一种基于汉明码和湿纸码的隐写算法

朱雪秀　　刘九芬　　张卫明
(信息工程大学信息工程学院　郑州　450002)

摘　要：该文通过将载体图像分割成矩阵块，重复利用载体矩阵块像素，结合汉明码和湿纸码构造了一种新的双层结构隐写算法。该算法首先利用汉明码在载体矩阵的行向量中嵌入信息，然后在列向量上根据是否影响前面嵌入结果以及是否需要 3 个修改引入“dry”和“wet”的概念，并通过构造伪随机二值矩阵，采用湿纸码在列向量上实现二次嵌入，从而获得较好的隐写性能。实验结果表明：该隐写算法在嵌入率为 0.1 至 0.8 bpp 范围内，性能优于 PPC 和 F5 算法。

关键词：隐写术；矩阵编码；汉明码；湿纸码
中图分类号：TN918　　**文献标识码**：A　　**文章编号**：1009-5896(2010)01-0162-04
DOI: 10.3724/SP.J.1146.2009.00030

A Steganographic Algorithm Based on Hamming Code and Wet Paper Code

Zhu Xue-xiu　　Liu Jiu-fen　　Zhang Wei-ming
(*Institute of Information Engineering*, *Information Engineering University*, *Zhengzhou* 450002, *China*)

Abstract: Dividing the cover image into matrix blocks, using the pixels in the matrix blocks repeatedly, this paper constructs a new double-layered steganographic algorithm based on Hamming code and wet paper code. For each matrix block, the secret message is first embedded by using Hamming code in rows. Then according to whether the previous embedding result is influenced or whether 3 changes are needed when embedding message in columns, the notation of “dry” and “wet” is introduced. Finally wet paper code is used for second embedding. The proposed setganographic algorithm can obtain better performance than PPC and F5 algorithm for 0.1-0.8bpp of embedding rate.

Key words: Steganography; Matrix embedding; Hamming code; Wet paper code

1　引言

作为信息安全的一种新兴技术，信息隐藏已经成为信息技术领域研究的一个热门课题。隐写术是信息隐藏技术的一个主要分支，主要研究如何在可公开的数字媒体中隐藏信息实现隐蔽通信。秘密信息的伪装性和与载体的不可分离性使隐写术除了具有传统加密通信的优点外，还大大降低了受攻击的可能性。

提高隐藏效率，即对载体作尽可能少的修改而嵌入尽可能多的信息是隐写术的一个核心问题。一般而言，在相同嵌入率的情况下，隐写术对载体的修改越少，隐藏信息被检测到的可能性越小，即安全性越高。

矩阵编码最早由 Crandall[1]提出。它是用来提高编码效率的一种隐写方案。这种技术是在信息嵌入的时候以占用较多的载体数据为代价，达到减少对载体数据的修改来提高嵌入效率。矩阵编码最早应用于 F5 算法[2]。该算法通过在 2^k-1 个 DCT 系数上最多修改一个比特就可以嵌入 k 比特信息，大大减少了对载体数据的修改。基于二值图像的 CPT 算法[3]实际上是 GF(2) 上的一个 $(2^r-1,r,2)$ 隐写码，此算法通过将载体二值图像分成 $m\times n$ 的小块，在每个小块中最多修改两个像素值就可以嵌入 $\lfloor\log_2(mn+1)\rfloor$ 比特秘密消息。2005 年，Fridrich 等[4]提出了湿纸编码方案。湿纸编码通过一条只有发送方知道的选择规则决定图像中哪些是可作修改的位置，使得发送方可以在避免对载体的敏感区域做修改的情况下嵌入信息，从而提高了隐写算法的安全性。

PPC(Product Perfect Code)[5]算法将载体数据分割成大小为 $(2^x-1)\times(2^y-1)$ 的矩阵块，在行和列上均使用汉明码，获得比 F5 算法更好的隐写性能。本文在 PPC 算法的基础上提出了一种基于汉明码和湿纸码的隐写算法，该算法将载体数据分成大小为 $N\times(2^x-1)$ 的矩阵块(N 为湿纸码的分组长度)，对矩阵块的行仍使用汉明码，而对列使用湿纸码避开要引入 3 个修改的位置，在相同嵌入率的情况下，相对 PPC 算法而言减少了对图像像素的修改个数，

