实现的。

以下将说明，针对上述安全需求，研究人员建立的基于信息隐藏的理论基础与方法体系。

1.1.2　信息隐藏基本概念

信息隐藏又名数据隐藏(data hiding)，是指将特定用途的信息隐蔽地存储在其他载体(cover)中，使得它们难以被发现或者消除。其中，载体一般是可公开的数字内容，包括多媒体或者网络包等，隐藏的信息一般指保密通信的加密消息、内容所有权标识或内容用户标识等验证信息，信息可以被授权用户提取或验证，实现隐蔽通信、内容认证或内容保护等功能。载体内容存在的冗余性是信息隐藏获得存储空间的前提，当前的主要载体是多媒体，是因为多媒体信息冗余更多。信息隐藏的基本研究模型见图 1.1。根据所隐藏信息用途不同，信息隐藏主要分为面向隐蔽通信的隐写[5]与面向内容认证和内容保护的数字水印(watermarking，以下简称水印)[6]两类方法，其中，水印又可分为鲁棒(robust)水印与脆弱(fragile)水印等。在安全性与可靠性等主要性能指标的定义上，这些不同的信息隐藏方法大不相同。

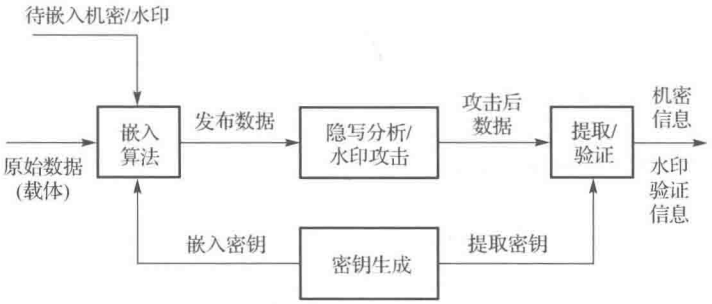

图 1.1　信息隐藏基本研究模型

1. 隐写

隐写是基于信息隐藏的隐蔽通信或者隐蔽存储方法，它将机密信息难以感知地隐藏在内容可公开的载体中，在保护保密通信或者存储内容的同时保护了这种行为事实。一般称隐写后的载体为隐文(stego-text)或者隐写媒体(stego-media)，也称为含密载体或者隐密载体。隐写技术历史悠久，英文 steganography 一词源于希腊语词根 στεγανός 和 γράφειν，意思是密写，说明在古希腊人们就已经使用隐写技术。但是，相比密码方法，隐写一直在信息传输率上有较大的劣势，因此长时期没有获得显著的发展。当前，随着网络与数字媒体应用的普及，这种情况正在迅速改变，隐写已经获得了非常好的载体来源与传输条件，因此，隐写的研究非常活跃，研究人员已经提出了很多以数字媒体为载体的隐写方法，本书称这类隐写为现代隐写，它们的主要性能如下。

(1)安全性。隐写的首要安全性是特征隐蔽性，因此，隐写的安全性一般就是指隐写后媒体特征变化的隐蔽性，即载体经过隐写后各种特征的变化难以被检测方法所发现。

(2)隐写容量。隐写容量指隐写传输的信息量。隐写容量可用负载率(payload)表示，负载率也称为嵌入率(embedding rate)，它表示平均每一个嵌入位置所承载的隐蔽消息量；令 m 表示传输的消息量，n 为嵌入位置的数量，则负载率 α 的计算公式为

$$\alpha=\frac{m}{n} \tag{1.1}$$

一般将每个可用的信号样点作为一个承载位置，普遍使用的负载率单位是 bpp(bits per pixel)与 bpnac(bits per nonzero alternating-current coefficient)等，bpnac 也常记为 bpnzac。

(3)嵌入效率(embedding efficiency)。嵌入效率 e 的含义是，平均每修改一个位置单元(一般是信号样点)所能传输的消息量，若消息量用比特表示，则其计算公式为

$$e=\frac{\text{平均每个载体样点承载的消息比特}}{\text{平均每个载体样点被修改量}}=\frac{\alpha}{d}=\frac{\frac{m}{n}}{\frac{E(K)}{n}}=\frac{m}{E(K)}\ \text{bit/次} \tag{1.2}$$

其中，d 称为平均每个载体样点被修改量；$E(K)$ 表示嵌入过程中总修改次数的期望值。显然，在传输相同消息量的前提下，提高 e 有助于减少修改次数，从而增加安全性。

