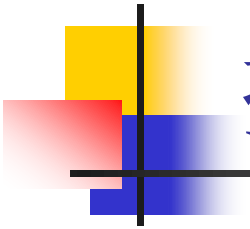
3.1、隐写术基本理论

信息安全中心
钮心忻、杨榆、雷敏



提纲

- 隐写系统
- 隐写系统分类
- 隐写术性能指标
- 隐写系统的攻击方法

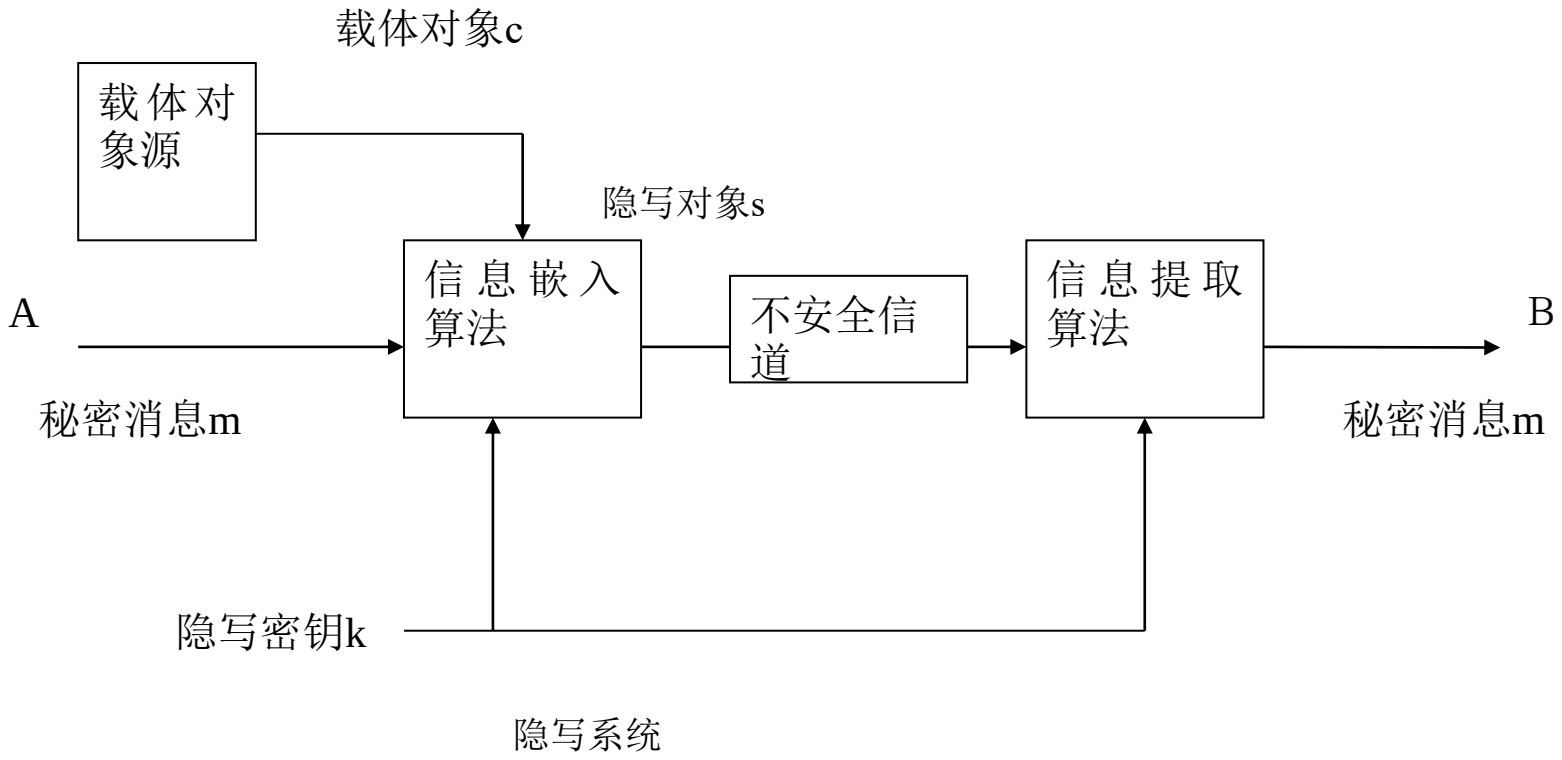
提纲

- ***隐写系统***
- 隐写系统分类
- 隐写术性能指标
- 隐写系统的攻击方法

囚犯问题

- 保密通信双方
  - 两个囚犯A和B被关押在监狱的不同牢房，他们想通过一种隐蔽的方式交换信息，但是交换信息必须要通过看守的检查。因此，他们要想办法在不引起看守者怀疑的情况下，在看似正常的信息中，传递他们之间的秘密信息。
- 被动看守者
  - 只是检查传递的信息有没有可疑的地方。
- 主动看守者
  - 故意去修改一些可能隐藏有信息的地方，或者假装自己是其中的一个囚犯，隐藏进伪造的消息，传递给另一个囚犯。

隐写系统



隐写系统

隐写系统——术语

- A打算秘密传递一些信息m给B，A需要从一个载体对象源中随机地选取一个无关紧要的载体对象c。当这个对象公开传递时，不会引起怀疑，称这个对象c为**载体对象** (carrier)。
- 把需要秘密传递的消息m隐藏到载体对象c后，载体对象c就变为**隐写对象**s (stego)。
- 把秘密消息的嵌入载体对象的过程需要密钥，此密钥称为**隐写密钥**k(stego key)。

