对空域图像 LSB 隐写术的提取攻击

张卫明[1),2)] 李世取[1)] 刘九芬[1)]

[1)](信息工程大学信息研究系 郑州 450002)

[2)](上海大学通信与信息工程学院 上海 200072)

摘 要 隐写分析的研究一直集中于检测隐藏信息的存在性,而关于如何提取隐藏信息(即对隐写术的提取攻击)的研究还非常少.对于使用密钥的隐写术,提取攻击等价于恢复隐写密钥.文中结合隐写分析中的检测技术和密码分析中的相关攻击技术,对空域图像 LSB 隐写术提出了一种隐写密钥恢复方法.理论分析说明:此攻击方法的计算复杂度主要由所需的样本量决定,并且当嵌入率 r 接近 0 或 1 时攻击将失败.作者通过混合高斯模型给出了一个估计最小样本量的方法.针对隐写软件"Hide and Seek 4.1"的实验表明:此攻击方法可以成功恢复隐写密钥,从而提取隐藏的消息.如果消息长度 L 未知,当嵌入率 $5.3\%<r<94.7\%$ 时攻击可以成功;如果 L 已知,当 $1.1\%<r<98.4\%$ 时攻击可以成功,并且当 $11\%<r<50\%$ 时,使用估计的最小样本量可以将攻击速度提高 10%~45%.

关键词 隐写术;隐写分析;提取攻击;相关攻击;混合分布

中图法分类号 TP309

Extracting Attack to LSB Steganography in Spatial Domain

ZHANG Wei-Ming[1),2)] LI Shi-Qu[1)] LIU Jiu-Fen[1)]

[1)](*Department of Information Research, Information Engineering Institute, Zhengzhou* 450002)

[2)](*School of Communication and Information Engineering, Shanghai University, Shanghai* 200072)

Abstract Many steganalysis techniques have been developed to detect the existence of hidden information. However, there are few approaches about how to extract the hidden information, i.e., extracting attack to steganography. For key-dependent steganography, extracting attack is equivalent to recovering the stego key. By combining detecting techniques in steganalysis and correlation attack in cryptanalysis, the authors propose a new method to recover the stego key of LSB steganography in spatial domain of images. Theoretic analysis is made for this method, which establishes that its complexity is mainly determined by the number of samples and the attack will fail when the embedding rate r tends to 0 or 1. The authors introduce an estimator for the minimum number of samples n based on a mixture Gaussian model. The experiments on the stego software "Hide and Seek 4.1" show that the proposed method can recover the stego key and extract the hidden message. If the length of messages L is not known, the attack can success for $5.3\%<r<94.7\%$; if L is known, the attack can success for $1.1\%<r<98.4\%$, and by using the estimator for n, the attacking speed can be increased by 10%~45% when $11\%<r<50\%$.

Keywords steganography; steganalysis; extracting attack; correlation attack; mixture distribution

收稿日期:2004-05-17;最终修改稿收到日期:2007-06-06.本课题得到国家自然科学基金(60473022)资助.**张卫明**,男,1976年生,博士,讲师,主要研究方向为密码学和信息隐藏.E-mail:nlxdweiming@sohu.com.**李世取**,男,1945年生,教授,博士生导师,主要研究领域为密码学和概率论.**刘九芬**,女,1963年生,博士,副教授,主要研究方向为信息隐藏.

1 引 言

隐写术是信息隐藏的一个重要分支,其内容是研究如何把消息隐藏于数字多媒体数据(如文本、图像、音频或视频)中来实现隐蔽通信.本文是针对图像隐写术的.图像隐写术指:发送者首先选择一幅"载体图像",把要发送的消息通过某种隐写算法嵌入其中生成"载密图像",然后把"载密图像"发送给接收者,嵌入过程由发送者和接收者所共享的"隐写密钥"所控制,使得只有合法的接收者才可以检测并提取出消息.

对隐写术的攻击被称为隐写分析,狭义的隐写分析专指检测技术,即只要能检测到嵌入消息的存在,就认为攻击成功.文献[1-2]对隐写分析过程作了细致划分:

(1)区分载体图像和载密图像,即检测是否存在嵌入消息;

(2)判断隐写术类型;

(3)判断所使用的隐写软件;

(4)恢复隐写密钥,提取嵌入消息;

(5)若嵌入消息为密文,则解密所提取的消息,获得明文消息.

其中第(5)步实际上是密码分析.传统的隐写分析主要集中于对前3步的研究,尤其是第(3)步,即检测嵌入消息的存在.现在有些算法不仅可以可靠地检测图像中是否隐藏了消息,而且可以估计消息的长度,如文献[3-4];另外,在检测的基础上,可以通过某些特征判断发送者使用了什么类型的嵌入方法,甚至可确定他使用的何种隐写软件[5-6].

