隐写分析

雷敏

北京邮电大学 网络空间安全学院

leimin@bupt.edu.cn

前言

- 隐写（steganography）
  - 目的：以表面正常的数字载体如静止图象、数字音频和视频信号等作为掩护，在其中隐藏秘密信息。额外数据的嵌入既不改变载体信号的视、听觉效果，也不改变计算机文件的大小和格式（包括文件头），使隐蔽信息能以不为人知的方式进行传输
- 隐写分析（steganalysis）

前言

- 早在2001 年初，震惊世界的9. 11 事件发生半年多以前，美国报纸就曾刊登文章，指出本·拉登及其同伙可能利用某些网站上的大量数字图像秘密传递与其恐怖行动有关的信息如指令、地图、攻击目标的资料等
- 当时还有报道指出，一些著名的网站等已成为传播隐写信息的隐蔽渠道

前言

- 有报道称，首先将科学家在隐写研究中取得的早期成果用于实践的就有基地和哈马斯等国际恐怖组织
- 一些国家的警方也曾在恐怖组织的计算机内查获大量可疑图像和视频文件，据分析可能藏有与恐怖行动有关的信息

前言

- 一些研究者开始对著名网站上数以百万计的图像展开搜索和检测，试图寻找可能存在的敌对隐蔽信息
- 他们又用所谓字典式攻击法分析了USENET上数以百万计的文档
- 这些工作虽然未能找到隐蔽恐怖信息的确凿证据，却推动了隐写和隐写分析的研究

前言

- 隐写和隐写分析在军事、情报、国家安全方面的重要意义是不言而喻的
- 设计高度安全的隐写方法是一项富于挑战性的课题
- 对隐写的准确分析往往比隐写本身更加困难

隐写分析的目标

- 判断是否隐写
- 估计隐写量多少
- 提取隐藏信息

隐写分析方法

- 感观分析
- 特征分析
- 统计分析
- 通用分析

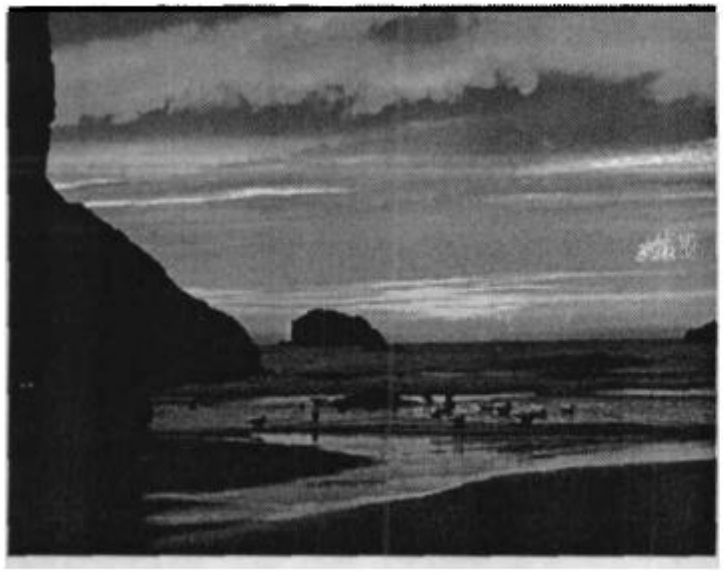
感观分析

carrier-image

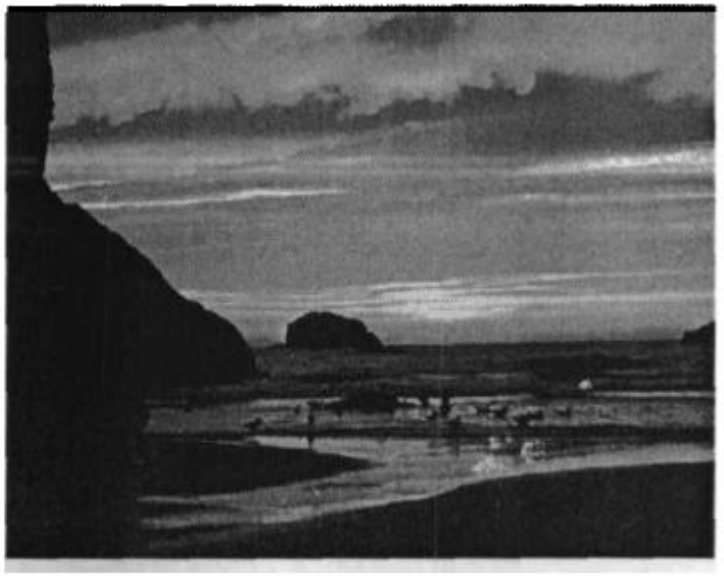
stego-image

the 1th plane of carrier-image

the 1th plane of stego-image



(a)



(b)

(c)

(d)

