

# 基于 SPAM 和特征优化的通用隐写分析算法改进

杨雪<sup>1</sup>, 杨榆<sup>2</sup>, 雷敏<sup>2</sup>

(1. 北方工业大学电子信息工程学院, 北京 100144; 2. 北京邮电大学信息安全中心, 北京 100876)

**摘要:**通用隐写分析特征的高维化趋势加剧,导致算法时间复杂度和空间复杂度急速上升。因此,研究如何在维持检测率水平的同时降低特征维度,对隐写分析的实用性有重要意义。本研究通过主成分分析确定特征矢量的最优维度;借用费希尔线性判别式思想,以“类内聚合”和“类间离散”程度评价各维特征区分自然和隐写载体的能力,进而选取最优子集。分析针对主流通用检测模型的基础——SPAM 模型进行,仿真实验证明,优化子集具有良好检测性和较低计算复杂度。

**关键词:**通用隐写分析;SPAM;Fisher score;维度规约

**中图分类号:**TP309.2

**文献标志码:**A

## 1 研究背景

隐写分析指对截获的信号进行分析,从而判断其是否含有秘密信息的技术<sup>[1]</sup>,可分为两类:专用隐写分析和通用隐写分析<sup>[2]</sup>。考虑到当前隐写技术的多样性,通用隐写分析明显适应性更强,因而越来越受到关注。作为目前典型的隐写分析方法之一,文献[3]提出的 SPAM 模型受到了学术界的广泛关注。文献[4]针对一种新的隐写方法自适应像素匹配(Adaptive Pixel Pair Matching, APPM)<sup>[5]</sup>,在 SPAM 模型的基础上利用二阶马尔可夫链,提出了二次差分模型,并从水平、垂直、对角和反对角 4 个方向提取特征。文献[6]在 SPAM 模型的基础上,提出了差分像素联合分布特征,并结合图像内容分割对算法进行了改进。文献[7]组合了 SPAM 特征和从 DCT 域量化系数中提取的 548 维特征,构成了一个 1234 维的特征向量,实现了变换域的通用隐写分析。文献[8]在 SPAM 模型的基础上增加平行差分像素,并使用了一种新的邻接矩阵,减少了约 3/4 的特征维数,提升了算法的检测性能。因此,分析并优化 SPAM 模型对隐写分析的实用性有重要价值。

随着隐写算法的迅速发展,各种高性能的隐写分析算法也不断涌现。特征维数高、分类器培训时间长、检测效率低等,是目前大多数算法所存在的问题。很多情形下,隐写分析算法所提取的高维特征集中,特征彼此之间有一定的相关关系,可以解释为其中一些特征是冗余的,对检测性能的贡献有一定的重叠。因此,选择具有良好可分性的特征,从而减少特征空间维数,对提高隐写分析的效率和实用性具有很大意义。图 1

是研究方案基本框图。

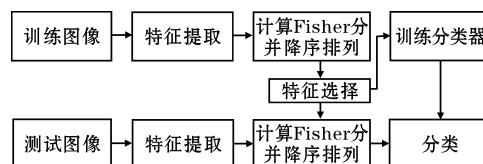


图1 隐写分析流程框图

## 2 研究基础

### 2.1 SPAM 模型简介

文献[3]提出差分像素邻接矩阵(subtractive pixel adjacency matrix, SPAM)特征实现对 LSB 匹配算法的有效检测。作为目前性能最好的隐写分析方法之一,在其基础上进行改进,通过降低特征维数的方法,来提高 SPAM 特征的通用检测性能。本节对 SPAM 特征进行介绍。

对于一幅给定的尺寸为  $W \times H$  的图像,令其像素矩阵为  $I$ ,则图像的 SPAM 特征计算步骤如下(其中  $\rightarrow$  代表水平正向,  $\leftarrow$  代表水平反向,  $\downarrow$  代表垂直正向,  $\uparrow$  代表垂直反向,  $\searrow$  代表对角正向,  $\swarrow$  代表对角反向,  $\nearrow$  代表反对角正向,  $\nwarrow$  代表反对角反向)。

