2018-2019春季
信息隐藏课程
第6讲 专用隐写分析

中国科学院信息工程研究所
INSITUTE OF INFORMATION ENGNEERING CAS

SKLOIS
信息安全国家重点实验室

赵险峰

中国科学院信息工程研究所
信息安全国家重点实验室

2018年10月

纲要

1. 基本概念
2. 对彩色图像LSBR的分析
3. 对分布恢复隐写OutGuess的分析
4. 对MB隐写的分析
5. 对矩阵编码隐写F5的分析
6. 文献阅读推荐

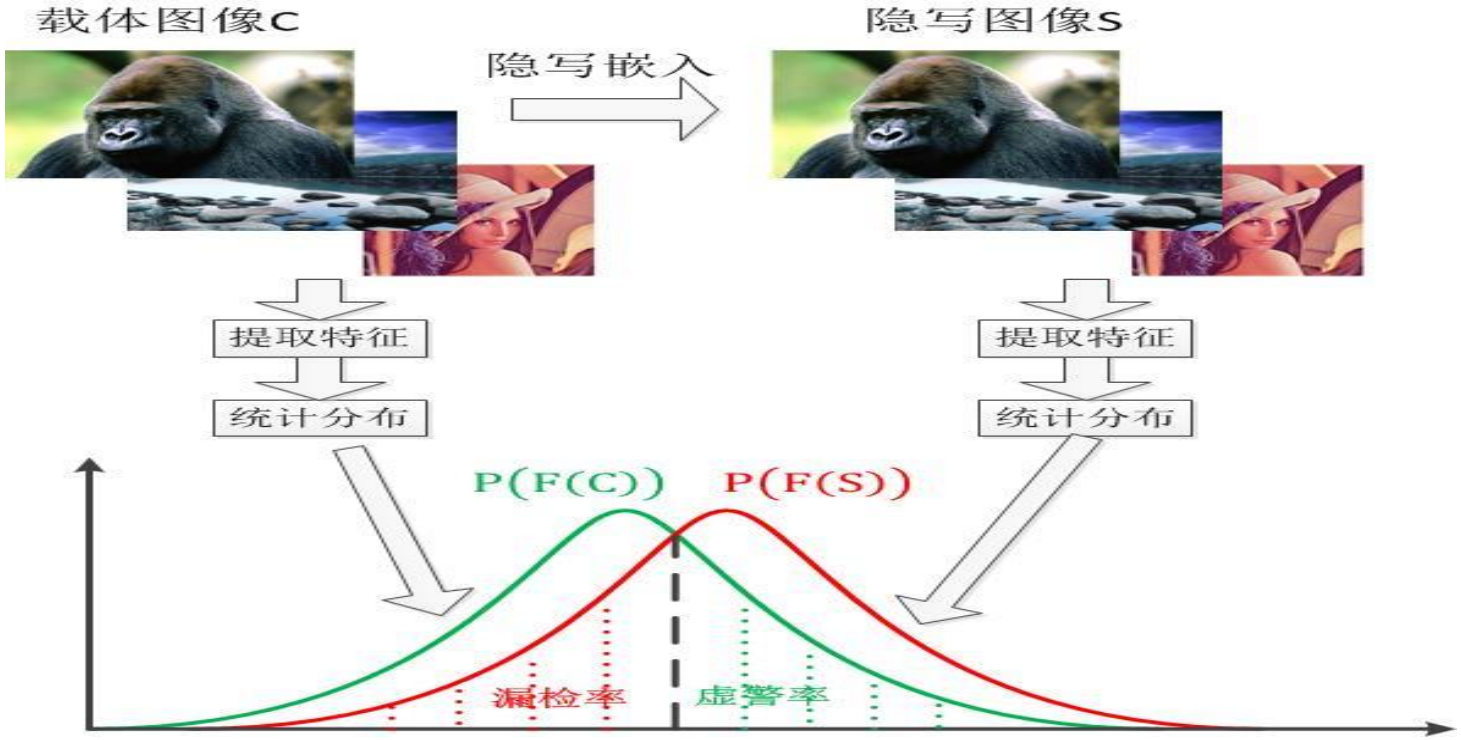
SKLOIS
信息安全国家重点实验室

1-1 上五讲回顾

- 上五讲要点
  - 基本嵌入方法：LSBR、LSBM，基本的概念与性质介绍
  - 基于统计特征恢复的隐写OutGuess、基于模型的隐写MB，F3与F4等基于修改方式的统计保持方法
  - 矩阵编码隐写及其典型代表F5隐写，具有自适应思想的MMEx
  - 有关LSBR（Jsteg）、OutGuess、F5隐写与MME隐写的实验

1-2 专用隐写分析基本概念

- 专用（Specific）隐写分析是指针对某一种或者某一类隐写有效的分析方法。本讲讲授用专用分析攻击之前介绍的隐写
  - χ^2 隐写分析就只对连续嵌入的LSBR有效
- 专用隐写分析原理：在知道隐写算法前提下，通过分析与实验隐写性质，得到专门用于识别该隐写的特征并构造识别方法
  - 一般基于假设检验等统计推断法构造识别方法，需选定阈值



1-3 专用隐写分析 vs 通用隐写分析

- 通用（Universal）隐写分析：对多个或多类隐写分析有效的分析方法；相比较，专用隐写分析在实际使用中存在适用面小的问题，但是，这类分析也具有特殊的价值
- 由于主流通用隐写分析也需将隐写媒体作为训练样本实施监督学习，这等价于知道隐写算法，因此，专有分析与通用隐写分析并没有特别清晰的界限，只不过通用隐写分析方法的**分析特征适用面更宽**
- 专用分析方法的一个优势是，**有一些方法的错误率很低**，能够较为确凿地反映隐写事实，例如，前面介绍的 χ^2 隐写分析方法，在针对连续LSBR的检测中准确率非常高
- 但专用分析比较缺乏统一的设计规律
- 本讲将介绍更多的专用分析方法，**这些方法能够对前面介绍过的主要隐写方法形成有效攻击**