隐写系统示例-LSB隐写系统

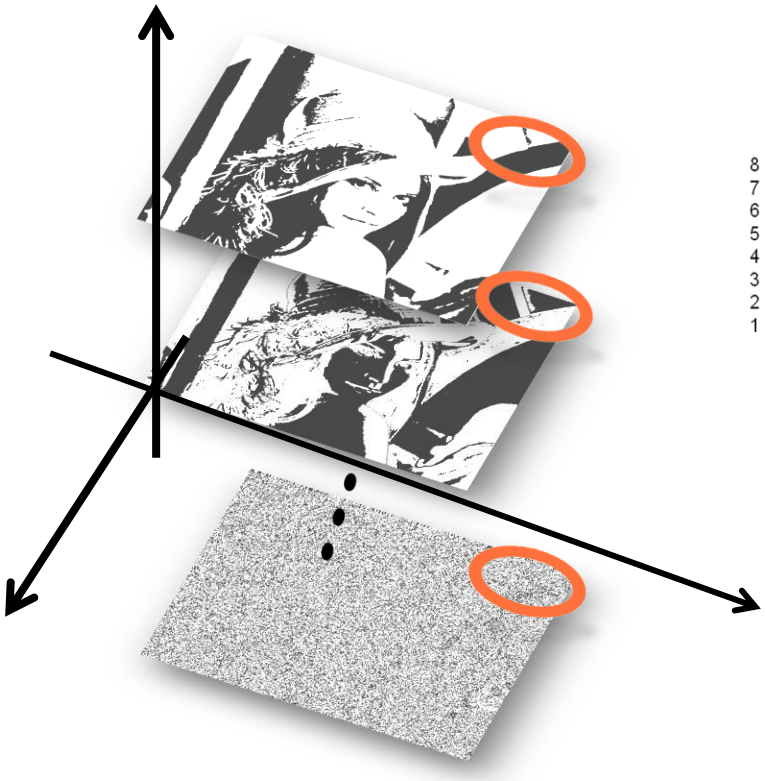
- 图像的数字化表示

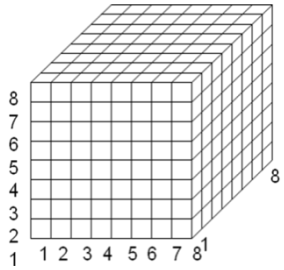
$I(x,y)$

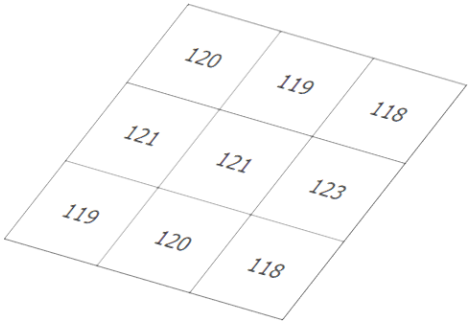
120	119	118
121	121	123
119	120	118

隐写系统示例-LSB隐写系统

- 图像的数字化表示

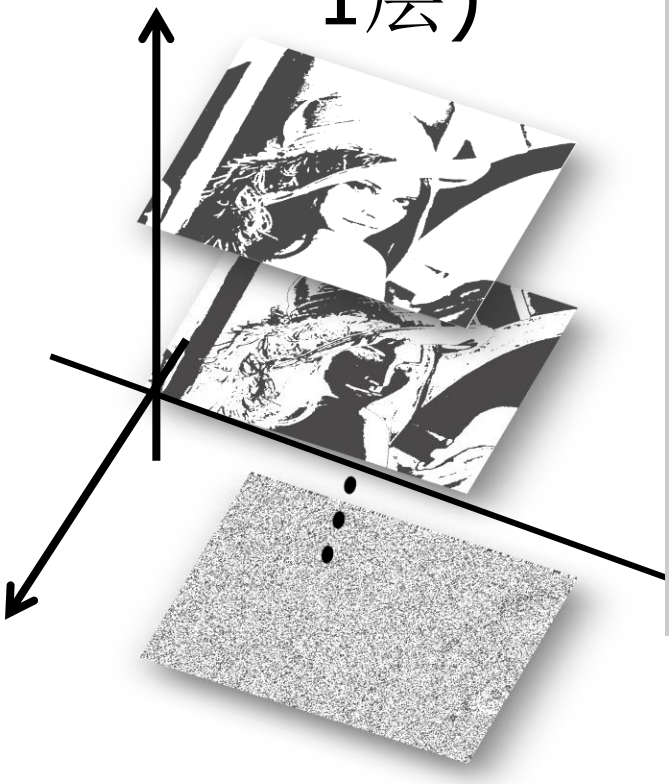


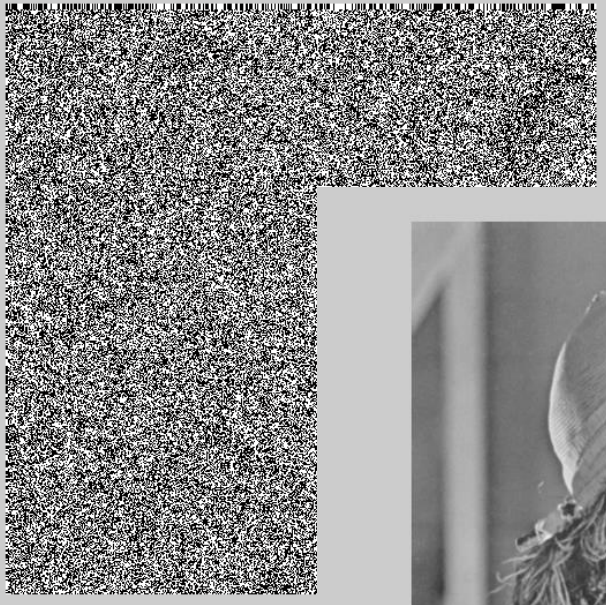




隐写系统示例-LSB隐写系统

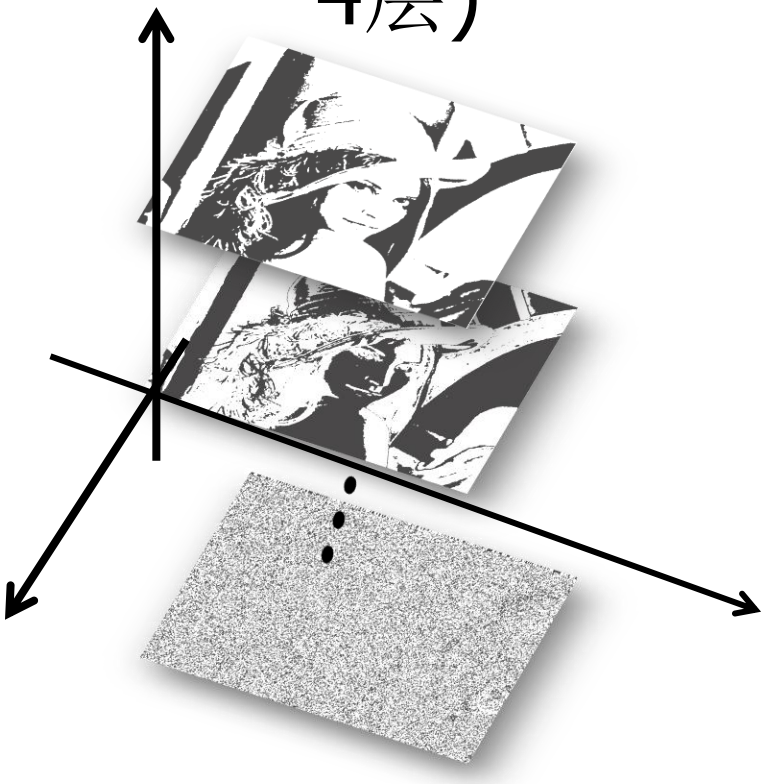
- 不同比特平面（层）对图像视觉质量(第1层)

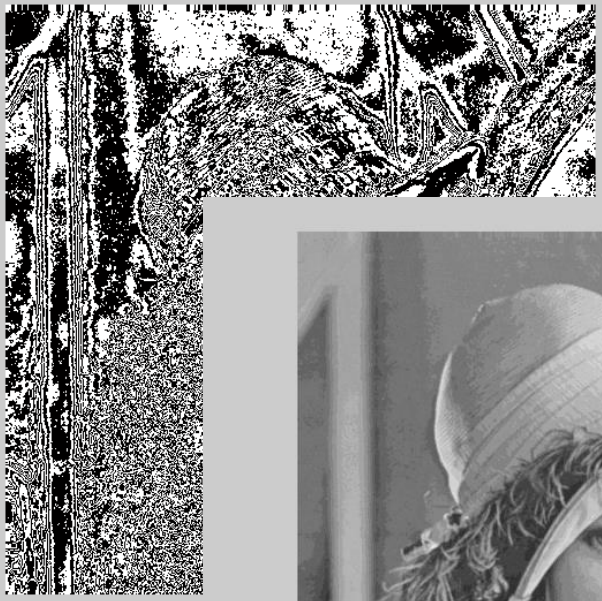




隐写系统示例-LSB隐写系统

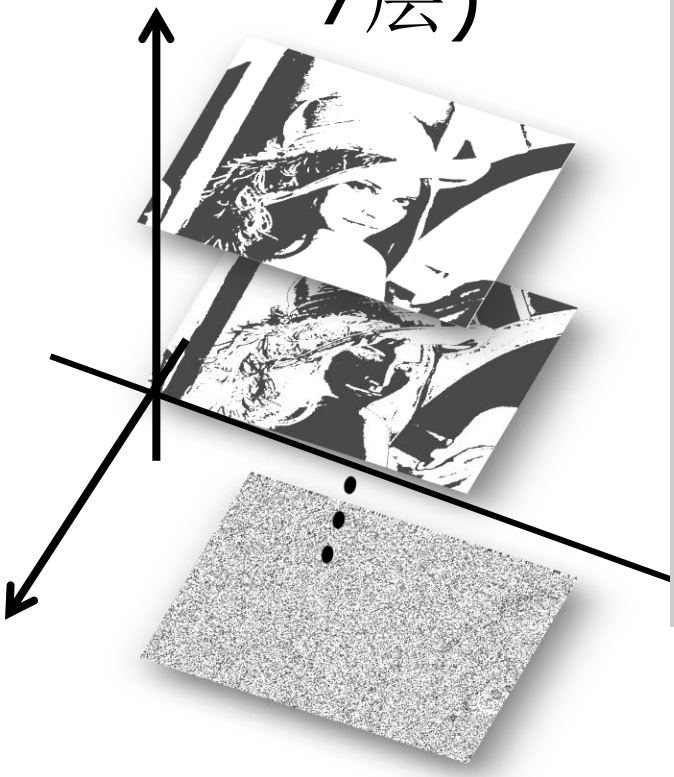
- 不同比特平面（层）对图像视觉质量(第4层)





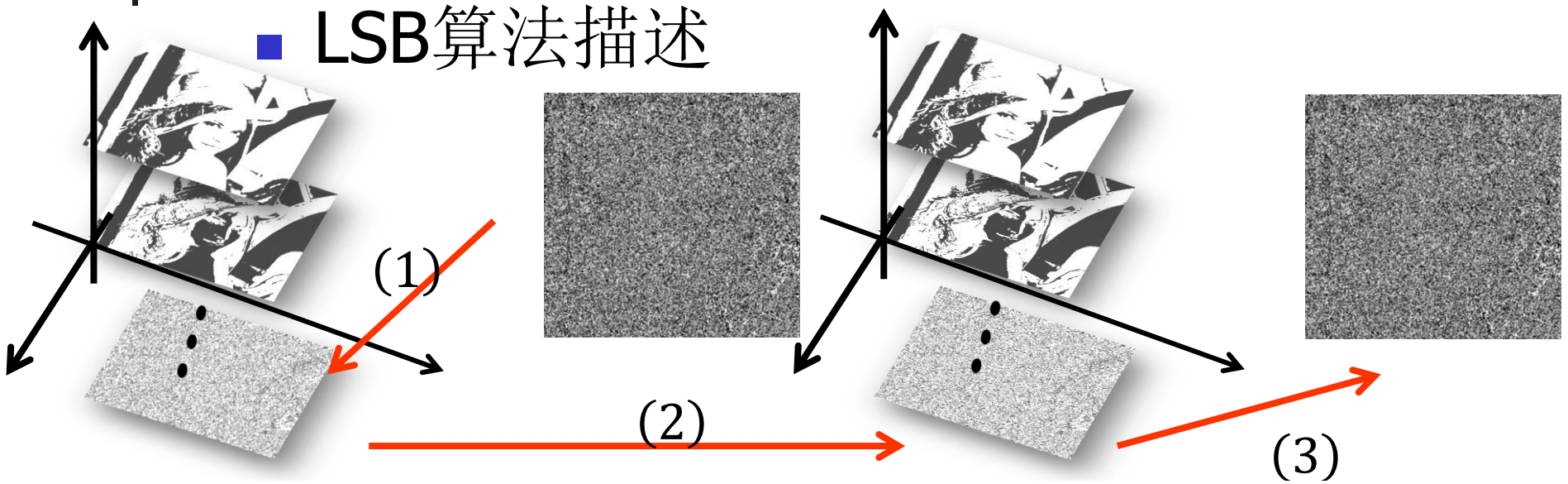
隐写系统示例-LSB隐写系统

- 不同比特平面（层）对图像视觉质量(第7层)