2009-01-09 收到，2009-06-17 改回
国家自然科学基金(60803155)资助课题
通信作者：朱雪秀　　xiuzi0305@163.com

从而获得比 PPC 算法更好的隐写性能。另外，我们还讨论了本文算法的计算复杂度。

2　F5 算法和湿纸码

本节介绍相关知识。首先给出隐写码和向量支集的定义。以下用黑斜体符号代表向量或矩阵。

定义 1[6]　令 n 和 k 为正整数，$k \le n$，X 是一个有限集合。一个 X 上 $[n,k]$ 隐写码是一组映射：$e: X^k \times X^n \to X^n$ 和 $r: X^n \to X^k$，使得 $r(e(\boldsymbol{s},\boldsymbol{x})) = \boldsymbol{s}$，对所有的 $\boldsymbol{s} \in X^k$ 和 $\boldsymbol{x} \in X^n$，e 和 r 分别为嵌入映射和提取映射。$\rho = \max\{d(\boldsymbol{x}, e(\boldsymbol{s},\boldsymbol{x})) \mid \boldsymbol{s} \in X^k, \boldsymbol{x} \in X^n\}$ 为 $[n,k]$ 隐写码的隐写半径。

定义 2[5]　令 $\boldsymbol{x}=(x_1,x_2,,x_n) \in \mathrm{GF}(q)^n$，则称 $\mathrm{supp}(\boldsymbol{x}) = \{i \mid x_i \ne 0\}$ 为向量 $\boldsymbol{x}$ 的支集。

下面简要介绍 F5 算法和湿纸码。

2.1 F5 算法

F5 编码算法实际上是一个 $[2^k-1,k,1]$ 线性隐写码[7]。设载体数据的 LSB(Least Significant Bit)块为长为 n 的向量 $\boldsymbol{x}$，$\boldsymbol{x} = (x_1,x_2,\cdots,x_n) \in \mathrm{GF}(2^n)$，$n = 2^k-1$，要隐藏的消息分组为 $\boldsymbol{s}$，$\boldsymbol{s} \in \mathrm{GF}(2^k)$，F5 算法的嵌入函数可为

$$e(\boldsymbol{s},\boldsymbol{x}) = \left(\bigoplus_{i=1}^{n} x_i \cdot i\right) \oplus \boldsymbol{s} \tag{1}$$

其中 $\oplus$ 均指模 2 加运算，$x_i \cdot i$ 和 $\boldsymbol{s}$ 均为整数，将其转化为二进制向量然后做 $\oplus$ 运算。给定一个向量 $\boldsymbol{x}'$，F5 算法的提取函数定义为

$$r(\boldsymbol{x}') = e(0,\boldsymbol{x}') = e(\boldsymbol{s},\boldsymbol{x}) \oplus e(0,\boldsymbol{x}) = \boldsymbol{s} \tag{2}$$

2.2 湿纸码

湿纸码可以用来构造具有任意选择信道的隐写机制[8]。假设载体图像 X 由 n 个像素 x_i 组成，其中 $x_i \in \{0,1,\cdots,255\}$，令 $\boldsymbol{b} = \mathrm{LSB}(x_1,\cdots,x_n)$。发送方选择 k 个可用来嵌入消息的像素 x_j 作为选择信道，这里 $j \in J \subset \{0,1,\cdots,255\}$，$|J| = k$。令 $Y = \{y_i\}$ 为在图像 X 嵌入秘密消息后的载密图像，$\boldsymbol{b}' = \mathrm{LSB}(y_1,\cdots,y_n)$，接收方通过计算

$$\boldsymbol{D}\boldsymbol{b}' = \boldsymbol{m} \tag{3}$$

提取 p 比特的秘密信息 $\boldsymbol{m}$，其中 $\boldsymbol{D}$ 是由隐写密钥生成的 $p \times n$ 的伪随机二值矩阵。发送方的主要任务是求解变形后的式(3)，即

$$\boldsymbol{H}\boldsymbol{v} = \boldsymbol{m} - \boldsymbol{D}\boldsymbol{b} \tag{4}$$