2-1 对彩色图像LSBR的分析（现象观察）

- 光栅格式的空域编码彩色图像中，色彩分量(R,G,B)构成了常用的嵌入域，一些软件采用LSBR方法在其中嵌入秘密信息
- 若是LSBR在彩色图像像素的3个通道上随机选择位置嵌入，χ^2 分析方法就不能有效检测
- J. Fridrich等人发现，彩色图像被LSBR嵌入后，由于存在大量加减1操作，接近色彩对（Close Color Pairs）的数量 P 将有增长，其中，接近色彩对的定义是：如果(R_i,G_i,B_i)与(R_j,G_j,B_j)是接近色彩对，则满足

$$|R_i-R_j|\le 1,\qquad |G_i-G_j|\le 1,\qquad |B_i-B_j|\le 1$$

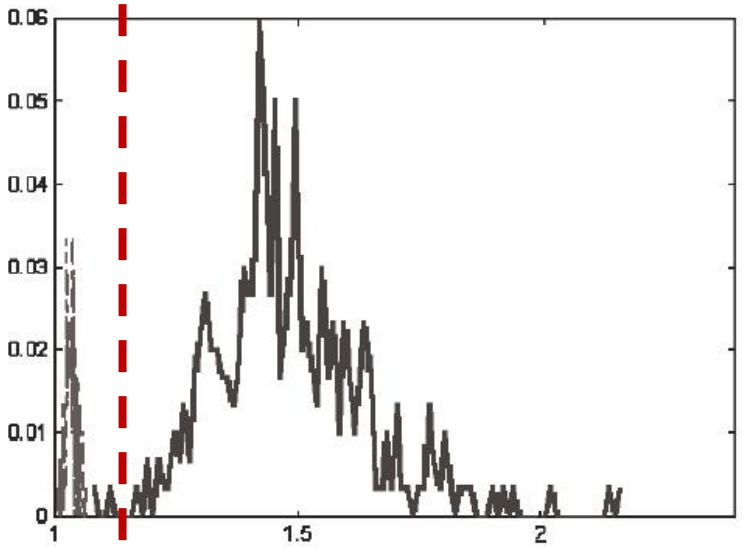
或等价的：$(R_i-R_j)^2+(G_i-G_j)^2+(B_i-B_j)^2\le 3$

- 设 U 为图像颜色数，则$R=P/\binom{U}{2}$为接近色彩对数量与全部可能颜色对之间的比例，以上现象使该比例在存在隐写的情况下有提高

2-2 对彩色图像LSBR的分析（算法）

- 基于以上观察，Fridrich等人提出了RQP（Raw Quick Pair）分析。它没用 R 为检验统计量，而对图像主动隐写，得到 R'，一个显著事实是，如图像已隐写，则 R'变化较小，否则较大，因此采用 R'/R 为检验统计量。设 T 为假设检验阈值，以下是分析步骤：
  - 对待检的 $M \times N$ 图像，计算 R
  - 对待检图像的拷贝，其中有 $3MN$ 个像素分量，在其中随机选择 α 比例的位置，用LSBR方法嵌入 $3\alpha MN$ 个比特（下称 α 为测试性负载率）
  - 重新计算 R'，得到统计量R'/R
  - 如果 $R'/R < T$，推断待检测图像存在隐写，否则没有

T=1.0736



2-3 对彩色图像LSBR的分析（性能）

- R'/R 在隐写与未隐写两情况下分布（按高斯分布）（s 为原负载率）：

$$f_{\mu(s),\sigma(s)}(x)=\frac{1}{\sqrt{2\pi\sigma^2(s)}}e^{-\frac{(x-\mu(s))^2}{2\sigma^2(s)}},\quad f_{\mu,\sigma}(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- 为求阈值 T，令I型错误（虚警率）= II型错误（漏检率）：

$$P(\mathrm{I})=\int_{-\infty}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\int_{T}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2(s)}}e^{-\frac{(x-\mu(s))^2}{2\sigma^2(s)}}dx=P(\mathrm{II})$$

- 代换：$w=(x-\mu)/\sigma$，$w'=(x-\mu(s))/\sigma(s)$，$w,w'\sim N(0,1)$，则有

$$\int_{-\infty}^{(T-\mu)/\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{w^2}{2}}dw=\int_{(T-\mu(s))/\sigma(s)}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{w\prime^2}{2}}dw'$$

  - 有 $\frac{T-\mu(s)}{\sigma(s)}=-\frac{T-\mu}{\sigma}$，即 $T=\frac{\mu\sigma(s)+\sigma\mu(s)}{\sigma+\sigma(s)}$

- 实验结果显示，在 $s=0.2$、$\alpha=0.2$ 与 $T=1.0736$ 时，RQP分析方法的错误率仅有0.82%
- **针对LSBR的检测方法还有RS、SPA，请参阅讲义**

3-1 对OutGuess的分析（现象观察）

- 基于提取一阶分布特征的分析方法都难以奏效
- J. Fridrich等人发现，OutGuess在相同负载率下需要修正直方图，因此**嵌入次数较多，造成了分块之间的相关性下降**。考虑分块之间的相关性实际上已经等价于考虑图像的**二阶分布，这已经不是OutGuess隐写的防护范围了**。对 $M \times N$ 图像空间域像素 $g_{i,j}$，可以定义块效应（Blockiness）指标为

$$B = \sum_{i=1}^{\left\lfloor \frac{M-1}{8} \right\rfloor} \sum_{j=1}^{N} \left| g_{8i,j} - g_{8i+1,j} \right| + \sum_{j=1}^{\left\lfloor \frac{N-1}{8} \right\rfloor} \sum_{i=1}^{M} \left| g_{i,8j} - g_{i,8j+1} \right|$$