隐写系统示例-LSB隐写系统

- LSB算法描述



隐藏：用秘密信息替换最低有效比特；提取：提取最低有效比特还原信息

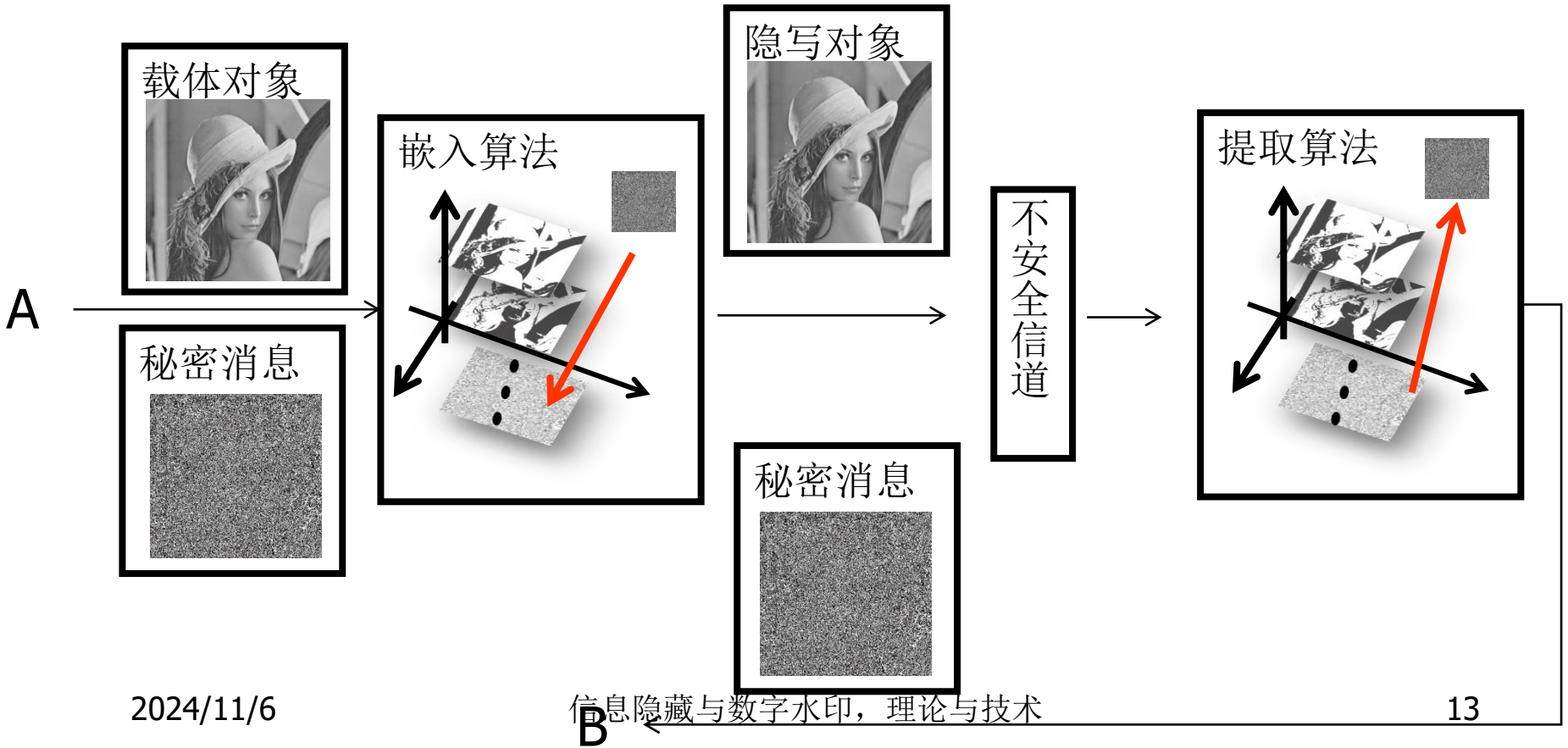
载体样点：0110 001**1**　0101 011**1**　0111 011**0**

秘密消息：　　**0**　　**0**　　**1**

隐写样点：0110 001**0**　0101 011**0**　0111 011**1**

隐写系统示例-LSB隐写系统

- 基于LSB算法的隐写系统



提纲

- 隐写系统
- ***隐写系统分类***
- 隐写术性能指标
- 隐写系统的攻击方法

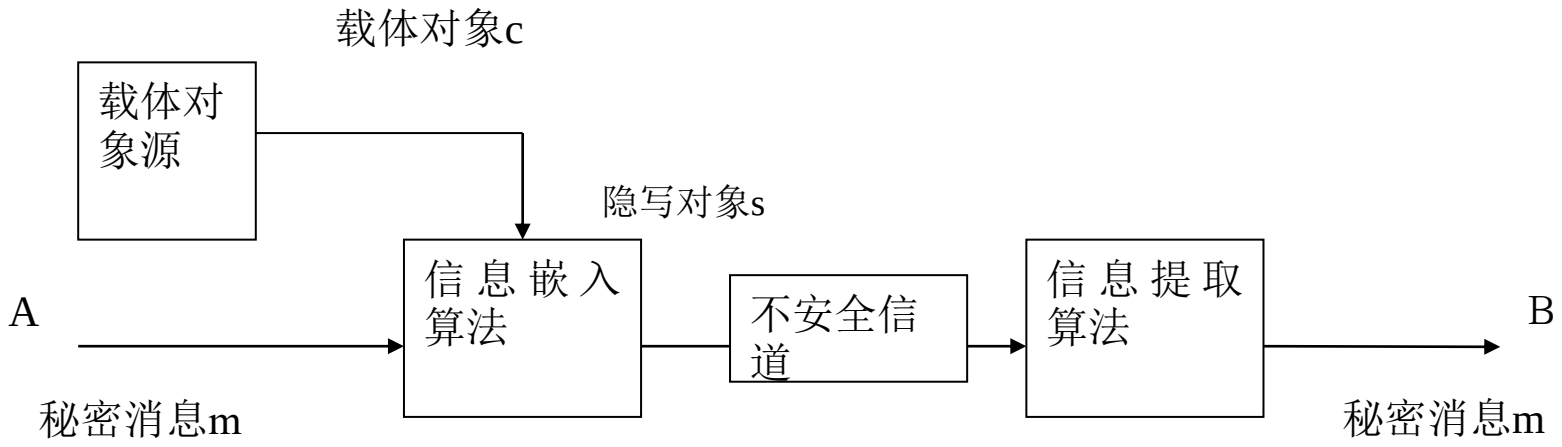
隐写系统分类

- 按载体
  - 视频、文本、图像...
- 按采用的技术
  - 空间域、变换域、基于扩频、基于结构
- 按隐藏时是否使用载体
  - 有载体隐写，无载体隐写
- 按提取时是否使用原始载体
  - 盲和非盲提取

隐写系统分类

- 按隐藏协议
  - 无密钥隐写系统
  - 私钥隐写系统
  - 公钥隐写系统

无密钥隐写系统



无密钥隐写系统

略过深入讨论

无密钥隐写系统（深入讨论）

- 隐藏过程：映射$E: C \times M \rightarrow S$
  - C: 所有可能载体对象的集合
  - M: 所有可能秘密消息的集合
  - S: 所有隐写对象的集合
- 提取过程：映射$D: S \rightarrow M$
- 双方约定嵌入算法和提取算法，算法要求保密

无密钥隐写系统（深入讨论）

- **定义：**对一个五元组Σ=〈C，M，S，D，E〉，其中C是所有可能载体对象的集合，M是所有可能秘密消息的集合，S是所有可能隐写对象的集合
- E：C×M→S是嵌入函数
- D：S→M是提取函数
- 若满足性质：对所有m∈M和c∈C，恒有：D(E(c，m))=m，
- 则称该五元组为无密钥隐写系统

无密钥隐写系统（深入讨论）——相似性函数

- 载体对象和隐写对象在感觉上不可区分，如何度量？
  - **定义：**设C是一个非空集合，一个函数 sim：$C^2 \rightarrow (-\infty,1)$，对x，y∈C，若满足：

$$sim(x,y)\begin{cases}=1 & x=y\\ <1 & x\neq y\end{cases}$$

  - 则sim称为C上的相似性函数
  - 相似度应尽可能接近1

无密钥隐写系统（深入讨论）——相似性函数

- 相似函数实例
  - 设 $C \subset R^n$ 是非空集合，$\boldsymbol{x} = (x_1 x_2 \cdots x_n) \in C, \boldsymbol{y} = (y_1 y_2 \cdots y_n) \in C$，则 $f = \dfrac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$ 为 C 上的相似函数。
  - 证明：
    - 由 $f = \dfrac{\langle \boldsymbol{x} \cdot \boldsymbol{y} \rangle}{|\boldsymbol{x}||\boldsymbol{y}|} = \cos\theta$（$\theta$ 为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间夹角）可知，$-1 \le f \le 1$；当 $\boldsymbol{x}=\boldsymbol{y}$ 时，$f = \dfrac{\langle \boldsymbol{x} \cdot \boldsymbol{x} \rangle}{|\boldsymbol{x}||\boldsymbol{x}|} = 1$

