基于乘同余的 LSB 数字图像信息隐写

聂 鹏
(江西财经大学现代教育技术中心 江西 南昌 330013)

摘 要 为了提高针对 LSB 隐写的抗隐写分析能力,提出在数字图像上利用乘同余的离散特性实现 LSB 信息隐写,使嵌入信息在载体图像噪声空间内得到均布,提高隐写信息的抗分析能力的方案;研究乘同余和载体图像位面噪声特征;对比分析信息隐写前后的图像噪声特征,并通过实验验证基于乘同余的 LSB 数字图像信息隐写方法可加强隐藏信息的隐秘性。

关键词 信息安全 信息隐写 图像噪声 乘同余

中图分类号 TP301 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2013.07.052

MCM-BASED LSB STEGANOGRAPHY FOR DIGITAL IMAGES

Nie Peng
(*Center of Modern Education Technology, Jiangxi University of Finance and Economics, Nanchang 330013, Jiangxi, China*)

Abstract For improving the ability of anti-steganalysis against LSB steganography, we propose that to implement the LSB information steganography on digital image using the discrete characteristics of the multiplicative congruential method (MCM) to make the embedding information to be distributed uniformly in noise space of carrier image so as to improve the anti-analysis ability of the steganography information; we studied the MCM and the noise features on bit-plane of the carrier image; comparatively analysed the noise features of image before and after the steganography, and verified through the experiment that the MCM-based LSB information steganography for digital images can enhance the secrecy of information hiding.

Keywords Information security Steganography Image noise Multiplicative congruential method

0 引 言

随着计算机网络的发展和普遍应用,大量信息在大小不同的公共网络上频繁传输,其信息安全性因为网络环境的开放性和复杂性受到严重威胁。随着政府上网、电子商务、电子政务、网络银行等应用,越来越多的重要信息需要安全的传输,单纯地使用密码技术仍显单薄。为了确保信息能够安全传递并掩盖信息传递的事实,需要对网络上传输的信息进行隐写。由于数字图像在网络上大量存在,因此成为信息隐写的重要载体。

在数字图像中隐藏信息方法很多,最低有效位算法 LSB[1] 由于信息隐写容量大,易于实现受到广泛青睐。LSB 算法除可应用与数字图像隐写外,还可广泛用于音频的信息隐写。LSB 主要是用秘密信息比特替换载体图像的量化噪声和可能的信道噪声,正常的量化噪声应该是高斯分布的白噪声,而使用简单 LSB 隐写后的图像噪声分布就可能不在满足高斯分布,因此可以通过噪声能量分布发现载体图像是否嵌入了隐藏信息[2-4]。

基于乘同余[5] 的 LSB 数字图像信息隐写将简单 LSB 隐写加以改造,使载体图像在加入隐藏信息后的噪声分布仍然可以较好地得到保持,从而具备较好的抗信息隐藏分析能力。

1 LSB 数字图像信息隐写分析

LSB 数字图像隐写[6] 是利用人眼对图像的分辨力受限制这一生理现象作为理论基础[7]。人眼的分辨力是指人眼在一定距离上能区分相邻两个图像像素的能力,对于色差较小的两个像素人眼将无法分辨它们的不同。LSB 则是利用这一点将信息写入图像的噪声位面(人眼感兴趣的图像能量通常不分布在噪声位面中),使写入位于噪声位面中的信息造成的像素色差发生可控的轻微改变,但不足以让人眼察觉,从而达到了隐匿信息的目的。但是 LSB 信息隐写事实上仍然改变了图像噪声的能量分布,往往简单的 LSB 信息隐写方法遗留在噪声平面上的能量分布不是均匀的,可以通过对图像噪声能量的统计分析发现图像中的隐写信息。

图 1 是在 google 图像检索中分别以“人像”、“风景”、“生活”三个关键字得到的 150 幅 24 位 BMP 图像的基础上生成的

收稿日期:2012-07-17。江西省科技厅基金项目(2009BHA19200)。聂鹏,讲师,主研领域:信息安全,信息隐写,数字水印。

在各位面图像能量累计分布比例。

噪声位面的选取可以由图 1 的数据得出。如图 1 所示,1 号 -2 号位面上的能量占图像全部能量的 1.3%,水平非常低;位面 3 为 3.1%,已经开始具备较高的能量水平;其后的位面能量约以 2 的指数开始迅速上升。因此在本文方法中噪声位面定位为 1 号、2 号位面。