- J. Fridrich等人还发现，B 与嵌入消息引发的修改次数呈线性关系，据此可以估算消息的长度，他们根据以上观察提出了相应的隐写分析方法，该方法输出的是，嵌入消息长度占可嵌入最大消息长度的比例 p，**是OutGuess负载率的一种表示**

3-2 对OutGuess的分析（1次隐写修改次数）

- 为说明消息长度估算方法，需推导修改次数的表达式。设 h_i 为JPEG量化DCT系数的直方图，P 为可修改的系数数量，由于OutGuess不使用值为0和1的系数，因此有$P = \sum_{\substack{i\neq 0,\\ i\neq 1}} h_i$
- 设值对的直方图为(h_{2i}, h_{2i+1})，$h_{2i} > h_{2i+1}$，统计上，嵌入消息后发生 $h_{2i} \rightarrow h_{2i} - \alpha(h_{2i} - h_{2i+1})$，$h_{2i+1} \rightarrow h_{2i+1} + \alpha(h_{2i} - h_{2i+1})$
- $\alpha = 0.5m/P$ 为修改率，m 为消息长，$m/P = 2\alpha$为负载率。OutGuess保证在值为 $2i+1$ 的系数中预留足够的位置恢复 h_{2i}，有 $(1-2\alpha)\, h_{2i+1} \geq \alpha(h_{2i} - h_{2i+1})$，即$\alpha_i \leq \frac{h_{2i+1}}{h_{2i} + h_{2i+1}}$
- 因此总修改率应满足 $\alpha = \min_i \alpha_i$，使有效消息长为 $2\alpha P$。实际消息长 $2p\alpha P$ 是这个长度的一个比例 $0 \leq p \leq 1$（注：随机跳跃），令 $\bar{h}_{2i} = \max(h_{2i}, h_{2i+1})$，$\underline{h}_{2i} = \underline{h}_{2i+1} = \min(h_{2i}, h_{2i+1})$，考虑直方图修正后，值对的两个值上修改次数均为 $p\alpha\bar{h}_{2i}$，因此各个值对上的修改总次数为（以下右式第1、2项是为嵌入消息、修正直方图分别做出的修改次数）：

$$T_p = 2p\alpha \sum_{i\neq 0} \bar{h}_{2i} = p\alpha P + p\alpha \sum_{i\neq 0} \left|\bar{h}_{2i} - \underline{h}_{2i}\right|$$

3-3 对OutGuess的分析（2次修改嵌入次数）

- 在J. Fridrich等人提出的分析方法中，也使用了以上RQP方法中的**再次嵌入密文的检测方法**，因此，这里推导两次嵌入密文后的修改次数
- 设有一个包含 n 个整数的集合，如果随机选择集合中由 s 个整数组成的子集 S，对子集中的整数进行LSB修改操作，并接着随机选择集合中由 r 个整数组成的子集 R，对其中的整数进行翻转，则显然最后实际发生修改的次数为 $s - s \times \frac{r}{n} + r - r \times \frac{s}{n} = s + r - \frac{2sr}{n}$
- 对一幅已经由OutGuess嵌入了 $2p\alpha P$ 长度消息的图像，如果再次嵌入 $2q\alpha P$ 长度的消息，$0 \le q \le 1$，则在数值 $2i$ 与 $2i+1$ 上发生的修改次数分别为：$p\alpha\,\bar{h}_{2i} + q\alpha\,\bar{h}_{2i} - \frac{2pq\alpha^2\bar{h}_{2i}^2}{\bar{h}_{2i}} = \alpha\,\bar{h}_{2i}\,(p+q-2pq\alpha)$，

  $p\alpha\,\bar{h}_{2i} + q\alpha\,\bar{h}_{2i} - \frac{2pq\alpha^2\bar{h}_{2i}^2}{\underline{h}_{2i+1}} = \alpha\,\bar{h}_{2i}\,(p+q-2pq\alpha\,\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}})$
- 两次用OutGuess嵌入后，修改次数总和是全部值对上以上两个次数的和（$q=0$ 为1次）：$T_{pq} = 2\alpha\sum_{i\neq 0}\bar{h}_{2i}\left(p+q-\alpha pq\left(1+\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}}\right)\right)$

3-4 对OutGuess的分析（参数化与多种块效应计算）

- 实验表明，块效应指标 B 与嵌入消息引发的修改次数呈线性关系，它们都与消息的长度有关，这可以描述为$B(p) = c + dT_p$
- 前面定义的 p 是表征消息长度的参数，因此，以下推导通过各种情况下的块效应指标估计 p 的方法。设在以下情况下计算块效应指标：
  - 对待检测的JPEG图像，解码至空间域并计算块效应指标$B_{s(0)}$
  - 对待检测的JPEG图像，用OutGuess嵌入最大长度 $2\alpha P$ 比特的消息，解码至解码至空间域并计算块效应指标$B_{s(1)}$
  - 对待检测的JPEG图像，在空间域裁剪掉4行与4列，重新按照相同的质量因子压缩，再解码到空间域并计算块效应指标$B_{(0)}$。需注意，裁剪图像的分块从原来分块的中点开始，较大程度消除了原来的分块效应，可以将其看作是原始图像的近似，这类技术被称为校准（Calibration）
  - 对以上裁剪的图像，用OutGuess嵌入最大长度的消息，解码至解码至空间域并计算块效应指标$B_{(1)}$，用同样的质量因子压缩得到JPEG图像
  - 对上一步得到的嵌入后JPEG图像，用OutGuess再次嵌入最大长度的消息，解码至空间域并计算块效应指标$B_{1(1)}$

