信息隐藏技术

第一章 概论

1. 国际上首次正式提出信息隐形性研究是在 **1992年** 国际上第一届信息隐藏研讨会学术会议于 **1996年** 在 **剑桥大学** 举行 中国于 **1999年** 召开了第一次全国信息隐藏学术研讨会
2. 隐蔽信息的方法可以分成两个技术路线：一条从 **古典密码术** 发展到 **现代密码学** 一条从 **古典隐写术** 发展到现在的 **伪装式信息安全 信息隐藏** 和 **数字水印**
3. 信息隐藏的原理是利用 **载体** 中的 **冗余信息** 来隐藏秘密信息 信息隐藏的载体包括：图片 视频 音频 文本 协议 其他数据等
4. **技术性隐写技术**包括：文头送信 书记板 实物隐藏 标记法 微缩胶片法 化学方法的隐写术 艺术作品的隐写术 回声法 嵌入隐匿标记
5. **语言学隐写术**：藏头藏尾诗，首字母（次字母），卡登格子，乐谱
6. **保护版权隐写术**：核对校验图，纸张中的水印，纸币中的水印
7. **信息隐藏的基本特征**：**误码不扩散 隐藏信息和载体物理上不可分割 核心思想是使秘密信息不可见 让隐藏载体更加"自然"**
8. Q:**信息隐藏和密码学有什么区别?**

A:（简记）：密码学：**乱码 签名和消息分离 雪崩效应**；信息隐藏：**自然载体 不易隔离 影响范围小**

（详细解答）：密码学的主要思路是使秘密信息 **不可懂** 秘密信息经过加密后变成乱码 很容易引起攻击者怀疑。并且密码学产生的签名和秘密信息分别存储在 **不同的数据结构** 中，物理上 **可以剥离** 攻击者甚至不需要改写信息 只要删除签名就能使接受者无法使用没有被篡改过的秘密信息 密码学加密的秘密信息哪怕错1bit 其他信息都无法恢复
信息隐藏的主要思路是使秘密信息 **不可见** 携带秘密信息的载体与普通载体相似不会引起攻击者的怀疑。秘密信息是掩蔽载体的 **一部分** 在保证载体使用价值的情况下 难以去除秘密信息，部分区域中的秘密信息无法正常提取不会影响到其他区域秘密信息的提取

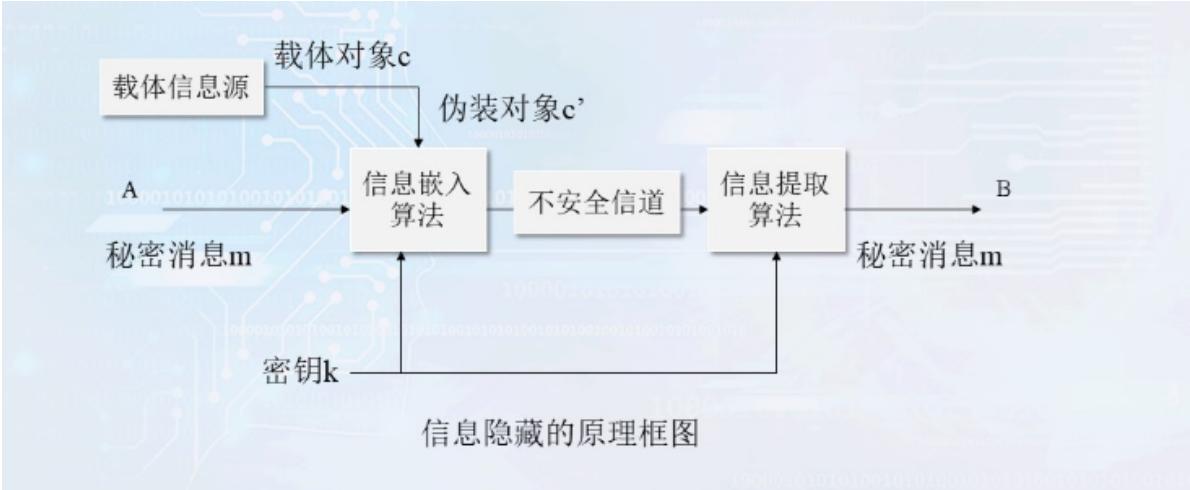
9. 信息隐藏的研究也分为三个层次，分别是基础理论研究、应用基础研究和应用技术研究，简述每个研究层次的研究内容。
   Q:基础理论研究主要针对感知理论、信息隐藏及其数字水印模型、理论框架和安全性理论等;应用基础研究的主要针对图像、声音、水印等载体，研究相应的数字水印隐藏算法和检测算法;应用技术研究以实用化为主要目的，研究各种多媒体格式的信息隐藏和数字水印技术在实际中的应用。

