**大连理工大学本科毕业设计（论文）**

**基于SVM的LSB信息隐藏算法研究与优化**

**The Study and Improvement of SVM-based LSB steganography**

学 院（系）： 软件学院

专 业： 网络工程

学 生 姓 名： 李欣宜

学 号： 201292241

指 导 教 师：

评 阅 教 师：

完 成 日 期：

大连理工大学

Dalian University of Technology

# 摘 要

注：页眉，居中，宋体，五号，论文中文题目。阅后删除此文本框。

早在信息时代以前，人们就开始研究交换数据而不让其他人察觉的方法，这样的需求催生了隐写技术的研究。而随着科学技术的发展，人们对于秘密隐藏的要求也迅速增加，同时，一系列用于探查可疑文件是否经过隐写操作的隐写分析算法也应运而生，因此，找到可以对抗这些分析方法的安全有效的隐写方法对于隐写技术的发展十分重要。本文针对于隐写技术中最为流行的LSB隐写技术，将信息交流中通信双方无法通过安全信道进行秘钥交换的情况作为使用场景，在此情况下，传统的LSB隐写技术的顺序或均匀隐藏的方法在安全性方面十分薄弱，对此，本文建议使用SVM分类器自适应地选择灰度图像中可以安全隐藏数据的位置进行LSB隐写。本文提出了使用4个能够较为全面地描述图像块、图像和秘密消息三者间关系的特征作为SVM训练和预测时使用的特征集，使用现有的隐写分析算法评估这些样本的安全性作为标签集，使用SVM对图像中一些易于提取的特征来估计隐写的安全性并帮助隐写者作出隐写位置的选择。为了验证优化的实现效果，本研究使用图像数据库设计了一组实验对于嵌入率为5%-50%的伪装图像进行分析，发现相较传统的顺序LSB替换，本文的方法确实可以在低嵌入率的情况下以10%以上的概率提高隐写的安全性。

关键词**：**信息隐藏；LSB嵌入；机器学习; 支持向量机;

**The Study and Improvement of SVM-based LSB steganography**

# Abstract

Before information age, people began to study the ways to exchange information without drawing attention from others, which gave birth to steganography techniques. As science and technology develop, people has much more requirements of data hiding and their requests develop rapidly. At the same time, a series of steganalysis algorithms to detect whether a suspicious file contains secret message or not come to being as time requires. Therefore, it is significant for steganography to find a safe and effective steganography method against those steganalyses. This paper focuses on the most popular steganography method, LSB steganography, and consider the case that there are no available safe channels for both sides in the communication to exchange keys. In this case, traditional consecutive or uniform embedding method in LSB steganography is extremely vulnerable. To solve the problem, this paper proposes to select an image block for embedding in a gray-scale image adaptively with a SVM classifier. In this paper, four features are suggested to comprehensively describe the relationship among image, image block and secret message, which are extracted as feature sets for the training and prediction of SVMs. Then the label sets are generated as the security evaluation results of present steganalysis methods. In the application process, the new scheme uses a SVM to estimate the security with those accessible features, and help people make decisions on the location to hide secret message. To verify the performance of this improvement, the research designs a group of experiments on image datasets to analyze the stego images with their embedding rates range from 5% to 50%. It can be easily found that comparing to traditional LSB replacement, the method in this paper do improve the steganography security by the probability of at least 10% under low embedding rate situations.