3-5 对OutGuess的分析（多种块效应比较）

- **如果图像本来就嵌入了秘密消息，则再次嵌入后，块效应指标的变化相比图像是自然图像的情况相对轻微。**这种差别可以通过 $S = B_{s(1)} - B_{s(0)}$，$S_0 = B_{(1)} - B_{(0)}$，$S_1 = B_{1(1)} - B_{(1)}$表示，有

$$S_1 = B_{1(1)} - B_{(1)} = d(T_{11} - T_{10}) = 2\alpha d\left(\sum_{i\neq 0}\bar{h}_{2i}\left(2-\alpha\left(1+\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}}\right)\right)-\sum_{i\neq 0}\bar{h}_{2i}\right)$$

$$= 2\alpha d\sum_{i\neq 0}\bar{h}_{2i}\left(1-\alpha\left(1+\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}}\right)\right)$$

$$S_0 = B_{(1)} - B_{(0)} = d(T_{10} - T_{00}) = 2\alpha d\left(\sum_{i\neq 0}\bar{h}_{2i}\,(1+0)-0\right) = 2\alpha d\sum_{i\neq 0}\bar{h}_{2i}$$

$$S = B_{s(1)} - B_{s(0)} = d(T_{p1} - T_{p0}) = 2\alpha d\left(\sum_{i\neq 0}\bar{h}_{2i}\left(p+1-\alpha p\left(1+\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}}\right)\right)-\sum_{i\neq 0}\bar{h}_{2i}\,(p+0)\right)$$

$$= 2\alpha d\left(\sum_{i\neq 0}\bar{h}_{2i}\left(1-\alpha p\left(1+\frac{\bar{h}_{2i}}{\underline{h}_{2i+1}}\right)\right)\right)$$

得到： $p = \frac{S_0 - S}{S_0 - S_1}$

在J. Fridrich等人的实验中，在估计了 p 之后，表达为相对修改次数（Relative Number of Changes/Modification） $T_p/\alpha P$ 的形式

αP为OutGuess可能的最大修改次数

3-6 对OutGuess的分析（二次压缩的特殊处理）

- 若OutGuess软件的输入不是一个位图而是一个质量因子为Q_c的JPEG图像，则其会把它解码后，以固定的质量因子Q_s重新压缩后进行隐写，**由于载体图像经过了两次JPEG压缩，分布独特，以上校准图像不能逼近块效应性质**，尤其在 $Q_c < Q_s$，会造成较大的错误率
- 为了在一定程度上解决这个问题，可以估计出Q_c，在以上裁剪校准后，将图像先用假设的 Q_c压缩，接着解码后用 Q_s压缩，之后解压缩并计算 $B_{(0)}$，这样可认为 $B_{(0)}$ 是在近似的两次压缩载体图像上计算的
- 估计 Q_c：设$h_d(i,j)$ 为待检测JPEG图像中分块DCT系数频率分量为(i,j)且数值为 d 的直方图值，被裁剪校准后用质量因子 Q 压缩，解压并继而用Q_s压缩（模拟嵌入处处理理处理处理），记以上直方图值变为 $h_d(i,j,Q)$。由于裁剪校准后的图像在空间域逼近原载体图像，则可以认为若$Q = Q_c$，$h_d(i,j,Q)$ 与 $h_d(i,j)$ 之间更加接近，因此

