# 针对序贯 JSteg的提取攻击

周 涵, 刘九芬, 张卫明

(信息工程大学 信息工程学院, 河南 郑州 450002)

摘要: JSteg是一种基于 JPEG图像的常见隐写算法。对序贯 JSteg算法的提取攻击本质上是对连续嵌入的隐蔽信息起止点的估计问题。文章提出一种针对序贯 JSteg的提取攻击方法, 该方法首先通过多次实验选择合适的窗口, 在窗口内执行 $\chi^2$检验; 然后通过滑动窗口, 估计出起止点可能的存在区间; 最后使用 CPA(Change-Point Analysis)法, 进一步精确估计隐蔽信息的起止点。实验结果表明, 误差范围能达到真实起止点的 -190bits~190bits 对序贯 JSteg算法来说, 该方法还可以作为一种隐蔽信息存在性的检验手段。

关键词: $\chi^2$检验; CPA; JSteg; 提取攻击

中图分类号: TN918.1; O157.4 文献标识码: A 文章编号: 1671-0673(2007)03-0265-04

## New Extracting Method for Sequential JSteg

ZHOU Han, LIU Jiu-fen, ZHANG Wei-ming

(Institute of Information Engineering, Information Engineering University, Zhengzhou 450002, China)

**Abstract**: JSteg is a popular steganographic algorithm based on JPEG. The aim of the extraction to sequential JSteg is to get the beginning and end of the secret message. This paper presents a new extracting method for sequential JSteg. First, we do many experiments to get proper number of the samples for chi-square test, and call this number for the slipping window. Second, we slip the window throughout the whole image. In every window, chi-square test must be carried out one time, and then the area which may conclude the beginning and end could be estimated. Third, we make sure the exact beginning and end through CPA (Change-Point Analysis). The results of experiments show that the error range can reduce to -190bits~190bits. For sequential JSteg, this method is also a detecting method of the secret message.

**Key words**: Chi-square test; CPA; JSteg; extracting attack

信息隐藏作为信息安全领域一个新的前沿技术, 已经成为信息处理领域的一个研究热点。隐写分析是信息隐藏技术的重要分支, 主要研究如何检测、提取、还原破坏隐藏的秘密信息。目前隐写分析的研究主要集中在隐藏信息的存在性检测和隐蔽数据量的估计, 对于隐藏信息提取技术的研究还很薄弱。但是, 是否提取出隐写信息对于信息战中能否最终获取情报、电子辩论中能否有效取证打击网络犯罪起着决定性作用。因此对"隐藏信息提取技术"的研究非常必要。

LSB(Least Significant Bits)替换是目前使用最为广泛的隐写算法。JSteg[1]是针对 JPEG图像的 LSB替换隐写算法, 分为序贯 JSteg和随机 JSteg。序贯 JSteg是将隐蔽信息连续替换载体图像的有效 DCT系数(非 0, 1的量化后 DCT系数)的 LSB, 对于它的提取攻击就是寻找隐蔽信息的起止点。目前关于序贯 JSteg提取攻击公开的工作为 Westfeld[2]所提出的 $\chi^2$检验算法和马宁[3]结合 $\chi^2$检验

收稿日期: 2007-03-09; 修回日期: 2007-04-05

基金项目: 国家自然科学基金资助项目(60473022); 河南省自然科学基金资助项目(0511011300)

作者简介: 周 涵(1982-), 女, 辽宁锦州人, 信息工程大学硕士研究生, 主要研究方向为信息隐藏技术。

算法和最小二乘法设计的方法。这两种方法都能估计出隐蔽信息的终止点，但是它们必须在已知起始点的条件下才能够执行并且估计结果误差比较大。本文提出一种新的提取攻击方法。该方法先通过大量的实验选择出合适的窗口，在窗口内执行 $\chi^2$ 检验；然后通过滑动窗口，估计出起止点可能存在的区间，在此过程中可以判断隐蔽信息的存在性；最后利用 CPA 进一步确定起止点的准确位置。实验结果表明，误差范围能减少到 -190bits ~ 190bits