**Key Words：**Secret hiding；LSB embedding；Machine learning；Support Vector Machine

目 录

[摘 要 I](#_Toc453003340)

[Abstract II](#_Toc453003341)

[引 言 1](#_Toc453003342)

[1 绪论 2](#_Toc453003343)

[1.1 隐写术的问题背景 2](#_Toc453003344)

[5.1.1 第一节一级题目（黑体，小四，1.5倍行距，段前0.5行） 3](#_Toc453003345)

[5.2 第二节题目 3](#_Toc453003346)

[5.2.1 第二节一级题目 3](#_Toc453003347)

[结 论（设计类为设计总结） 4](#_Toc453003348)

[参 考 文 献 5](#_Toc453003349)

[附录A 附录内容名称 8](#_Toc453003350)

[致 谢 9](#_Toc453003351)

# 引 言

理工文科所有专业本科生的毕业设计（论文）都应有“引言”的内容。如果引言部分省略，该部分内容在正文中单独成章，标题改为文献综述，用足够的文字叙述。从引言开始，是正文的起始页，页码从1开始顺序编排。

针对做毕业设计：说明毕业设计的方案理解，阐述设计方法和设计依据，讨论对设计重点的理解和解决思路。

针对做毕业论文：说明论文的主题和选题的范围；对本论文研究主要范围内已有文献的评述；说明本论文所要解决的问题。建议与相关历史回顾、前人工作的文献评论、理论分析等相结合。

注意：是否如实引用前人结果反映的是学术道德问题，应明确写出同行相近的和已取得的成果，避免抄袭之嫌。注意不要与摘要内容雷同。

书写格式说明：

标题“引言”选用模板中的样式所定义的“引言”；或者手动设置成字体：黑体，居中，字号：小三，1.5倍行距，段后1行，段前为0行。

引言的字数在3000字左右（毕业设计类引言可适当调整为800字左右）。引言正文选用模板中的样式所定义的“正文”，每段落首行缩进2字；或者手动设置成每段落首行缩进2字，宋体，小四，多倍行距 1.25，段前、段后均为0行，取消网格对齐选项。

1 绪论

## 隐写术的问题背景

隐写是指把一个文件、消息、图像或者视频隐藏到另一个文件、消息、图像或者视频的行为。与密码学不同的是，隐写术旨在隐藏消息或其他形式的信息本身的存在，不引起发送方和接收方以外的人的怀疑而完成信息的交流，而密码学则用于隐藏这些信息的内容，使得非发送方或接收方即使截获消息也无法得到所交流的信息的真实内容。隐写术的主要任务是使发生在公共信道上的秘密信息交流不被察觉，隐藏了秘密信息的图片或其他格式的载体与隐藏之前的原始数据在视觉上以及其他几个重要特征一致。

隐写术是一种秘密通信的艺术，这个术语最早在从几千年前开始，人类就着迷于密文书写，并出于多种原因和动机学习这种和研究这种技术[1]，早期密写通常是指密码学，而随着时代发展现在密写也涵括了隐写术，两者用不同的方式实现密写的目的。无论在哪个时代，隐写术广泛应用于各个领域。早期的隐写实践是使用不可见的墨水在信件中书写消息，显而易见地，军事和政治中，在不被敌方察觉的前提下向友方传递信息的能力十分关键。在数字时代，这种思想发展为在多媒体文件中隐藏其他数字消息。现代隐写术在1985年以后随着个人计算机的推广而问世，随着计算机科学技术和数学的发展以及深入研究而迅速发展并投入更广泛的应用，例如电子通信包含了传输层内的隐写码以传送如文档文件或图片文件等多媒体数据。