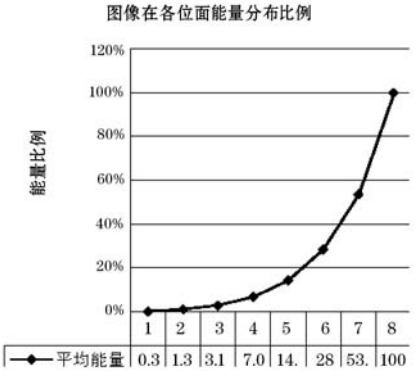
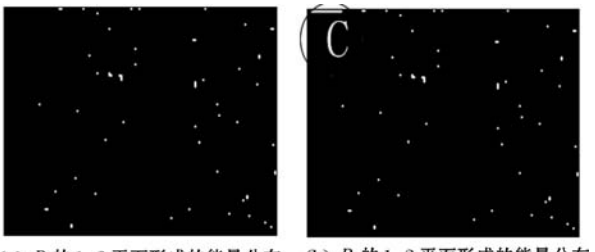


图 1 图像在各位面能量分布比例

从 150 张图像中随机抽取一幅图像并命名为 P_1,采用简单 LSB 将一张 40×30 像素的全白 BMP 图像隐写入 P_1,生成的图像为 P_2,分别绘制 P_1、P_2 的 1 号、2 号位面能量在二维空间的分布情况,如图 2 所示。



(a) P_1的1、2平面形成的能量分布 (b) P_2的1、2平面形成的能量分布

图 2 P_1、P_2 的 1、2 号位平面能量分布

图 2(a)为 P_1 的 1、2 平面形成的能量分布情况,白色点表示有能量;图 2(b)为 P_2 的 1、2 平面形成的能量分布情况,易见在图 2(b)中存在标志为 C 的能量异常区域,即存在隐写造成的能量异常,这种异常能量成连续分布,易被分析工具发现[8,9]。

2 乘同余的 LSB 隐写分析

解决上述问题的核心在于将写入图像噪声空间的秘密信息能量尽量在噪声空间中随机离散地分布,使得隐写分析更难于发现图像的能量异常。

将隐写信息进行离散化使用的随机数可以用专门的物理装置产生,如放射性粒子计数器,电子管随机数发生器等;但这些方法的成本很高而且使用不方便,因此通常使用的随机数是由计算机生成的伪随机数。乘同余算法是生成伪随机数的一种方法,本文使用乘同余算法作为改进数字图像隐写算法的伪随机数发生器。

乘同余算法递推计算公式如式(1)所示:

$$x_n = ax_{n-1} \bmod k \quad n \geqslant 0 \tag{1}$$

其中 $n \in (1,2,\cdots)$;x_0 为随机数初值;a 为非负乘子,k 为模数;$x \bmod k$ 表示 x 对 k 取模的最小非负数。

设随机变量 X 的取值范围是(0,1),且对任意的 $0<a<1$ 存在 $P\{0<X<1\}=a$,则称 X 服从 (0,1) 上的均匀分布,乘同余算法符合上述分布。如:$k=2^{31}-1$,$a=7^5$,$x_0=1$ 时,产生的随机数列为 {16807, 282475249, 1622650073, 984943658, 1144108930,…}。

经乘同余算法处理的秘密信息降低了隐写信息的相关性,在秘密信息替换图像最低位平面的过程中可保持载体图像的复杂度正态分布特征,可降低检测软件发现载体图像中含有秘密信息的概率。

3 乘同余 LSB 隐写算法(MCM LSB)

3.1 算法符号描述

符号	描述
M	需要被隐写的秘密信息
L_m	M 的位长度
P	载体图像
W_p	载体图像宽度
H_p	载体图像高度
t	载体图像噪声位面数
P'	载体图像的噪声空间
P'_i	载体图像噪声空间中第 i 个元素
M'	秘密信息所在的隐写元素空间
M'_i	载体图像中第 i 个隐写元素

