第34卷 第10期 计 算 机 工 程 2008年5月
Vol.34 No.10 Computer Engineering May 2008
·安全技术· 文章编号：1000—3428(2008)10—0153—02 文献标识码：A 中图分类号：TP309

基于ICA模型的主动隐写分析

徐 波，王嘉祯，张政保，刘晓芹
(军械工程学院计算机工程系，石家庄 050003)

摘 要：提出一种基于独立成分分析(ICA)模型的主动隐写分析方案。该方案假设秘密信息是独立同分布序列且统计独立于载体图像，将隐写分析过程视为 ICA 模型的求解问题。借助于最大后验概率估计器，该方案仅使用一幅隐写图像即能提取出秘密信息，克服了Chandramouli所提方案的局限性。仿真实验结果表明，该方案能提取大约80%的秘密信息，且性能随嵌入长度的增加而提高。

关键词：主动隐写分析；独立成分分析模型；最大后验概率估计器

Active Steganalysis Based on ICA Model

XU Bo, WANG Jia-zhen, ZHANG Zheng-bao, LIU Xiao-qin
(Department of Computer Engineering, Ordnance Engineering College, Shijiazhuang 050003)

【Abstract】 A new active steganalysis scheme which only uses one copy of stego image is presented. This paper views active steganalysis as an Independent Component Analysis(ICA) problem under the assumption that embedded secret message is an independent, identically distributed (i.i.d) random sequence and independent to cover image. With only one copy of stego image, it adoptes Maximum Posteriori(MAP) estimator to obtain an estimation of cover image and generates another two signals with the estimated version. All the three signals are as input signals of ICA algorithm. The proposed scheme overcomes the constraint of Chandramouli's method which needs two copies of stego image. Experimental results show that the proposed method achieves acceptable performance and improves its performance with larger message length.

【Key words】 active steganalysis; Independent Component Analysis(ICA) model; Maximum Posteriori(MAP) estimator

1 概述

隐写术(steganography)通过在载体(图像、视频或音频文件)上嵌入秘密信息以达到隐藏信息的目的。它的一个重要特性是不可感知性，虽然载体中嵌入了机密信息，但是并不影响载体的视觉或听觉效果。隐写分析(steganalysis)就是研究如何通过某种检测方法发现隐藏于载体的信息，目的是破坏信息隐藏系统。隐写分析可分为被动分析和主动分析。被动分析的目的只是检查通信中是否含有隐蔽信息；主动分析的目的则是试图提取、删除、篡改或覆盖部分或全部隐藏信息，使隐藏系统无法正常工作。

目前的研究主要集中在被动分析技术[1-2]上面，对主动分析技术研究较少。许多研究者仅针对某种特定的隐写算法或某类隐写算法进行分析[3]，一般只能估计出秘密信息的长度或位置。文献[4]提出了一个主动分析的通用数学模型，使用独立成分分析(Independent Component Analysis, ICA)方法求解，实验结果表明能提取大约95%的秘密信息。但是该模型需要两幅隐写后的图像，这2幅图像是由同样的原始载体、秘密信息和密钥采用同样的隐写算法而得到的，唯一不同的是嵌入强度。这个要求在现实环境中很难达到，因为隐写分析人员一般只能得到一幅隐写后的图像。

为了更切合现实应用，本文提出一种只使用一幅隐写图像的主动分析方案。该方案沿用Chandramouli的ICA模型，使用最大后验概率(Maximum Posteriori, MAP)估计器估计出一幅载体图像，并生成另外2幅图像一起作为ICA算法的输入信号，提取出秘密信息。

2 主动隐写分析的ICA模型

$\{s(k)\}$, $\{w(k)\}$和$\{z(k)\}$分别指原始图像、秘密信息载体和隐写后的图像，$s(k)$, $w(k)$和$z(k)$分别指它们的第k个采样。一般假设$\{s(k)\}$是独立同分布(i.i.d)的随机序列，$\{w(k)\}$同样是独立同分布随机序列，且统计独立于$\{s(k)\}$，则隐写过程可以表示为

$$z(k)=s(k)+\alpha w(k), k=1,2,\cdots,N \tag{1}$$

其中，α 是嵌入强度，可以根据不可见性和鲁棒性要求进行调整。

如果隐写分析人员可以得到 2 幅隐写图像，$\{z_1(k)\}$和$\{z_2(k)\}$，那么可将式(1)写成矩阵形式：

$$\underbrace{\begin{pmatrix}\mathbf{z}_1(k)\\ \mathbf{z}_2(k)\end{pmatrix}}_{\boldsymbol{z}}=\underbrace{\begin{pmatrix}1 & \alpha_1\\ 1 & \alpha_2\end{pmatrix}}_{\boldsymbol{A}}\underbrace{\begin{pmatrix}s(k)\\ w(k)\end{pmatrix}}_{\boldsymbol{r}} \tag{2}$$