首先,沿着 8 个方向计算相邻像素差值矩阵。以水平正向为例进行说明,其他方向使用相同的规则。定义水平正向(从左至右)方向的差分矩阵为

$$D_{i,j}^{\rightarrow} = I_{i,j} - I_{i+1,j} \quad (1)$$

其中  $i=1, \dots, W, j=1, \dots, H$ 。其他 7 个方向的差分矩阵分别记为  $D_{i,j}^{\leftarrow}, D_{i,j}^{\downarrow}, D_{i,j}^{\uparrow}, D_{i,j}^{\searrow}, D_{i,j}^{\swarrow}, D_{i,j}^{\nearrow}, D_{i,j}^{\nwarrow}$ 。

其次,根据差分矩阵  $D_{i,j}^{\rightarrow}$  提取 Markov 特征,定义

一阶概率转移矩阵为

$$M_{u,v}^{\rightarrow} = P(D_{i,j+1}^{\rightarrow} = u \mid D_{i,j}^{\rightarrow} = v) \quad (2)$$

其中  $u, v \in \{-T_1, -T_1 + 1, \dots, T_1 - 1, T_1\}$ 。定义二阶概率转移矩阵为

$$M_{u,v,w}^{\rightarrow} = P(D_{i,j+2}^{\rightarrow} = u \mid D_{i,j+1}^{\rightarrow} = v, D_{i,j}^{\rightarrow} = w) \quad (3)$$

其中  $u, v, w \in \{-T_2, -T_2 + 1, \dots, T_2 - 1, T_2\}$ 。

其中,  $T_1$  和  $T_2$  为特征统计时为了达到降低维数的目的, 对统计差值设置的阈值。同理可以定义其他 7 个方向的一阶、二阶概率转移矩阵。

最后, 将转移概率矩阵转换为一维向量, 并将水平和垂直方向的 4 个结果平均, 并将结果矩阵转换为一维向量, 即

$$F_{1,\dots,k} = \frac{1}{4}(M^{\rightarrow} + M^{\leftarrow} + M^{\downarrow} + M^{\uparrow})$$

$$F_{k+1,\dots,2k} = \frac{1}{4}(M^{\nearrow} + M^{\nwarrow} + M^{\swarrow} + M^{\searrow}) \quad (4)$$

分别对一阶和二阶概率转移矩阵求取相应阶数的 SPAM 特征。对于一阶 SPAM 特征,  $k = (2T_1 + 1)^2$ , 对于二阶 SPAM 特征,  $k = (2T_2 + 1)^3$ 。

文献[3]得出使用二阶 SPAM 特征比使用一阶 SPAM 特征可以得到更高的检测率, 因此文中的特征优化主要针对二阶 SPAM 特征进行研究。

## 2.2 特征优化算法

通用隐写分析算法中, 使用的特征集维数较大, 其中可能存在不相关的特征, 特征之间也可能存在相互依赖, 容易导致如下的后果: 特征个数越多, 分析特征、训练模型所需的时间就越长; 特征个数越多, 容易引起“维数灾难”<sup>[9]</sup>, 模型也会越复杂, 对检测效率产生严重影响。特征优化能剔除不相关或冗余的特征, 从而达到减少特征个数, 提高模型精确度, 减少运行时间的目的。

PCA (principal component analysis) 与 FLD (fisher linear discriminant) 作为常用的降维技术, 其侧重点各有不同。

PCA 的基本思想是在尽可能好的代表原特征的情况下, 将原特征用数目更少的特征取代, 新特征是旧特征的线性组合, 这些线性组合最大化样本方差, 尽量使新特征互不相关。

FLD 的基本思想是将高维的样本投影到最佳鉴别矢量空间, 以达到抽取分类信息和压缩特征空间维数的效果, 投影后保证样本在新的子空间有最大的类间距离和最小的类内距离, 即特征在该空间中有最佳的可分离性。

在隐写分析过程中, 需要同时考虑自然图像和隐写图像两类样本的特征优化。PCA 将所有的样本(特

征向量集合) 作为一个整体对待, 去寻找一个均方误差最小意义下的最优线性映射投影, 而忽略了类别属性, 而它所忽略的投影方向有可能刚好包含了重要的可分性信息。从这方面来讲, 用 PCA 对特征进行优化是不充分的。而 FLD 对特征空间维数进行了压缩, 但是在样本投影的过程中忽略了向量方向, 因此, 用 FLD 对特征进行优化也是不充分的。在此基础上, 通过结合 PCA 与 FLD 的基本思想, 对特征进行分析, 从而寻找最优特征子集对 SPAM 模型进行优化。