3.2 算法步骤

Step1 选取 24 位 BMP 图像 P(不考虑 RLE 压缩格式),并根据式(2)计算图像 P 的噪声空间 P'的大小。

$$|P'| = W_p \times H_p \times 3 \tag{2}$$

Step2 考虑秘密信息 M 需要存入的噪声空间 P' 的位面大小为 t,则根据式(3)计算 M'_i 所在空间大小。

$$|M'| = L_m / t \tag{3}$$

其中 $\{t \mid 1 \leqslant t \leqslant 8, t \in N\}$,且 $|P'| \geqslant |M'|$。

Step3 考虑乘同余算法式(4)对于 P' 和 M' 的参数选取。

$$x_n = ax_{n-1} \bmod k \tag{4}$$

将秘密信息 M 中单个字节的信息隐写到 1 至 t 号噪声平面中,需要将该字节信息拆分为 $8/t$ 个信息片段,每个信息片段即一个隐写元素 M'_i。因为 $|P'| \geqslant |M'|$,故 M'_i 在 $|P'|$ 中可用于隐藏信息的噪声空间元素 P'_i 的个数不小于 1,为 $\lfloor |P'|/(L_m/t) \rfloor$ 个。乘同余 LSB 隐写算法在 $\lfloor |P'|/(L_m/t) \rfloor$ 个 P'_i 元素中选择一个分配给 M_i 用于信息隐写,因此式(4)中 $k=\lfloor |P'|/(L_m/t) \rfloor$,$x_0$ 与 a 取正整数。

Step4 确立 M' 到 P' 的空间映射关系。每 $\lfloor |P'|/(L_m/t) \rfloor$ 个 P'_i 中只有一个元素保存隐写的秘密信息,因此 M' 到 P' 的空间映射关系为如式(5)所示。

$$M'_i \rightarrow P'_{k(i-1)+1+(ax_{i-1} \bmod k)} \tag{5}$$

其中,$k=\lfloor |P'|/(L_m/t) \rfloor$。

隐藏信息的提取是本算法的逆过程,只需确定 a、k 和 x_0,即可将隐写的秘密信息提取还原。

3.3 软件编程与实现方法

图 3 为 MCM LSB 隐写算法的软件流程图。

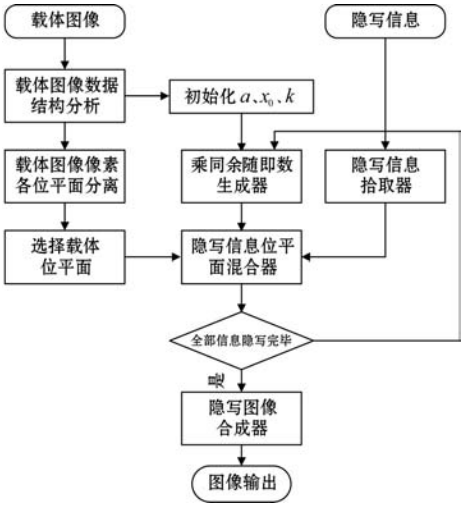


图 3　MCM LSB 软件流程图

MCM LSB 隐写算法由 Delphi 语言具体实现,输入信息包括载体图像输入和隐写信息输入,输出为隐写后的图像。首先 MCM LSB 隐写算法对输入的载体图像结构信息进行分析,包括图像宽度、高度、色彩位平面数。软件根据这些信息确定乘同余算法的初始参数,乘同余随机算法生成器产生信息隐写位置,隐写信息拾取器提取隐写信息内容,然后由隐写信息位平面混合器生成含有隐写信息的图像位平面。最终隐写图像合成器将全部位平面合成为完整图形输出。图像还原算法为隐写算法的逆过程。隐写信息位平面混合器核心伪代码如下:

```
p: =0;
for i: = 1 to byteSizeOf(data) do begin
  dataByte : = ord(data[i]);
  for j: = 0 to 3 do begin
    x: = (a * xOld) mod k;
    xOld: = x;
    eBmpFile. Position: = bmpHeadOffSet + k * p + x;
    eBmpFile. Read(bmpData,1);
    asc: = pickup(dataByte);
    asc: = (ord(bmpData) and bitPlaneMask) + asc;
    writeBuffer(asc);
    inc(p);
  end;
end;
```

隐写信息位平面混合器中 data 为隐写信息,eBmpFile 为载体图像,bmpHeadOffSet 为载体图像位平面数据区偏移量,bmpData 为载体图像位平面数据,pickup( )为隐写信息拾取器,bitPlaneMask 为位平面选择掩码,writeBuffer( )将混合完成的数据写入缓冲区等待隐写图像合成器生成最终结果。

4　实验与分析

本节对 MCM LSB 的隐写抗分析性能进行验证分析,具体实验环境如下:采用 IBM P5 作为实验服务器,配置为 4 × POWER5 + CPU、4G 内存、操作系统采用 Windows 7 。

4.1　实验设计

图 4 中的载体图像和隐写信息来自 google 图片的随机检索,选取图 4(a)中的 film. bmp 作为隐写载体图像 P,选取图 4(b)中 info. jpg 作为隐写信息。其中 P 为宽度 800 像素、高度 600 像素、24 位彩色图像、分辨率为 72dpi;info. jpg 为宽度 127 像素、高度 95 像素、24 位彩色图像、分辨率为 72dpi、3820 字节大小。MCM LSB 算法各参数计算如下:

