文章编号: 1671-8860(2006)04-0371-04 文献标志码: A

chi-square检测算法的特性分析研究

周继军[1] 陈 钟[1]

(1 北京大学计算机系信息安全实验室, 北京市海淀路75号, 100871)

摘 要: 对图像信息隐藏chi-square检测算法的特性进行了剖析, 介绍了改变隐藏载体的分布特性、非LSB(least significant bits)嵌入和随机嵌入三种抗检测的算法思想, 分析了EzStego v2.0b3和Jsteg v4.0隐写软件隐藏算法存在的弱点, 并进行了改进, 达到了抵御chi-square检测法的目的。

关键词: 信息隐藏; 检测; LSB; chi-square

中图法分类号: TP393

chi-square[1]检测算法是Andreas等人在第三届国际信息隐藏学术研讨会上提出的, 现在已成为国际上一种较为流行和有效的图像隐藏检测算法, 很多隐写软件都能被该算法检测出来, 因此, 如何改进隐写术的算法, 使之能抵抗这种检测算法成为了一个研究热点。

1 抵御chi-square检测算法

chi-square检测法[2]又称 x^2 检测法或POVs(pairs of values)检测法。有多种方法可以使chi-square检测算法失效。一种是改变被检测数据的分布特性。由于chi-square检测的理论基础之一是检测数据必须符合广义拉普拉斯分布, 因此, 改变隐藏信息后的索引值或DCT系数为其他分布, 将有效抵御这种检测算法。另一种是采用非LSB嵌入方式, 并且修正嵌入信息导致的嵌入密度等图像重要特性参数的规律性改变。如F5[3]算法, 该算法的目的是设计一种大容量、高安全性的隐写工具, 其算法的实现过程首先是得到量化后的DCT系数, 然后根据用户输入的口令产生随机数, 并对DCT系数进行置换, 最后估计非矩阵编码[4]的嵌入容量, 并根据这个值和嵌入信息的长度设计矩阵编码。矩阵编码具体的参数为(C, N, K), 其中, C 代表每 N 个DCT系数中需要改变的DCT系数的个数; K 为要嵌入信息的比特长度。矩阵编码将使得对嵌入载体的改变达到最小, 再对修改的DCT系数进行Huffman编码, 生成JPEG图像。还有一种是Provos提出的利用随机比特嵌入和纠错编码来达到最小程度地改变嵌入载体[5]。

尽管Provos采用了随机序列的嵌入方式, 但是嵌入的信息在某些情况下仍然有可能暴露隐藏信息的存在, 如一个字节范围内的连续比特1或0。为了更高的安全性, 该算法在嵌入隐藏信息时, 故意引入适当的错误, 在提取隐藏信息时, 使用纠错编码技术纠正错误, 从而使嵌入信息具有更强的随机性和扰乱性, 付出的代价只是增加了嵌入信息的长度。

下面给出针对GIF图像调色板隐藏信息的隐写工具EzStego v2.0b3[6]的算法改进和针对JPEG图像DCT域隐藏信息的隐写工具Jsteg v4.0[7]的算法改进, 使之能够抵御chi-square检测。

2 EzStego算法改进

EzStego是Machado制作的针对GIF调色板的隐写工具, 其算法可描述为: ① 拷贝并排序隐藏载体的调色板, 使相邻色彩更加接近; ② 找到排序调色板色彩的索引值, 并从隐藏信息中取出1比特修改该索引值的LSB; ③ 根据修改的索引值在排序的调色板中找到对应的色彩值, 并根据该色彩值找到隐藏载体的调色板的索引值, 用

收稿日期: 2006-01-12。

项目来源: 中国博士后科学基金资助项目(2005038305)。

该值替换第②步中含有 1 比特隐藏信息的索引值,依次类推,将信息隐藏进去;④ 信息提出时,只需根据隐写载体的索引值找到与隐藏载体的调色板相同的排序调色板的色彩值的索引值,取出该值的 LSB 作为隐藏信息的 1 个比特。

由于 EzStego 使用的是 LSB 的顺序嵌入,且没有改变调色板的值,所以其相邻颜色对值会呈现近似的等值性。图 1 为对 3 色深 GIF 图像用 EzStego v2.0b3 隐藏 50 字节信息前后色彩直方图的对比。