本文把上述第(4)步的工作称作"提取攻击",目前关于"提取攻击"的公开文献还很少,Chandramouli[1]针对基于扩频通信的隐写,就一种特殊情况给出了提取攻击方法.他考察的情况是同一消息使用同一载体发送了两次,前后两次的差别仅在于嵌入消息时所用的强度因子不同.Fridrich等借鉴隐写分析中区分载体和载密图像的卡方检验来区分真伪密钥,他们最先用这种方法恢复JPEG图像LSB隐写的密钥[2],后来又将其一般化,用于攻击空域图像LSB嵌入和±1嵌入[7].

我们认为对隐写算法的提取攻击本质上是一种特殊的密码分析,它需要结合使用传统的隐写分析技术(检测技术)和密码分析技术.基于这种观点,我们针对空域图像LSB隐写提出了一种提取攻击方法,基本思路如下:首先利用已有的检测方法估计嵌入消息的长度(也即估计嵌入率);然后对图像进行均值滤波,获取噪声数据,经过适当的处理,把噪声数据看成是来自一个混合分布[8]的样本;通过对混合模型的分析,寻找某种相关优势;最后基于此相关优势,使用密码分析中的相关攻击方法恢复隐写密钥,从而提取嵌入消息.利用此方法,我们实现了对隐写软件Hide and Seek 4.1①的成功攻击.

2 统计模型

2.1 随机LSB隐写模型

用$\boldsymbol{C}=\{c_{ij}\}$表示载体图像,$\boldsymbol{S}=\{s_{ij}\}$表示对应的载密图像,$1\leq i\leq w, 1\leq j\leq h, wh=N$.其中$c_{ij}$和$s_{ij}$都是取值于[0,255]的整数.用$\boldsymbol{M}=\{m_1,m_2,\cdots,m_L\}$($L\leq N$)表示嵌入消息(一般为加密后的伪随机序列),$m_i\in\{0,1\}$,$1\leq i\leq L$.用$k$表示隐写密钥,它取值于密钥空间$\mathcal{K}$.

随机LSB隐写算法的嵌入过程一般如下:首先利用隐写密钥$k_0\in\mathcal{K}$通过一个伪随机数发生器G生成一条长为L的嵌入路径,然后用消息$\boldsymbol{M}=\{m_1,m_2,\cdots,m_L\}$替换此路径上对应像素的LSB(最不重要比特)位,从而得到载密图像$\boldsymbol{S}$.合法的接收者拥有隐写密钥k_0,所以可以很容易从载密图像$\boldsymbol{S}$中读出嵌入的消息.本文假设攻击者已知隐写算法的全部细节,但是不知发送者使用的密钥k_0,我们的目的是研究在此条件下,如何只利用载密图像$\boldsymbol{S}$恢复出密钥k_0,从而获得嵌入的消息.

为增加隐蔽性,一般嵌入算法要把L比特消息近似均匀地扩散到整幅图像上,所以若记嵌入率为$r=\frac{L}{N}$,则一个像素被选到来承载消息的概率是r,进一步可假设消息与图像的LSB位独立,从而消息和图像的LSB位相等的概率是$\frac{1}{2}$,所以一个像素被修改的概率是$\frac{r}{2}$.

2.2 噪声数据的混合分布模型

用LSB方法隐藏的消息对图像而言是一种加性噪声,所以下面我们以隐写图像的噪声数据为对象进行分析.首先对载密图像$\boldsymbol{S}=\{s_{ij}\}$进行如下均值滤波:

$$\bar{s}_{ij}=\frac{1}{4}(s_{i,j+1}+s_{i,j-1}+s_{i-1,j}+s_{i+1,j}),\quad 1<i<w,\ 1<j<h.$$

① Hide and Seek 4.1. Available: http://71.6.196.237/fravia/stego.htm

特别地，对于图像 4 条边界上的点取其周围 3 个像素的均值，4 个角上的点则取其周围 2 个点的均值，注意此处的均值不是取整数而是保留 3 位小数的实数. 滤波后的图像记作 $\bar{\boldsymbol{S}}=\{\bar{s}_{ij}\}$. 下面为了方便，我们将图像数据矢量化，即 $\boldsymbol{C}=\{c_1,c_2,\cdots,c_N\}$，$\boldsymbol{S}=\{s_1,s_2,\cdots,s_N\}$，$\bar{\boldsymbol{S}}=\{\bar{s}_1,\bar{s}_2,\cdots,\bar{s}_N\}$. 对应的噪声数据 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_N\}$ 按如下方式得到：如果 s_i 是奇数，则 $v_i=s_i-\bar{s}_i$；如果 s_i 是偶数，则 $v_i=\bar{s}_i-s_i$，$i\leqslant 1\leqslant N$.