感观分析

- 优点
  - 简单、直观
- 弱点
  - 自动化程度弱
  - 可靠性弱

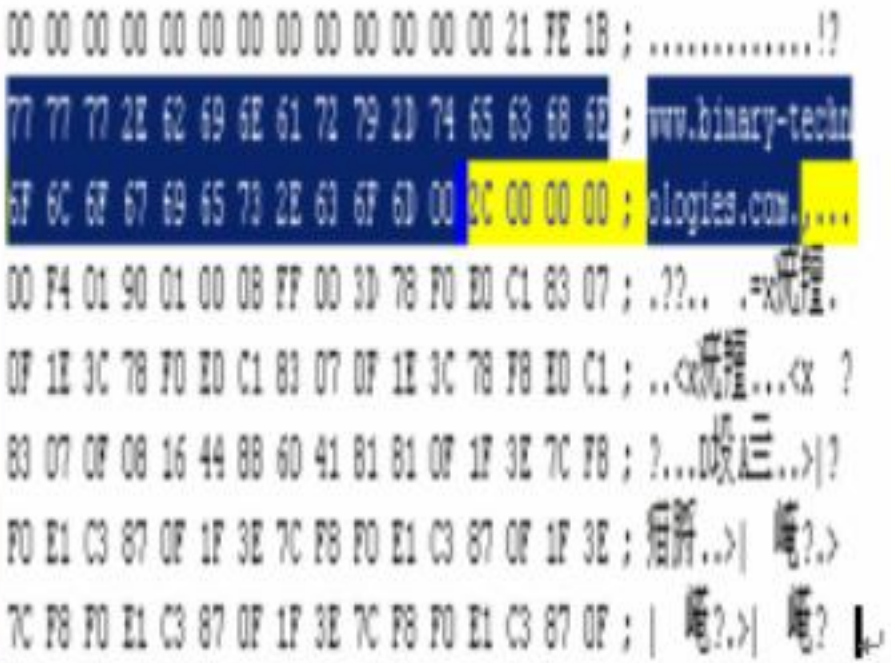
特征分析

- 基于文件结构的隐写特征
  - 文件大小异常
  - 调色板中有像素没有使用的颜色
  - 。 。 。
- 软件特征
  - 2006年高清晰度DVD视频播放器面世时，使用了强度较高的加密算法。可仅仅6个月之后，系统就被破解了，问题不是出在算法之上，而是算法的实现，攻击者能从内存获取密钥。

特征分析

- 隐写软件特征例1

特征分析



隐写软件TheThirdEye的隐写标记：www.binary-techNologies.com

特征分析

- 隐写软件特征例3
  - 隐写软件Securengin3.0特征码
  - 0111 0000 1101 1110 1011 0010 0100 0110 1101 1010 1110 1111
  - 1111 0111 0100 0110 0011 1101 0010 0100 0001 1110 1010 1000
- 隐写软件特征例4
  - 早期F5算法总插入“JPEG Encoder Copyright 1998, James R. Weeks and BioElectroMech”，而普通图像编辑器几乎不会插入这条信息。

统计分析

- 载体感观效果没有变化，但统计特征改变
- 分析待检测载体的统计特征，可以判断载体是否经过隐写
- 典型方法：
  - 卡方、RS检测等
  - JPEG检测等

通用分析方法

- 通用分析方法举例：
  - 原理：自然信号与其去噪信号的“距离”，隐写信号与其去噪信号的“距离”，两者不同
  - 提取出对隐写敏感的若干统计特征构成矢量
  - 获取待检测信号与其去噪版本的特征矢量
  - 判决

通用隐写分析方法

- 通用隐写分析方法一般分为两个步骤
  - 特征量选取
    - 利用特征量来描述自然和隐写载体之间的差异
    - 考虑到DWT的多分辨率辨析能力，图像常用的特常用统计量有DWT系数的均值方差偏离度峰度等
    - 特征量一般是由若干统计量构成的集合
    - 常用方差分析（ANOVA，analysis of variance）来选取特征量（即找出对隐写和原始载体的差异最敏感的统计量）
  - 选定特征量后，通用隐写分析就转变为一个分类问题

通用隐写分析方法

- 分类
  - 通常使用支持向量机（SVM，support vector machine）来进行分类。
  - 常用两类支持向量机，线性和非线性。
  - 支持向量机分类的思路是，通过培训集，即已知是自然和隐写的载体中提取的特征（可视为矢量空间中的点），找到一个判决阈值（可视为一个超平面，把代表隐写的点和自然载体的点分开）。并且，检测是使用这个阈值进行判决，载体是否携带秘密信息。

通用隐写分析方法

- 分类
  - 判定阈值时，自然载体和隐写载体特征不一定能被严格区分，也就是说，一部份隐写载体可能被划分为“自然载体”，反之亦然。
  - 因此，阈值划分是一个最优化的过程，在特定集合情况下，选择误判（虚警，false positive）和漏检（漏警，false negative）尽可能小。
  - 既可用平面分割特征集，也可用曲面分割。前者称为线性SVM，后者称为非线性SVM