$$Q_c = \arg\min_{Q} \sum_{(i,j)} \sum_{d} |h_d(i,j) - h_d(i,j,Q)|^2$$

(i,j) 仅仅限于$(1,2),(2,1),(2,2)$

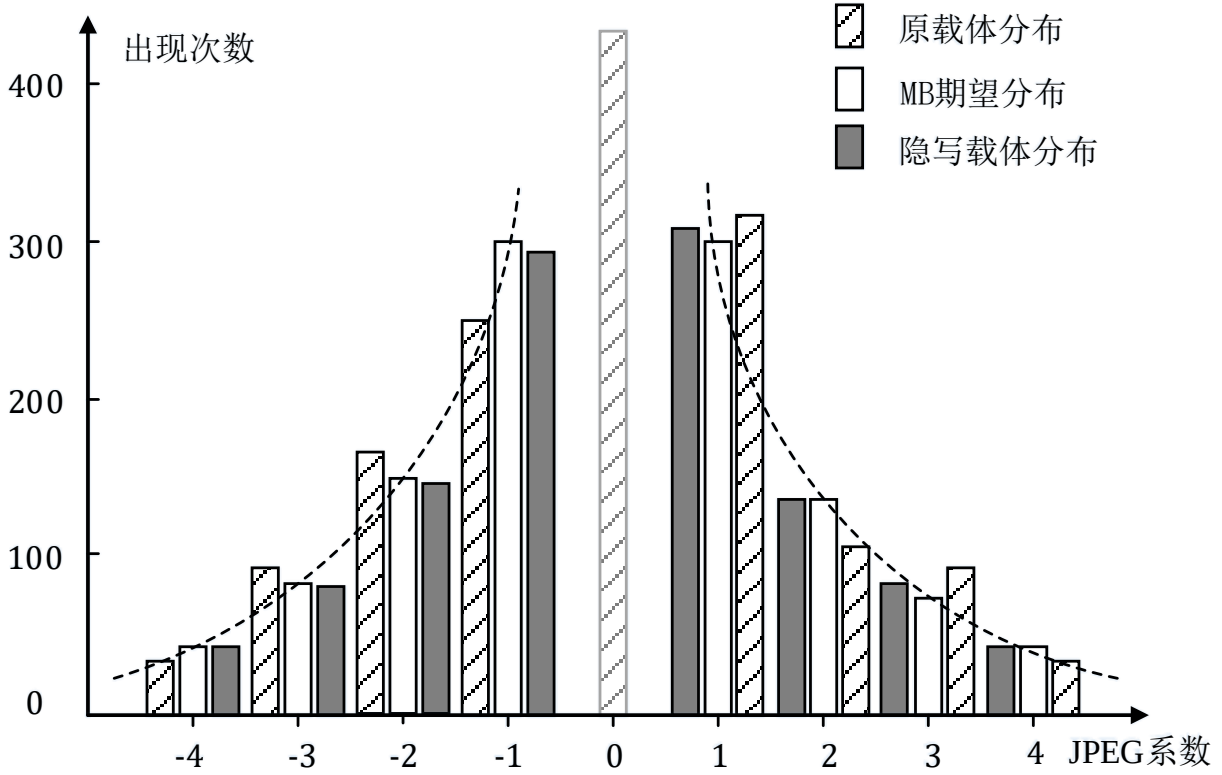
4-1 对MB隐写的分析（现象观察）

- 以JPEG量化系数为嵌入域的MB隐写保持了载体一阶统计分布的大致轮廓，这个轮廓由载体的不修改部分决定。可通过参数估计拟合Cauchy曲线确定这个轮廓，作为隐写消息收发操作的基本参数
- R. Böhme与A. Westfeld发现，隐写分析者也可以使用这个估计的分布轮廓，如果分析方法发现JPEG量化系数的分布过于接近这个轮廓，就可以判定这个图像被MB隐写处理过的
- 以上发现反映了MB隐写在基本设计原则上的失误，对正确建立隐写的设计原则有重要意义。显然，自然的统计分布一般不会平滑地接近一个拟合后的分布，如果隐写算法将统计分布曲线调整到一个人工估计的曲线上，有显著的人工处理痕迹

4-2 对MB隐写的分析（现象观察——图示）

JPEG图像量化系数的在MB隐写前后的对比：

过度吻合现象



4-3 对MB隐写的分析（分布曲线拟合回顾）

- 设基于图像系数不修改部分X_α得到的“粗糙”直方图记为$b_k^{(i,j)}$，其中，$(i,j),0<i,j\le 8$表示频率分量，被用于嵌入的系数不包括DC系数，因此一共63个分量，需要估计63个直方图；设$h_{2k}^{(i,j)}$表示图像(i,j)分量分块DCT系数上的数值数量（直方图值），因此根据值对封闭对流的事实有：$$b_k^{(i,j)}=\begin{cases}\frac{h_{2k+1}^{(i,j)}+h_{2k}^{(i,j)}}{2}, & k<0\\ h_0^{(i,j)}, & k=0\\ \frac{h_{2k-1}^{(i,j)}+h_{2k}^{(i,j)}}{2}, & k>0\end{cases}$$
- 由于0值系数不用，因此$h_0^{(i,j)}$单独统计。假设分布服从广义Cauchy分布，根据$b_k^{(i,j)}$可以估计分布参数π,s，最后得到广义Cauchy分布曲线$f(x,\pi,s)$。设$p_k^{(i,j)}=P(X_\beta=1|X_\alpha=2k)=\frac{f(2k-1,\pi,s)}{f(2k-1,\pi,s)+f(2k,\pi,s)}$
- 则对应$X_\alpha=2k$的(i,j)上频率分量系数，按概率$p_k^{(i,j)}$嵌入1，按概率$1-p_k^{(i,j)}$嵌入0，由Huffman解码器“解码”密文消息完成

4-4 对MB隐写的分析（分析算法思路）

- MB隐写图像的分布过于拟合以上广义Cauchy分布曲线。如果存在度量这种拟合程度的方法，实际就存在一种检测MB隐写的方法
  - 设 $\hat{h}_{2k}^{(i,j)}$ 为表示隐写图像(i,j)分量分块DCT系数上的数值$\hat{h}_{2k}^{(i,j)}$数量，则可认为$\hat{h}_{2k}^{(i,j)}$与$\hat{h}_{2k-1}^{(i,j)}$服从以下$2b_k^{(i,j)}$ 重Bernoulli分布：
    $\hat{h}_{2k-1}^{(i,j)} \sim b\left(2b_k^{(i,j)}, p_k^{(i,j)}\right)$，$\hat{h}_{2k}^{(i,j)} \sim b\left(2b_k^{(i,j)}, 1-p_k^{(i,j)}\right)$
  - 它们的期望值分别是：$\bar{h}_{2k-1}^{(i,j)} = 2b_k^{(i,j)} p_k^{(i,j)}$，$\bar{h}_{2k}^{(i,j)} = 2b_k^{(i,j)}\left(1-p_k^{(i,j)}\right)$
  - 对一个图像，可以观察$h_{2k}^{(i,j)}$与$h_{2k-1}^{(i,j)}$的数值，推断其是否与$\bar{h}_{2k}^{(i,j)}$与$\bar{h}_{2k-1}^{(i,j)}$较为符合，如果符合则认为存在MB隐写。

	嵌入1的频次	嵌入0的频次	观测总数
观测频次	$h_{2k-1}^{(i,j)}$	$h_{2k}^{(i,j)}$	$2b_k^{(i,j)}$
期望频次	$\bar{h}_{2k-1}^{(i,j)}$	$\bar{h}_{2k}^{(i,j)}$	$2b_k^{(i,j)}$