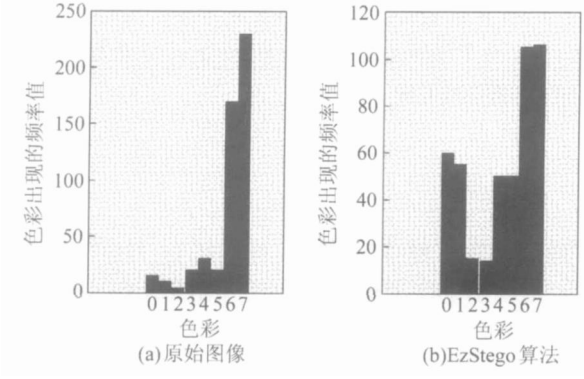

图 1 GIF 图像隐藏信息前后色彩直方图的对比

Fig. 1 Comparison of Color Histogram Pre- and Post-processing of Information Hiding

从图 1 中可以看到,相邻色彩频率近似等值出现,特别是在未排序的调色板色彩对 110 和 111,在隐藏前出现的次数相差大约 70 次,而在隐藏后相差却不到 5 次,所以当使用 chi-square 检测时,检测成功率较高。

在 EzStego 算法中,相邻的 RGB 色彩值可能具有较大范围的亮度差,即同一亮度值可能对应完全不同的色彩值。根据这个特性,在色彩值进行 LSB 修改后,有条件地选择相邻亮度值作为置换对象,其结果扰乱了色彩对值相似频率出现的概率,从而打破了 chi-square 检测的攻击。改进的算法描述如下。

1) 定义相邻色彩之间的 RGB 距离为:

$$D = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}$$

并定义色彩所对应的亮度值为:

$$Y = 0.229R + 0.587G + 0.114B$$

2) 拷贝并排序隐藏载体的调色板,使相邻色彩更加接近。

3) 找到排序调色板色彩的索引值,并从隐藏信息中取出 1 比特修改该索引值的 LSB。

4) 根据修改的索引值,在排序的调色板中找到对应的色彩值,并度量该色彩值与索引值修改前的色彩值的距离 D。如果 D 小于规定的阈值,则按照与 EzStego 算法一样的规则处理;如果 D 大于规定的阈值,则找到隐藏载体中与其亮度近似的色彩值,并根据该色彩值找到隐藏载体的调色板的索引值,用该值替换第 3) 步中含有 1 比特隐藏信息的索引值,依次类推,将信息隐藏进去。

5) 信息提出时,只需根据隐写载体的索引值找到与隐藏载体的调色板相同的排序调色板色彩值的索引值,取出该值的 LSB 作为隐藏信息的 1 个比特。

图 2 为 EzStego 算法和改进后在 D 和亮度的差值门限均为 50 的情况下,针对同一个 3 色深 GIF 图像隐藏信息后色彩直方图的对比。从图中可以看出,改进的算法并没有使用常规的 PNRG,而是利用了一个色彩与亮度比较置换的小技巧。尽管该算法在 EzStego 的抗视觉攻击上并没有提高多少,但就抗 chi-square 检测是有效的,且工程实现简单。

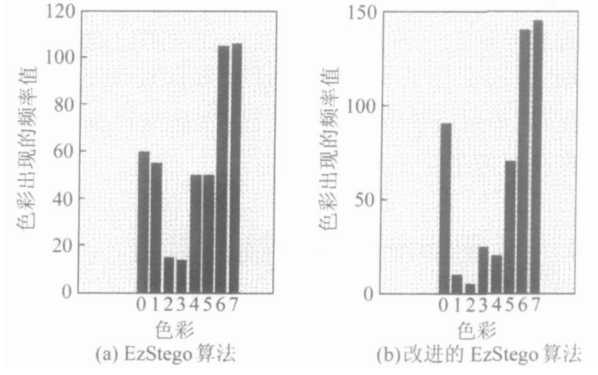

图 2 EzStego 算法改进前后色彩直方图的对比

Fig. 2 Comparison of Color Histogram Pre- and Post-processing of EzStego Algorithm Improvement

3 Jsteg 算法改进

Jsteg v4.0 是 Derek 制作的 JPEG 图像信息隐藏软件,采用 LSB 方法嵌入 DCT 系数,其嵌入过程介于 JPEG 量化和编码之间。具体算法可描述为:先对隐藏信息进行压缩和加密预处理,反解 Huffman 编码生成 DCT 系数,然后将信息按顺序嵌入在交流 DCT 系数的 LSB 上,且保持 0、1 系数值不变,再 Huffman 编码生成隐写载体。隐写载体表面上很难看出隐藏数据的痕迹,但对比添加信息前后的 DCT 谱,却呈现明显的阶梯效应,即相邻 DCT 系数出现的频率非常接近,使用 chi-square 检测很容易检测出来。图 3 为使用 Jsteg 隐藏前后的 DCT 系数直方图比较。