第二章 基础知识

1. 语音的质量一般从两个方面来衡量 分别是 **清晰度** 和 **自然度** 前者衡量语音中字，单词和句子的 **清晰程度** 后者是衡量通过语音识别人讲话的 **难易程度**
2. 语音的数字模型是一个 **缓慢时变** 的线性结构 在 10-20ms内 是 **近似不变的**
3. 语音的质量评价包括 **主观评价（人）** 和 **客观评价(机器)** 两个部分
4. 主观评价的基本方法有： **平均意见分（MOS）** 4分以上为高质量 3.5分左右直来那个下降但不妨碍正常通话 3分一下为合成语音 优点：**真实 比较准确** 缺点：**费时费力 不够灵活 重复性和稳定性较低 且受受试者的主观影响较大**
5. 客观评价方法 对输入和输出的信号进行分析和处理 设计失真距离 由此值来作为语音客观评价值 基本方法有：PSNR（峰值信噪比）MSE SSIM LPC （PSNR值越大 失真越小）优点：**使用方便 可重复性强** 缺点：**无法达到与主观评价一样的效果。**
6. 分辨力是指人眼在距离上分开两点的能力 **速度大分辨力小 彩色分辨力比黑白低 照度过高过低都会使分辨力下降** 暗适应性和光适应性顾名思义 暗慢
7. 人类的听觉系统（HAS）要比人类视觉系统(HVS) **灵敏** 正常人类听觉系统的频率在0.016-16Khz（随年龄变化）

8. 常用的图像处理方法包括 **二维离散傅里叶变换（2DFT） 二位离散余弦变换（2DCT） 二位离散小波变换（DWT2）**
9. 二位离散小波变换处理图像，一级分解后的图像包括四个部分 **近似部分(LL) 水平方向细节部分(HL) 垂直方向细节部分（LH） 对角线方向细节部分(HH)**
10. 语音编码的分类包括：**波形编码 参数编码** 波形编码：语音质量好 编码速率比较高（64-16kb/s） 参数编码：编码速率低 2.4kb/s甚至更低 能听懂语音 但自然度低
11. 常见的波形编码包括：脉冲编码调制PCM，自适应差分编码ADPCM，自适应子带编码ASBC，自适应增量调制ADM，自适应预测编码APC，自适应变换编码ATC (看懂英文会翻译就行)
12. 常用的语音处理算法包括：傅里叶变换与短时傅里叶变换 小波变换 离散余弦变换
13. **低频信号变化缓慢，高频信号变化迅速** 小波变换在低频部分频率分辨率较高 高频部分时间分辨率较高 **小波分析法**：窗口大小固定但形状可变的时--频局部化分析方法
14. 按照噪声的性质分类，信息隐藏通信模型可分为： **加性噪声信道模型** 和 **非加性噪声信道模型**

第三章 信息隐藏的基本原理



（这个图可能会考 需要记住）
信息隐藏的分类:**无密钥信息隐藏 公钥信息隐藏 私钥信息隐藏**

1. Q:什么是无密钥信息隐藏?

A:如果一个信息隐藏系统不需要预先的密钥，称为元密钥信息隐藏系统。在数学上信息隐藏过程可以表示为一个映射E:CXM -> C' 这里C标识可能的载体集合 M表示所有可能秘密的信息集合 C' 表示所有伪装对象的集合。信息提取的过程也是一个映射D:C'-> M 发送方和接收方需要事先预定好嵌入算法和提取算法，但这些算法都是要求保密的。