无密钥隐写系统（深入讨论）——载体的选择

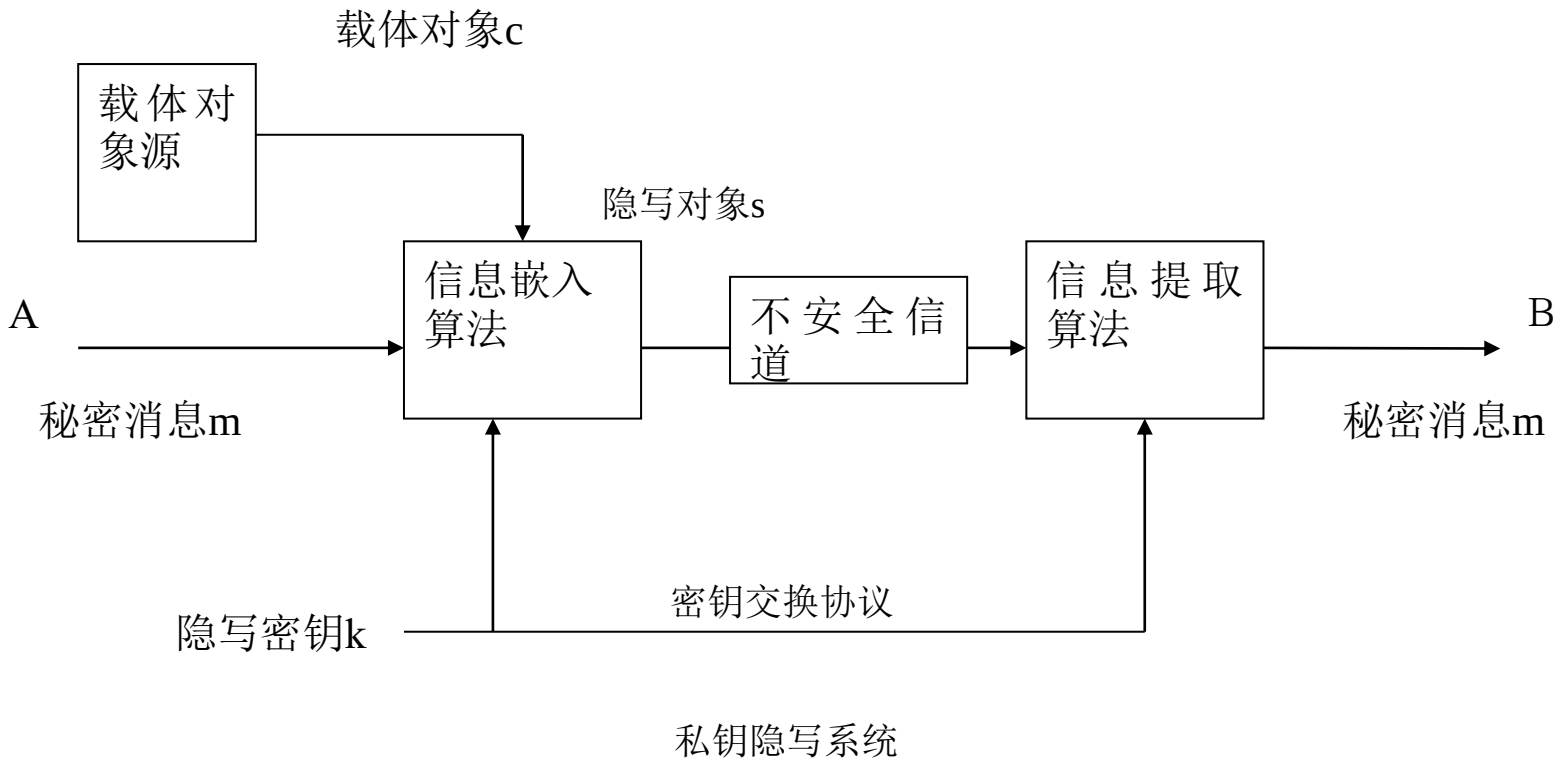
- 不同的嵌入算法，对载体的影响不同。
- 不同的载体，能隐藏秘密信息数量不同。
- 选择最合适的载体，信息嵌入对其影响最小，即载体对象与隐写对象的相似度最大。

$$c = \underset{x \in C}{Max}\, sim(x, E(x, m))$$

私钥隐写系统

- Kerckhoffs准则
  - 密码设计者应该假设对手知道数据加密的方法，数据的安全性必须仅依赖于密钥的安全性 。
- 无密钥隐写系统
  - 违反了Kerckhoffs准则。

私钥隐写系统

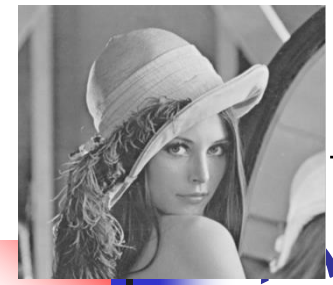


私钥隐写系统

略过深入讨论

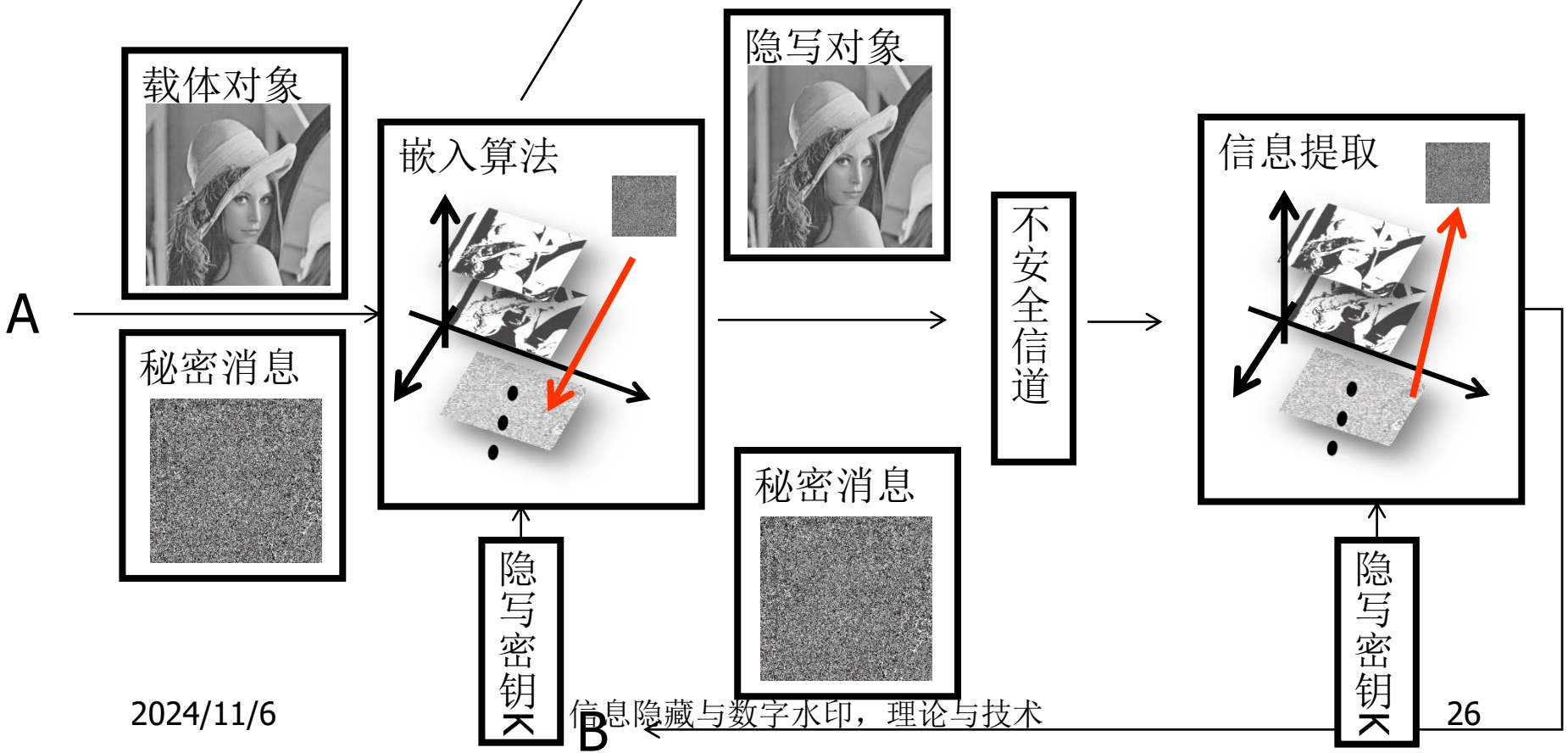
私钥隐写系统（深入讨论）

- 定义：
  - 对一个六元组$\Sigma=\langle C，M，K，S，D_K，E_K\rangle$，其中C是所有可能载体对象的集合，M是所有可能秘密消息的集合，K是所有可能密钥的集合，E_K：$C\times M\times K\rightarrow S$是嵌入函数，$D_K$：$S\times K\rightarrow M$是提取函数，若满足性质：对所有$m\in M$，$c\in C$和$k\in K$，恒有：$D_K(E_K(c，m，k)，k)=m$，则称该六元组为私钥隐写系统
- 私钥的传递：密钥交换协议

