基于整体性的隐密分析特征提取和融合方法研究

郭艳卿[1], 何德全[1], 尤新刚[1,2], 孔祥维[1], 王 波[1]
(1. 大连理工大学信息安全研究中心, 辽宁大连 116024; 2 北京电子技术应用研究所, 北京 100091)

摘 要: 在Simmons的"囚犯问题"模型和Cachin的安全性理论模型下, 如何提取并融合统计特征是隐密分析技术亟待解决的关键问题之一. 基于对已有隐密分析技术及其所涉及的图像统计特征的分析, 本文将隐密分析技术所涉及的统计特征定义为载体数据固有特征和隐密方法引入特征, 并分别讨论了这两类特征的重要性. 在阐述隐密分析技术的整体性研究方法的基础上, 提出了一种基于整体性思想的特征提取及融合方法, 并以研制的图像隐密分析系统(StegDetect)验证了此整体性特征提取及融合方法的有效性.

关键词: 隐密分析; 固有特征; 引入特征; 整体性; 特征提取; 特征融合
中图分类号: TN918.1; TN911.73 文献标识码: A 文章编号: 0372-2112 (2006) 12A-2443-04

Research on Holism-Based Feature Extraction and Fusion for Steganalysis

GUO Yan-qing[1], HE Dequan[1], YOU Xin-gang[1,2], KONG Xiang-wei[1], WANG Bo[1]
(1. *Information Security Research Center of Dalian University of Technology, Dalian, Liaoning* 116024, *China*;
2. *Beijing Institute of Electronic Technology and Application, Beijing* 100091, *China*)

Abstract: The extraction and fusion of statistical features has been a key problem demanding solving for steganalytic technology in the context of the prisoner's problem presented by Simmons and Cachin's security theory model. We divide the statistical features into two categories, i.e., inherited features and introduced features, and the importance of these features is further discussed. Furthermore, based on the research of steganalytic methodology and the theory of holism, a fusion strategy of statistical features is proposed, and the availability of the proposed fusion strategy is verified by our image steganalysis system, which is developed using the strategy.

Key words: steganalysis; inherent feature; introduced feature; holism; feature extraction; feature fusion

1 引言

隐密(Steganography)及隐密分析(Steganalysis)技术的对抗是近些年来学术界关注的热点. 一般认为, 目前此领域的研究主要是围绕Simmons提出的"囚犯问题"模型[1]和Cachin提出的安全性理论模型[2]展开的. 基于这两个模型, 隐密技术已超越了仅"感官不可察觉"层次的安全性, 发展为具有保持载体部分统计特征[3,4]层次的安全性; 隐密分析技术也已从早期基于统计学的χ^2检测[5]发展为基于统计特征的定量分析方法[6~9]和通用性(有训练)分析方法[10,11].

从隐密分析技术的发展来看, 不仅需要研究载体数据的本质特性并以此为依据构建分析框架[12], 而且需要研究特定隐密方法向载体数据引入的新特征. 目前, 对载体数据本质特性和隐密方法引入特征的研究已取得一些研究成果: 载体数据本质特性有自然图像DCT系数服从特定的统计分布[13,14]、JPG图像空域块与块之间存在的不连续性[8]等; 隐密方法引入特征有LSB方法引入的直方图"对效应"[5]、RS特征的变化规律[6]、F5方法引入的直方图"收缩效应"[7]等. 然而, 隐密分析技术所涉及的统计特征具有复杂的层次性[15], 而上述研究成果仅描述了这两类统计特征的某些侧面. 本文将研究如何系统地提取及融合用于隐密分析的统计特征.

全文共分四部分: 第一部分综述了隐密分析技术的研究现状, 并阐明了统计特征对隐密分析的重要意义及其分类方法; 第二部分分析隐密分析技术研究思想, 并提出研究隐密分析技术的整体性思想; 第三部分给出一种基于整体性思想的特征提取及融合方法, 并以基于此方法研制的图像隐密分析系统(StegDetect)验证了此方法的有效性. 第四部分是本文的结论和展望.