再通过修改对应于 $\boldsymbol{v}$ 中非零元的载体像素来嵌入消息 $\boldsymbol{m}$。这里 $\boldsymbol{v} = \boldsymbol{b}' - \boldsymbol{b}$，$\boldsymbol{H}$ 为由 $\boldsymbol{D}$ 的 k 个列向量组成的子矩阵。

3　基于汉明码和湿纸码的隐写算法

先给出有关汉明码的两个重要引理。

引理 1[5]　给定一个长为 2^m-1 的线性完备码 C 及其奇偶校验矩阵 $\boldsymbol{H}$，对任意的第 i 个分量，$1 \le i \le 2^{m-1}-1$，存在第 $(2^{m-1}-1+i)$ 和第 (2^m-1) 个分量使得具有支集 $\mathrm{supp}(\boldsymbol{v}) = \{i, 2^{m-1}-1+i, 2^m-1\}$ 的向量 $\boldsymbol{v}$ 属于码 C。

引理 2[5]　设 C 是一个线性完备码，令 $\boldsymbol{x} = (x_1, x_2,\cdots,x_n) \in \mathrm{GF}(2^n)$，且 $\mathrm{supp}(\boldsymbol{x}) = \{i_1,i_2,i_3\}$，由 F5 嵌入函数(式(1))可知，存在第 4 个分量 i_4 使得具有支集 $\mathrm{supp}(\boldsymbol{x}') = \{i_1,i_2,i_3,i_4\}$ 的向量 $\boldsymbol{x}'$ 属于 C，其中 $i_4 = i_1 \oplus i_2 \oplus i_3$。

下面我们来构造基于汉明码和湿纸码的隐写算法。先将载体按矩阵排列，在矩阵块上对行使用汉明码，对列使用基于 meet-in-the-middle 算法实现的湿纸码[8]进行二次嵌入操作。在列嵌入时，利用汉明码的性质，可以保持行嵌入的结果，且不需引入过多的修改，大多数情况只需 1 个修改，极小概率需 3 个修改。然后在列嵌入时，利用湿纸码的性质避开将会引入 3 个修改的位置，在给定的嵌入率下，减少对载体的修改个数。

假设发送方和接收方共同协商好一个整数 x 及湿纸码的嵌入率 a，将载体数据分割成大小为 $N \times (2^x-1)$ 的矩阵，N 是发送方所选的湿纸编码的分组长度，且选择的湿纸码嵌入率为 a，平均修改是 D (在每个载体像素上平均修改的比特数)。在对载体矩阵块使用湿纸码进行列嵌入时最关键的步骤是确定“dry”位置的比例。由引理 1 及引理 2，只能对矩阵块第 1 至第 $2^{x-1}-1$ 列进行二次嵌入。

在第 1 列嵌入时，如果第 j 行嵌入时没做修改，那么把第 1 列第 j 个位置标记为“wet”，则“dry”位置的比例是 $(2^x-1)/2^x$，所以第 1 列平均可嵌入 $aN(2^x-1)/2^x$ 比特的秘密信息，平均要做 $DN(2^x-1)/2^x$ 个修改。假设在第 i 行使用汉明码嵌入时修改了第 r_i 个比特，在第 1 列使用湿纸码嵌入消息时需要修改第 i 个，为了不影响第 i 行的嵌入结果，由引理 1，还要修改第 2^{x-1} 和第 2^x-1 个比特，为了减少修改个数，根据引理 2，可用修改第 $r_i \oplus 2^{x-1} \oplus (2^x-1)$ 个比特来代替修改第 r_i，第 2^{x-1} 及第 2^x-1 这 3 个比特。

对矩阵块第 2 列嵌入时，最多也只有 $(2^x-1)/2^x$ 比例的位置可用(同上，如第 i 行利用汉明码嵌入时没做修改，则把第 2 列第 i 个位置标记为“wet”)，在这些位置中还要去掉那些可能影响到第 1 列嵌入的位置，如第 2 列嵌入需要修改第 j 个，假设第 j 行嵌入时修改了第 r_j 个，同上，为了不影响第 j 行嵌入，需把第 r_j 个修复，并修改第 $r_j \oplus (2^{x-1}+1) \oplus (2^x-1)$ 个，但如果 $r_j \oplus (2^{x-1}+1) \oplus (2^x-1) = 1$ 则会影响第 1 列的嵌入结果，所以要把第 2 列第 j 个位置标记为