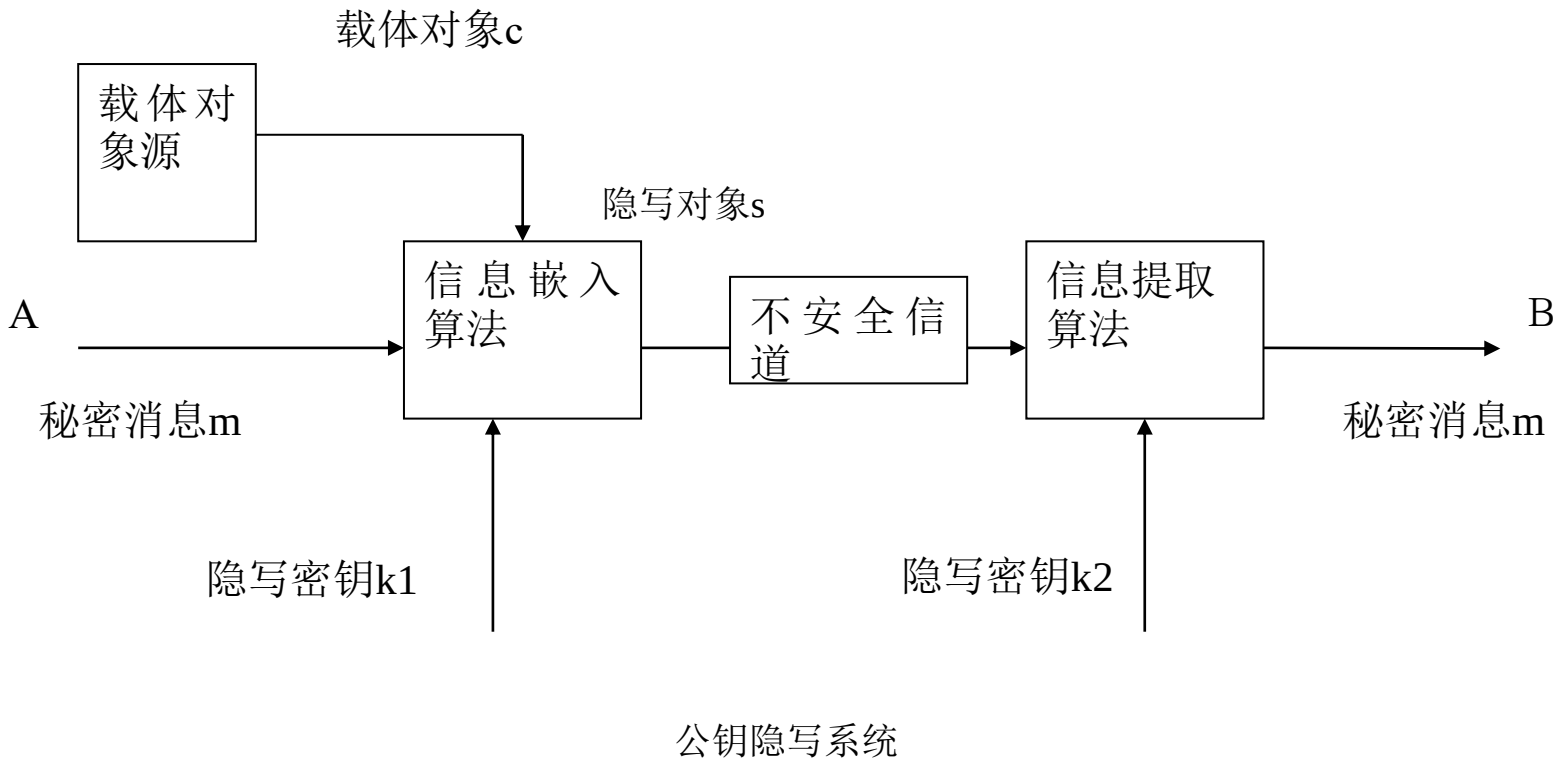


钥隐 统

- 基于LSB的私钥隐写系统



公钥隐写系统



公钥隐写系统

略过深入讨论

公钥隐写系统（深入讨论）

- **定义：**
  - 对一个七元组$\Sigma=\langle C，M，K1，K2，S，D_{K2}，E_{K1}\rangle$，其中C是所有可能载体的集合，M是所有可能秘密消息的集合，K1是公开密钥集合，K2是私有密钥集合，E_{K1}：$C\times M\times K1\rightarrow S$是嵌入函数，$D_{K2}$：$S\times K2\rightarrow M$是提取函数，若满足性质：对所有$m\in M$，$c\in C$, $k1\in K1$，$k2\in K2$，（k1和k2是一对）恒有：$D_{K2}(E_{K1}(c，m，k1)，k2)=m$，则称该七元组为公钥隐写系统

公钥隐写系统

- 类似于公钥密码
- 通信各方使用约定的公钥体制，各自产生自己的公开钥和秘密钥，将公开钥存储在一个公开的数据库中，通信各方可以随时取用，秘密钥由通信各方自己保存，不予公开.
- 发送方：用对方的公开钥将需要传递的秘密信息进行加密，再隐藏.
- 接收方：提取隐藏信息，用自己的私有密钥解密.

公钥隐写系统——问题

- 公钥隐写术只是借用公钥密码的思想，对秘密信息先加密，再隐藏。
- 类似于公钥密码算法的公钥隐藏算法？？目前还没有

提纲

- 隐写系统
- 隐写系统分类
- ***隐写算法性能指标***
- 隐写系统的攻击方法

隐写算法性能指标

- 容量（Capacity）
  - 负载（Payload），载体数据利用率，嵌入效率（Embedding Efficiency）
- 不可感知性（Imperceptibility）
  - 透明性（Transparency）保真性（Fidelity）
- 稳健性（Robustness）
  - 鲁棒性、健壮性
- 安全性(Security)
  - 统计不可检测性(Statistical Undetectability)

隐写算法性能指标

- **Capacity**
  - This refers **to the number of bits** of information that are embedded in the host signal.
  - The payload is often normalized by the number of samples of the host signal, resulting in a bit rate R per sample of the host.

- 容量
  - 隐写算法容量指，算法在载体对象中能嵌入的消息总数。
  - 容量也常用平均每样点能嵌入的信息比特来衡量算法容量，即，载体数据利用率=嵌入消息总比特数/样点总数，单位为比特每样点(bits per sample, bps)。对于图像，样点即像素，单位为bpp(bits per pixel)。

隐写算法性能指标

- **案例：容量分析**
  - 512*512规格的图像，使用LSB（只替换最低比特），那么其容量为？
    - 解：
    - 512*512的图像有256k个像素，而每个像素能隐藏1比特消息，所以，容量为：256k比特。
    - 因为每个像素能隐藏1比特消息，容量（载体数据利用率）也可以表示为1bpp。

隐写算法性能指标

- **案例：容量分析**
  - 例2. 已知MLSB是一种LSB改进算法，每个像素的低M比特都用于隐藏秘密消息，那么，当M选为3时，算法的载体数据利用率为？
    - 解：
    - 设载体有N个像素，由算法原理可知，可隐藏3×N比特消息，则：

    载体数据利用率=秘密信息总比特数/载体样点总数

    =3*N/N=3bpp

隐写算法性能指标

- **案例：容量分析**
  - 例3. 已知某算法在变换域隐藏信息。变换域有N个系数，其中有L个非零系数。若算法能在这些系数中隐藏M比特隐藏秘密消息，那么，算法的载体数据利用率为？
    - 解：

载体数据利用率=秘密信息总比特数/载体样点总数

=M/N(bpc)　(bits per coefficient)

或=M/(L) (bpnc)　(bits per　non-zero coefficient)

隐写算法性能指标

指标分类	约定	定义	单位
通用	m比特信息藏入n个样点	c=m/n (bps)	bits per sample
音频（时域）	m比特信息藏入n秒音频	c=m/n (bps)	bits per second
图像（空域）	m比特信息藏入n个像素	c=m/n (bpp)	bits per pixel
图像（变换域）	m比特信息藏入n个系数	c=m/n (bpc)	bits per coefficient
	m比特信息藏入n个非零系数	c=m/n (bpnc)	bits per non-zero coefficient
	m比特信息藏入n个非零交流系数	c=m/n (bpnzAC)	bits per non-zero AC coefficient
通用	m字节消息文件藏入n字节载体文件	r=m/n	%

隐写算法性能指标

- **Transparency (Fidelity, Imperceptibility)**
  - In most applications, embedding of information **should not cause perceptual degradation** of the host signal.
- 透明性（保真性，不可感知性）
  - 透明性指算法对载体对象感官质量的影响程度。
  - 通常，算法引入的失真应该是不可感知的。

