一种安全的信息隐藏范式及其在二值图像上的实现

林代茂[1], 郭云彪[1], 胡 岚[1], 周琳娜[1], 牛夏牧[2]
(1. 北京电子技术应用研究所, 北京 100091; 2. 哈尔滨工业大学, 黑龙江哈尔滨 150001)

摘 要: 本文提出了一种安全的信息隐藏范式, 这种范式是让信息隐藏过程模拟一种其他的正常操作来实现的. 针对文字扫描操作, 给出了这种安全范式在二值图像上的具体实现方法: 利用扫描过程中测量噪声的不确定性来携带信息. 并根据形态学滤波理论提出了该方法的快速实现, 同时采用矩阵编码来提高隐藏信息的容量.

关键词: 信息隐藏; 安全范式; 扫描; 二值图像
中图分类号: TN918.2 文献标识码: A 文章编号: 0372-2112 (2005) 09-1537-04

A Secure Paradigm of Information Hiding and the Realization in Binary Images

LIN Dai-mao[1], GUO Yun-biao[1], HU Lan[1], ZHOU Lin-na[1], NIU Xia-mu[2]
(1. *Beijing Application Institute of Electronic Technology, Beijing* 100091, *China*;
2. *Information Countermeasure Technique Research Institute, Harbin, Helongjiang* 150001, *China*)

Abstract: The security of information hiding is of great importance. A secure paradigm of information hiding which is fulfilled by simulating natural operation is suggested. A scheme to realize this secure paradigm in binary images using scanning operation is proposed and a more efficient practical method is also presented based on morphological theory. The matrix encoding is applied to increase the embedding capacity.

Key words: information hiding; secure paradigm; scanning; binary image

1 引言

对于隐蔽通信而言, 信息隐藏技术的安全性与隐藏量的关系是最为突出的矛盾. 为了隐藏更多的秘密信息, 人们希望掩盖载体具有更大的冗余, 以便隐藏一定量的秘密信息以后, 载体本身的特性不会发生较大的改变. 然而这样做未必是安全的, 因为信息隐藏技术的现代攻击方法不仅局限于分析载体的统计特性, 还会从各种不同的角度进行分析判断, 以发现隐藏在载体信息中的秘密. 例如, 虽然 BMP 格式真彩色图像存在较大的冗余空间, 但是在带宽不十分宽裕的信道上传输带有大量冗余信息的 BMP 格式图像就不是安全的好方法, 因为攻击者不用做更深层次的攻击就可能怀疑有秘密信息隐藏其中, 所以这种方法与其说是信息隐藏, 倒不如说是一种信息加密. 因此, 在考虑信息隐藏的安全性时, 必须全面地考虑可能的攻击方法.

在文献[1]中对信息隐藏的攻击方法进行了较为详尽的综述(参见图 1).

在图 1 的诸多攻击方法中, 唯携密载体的被动攻击是隐蔽通信面临的最重要的攻击方式. 然而, 这种攻击方式也可能对信息隐藏采用的原始载体作必要的推测, 所以, 在建立安全的信息隐藏系统时, 主要考虑针对载体的攻击.

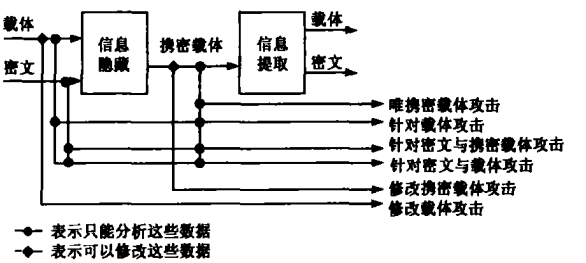

图 1 信息隐藏的攻击模型

2 一种安全的信息隐藏范式

目前多数的实现方法都使用某种信息作为信息隐藏的载体, 对这种情况来说, 载体不仅是传输秘密信息的依托, 也是攻击者极力获取的数据. 攻击者希望通过对比原始载体与携密载体来发现二者的差异. 然而, 对同一信息所做的不同原始记录也未必相同, 这样, 攻击者就必须分析原始载体与携密载体间差异的性质, 判断这种差异是隐藏信息的处理给载体留下的特征痕迹, 还是对载体的某种其他正常处理(例如扫描)所造成的差别. 假如隐藏信息的处理没有给载体留下痕迹, 或者留下的痕迹没有明显的特征, 那么从针对载体攻击的角度而言, 这个信息隐藏系统就是安全的了. 在载体中嵌入信息势

收稿日期: 2004-10-12; 修回日期: 2005-03-15
基金项目: 国家 863(No. 2002AA145010); 国家 973(No. TG1999035804)

必改变载体, 信息隐藏不给载体留下痕迹是不可能的, 因此, 我们只能在消除隐藏特征上下工夫.