(4)应用安全性。应用安全性指敌手难以从隐写应用协议与实现上发现有利于检测隐写媒体的方法；对于应用协议，这里也分为敌手知晓或者不知协议设计两种情况。

(5)计算效率。计算效率指隐写的算法执行效率；实际上任何方法都存在这个指标，但是，由于隐写多用于不安全的物理环境，这个指标也部分关系到隐写的应用安全。

(6)鲁棒性。隐写信道存在无损和有损两种情况，在有损情况下，含密载体面临有意或无意的干扰，隐写在这类条件下需要有抗干扰能力，但目前的相关研究普遍基于无损情况，并不假设信道存在干扰。

2. 水印

在互联网环境下，图像、音乐、影视和书籍等逐渐以数字内容的形式出现，这使复制品易于获得和传播，造成了娱乐业和出版业巨大的经济损失。20 世纪末期，社会对数字内容版权保护问题日益关注，并且人们开始认识到仅靠法律保护版权是不够的，因此出现了数字产权管理(digital rights management，DRM)技术。鲁棒水印是重要的版权保护与安全标识技术之一，它指将与数字媒体版权或者购买者有关的信息嵌入数字媒体本身中，使得攻击者难以在载体不遭到显著破坏的情况下消除水印，而授权者可以通过检测水印实现对版权所有者或内容购买者的认定，这种认定有助于判定版权权益或者侵权责任。需要指出，如果鲁棒水印隐藏的信息是数字内容购买者或者消费者的信息，有时也将这类水印技术称为指纹化(fingerprinting)技术。鲁棒水印(robust water marking)的嵌入方法是一般将水印信息进行信号调制，之后嵌入载体内容的相对稳定成分中，其性能主要体现在以下几方面。

(1)鲁棒性(robustness)。鲁棒性指在主动攻击下，授权用户仍然能够提取水印信息。主动攻击是指允许对含水印媒体进行一定的改动，包括对含水印媒体实施信号处理、添加噪声、有损压缩编码、尺寸缩放和裁剪等，但从攻击者的角度看，这种改动是适度的，不应该显著破坏数字内容的质量或者可用性。

(2)水印容量(capacity)。水印容量指水印能够可靠传输的信息量。有的水印方案只嵌入 1bit 信息，标识相应水印的有或无，有的方案允许嵌入更多信息，显然，在保持其他性能的前提下，后者往往更受欢迎。

(3)安全性(security)。安全性指水印攻击者难以从水印的算法、应用协议或者实现方法上获得有益于攻击的信息。

(4)盲性(blindness)。盲性指水印检测不依赖于原始媒体或其相关信息的存在，这样的水印方案也称为是公有的(public)，反之则为私有的(private)；当前信息隐藏领域普遍认为，具备盲性性质的水印方法才是有实用价值的。

由于存在大量攻击，当前实现完全有效的鲁棒水印难度非常大。尤其是在尺寸缩放和裁剪等几何攻击(geometric attacks)下，水印信息的检测或提取非常困难，设计抗几何攻击的水印方案已经成为一个挑战。但是，如果不那么理想化，可以认为，由于水印攻击降低了媒体的感知质量，水印在某种程度上还是成功的，这可能就像门锁一样，虽然从严格意义上防止不了坏人破门而入，但还是起到了防范作用。

目前已经出现了大量针对数字媒体的处理工具，它们可以修改数字内容而不易被人察觉，这为内容造假提供了方便。为保证数据内容的真实性和完整性，需要对许多来自政府、司法、军事和商业等部门的重要数据进行防伪处理。这显然可以通过密码技术中的数字签名实现，但是，数字签名产生了单独的签名数据，需要在应用中专门进行管理，而脆弱水印(fragile watermarking)技术可将防伪信息隐藏在数字内容本身中，以后通过水印检测发现篡改，还可以发现篡改的位置，更方便地支持了被保护内容的安全流转。脆弱水印方法隐藏在被保护内容中的信息会随着内容的改动而变化，这就是它能够进行内容认证与篡改定位的基础。脆弱水印的基本性质如下。

(1)脆弱性(fragileness)。嵌入被保护内容的水印应随着内容的改动而变化，其变化要反映内容被篡改的事实。

(2)定位精度。被嵌入水印随着内容改动的变化需要反映内容被篡改的位置。

(3)可逆性(reversibility)。可逆性指嵌入的水印可以被授权者完全消除，使原始媒体能够得到还原，具有这个性质的水印称为可逆水印(reversible watermarking)。