与密码学相似，隐写术的使用场景决定了它必须满足以下三个要求：

* 保密性：不容易被探查到隐藏消息的存在
* 可获得性：不会出现由于修改数据的载体导致秘密消息的丢失，秘密消息可以被恢复
* 完整性： 其他人无法伪造出错误信息

隐写系统可以被视为加密系统的一个特例[2]，在这个系统中我们要求密文与明文对于其他人来说难以区分。值得注意的是，隐写加密方必须首先合成一个与秘密消息无关的无害的文件。载体合成具有很大的挑战性，高效合成载体的隐写机制很少。对于这个问题，使用通过修改实现隐写的方法很好地避免了这个挑战。这种系统使用一个现有的与秘密消息无关的文件，也就是载体，作为隐写系统的输入的一部分，接下来秘密消息转换为以让人难以觉察的载体的修改，得到与载体极其相似的结果，即伪装（stego）。理论上，载体合成是最自由且强大的隐写方法，因为它可以不受限制地适用于任意的秘密消息。但在实践中，无害文件的合成是一个非常复杂且低效的过程，所以大多数众所周知的实用系统实用的都是修改载体的方法。当然除了合成和修改两种模式以外，载体选择[3]也是可行的方案，可以类比为传统密码学中使用的编码本，有多个载体或密文分别对应特定的秘密消息。然而，这种模式需要的可用载体数量太大，并不是在所有场景下都实用。所以我们在大多数情况下会选择通过修改进行隐写的模式，也就是说隐写的过程需要使用已经存在的文件作为原始的输入，而多媒体文件（如图像、音频和视频等）往往较大，包含了大规模的数据，可以找到足够的空间隐藏消息同时在不表达出可以被察觉的异常效果，是理想的载体。其中，数字图像的应用场景广泛，修改方便，且容易在互联网快速传播，成为了应用最多的载体。本文也将围绕图像隐写技术展开。

早期的隐写安全完全依赖于隐写算法，只将载体和秘密消息作为输入而不使用秘钥，被称为纯隐写系统，一旦隐写算法泄露则整个系统被破解。将秘钥引入作为输入的隐写系统则被称为秘钥隐写系统。在密码学中，加密者和解密者共享秘钥，关于算法的知识对双方区分伪装和正常消息没有帮助，这个结论即kerckhoffs原则，在隐写系统中并不总是成立[4]。符合kerckhoffs原则的系统在使用一个秘钥的实例被攻破后，使用其他秘钥的实例的安全性不受影响，与其他系统相比这样的系统具有巨大的优势，因为生成不同的秘钥恨容易，而重新设计一个算法却很困难。但事实上，基于kerckhoffs原则安全的隐写系统很难设计，同时大多数应用场景下，除了发送方和接收方外其他人对于他们的隐写系统通常一无所知，因此，不遵循kerckhoffs原则的系统也是可行的。

## 1.2 LSB图像隐写技术

对于一个二进制整数来说，最低有效位（LSB）是最低的比特位（即第0位），决定了这个数是奇数还是偶数，这个比特位相比于其他位置的变化对于整个数值变化影响是最小的。

LSB嵌入方法是一种经典的图像隐写算法。这种方法最早被用于像素图像，在像素图像中，每个像素都是代表该点颜色强度的整数。在灰度图像中，每个点的像素值表达了该点色彩介于黑白之间的程度，而在具有三个色彩信道的RGB图像中，每个像素点由三个独立个代表红、绿、蓝三种颜色强度的值合成单元，这些像素值的取值范围通常是8bit的整数值，也就是。在颜色强度上的微小改动被察觉的可能性很小，LSB隐写算法正是利用了像素图像的这个特性，舍弃每个像素原来的最低有效位（LSB），并替换为需要隐藏的消息。接收者得到伪装完成的图像后可以通过模2操作提取新的LSB并将之还原为完整的消息。



最基础的基于LSB的隐写系统从图像的左上角开始逐位嵌入秘密消息，推进的顺序是由发送者和编程语言决定的最自然的方向。为了方便接受者确定隐藏的消息在何处结束，我们可以将消息的大小作为头部隐藏在前像素中，这里的是一个双方已经约定好的整数。



### 5.1.1 第一节一级题目（黑体，小四，1.5倍行距，段前0.5行）

## 5.2 第二节题目

### 5.2.1 第二节一级题目

# 结 论（设计类为设计总结）

结论是理论分析和实验结果的逻辑发展，是整篇论文的归宿。结论是在理论分析、试验结果的基础上，经过分析、推理、判断、归纳的过程而形成的总观点。结论必须完整、准确、鲜明、并突出与前人不同的新见解。