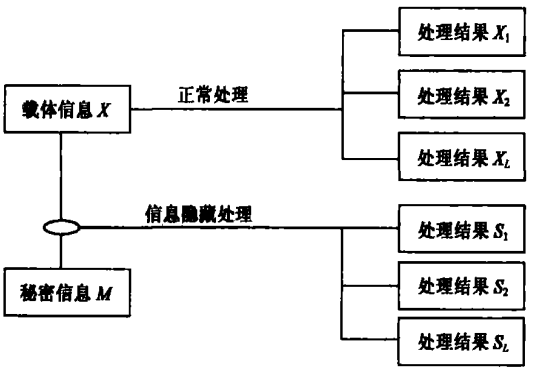

图 2 一种安全信息隐藏范式

下面, 我们用图 2 来说明建议的安全信息隐藏范式. 对原始载体的不同处理可能产生不同的结果如 X_1, X_2, X_l 等, 不同的信息隐藏方法也会有不同的结果如 S_1, S_2, S_l 等, 如果能找到一种信息隐藏方法, 使得 S_i 与某一种正常处理的 X_i 基本一致, 就很难判断其是否隐藏了秘密信息.

从信息论角度来看, 原始载体的信息量是 $I(X)$, 而正常处理结果的信息量是: $I(X_j)=I(X_j;X)+H(X_j)$, 其中互信息项 $I(X_j;X)$ 表示从 X 得到的关于 X_j 的信息, 而第二项是由于处理结果引入的不确定成分, 同样信息隐藏处理也有类似的结果 $I(S_i)=I(S_i;X)+H(S_i)$, 因此, 上述安全信息隐藏范式应以 $H(X_j)$ 为约束.

3 安全信息隐藏范式在二值图像上的实现

上述安全信息隐藏范式的实现, 首先要确定信息隐藏处理要模拟哪一种正常操作. 然后对这种正常操作进行必要的研究, 以便使信息隐藏处理符合正常的规律.

3.1 实现思路

设有二值图像 X(其各像素点为 $X(i,j)=0,1$), 对其进行扫描的结果作为信息隐藏的载体数据. 实践证明, 在同一设备上进行多次扫描得到的输出图像 $X_i(i=1,\cdots,l)$ 并不相同, 这种差异是测量过程引入的噪声, 利用图像扫描差异隐藏秘密信息必须符合测量噪声的规律. 为此, 首先定义两幅图像的距离为这两幅图像对应像素差之和:

$$D=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}|x_1(i,j)-x_2(i,j)| \tag{1}$$

其中 M, N 表示图像的大小; i, j 表示图像像素点的位置. 然后进行如下操作:

(1) 首先用一台扫描仪在同样的参数设置下, 对同一幅图像进行 L 次扫描, 得到 L 组图像数据 X_1, X_2, …, X_L

(2) 对所得到的 L 幅图像在二维空间求平均, 求得均值图像 $\bar{X}$, 其各像素点为:

$$\bar{x}(i,j)=\frac{1}{L}\sum_{l=1}^{L}x_l(i,j) \tag{2}$$

(3) 求每幅扫描图像与均值图像 $\bar{X}$ 之间的距离 D_1, D_2, …, D_L, 并计算平均距离作为扫描图像的标准误差:

$$\bar{D}=\frac{1}{L}\sum_{l=1}^{L}D_l$$

(4) 以第 3 步的标准误差为约束控制隐藏算法, 以保证嵌入秘密信息时引入的噪声与扫描引入的随机噪声处于同等水平.