## 3 基于 Fisher 分的 SPAM 算法优化

选取 FLD 作为降维依据, 以“类内聚合”和“类间离散”程度评价各维特征区分自然载体和隐写载体的能力, 进而选取最优子集。同时, 为了确定最优子集的维数, 通过 PCA 确定特征矢量的最优维度。

### 3.1 基于 Fisher 分的特征选择

Fisher 准则<sup>[10]</sup>是特征选择的有效方法之一, 其主要思想是鉴别能力较强的特征表现为类内距离尽可能小, 类间距离尽可能大。这里采用单个特征的 Fisher 分作为准则, 对特征进行排序, 选出鉴别性能较强的特征, 从而达到降维的目的, 并得到较优的识别性能。某个特征的 Fisher 分值越大, 说明该特征的分类能力越强。本节对 Fisher 分的计算过程进行介绍。

考虑训练样本  $X = (C_1, S_1), (C_2, S_2), \dots, (C_m, S_m)$ ,  $C_i, S_i \in R^k, i = 1, 2, \dots, m$ , 其中  $m$  为样本数量,  $k$  为特征向量维数;  $C_i$  表示自然图像特征,  $S_i$  表示隐写图像特征。将  $X$  中自然图像特征样本集合记为  $X_1$ , 个数记为  $N_1$ ; 隐写图像特征样本集合记为  $X_2$ , 个数记为  $N_2$ 。则 Fisher 分值定义为

$$F = \frac{S_b}{S_w} \quad (5)$$

式(5)中:  $S_b$  为类间离散度, 描述两类样本间的距离;  $S_w$  为类内离散度, 描述同类样本间的距离。其定义如下

$$S_b = (\bar{m}_1 - \bar{m}_2)^2 \quad (6)$$

$$S_w = \delta_1^2 + \delta_2^2 \quad (7)$$

式(6)和式(7)中:  $\bar{m}_1$  为自然图像特征样本均值;  $\bar{m}_2$  为隐写图像特征样本均值;  $\bar{m}$  为所有样本的均值。

其中,  $\bar{m}_1 = \frac{1}{N_1} \sum_{X \in X_1} X$ ,  $\bar{m}_2 = \frac{1}{N_2} \sum_{X \in X_2} X$ ;  $\delta_1^2$  和  $\delta_2^2$  分别为自然图像特征样本和隐写图像特征样本的方差, 其中,  $\delta_1^2 = \frac{1}{N_1} \sum_{X \in X_1} (X - \bar{m}_1)^2$ 。则第  $r$  个特征的 Fisher 分值可表示为

$$F_r = \frac{S_{b,r}}{S_{w,r}} = \frac{(\bar{m}_{1,r} - \bar{m}_{2,r})^2}{\delta_{1,r}^2 + \delta_{2,r}^2} \quad (8)$$

式(8)中,和分别表示自然图像特征和隐写图像特征中第  $r$  个特征属性的均值。

### 3.2 优化前后特征分析

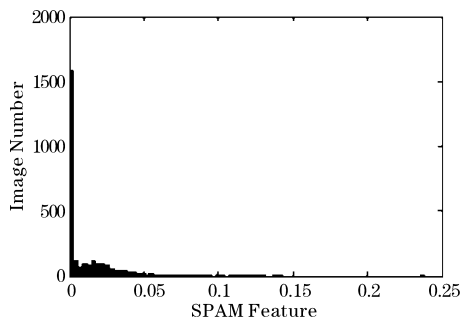
考虑到图像的 SPAM 特征值相差较大,为解决不同特征的取值范围相差太大而可能出现的“大数吃小数”现象,不忽略每一维特征对 Fisher 分计算的贡献,在提取图像特征后,对 SPAM 特征进行归一化处理。