(4)安全性。类似地，亦指水印攻击者难以从水印的算法、应用协议或者实现方法获得有益于攻击的信息。

(5)盲性。类似地，亦指水印的检测不依赖于原始媒体的存在。

相比鲁棒水印，当前脆弱水印在技术上相对更成熟。内容的变化一般会改变水印的局部形态，因此，当前的脆弱水印能够较好地保护内容数据的完整性，并能够定位篡改区域；尤其是可逆水印技术使得授权者可以恢复原始载体，这也使得其更适合保护珍贵的影像资料。但是，一些脆弱水印方案距离内容(而不是数据)保护的需求还有一定概念上的偏差，理想情况下，水印检测不应该对格式转换等正常操作敏感，按照这类要求设计的脆弱水印称为半脆弱水印(semi-fragile watermarking)，它们只对内容的变化敏感而允许内容接受转码等正常处理，显然这样的水印更接近于实现内容保护的目标，但实现难度也更大。

在以上两大类信息隐藏方法中，本书专门讲述隐写的原理与技术。按照信息安全学科“盾”与“矛”相辅相成的原则，书中也将重点介绍隐写分析(steganalysis)。本书将逐渐展示，现代隐写与隐写分析已经成功运用了数学、信息论、编码、模式识别与信号处理等学科的成果，得到了一些经过优化的方法，产生了独有的核心理论与方法，因此，本书用隐写学总称隐写与隐写分析的相关理论与方法体系。

1.1.3 隐写与隐写分析对抗模型

一项信息安全技术的出现必然伴随着相应的攻击，攻守双方不断对抗发展。前面已经描述了隐写的定义，可以看出，隐写的主要目的是掩盖保密通信或存储的行为事实，技术上主要追求特征变化的隐蔽性，因此，隐写失败的标志是隐写事实的暴露。

隐写分析泛指针对隐写的攻击，它主要通过检测隐写后载体特征的变化判定隐写的存在；也有少量隐写分析的技术目的还包括对隐写算法、参数的估计或者对隐藏信息的非授权提取等，例如，定量隐写分析(quantitative steganalysis)的目标是得到隐写负载率。隐写与隐写分析的对抗模型示意请参见图 1.2。

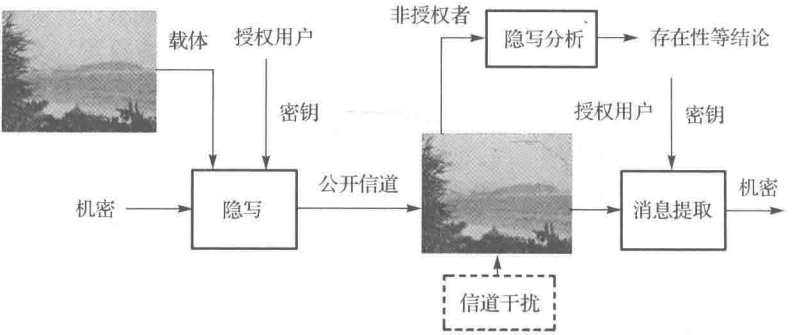

图 1.2　隐写与隐写分析对抗模型示意

图中隐写分析的主要性能指标如下。

(1)漏检率(miss detection rate)，也称为假阴性率(false negative rate)，指将隐写媒体判断为自然媒体的比率；与之相对的概念是真阳性率(true positive rate)或检测率(detection rate)=1−漏检率，即将隐写媒体判断为隐写媒体的比率。

(2)虚警率(false alarm rate)，也称为假阳性率(false positive rate)，指将自然媒体判断为隐写媒体的比率；与之相对的概念是真阴性率(true negative rate)=1−虚警率，即将自然媒体判断为自然媒体的比率。

(3)正确率(accuracy rate)，亦称精度或精确度，是隐写分析的主要技术指标，一般认为真阳性率与真阴性率同等重要，可表示为

$$\text{正确率}=1-\frac{\text{漏检率}+\text{虚警率}}{2}=\frac{\text{真阳性率}+\text{真阴性率}}{2} \tag{1.3}$$

与正确率相对应的是

$$\text{错误率}=\frac{\text{漏检率}+\text{虚警率}}{2}=1-\frac{\text{真阳性率}+\text{真阴性率}}{2} \tag{1.4}$$

隐写分析实验一般基于检测一组原始载体得到虚警率或真阳性率，并基于检测一组隐写样本(一般在一个负载率下)得到检测率或漏检率，进而得到正确率。以负载率为横轴、正确率为纵轴可得到相应的正确率曲线，但是它缺乏对以上两类错误率的综合描述。接收操作特性(receiver operating characteristic，ROC)曲线经常用来描述隐写分析分类器对阴(无隐写)、阳(有隐写)检测样本的综合分析性能(图 1.3)，它的横轴是虚警率，纵轴是检测率，阳性样本的性质(如负载率)是其总体属性。显