2018-2019春季
信息隐藏课程
第12讲 自适应隐写

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

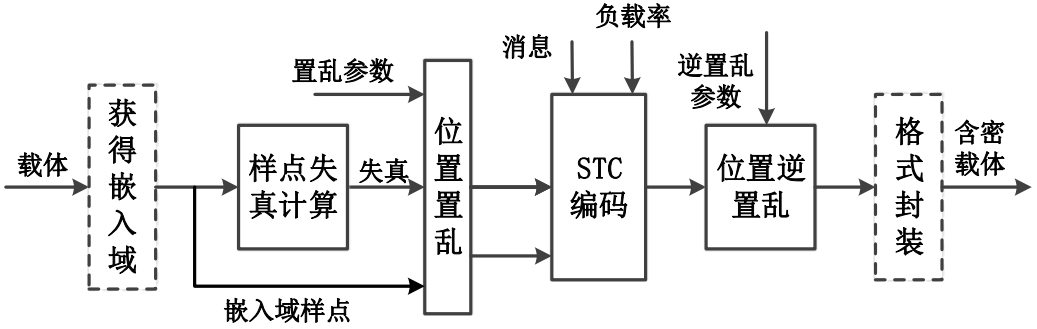
2018年12月

纲要

1. 基本设计方法
2. 空间域自适应隐写
    1. HUGO
    2. WOW
    3. S-UNIWARD
3. JPEG域自适应隐写
    1. J-UNIWARD
    2. UED
4. 文献阅读推荐

1-1 基本设计方法（概念）

- 自适应隐写的主要目的是在负载率一定条件下最小化总失真
- 一般来说，由于隐写修改样点之间的相互干扰，总失真只能基于一定的假设估算，前面提到的加性模型是最常用的总失真估算模型，STC编码是实现该模型下总失真最小的基本途径
- 当前能够进行总体优化的自适应隐写均采用STC编码。自适应隐写的设计主要包含**失真函数的设计与应用STC编码两个环节**，其中，失真函数的设计与隐写的嵌入域非常相关
- 由于隐写分析在纹理复杂区的分类效果较差，因此，当前的失真函数设计普遍采用高纹理区失真值较低的原则；一些失真函数的设计也考虑了对统计分布特性的影响，对引起分布变化更明显的区域，失真函数一般被设计为输出较高的值



1-2 基本设计方法（加性总失真与单点失真）

- 在限定负载率下，加性模型下的自适应隐写基于STC编码最小化各个样点上的失真和

$$D(\boldsymbol{X},\boldsymbol{Y})=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\rho_{ij}\left|X_{i,j}-Y_{i,j}\right|$$

- 以上$D(\boldsymbol{X},\boldsymbol{Y})$是对实际总失真的估计，载体$\boldsymbol{X}$与含密载体$\boldsymbol{Y}$的尺寸是$n_1\times n_2$，$\rho_{ij}$是单点失真函数，它估计仅单点被修改的失真。STC编码中，需要知道ρ_{ij}，这需要隐写设计者对ρ_{ij}进行定义，或者，若能先定义总体失真函数$D(\boldsymbol{X},\boldsymbol{Y})$，则可以得到

$$\rho_{i,j}=D(\boldsymbol{X},\boldsymbol{X}_{\sim(i,j)}Y_{i,j})$$

- 其中，$\boldsymbol{X}_{\sim(i,j)}Y_{i,j}$表示仅仅将$\boldsymbol{X}$中$X_{i,j}$改为$Y_{i,j}$的版本
- 以下主要通过介绍$\rho_{i,j}$的定义方法介绍限定负载率下基于加性模型的自适应隐写

2 空间域自适应隐写

- 空域自适应隐写的失真函数一般在空间域设计，典型的是HUGO（Highly Undetectable steGO）隐写
- 由于小波变换是时频变换，同时包含载体局部的空间信息与频率信息，因此，也可以在小波域设计失真函数，典型的是WOW（Wavelet Obtained Weights）与S-UNIWARD（Spatial-UnIversal Wavelet Relative Distortion）隐写
- 以上两种情况下设计的**失真函数均在较高纹理区输出较小值**，但是在具体计算上有不同，造成安全性的差异

2.1-1 HUGO（技术目标）

- HUGO是最早采用STC的自适应隐写，显著提高了空间域图像隐写的安全。由于SPAM二阶特征能够很好地反映空间域隐写带来的扰动，具有一定的通用性，因此，HUGO的直接设计目的是使得空间域隐写能够抵御基于SPAM特征的隐写分析。为此，**HUGO隐写基于SPAM二阶特征构造失真函数**