$|P'| = 800 \times 600 \times 3 \times 8 = 1440000$ bit ;

$L_m = 3820 \times 8 = 30560$ bit ;

$t = 2$(由本文对 LSB 图像能量在各个位面的分布情况分析可知 $t = 2$ 为合理取值);

$a = 8 \quad x_0 = 2$;

$|M'| = L_m / t = 30560/2 = 15280$,即 M'拥有 15280 个需要写入 P'的元素;

$k = \lfloor |P'| / (L_m / t) \rfloor = \lfloor 1440000/15280 \rfloor = 94$,即 94 个元素 P'_i中含一个 M'_i;

故 M'到 P'的空间映射关系为 $M'_i \rightarrow P'_{94(i-1)+1+(94x_{i-1} \bmod 94)}$。

图 4　MCM LSB 隐写算法实验数据

4.2　统计特性分析

载体图像 P 的最低位平面近似随机噪声,相邻比特位能量表现为弱相关。复杂度是位平面小块中能量相异的像素对的总数,通常满足中心极限定理,即载体图像 P 最低比特位复杂度近似服从正态分布,而原始图像嵌入未经乘同余处理的秘密信息后,最低位平面的正态分布分布特性将会被破坏,如图 5 所示。

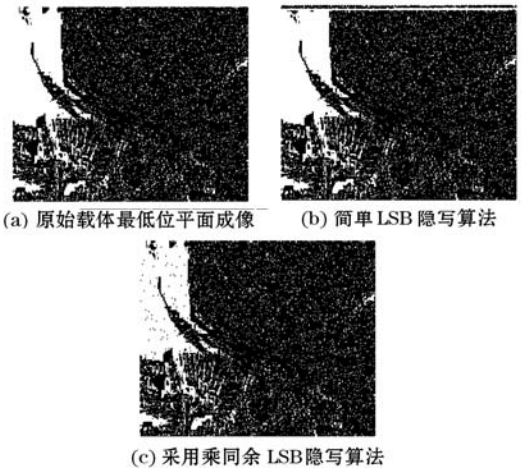


图 5　原始载体图像、LSB 隐写算法和 MCM LSB 隐写算法的低位面能量图像

将图 5 所示各低位面能量成像分割为 4 × 4 像素大小的块,并对图 5 中的(a)至(c)进行复杂度分布统计,统计结果如图 6 所示;其中横坐标为复杂度,纵坐标为特定复杂度在低位面中出

(下转第 212 页)

表 1 网络安全模型加密速度测试表

日期与时间	Logistic映射迭代次数	密码明文长度	随机抽取密码源生成函数	p、q取值长度	加密时间（秒）
2010-11-02 09：32：42	300次	320bit	异或+一元函数	6位素数	0.856
2010-11-11 18：23：06	300次	320bit	双变序函数	8位素数	1.35
2010-11-22 21：45：37	300次	320bit	异或+变序函数	10位素数	3.687

显然，本模型加密速度主要受到 p、q 取值的影响，我们再以自动生成 6、8、10 位素数的方法对 320bit 的密码明文作单独的 RSA 加密测试，其加密时间分别为 0.738、1.186、3.553 秒，可见整个模型并不因为多次加密处理而过多影响其加密速度。

从文献[3]看，目前单独的 RSA 加密，其密钥长度一般达到 1024 位以上才被认为是安全的，当前主流配置的 PC 机破解 1024 位密钥所需要时间约 1 年以上，即使破解 64～256 位公认的弱密钥最少也在几个小时以上，远未达到“秒”级。由于本安全模型的密钥是随着时间动态变换并随机生成的，其时效性极短，且模型的加密速度非常快，密钥生成至明文加密整个过程的速度均为“秒”级，因此完全可以在密码时效内抵御各类攻击。

5 结 语

本安全模型并不适应所有场景的加密，仅适合对少量明文的加密，如网络密码、网络消息等。由于本模型核心思想在于动态加密和动态密码源。如能在动态密码源的生成中引入硬件辅助，如 USB-KEY 等，那么动态密码源的安全性会更高，但成本也会更大。在以后的改进中我们还可以引入双混沌算法，以增加模型随机数的不确定性和实现随机抽取事务，这将会使本模型加密的安全性更高。

参 考 文 献

[1] Kocarev L, Jakimoski G, Stojanovski T, et al. From chaotic maps to encryption schemes[C]//Proc. IEEE Int. Sym. CAS. 1998, 4: 514-517

[2] Gao H J, Zhang Y S, Liang S Y, et al. A New Chaotic Algorithm for Image Encryption[J]. Chaos, Solitons and Fractals, 2006, 29(2): 393-399.

