隐写分析*
——原理、现状与展望

梁小萍，何军辉，李健乾，黄继武
（中山大学电子与通信工程系，广东 广州 510275）

摘 要：信息隐藏是信息安全和多媒体信号处理领域一个非常年轻但又发展迅速的研究方向。隐写术和隐写分析是信息隐藏的重要研究内容。介绍隐写分析的原理，给出隐写分析方法的四个评价指标，从隐写分析类型的角度给出图像隐写分析通用原型系统。将国内外隐写分析研究方法分为专用和通用隐写分析两类，分别介绍其研究现状，对每类中典型的隐写分析方法进行评价，并对隐写分析中存在的 3 个主要问题进行讨论，指出隐写分析未来的发展方向，展望隐写分析的前景。

关键词：信息安全；多媒体信号处理；信息隐藏；隐写术；隐写分析
中图分类号：TP391 文献标识码：A 文章编号：0529-6579（2004）06-0093-04

密码技术是以往解决通信安全的一个主要手段，而作为信息加密补充办法的信息隐藏，是近年来多媒体信号处理领域提出的一种解决通信安全的新方法。信息隐藏是将某些特殊信息隐藏于正常载体之中，从而掩盖特殊信息存在的事实。隐藏了特殊信息的载体经由不安全信道传送称为隐秘通信。经过加密的信息由不安全信道传送称为加密通信。密码技术保护了信息内容，但却暴露了通信的行为，容易引起怀疑，也给攻击者留下了追踪的线索。隐秘通信掩盖了秘密通信的行为，不易引致攻击者的怀疑，攻击者也很难追踪发信者。

实现隐秘通信的技术手段主要是隐写术，隐写术是信息隐藏的重要分支。隐写术主要研究如何将实际存在的信息隐藏于正常载体中。最近 10 年 Internet 获得广泛使用，隐写术获得了蓬勃发展，已经被军事机构、政府部门、金融机构等涉及国计民生的重要部门采用，更被恐怖分子用来在 Internet 上互通消息[1]。隐写术的滥用给国家和社会带来了潜在的严重危害，如何有效监督隐写术的使用、防止隐写术的非法应用，成为国家安全等部门关切的问题。

隐写分析是对隐写术的攻击，目的是为了检测秘密消息的存在以至破坏隐秘通信，隐写分析是解决非法使用隐写术问题的关键技术。近几年来由于恐怖活动猖獗。而受到了较多的关注，获得了较大的发展，但还没有形成成熟的、系统化的理论体系。隐写分析技术的提高有利于防止隐写术的非法应用，可以起到防止机密资料流失、揭示非法信息、打击恐怖主义、预防灾难发生的作用，从而保证国家的安全和社会的稳定。隐写分析不仅具有重要的应用价值，更具有重要的学术意义。隐写分析研究可以揭示当前隐写术的缺陷，对隐写术的安全性进行测试与评价，这是信息隐藏技术发展与完善的一条有效途径。

1 隐写分析原理

1.1 隐写分析原理

隐写术通用的隐写过程可表示如下：

$$S' = S + f(S, M) \tag{1}$$

S 和 S' 分别代表载体消息和嵌入秘密消息后的隐藏消息，M 为待嵌入的秘密消息。隐写分析的过程就是从 S' 中检测出 M 以至提取 M。

我们用阐述隐写术的“囚犯”问题从攻击的角度对隐写分析进行分类。进行秘密通信的囚犯的来往信件都要经过看守的检查，看守检查消息后判断是否存在秘密消息并作不同的处理，称为被动攻击；如果看守不经判断就对消息进行修改的攻击称为主动攻击。被动攻击又分为被动隐写分析与主动隐写分析两种。被动隐写分析是指检测秘密消息的存在与否，并确定隐藏嵌入算法。主动隐写分析是指估计嵌入的秘密信息的长度，嵌入的位置，以及嵌入算法中使用的密钥和某些参数，最终提取秘密

* 收稿日期：2004－08－30
基金项目：国家自然科学基金资助项目（60172067）
作者简介：梁小萍（1977 年生），女，助教，研究生；通讯联系人：黄继武，博士生导师；E-mail：isshjw@zsu.edu.cn

消息。主动隐写分析相对于被动隐写分析要困难得多。目前国内外主要研究被动隐写分析，主动隐写分析的研究才刚刚起步。

不可感知和信息隐藏容量大是隐写术的基本要求。由于图像在互联网上应用广泛，且本身冗余度大，可嵌入较大的信息量，因此成为隐写术的主要载体。本文主要探讨未知载体消息或无法获得载体消息的情况下以图像为载体的隐写分析，即图像隐写分析，重点探讨图像的被动隐写分析。