## 1 $\chi^2$ 检验算法

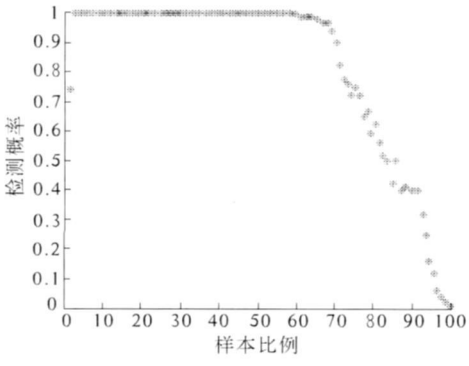
采用 JSteg算法嵌入了秘密消息的 JPEG 图像的 DCT系数的统计特性发生了变化。嵌入过程中相互交换的 DCT系数值对的出现频次开始趋向一致。所谓值对是指在秘密消息嵌入时相互交换的两个值，对于 JSteg算法来说，仅最低位比特不同的两个量化后的有效 DCT系数构成一个值对。例如，2和 3、4和 5、…、-1和 -2、-3和 -4、…，等等都是值对。原始载体图像中，每个值对中两个值的出现频次不一致，但是如果隐蔽信息的比特位 0、1是均匀分布的，在嵌入了秘密数据的载密图像中，每个值对中两个值的出现频次逐渐趋向一致。根据载密图像区别于载体图像的这一统计特征，Westfeld采用 $\chi^2$ 检验统计量来判断被测试图像是否具有载密图像的特征，从而可以判断图像中是否嵌入了秘密信息。

通过不断累加样本连续进行多次 $\chi^2$ 检验，此方法还可以大致估计出秘密消息的大小。

但 $\chi^2$ 检验算法本身是需要已知隐蔽信息的起始点，否则不知从何处开始采样。一个极端的例子，信息隐藏在图像的末尾，采样从图像头开始，样本中有太多未改变的有效 DCT系数致使 $p$ 值几乎降至 0，导致检测失败。那么直接利用 $\chi^2$ 检验算法对序贯 JSteg嵌入的载密图像进行提取攻击，仅能估计隐蔽信息的终止点。

另一个问题，以图 1为例，检测样本数量超出隐蔽信息长度后，随着参杂的载体数据数量越来越多，检测概率也越来越趋近于 0。但是当检测样本数量未超出隐蔽信息长度的一定比例时，检测概率与之前检测样本数量未超出隐蔽信息时的检测概率仍然十分接近，几乎难以区分。那么根据给定的阈值判决终止点位置，估计的终止点必将滞后于真实终止点许多。

(a)嵌入率为 50% 的 Lena 图像




(b)用 LSB 替换嵌入的概率

图 1

由此可见，用 $\chi^2$ 检验算法直接进行提取攻击有两点不足，一是需要知道隐蔽信息的起始点，二是估计的终止点滞后于真实终止点。第 1个不足是 $\chi^2$ 检验算法自身固有的缺陷，而造成第 2个不足的根本原因是用于 $\chi^2$ 检验算法的样本需不断累加，其数量超出隐蔽信息长度后会参杂进载体数据，当载体数据在样本中所占的比例较小时，检测概率很难反应出样本已参杂进载体数据的这种变化。马宁[2]的提取攻击方法第 1步也采用 $\chi^2$ 检验算法，所以也有 $\chi^2$ 检验算法的第 1个不足。她采用最小二乘法和线性回归模型在一定程度克服了 $\chi^2$ 检验算法的第 2个不足。

## 2 提取攻击方法

$\chi^2$ 检验算法的第 2个不足在样本数量减小到一定数量时就能够得到改善。我们将使 $\chi^2$ 检验成功执行的最小样本量称为窗口。利用窗口不仅可以改善 $\chi^2$ 检验算法的第 2个不足，同时也能通过滑动窗口克服 $\chi^2$ 检验算法的第 1个不足。此外，利用该方法还可以检测较少隐蔽信息的存在性。

### 2.1 窗口的选择

我们将窗口记为 $M$。由于不同图像窗口也可能不同，我们利用实验的方法得到合适的窗口。给

定显著性水平 $\alpha=0.05$, 对南加利福尼亚州大学标准图像库[5]中的 352幅图像做 $\chi^2$ 检验。具体实验步骤如下:

① 设定 $M=50$。因为统计学理论上要求 $\chi^2$ 检验的样本量不小于 50, 所以在实验中将 $M$ 的初始值设为 50。

② 在每一幅图像未嵌隐蔽信息的情况下执行 $\chi^2$ 检验。一幅图像需要执行多次 $\chi^2$ 检验, 每次的检验样本取 $M$ 个连续有效 DCT系数, 一个有效 DCT系数只能参与一次 $\chi^2$ 检验。每次 $\chi^2$ 检验得到检测概率 $p$, 记录 $p\geq\alpha$ 的次数, 计算出取伪概率。