“wet”。若矩阵块第 2 列的载体元素为 $(x_{12}, x_{22}, \cdots, x_{N2})^{\mathrm{T}}$，且在对矩阵块所有行和第 1 列完成嵌入之后 x_{12} 被标记为“wet”，则发送方需逐一对剩下没标记为“wet”的位置做如上假设进行计算，将所有会影响第 1 列嵌入的位置先标记为“wet”，其比例是 $1/(2^x-1)$，最终第 2 列“dry”位置的比例是

$$\frac{2^x-1}{2^x}\times\left(1-\frac{1}{2^x-1}\right)=\frac{2^x-2}{2^x} \tag{5}$$

平均可嵌入 $aN(2^x-2)/2^x$ 比特秘密消息。

类似地，可算出第 3 列至第 $2^{x-1}-1$ 列“dry”位置的比例为 $(2^x-t)/2^x$，这里 $t=3,4,\cdots,2^{x-1}-1$，因为除了去掉那些行不需修改的位置之外，还要去掉会影响前 $t-1$ 列嵌入结果的位置，平均可嵌入的秘密消息比特数为 $aN(2^x-t)/2^x$。

3.1 消息嵌入算法

基于前面介绍的隐写算法思想，本小节给出具体的消息嵌入算法。将载体数据分割成大小为 $N\times(2^x-1)$ 的矩阵块，取湿纸码分组长度为 $N=20/a$，嵌入率为 a (对矩阵块第 1 至第 $2^{x-1}-1$ 列均取嵌入率 a)，第 t 列平均修改为 D_t (由于各列“dry”位置的比例不同，可嵌入的秘密消息比特数也不同，所以各列平均修改均不等)，$t=1,\cdots,2^{x-1}-1$，发送方进行消息嵌入时主要分为两层：第一层是对载体矩阵块的行进行嵌入操作；第 2 层是对载体矩阵块的列进行嵌入操作。

第 1 层使用汉明码嵌入是按从第 1 行到第 N 行的顺序进行，具体步骤如下：

步骤1 读取 x 比特的秘密消息 $\boldsymbol{m}$；

步骤2 使用式(1)(F5 嵌入函数)嵌入 $\boldsymbol{m}$；

步骤3 重复步骤 1 和步骤 2 直到所有行嵌入完毕。

第 2 层使用湿纸码嵌入是按从第 1 列到第 $2^{x-1}-1$ 列的顺序进行，在对第 t 列进行嵌入操作时，$t=1,\cdots,2^{x-1}-1$，具体步骤如下：

步骤1 确定“dry”位置的个数 k_t；

步骤2 计算要嵌入的消息比特数 p_t，$p_t=aN(2^x-t)/2^x$，读取 p_t 比特秘密信息 $\boldsymbol{m}$；

步骤3 利用隐写密钥生成 $\boldsymbol{p}_t\times N$ 的伪随机二值矩阵 $\boldsymbol{D}$，并由 $\boldsymbol{D}$ 得到 $p_t\times k_t$ 的子矩阵 $\boldsymbol{H}$，通过求解 $\boldsymbol{Hv}=\boldsymbol{s}$ 嵌入消息 $\boldsymbol{m}$。

在列嵌入时使用的湿纸编码方法具体参阅文献[8]。在实际嵌入操作中，由于各个矩阵块相同列的“dry”位置个数不定，所以嵌入的消息比特数有可能小于 p_t (因 $\boldsymbol{H}$ 行不满秩，具体参见文献[7])。假设载体数据被分割成 n_B 块，则记所有块的相同列嵌入的消息比特数序列为 $p_{1,t},\cdots,p_{n_B,t}$，$t=1,\cdots,2^{x-1}-1$，要把这些序列传递给接收方，可以先采用自适应算法编码压缩后再传递。