#### 3.2.1 Fisher 分的分类性能分析

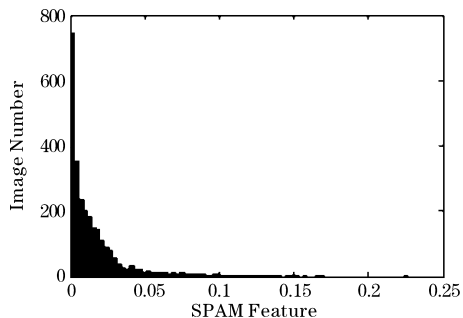
为说明 Fisher 分的分类性能,选择 Fisher 分最高和最低的特征,通过绘制直方图对特征的检测性能进行分析。

随机选取 2700 张自然图像和 2700 张隐写图像,提取 SPAM 特征,计算每一维特征的 Fisher 分,对其进行降序排列。分别绘制出 Fisher 分最高和最低的自然图像 SPAM 特征和隐写图像 SPAM 特征对应的直方图,仿真结果如图 2 所示。

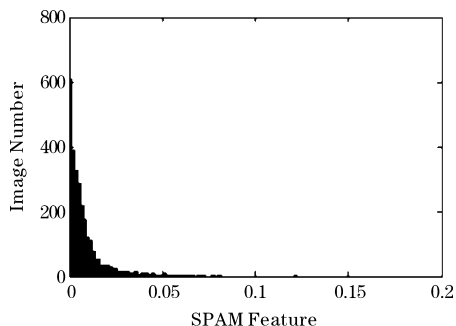
图 2 中的(a)、(b)表示 Fisher 值最大的自然特征和隐写特征所对应的直方图,(c)、(d)表示 Fisher 值最小的自然特征和隐写特征所对应的直方图。可以看出,图 2(a)、(b)两图有很大的区别,图 2(c)、(d)两图差别很小。而自然和隐写载体的直方图就是对其概率密度函数的近似,两分布离得越远,对应特征的检测性能越好。因此,可以说明某个特征的 Fisher 分越大,该特征的分类能力越强。



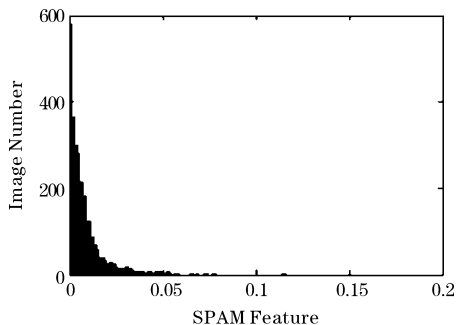
(a) Fisher 值最大的自然特征



(b) Fisher 值最大的隐写特征



(c) Fisher 值最小的自然特征



(d) Fisher 值最小的隐写特征

图 2 不同 Fisher 值所对应特征的直方图

#### 3.2.2 基于 PCA 的特征优化分析

为说明 SPAM 模型提取的高维特征集中,特征之间存在冗余,以 686 张图像的 SPAM 特征组成方阵,计算方阵的特征值与特征向量,并进行分析,具体步骤如下。

(1) 随机选取 686 张图像,提取 SPAM 特征,组成一个 686686 的方阵(方阵的每一行代表一张图像的特征)。计算该方阵的特征值和特征向量。

(2) 假设该矩阵共有  $K$  个特征值  $V$ ,依次计算前  $m$  ( $m < K$ ) 个特征值之和与  $K$  个特征值之和的比值,画出斜坡图。

(3) 计算特征向量之间的夹角,画出散点图。

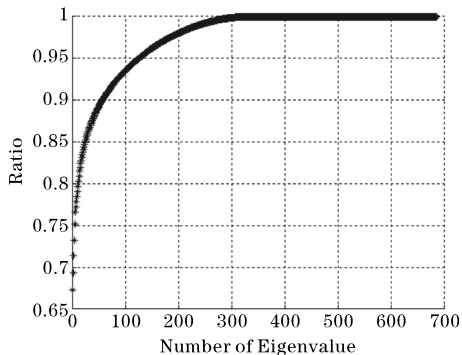


图 3 特征值比值斜坡图

图 3 为特征值比值的斜坡图,可以看出,随着特征的增加,比值逐渐趋向于 1,说明 300 维左右之后的特



征值信息量小到可以忽略,因此,特征集存在很大的优化空间。通过拟合斜坡图的曲线,计算曲线拐点,有助于寻找最优优化维度。

图 4 为固定的一个特征向量分别与其他 685 个特征向量之间夹角的散点图,可以看出,很多夹角都接近,说明大部分 SPAM 特征之间相关性很强,可以选取特征集的子集对原始特征进行优化。