2.1-2 HUGO（SPAM回顾）

- 记$\{\leftarrow, \rightarrow, \downarrow, \uparrow, \nwarrow, \searrow, \swarrow, \nearrow\}$为像素的8个方向，用于标记相邻像素在这些方向上的差值，如$D_{i,j}^{\rightarrow} = I_{i,j} - I_{i,j+1}$表示水平方向相邻像素差，**统计8个方向的相邻像素差值一阶与二阶转移概率**，如在从左向右水平方向统计一阶转移概率$M_{d_1,d_2}^{\rightarrow} = \Pr(D_{i,j+1}^{\rightarrow} = d_2 | D_{i,j}^{\rightarrow} = d_1)$与二阶转移概率$M_{d_1,d_2,d_3}^{\rightarrow} = \Pr(D_{i,j+2}^{\rightarrow} = d_3 | D_{i,j+1}^{\rightarrow} = d_2, D_{i,j}^{\rightarrow} = d_1)$，$d_i \in [-T, \cdots, T]$，$i = 1,2,3$；
- 最后，对一阶与二阶特征分别按照水平与对角线方向合并，其中，对二阶特征合并如下：

$$F_{d_1,d_2,d_3}^{+} = \frac{1}{4}(M_{d_1,d_2,d_3}^{\rightarrow} + M_{d_1,d_2,d_3}^{\leftarrow} + M_{d_1,d_2,d_3}^{\downarrow} + M_{d_1,d_2,d_3}^{\uparrow})$$

$$F_{d_1,d_2,d_3}^{\times} = \frac{1}{4}(M_{d_1,d_2,d_3}^{\searrow} + M_{d_1,d_2,d_3}^{\nwarrow} + M_{d_1,d_2,d_3}^{\swarrow} + M_{d_1,d_2,d_3}^{\nearrow})$$

- 在提取SPAM二阶特征中，一般取$T = 3$，因此，二阶特征的维度是$2(2T+1)^3 = 686$。实验表明，**SPAM二阶特征具有更好的分析效果**

2.1-3 HUGO的总失真函数

- 定义水平方向邻域像素差共生矩阵（Co-occurrence Matrices）：

$$C^{\rightarrow}_{d_1,d_2} = \Pr\left(D^{\rightarrow}_{i,j} = d_1, D^{\rightarrow}_{i,j+1} = d_2\right)$$

$$C^{\rightarrow}_{d_1,d_2,d_3} = \Pr\left(D^{\rightarrow}_{i,j} = d_1, D^{\rightarrow}_{i,j+1} = d_2, D^{\rightarrow}_{i,j+2} = d_3\right)$$

- 类似地可以定义其他方向的共生矩阵，当 $T = 3$ 时有

$$\left\{C^{\Gamma}_{d_1,d_2}, C^{\Gamma}_{d_1,d_2,d_3} \middle| \Gamma \in \{\rightarrow, \uparrow, \nwarrow, \nearrow\}, -3 \le d_i \le 3\right\}$$

- 其中包含 $4 \times (7^2 + 7^3) = 1568$ 维特征。由于$M^{\Gamma}_{d_1,d_2,d_3} = C^{\Gamma}_{d_1,d_2,d_3} / C^{\Gamma}_{d_1,d_2}$，并且$C^{\rightarrow}_{d_1,d_2} = C^{\leftarrow}_{-d_1,-d_2}, C^{\rightarrow}_{d_1,d_2,d_3} = C^{\leftarrow}_{-d_1,-d_2,-d_3}$（即方向可合并为），显然对以上共生矩阵特征的保持有利于抵抗基于$M^{\Gamma}_{d_1,d_2,d_3}$特征的SPAM隐写分析，因此HUGO的总体失真函数为：

$$D(\boldsymbol{X},\boldsymbol{Y}) = \sum_{d_1,d_2,d_3=-T}^{T} \left[w(d_1,d_2,d_3) \left| \sum_{\Gamma\in\{\rightarrow,\leftarrow,\uparrow,\downarrow\}} C^{\boldsymbol{X},\Gamma}_{d_1,d_2,d_3} - C^{\boldsymbol{Y},\Gamma}_{d_1,d_2,d_3} \right| + w(d_1,d_2,d_3) \left| \sum_{\Gamma\in\{\searrow,\nwarrow,\swarrow,\nearrow\}} C^{\boldsymbol{X},\Gamma}_{d_1,d_2,d_3} - C^{\boldsymbol{Y},\Gamma}_{d_1,d_2,d_3} \right| \right]$$

2.1-4 HUGO的单点失真函数

- 以上总失真函数表示从载体 $\boldsymbol{X}$ 修改为含密载体 $\boldsymbol{Y}$ 的失真，其中，$w(d_1,d_2,d_3)$ 是权值函数，定义为

$$w(d_1,d_2,d_3)=\frac{1}{\left(\sqrt{d_1^2+d_2^2+d_3^2}+\sigma\right)^{\gamma}}$$

- $\sigma,\gamma>0$ 需要确定的系数，实验中通过比对特征的变化搜索得到 $\sigma=10,\gamma=4$
- 在以上定义下，1）对3阶共生矩阵特征扰动总和越大失真越大，2）当d_1,d_2,d_3较小时（相对平滑区），相应共生矩阵特征扰动的相加权值较大
- 在STC编码中，需要计算单点失真函数 $\rho_{i,j}$，这里显然有 $\rho_{i,j}=D(\boldsymbol{X},\boldsymbol{X}_{\sim(i,j)}Y_{i,j})$

2.1-5 HUGO嵌入算法

- 基于以上失真函数，HUGO算法基于STC编码实现空域自适应±1隐写，为了实现部分非加性效果技术，还提供了模型矫正（Model Correction）选项
- 模型矫正功能：先进行STC计算但暂不修改，而是确定需要修改位置；在嵌入过程中逐步针对每个待修改样点再次计算当前的+1失真与-1失真，选择其中失真小者对应的修改方式