[3] 钟诚，赵跃华. 信息安全概论[M]. 武汉：武汉理工大学出版社，2003：91-109.

[4] 郑伟谋. 实用符号动力学[M]. 上海：上海科技教育出版社，1994：159-167.

[5] 丁存生，萧镇国. 流密码学及其应用[M]. 北京：国防工业出版社，1994：1-15.

[6] 吕金虎，陆君安，陈士华. 混沌时间序列分析及其应用[M]. 武汉：武汉大学出版社，2005：31-45.

[7] 盛利元，曹莉凌，孙克辉，等. 基于 TD-ERCS 混沌系统的伪随机数发生器及其统计特性分析[J]. 物理学报，2005：4031-4036.

[8] 范九伦，张雪峰. 分段 Logistic 混沌映射及其性能分析[J]. 电子学报，2009：720-725.

[9] 李恩，吴敏，熊永华. 一种基于双混沌映射的加密算法设计与应用[J]. 计算机应用研究，2009：1512-1514.

（上接第 199 页）

现的次数。

如图 6 所示，未经本文算法处理的秘密信息由于具有较强的相关性，在秘密信息替换图像的最低位平面的过程中容易改变载体图像（图 5(a)）的复杂度分布特征，呈现出明显的非正态分布，易于被检测软件发现载体图像 P 中写有秘密信息。经本文算法隐写的秘密信息，由于经过离散处理，最大程度地保持了载体图像 P 的复杂度分布特征，在图 6 中图 5(c)的复杂度分布仍表现为正态分布。

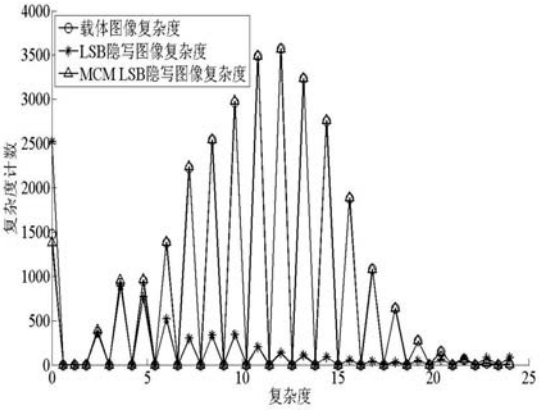


图 6 载体、LSB 隐写和 MCM LSB 隐写图像的低位面复杂度分布

综上所述，采用简单 LSB 算法的图 5(b)有明显的非离散能量分布，在图 6 的复杂度分布统计特性中表现为复杂度的非正态化，容易暴露隐藏的秘密信息；采用了乘同余 LSB 隐写算法的图 5(c)能量分布离散性好，在复杂度分布统计特性中也与载体图像接近，从而使隐写的秘密信息具有更高的隐蔽性。

5 结 语

本文提出了基于乘同余的 LSB 数字图像隐写算法，该算法利用乘同余的特性将隐写信息的能量离散随机地均布在载体图像噪声空间中；实验证明了该算法可以降低隐写信息的可探测性，有效地提高了秘密信息在开放网络环境中的传输安全性。

参 考 文 献

[1] 杨义先. 数字水印基础教程[M]. 北京：人民邮电出版社，2007：274.

[2] 张新鹏，王朔中，张开文. 基于统计特性的 LSB 密写分析[J]. 应用科学学报，2004，22(1)：16-19.

[3] 孙露霞，陈丽亚，李建华. 彩色 BMP 图像 LSB 隐藏的检测算法[J]. 计算机工程与应用，2005，41(30)：43-45.

[4] Yu X, Babaguchi N. Run length based steganalysis for LSB matching steganography[M]. Hannover, Germany: IEEE Press, 2008.

[5] 耿素云，屈婉玲，张立昂. 离散数学[M]. 北京：清华大学出版社，2008：257.

[6] 徐旭，平西建，张涛，等. 针对 LSB 匹配隐写的图像复原隐写分析[J]. 计算机辅助设计与图形学学报，2009，21(2)：262-267.

[7] 张春田. 数字图像压缩编码[M]. 北京：清华大学出版社，2006：471.

[8] Fridrich J, Goljan M. Practical Steganalysis of Digital Images - State of the Art[C]//San Jose, California: International Society for Optical Engineering, 2002.

[9] Chandramouli R, Kharrazi M, Memon N. Image Steganography and Steganalysis: Concepts and Practice[M]. Seoul, Korea: Springer, 2004.