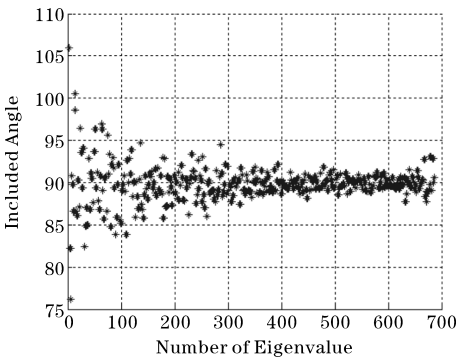


图 4 特征向量夹角散点图

4 实验仿真

4.1 数据预处理

由于不同隐写图像提取的 SPAM 特征有较大区别,为减小不同算法之间的影响,得到较为准确的 Fisher 分值,实验采取分别计算不同隐写算法得到的 Fisher 分,然后求取平均值的方法。具体步骤为:提取不同隐写图像及其自然图像的 SPAM 特征,分别计算两者之间的 Fisher 分,将得到的多组结果取均值,然后将该组均值按照从大到小进行排序。根据排序后的 Fisher 分值,对原 SPAM 特征进行排列,得到新的特征集。在实验阶段,根据前面 PCA 分析所得到的最优维数对新特征集进行降维,寻找最优优化维度。

实验所用电脑配置为 Intel (R) Core (TM) CPU (2.30GHz)、6.00GBRAM,并使用 MATLAB R2012a 进行仿真实验。

对 LSB<sup>[11]</sup>、LSBM (LSB Matching)<sup>[12]</sup> 和 EA (Edge Adaptive)<sup>[13]</sup> 算法分别用提出的通用分析算法进行仿真测试,其中,EA 为自适应隐写算法。实验使用 Boss-Base-1.01 图像库<sup>[14]</sup>,库中包括 10000 张 256KB 的未经过 JPEG 压缩的灰度 PGM 图像,大小均为 512512,图像内容涉及自然风光、人物和建筑等范围。用 LSB、LSBM 和 EA 算法分别以 0.2bpp (bits per pixel)、0.4bpp 和 0.6bpp 的嵌入率生成隐写图像。从隐写图像库中分别随机抽取 300 幅图像,共 33300=2700 张隐写图像组成的混合图像用于训练,在剩下的隐写图像中随机抽取 2700 幅图像用于测试。采用集成分类器

(Ensemble Classifier)<sup>[15]</sup> 进行分类。

表 1 算法优化前后性能对比

隐写 算法	嵌入率/bpp	SPAM(686)		优化后(300)	
		错误检测率	时间/s	错误检测率	时间/s
LSB	0.2	0.0577	17.38	0.0462	7.98
	0.4	0.0321	17.04	0.0275	6.74
	0.6	0.0303	16.26	0.0256	6.64
LSBM	0.2	0.2383	16.06	0.2269	7.88
	0.4	0.0619	15.56	0.0625	6.98
	0.6	0.0463	17.68	0.0406	7.24
EA	0.2	0.4466	20.88	0.4761	7.20
	0.4	0.2909	16.90	0.3150	6.12
	0.6	0.1281	12.59	0.1369	6.06
均值		0.1480	16.7056	0.1508	6.9822

4.2 仿真测试与结果分析

为检测优化后算法的性能,分别用文中算法与 SPAM 算法(计算二阶马尔科夫转移矩阵, $T=3$ )在相同的条件下对 LSB、LSBM 和 EA 算法进行仿真测试。为了保证实验数据的可信性,测试 10 次取平均值。平均错误检测率、训练分类器所用的平均时间以及检测算法的特征维数见表 1。

从表 1 中可以看出,对 SPAM 特征进行优化后训练分类器所用时间大幅度缩短,算法的检测效率得到了很大提升,同时,算法的错误检测率基本保持不变。

4 结束语

针对目前典型的通用隐写分析方法之一,SPAM 模型进行特征优化分析,通过计算特征的 Fisher 分值,减少特征维数。实验结果证明,与 SPAM 算法相比,文中算法减少原 SPAM 算法检测时间的 58.20%,同时,算法的错误检测率只提升原 SPAM 算法的 1.89%。因此,对 SPAM 特征进行优化后,大幅度提升算法的检测效率,增强算法的实用性。