对于自然图像而言，上述滤波是对原始图像的较好的估计，估计的误差服从均值接近零的对称分布，可以用广义高斯分布来刻画[9]. 在消息嵌入过程中，像素 c_i 周围的像素被加 1 和减 1 的概率是相等的，所以载密图像的滤波值 $\bar{s}_i$ 依然是对 c_i 的较好的估计. 如果嵌入过程中 c_i 没有被修改，即 $s_i=c_i$，差值 $s_i-\bar{s}_i=c_i-\bar{s}_i$ 可看成是取自一个零均值对称分布 $F(x)$ 的样本. 设随机变量 X 服从分布 $F(x)$，因 $F(x)$ 是均值为零的对称分布，所以 $-X$ 与 X 同分布，因而当 $s_i=c_i$ 时，无论噪声值取 $v_i=s_i-\bar{s}_i$ 还是 $v_i=\bar{s}_i-s_i$，都有 v_i 与 X 同分布. 另一方面，若嵌入过程中 c_i 被修改，则有两种可能：(1) c_i 为偶数且其 LSB 位由 0 替换为 1 得到 s_i，所以 $s_i=c_i+1$ 为奇数，此时 $v_i=s_i-\bar{s}_i=c_i-\bar{s}_i+1$，因而 v_i 与 $X+1$ 同分布，记此分布为 $G(x)$；(2) c_i 为奇数且其 LSB 位由 1 替换为 0 得到 s_i，所以 $s_i=c_i-1$ 为偶数，此时 $v_i=\bar{s}_i-s_i=\bar{s}_i-c_i+1$，因而 v_i 同样与 $X+1$ 同分布. 注意到当嵌入率为 r 时，像素以 $\frac{r}{2}$ 的概率被修改，所以噪声数据 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_N\}$ 可看成是来自一个混合率为 $\frac{r}{2}$ 的混合分布的样本，其分布为

$$F_{\frac{r}{2}}(x)=\left(1-\frac{r}{2}\right)F(x)+\frac{r}{2}G(x) \tag{1}$$

现我们从密钥空间 $\mathcal{K}$ 中选择一个密钥 k，以 k 为种子用伪随机数发生器 G 生成一条随机路径 $\{j_1,j_2,\cdots,j_L\}\subset\{1,2,\cdots,N\}$. 若 k 是伪密钥(即 k 不是嵌入消息时用的密钥 k_0)，则对应的噪声数据子集 $\boldsymbol{V}_k=\{w_{j_1},w_{j_2},\cdots,w_{j_L}\}$ 是对噪声数据 $\boldsymbol{V}$ 的一个随机抽样，所以 $\boldsymbol{V}_k$ 中的数据仍服从混合分布(1)；若选到的密钥恰好是真密钥 k_0，则此时路径 $\{j_1,j_2,\cdots,j_L\}$ 对应的载密图像的像素子集 $\boldsymbol{S}_{k_0}=\{s_{j_1},s_{j_2},\cdots,s_{j_L}\}$ 中，大约有一半与载体图像对应位置的像素值相同，而另一半是经过修改得到的. 所以对应的噪声数据子集 $\boldsymbol{V}_{k_0}=\{v_{j_1},v_{j_2},\cdots,v_{j_L}\}$ 服从混合率为 $\frac{1}{2}$ 的混合分布：

$$F_{\frac{1}{2}}(x)=\frac{1}{2}F(x)+\frac{1}{2}G(x) \tag{2}$$

分布式(1)与式(2)的差异是我们区分真伪密钥的依据.

2.3　碰撞优势

取实数 $A>0$，记 $\alpha_0=P\{X>A\}$，$\alpha_1=P\{X+1>A\}$，因为 $X\sim F(x)$，$X+1\sim G(x)$，所以

$$\alpha_0=\int_A^{+\infty}\mathrm{d}F(x),\quad \alpha_1=\int_A^{+\infty}\mathrm{d}G(x) \tag{3}$$