2.1-6 HUGO嵌入算法

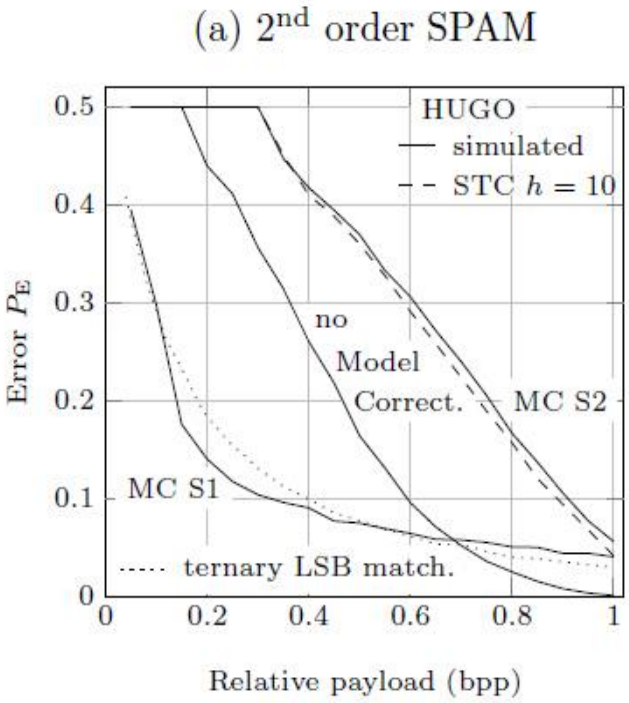
```
输入：载体X，消息message；输出：含密载体Y
1. For (i,j) in PIXELS {    //每个像素计算一次
2.     Yp = X; Yp(i,j)++; rho_p(i,j) = D(X, Yp); //计算像素+1的单点失真
3.     Ym = X; Ym(i,j)--; rho_m(i,j) = D(X, Ym); //计算像素-1的单点失真
4. }
5. rho_min = min(rho_p, rho_m);//采用+1与-1中小的单点失真进行STC编码
   //以下隐含进行了STC编码，但是仅仅返回需要修改的像素位置
6. PIXELS_TO_CHANGE = minimize_emb_impact(LSB(X), rho_min, message);
7. Y = X;  //含密载体的初态
8. For (i,j) in PIXELS_TO_CHANGE {//以上确定了修改位置，现在确定+1还是-1
9.     If (model_correction_step_enabled) { //模型矫正开关开，则新计算+-失真
          //计算当前修改程度下的当前位置上+1与-1单点失真
10.        Yp = Y; Yp(i,j)++; dp = D(X, Yp); Ym = Y; Ym(i,j)--; dm = D(X, Ym);
11.        If  (dp < dm) { Y(i,j)++;} else { Y(i,j)--;}
12.    }
13.    else {//无模型矫正开关则用之前的+-失真
14.        If (rho_p(i,j) < rho_m(i,j)) { Y(i,j)++;} else { Y(i,j)--;}
15.    }
16.}
```

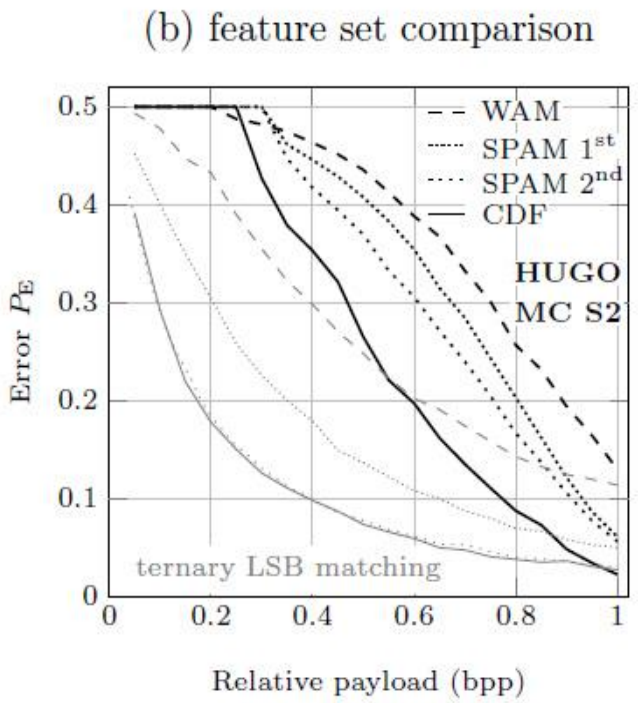
2.1-7 HUGO的安全性实验

- 在模型矫正中，存在空间域矫正方向的选择。它实际就是修改的方向，HUGO分为以下几种：（S1）从左到右、从上到下；（S2）从最大的 $\rho_{i,j}$ 到最小的顺序；（S3）从最小的 $\rho_{i,j}$ 到最大的顺序；（S4）随机顺序。实验发现，顺序S2的安全性最好（图）
- HUGO的抗检测性能：（a）SPAM攻击下LSBM、HUGO与理想加性嵌入的性能，$T = 3$，simulated表示对理想加性模型下最优嵌入的模拟；（b）HUGO S2与LSBM在4种攻击下的性能比较