③ 在每一幅图像满嵌隐蔽信息的情况下执行 $\chi^2$ 检验, 记录 $p<\alpha$的次数, 计算弃真概率。

按照保证低取伪概率的前提下尽可能降低弃真概率的原则来选择 $M$。从表 1的实验结果可以看出 $M=200$较好。

表 1　图像的滑动窗口选择

| $M$ | 弃真概率 | 取伪概率 |
|---|---|---|
| 50 | 0.9% | 23.9% |
| 100 | 1.4% | 5.7% |
| 150 | 1.1% | 2.1% |
| 200 | 3.3% | 0.8% |

### 2.2　隐蔽信息起止点的可能存在区间估计

确定窗口 $M$ 之后, 我们利用 $\chi^2$ 检验做初步的提取攻击。对一幅待检测图像通过滑动窗口 $M$ 执行多次 $\chi^2$ 检验。设 $i$为检验样本的序号, $p_i$ 为相应的检测概率, 集合 $B$ 为含可疑起始点的样本序号集合, 集合 $E$ 为含可疑终止点的样本序号集合。具体步骤如下:

① 取 $M$ 个连续的有效 DCT系数作为检验样本。选取的原则是同一个有效 DCT系数只参与一次 $\chi^2$ 检验, 该次样本与前次样本相邻且不相交。

② 执行 $\chi^2$ 检验。给定阈值 $\alpha$。当 $p_i\geq\alpha$时, 我们认为样本含有载密数据, 当 $p_i<\alpha$时, 则认为样本未含有载密数据。

③ 当 $p_{i-1}<\alpha$且 $p_i\geq\alpha$时, 则 $i\in B$; 当 $p_i\geq\alpha$且 $p_{i+1}<\alpha$时, 则 $i\in E$。

### 2.3　隐蔽信息起止点估计

2.1和 2.2的结果是嵌入信息的起始点或终止点可能存在于某一段有效 DCT系数中, 但不能确定具体是哪个有效 DCT系数。我们利用 CPA理论中的 CUSUM-Bootstrap[4]法, 将估计起始点或终止点精度提高到具体 DCT系数的水平上。

按照上节的结果, 起始点可能存在于 $B$ 所含的样本中, 终止点则存在于 $E$ 所含的样本中。按照下面的操作步骤估计出起止点。

① 对于 $B$ 中的任意一个元素 $i$, 以 $i-1$, $i$, $i+1$的范围内的每一个有效 DCT系数作为窗口的起始点, 向后取连续 $M$ 个有效 DCT系数执行 $\chi^2$ 检验, 记录下每次检验概率值。

② 将①中得到的检测概率看作是关于起始点的污染数据, 计算 CUSUM 得到估计起始点, 记为 $b_i$, 再用 Bootstrap求得相应的置信度 $con(b_i)$。

③ 对于 $E$中的任意一个元素 $i$, 分别以 $i-1$, $i$, $i+1$的范围内的每一个有效 DCT系数作为滑动窗口的终止点, 向前取连续 $M$个有效 DCT系数执行 $\chi^2$检验, 记录下检验概率值。

④ 利用③中得到的检测概率做 CPA。记估计终止点为 $e_i$, 相应的置信度为 $con(e_i)$。

⑤ 对于 $i\in B\cap E$, 因为做 $\chi^2$检验的样本是相同的, 所以必有 $b_i=e_i$。那么最终估计起始点为 $b_i$, 估计终止点为 $e_i$, 其中

$$con(b_i)=\max\{con(b_i)\mid i\in B-B\cap E\},$$
$$con(e_i)=\max\{con(e_i)\mid i\in B-B\cap E\}。$$

## 3　实验

以 pepper.JPEG为例简述本文提取攻击方法的具体实验过程。pepper.JPEG的有效 DCT系数个数为 34297。将有效 DCT系数读取成一维数组, 隐蔽信息的嵌入起始点为 2000, 结束点是 21999。首先通过实验选择滑动窗口 $M=200$。然后对全部有效 DCT系数进行多次采样, 每次读取连续 200个有效 DCT系数做为检验样本, 对其执行 $\chi^2$ 检验, 记录样本序号和检验概率 $p_i$。设阈值 $\alpha=0.05$, 判断 $p_i$与 $\alpha$的大小关系得 $B=\{0, 3, 10\}$, $E=\{0, 3, 109\}$。