令 $\Delta\alpha=\alpha_1-\alpha_0$，易知 $\Delta\alpha>0$，即修改点对应的噪声值大于 A 的概率比未修改点对应的噪声值大于 A 的概率要大. 进一步，设真密钥生成的路径上噪声值大于 A 的概率为 p_0，伪密钥生成的路径上噪声值大于 A 的概率为 p_1，则

$$p_0=\int_A^{+\infty}\mathrm{d}F_{\frac{1}{2}}(x)=\frac{1}{2}\alpha_0+\frac{1}{2}\alpha_1 \tag{4}$$

$$p_1=\int_A^{+\infty}\mathrm{d}F_{\frac{r}{2}}(x)=\left(1-\frac{r}{2}\right)\alpha_0+\frac{r}{2}\alpha_1 \tag{5}$$

二者的差

$$\Delta p=p_0-p_1=\frac{1}{2}(1-r)(\alpha_1-\alpha_0)=\frac{1}{2}(1-r)\Delta\alpha \tag{6}$$

因为随机 LSB 隐写算法要求嵌入率 $r<1$，又由上面的分析知 $\Delta\alpha>0$，所以 $\Delta p>0$，即用真密钥对噪声数据抽样较之伪密钥的抽样，碰到大于 A 的值的概率大. 我们把 Δp 称作“碰撞优势”，它类似于密码分析中的“符合优势”，当我们有足够大的优势时，就可以恢复出真密钥. 当 r 给定，欲使 Δp 大，我们需要选取合适的 A 使 $\Delta\alpha$ 尽可能大. 而为了计算 A，p_0 和 p_1，我们需要知道分布 $F(x)$. 如 2.2 节提到的，对于自然图像，$F(x)$ 近似满足零均值的广义高斯分布[9]. 值得注意的是我们只能观测到混合分布的数据，而通过混合的广义高斯分布估计参数较困难. 为了分析方便，我们采用简化的假设，设 $F(x)$ 是均值为 0、方差为 σ^2 的高斯分布 $N(0,\sigma^2)$ 的分布函数，则相应地，$G(x)$ 是高斯分布 $N(1,\sigma^2)$ 的分布函数. 定义标准正态分布函数：

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{+\infty}\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y \tag{7}$$

则由式(3)得 $\alpha_0=Q\left(\frac{A}{\sigma}\right)$，$\alpha_1=Q\left(\frac{A-1}{\sigma}\right)$，所以 $\Delta\alpha=Q\left(\frac{A-1}{\sigma}\right)-Q\left(\frac{A}{\sigma}\right)$，易知当 $\frac{A-1}{\sigma}=-\frac{A}{\sigma}$，即 $A=\frac{1}{2}$ 时，$\Delta\alpha$ 达到最大，此时

$$\Delta\alpha=Q\left(-\frac{1}{2\sigma}\right)-Q\left(\frac{1}{2\sigma}\right)=1-2Q\left(\frac{1}{2\sigma}\right) \tag{8}$$

为了具体算出 p_0 和 p_1，我们还需要估计方差 σ^2. 注意到噪声数据 V 服从混合分布(1)，若令 $\bar{a}_2$ 表示样本 V 的二阶原点矩，即 $\bar{a}_2=\frac{1}{N}\sum_{i=1}^{N}v_i^2$，则由文献[8]关于混合高斯分布参数估计的结果知

$$\bar{a}_2=\left(1-\frac{r}{2}\right)(\hat{\sigma}^2+0^2)+\frac{r}{2}(\hat{\sigma}^2+1^2) \quad (9)$$

所以

$$\hat{\sigma}^2=\bar{a}^2-\frac{r}{2} \quad (10)$$

我们用式(10)估计方差 σ^2.

3 碰撞攻击

本节利用上面分析得到的“碰撞优势”Δp，借鉴相关攻击的思想，给出一种恢复隐写密钥的方法，我们称之为“碰撞攻击”.

同上，以密钥 $k\in\mathcal{K}$ 为种子驱动伪随机数发生器 G，生成嵌入路径 $\{j_1,j_2,\cdots,j_L\}$，按此路径对噪声数据 V 的抽样记为 $\boldsymbol{V}_k=\{v_{j_1},v_{j_2},\cdots,v_{j_L}\}$. 进一步，定义随机变量 $Z_i(1\leq i\leq L)$：

$$Z_i=\begin{cases}1, & \text{若 } v_{j_i}>A\\ 0, & \text{若 } v_{j_i}\leq A\end{cases},\quad 1\leq i\leq L \quad (11)$$

构造统计量 $\eta_n=\sum_{i=1}^{n}Z_i,1\leq n\leq L$. 由第 2 节的分析知，对于真密钥 k_0，$P\{Z_i=1\}=p_0$，当 n 充分大时，η_n 近似服从正态分布 $N(np_0,np_0(1-p_0))$；对于伪密钥 k，$P\{Z_i=1\}=p_1$，当 n 充分大时，η_n 近似服从正态分布 $N(np_1,np_1(1-p_1))$，从而隐写密钥恢复问题转化成如下的假设检验问题：

H_0：所选密钥 k 是真密钥 k_0；

H_1：所选密钥 k 是伪密钥.