CDF（Cross-Domain Feature）：融合了SPAM与Pev-274

WAM：Wavelet Absolute Moments





SKLOIS
信息安全国家重点实验室

2.2-1 WOW（小波滤波器组）

- 小波系数反映了信号局部区域的时频特性，失真函数的设计可以以隐写后这种特性得到更多保持为原则。WOW是空间域图像自适应隐写，它在小波变换的一级分解LH、HL与HH子带上对系数改动量与原系数做相关运算，总体相关性越大则失真越低
- WOW选择了Daubechies 8小波低通、高通滤波器$\boldsymbol{h}$，$\boldsymbol{g}$进行小波滤波，则获得一级分解LH、HL与HH三个子带系数的二维滤波器系数矩阵是：$\boldsymbol{K}^{(1)} = \boldsymbol{h} \cdot \boldsymbol{g}^{\mathrm{T}}$，$\boldsymbol{K}^{(2)} = \boldsymbol{g} \cdot \boldsymbol{h}^{\mathrm{T}}$，$\boldsymbol{K}^{(3)} = \boldsymbol{g} \cdot \boldsymbol{g}^{\mathrm{T}}$。它们滤波后得到的三个子带系数也称为三组残差（Residual），分别为$\boldsymbol{R}^{(k)} = \boldsymbol{K}^{(k)} * \boldsymbol{X}$，$k = 1,2,3$，“$*$”表示二维卷积
- 在WOW中，先对载体进行镜像填充（Mirror-padded），之后在计算卷积，因此，$\boldsymbol{R}^{(k)}$的尺寸与原图$\boldsymbol{X}$相同

2.2-2 WOW 失真函数

- 设 $\boldsymbol{R}_{ij}^{(k)}$为仅仅修改了位置$(i,j)$上一个像素后计算得到的第$k$ 组残差，$k=1,2,3$，则WOW定义的一般单点失真函数定义为

$$\rho_{i,j}=\left(\sum_{k=1}^{n}\left|\xi_{ij}^{(k)}\right|^{p}\right)^{-\frac{1}{p}}$$

- 其中，n 为二维滤波器数量，$p<0$ 是待定常数，并且

$$\xi_{ij}^{(k)}=\left|\boldsymbol{R}^{(k)}\right|\circledast\left|\boldsymbol{R}^{(k)}-\boldsymbol{R}_{ij}^{(k)}\right|;\ \text{//更希望修改纹理复杂高频区}$$

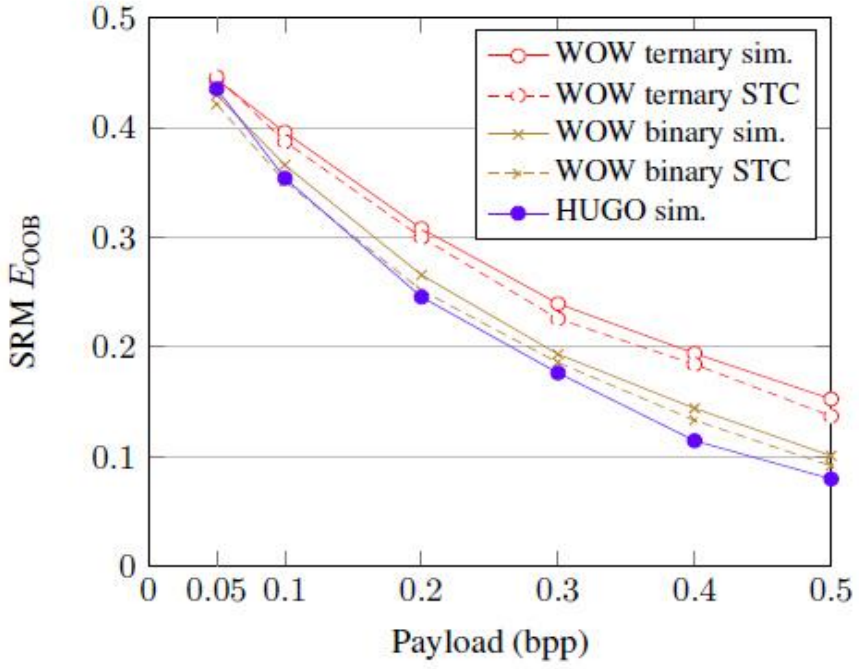
- $\circledast$ 表示计算相关性。在实际中，WOW取 $n=3$ 以及 $p=-1$，有

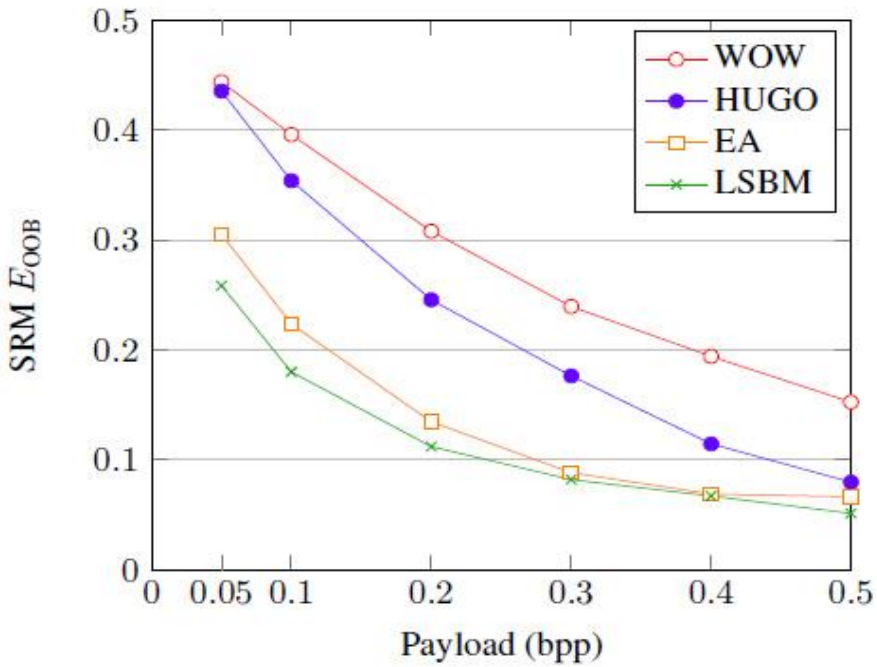
$$\rho_{i,j}=\sum_{k=1}^{3}\left|\xi_{ij}^{(k)}\right|^{-1}$$