从图 3 可以很明显地看出,DCT 值对系数

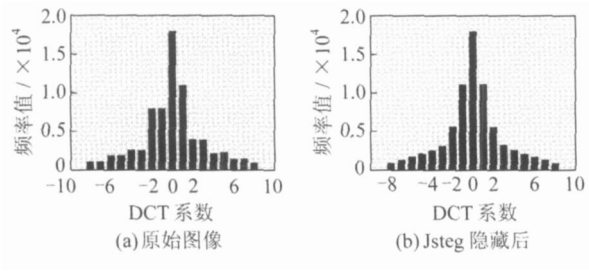

图 3　Jsteg 隐藏前后的 DCT 系数直方图

Fig. 3　Comparison of DCT Coefficient Histogram Preand Post-processing of Jsteg hidding

(-1, -2) 和(2, 3) 在隐藏信息的 JPEG 图像中出现的频率几乎一样, 因此, 使用 chi-square 检测很有效。为了抵抗这种检测方法, Provos 在 Jsteg 的基础上改进了其算法, 并制作了 outguess 0.2[8] 来抵抗 chi-square 检测。该算法的思想是: chi-square 及其扩展检测法只能对那些仅仅修改了一阶图像的统计量有效, 因此, 如果对随机嵌入的信息引起的图像特征进行二阶图像统计处理, 即使用纠错编码技术使得隐写载体的分布尽量与原图一样, 便可以彻底抵抗 chi-square 检测法了。

笔者采用的算法是 LSB 嵌入后, 根据每一块 DCT 相应的系数值的统计结果按规则作修改值对, 以使 DCT 系数出现的频率接近未嵌入信息前的广义拉普拉斯分布。该分布的概率密度函数为:

$$f(x) = \frac{1}{z} \cdot e^{-|x/s|^{\alpha}} \qquad (1)$$

式中, 方差 $\sigma^2 = s^2\Gamma\left(\frac{3}{\alpha}\right)/\Gamma\left(\frac{1}{2}\right)$; 峰值 $K = \Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{5}{2}\right)/\Gamma^2\left(\frac{3}{2}\right)$; z 为归一化常数。根据式(1), 在 8×8 DCT 系数块中, 0 的比例占绝大多数, 而且 Jsteg 算法中没有对 0、1 进行任何修改, 因此, 0、1 出现的频率值在隐藏前后将保持不变。对于其他的 DCT 系数, 正数 2 ↔3、4 ↔5、6 ↔7 等构成值对, 负数 -1 ↔ -2、-3 ↔ -4、-5 ↔ -6 等构成值对, 其原因是负数在计算机中用补码表示, 因此, -1 表示为 11111111, -2 表示为 11111110, 两者之间的差异可由 LSB 改变实现。Jsteg 改进的算法可描述如下。

1) 在 Jsteg 嵌入信息后, 统计每一块 DCT 系数对值, 当 $N = 1, 2, \cdots, 127$ 时, 值对为$(2N, 2N+1)$, 出现的频率为 v_{2N}、v_{2N+1}; 当 $N = -1, -2, \cdots, -253$ 时, 值对为$(N, N-1)$, 出现的频率为 v_N、v_{N-1}。

2) 定义正的 DCT 系数对值频率差求解式(2), 负的 DCT 系数对值频率差求解式(3):

$$D_+ = |v_{2N} - v_{2N+1}| \qquad (2)$$

$$D_- = |v_N - v_{N-1}| \qquad (3)$$

其频率差值门限为 3。该门限是根据对 USC-SIPI[9] 图像数据库 50 张标准图像进行 Jsteg 隐藏前后对比的测试值, 平均在每一块 DCT 中, 值对频率差值门限大于 3 时, 这种差异在多个 DCT 系数块的累积作用下将使得相邻值对频率的差异明显, 从而使 chi-square 检测无效。但当频率差值门限在 3 以下时, 这种差异将减弱, 从而容易受到 chi-square 检测的攻击。

3) 由于 Jsteg 没有对 DCT 直流系数进行修改, 因此, 将利用直流系数的后 3 位作为修改值对中绝对值较小的个数计数, 修改时, 尽可能增加绝对值大的个数而减少绝对值小的个数, 其目的是使得 DCT 系数直方图能尽可能符合未隐藏信息前的广义拉普拉斯分布。