2 隐密分析技术的研究现状

2.1 隐密与隐密分析技术的对抗

简单地说, 隐密的目的是在正常通信数据的掩护下实现隐密双方的秘密通信; 而目前隐密分析的目的是通过考察通信双方正常的通信数据, 发现潜在的秘密通信行为. 早期人们认为, 在相同隐密量的情况下, 对载体数据改动越小则隐密方法越安全. 因此设计了相同条件下对数据改动最小的LSB(最低比特位替换)隐密方法[16]. 然而, 人们发现可以通过图像统计特征的变化来判断图像中是否含有秘密信息. 基于LSB方法造成的直方图"对效应", 文献[5]提出了能从统计意义上给出有无LSB隐密信息的χ^2检测方法. 通过对载体数据统计特征的深入研究, 文献[6]提出了能估计图像中含有LSB隐密信息长度的RS隐密分析方法. 这也标志着隐密分析技术已发

收稿日期: 2006-06-10; 修回日期: 2006-11-27
基金项目: 国家自然科学基金(No. 60572111); 国家高技术研究发展计划(863计划)(No. 2005AA145010)

展为精确的定量分析阶段.

近些年来,隐密及隐密分析技术的研究是大都围绕文献[2]的安全性理论模型展开.此理论认为,如果隐密方法不改变载体数据的任何统计特征,则此隐密方法就是安全的.因此,隐密方法的设计应尽量保持载体数据固有的统计特征,以逼近"绝对安全";隐密分析方法将寻找统计特征是否存在异常的判据或寻找统计特征值与隐密信息量之间的变化关系,进而对待测图像进行是否含密判决或含密量估计.如 A. Westfeld 提出的不引入直方图"对效应"的 F5 方法, Niels Provos 在 Jsteg 基本原理的基础上提出的能够保持量化后 DCT 系数一阶统计特征的 OutGuess 方法; Jessica Fridrich 先后利用 F5 方法引入的直方图"收缩效应"[7]和 Outguess 方法引入的"块不连续性"变化规律[8]设计了分别针对 F5 和 Outguess 的定量隐密分析方法; Phil Sallee 应用将秘密信息通过熵编码器编码成近似给定统计分布的思想,提出 MB1[4]和 MB2[17]方法; Andreas Westfeld[18]发现柯西分布不能很好地拟合 JPG 图像量化后 DCT 系数直方图,并基于此提出了针对 MB1 的隐密分析方法.

2.2 固有特征和引入特征

上述分析表明,图像数据统计特征是研究隐密及隐密分析技术的关键.为进一步加深对统计特征的认识,我们将隐密分析技术涉及的图像数据统计特征区分为载体数据固有特征(简称固有特征)和隐密方法引入特征(简称引入特征),并作如下定义:

定义 1 载体数据固有特征是指客体经信息获取、存储以及正常数据处理等操作后所具有的特征,它是客观事物本源特征、获取方式特征、存储方式特征、正常处理特征等共同作用的结果.如量化后 DCT 系数的统计模型、JPG 图像的块不连续性等.

定义 2 隐密方法引入特征是指载体数据经特定隐密方法隐藏一定量的秘密信息后所具有的新特征,它是秘密信息统计特征、隐密位置特征、隐密方式特征等共同作用的结果.如 LSB 算法引入的"对效应"、F5 方法引入的直方图"收缩效应"等.

3 隐密分析技术的研究方法

3.1 隐密分析的三类技术方法

基于前述分析,目前已有的隐密分析技术可分为三类:基于统计学的判决方法、针对特定隐密方法的定量分析方法、基于训练的通用性分析方法.

基于统计学的判决方法是首先对某种载体数据固有特征或隐密方法引入特征进行合理的统计假设,然后再通过对样本数据的考察来验证统计假设的合理性,进而判决待测数据是否含有隐密信息.由于这类方法以统计学假设检验理论为基础,因此其输出结果仅为统计意义上是否含密的可能性.例如, χ^2 检测[5]是典型的基于统计学的判决方法.