因此，可以把主动隐写分析过程视为盲源分离问题：已知观测信号$\{z_1(k)\}$和$\{z_2(k)\}$，估计统计独立的原始信号$\{s(k)\}$和$\{w(k)\}$。当$\alpha_1\neq\alpha_2$时，$\boldsymbol{A}$是满秩矩阵，可使用ICA方法求解。更详细的内容见参考文献[4]。

3 ICA模型的改进

在实际情况下，分析人员一般只能得到一幅隐写图像，很难得到符合ICA模型要求的2幅图像，因此有必要对该模型进行改进，放宽条件，使之更切合实际。思路是：使用估计理论从$\{z(k)\}$中估计出一个$\{s(k)\}$的近似版本，使用这个版本生成新的观测信号，作为ICA算法的输入信号。

基金项目：河北省科技厅基金资助项目(05213579)
作者简介：徐 波(1976－)，男，工程师、博士研究生，主研方向：信息隐藏；王嘉祯，教授、博士生导师；张政保，副教授；刘晓芹，硕士研究生
收稿日期：2007-05-25 **E-mail：**xubohxl@hotmail.com

3.1 模型的前提条件

在嵌入操作前，一般需要对秘密信息的比特序列进行加密，常用的方法是使用一个随机序列进行调制，得到秘密信息载体$\{w(k)\}$。本研究的目的是从一幅隐写图像$\{z(k)\}$中提取出秘密信息载体，即加密后的随机序列。至于从提取的序列中解调出秘密信息的比特序列，这属于加解密的范畴，不是本文的研究目标。为了能够定性地衡量提取的性能，本文对加解密部分进行了简化，假设了一个简单的应用场景，即调制和解调时，秘密信息载体的正值表示比特 1，负值代表比特 0。这样做的好处是只需要提取出秘密信息载体的符号即可。如何提取出信号的幅值是本课题下一步的研究目标。

同文献[4]类似，这里同样假设秘密信息载体$\{w(k)\}$是 i.i.d 随机序列，且统计独立于$\{s(k)\}$。为简化推导和计算，不妨假设 $w(k)$服从零均值高斯分布。

需要指出的是，式(1)所表示的隐写过程在空域和变换域是等价的。例如，对于离散余弦变换(DCT)，设$\boldsymbol{s}_0$是原始图像的像素值矩阵：

$\boldsymbol{s}$=DCT($\boldsymbol{s}_0$)

当在 DCT 域嵌入信息时，隐写图像的 DCT 系数矩阵是：

$\boldsymbol{z}=\boldsymbol{s}+\alpha w$

则隐写图像的像素值矩阵为

$$\boldsymbol{z}_0=\mathrm{IDCT}(\boldsymbol{z})=\mathrm{IDCT}(\boldsymbol{s}+\alpha w)=\boldsymbol{s}_0+\mathrm{IDCT}(\alpha w)=\boldsymbol{s}_0+\alpha w$$

其中，IDCT 指反离散余弦变换。

因此，本文只关注在 DCT 域中嵌入秘密信息。但是文中提出的方法同样可应用于空域和其他变换域。

3.2 MAP 估计器

从$\{z(k)\}$中抽取$\{s(k)\}$是一个典型的估计理论问题，有许多方法可以达到这个目的，这里选择 MAP，因为 MAP 估计器推导简单、计算量小。

经典的 MAP 估计器由式(3)给出：

$$\hat{s}=\arg\max_{s}\{\ln p_w(\boldsymbol{z}\mid\boldsymbol{s})+\ln p_s(\boldsymbol{s})\} \tag{3}$$

因此，估计 $\hat{S}$ 依赖于$\{\boldsymbol{s}(k)\}$和$\{w(k)\}$的概率分布。

由假设可知，$w(k)$～Gaussian(0, δ_n^2)，即

$$p_w(w)=\frac{1}{\delta_n\sqrt{2\pi}}\exp(-\frac{w^2}{2\delta_n^2}) \tag{4}$$

在图像处理中，一般认为自然图像的 DCT 系数服从 Laplacian 分布，这里也同样采用这个假定，即

$$p_s(\boldsymbol{s})=\frac{1}{\sqrt{2}\delta}\exp(-\frac{\sqrt{2}|\boldsymbol{s}|}{\delta}) \tag{5}$$