书写格式说明：

标题“结论”选用模板中的样式所定义的“结论”，或者手动设置成字体：黑体，居中，字号：小三，1.5倍行距，段后1行，段前为0行。

结论正文选用模板中的样式所定义的“正文”，每段落首行缩进2字；或者手动设置成每段落首行缩进2字，字体：宋体，字号：小四，行距：多倍行距 1.25，间距：段前、段后均为0行。

参 考 文 献

标题“参考文献”不可省略，选用模板中的样式所定义的“参考文献”；或者手动设置成字体：黑体，居中，字号：小三，1.5倍行距，段后1行，段前为0行。

参考文献内容设置成字体：宋体，字号：五号，多倍行距1.25，段前、段后均为0行，取消网格对齐选项。

参考文献的著录，按论文中引用顺序排列。

参考文献数量不少于10篇，其中期刊不少于5篇，并且包含一定数量的外文期刊。

文献类型标志参考国家标准 GB/T 7714－2005，如下表：

|  |  |
| --- | --- |
| 文献类型 | 标志代码 |
| 普通图书 | M |
| 会议录 | C |
| 汇编 | G |
| 报纸 | N |
| 期刊 | J |
| 学位论文 | D |
| 报告 | R |
| 标准 | S |
| 专利 | P |
| 数据库 | DB |
| 计算机程序 | CP |
| 电子公告 | EB |

按照引用的文献类型不同使用不同的方法，示例如下：

1 普通图书

[1] 广西壮族自治区林业厅.广西自然保护区[M].北京:中国林业出版社,1993.

[2] 蒋有绪,郭泉水,马娟,等.中国森林群落分类及其群落学特征[M].北京:科学出版社,1998.

[3] International Federation of library Association and Institutions. Names of persons: national usages for entry in catalogues [M].3rd ed. London: IFLA International office for UBC, 1977.

2 期刊中析出的文献

[1] 李炳穆.理想的图书馆员和信息专家的素质与形象[J].图书情报工作,2000,(2):5-8.

[2] 陶仁骥.密码学与数学[J].自然杂志,1984,7(7):527.

[3] 亚洲地质图编目组. 亚洲地层与地质历史概述[J].地质学报,1978,3:104-208.

[4] DES MARAIS D J, STRAUSS H , SUMMONS R E, et al. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment [J].Nature ,1992,359:605-609.

3 论文集、会议录

[1] 中国力学学会.第3届全国实验流体力学学术会议论文集[C].天津:[出版者不祥],1990.

[2] ROSENTHALL E M. Proceedings of the Fifth Canadian Mathematical Congress, University of Montreal, 1961[C]. Toronto: University of Toronto Press,1963.

4 专著中析出的文献

[1] 国家标准局信息分类编码研究所.GB/T 2659-1986 世界各国和地区名称代码[S]//全国文献工作标准化技术委员会.文献工作国家标准汇编:3.北京:中国标准出版社,1988:59-92.

[2] 韩吉人.论职工教育的特点[G]//中国职工教育研究会.职工教育研究论文集.北京:人民教育出版社,1985:90-99.

[3] FOURNEY M E. Advances in holographic photoelasticity [C]//American Society of Mechanical Engineers．Applied Mechanics Division．Symposium on Applications of Holography in Mechanics, August 23-25,1971,University of Southern California, Los Angeles, California. New York：ASME,c1971:17-38.

[4] MARTIN G. Control of electronic resources in Australia[M]//PATTLE L W , COX B J. Electronic resources: selection and bibliographic control. New York : The Haworth Press,1966:85-96.

5 学位论文

[1] 张志祥. 间断动力系统的随机扰动及其在守恒律方程中的应用[D].北京:北京大学数学学院,1998.

[2] CALMS R B. Infrared spectroscopic studies on solid oxygen [D]. Berkeley: Univ．of California ,1965.