针对性定量分析方法是对某一载体数据固有特征随特定隐密方法隐密信息量的变化规律进行假设,并根据待测数据的这一特征值来估计隐密信息量.这类方法有两个特点:(1)对待测数据已含有特定隐密方法隐密信息进行强假设,因此输出结果仅对判决待测数据是否含有特定隐密方法的隐密信息有意义;(2)所采用的规律不仅源于理论分析,更源于大量的数值实验,因此此类方法的稳定性可能受到最初实验环境的影响.文献[8~11]中的方法是典型的针对性定量分析方法.

基于训练的通用性分析方法主要是应用机器学习理论,建立由载体数据固有特征或隐密方法引入特征组成的特征空间,然后对载体数据样本和含密体数据样本在特征空间中的映射特征值进行机器学习,构建此特征空间的分类器来对待测数据是否含密进行判决.这类方法的特点是:能弥补人类在计算能力方面的不足,通过大数据量的机器训练来拟合具有复杂结构的决策系统.例如,文献[10,11]中的方法是典型的基于训练的通用性分析方法.

3.2 隐密分析的整体性(Holism)研究思想

上述三类技术方法中,基于统计学的判决方法和针对性定量分析方法是按还原论(Reductionism)思想进行研究的,而基于训练的通用性分析方法走的是"人机结合,以机器为主"的研究路线.这两种研究方法都有各自的优势和局限性.

还原论的思想是指为认识事物的整体,采用分析、分解、还原的方法进行研究,然后根据研究的结果把握整体的特性,即所谓分析——重构的方法[19].应用还原论思想,统计学的判决方法和针对性定量分析方法通过对众多统计特征的分析、理解,很好地解决了隐密分析的一些子问题,如待测数据是否含有载体或含密体的某一特定特征、假定使用了特定隐密方法的待测数据含有多少隐密信息.但想通过这两类方法对待测数据"是否含密"的最终判决是十分困难的,因为众多单一的特征只是整体的众多侧面,并不是整体本身.试想沿着这种研究方法,我们只能通过对众多特征进行穷举的综合方法进行最终判决,而特征的多样性和统计的不确定性必然造成判决的混乱.

基于训练的通用性分析方法试图靠机器(计算机)来解决隐密分析技术所面临的复杂性.在研究载体数据固有特征时,我们认识到了源于客体本源特征多样性、信息获取和存储时变换和量化等非线性关系的载体数据复杂性[17].同样,隐密方法引入特征也是复杂的,其复杂性来源于隐密方法的多样性,来源于复杂的密信处理方法、隐密位置选择方法和隐密方式.正是这些统计特征的复杂性给隐密分析带来了困难.基于训练的通用性分析方法是一种"人机结合,以机器为主"的研究方法.这种研究方法能弥补人类在计算能力方面的不足,通过大数据量的机器训练来拟合具有复杂结构的决策系统.它适合解决隐密分析的另一类子问题,如在已知部分载体和含密体样本来验证新特征的有效性、对无法用解析形式描述的决策系统进行拟合.但这种研究方法无法避免机器学习方法处理复杂问题时遇到的"维数危机"问题和泛化能力问题.

实际上,研究隐密分析技术应采用从整体上分析和解决问题的方法,即采用"整体性"研究方法,对不同层次的信息和知识进行综合集成,达到对整体的定量认识[19].与还原论研究方法相同的是,整体性研究方法也需要进行分解,不同的

是在整体指导下分解, 在分解后研究的基础上再综合集成到整体. 与“人机结合, 以机器为主”的方法相同的是, 整体性研究方法也需要计算机的参与, 不同的是整体性研究方法采用的是“人机结合, 以人为主”的方法来对特征进行融合.

4　基于整体性的特征提取和融合方法

4.1　固有特征和引入特征的提取

隐密分析技术的整体性研究方法首先体现在对统计特征的提取上, 对固有特征和引入特征的整体性把握也是隐密分析技术特征融合方法的重要基础. 固有特征和引入特征的优选不单是研究一个个的载体或含密体的集合、不单是研究一个个隐密方法, 而是通过对客观实际的理解和经验研究这些集合、方法相互之间的关系.