最后我们来精确估计隐蔽信息的起止点。因 $B\cap E=\{0, 3\}$, 所以仅对 $i=10$做起始点估计, $i=109$做终止点估计。$i=10$所代表的样本是第 2000个至第 2199个有效 DCT系数, 估计起始点时以第 1800个至第 2399个有效 DCT系数中的每一个为起始点, 向后读取连续 200个有效 DCT系数做为检验样本, 得到检验概率, 用这些检验概率计算 CUSUM得到估计起始点 $b=1970$, 求出相应的置信度 $con(b)=100\%$。因为 $B-B\cap E$中仅包含 $i=10$一个可疑起始点存在区间, 所以 $b=1970$就是

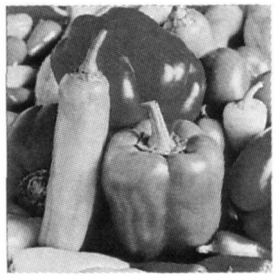


图 2 pepper. JPEG

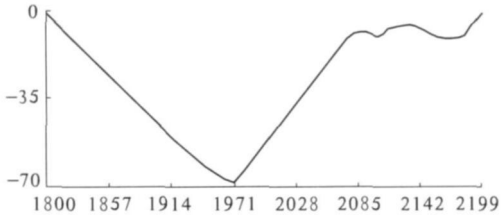



图 3 $i=10$的 CUSUM

最终的估计起始点。对 $i=109$我们重复与 $i=10$同样的过程,得到 $e=22051$。

表 2 检测起始点

| 起始点 | 置信区间 | 置信度 |
|---|---|---|
| 1970 | (1970, 1970) | 100% |

对南加利福尼亚州大学标准图像库[5]中的512×512图像库的 150幅图像,嵌入相同长度隐蔽信息,嵌入起始点为第 2000个有效 DCT系数,终止点为第 21999个有效 DCT系数,长度为20000bits。估计起始点的平均值为 2023.06,标准差值为 63.25,估计终止点的平均值为 22025.4,标准差值是 65.06。而用马宁的 $\chi^2$-最小二乘法不能估计起始点,估计终止点的平均值为 22172.4,标准差是 113.4。由此可以看出,估计结果精度提高很多。

表 3 起始点和终止点的估计

| 图像名称 | 马宁方法起始点 | 马宁方法终止点 | 本文方法起始点 | 本文方法终止点 |
|---|---|---|---|---|
| Lena JPEG | — | 21463 | 2139 | 22028 |
| Pepper JPEG | — | 21373 | 1970 | 22048 |
| Feifei JPEG | — | 22341 | 1971 | 21884 |

## 4 结论

本文研究序贯 JSteg算法的提取攻击问题,首先指出用 $\chi^2$ 检验算法直接进行提取攻击有 2点不足:一是需要知道隐蔽信息的起始点,二是估计终止点滞后于真实终止点。然后提出一种提取攻击方法。该方法首先通过多次实验选择合适的窗口,在窗口内做 $\chi^2$ 检验,不仅可以改善 $\chi^2$ 检验算法的第 2个不足,还可以检测较少嵌入信息的存在性;其次通过滑动窗口估计出起止点可能存在的区间,还可以改善 $\chi^2$ 检验算法的第 1个不足;最后使用 CPA (Change-Point Analysis)法,将估计起止点精确到单个有效 DCT系数水平。与 $\chi^2$ 检验算法[1]和马宁[2]的 $\chi^2$-最小二乘法估计方法相比,不仅可以确定嵌入信息的起始点,而且嵌入信息的终止点位置精度又很大的提高,可以达到真实起止点的 -190bits~190bits。由于该方法的第 1步执行时可以判断出隐蔽信息是否存在,所以该方法也可以作为一种检测方法。

参考文献:

[1] JPEG-JSteg V4[EB/OL]. [2006-09-10]. http://www.funet.fi/pub/crypt/steganography/jpeg-jsteg-v4.iff.gz.

[2] Westfeld A, Pfitzmann A. Attacks on Steganographic System[C]//Proc. of the 3rd Information Hiding Workshop. Dresden Germany, 1999. LNSC. 1768: 61-75.

[3] Ning Ma, Weiming Zhang, Wei Guan and Wenfen Liu. Extracting Attack to Sequential Jsteg Stegostems[C]// The 10th Joint International Computer Conference. Kunming, 2004: 277-281.

[4] Taylor Wayne (2000a). Change-Point Analyzer 2.0 shareware program. Taylor Enterprises, Libertyville, Illinois[EB/OL]. [2006-05-17]. http://www.variation.com/cpa.

[5] CBIR Image Database[EB/OL]. [2006-01-07]. http://www.cs.washington.edu/research/imagedatabase/groundtruth/, 1997.