- 因此，当修改位置(i,j)上像素后，如果残差上的噪声与原残差更相关，则WOW的单点失真函数较小，(i,j)位置更可能被修改
- WOW采用随机±1的方式进行修改，采用STC。由于 $+1$ 与-1的失真是一样的，因此，支持采用双层STC进行嵌入

2.2-3 WOW 安全性比较

- 从实验结果看，单层WOW的安全性高于HUGO，显著高于非STC自适应的方法（sim表示在加性模型下对基于WOW失真函数的理想嵌入模拟）
- SRM采用了集成分类器投票的判决方法，图中的 E_{OOB}表示Error of Out of Bag，它是通常评价集成分类器错误率的指标，主要特点是，将训练样本（对每个子分类器采用放回抽样的方法在标注样本集合中选取）与未选入训练的样本一并作为测试样本，统计得到的错误率





2.3-1 S-UNIWARD（失真函数）

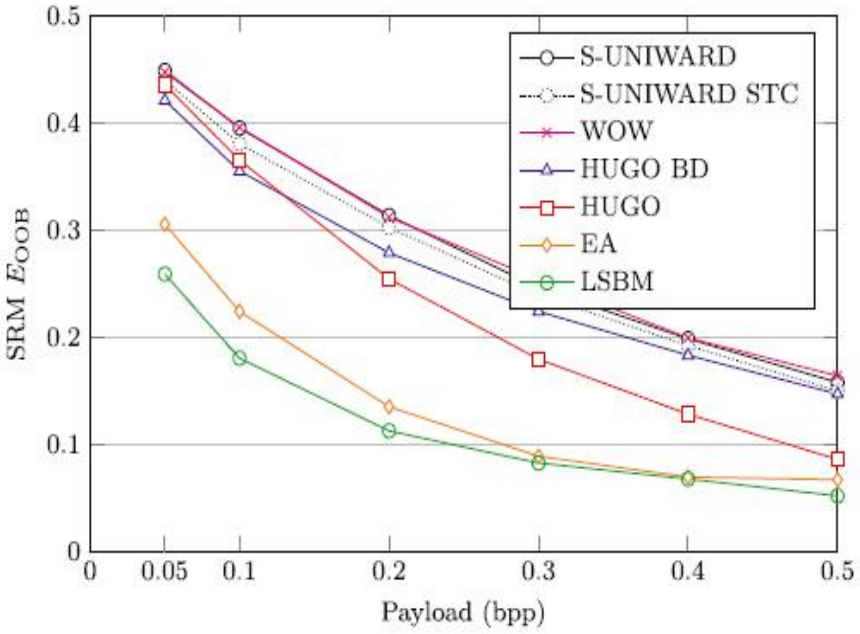
- S-UNIWARD也是一种图像空间域自适应隐写，它的失真函数基于小波系数定义。设 $\boldsymbol{R}^{(k)}(\boldsymbol{X}) = \boldsymbol{K}^{(k)} * \boldsymbol{X}$，$\boldsymbol{R}^{(k)}(\boldsymbol{Y}) = \boldsymbol{K}^{(k)} * \boldsymbol{Y}, k = 1,2,3$ 分别是原图与含密图的第 k 组残差，$W_{uv}^{(k)}(\boldsymbol{X})$，$W_{uv}^{(k)}(\boldsymbol{Y}), u \in [1, \cdots, n_1]$，$v \in [1, \cdots, n_2]$ 分别表示$\boldsymbol{R}^{(k)}(\boldsymbol{X})$、$\boldsymbol{R}^{(k)}(\boldsymbol{Y})$ 中位置(u, v)上的小波系数，则S-UNIWARD算法定义的总体失真函数为：