载体数据固有特征可以通过研究本源特征、获取方式特征和存储方式特征及转化关系来获得; 而隐密方法引入特征可以通过对隐密方法按密信处理方式、隐密位置选择方式和隐密方式进行归类分析. 固有特征和引入特征的数学模型不但要建立在对两类特征的实际理解和经验上, 而且要通过大量科学的数值实验检验模型的合理性和精度, 进行必要的参数修正.

4.2　基于整体性思想的特征融合方法

设已优选特征集合为 F, 其中固有特征集合 $F_{固}$, 引入特征集合为 $F_{引}$, 即 $F_{固}\cup F_{引}= F$.

设 $F_{inh}\in F_{固}$, 且对载体图像的批量估计量为 $F_{inh}(0)$, 均值、方差分别为 $\mu_{inh}(0)$ 和 $\sigma^2_{inh}(0)$; $F_{int}\in F_{引}$, 针对一个或者一类隐密算法秘密信息嵌入量为 α 的含密图像批量估计量为 $F_{int}(\alpha)$, 均值、方差分别为 $\mu_{int}(\alpha)$ 和 $\sigma^2_{int}(\alpha)$, 其中, $0<\alpha\leqslant 1$. 实验表明 $F_{inh}(0)$ 和 $F_{int}(\alpha)$ 近似服从高斯分布, 即:

$$F_{inh}(0)\sim N(\mu_{inh}(0),\ \sigma^2_{inh}(0))$$

$$F_{int}(\alpha)\sim N(\mu_{int}(\alpha),\ \sigma^2_{int}(\alpha))$$

则可计算当虚警率近似小于 p_f 时固有特征 $F_{inh}(0)$ 的判决门限 $t_{inh}(p_f)$, 如图 1 所示. 同理可得对于隐密信息量大于等于 α 时, 漏警率近似小于 p_l 时引入特征 $F_{int}(\alpha)$ 的判决门限 $t_{int}(p_l)$.

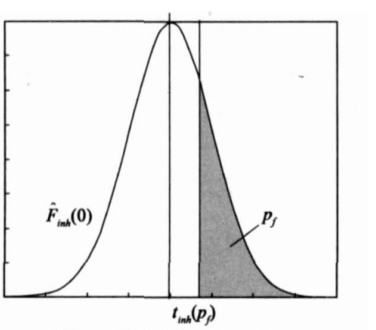

图 1　固有特征判决门限

设一幅待测图像的 M 个固有特征和 N 个引入特征分别为 $F_{inh(m)}$ 和 $F_{int(n)}$, 对应的判决门限分别为 $t_{inh(m)}(p_f)$ 和 $t_{int(n)}(p_l)$. 其中 $1\leqslant m\leqslant M$, $1\leqslant n\leqslant N$. 则融合决策准则可表示为:

$$\mathrm{Dec}=\begin{cases}\text{不含密}, & \text{当}\prec_{1\leqslant m\leqslant M}(F_{inh(m)}< t_{inh(m)}(p_f))=\text{true 时}\\ \text{含密}, & \text{当}\prec_{1\leqslant n\leqslant N}(F_{int(n)}< t_{int(n)}(p_l))=\text{true 时}\\ \text{不确定}, & \text{else}\end{cases}$$

4.3　图像隐密分析系统(StegDetect)的构建

我们以对 BMP(位图)格式图像为例来说明 StegDetect 图像隐密分析系统中特征提取及融合方法的具体实现过程.

首先, 我们对多种 BMP(位图)格式图像隐密分析方法进行原理分析, 研究各隐密分析方法理论上存在的误差来源(如 RS[8] 隐密分析方法的初始偏差等问题); 并通过基于载体图像库的大量数值实验检验各隐密分析方法的准确性和稳定性(检验指标为定量隐密分析方法对批量图像隐密量估计值的均值、方差和基于统计学判决方法的虚警率、漏警率等); 然后, 剔除 χ^2 检测等性能较差的隐密分析方法, 并根据具体实验结果对性能较好的隐密分析方法进行模型修正, 如修正定量分析方法的曲线参数、根据复杂度预判减小初始偏差对 RS 分析性能的影响等; 再优选出基于复杂度的 RS、Sample-Pair[20]、Fridrich 最优逼近[21] 等隐密方法的统计特征组成特征集合; 最后按 4.2 中的融合准则进行最终判决.