设定门限值 T，当 $\eta_n\geq T$ 时，接受 H_0；当 $\eta_n<T$ 时接受 H_1.

现在的问题是我们需要多大的样本容量(即 n 选为多大)以及门限值 T 应设为多大时，可以确保按上述规则能得到真密钥. 一般来说，n 越大所作的判决越可靠，但 n 过大会影响攻击的速度. 我们对 n 和 T 的取值作如下分析.

在上述假设检验问题中我们可能犯两类错误：“取伪错误”和“弃真错误”，记取伪错误概率为 P_f，弃真错误概率为 P_m，则

$$p_f=\int_T^{+\infty}\frac{1}{\sqrt{2\pi np_1(1-p_1)}}\exp\left\{-\frac{(x-np_1)}{2np_1(1-p_1)}\right\}\mathrm{d}x \quad (12)$$

$$p_m=\int_{-\infty}^{T}\frac{1}{\sqrt{2\pi np_0(1-p_0)}}\exp\left\{-\frac{(x-np_0)}{2np_0(1-p_0)}\right\}\mathrm{d}x \quad (13)$$

利用式(7)，式(12)和式(13)可转化为

$$p_f=Q\left(\frac{T-np_1}{\sqrt{np_1(1-p_1)}}\right) \quad (14)$$

$$p_m=Q\left(\frac{np_0-T}{\sqrt{np_0(1-p_0)}}\right) \quad (15)$$

此处我们主要关心取伪错误概率 P_f，因为密钥空间的势为 $|\mathcal{K}|$，为了保证能确定唯一的真密钥，需要 $P_f\leq\frac{1}{2^{|\mathcal{K}|}}$，一般取 $P_f=\frac{1}{2^{|\mathcal{K}|}}$ 即可，同时适当限定弃真错误概率 P_m，并查表得 w_f 和 w_m 满足：$\frac{1}{2^{|\mathcal{K}|}}=Q(w_f)$，$P_m=Q(w_m)$. 由此利用式(14)和式(15)就可以估计攻击所需的样本量 n 以及对应的门限值 T 如下：

$$n=\left[\frac{w_m\sqrt{p_0(1-p_0)}+w_f\sqrt{p_1(1-p_1)}}{\Delta p}\right]^2 \quad (16)$$

$$T=w_f\sqrt{np_1(1-p_1)}+np_1 \quad (17)$$

进一步，

$$n\leq\left[\frac{w_m+w_f}{2\Delta p}\right]^2 \quad (18)$$

$$T\leq\frac{1}{2}w_f\sqrt{n}+n \quad (19)$$

下面对攻击所需的样本容量 n 作分析. 由式(18)知样本容量 n 近似正比于 $(w_m+w_f)^2$，因弃真错误概率 P_m 可适当取定(一般取 0.01 或 0.001 即可)，所以 w_m 是确定的，而 w_f 随着密钥量 $|\mathcal{K}|$ 的增大而增大，所以密钥空间越大，攻击所需的数据量越大. 另一方面，n 近似反比于 Δp^2，即碰撞优势越大，攻击所需的数据量越少. 把式(6)代入式(16)得

$$n=\frac{\left(2w_m\sqrt{p_0(1-p_0)}+2w_f\sqrt{p_1(1-p_1)}\right)^2}{[(1-r)\Delta\alpha]^2}\leq\frac{(w_m+w_f)^2}{[(1-r)\Delta\alpha]^2} \quad (20)$$

由式(20)知，样本容量 n 主要受密钥量 $|\mathcal{K}|$，嵌入率 r 和 $\Delta\alpha$ 的影响. $|\mathcal{K}|$ 是假设已知的，r 可以由检测算法比较精确地估计. 对 $\Delta\alpha$ 我们采用了正态分布假设做近似估计，这可能会带来相对大的误差. 注意到 r 越大，$\Delta\alpha$ 的误差对式(20)的影响越显著，所以当 r 较大时，对 n 和 T 的估计误差也会较大. 另外，对于嵌入率 r，欲得到 n 个样本需要的像素点数 n^* 约为 $\frac{n}{r}$，所以由式(20)知

$$n^* = \frac{\left(2w_m\sqrt{p_0(1-p_0)}+2w_f\sqrt{p_1(1-p_1)}\right)^2}{r[(1-r)\Delta\alpha]^2} \tag{21}$$

由式(21)知当 $r\to 0$ 或 $r\to 1$ 时，攻击所需的像素点数 $n^*\to\infty$，即当嵌入率接近 0 或 1 时，图像能提供的样本数不足以区分真伪密钥，所以攻击将难以成功. 下一节的实验结果将验证上述理论分析.