3.2 误差控制隐藏算法

为实现上节第 4 步的要求, 本文采用分块法嵌入信息, 将选定的扫描图像分成大小为 $K=m\times n$ 的若干个子块(m, n 表示子块的长度和宽度), 在每个子块中至多修改一个像素, 通过调整子块的大小, 保证信息隐藏引入的误差与扫描引入的随机噪声保持在同一水平. 具体作法是:

首先计算标准误差的倒数 $\frac{1}{\bar{D}}$ 和正整数 $K=2^{\lfloor\log_2(\frac{1}{\bar{D}})\rfloor}$, 然后将图像分成大小为 $K=m\times n$ 的子块, 再对每一子块分别采用矩阵嵌入法嵌入秘密信息.

矩阵嵌入法是 Ron Crandall 提出的一种提高信息携带效率的方法[2]. 定义携带效率为携带信息量与修改信息量之比, 则直接嵌入的效率是 2, 即修改 1 比特数据平均携带两比特信息, 而矩阵嵌入法的携带效率可以根据矩阵规模有不同程度的提高. 下面我们用一个简单的例子来说明这个问题.

假如我们希望利用 a_1、a_2、a_3 三个 1 比特数据来携带两比特信息 x_1、x_2, 可以先计算两个参数: $x'_1=a_1\oplus a_3$ 和 $x'_2=a_2\oplus a_3$, 然后将它们与 x_1 和 x_2 进行比较:

当 $x_1=x'_1$ 且 $x_2=x'_2$ 时, 可以不对原始数据做任何修改;

当 $x_1\neq x'_1$ 且 $x_2=x'_2$ 时, 只要修改 a_1;

当 $x_1=x'_1$ 且 $x_2\neq x'_2$ 时, 只要修改 a_2;

当 $x_1\neq x'_1$ 且 $x_2\neq x'_2$ 时, 只要修改 a_3.

这个例子至多修改 1 比特数据就可以携带两比特信息, 明显提高了携带效率. 可以将其推广为一般情况: 在 n 个 1 比特数据 $a_1\cdots a_n$ 中至多修改 m 个数据可以携带 k 比特信息, 则修改后的数据 a'_1, …, a'_n 与修改前的数据之间的汉明距离满足 $d(a,a')\leqslant m$. 根据汉明编码理论, 我们很容易实现在 $n=2^k-1$ 比特数据中至多修改 1 比特来携带 k 比特信息, 记作 $(1,k,2^k-1)$, 携带效率为 $w=\frac{2^k}{2^{k-1}-1}k$.

4 基于形态学滤波的安全隐藏快速实现

在上一节中, 我们叙述了基于扫描随机误差的安全信息隐藏范式在二值图像中的实现方法. 由于需要经过多次扫描、求均值图像、误差统计等步骤, 显得十分繁琐.

为了解决这个问题, 我们引用形态学非线性滤波理论来估计原始图像. 扫描过程引入的随机噪声在视觉上的反映是: 文字的笔画不如原图光滑, 起笔存在凸凹破缺. 根据形态学理论, 对二值图像进行腐蚀和膨胀处理可以滤去这些噪声. 腐蚀和膨胀运算的公式分别为:

$$A\circ B=\{X:B+X\in A\} \tag{3}$$

和

$$A\ominus B=\bigcup\{A+X:X\in B\} \tag{4}$$

式(3)中的 $A\circ B$ 表示用结构元素 B 来腐蚀图像 A, 将结构元

素 B 在图像 A 中平移, 所有可以填入图像 A 内部的结构元素的原点组成的新图像即为腐蚀的结果图像. 式(4)中的 $A \ominus B$ 表示用结构元素 B 来膨胀图像 A, 用结构元素 B 内的所有点作为参考点来平移输入图像 A, 然后计算平移结果的并集. 从上述公式可以看出, 只要选择合适的滤波结构, 就可以去除信息嵌入过程引入的噪声颗粒.

注意到腐蚀会使图像收缩, 膨胀会使图像扩张, 因此不能单独采用腐蚀或膨胀对图像进行滤波, 应该将二者结合起来. 可以对图像先做腐蚀再做膨胀(形态学称之为开运算), 也可以先膨胀再腐蚀(形态学称之为闭运算), 使得膨胀与腐蚀运算改变图像拓扑的缺陷相互抵消.