1.2 隐写分析评价

隐写分析方法/技术的评价，这里特指对被动隐写分析方法的评价，可以采用 4 个评价指标：准确性、适用性、实用性和复杂度。

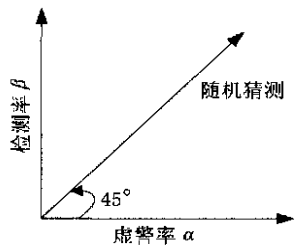

图 1 检测器 ROC 平面

Fig 1 Detector ROC plane

准确性指检测的准确程度，是隐写分析最重要的一个评价指标，一般采用虚警率和检测率表示，两个概率的关系可以描绘成如图 1 所示的检测器接收操作特性（detector's receiver operating characteristic，ROC）二维平面。虚警率是把非隐藏消息误判为隐藏消息的概率，表示为 $\alpha = P$（隐藏消息| 非隐藏消息）；检测率是把隐藏消息正确判为隐藏消息的概率，表示为 $\beta = P$（隐藏消息| 隐藏消息）。还需要考虑漏报率，即把隐藏信息错误判为非隐藏信息的概率，表示为 $\eta = 1 - \beta = P$（非隐藏消息| 隐藏消息）。

隐写分析要求在尽量减少虚警率和漏报率的条件下取得最佳检测率。在虚警率和漏报率无法同时减少的情况下，着重减少漏报率。

全面衡量隐写分析准确性的一个量是全局检测率 $P_r = 1 - P_e$，其中 P_e 为平均错误概率：

$$P_e = (1-\beta)P(\text{隐藏消息}) + \alpha P(\text{非隐藏消息}) = \eta P(\text{隐藏消息}) + \alpha P(\text{非隐藏消息}) \quad (2)$$

当 $\alpha = \beta$ 即点(α, β)落在图 1 的 45°对角线上时，全局测率为 50%，属于随机猜测，也即瞎猜，此时隐写分析检测器无效。当全局检测率达到 85%或以上，可以认为检测器性能良好。

适用性指检测算法对不同嵌入算法的有效性，可由检测算法能够有效检测多少种、多少类隐写术或嵌入算法来衡量。实用性，指检测算法可实际应用的程度，可由现实条件允许与否、检测结果稳定与否、自动化程度和实时性等来衡量。复杂度是针对检测算法本身而言的，可由检测算法实现所需要的资源开销、软硬件条件等来衡量。

到目前为止，还没有人给出适用性、实用性和复杂度的定量度量，只能通过比较不同检测算法之间的实现情况和检测效果得出一个相对的结论。

1.3 隐写分析通用原型系统

图 2 所示是从隐写分析类型的角度给出的图像隐写分析通用原型系统。隐藏秘密消息后的载体图像称为隐藏图像。待测图像输入后进行特征提取，根据图像的特征是否被改变以及改变的程度来判别图像是否隐藏图像。

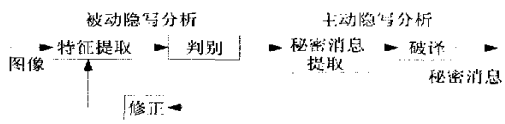

图 2 图像隐写分析通用原型系统

Fig 2 General prototype system of image steganalysis

特征提取包括特征的寻找与选择。根据特征提取与嵌入算法的关系，图像隐写分析有两条途径。一是针对某种具体的嵌入方法提取其专有特征，根据这些专有特征进行判别，可称为专用隐写分析技术；二是寻找独立于具体的嵌入算法之外的特征，根据这些特征进行判别，可称为通用隐写分析技术。专用隐写分析技术可以准确检测采用某种嵌入方法的隐藏图像，准确性高但适用性低。通用隐写分析技术在整体上准确性也许不如专用隐写分析技术，但适用性高。寻找对隐写术敏感的特征是隐写分析实现的关键。判别是根据提取的特征对图像进行归类、判断。修正是根据判别结果的好坏对提取的特征、判别系数或阈值作改动，最终达到正确判断或提高判别的准确性。秘密消息的提取包括估计秘密消息的嵌入量和秘密消息的嵌入位置。虽然该过程的实现难度更大，但是已经有若干检测算法在检测是否存在秘密消息的同时，可以比较准确估计秘密消息的嵌入量。到目前为止仍然没有能够准确确定秘密消息的嵌入位置的研究报道。成功破译是在秘密消息提取后的解密工作，到目前为止还没有隐写分析成功破译的报道，有的仅仅是初步的探索与分析。

2　隐写分析研究方法与现状

2.1　专用隐写分析方法

根据提取特征所在的域不同，专用隐写分析可分为空域隐写分析和变换域隐写分析。

空域隐写分析的攻击对象主要是空域 LSB 隐写术，包括 EzStego、S-Tools、Stash、Steghide、Gifshuffle、Stagano、BPCS 等，是隐写分析研究初期非常活跃的部分。