碰撞攻击算法.

有了以上准备，我们下面描述“碰撞攻击”的具体步骤. 假设已获得一幅用空域随机 LSB 算法隐藏了消息的载密图像 $\boldsymbol{S}$，它含有 N 个像素，并假设隐写算法已知，密钥空间为 $\mathcal{K}$. 攻击过程如下：

1. (1)用文献[4]的算法估计消息的嵌入率 r，并计算消息的长度 $L=rN$；

(2)用 2.2 节描述的方法计算噪声集 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_N\}$；

(3)用式(10)估计方差 σ^2，设定 $A=0.5$，然后用式(3)～(5)计算 p_0 和 p_1；

(4)令 $P_f=\frac{1}{2^{|\mathcal{K}|}}$，设定 p_m(如 $p_m=0.01$)，查标准正态分布表得 w_f 和 w_m 满足：$\frac{1}{2^{|\mathcal{K}|}}=Q(w_f)$，$P_m=Q(w_m)$，然后用式(16)和式(17)计算样本容量 n 和门限值 T.

2. 若 $n>L$，转到步 4；否则穷举密钥空间 $\mathcal{K}$ 中的密钥，对每个密钥 k，以 k 为种子利用隐写算法的伪随机数发生器生成前 n 个随机位置 $\boldsymbol{I}(k)=\{j_1,j_2,\cdots,j_n\}$，并对噪声集 $\boldsymbol{V}$ 抽取对应的 n 个元素 $\{v_{j_1},v_{j_2},\cdots,v_{j_n}\}$，然后计算其中大于 0.5 的值的个数 T_k，即 $T_k=|\{v_{j_i}\,|\,v_{j_i}>0.5,1\le i\le n\}|$，若 $T_k\ge T$，则把密钥 k 存入备选集 $\mathcal{B}$；否则抛弃 k.

3. 若 $|\mathcal{B}|=1$，则以 $\mathcal{B}$ 中的密钥为真密钥提取嵌入消息，攻击结束；若 $|\mathcal{B}|=0$，转到步 4；若 $|\mathcal{B}|>1$，令密钥空间 $\mathcal{K}=\mathcal{B}$，转到步 4.

4. 令 $n=L$，穷举密钥空间 $\mathcal{K}$ 中的密钥，对每个密钥 k，按照步 2 中的方法计算 T_k，把使 T_k 达到最大值的密钥 k 存入备选集 $\mathcal{D}$.

5. 若 $|\mathcal{D}|=1$ 用 $\mathcal{D}$ 中的密钥提取消息，攻击结束；若 $|\mathcal{D}|>1$，攻击失败，结束.

4　实验结果

我们用上述算法分析了隐写软件 Hide and Seek. Hide and Seek 是一个典型的空域图像 LSB 隐写算法. 本文针对 Hide and Seek 4.1 来检验“碰撞攻击”的效果. Hide and Seek 4.1 仅以 320×480 的 256 色 GIF 图像为载体，用 Borland C++3.1 中的 *random*(*num*)函数生成随机数，*random*(*num*)以一个 16bit 的种子 k 初态，参数 *num* 用来控制最大偏移步长，它动态变化，与当前所剩消息的长度以及尚未使用的像素的个数有关，这样可使消息近似均匀地散布到整幅图像上. 由于(320×480)/8=19200，所以算法中限定最大消息长度为 19000 字节. 嵌入过程首先用加密算法 IDEA 加密消息长度、随机数发生器种子和版本信息，生成 64bit 密文，并将这 64bit 密文藏于图像的前 64 个像素的 LSB 位，然后从第 65 个像素开始，用随机数发生器生成随机位置嵌入消息. IDEA 的密钥由最长为 8 个字符的口令生成，所以 Hide and Seek 4.1 的密钥长度为 64bit(也即密钥量为 2^{64}). 合法的接收者知道口令，可以解密获得消息长度和种子，从而提取出消息.

欲攻击 64bit 的口令信息是困难的，但是我们可以跳过图像的前 64 个像素，以第 65 个像素为起点进行上节描述的“碰撞攻击”，这样只需恢复出 16bit 的种子和确切的消息长度即可，由于用文献[4]的方法估计消息嵌入率的误差可控制在 ±0.02 以内，所以有大约 $19000\times 0.04=760$ 个可能长度需要测试，由此我们将其有效密钥长度降到了 $16+\log_2 760\approx 25.57$bit.