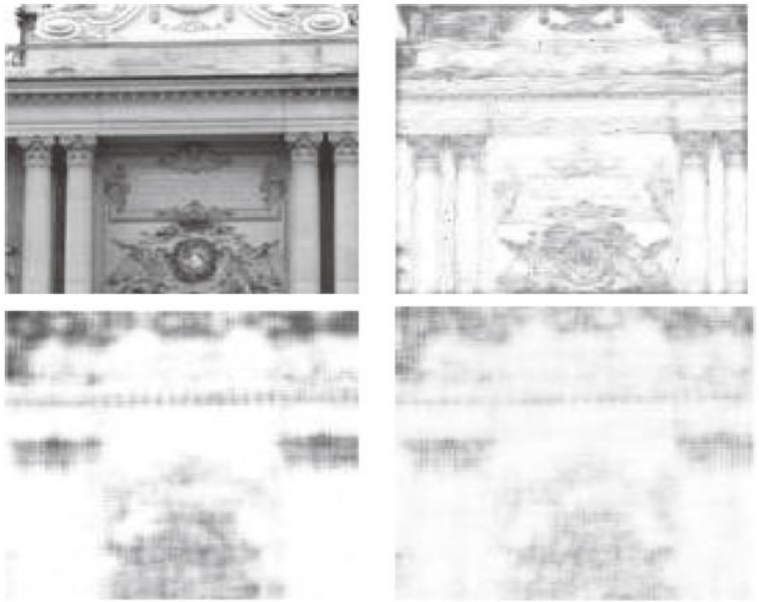
$$D(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{k=1}^{3} \sum_{u=1}^{n_1} \sum_{v=1}^{n_2} \frac{\left| W_{uv}^{(k)}(\boldsymbol{X}) - W_{uv}^{(k)}(\boldsymbol{Y}) \right|}{\sigma + \left| W_{uv}^{(k)}(\boldsymbol{X}) \right|}$$

- $\sigma > 0$ 为调节参数，在S-UNIWARD的实验中取为1。这样，单点失真函数为 $\rho_{i,j} = D(\boldsymbol{X}, \boldsymbol{X}_{\sim(i,j)} Y_{i,j})$，由于以上小波系数系数主要分布在中高频，在纹理区数值较大，使得上式的分母较大，整个失真下降
- S-UNIWARD对像素的修改方法类似于WOW。由于$Y_{i,j}$加1还是减1得到的 $\rho_{i,j}$相同，因此，S-UNIWARD的修改方式可以是在STC编码中随机的± 1，也支持双层的嵌入

2.3-2 S-UNIWARD（安全性）

- 隐写分析实验结果表明，相比HUGO与WOW，S-UNIWARD进一步提高了安全性（图）。在嵌入位置的选择上，HUGO比较趋向于选择轮廓附近，而WOW、S-UNIWARD更趋向于选择纹理复杂区，后者的选择在纹理区范围更均匀





S-UNIWARD STC在SRM隐写分析下的性能（其中S-UNIWARD指理想模拟）；（b）HUGO、WOW、S-UNIWARD的嵌入区域选择对比

3 JPEG域自适应隐写

- JPEG域自适应隐写需要估计修改每个可嵌入JPEG系数的单点失真
- 一般可以变换到空间域或者小波域进行设计，原则也是将扰动尽量限制在纹理区，典型的是J-UNIWARD（JPEG-UNIversal Wavelet Relative Distortion）隐写
- 也可以直接在JPEG域进行失真函数的设计，设计的原则是，尽量选择对JPEG系数分布特性扰动小的位置进行嵌入，典型的是UED（Uniform Embedding Distortion）隐写

3.1-1 J-UNIWARD（失真函数）

- J-UNIWARD是一种JPEG域自适应隐写。由于隐写的输入与输出都是JPEG文件，因此，以上的载体图像 $\boldsymbol{X}$ 与含密图像 $\boldsymbol{Y}$ 均在JPEG域。设 $J^{-1}(\boldsymbol{X})$与$J^{-1}(\boldsymbol{Y})$表示JPEG解码到空间域的操作，则J-UNIWARD的总失真函数定义为：