隐写算法性能指标

- 透明性 (保真性，不可感知性)
  - 实例，以峰值信噪比衡量透明性，

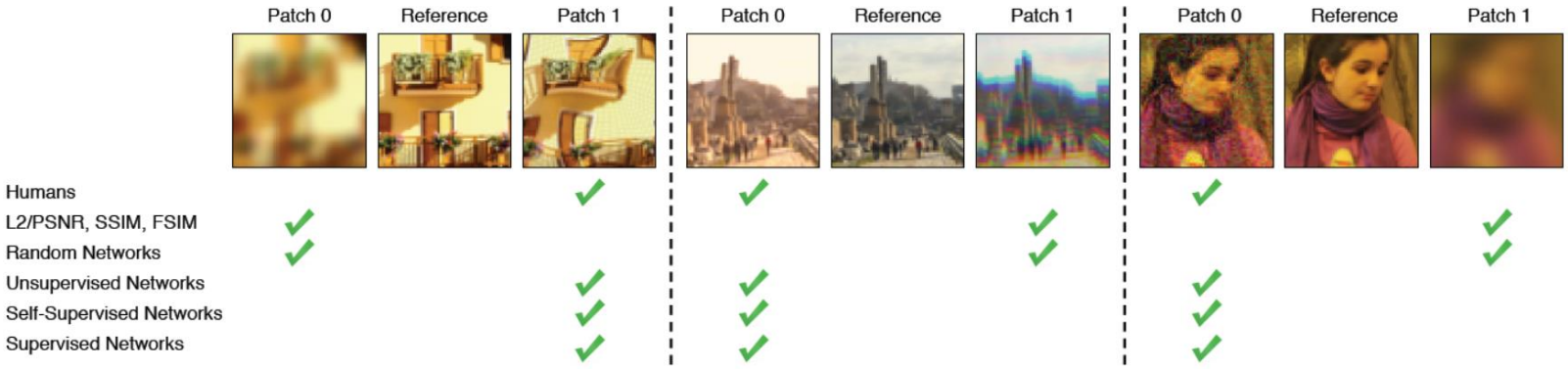
$$PSNR = XY \max_{x,y} \frac{p_{x,y}^2}{\sum_{x,y}\left(p_{x,y} - \tilde{p}_{x,y}\right)^2}$$

（替换最低比特平面）PSNR=51dB

（替换第7比特平面）PSNR=14dB

隐写算法性能指标

- 透明性 (保真性，不可感知性)



可学习感知图像块相似度**(Learned Perceptual Image Patch Similarity, LPIPS)**

1. 也称感知损失**.**

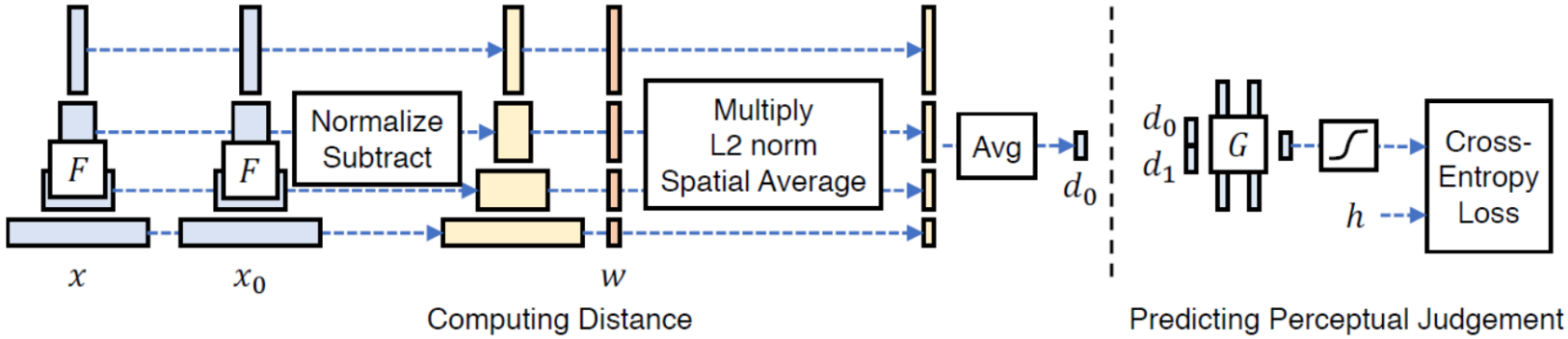
2. 使用**SSIM**和**PSNR**等客观指标评价图像质量时，其结果可能与人的感知存在偏差。从机器学习的角度，**PSNR**和**SSIM**是浅层函数，难以解释人类的感知。

3. 研究表明，用**VGG**为代表的卷积神经网络来度量感知损失可以得到高质量合成图像。受此启发设计感知度量函数。

隐写算法性能指标

- 透明性 (保真性，不可感知性)

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw})||_2^2$$



可学习感知图像块相似度**(Learned Perceptual Image Patch Similarity, LPIPS)**

1. 网络的训练包括两个部份，首先使用参数固定的深度网络$\mathcal{F}$（例如**VGG**）计算图像块特征距离然后训练一个简单网络$\mathcal{G}$来预测感知评分。

2. 特征距离计算公式如上。计算图像块x和x_0对于深度网络$\mathcal{F}$的l层特征，以通道为粒度对特征进行规范化得到$\hat{y}^l_{hw}$和$\hat{y}^l_{0hw}$，再用参数ω_l对其缩放，计算L_2范数后逐点逐层累加得到两者特征距离。

隐写算法性能指标

- **Robustness**
  - This refers to the ability of the embedding algorithm to survive common signal processing operations。
- 稳健性（鲁棒性，健壮性）
  - 稳健性指算法抵抗常规信号处理操作的能力。
  - 隐写对象经滤波等操作处理后，会产生失真。如果信息提取算法仍能从这样的载体中提取消息，那么就称算法对滤波等操作稳健。

隐写算法性能指标

- **稳健性（鲁棒性、健壮性）** 略过深入讨论
  - **案例：LSB算法对噪声添加处理不稳健。**

隐写算法性能指标（深入讨论）

- 稳健性（鲁棒性，健壮性）robustness
- **定义：**设Σ是一个隐写系统，P是一类映射：C → C，若对所有的p ∈ P，
- 对私钥隐写系统，恒有：

$$D_K(p(E_K(c, m, k)),k)= D_K(E_K(c, m, k), k) = m$$

- 对无密钥隐写系统，恒有：

$$D(p(E(c, m)))= D(E(c, m)) = m$$

且不管如何选择：m ∈M，c ∈C，k ∈K，则称该系统为P-健壮性的隐写系统

隐写算法性能指标（深入讨论）

- 理想的隐写系统应该对所有的“**保持α－相似性**”的映射具有健壮性
  - 映射$p: C \rightarrow C$具有性质$sim(c, p(c)) \geqslant \alpha$ 且 $\alpha \approx 1$
- 一般情况下，只能针对某一类特殊的映射具有健壮性
  - 如JPEG压缩与解压缩、滤波、加入白噪声等

隐写算法性能指标

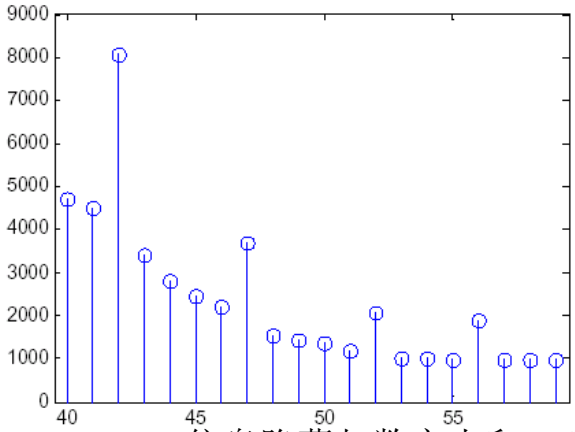
- Security（安全性）
  - 隐写算法主要用于保密通信，因此，仅确保对载体的改动不可感知还不够，还需要同时保持载体的统计特性尽可能不变。
  - 隐写算法对载体的统计特性的影响程度称为安全性。算法安全性越高，其对载体统计特性的影响程度越小。
  - 不可检测性与不可感知性
    - 评价感官质量变化不可检测（感知）程度的指标是算法透明性。
    - 评价载体统计特性变化不可检测程度的指标是算法安全性。

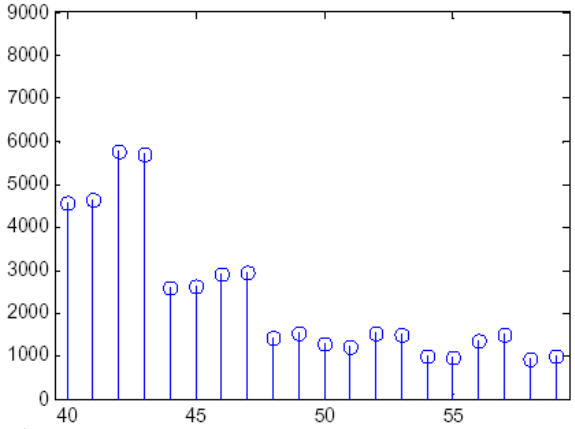
隐写算法性能指标

略过深入讨论

- Security（安全性）
  - 实例

LSB算法透明性虽然好，但安全性不足。右图显示，LSB隐写后，图像的直方图发生明显变化。灰度值为2n和2n+1的像素成对出现。





隐写算法性能指标（深入讨论）

- 熵理论和安全性度量
  - 定义在集合Q上的两个分布P_1和P_2之间的**KL散度**（KL距离或者相对熵）为

$$D(P_1 \| P_2) = \sum_{q \in Q} P_1(q) \log_2 \frac{P_1(q)}{P_2(q)}$$

  - KL散度可视作概率分布P_1与P_2之间差异的度量。
  - 当且仅当P_1与P_2完全相同时，熵D为零，说两概率分布之间没有不确定性
  - 当P_1与P_2不同时，D给出了两分布之间不确定性的衡量，P_1和P_2之间差别越大，熵越大