Westfeld 等[2] 采用 Chi-square 统计量统计调色板图像嵌入秘密消息前后出现近似颜色对概率比，可以可靠检测连续嵌入秘密消息的调色板图像，但对随机嵌入的真彩色图像检测无效。

Fridrich 等[3] 提出的 RS 检测法（regular groups and singular groups）把图像像素分成规则类、异常类和不可使用类，根据待测图像 LSB 置换操作前后每类像素组的变化曲线可以可靠检测灰度和真彩色图像并估计嵌入量，但算法的检测结果直接受载体图像随机性、噪声和秘密信息嵌入位置影响。

Dumitrescu 等[4] 提出的样本对分析法达到了与 RS 最优检测等效的结果。算法根据相邻像素值的奇偶性质将象素对分为 4 种基本集合，秘密消息的嵌入导致像素对从一个集合转换到另一个集合，根据集合更改的比例采用二次方程来估计嵌入量。该方法适用于对连续信号采样的检测，但检测结果直接受秘密信息嵌入位置影响，对非随机嵌入无效。

张涛等[5] 定义差分直方图的转移系数作为 LSB 平面与图像其余比特平面之间的弱相关性度量，并在此基础上构造载体图像与隐藏图像的分类器。在嵌入量比较大的情况下该算法检测效果优于 RS，但检测效果受载体图分布、嵌入位置和秘密消息随机性的影响。

空域隐写分析比较多的围绕颜色对现象进行研究，研究的方法经历了从简单分析隐藏图像颜色对到采用比较复杂的实验手段如再次嵌入秘密消息、归类、划分集合等来获得颜色对变化量的过程，总体来说适用性与实用性比较低。

变换域隐写分析的攻击对象主要是 DCT 域隐写术，包括 JSteg、Jsteg-Shell、JPHide、F5、Outguess、MB，是隐写分析研究现时活跃的部分。

Westfeld 等[2] 使用 Chi-square 统计量统计颜色频度检测隐写术 Jsteg 连续嵌入秘密消息的 JPEG 图像，对于嵌入量较大的情况检测准确率较高。该方法对离散嵌入情况的检测无效。

Fridrich 等[6] 通过解压缩待测图像、裁剪、再压缩等步骤估计载体图像的 DCT 系数直方图，根据待测图像直方图和估计直方图的相关改变量估计 F5 算法的嵌入量。该方法能有效检测低至 10% 的嵌入量，但对于具有特殊网格结构的图像无效。

Fridrich 等[7] 对待测图像进行 Outguess 嵌入操作，根据载体图像与隐藏图像像素块边界的增量差来估计嵌入算法 Outguess 的嵌入量。该方法无需确定阈值，对不可以由嵌入秘密消息的长度预见图像的宏观改变量的情况以及以 DCT 系数的增/减量作嵌入算法的无效。

DCT 域隐写分析主要围绕 DCT 系数的统计特性及其对空域像素的影响进行研究，包括了对载体图像 DCT 系数的估计及空域像素块不连续性的计算。研究的方法经历了从简单的一阶统计分析到采用比较复杂的实验手段来获得相关变化量的过程，总体来说也是适用性较低，实用性不高。

DWT 域隐写分析的研究报道较少，Shaohui Liu 等[8] 针对 DWT 域 QIM 嵌入算法，提出了基于 DFT 域能量差分的检测算法，检测率达到 90%。该文是检测 DWT 域隐写术的有益尝试。

2.2　通用隐写分析方法

通用隐写分析的攻击对象包括了空域与变换域隐写术，是隐写分析领域不可忽视的部分。

Avcibas 等[9] 提出的 IQM's（image quality metrics）方法，采用变量分析技术来分析和选取可用于区分载体图像和隐藏图像的质量度量，根据选取的图像质量特征采用多元回归对图像进行分类。该方法对多种隐写术的检测有效，但是需要对分类器进行训练，性能一般。

Farid 等[10] 采用 QFM 分析图像小波域系数及其预测误差的高阶统计量，再分别采用 Fisher 线性判别式、线性与非线性支持矢量机来判别和归类的方法，对 DCT 域隐写术和以自然图像为载体的隐写术效果较好。该方法需要对分类器进行训练，对嵌入量低的空域隐写术和 OutGuess 的检测无效。

通用隐写分析主要围绕嵌入秘密消息前后待测图像的总体、局部、相关等特征值及具有训练模式的判别方法进行研究，但是通用特征的选取和阈值的确定非常困难，而且复杂度偏高，实用性不强，准确性较低，无法控制虚警率和漏报率。