我们对 20 幅 320×480 的 256 色 GIF 图像在用 Hide and Seek 4.1 嵌入各种长度的消息后，做了攻击实验. 如果已知确切的消息长度 L(此时密钥量为 2^{16})，我们令 $p_f=\frac{1}{2^{16}}$，$p_m=0.01$，实验表明当嵌入率 r 满足 $1.1\%<r<98.4\%$ 时，攻击可以成功，即可以恢复出随机种子. 当嵌入率 r 满足 $11\%<r<50\%$时，估计的样本容量 n 小于允许的最大样本量(即消息长度 L)，攻击过程一般在算法的第 3 步就可结束，攻击速度比直接设置成最大样本量可提高 10%～45% . 这与前面的理论分析相吻合，即当嵌入率 r 接近 0 或 1 时攻击将因数据量不足而失败，当 r 较大时，对 $\Delta\alpha$ 的估计误差会显著影响对 n 的估计，从而使我们得不到可用的 n.

表 1 列出了以 Lena.gif 为载体时，各种嵌入率对应的估计样本容量 n 和门限值 T 以及用某一随机种子 k_0 嵌入消息时的实验结果. 其中 T_{k_0} 表示真密钥对应的大于 0.5 的样本数；“－”表示估计的样本容量 n 大于允许样本容量 L，攻击算法需运行第 4 步；带“*”的 T_{k_0} 小于对应的 T，在算法的第 3 步 $|\mathcal{B}|=0$，此时也需运行第 4 步.

图 1 Lena.gif

表 1 以 Lena.gif 为载体时,对 Hide and Seek 4.1 进行碰撞攻击的实验数据

嵌入率	n	T	T_{k_0}	结果
0.005	—	—	—	失败
0.015	—	—	—	成功
0.053	—	—	—	成功
0.105	14640	6925	7086	成功
0.158	16528	7845	8040	成功
0.263	21592	10319	10466	成功
0.421	34992	16888	16922	成功
0.474	42344	20503	20560	成功
0.526	52280	25398	24086*	成功
0.632	86440	42265	40354*	成功
0.684	—	—	—	成功
0.895	—	—	—	成功
0.950	—	—	—	成功
0.985	—	—	—	成功
0.990	—	—	—	失败

如果消息长度未知,首先要估计消息长度的范围,如前所述,此时的密钥量增大到 $2^{25.57}$,而相对的 Hide and Seek 使用的图像较小,这使得估计的样本容量一般会接近或大于允许的最大样本容量,所以攻击过程总要执行第 4 步.以 10 幅 GIF 图像为载体的实验结果表明,当嵌入率 r 满足 $5.3\%<r<94.7\%$时攻击可以成功.

另外我们将本文的方法与文献[7]中的方法做了比较.文献[7]中方法的出发点是直接利用隐写分析中的卡方检验来区分真伪密钥,而碰撞攻击是将隐写密钥恢复问题转化成一个序列密码分析问题,然后借鉴密码分析方法恢复密钥.若不考虑伪随机数发生器 G 的具体结构,二者都只能局限于穷举区分密钥;若考虑 G 的具体结构,碰撞攻击更适合与针对 G 的密码分析方法相结合设计快速密钥恢复算法.就区分密钥而言,二者都是基于真隐写路径与伪隐写路径上的数据统计差异,所以性能有相似之处.它们的计算复杂度都主要由所需的样本量决定,由于随着嵌入率增大,所需的样本量都迅速增大,所以测试密钥的速度也相应明显下降.我们以 50%的嵌入率对 20 幅 512×512 的灰度图像做了测试,随机嵌入路径由移位寄存器序列生成,密钥长度为 16bit.在配置为 Pentium IV 2.8GHz, 512MB, 3200 DDR RAM 的 PC 机上进行密钥搜索,碰撞攻击的密钥搜索速度为 149 个密钥/s,卡方检验方法的速度为 51 个密钥/s.但应该指出的是碰撞攻击只适用于 LSB 嵌入,而卡方检验方法可应用于 LSB 嵌入和±1 嵌入.

5 结束语

我们在第 2 节假设了空域图像噪声数据近似服从 0 均值正态分布.如果图像的噪声数据的均值偏离 0 较大,会一定程度上影响攻击的速度,因为此时,在攻击算法中设置 $A=0.5$ 并不是最优,并且估计的样本容量 n 和门限值 T 也不精确.一种解决办法是利用类似式(10)估计方差的方法,用多阶矩,列方程组估计两个正态分布的均值,但此时解高次方程也会带来一些误差;另一种方法是剔除异常值,在取噪声数据时,丢掉那些滤波值与像素值差异较大的点,这样可以使获得的数据较好地满足假设,但需注意这会使数据量减少,对于较小的图像(如 Hide and Seek 4.1 使用的图像)不宜采用这种方法.