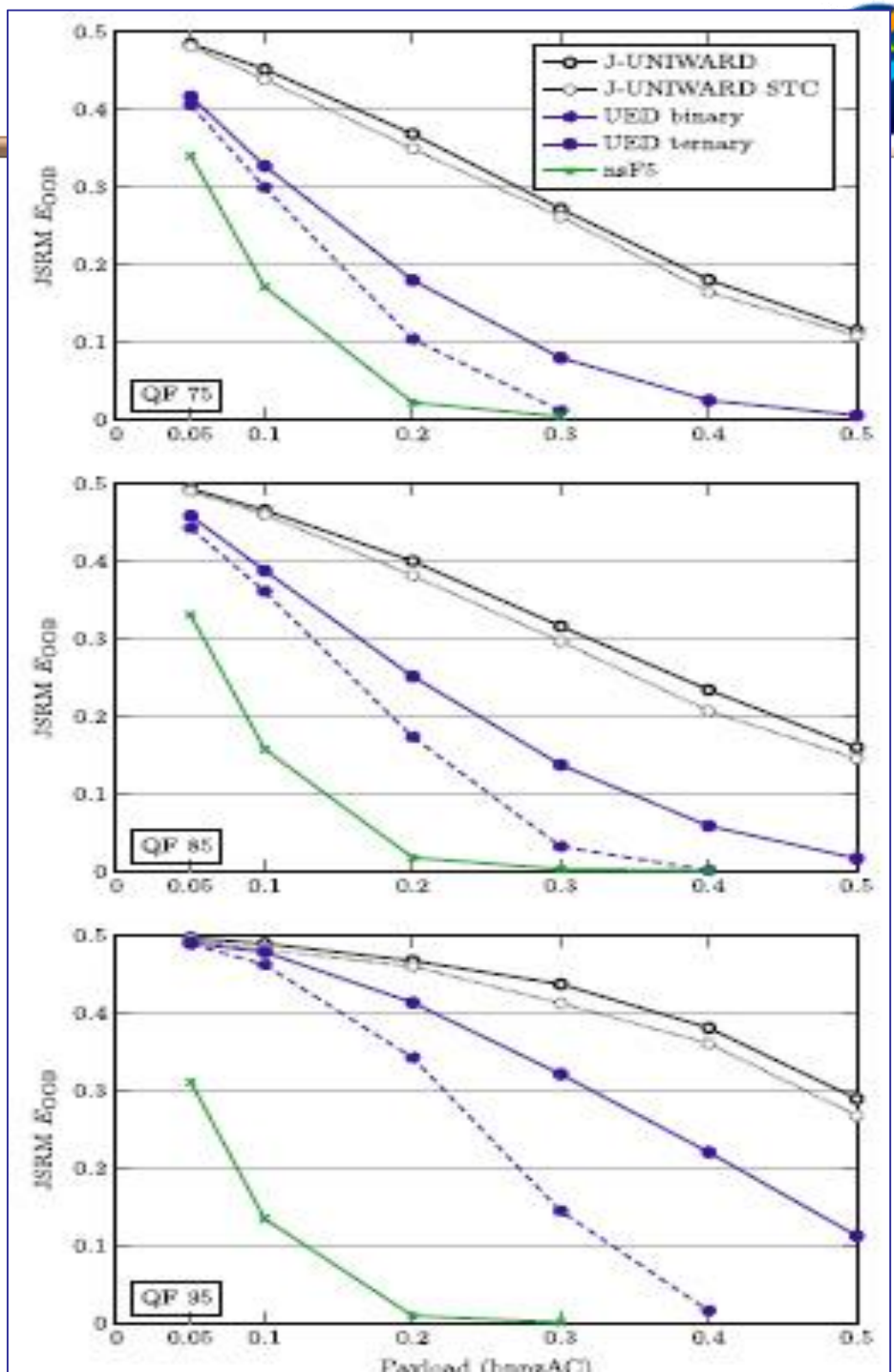
$$D'(\boldsymbol{X},\boldsymbol{Y}) = D(J^{-1}(\boldsymbol{X}),J^{-1}(\boldsymbol{Y}))$$

- 其中，D 是S-UNIWARD的总失真函数，待定系数$\sigma > 0$ 在实验中取为2^{-6}
- 这样，单点失真函数为 $\rho_{i,j}= D'(\boldsymbol{X},\boldsymbol{X}_{\sim(i,j)}Y_{i,j})$；这里，$i,j$为JPEG量化DCT系数的坐标

3.1-2 J-UNIWARD安全

在JSRM通用隐写分析下（图），J-UNIWARD比UED的抗检能力更强，安全性显著超出非自适应的JPEG隐写nsF5.

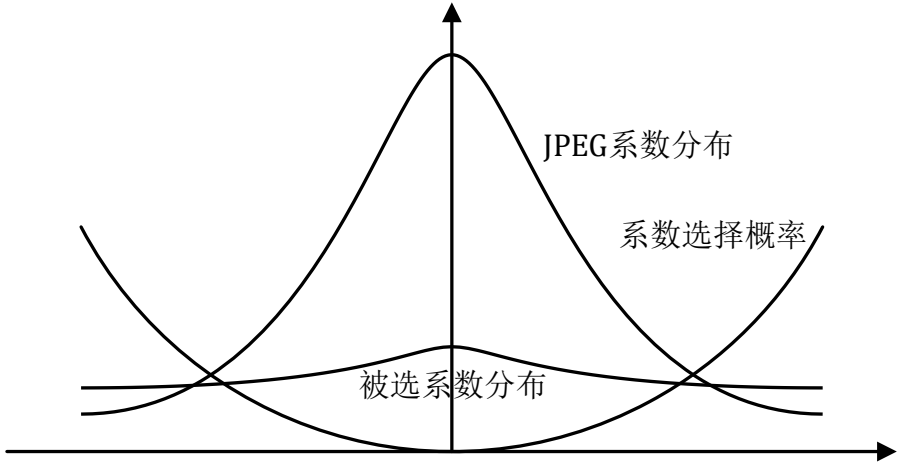
分别在质量因子QF为75、85与95三组样本下测试



SKLOIS
信息安全国家重点实验室

3.2-1 UED（设计思想）

- UED是一种JPEG域的自适应隐写。UED的设计者观察到，若任意选择嵌入位置，由于小值JPEG系数的数量较多，则被选择到的可能性更大，这样，小值系数上的分布会变化更剧烈，更可能造成系数分布的形变
- 而如果每个值上的变化接近，则更可能保持JPEG系数的分布。因此，UED的设计目标是，压制在更多出现的小值系数上修改，优先选择更少出现的大值系数修改（图）