3　问题与展望

国内外的相关研究人员和学者在隐写分析方面已经取得了一些研究成果，但是留待大家去继续研究和解决的问题还不少，主要集中在隐写分析研究方法、隐写分析方法的评价和隐写分析理论构建与实用系统的实现三方面。

隐写分析研究现时主要采用统计分析方法，隐藏图像、秘密消息的统计特性仍有待仔细分析。近年来出现的抗统计分析隐写术旨在嵌入秘密消息的同时保持载体的统计特征不变，这给采用统计分析方法的隐写分析带来了新的挑战。隐写分析方法的评价方面还没有形成系统而又十分有效的隐写分析评价标准，有必要建立用于检测的测试图像库和相关的一系列评价量与评价手段。隐写分析理论构建方面现时把隐写分析简化为检测载体的噪声甚至去噪，如何区分随机噪声和秘密消息是一有待解决的问题，建立合理并符合实际的隐写与隐写分析模型更是亟待解决的问题。鉴于检测准确性、实用性和适用性等要求，构建行之有效的隐写分析系统并非容事，将统计分析和归类判断的方法相结合，实现全自动检测是构建实用检测系统的方向。

隐写分析有待研究的内容还包括：可用于检测的新特征；更有效、更准确的判别、分类方法；隐秘通信可靠性与安全性的度量；隐写分析的应用如图像隐写分析、视频序列隐写分析、音频隐写分析、文档隐写分析、一般文件隐写分析等。

隐写分析将不可避免地涉及社会道德与法律问题。私人使用隐写分析技术有可能遭受侵犯公民隐私权的控告，违法犯罪分子使用隐写分析技术获取私人、商业机构或国家机关的秘密消息以进行犯罪活动则可能会导致严重的后果，因此隐写分析的使用将会受到社会道德的规范与国家法律的限制。

隐写分析研究和发展的空间还很大。随着研究的广泛与深入，我们相信隐写分析的实时应用将为期不远。

参考文献：

[1] KELLEY J. Terror groups hide behind Web encryption[N]. USA Today, Feb. 2001, http://www.usatoday.com/tech/news/2001-02-05-binladen.htm.

[2] WESTFELD A, PFITZMANN A. Attacks on steganographic systems[C]. Proc. 3rd Int'l Workshop in Information Hiding. Springer Verlag, 1999: 61-76.

[3] FRIDRICH J, GOLJAN M, DU R. Reliable detection of LSB steganography in grayscale and color images[C]. Proc. ACM, Special Session on Multimedia Security and Watermarking. Ottawa CA, 2001: 27-30.

[4] DUMITRESCU S, WU X L, WANG Z. Detection of LSB steganography via sample pair analysis[J]. IEEE Transactions on Signal Processing, 2003, 51(7): 1995-2007.

[5] 张涛，平建西. 基于差分直方图实现 LSB 信息伪装的可靠检测[J]. 软件学报. 2004, 15(1): 151-158.

[6] FRIDRICH J, GOLJAN M, HOGEA D. Steganalysis of JPEG images: breaking the F5 algorithm[J]. 5th Information Hiding Workshop. Noordwijkerhout Netherlands, 2002: 10-323.

[7] FRIDRICH J, GOLJAN M, HOGEA D. Attacking the OutGuess[C]. Proc. ACM Workshop Multimedia and Security. Juan-les-Pins, France, 2002.

[8] LIU S H, YAO H X, GAO W. Steganalysis of data hiding techniques in wavelet domain[C]. International Conference on Information Technology: Coding Computing, ITCC. Las Vegas USA, 2004(1): 751-754.

[9] AVCIBAS I, MEMON N, SANKUR B. Steganalysis using image quality metrics[J]. IEEE Transactions on Image Processing, 2003, 12(2): 221-229.

[10] LYU S, FARID H. Steganalysis using color wavelet statistics and one-class support vector machines[C]. SPIE Symposium on Electronic Imaging. San Jose CA, 2004.

Steganalysis ——Principle, Actuality and Prospect

LIANG Xiao-ping, HE Jun-hui, LI Jian-qian, HUANG Ji-wu

(Department of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Information hiding has become the forefront of information security and multimedia image processing. Steganography and steganalysis are important research content of information hiding. Steganalysis principle, gives four estimated values for steganalysis method and general prototype system of image steganalysis based on types of steganalysis from the view of warden are introduced. Methods of steganalysis research home and abroad are divided into two types, special and universal steganalysis. Research actuality of either type is presented and representative methods of either type are evaluated. Three main problems existed in steganalysis are discussed, future direction in steganalysis development is pointed out, and prospect of steganalysis is expected.

Key words: information security; multimedia signal processing; information hiding; steganography; steganalysis