隐写算法性能指标（深入讨论）

- 熵理论和安全性度量
  - 例，已知投正常和问题骰子时，点数及对应概率如下所示，请计算这两个分布的KL散度。

	1	2	3	4	5	6
P_1	1/6	1/6	1/6	1/6	1/6	1/6
P_2	7/24	6/24	5/24	3/24	2/24	1/24

$$\text{解}: D(P_1||P_2) = \sum_{q\in Q} P_1(q)\log_2\frac{P_1(q)}{P_2(q)}$$

$$= P_1(1)\log_2\frac{P_1(1)}{P_2(1)} + P_1(2)\log_2\frac{P_1(2)}{P_2(2)} + \cdots + P_1(6)\log_2\frac{P_1(6)}{P_2(6)}$$

$$= \frac{1}{6}\left(\log_2\frac{1/6}{7/24} + \log_2\frac{1/6}{6/24} + \cdots + \log_2\frac{1/6}{1/24}\right) \approx 0.284$$

隐写算法性能指标（深入讨论）

- 根据KL散度定义隐写算法的安全性
  - 绝对安全性
  - **定义：**设Σ是一个隐写系统，P_S是隐写对象的概率分布，P_C 是载体对象的概率分布
    - 若有：$D(P_C \| P_S) \leqq \varepsilon$，则称Σ抵御被动攻击是ε-安全的。
    - 若有：$\varepsilon = 0$，则称Σ是绝对安全的
  - 如果一个隐写系统嵌入一个秘密消息到载体中去的过程不改变*C*的概率分布，则该系统是（理论上）绝对安全的

隐写算法性能指标（深入讨论）

- **定理：**存在绝对安全的隐写系统
  - 构造性证明：
    - 设C是所有长度为n的比特串的集合，P_C 是C上的均匀分布，e是秘密消息（$e \in C$）
    - 发送者随机选择一个载体$c \in C$，产生隐写对象$s = c \oplus e$ ，s在C上也是均匀分布的，因此$P_C = P_S$，并且$D(P_C \| P_S) = 0$

隐写算法性能指标（深入讨论）

- 被动攻击者：判断是否有隐藏
- 定义一个检验函数 $f: C \to \{0, 1\}$

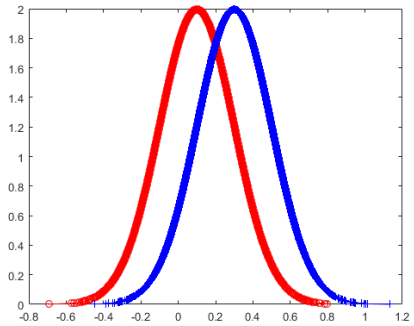
$$f(c) = \begin{cases} 1 & c\text{中含有秘密消息} \\ 0 & \text{其它} \end{cases}$$

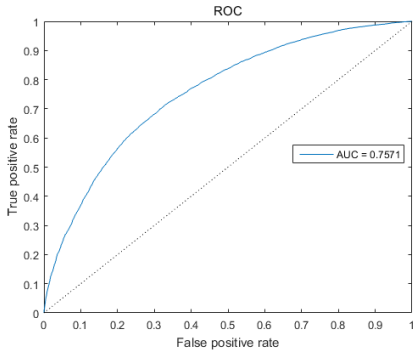
隐写算法性能指标（深入讨论）

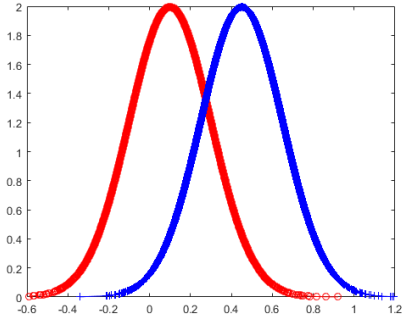
- 判断结果
  - 实际有隐藏，判断有隐藏——正确
  - 实际无隐藏，判断无隐藏——正确
  - 实际无隐藏，判断有隐藏——错误
    - 虚警（ false positive，误判，纳伪错误）
  - 实际有隐藏，判断无隐藏——错误
    - 漏检（false negative，漏判，弃真错误）

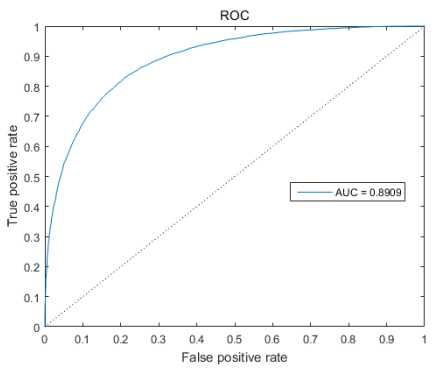
隐写算法性能指标（深入讨论）

ROC（Receive Operating Characteristic，接收者操作特性）定义为灵敏性（Sensitivity）相对于虚警率（1-Specificity）的函数。ROC曲线能更好地描述被动攻击（检测器）的能力。下图示意不同检测器提取的正负类特征分布和对应ROC









判决 样本	自然	隐写
自然	TN	FP
隐写	FN	TP

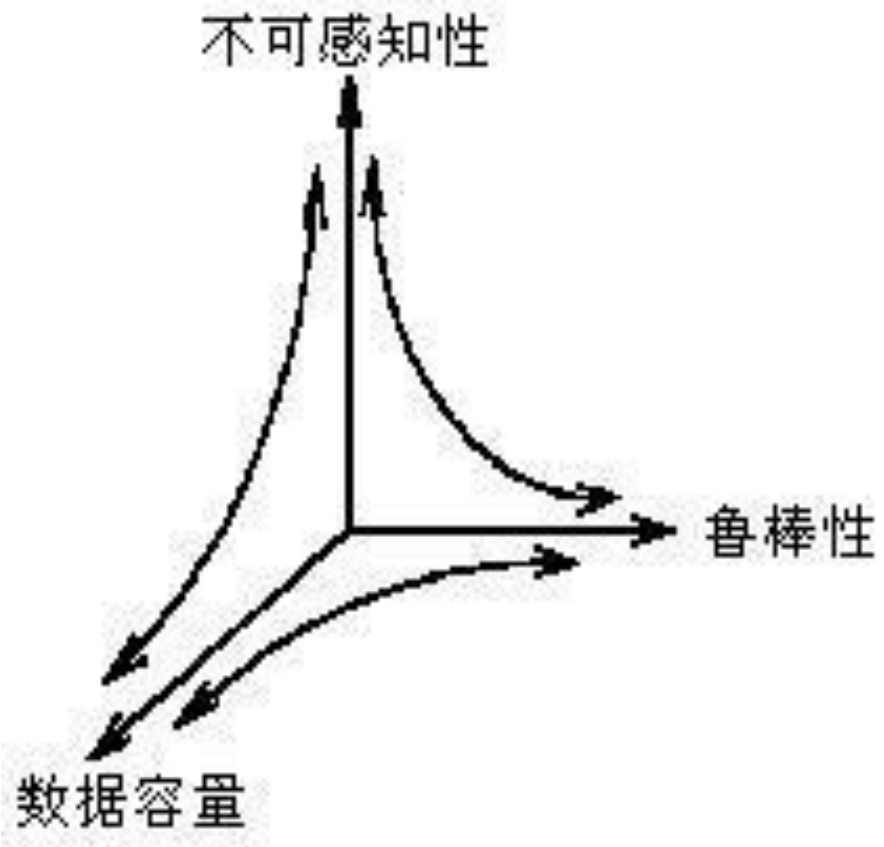
Precesion=TP/(TP+FP)	精度:检出的隐写样本的准确程度
Specificity=TN/(TN+FP)	特异性:正确检出的自然样本的能力
Sensitivity=TP/(TP+FN)	灵敏性:正确检出的隐写样本的能力
Accuracy=(TP+TN)/(TP+TN+FP+FN)	准确率:正确判别样本的能力
FPR=1-Specificity=FP/(TN+FP)	虚警率

隐写算法性能指标（深入讨论）

- 实用的隐写系统
  - 一个抵御被动攻击是 ε -安全隐写系统
  - 攻击者犯虚警错误的概率为 α
  - 攻击者犯漏检错误的概率为 β
  - 一个实用的隐写系统应该尽可能使 β 最大
  - 一个理想的隐写系统应该有 β =1
  - 即，所有藏有信息的载体都被认为没有隐藏信息而被放过，达到了隐写术、迷惑攻击者的目的

隐写算法性能指标

- 隐写算法三个最关键的指标：
  - 不可感知性
  - 鲁棒性
  - 容量



提纲

- 隐写系统
- 隐写系统分类
- 隐写算法性能指标
- ***隐写系统的攻击方法***

隐写系统的攻击方法

- 被动攻击
  - 监视和破译隐藏的秘密信息
- 主动攻击
  - 破坏隐藏的秘密信息
  - 篡改秘密信息
- 非恶意修改
  - 压缩编码，信号处理技术，格式转换，等