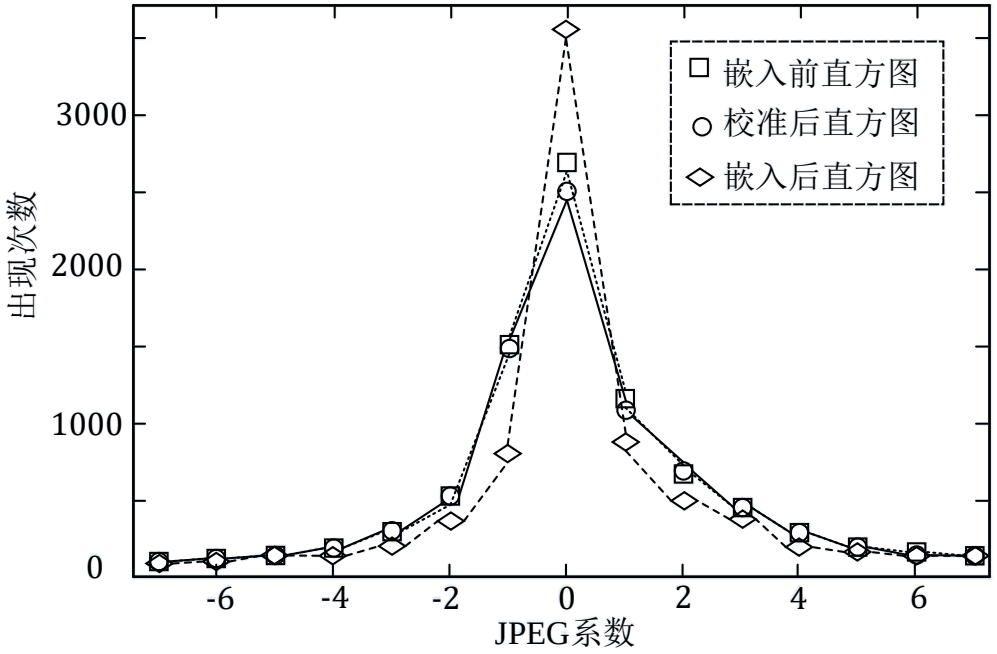
4-5 对MB隐写的分析（χ^2 检测法的再应用）

- 从前图看，当$2k-1=1$与$2k+1=-1$时，以上期望值的与观测值的一致性差别在隐写与未隐写图像之间达到最大。因此，R. Böhme与A. Westfeld仅基于值为±1的两个系数比较，这样，在63个频率分量下每个分量有2个比较序列，因此每个图像首先做$63\times2=126$次比较，评估观测与期望频次间的差异，采用差异比较方法是自由度为1的Pearsons χ^2 检测法。即每次计算 $\chi^2=\frac{\left(h_u^{(i,j)}-\bar{h}_u^{(i,j)}\right)^2}{\bar{h}_u^{(i,j)}}$，$u=-1,1$
- 设置一阈值，当以上计算值大于阈值则认为不符合拟合曲线，小于阈值则认为符合拟合曲线（即隐写）。这个阈值作为分位点对应一个χ^2分布积分（从阈值到无穷）概率$p_{\lim}$，隐写后曲线左移，当积分值$p<p_{\lim}$时可以认为符合拟合曲线（即隐写）。
- 在126次中分布符合的次数可以作为最终隐写判定的依据（决策融合），正常图像一般这个值很小，只有2或者3，也说明这个方法检测效果很稳定

5-1 对F5的分析（基本思路）

- 前面，通过裁掉4行4列对图像校准，可以得到一个统计特性近似原载体的图像，基于与这个图像特征的比较分析，可以估计OutGuess嵌入的消息长度
- J. Fridrich等人类似地将这个方法用于分析F5隐写。在校准中，不但仍然采用在空间域裁剪4行4列的方法，紧接着也采用了一个3 × 3的滤波器进行平滑操作。如果F5软件输入的是光栅格式图像，校准后图像的直方图非常逼近原载体图像的直方图

隐写图像、校准图像与原载体图像直方图对比（系数（2,1））



5-2 对F5的分析（F5两个性质回顾）

- F5采用$(b=2^k-1,k)$的海明矩阵码，负载率$R(k)=k/b=k/(2^k-1)$，平均每比特修改次数为$D(k)=\frac{1}{b+1}\cdot\frac{0}{b}+\frac{b}{b+1}\cdot\frac{1}{b}=\frac{1}{b+1}=\frac{1}{2^k}$，因此，嵌入效率为$W(k)=\frac{R(k)}{D(k)}=\frac{2^k}{2^k-1}k$
- 若需要修改系数，F5采用绝对值-1的方法
- C是F5算法估计的有效非零AC系数容量，估算方法是：$C=h_{DCT}-\frac{h_{DCT}}{64}-h(0)-0.51h(1)$，其中，$h_{DCT}$表示量化DCT系数的总数，$h_{DCT}/64$是DC系数的数量，$0.51h(1)$是考虑因为收缩嵌入失败的情况；由于矩阵编码负载率越低嵌入效率越高，F5通过估计以上容量，结合需要嵌入的消息长度，确定合适的编码参数k

5-3 对F5的分析（估计系数被修改的可能 β ）

- 令 β 表示一个非零交流系数被修改的可能，J. Fridrich等人发现可以用最小二乘法得到它的一个估计量
- 设 $h_{kl}(d)$ 是原载体分块DCT变换频率分量(k,l) 上绝对值为 d 的直方图值，$1 \le k,l \le 8$，令$\hat{h}_{kl}(d)$ 为校准图像的相应直方图值，$H_{kl}(d)$ 为F5隐写后图像的相应直方图值，n 为修改总次数，$P = h(1) + h(2) + \cdots$ 为非零AC系数总数，有 $\beta = n/P$ 。由于F5通过置乱位置次序进行嵌入，实际上是随机选择嵌入系数（修改方式是减小绝对值），这样有 $H_{kl}(d) = (1-\beta)h_{kl}(d) + \beta h_{kl}(d+1), d > 0$，$H_{kl}(0) = h_{kl}(0) + \beta h_{kl}(1)$
- 在上两式右侧，第一项表示未修改的 d 值位置，第二项是临近值修改后流入的。由于一般JPEG分块DCT量化系数在0与1上分布受隐写的扰动最大，可定义以下线性最小二乘问题（减数是对被减数的逼近）：

$$\beta_{kl} = \arg\min_{\beta} \left[H_{kl}(0) - \hat{h}_{kl}(0) - \beta\hat{h}_{kl}(1)\right]^2 + \left[H_{kl}(1) - (1-\beta)\hat{h}_{kl}(1) - \beta\hat{h}_{kl}(2)\right]^2$$

- 求导令为0，求解：$\beta_{kl} = \dfrac{\hat{h}_{kl}(1)\left[H_{kl}(0)-\hat{h}_{kl}(0)\right]+\left[H_{kl}(1)-\hat{h}_{kl}(1)\right]\left[\hat{h}_{kl}(2)-\hat{h}_{kl}(1)\right]}{\hat{h}_{kl}^2(1)+\left[\hat{h}_{kl}(2)-\hat{h}_{kl}(1)\right]^2}$
- 最后，β 的估计值取$\beta_{12},\beta_{21},\beta_{22}$的平均值；尖角表校准