3.2 消息提取算法

接收方要提取秘密信息很简单，因为知道了 x 和 a (双方共同协商)，从而也知道了 N 和 p_t，$t=1,\cdots,2^{x-1}-1$。由于隐写密钥是由发送方和接收方共享的，接收方知道了矩阵块每列嵌入的消息比特数之后，很容易利用隐写密钥将载密图像分割为 $N\times(2^x-1)$ 的矩阵块，然后对矩阵块的行和列进行秘密信息的提取。

本文将对矩阵块第 1 层和第 2 层的提取看作一个整体，接收方通过如下步骤提取出秘密信息。

步骤1 对矩阵块的每一行使用式(2)(F5 提取函数)提取出 x 比特的秘密信息。

步骤2 对矩阵块的每一列进行秘密信息提取。利用隐写密钥生成 $p_t\times N$ 的伪随机二值矩阵 $\boldsymbol{D}$，利用 $\boldsymbol{m}=\boldsymbol{Dy}$ 提取出 p_t 比特的秘密信息($\boldsymbol{y}$ 为载密矩阵块列向量)。

注意，步骤 1 与步骤 2 可交换。

4 性能分析

下面用平均失真和嵌入率来度量隐写算法的性能。在本文嵌入算法中，对所有行处理完之后，平均所作修改的个数为 $N(2^x-1)/2^x$；对载体矩阵块进行列嵌入时，各列平均所作的修改个数为 $D_tN(2^x-t)/2^x$，$t=1,\cdots,2^{x-1}-1$。故矩阵块的平均修改个数为

$$R_a=\sum_{t=1}^{2^{x-1}-1}\frac{D_tN(2^x-t)}{2^x}+N\times\frac{2^x-1}{2^x} \tag{6}$$

即平均失真为

$$\frac{R_a}{N\times(2^x-1)} \tag{7}$$

嵌入率为

$$\frac{\sum_{t=1}^{2^{x-1}-1}aN(2^x-t)/2^x+Nx}{N\times(2^x-1)} \tag{8}$$

下面给出 Bernoulli(1/2)-Hamming 情况下隐藏容量的理论界 $C(D)$[9]。它是在给定一个嵌入失真 D 零错误概率(即除了由隐写算法引起的修改之外，没有其他的修改被引入)的情况下嵌入率可达的上确界

$$C(D)=-D\log_2(D)-(1-D)\log_2(1-D) \tag{9}$$

这里 $0\le D\le 1/2$。

图 1 给出了本文算法和 PPC 及 F5 算法的隐写性能，其中横坐标是平均失真，纵坐标是嵌入率。从图 1 可以看出，在嵌入率为 0.1 至 0.8 bpp 的范围内本文算法性能明显优于 PPC 及 F5 算法。但其在

嵌入率小于0.1时不及 PPC 及 F5 算法，这是因为当参数 x 逐渐增大时，$[2^x-1,x,1]$ 汉明码嵌入率逐渐减小，且每在 2^x-1 比特载体数据中嵌入 x 比特秘密消息需要做一个修改的概率逐渐增大，当 x 增大到一定程度时，PPC 在列向量上要做 3 个修改的概率趋于零，若此时在列向量上使用湿纸码反而会降低性能。在本文构造的隐写算法中，若矩阵块的行选用 $[2^x-1,x,1]$ 汉明码，则湿纸码选择分组长度为 $N=20x$，嵌入率为 $1/x$ (即 $a=1/x$)时算法的隐写性能最好。