4.4　与 StegoSuite 的检测性能对比实验

为验证此分析系统的检测性能, 我们将研制的 StegDetect 与美国 WetStone 公司开发的商业隐密分析软件——StegoSuite 进行了检测性能对比. 在对已发布的隐密分析软件[22-24]的对比中, StegoSuite 的功能相对完备且综合性能较强. 对比实验所使用的图像为不同尺寸、不同参数的 JPG 文件格式的载体图像(由多种相机按不同参数拍摄得到)及其经过流行隐密软件和方法按不同隐密量得到的含密图像. 比较结果如图 2 所示:

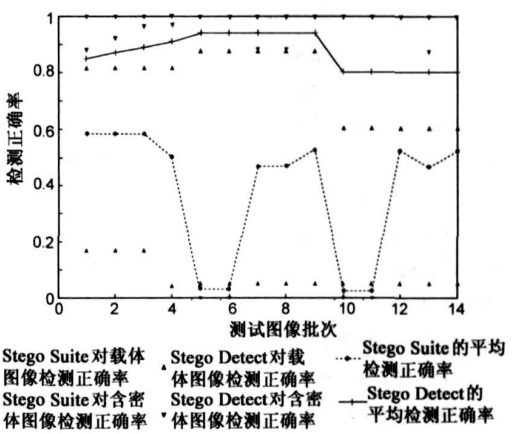

图 2　JPEG 图像比较测试结果

从图中可以看出, StegDetect 对 JPG 格式载体图像的检测正确率在 80% 以上, 对隐密图像的检测正确率高于 85%; 整体检测性能远优于 StegoSuite.

5　结论与展望

针对目前隐密分析技术亟待解决的关键问题, 本文首先将隐密分析技术所涉及的统计特征定义为载体数据固有特征和隐密方法引入特征, 然后在讨论隐密分析技术研究思想的基础上提出了一种基于整体性思想的隐密分析技术特征提取及融合方法, 并按此方法开发了图像隐密分析系统(StegDetect). 与美国 WetStone 公司开发的商业隐密分析软件 StegoSuite 的检测性能比较结果验证了此方法的有效性.

目前已开发的图像隐密分析系统仅是对本文所提出方法的初步探索, 后续工作包括对统计特征的不同建模方法、模型合理性的验证方法和精度分析方法, 以及基于知识工程和专

家系统等人工智能技术的特征融合方法的研究.

参考文献:

[1] Simmons G J. The prisoner's problem and the subliminal channel[A]. Advances in Cryptography: Proceedings of CRYPTO'83[C]. London: Plenum Press, 1983. 51-67.

[2] Cachin C. An information-theoretic model for steganography [A]. Information Hiding: Second International Workshop, vol. 1525 of Lecture Notes in Computer Science[C]. Portland, Oregon, U. S. A: Springer-Verlag Press, 1998. 306-318.

[3] R Chu, X You, X Kong. A DCT-based image steganographic method resisting statistical attacks[A]. The 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Montreal, Canada: Springer-Verlag Press, 2004. 953-956.

[4] Phil Sallee. Model-based steganography [A]. International Workshop on Digital Watermarking [C]. Seoul, Korea: Springer-Verlag Press, 2003. 154-167.

[5] A Westfeld, A Pfitzmann. Attacks on steganographic systems: Breaking the steganographic utilities EzStego, Jsteg, Stegnos, and S-tools-and some lessons learned[A]. The 3rd Information Hiding Workshop, Lecture Notes in Computer Science 1768 [C]. New York: Springer-Verlag Press, 2000. 61-76.