由于在进行特征优化时只考虑了贪婪算法,忽略了特征方向对实验结果的影响。因此,在下一步工作中,考虑结合 PCA 与 FLD 对隐写分析算法进行改进。利用 PCA 的降维思想,从特征集中遍历选取多个特征组合,结合 Fisher 分的分类性能,寻找最优特征子集。

参考文献:

[1] 黄炜,赵险峰,冯登国,等. 基于主成分分析进行特征融合的 JPEG 隐写分析[J]. 软件学报, 2012, 23(7): 1869-1879.

[2] Fridrich J, Goljan M. Practical steganalysis of dig-

- ital images—state of the art [C]. Security and Watermarking of Multimedia Contents, USA: SPIE 2001: 1–13.
- [3] Pevny T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix [J]. IEEE Transactions on Information Forensic and Security, 2010, 5(2): 215–224.
- [4] 侯瑜. 图像隐写分析[D]. 北京: 北京交通大学, 2015.
- [5] Hong W, T S Chen. A novel data embedding method using adaptive pixel pair matching [J]. IEEE Transaction on Information Forensics and Security, 2012, 7(1): 176–184.
- [6] 汪然, 许漫坤, 平西建, 等. 基于分割的空域图像隐写分析[J]. 自动化学报, 2014, 40(12): 2936–2943.
- [7] Kodovsky J, Pevny T, Fridrich J. Modern Steganalysis Can Detect YASS [C]. Proc. Of Conf. on Electronic Imaging, Media Forensics and Security. San Jose, USA: SPIE, 2010.
- [8] H Zhang, XJ Ping, XU Mankun, et al. Steganalysis by subtractive pixel adjacency matrix and dimensionality reduction [J]. Science China Information Sciences, 2014, 57(4): 1–7.
- [9] Veenu Bhasin, Punam Bedi, Anuradha Singhal: Feature selection for steganalysis based on modified stochastic diffusion search using fisher score [J]. International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2014: 2323–2330.
- [10] ZHANG Xue-Qin, GU Chun-Hua. A method to extract network intrusion detection feature [J]. Journal of South China University of Technology (Natural Science Edition). 2010, 38(1): 81–86.
- [11] R G van Schyndel, A Z Tirkel, C. F. Osborne. A digital watermark [J]. The IEEE International Conference on Image Processing. Austin, Texas, USA. 1994, (2): 86–90.
- [12] Sharp T. An Implementation of Key-based Digital Signal Steganography [C]. // In Proceedings of the 4th International Workshop on Information Hiding, London, UK, 2001, 13–26.
- [13] W Luo, F Huang, J Huang. Edge adaptive image steganography based on LSB matching revisited [J]. IEEE Trans. Inform. Forensics Security, 2010, 5(2): 201–214.
- [14] T Filler, T Pevny, P Bas. BOSS (Break Our Steganography System) [EB/OL]. Available: <http://agents.fel.cvut.cz/stegodata/BossBase-1.01-cover.tar.bz2>, 2014.
- [15] J Kodovsky, J Fridrich, V Holub. Ensemble classifiers for steganalysis of digital media [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2): 432–444.

## Improvement of Universal Steganalysis based on SPAM and Feature Optimization

YANG Xue<sup>1</sup>, YANG Yu<sup>2</sup>, LEI Min<sup>2</sup>

(1. Electronic Information Engineering Institute, North China University of Technology, Beijing 100144, China; 2. Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** The tendency for high-dimension of universal steganalysis characteristics toward intensifying, and lead to the rapid rise in complexity of algorithm in time and space domain. So maintain the level of detection rates, and reduce the dimension of features at the same time, have significance in research of steganalysis. This paper determine the optimal dimension of feature vectors by principal component analysis; using the concept of Fisher linear discriminant, with the degree of "aggregations within class" and "discreteness between classes" to evaluate the ability of each dimension feature to distinguish natural and hidden carrier, and then select the optimal subset. The analysis directs at the mainstream universal steganalysis model—SPAM model, and the simulation results show that optimal subset has a good detection and low computational complexity.

**Key words:** universal steganalysis; SPAM; fisher score; statute of the dimension