我们没有考虑针对具体伪随机数发生器的攻击,因为本文的重点不是密码分析而是如何将隐写密钥恢复问题转化成传统的密码分析问题.文中对 LSB 隐写的碰撞攻击只是这种转化的一个初步尝试,而如何利用密码分析方法做快速密钥恢复是有待进一步研究的问题.

参 考 文 献

[1] Chandramouli R. A mathematical framework for active steganalysis. ACM Multimedia Systems Journal, Special Issue on Multimedia Watermarking, 2003, 9(3): 303-311

[2] Fridrich J, Goljan M, Soukal D. Searching for the stego key//Proceedings of the SPIE-Security, Steganography and Watermarking of Multimedia Contents VI. San Jose, CA, 2004: 70-82

[3] Zhang T, Ping X J. A new approach to reliable detection of LSB steganography in natural images. Signal Processing, 2003, 83(10): 2085-2093

[4] Fridrich J, Goljan M. On estimation of secret message length in LSB steganography in spatial domain//Proceedings of the SPIE-Security, Steganography and Watermarking of Multimedia Contents VI. San Jose, CA, 2004: 23-34

[5] Johnson N F, Jajodia S. Steganalysis of images created using current steganography software//Proceedings of the 2nd In-

formation Hiding Workshop. Portland, Oregon, 1998: 273-289

[6] Provos N, Honeyman P. Detecting steganographic content on the Internet. University of Michigan, Michigan: Technical Report CITI 01-1a, 2001

[7] Fridrich J, Goljan M, Soukal D, Holotyak T. Forensic steganalysis: Determining the stego key in spatial domain steganography//Proceedings of the SPIE-Security, Steganography and Watermarking of Multimedia Contents VII. San Jose, CA, 2005: 631-642

[8] Wu Wei-Ren. Estimation of parameters in a mixture of two normal distributions. Journal of Fujian Agricultural College, 1989, 18(2): 236-243(in Chinese)
(吴为人. 两个正态分布的混合分布参数的估计. 福建农学院学报, 1989, 18(2): 236-243)

[9] Fridrich J, Soukal D, Goljan M. Maximum likelihood estimation of secret message length embedded using $\pm K$ steganography in spatial domain//Proceedings of the SPIE-Security, Steganography and Watermarking of Multimedia Contents VII. San Jose, CA, 2005: 595-606

ZHANG Wei-Ming, born in 1976, Ph. D., lecturer. His main research interests include cryptography and steganography.

LI Shi-Qu, born in 1945, professor, Ph. D. supervisor. His main research interests include cryptography and probability theory.

LIU Jiu-Fen, born in 1963, Ph. D., associate professor. Her main research interests include watermarking and steganography.

Background

This work is supported by the National Natural Science Foundation of China under grant No. 60473022. The purpose of this project is to construct theoretic and technical bases for applications of steganalysis, which is important for developing surveillance system of images with hidden information on networks.

Steganography is an important branch of information hiding, and it is about how to send secret message covertly. The attack on steganography, i. e. steganalysis, mainly considers detecting the existence of hidden message. There have been many reliable detecting methods for a variety of stegan ographic algorithms. How to extract the hidden message is a more difficult problem, which we call "extracting attacking", and is also referred as to forensic steganalysis because it is useful for forensic analysis. However, to the best of our knowledge, there are few literatures about extracting attack.

In fact, steganalysis is a systemic work, which consists of several steps: (1) identification of suspicious objects, (2) determining the steganographic method or software in use, (3) searching for the stego key and extracting the embedded bit-stream, (4) deciphering the bit-stream if it is cipher-text. Step 4 is cryptanalysis. The purpose of the authors' project is to solve the problems in Step 1～3, and they have done many work on the Step 1 and Step 2, including some detecting methods on image steganography in spatial and DCT domain and a detecting system on stego software. Step 3 is just the extracting attack which is the most difficult part in the project. The authors have constructed an information theoretic model for extracting attack.

This paper proposes a new solution for Step 3. The main idea is that recovering stego key is a special kind of cryptanalysis and it can be converted into a traditional cryptanalytic problem. Particularly, by combining detecting techniques in steganalysis and correlation attack in cryptanalysis, the authors propose a method to recover the stego key of LSB steganography in spatial domain of images. With this method, the message hidden by the stego software "Hide and Seek 4. 1" is extracted successfully.