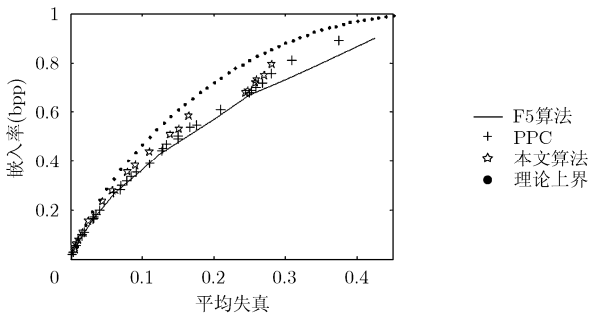

图 1 本文嵌入算法与 PPC，F5 算法的隐写性能比较

5 计算复杂度

下面讨论湿纸码分组长度的选取问题。从文献[8]的详细分析可知，当矩阵块每列可嵌入的消息比特数 p_t 逐渐增大直到 20 时，$t=1,\cdots,2^{x-1}-1$，对应于相同嵌入率的嵌入效率也逐渐增加，即平均修改率逐渐降低。若选择嵌入率为 a，分组长度为 $N=20/a$ 的湿纸码时，有 $10\le p_t<20$，此时在对列使用湿纸编码嵌入信息时嵌入容量的损失可忽略不计，且消息嵌入时间远远小于 $p_t=20$ 的情况。由于本文算法的主要计算量在使用湿纸码的列嵌入上，其计算复杂度为 $O({p_t}^2 2^{\beta p_t})$，其中 $\beta=H(H^{-1}\cdot(\alpha)/2)/\alpha<1$，略高于 PPC 的 $O(n)$，这里 $n=2^k-1$。但当 $p_t<20$ 时，这种计算复杂度换来的嵌入效率提高还是非常值得的。因此在选取了嵌入率为 a 的情况下，我们选择分组长度为 $N=20/a$ 的湿纸码以更好的降低计算复杂度。

6 结束语

本文通过将载体数据分割成矩阵块，在每个矩阵块上首先对行使用汉明码，并且在列上根据是否影响前面嵌入结果及是否需要 3 个修改引入“dry”和“wet”的概念，然后通过构造伪随机二值矩阵对列使用湿纸码来重复利用载体数据以提高性能。由于随机线性码普遍存在计算复杂度高的问题，所以本文采用选取相应于嵌入率 a 的分组长度为 $20/a$ 的湿纸码的技巧，使其符合小余数的随机线性码，且可利用计算复杂度仅为 $O({p_t}^2 2^{\beta p_t})$ 的算法[7]来求解。本文算法计算复杂度主要是使用湿纸码嵌入信息。寻找快速的求解方程 $\boldsymbol{Hv}=\boldsymbol{s}$ 方法降低算法计算复杂度是我们下一步将要研究的问题。

参 考 文 献

[1] Crandall R. Some notes on steganography. http://os.inf.tu-dresden.de/~westfeld/Crandall. pdf. 2007.12.

[2] Westfeld A. F5-A steganography algorithm[C]. Proc.4th International Workshop on Information Hiding, Lecture Notes in Computer Science, 2001, Vol.2137: 289–302.

[3] Tseng Y, Chen Y, and Pan H. A secure data hiding scheme for binary images[J]. *IEEE Transactions on Communications*, 2002, 50(8): 1227–1231.

[4] Fridrich J, Goljan M, and Lisonek P, *et al*.. Writing on wet paper[J]. *IEEE Transactions on Signal Processing*, *Third Supplement on Secure Media*, 2005, 53: 3923–3935.

[5] Rifa-Pous H and Rifa J. Product perfect codes and steganography[J]. *Digital Signal Processing*, 2009, 19(4): 764–769.

[6] Munuera C. Steganography and error-correcting codes[J]. *Signal Processing*, 2007, 87(6): 1528–1533.

[7] 张卫明, 李信然, 李世取. 线性隐写码的性质与构造[J]. 工程数学学报, 2007, 24(3): 547–550.
Zhang Wei-ming, Li Xin-ran, and Li Shi-qu. The properties and constructions of linear stegano- graphic codes [J]. *Chinese Journal of Engineering Mathematics*, 2007, 24(3): 547–550.

[8] Fridrich J, Goljan M, and Soukal D. Wet paper codes with improved embedding efficiency[J]. *IEEE Transactions on Information Security and Forensics*, 2006, 1(1): 102–110.

[9] Moulin P and Wang Y. New results on steganographic capacity[C]. Proceeding of 38th. Annual Conference on Information Sciences and Systems. Princeton, New Jersey, USA, 2004: 813–818.

朱雪秀：　女，1982 年生，硕士生，研究方向为信息隐藏.

刘九芬：　女，1963 年生，副教授，研究方向为小波理论及其应用和信息隐藏.

张卫明：　男，1976 年生，博士，讲师，研究方向为密码学和信息隐藏.