必须对结构元素(SE)的尺寸进行合理的选择. 大的结构元素可以去掉大的噪声颗粒, 但会影响原始图像的内容, 使图像拓扑产生退化; 考虑到二值图像在扫描过程引入的随机噪声大都是单象素噪声颗粒, 因此结构元素的尺寸选择应尽量小. 考虑到二值图像多用来表示文字, 图像内包含的基本图形单元较小, 采用闭运算方式能更好地保护图像的拓扑. 经多次实验, 选择 2× 2 的矩形结构:

$$\mathrm{SE}=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \tag{5}$$

进行降噪滤波所获得的图像与均值图像最为逼近.

这样便略去了扫描估计这些烦琐环节, 可以对二值图像直接进行信息隐藏, 具体步骤如下:

(1) 首先对二值图像进行闭运算滤波, 来估计均值或原始图像;

(2) 求估计图像与载体图像的图像距离.

(3) 以上一步计算的图像距离为约束确定嵌入深度, 根据嵌入深度进行分块.

(4) 利用矩阵编码进行信息嵌入.

完成信息嵌入后信息的提取就变得比较简单, 信息嵌入端和提取端就信息的嵌入策略早已形成约定, 提取时只需确定嵌入分块大小 K, 再按分块大小 K 进行分块, 然后应用嵌入矩阵的逆运算进行提取即可完成嵌入信息复原.

信息提取确定分块大小的过程与嵌入时相同, 首先对携密的二值图像进行闭运算滤波, 估计出均值图像(注: 此处估计出的均值图像虽然与嵌入端估计出均值或原始图像可能不完全相同, 但应基本相等), 然后求出均值图像与携密二值图像的图像距离 $\overline{D}$, 以图像距离为约束确定嵌入深度, 根据嵌入深度进行分块, 每块大小 $K=\left\lfloor \frac{1}{\overline{D}} \right\rfloor^{\log_2}$, 为了保证携密二值图像和嵌入前的原始图像与均值图像的距离保持在同一水平, 因此要求在进行隐藏时必须所有的图像块均进行携密处理(需要满嵌, 秘密信息不够满嵌时应以随机信息补足). 此外由于分块大小 K 是以 2^n 为模向上取整, 所以即使提取端所计算出的标准误差与嵌入端的稍有差异, 计算出的分块大小 K 也应与嵌入端的一致. 我们用紫光 B6210 扫描进行多次实验证明, 标准误差 $\overline{D}$ 在 0.5‰左右. 在此误差水平之下, 我们对扫描图像进行多次嵌入提取实验, 嵌入的信息均能准确复原.

5　实验结果

根据前面提出的安全隐藏范式, 我们对二值图像进行了模拟嵌入实验. 对一幅英文打印稿在相同设置下进行了多次扫描, 得到典型扫描图像如图 3(a)所示, 然后对多幅扫描结果求平均, 得到均值图像如图 3(b)所示, 求得图像的误差 0.0041, 然后分别以 2 倍、1 倍、1/2 倍图像误差为约束, 进行信息嵌入操作, 得到携密图像图 3(d)、图 3(e)、图 3(f). 利用数学形态学对图 3(a)进行开运算滤波, 得到滤波后图像图 3(c), 求得图像距离为 0.0054, 再分别以 2 倍、1 倍、1/2 倍图像距离为约束, 进行信息嵌入操作, 得到携密图像图 3(g)、图 3(h)、图 3(i). 从实验结果可以看出即使将嵌入约束扩大为 2 倍图像误差, 还是难以在主观视觉上区分原始扫描图像与隐藏携密图像.