UED的修改系数选择原则：优先选择大值系数修改。图中，a是JPEG量化系数分布，b是希望的选择概率，c是所选择系数的分布（更均匀）

3.2-2 UED（失真函数）

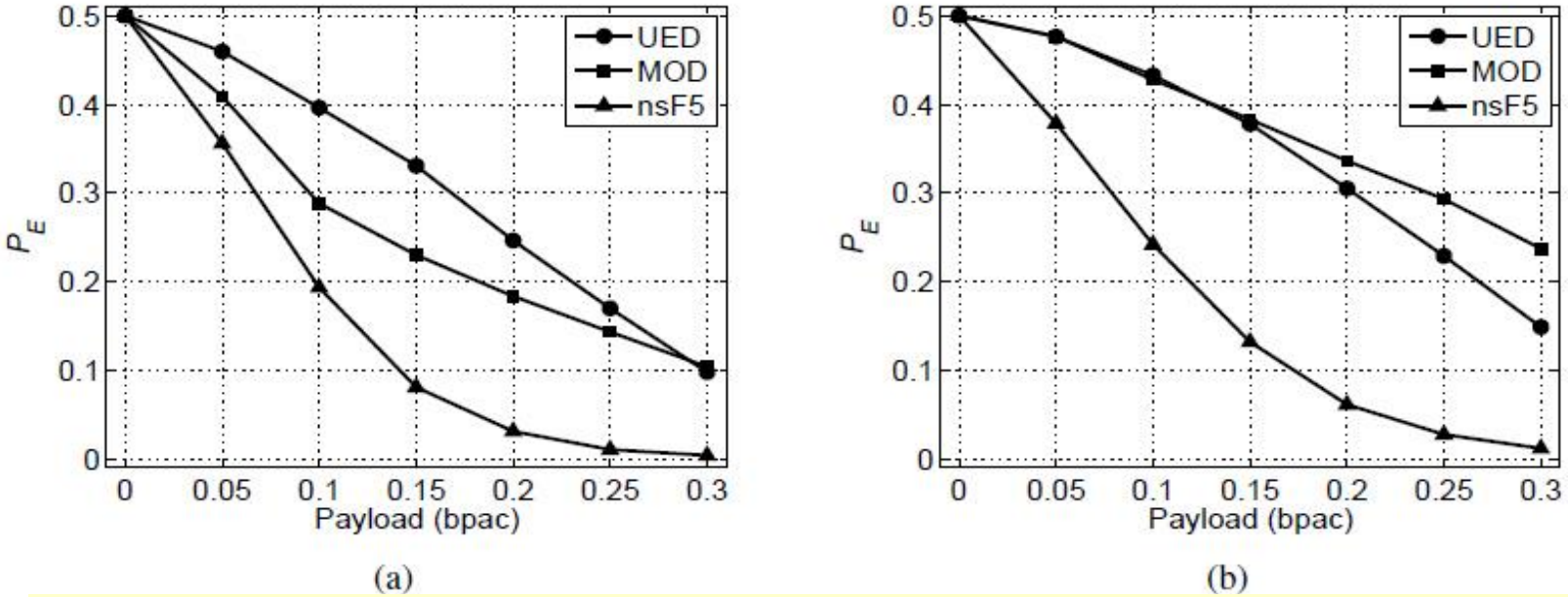
- 设c_{ij}表示位置(i,j)上的JPEG量化系数，定义$N_{\mathrm{ia}}=\{c_{i+1,j},c_{i-1,j},c_{i,j+1},c_{i,j-1}\}$是它的块内邻域，$N_{\mathrm{ir}}=\{c_{i+8,j},c_{i-8,j},c_{i,j+8},c_{i,j-8}\}$是它的块间邻域，则单点失真函数为：

$$\rho_{i,j}=\sum_{d_{\mathrm{ia}}\in N_{\mathrm{ia}}}\left(|c_{ij}|+|d_{\mathrm{ia}}|+\alpha_{\mathrm{ia}}\right)^{-1}+\sum_{d_{\mathrm{ir}}\in N_{\mathrm{ir}}}\left(|c_{ij}|+|d_{\mathrm{ir}}|+\alpha_{\mathrm{ir}}\right)^{-1}$$

- 以上α_{ia}与α_{ir}为待定常数，有实验结果分别确定为1.3与1；c_{ij}为系数值；以上失真函数压制了在小值区的取值

3.2-3 UED（安全性比较）

- 基于以上失真函数，UED采用STC优化嵌入，不使用系数为0的样点，当需要修改且c_{ij}绝对值大于1时，采用随机± 1的方式；当需要修改且c_{ij}绝对值为1时，将其改为$c_{ij} + \mathrm{sgn}(c_{ij})$，即+1改为2，$-1$改为$-2$
- 用CC-JRM与CC-PEV特征进行相应的隐写分析，结果显示，UED显著优于非自适应的方法（图），但是，前面已经说明，在主要攻击下它的安全性低于J-UNIWARD（不过UED速度显著更快）



用CC-JRM（左）与CC-PEV （右）特征对UED进行相应的隐写分析的比较结果

4 文献阅读推荐

- [1] 教材第12章
- [2] T. Pevny, T. Filler, P. Bas. Using high-dimensional image models to perform highly undetectable steganography. In Proc. IH 2010, LNCS 6387, pp. 161–177, Springer, 2010.
- [3] V. Holub, J. Fridrich. Designing steganography distortion using directional filters. In Proc. WIFS 2012, Tenerife, Spain, December 2-5, pp. 234-239, 2012.
- [4] V. Holub, J. Fridrich, and T. Denemark. Universal distortion function for steganography in an arbitrary domain. EURASIP Journal on Information Security 2014, 2014: 1
- [5] L. Guo, J. Ni, Y. Q. Shi. An efficient JPEG steganographic scheme using uniform embedding. In Proc. WIFS 2012, Tenerife, Spain, December 2-5, pp. 169-174, 2012.

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