[6] J Fridrich, R Du, M Long. Steganalysis of LSB encoding in color image[A]. Proceeding of IEEE International Conference on Multimedia and Expo[C]. New York, USA: IEEE Press, 2000. 1279-1282.

[7] J Fridrich, M Goljan, D Hogea. Steganalysis of JPEG images: Breaking the F5 algorithm[A]. Proceeding of 5th Information Hiding Workshop LNCS vol. 2578[C]. New York: Springer Verlag Press, 2002. 310-323.

[8] J Fridrich, M Goljan, D Hogea. Attacking the OutGuess[A]. Proceeding of the ACM Workshop on Multimedia and Security 2002[C]. Juan-les-Pins, France: ACM Press, 2002. 3-6.

[9] 刘畅. 针对信息隐藏的数字图像数据特性研究[D]. 辽宁大连: 大连理工大学, 2002.

[10] Hany Farid. Detecting Steganographic Messages in Digital Images[R]. USA: Dartmouth College, Computer Science, 1999. TR2001-412.

[11] J Fridrich. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes[A]. International Workshop on Information Hiding[C]. Toronto (CA): Springer Verlag Press, 2004. 67-81.

[12] 孔祥维. 信息隐藏技术安全性的攻击框架及方法[A]. 信息隐藏全国学术研讨会(CIHW2002) 论文集[C]. 北京: 机械工业出版社, 2002. 42-47.

[13] Edmund Y Lam, Joseph W Goodman. A mathematical analysis of the DCT coefficient distributions for images[J]. IEEE Transactions on Image Processing, 2000, 9(10): 1661-1666.

[14] Joon-Hyuk Chang. Image probability distribution based on generalized Gamma function[J]. IEEE Signal Processing Letters, 2005, 12(4): 325-328.

[15] 郭艳卿. 信息隐藏的层次安全性[J]. 中山大学学报(自然科学版), 2004, 43(增刊): 105-108.

[16] Walton S. Image authentication for a slippery new age[Z]. Dr. Dobb's J Software Tools Profess Program, 1995. 18-26.

[17] Sallee P. Model-based methods for steganography and steganalysis[J]. International Journal of Image and Graphics (IJIG), 2005, Volume 5: 167-190.

[18] Rainer B, Andreas Westfeld. Breaking Cauchy model-based JPEG steganography with first order statistics[A]. Computer Security (ESORICS 2004), LNCS 3193[C]. Sophia Antipolis, France: Springer Verlag, 2004. 125-140.

[19] 许国志. 系统科学[M]. 上海: 上海科技教育出版社, 2000. 310-319.

[20] Sorina Dumitrescu. Detection of LSB steganography via sample pair analysis[A] Proc of the 5th International Workshop on Information Hiding, LCNS 2578[C]. New York: Springer-Verlag, 2002. 55-372.

[21] Jessica Fridrich. On estimation of secret message length in LSB steganography in spatial domain[A]. Security, Steganography, and Watermarking of Multimedia Contents[C]. Saint Joey, USA: SPIE Press, 2004. 23-34.

[22] SpyHunter 软件[CP]. http://www.spy-hunter.com/index.html, 2002.

[23] Outguess 软件[CP]. http://www.outguess.org/, 2002.

[24] StegoSuit 软件[CP]. http://www.wetstonetech.com, 2001.

作者简介:

郭艳卿 男, 1980 年出生于辽宁省新民市, 2002 年毕业于大连理工大学电子工程系, 现为大连理工大学博士研究生. 研究方向: 信息隐藏、信号与信息处理.

E-mail: sagat823@yahoo.com.cn

何德全 男, 1933 年出生于北京, 1994 年当选为中国工程院院士, 信息技术专家. 国家信息化专家咨询委员会副主任. 在主持大型信息系统工程的建设中作出创造性贡献. 在信息防护与安全技术、新型显示与处理技术、信息光学与化学等多个领域, 有较高的学术造诣. 取得了 20 余项高难度、高水平的科技成果, 其中 10 项获国家发明二等奖、国家科技进步奖及部级科技进步奖. E-mail: hedq@dlut.edu.cn