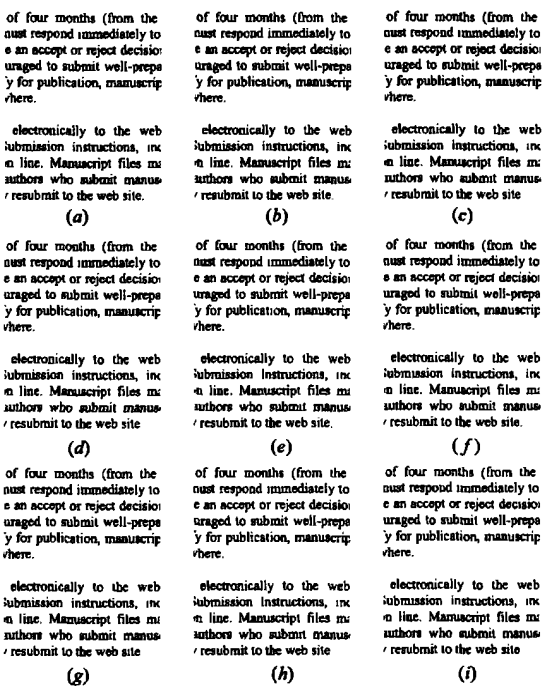

图 3　嵌入效果图

图 3 中(a)为典型扫描图像; (b)为多次扫描均值图像; (c)是图像(a)经过形态学滤波后的图像; (d)、(e)、(f)分别为对图像(b)以 2 倍、1 倍和 1/2 倍图像误差为约束进行嵌入的结果; (g)、(h)、(i)分别为对图像(c)以 2 倍、1 倍和 1/2 倍图像距离为约束进行嵌入得到的结果.

6　结论

本文提出了一种安全的信息隐藏范式, 让信息隐藏过程模拟一种其他的正常操作, 使攻击者难以区分信息隐藏产生的携密载体与正常操作产生的正常载体. 笔者还针对文字扫描操作给出了这种安全范式在二值图像上的具体实现方法, 当然实现方法仅仅保证正常操作与隐藏嵌入在引入误差幅度上在同一水平, 没有更为深入地考虑两个操作引入误差的分布情况, 具体算法还有待完善, 但嵌入操作模拟正常处理的隐藏算法设计思想对提高技术的抗检能力具有深远的意义.

参考文献:

[1] 胡岚, 周琳娜, 郭云彪. 信息隐藏分析与攻击[A]. 信息隐藏第四届全国学术研讨会(CIHW2002)论文集[C]. 北京: 机械工业出版社, 2002. 34- 41.
Hu Lan, Zhou Linna, Guo Yunbiao. Analysis and attack on information hiding[A]. Proceeding of CIHN2002[C]. Beijing: Mechanical Industry Press, 2002. 34- 41. (Chinese Source)

[2] Ron Crandall: Some notes on steganography. Posted on steganography mailing list[DB/OL]. http://os.inf.tu-dresden.de/~westfeld/crandall.pdf, 1998.

[3] 崔屹. 图像处理与分析——数学形态学方法及应用[M]. 北京: 科学出版社, 2000.
Cui Yi. Image Processing and Analysis-Mathematical Morphology Method and Its Applications[M]. Beijing: Scientific Press, 2002. (Chinese Source)

[4] 金淮斌. 基于数学形态学的图像信息隐藏检测研究[A]. 信息隐藏第二届全国学术研讨会(CIHW2000)论文集[C]. 西安: 西安电子科技大学出版社, 2000. 215- 220.
Jin Huaibin. Image steganalysis based on mathematical morphology [A]. Proceeding of CIHW2002[C]. Xi'an: Xi'an Electronic Technology Publishing House, 2000. 215- 220. (Chinese Source)

[5] F Cheng, A N Venetsanopoulos. An adaptive morphological filter for image processing[J]. IEEE Trans Image Processing, 1992, 1(4): 533- 539.

[6] C S Regazzoni, A N Venetsanopoulos, G L Foresti, et al. Statistical Morphological Filters for Binary Image Processing[A]. Proc of Int Conf. Acoustics Speech and Signal Processing ICASSP'94[C]. Adelaide, Australia, ICASSP'94. 1994. 77- 80.

作者简介:

林代茂 男, 辽宁沈阳人, 工学硕士, 1945 年生, 1969 年毕业与清华大学电子工程系, 北京电子技术应用研究所研究员, 兼职教授, 博士生导师, 电子学会高级会员, 多媒体信息安全专家委员会专家委员, 国家科技进步奖评审委员, 主要研究领域为信息安全、信息隐藏、信号处理、信号谱估计. E-mail: laol@sklao.ac.cn.

郭云彪 男, 河北赵县人, 1969 年生, 北京电子技术应用研究所副所长, 副研究员, 中国电子学会高级会员, 计算机学会高级会员, 天津大学在职博士研究生, 主要研究方向为信号处理、信息隐藏、图像处理. E-mail: gybgmx@hotmail.com.