5-4 对F5的分析（估计嵌入密文消息的长度 M）

- 将修改总次数 n 分解为 $n = s + m$，其中，s 为收缩的次数，即修改后没有形成有效嵌入的次数，m 表示嵌入消息的修改次数，由于每次嵌入产生收缩的可能 $P_S = h(1)/P$，因此有 $m + nP_S = n$
- 即得到 $m = n(1 - P_S)$。则密文消息的长度可以估计为：

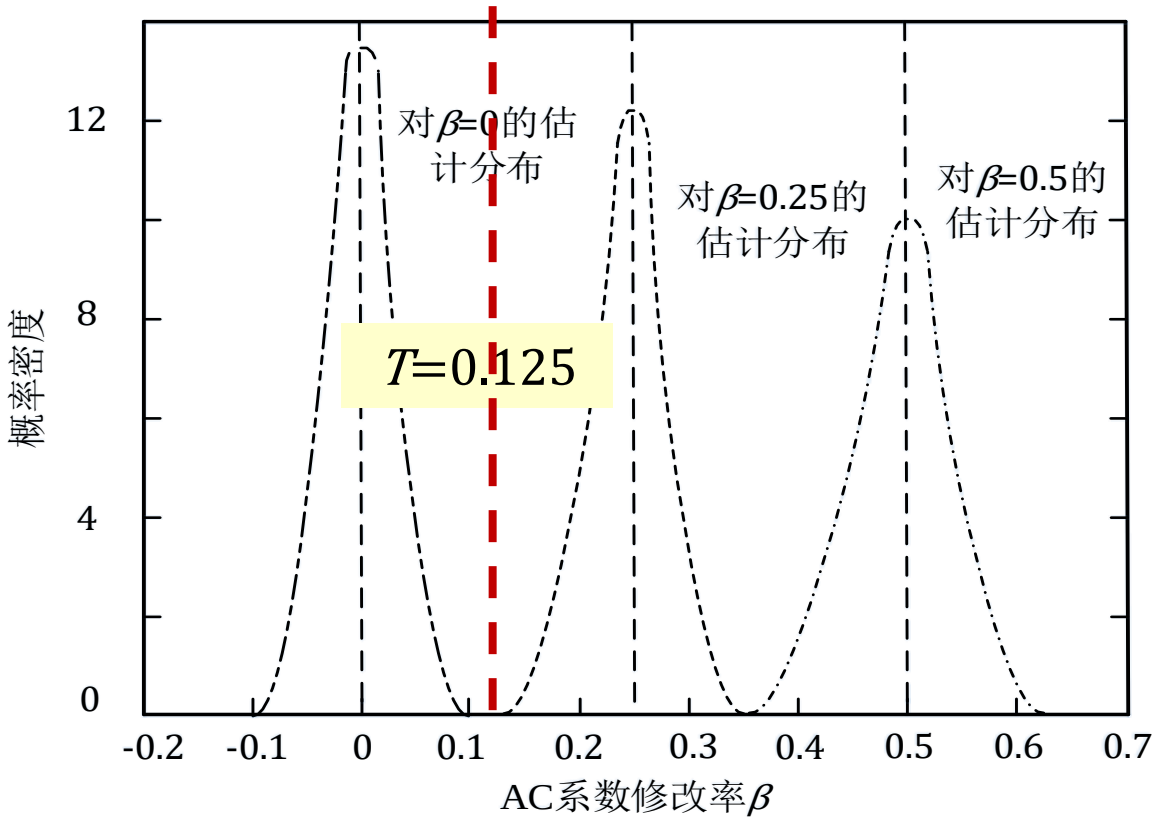
$$M = W(k)m = \frac{2^k}{2^k - 1}kn(1 - P_S) = \frac{2^k}{2^k - 1}k\beta P\left(1 - \frac{h(1)}{P}\right)$$

$$= \frac{2^k}{2^k - 1}k\beta(P - h(1))$$

- 其中，$h(1)$ 按照 $\hat{h}(1)$ 估算，β 前已估计；P 可参照校准图直方图估计：$P = \sum_{d>0} h(d) \approx \sum_{d>0} \sum_{\substack{k,l=1 \\ k+l>2}}^{8} \hat{h}_{kl}(d)$
- 其中，$k + l > 2$ 是要求不取直流系数；另外，由于进一步可得到 $n = \beta P$ 与 $m = n(1 - P_S)$，$P_S = \hat{h}(1)/P$，通过 $D(k) = m/\hat{C} = 1/2^k$，可确定海明矩阵编码参数 k，其中，$\hat{C}$ 是F5算法估计的不产生收缩的非零AC系数容量 $\hat{C} = \hat{h}_{DCT} - \frac{\hat{h}_{DCT}}{64} - \hat{h}(0) - 0.51\hat{h}(1)$，**最后得到** M