4) 统计 DCT 系数块中值对频率差小于 3 的最多值对数进行修改(如果有两组值对数相等, 将选取靠近 0 的值对), 由第 3) 步, 正 DCT 系数 $2N+1$ 将按顺序修改为 $2N$。当 $0 < v_{2N} - v_{2N+1} < 3$ 时, 修改个数为 $3 - D_+$; 当 $0 < v_{2N+1} - v_{2N} \leqslant 3$ 时, 修改个数为 $1 + D_+$; 当 $v_{2N+1} = v_{2N}$ 时, 修改个数为 $2 + D_+$。同理, 负 DCT 系数 $N-1$ 将按顺序修改为 N, 当 $0 < v_N - v_{N-1} < 3$ 时, 修改个数为 $3 - D_-$; 当 $0 < v_{N-1} - v_N \leqslant 3$ 时, 修改个数为 $1 + D_-$; 当 $v_N = v_{N-1}$ 时, 修改个数为 $2 + D_-$。

5) 信息提出时, 只需根据每块 DCT 系数的直流值的后 3 个 bit, 得到修改的 DCT 系数的个数, 然后按顺序恢复原有的 DCT 系数值, 再根据 Jsteg 的提取算法将隐藏信息提出。图 4 为 Jsteg 算法改进前后 DCT 系数直方图的对比。由图 4 可以看出, 改进的 Jsteg 算法在 DCT 系数直方图上基本与原始图像一致, 从而具有了抵御 chi-square 检测的能力。

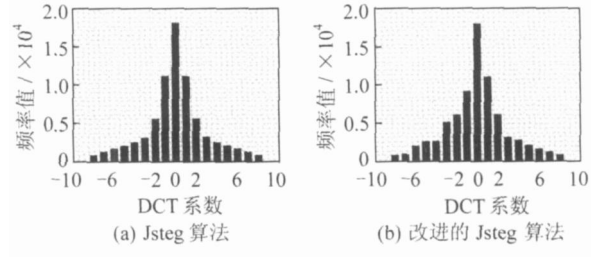

图 4　Jsteg 算法改进前后 DCT 系数直方图的对比

Fig. 4　Comparison of DCT Coefficient Histogram pre- and Post-processing of Jsteg Algorithm Improvement

参 考 文 献

[1] 盛骤, 谢式千, 潘承毅. 概率论与数理统计(2 版)[M]. 北京: 高等教育出版社, 1989

[2] Andreas W, Andreas P. Attacks on Steganographic Systems[M]. Heidelberg, Berlin: Springer-Verlag, 2000: 61-76

[3] Westfeld A. High Capacity Despite Better Steganalysis(F5-Astegano Graphic Algorithm)[M]. Berlin, NewYork, Heidelberg: Spring-Verlag, 2001: 289-302

[4] Ron C. Some Notes on Steganography[OL]. http://os.inf.tu-dresden.de/~westfeld/crandall.pdf, 1998

[5] Provos N. Defending Against Statistical Steganalysis[C]. The 10th USENIX Security Symposium, Washington D C, 2001

[6] Machado R. EzStego[OL]. http://www.stego.com, 1997

[7] Derek U. Jsteg[OL]. http://ftp.ntua.gr/pub/crypt/mirrors/idea.sec.dsi.unimi.it/code/jpeg-v4.tar.gz, 1997

[8] Provos N. Outguess[OL]. http://www.outguess.org/outguess-0.2.tar.gz, 2001

[9] University of Southern California Signal & Image Processing Institute. The USG-SIPI Image Database [OL]. http://sipi.usc.edu/services/database/database.html, 1977

第一作者简介: 周继军, 博士后, 高级工程师。现主要从事网络攻防、信息隐藏等网络与信息安全新技术研究。

E-mail: zjjinby@163.com

Characteristics Analysis on the chi-square Detection Algorithm

ZHOU Jijun[1] *CHEN Zhong*[1]

(1 Information Security Lab. of Department of Computer Science, Peking University, 75 Haidian Road, Beijing 100871, China)

Abstract: This paper analyzes the characteristics of chi-square detection algorithm for image information hiding and briefly introduces three defending detection algorithm ideas. They are transforming steganographic carry properties, inserting instead of LSB and random inserting algorithm. And then we analyze algorithmic weak points of EzStego v2.0b3 and Jsteg v4.0, improve them for defending chi-square detection algorithm successfully.

Key words: information hiding; detection; LSB; chi-square

About the first author: ZHOU Jijun, post-doctor, senior engineer. He is engaged in such new technologies on network and information security reseach work as information hiding, network attack and defence.

E-mail: zjjinby@163.com

(上接第 370 页)

coding and bite rate changing for different network bandwidths. With the key regional coding method for the key image optimization coding, and the variable transmission strategy by changing transmission bite rate according to bandwidth, we solve the distance teaching problem of real time in Internet network environmental.

Key words: distance teaching; video transmission; statified coding; bite rate control

About the author: GONG Zikang, associate researcher, majors in modern educational technology and modern distance educational technical.

E-mail: zkgong@mail.whut.edu.cn