将式(4)和式(5)代入式(3)中可得到

$$\hat{s}=\mathrm{sign}(z)(|z|-\frac{\sqrt{2}\delta_n^2}{\delta})_+ \tag{6}$$

这里$(y)_+$ 定义为

$$(y)_+=\begin{cases}0 & y<0\\ y & \text{otherwise}\end{cases} \tag{7}$$

3.3 改进的模型

使用上述的 MAP 估计器得到了$\{s(k)\}$的估计版本$\{\hat{s}(k)\}$。现在有 2 个信号$\{z(k)\}$和$\{\hat{s}(k)\}$。接着生成一个信号$\{f(k)\}$，这个信号可以是随机信号，也可以是一个特定的信号。将$\{\hat{s}(k)\}$和$\{f(k)\}$按照式(1)进行线性混合，强度系数分别为 β 和 λ，这样就得到了 3 个信号：

$$z_1(k)=\hat{s}(k)+\alpha w(k) \tag{8}$$

$$z_2(k)=\beta_1\hat{s}(k)+\lambda_1 f(k) \tag{9}$$

$$z_3(k)=\beta_2\hat{s}(k)+\lambda_2 f(k) \tag{10}$$

其中，$\beta_i+\lambda_i=1$, $i=1,2$。

因此可以将式(2)表示为

$$\underbrace{\begin{pmatrix}z_1(k)\\ z_2(k)\\ z_3(k)\end{pmatrix}}_{\boldsymbol{Z}}=\underbrace{\begin{pmatrix}1 & \alpha & 0\\ \beta_1 & 0 & \lambda_1\\ \beta_2 & 0 & \lambda_2\end{pmatrix}}_{\boldsymbol{A}}\underbrace{\begin{pmatrix}\hat{s}(k)\\ w(k)\\ f(k)\end{pmatrix}}_{\boldsymbol{r}} \tag{11}$$

如果信号 $f(k)$和强度系数 β, λ 满足以下 3 个条件：

(1) $\{\hat{s}(k)\}$, $\{w(k)\}$和$\{f(k)\}$统计独立；

(2)$\{f(k)\}$是非高斯分布的信号；

(3)矩阵 $\boldsymbol{A}$ 列满秩。

就可以使用 ICA 算法来求解式(11)。

ICA 算法的目的是找出一个分离矩阵 $\boldsymbol{W}$，使估计出的源信号之间统计相关性最小。可以从下式得到分离的原始信号：

$$\hat{r}=\boldsymbol{W}\boldsymbol{Z} \tag{12}$$

必须指出的是，由于 ICA 算法的固有缺陷，因此分离出的信号顺序并不确定。但是，对于分析人员来说，这并不存在什么问题，因为$\{\hat{s}(k)\}$和$\{f(k)\}$都是已知的。

4 仿真实验结果

笔者将提出的方法分别应用到空域、DCT 域和 DWT 域的加性扩频(SS)隐写，并在 MATLAB7.0 中进行仿真实验。测试图片采用标准 Lena 图像(256×256×8 bit 灰度图像)，秘密信息载体使用系统函数生成 $N(0, 1)$的高斯随机序列，长度为 L。采用式(1)分别在空域、DCT 域或 DWT 域中进行嵌入操作，操作的对象分别为像素值、DCT 系数和 DWT 系数。设定 α=0.1。具体嵌入方法如下：

(1)空域嵌入：将秘密信息载体嵌入到前 L 个像素中；

(2)DCT 域嵌入：将 Lena 图像进行 DCT 变换，去掉直流成分，将秘密信息载体嵌入到最大的 L 个 DCT 系数中；

(3)DWT 域嵌入：将 Lena 图像进行一层小波变换，将秘密信息载体嵌入到低频子带的最大的 L 个系数中。

使用FASTICA[5]算法来提取秘密信息载体。由于本文所推导的MAP估计器是一种典型的图像去噪技术，因此直接相减($z(k)-\hat{s}(k)$)同样可以提取出秘密信息载体。这里将这 2 种方法得到的结果作对比。分离出的信号，如果是正值就代表比特 1，负值代表比特 0。这里的阈值 0.15，是经过多次实验确定的。定义提取成功率=提取正确的比特位数/秘密信息的总比特位数，实验结果见表 1~表 3。

表 1 空域隐写的提取成功率

秘密信息长度	直接相减	FASTICA
100	0.740	0.840
2 000	0.763	0.872
5 000	0.775	0.913

表 2 DCT 域隐写的提取成功率

秘密信息长度	直接相减	FASTICA
100	0.670	0.760
2 000	0.656	0.819
5 000	0.675	0.844

表 3 DWT 域隐写的提取成功率

秘密信息长度	直接相减	FASTICA
100	0.520	0.680
2 000	0.564	0.716
5 000	0.591	0.767

为了验证本文方法的有效性，使用其他 49 幅同样规格的图像进行上述实验，结果发现，提取成功率都比较接近。定义平均提取成功率=所有 50 次实验 3 个域提取正确的比特位数之和/(秘密信息的总比特位数×3×50)，结果见表 4。

(下转第 160 页)