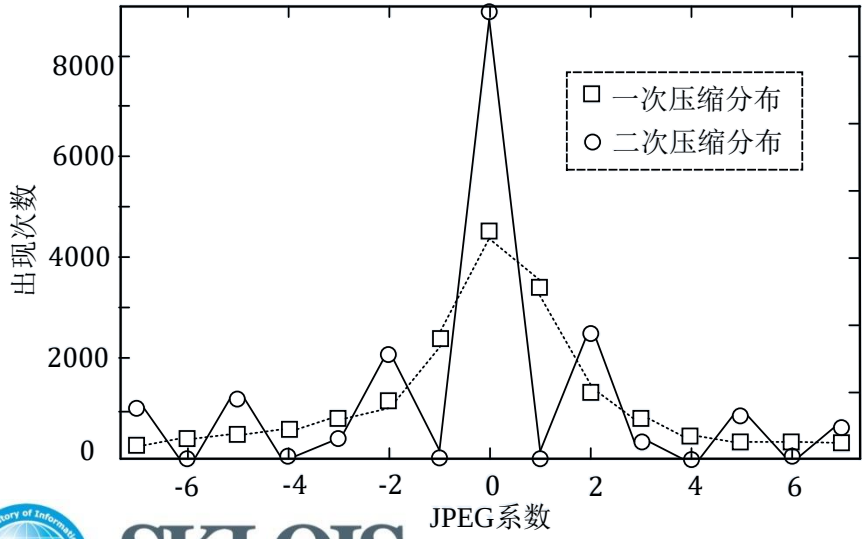
5-5 对F5的分析（隐写分析基本性能）

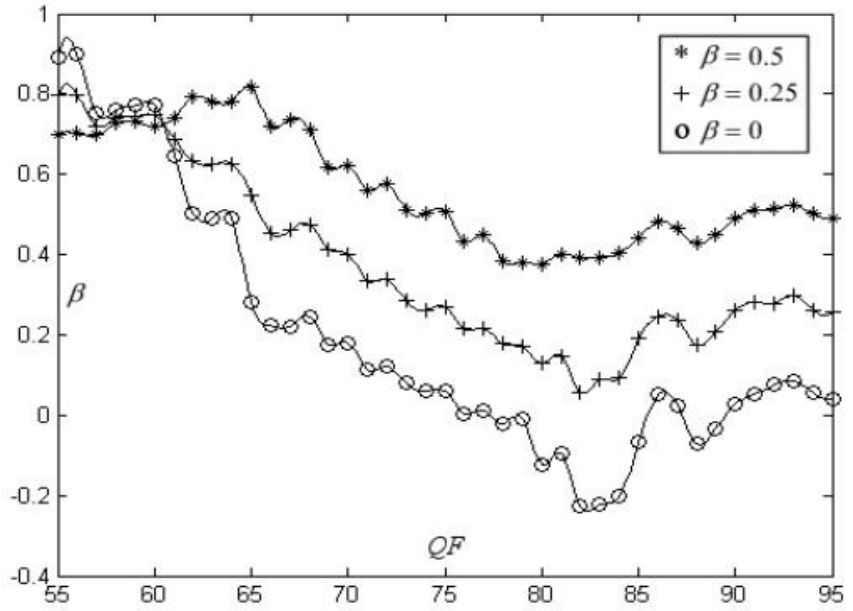
- 在没有双压缩（即F5的输入为光栅图像）的样本上，对消息长度的估计非常有效。下图给出了在系数修改率 $\beta = 0,0.25,0.50$情况下对它的估计情况，估计结果呈正态分布，在取阈值为0.125时，在 $\beta = 0.25,0.50$时，虚警率仅仅为10^{-8}，漏检率分别为10^{-7}与10^{-32}



5-6 对F5的分析（二次压缩干扰）

- 当将JPEG图像输入F5软件时，软件会将图像解压，并按照用户设定的质量因子再次压缩，之后嵌入消息。这样，消息实际是在二次压缩以后的图像中嵌入的，因此，校准图像的直方图与二次压缩原载体的直方图有很大差别（下图左）
- 在输入图像质量因子从55-95，F5二次压缩质量因子75，在 $\beta = 0,0.25,0.50$ 下对其进行估计的情况说明（下图右），如果2次质量因子接近，估计效果较好，否则很差





SKLOIS
信息安全国家重点实验室

5-7 对F5的分析（二次压缩干扰处理）

- 为降低双压缩造成的影响，J. Fridrich等人进行了特殊处理（类似于前面针对OutGuess的攻击）：
  - Q_s为第二次压缩的量化表，再准备一组量化表$\{Q_1, Q_2, \cdots, Q_r\}$，对应相应的质量因子；在运行前述分析方法时，仅仅做出以下改动：对裁剪与滤波后的图像（看做原载体的空间域），采用尝试性的最原始图的Q_i压缩，继而解压并用对应原待检图的Q_s压缩得到校准版本（模仿嵌入过程），此时再用以上方法估计载体直方图与$\beta_i, i = 1, \cdots, r$，在每次估计中，对频率分量$(k, l)$计算

$$
\begin{aligned}
&E_{kl}^{(i)} \\
&= \left[H_{kl}(0) - \hat{h}_{kl}(0) - \beta_i \hat{h}_{kl}(1)\right]^2 \\
&+ \sum\nolimits_j \left[H_{kl}(j) - (1-\beta_i)\hat{h}_{kl}(j) - \beta_i \hat{h}_{kl}(j+1)\right]^2
\end{aligned}
$$

  - 如果，Q_i是第一次的量化表，则以上$\hat{h}_{kl}$的逼近效果较好，$E_{kl}^{(i)}$较小。最后，$\beta = \beta_t$，$t = \arg\min_i \sum_{kl} E_{kl}^{(i)}$

以上方法假设：若过程猜准了，逼近效果更好

6 文献阅读推荐

- [1] 教材第5章
- [2] J. Fridrich, R. Du, and M. Long. Steganalysis of LSB encoding in color images. In Proc. ICME 2000, New York City, New York, USA, July 31 - August 2, vol.3: 1279-1282, 2000
- [3] J. Fridrich, M. Goljan, and D. Hogea. Attacking the OutGuess. In Proc. MM & Sec'02, Juan-Les-Pins, France, Dec.6, 2002.
- [4] R. Bohme and A. Westfeld. Breaking Cauchy model-based JPEG steganography with first order statistics. In Proc. ESORICS'04, LNCS 3193: 125–140, Springer-Verlag, 2004.
- [5] J. Fridrich, M. Goljan, and D. Hogea. Steganalysis of JPEG image: Breaking the F5 algorithm. In Proc. IH'02, LNCS 2578: 310-323, Springer-Verlag, 2002
- [6] 自学教材第5章中的RS、SPA与奇异颜色分析方法

谢谢！

中国科学院信息工程研究所
INSITUTE OF INFORMATION ENGNEERING CAS

SKLOIS
信息安全国家重点实验室