隐写术的应用

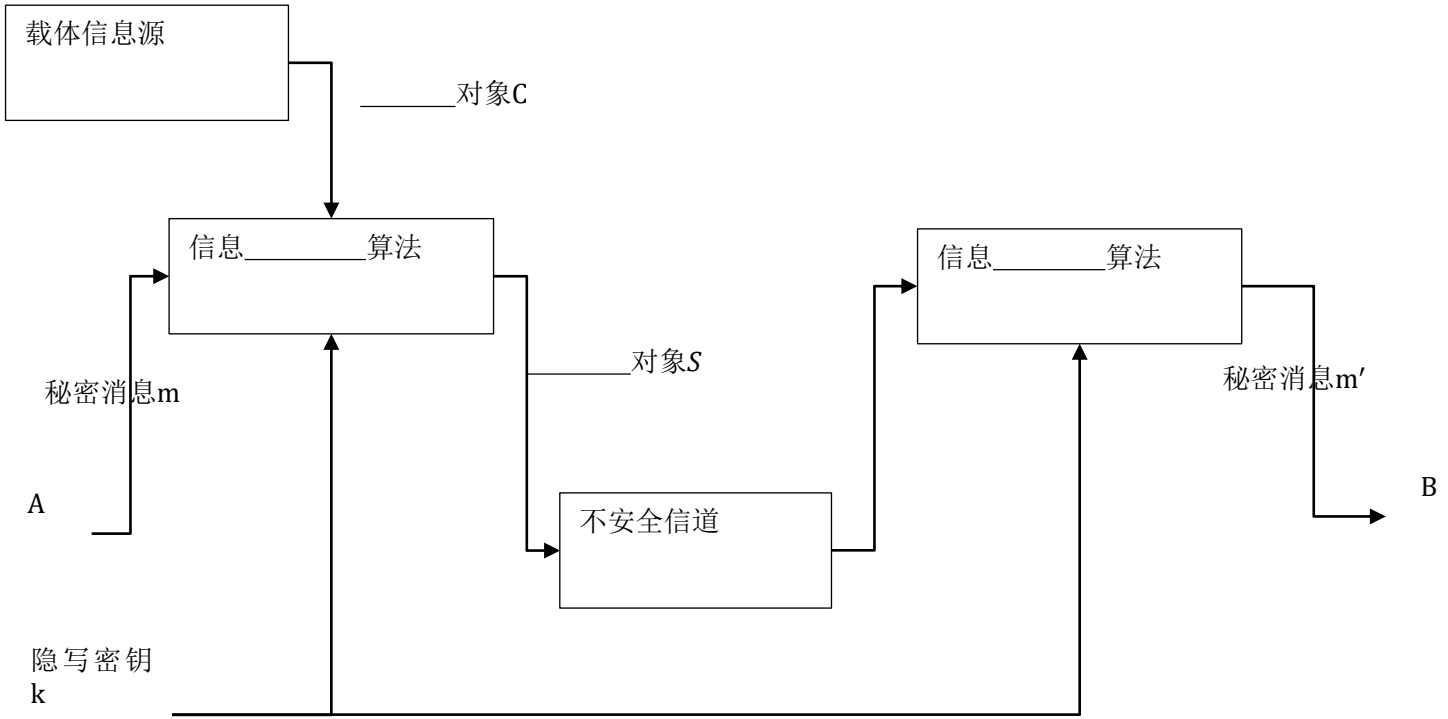
- 军事和情报部门
  - 现代化战争的胜负，越来越取决于对信息的掌握和控制权.
  - 军事通信中通常使用诸如扩展频谱调制或流星散射传输的技术使得信号很难被敌方检测到或破坏掉 .
  - 伪装式隐蔽通信正是可以达到不被敌方检测和破坏的目的.

隐写术的应用

- 需要匿名的场合
  - 包括很多合法的行为，如公平的在线选举、个人隐私的安全传递、保护在线自由发言、使用电子现金等 。
  - 非法的行为，如诽谤、敲诈勒索以及假冒的商业购买行为 。
  - 在隐写术技术的应用中，使用者的伦理道德水平并不是很清楚，所以提供隐写术技术时需要仔细考虑并尽量避免可能的滥用 。

基础练习题

- 1、请在框图中填上隐写系统各个部份名称。



基础练习题

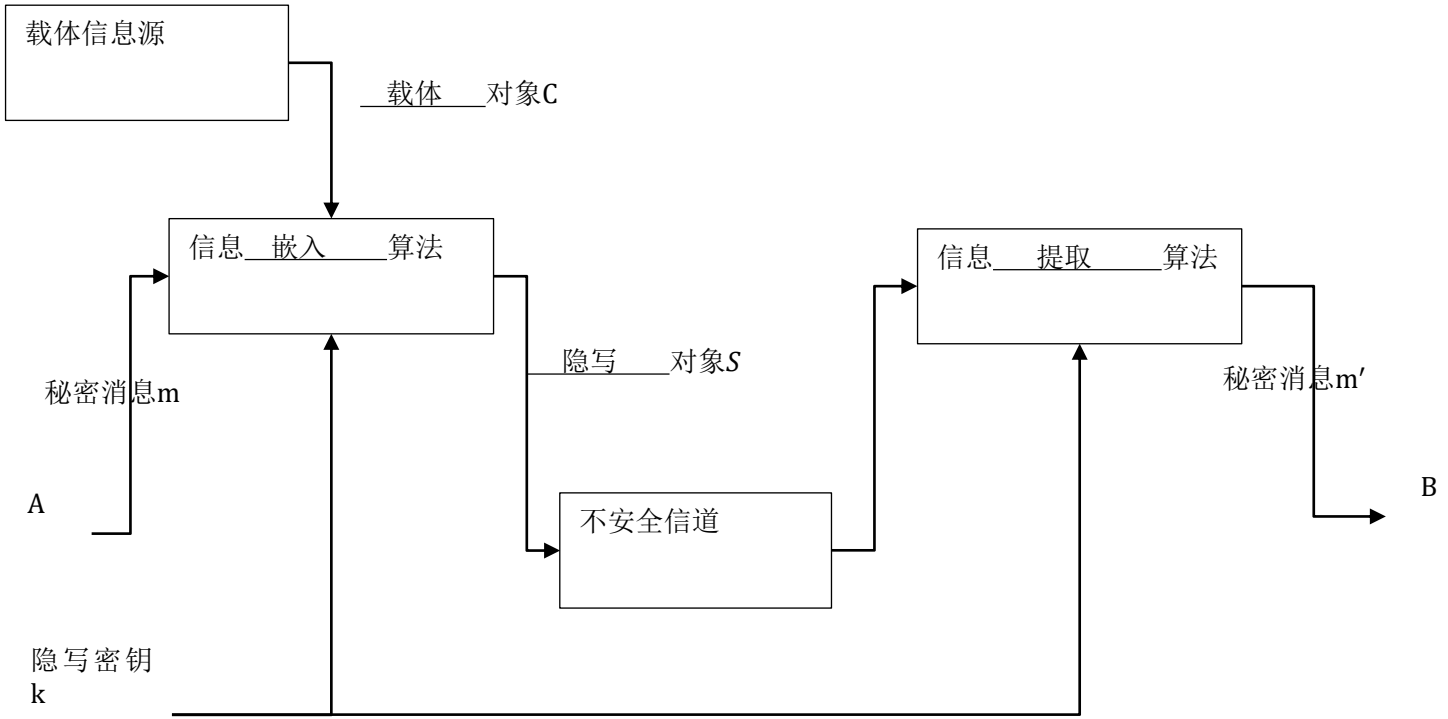
- 2、已知使用算法，在1000个样点中隐藏了100比特消息，请问算法容量是？
- 3、下面指标中，哪个不用于描述算法对载体感官质量的影响程度？（）
- A、不可感知性 B、透明性 C、健壮性 D、不可感知性
- 4、下面指标中，不用于描述隐写算法抵抗常规信号处理操作的能力？（）
- A、安全性 B、稳健性 C、健壮性 D、鲁棒性
- 5、简答：请简介隐写算法安全性，并说明安全性和透明性的区别。

附录

参考解答

基础练习题解答

- 1、请在框图中填上隐写系统各个部份名称。



基础练习题解答

- 2、已知使用算法，在1000个样点中隐藏了100比特消息，请问算法容量是？
  - 解：算法能隐藏100比特消息，因此容量为100bit。隐藏这100比特使用了1000个样点，因此容量也可表示为：
  - 嵌入消息总比特数/样点总数=100/1000=0.1 bps
  - 比特每样点（bit per sample, bps）
- 3、下面指标中，哪个不用于描述算法对载体感官质量的影响程度？（）
- A、不可感知性 B、透明性 C、健壮性 D、保真性
  - 解：C。

基础练习题解答

- 4、下面指标中，不用于描述隐写算法抵抗常规信号处理操作的能力？（）
- A、安全性 B、稳健性 C、健壮性 D、鲁棒性
  - 解：A
- 5、简答：请简介隐写算法安全性，并说明安全性和透明性的区别。
  - 解：隐写算法对载体的统计特征的影响程度称为安全性。算法安全性越高，其对载体统计特征的影响越小。
  - **透明性**度量算法对载体**感官质量**的影响，而**安全性**度量算法对载体**统计特征**的影响。