6 专利文献

[1] 刘加林. 多功能一次性压舌板:中国,92214985.2[P].1993,04,14.

[2] 河北绿洲生态环境科技有限公司.一种荒漠化地区生态植被综合培育种植方法:中国,01129210.5[P/OL].2001-10-24[2002-05-28].http://211.152.9.47/sipoasp/zlijs/hyjs-yx-new.asp?recid=01129210.5& leixin.

[3] KOSEKI A ,MOMOSE H, KAWAHITO M, et al .Compiler :US,828402[P/OL]. 2002-05-25[2002-02-28]. http://FF&p＝1 & u =netahtml/PTO/search-bool.html & r = 5 & f=G& l = 50& col = AND & d =PG01 & sl =IBM .AS. & 0S =AN/IBM & RS =AN/IBM.

7 科技报告

[1] U．S．Department of Transportation Federal Highway Administration. Guidelines for handling excavated acid-producing materials, PB 91-194001[R]. Springfield: U.S. Department of Commerce National Information Service,1990.

[2] World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.

8 报纸中析出文献

[1] 丁文祥. 数字革命与竞争国际化[N].中国青年报,2000 ,11,20(15).

[2] 张田勤. 罪犯DNA库与生命伦理学计划[N].大众科技报,2000,11,12(7).

9 电子文献（包括专著或连续出版物中析出的电子文献）

[1] 江向东.互联网环境下的信息处理与图书管理系统解决方案[J/OL].情报学报, 1999, 18(2):4[2000-01-18]. http://www.chinainfo.gov.cn/periodical/qbxb/qbxb99/qbxb990203.

[2] 萧钰.出版业信息化迈入快车道 [EB/OL]. (2001,12,19)[2002,04,15]. http://www.creader.com/news/20011219/200112190019.html.

[3] METCALF S W. The Tort Hall air emission study[C/OL] //The International Congress on Hazardous Waste, Atlanta Marriott Marquis Hotel, Atlanta, Georgia, June 5-8, 1995: impact on human and ecological health [1998,09,22]. http://atsdrl.atsdr.cdc.gov:8080/cong95.html.

[4] TURCOTTE D L. Fractals and chaos in geology and geophysics[M/OL]. Mew York: Cambridge University Press, 1992[1998,09,23]. http://www.seg.org/reviews/mccorm30.html.

# 附录A 附录内容名称

以下内容可放在附录之内：

（1） 正文内过于冗长的公式推导；

（2） 方便他人阅读所需的辅助性数学工具或表格；

（3） 重复性数据和图表；

（4） 论文使用的主要符号的意义和单位；

（5） 程序说明和程序全文

（6） 调研报告。

这部分内容可省略。如果省略，删掉此页。

书写格式说明：

标题“附录A 附录内容名称”选用模板中的样式所定义的“附录”；或者手动设置成字体：黑体，居中，字号：小三，1.5倍行距，段后1行，段前为0行。

附录正文选用模板中的样式所定义的“正文”，每段落首行缩进2字；或者手动设置成每段落首行缩进2字，字体：宋体，字号：小四，行距：多倍行距 1.25，间距：段前、段后均为0行。

# 致 谢

毕业设计（论文）致谢中不得书写与毕业设计（论文）工作无关的人和事，对指导老师的致谢要实事求是。

对其他在本研究工作中提出建议和给予帮助的老师和同学，应在论文中做明确的说明并表示谢意。

这部分内容不可省略。

书写格式说明：

标题“致谢”选用模板中的样式所定义的“致谢”；或者手动设置成字体：黑体，居中，字号：小三，1.5倍行距，段后1行，段前为0行。

致谢正文选用模板中的样式所定义的“正文”，每段落首行缩进2字；或者手动设置成每段落首行缩进2字，字体：宋体，字号：小四，行距：多倍行距 1.25，间距：段前、段后均为0行。