2. 公钥信息隐藏无法抵抗中间插入攻击 类似公钥密码
3. 判断是否有信息隐藏所犯的两种错误 弃真 纳伪 选择题注意鉴别
4. 信息隐藏的攻击
   被动攻击：监听破译秘密信息
   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
5. 鲁棒性（又称健壮性）安全性高 健壮性差；将隐藏信息与载体感官最重要的部分绑定在一起就会提高鲁棒性

第四章 音频信息隐藏

1. 因为人对随机噪声不敏感 由此产生了替换技术 替换技术就是试图用 **秘密信息比特** 替换掉 **随机噪声** 以达到信息隐藏的目的
2. 音频信号属于 **一维信号** 并且人类的听觉系统比视觉系统灵敏得多 对音频信号信息隐藏的要求：**透明性 鲁棒性 同步 盲检测**

3. Q:信息隐藏的透明性（不可感知性）表示的大致含义是什么?
   A:不可感知性：第一是指隐藏的秘密信息不对载体的视觉或听觉上产生影响 隐藏的信息附加在某种数字载体上，必须保证其存在不妨碍和破坏数字载体的欣赏价值和使用价值。第二是要求采用统计的方法也不能恢复隐藏的信息。

   Q:信息隐藏的三个指标及其含义：
   A:**透明性**：算法对载体感官质量造成的影响，算法应该不显著影响载体感官质量

   容量：是指在载体中能够嵌入秘密信息的总量，通常是将之除以样本总数得到样本平均嵌入量

   鲁棒性：指样本抵抗普通信号处理操作的能力
4. 数字化音频中，低有效比特对音质贡献弱，因此可采用LSB隐藏方法，此外还有回声隐藏法
5. LSB的参数包括样点和位置的选取 LSB算法：透明度高 容量大 鲁棒性差
6. 掩蔽效应可以被分为 **频域掩蔽** 和 **时域掩蔽** 或者也可以被分为 **同时掩蔽** 和 **异时掩蔽** 后者可以被分为 **超前掩蔽** 和 **滞后掩蔽**
7. Q:什么是 **回声隐藏法**?
   A:回声信息隐藏是利用人力类听觉系统中强信号在时域内的 **向后掩蔽效应**，即弱信号会在强信号消失后变得无法听见。弱信号可以在强信号消失后的50~200ms不被人耳察觉。音频信号和经过回声隐藏的秘密信息对于人耳朵来说，前者像是从耳机中听到的，后者像是从扬声器中听到的
8. 回声的数字音频信号可表示为：y[n]=s[n]+a*[n-t] 其中a是回声的 **幅度系数**，t是回声的 **时延参数。**
9. MP3编码的流程包括：**时频映射 心理学模型 量化编码 帧数据流格式优化**
10. 根据MP3隐藏算法的嵌入时间可以将其分为：**压缩编码前嵌入 压缩编码中嵌入 压缩编码后嵌入**
11. MIDI **乐器数字接口** 由 **头块** 和 **音轨块** 组成
12. MIDI消息可以被分为：**通道消息** 和 **音轨消息** 通道消息又被分为 **声音消息** 和 **模式消息**
13. 改变 **声音开启的最低比特 乐器编号的最低比特** 以及 **通道触压力的低4比特** 都不会引起听觉差异
14. Q:什么是音频文件的 **相位隐藏法**
    A:相位编码是利用人类听觉系统对声音的绝对相位不敏感，而对 **相对相位敏感** 的特性进行数字水印嵌入的。在相位编码中，载体信号首先被分成若干个短序列，然后进行DFT变换，修改所有信号片段中的 **绝对相位** 同时保证它们的 **相对相位差** 不变，然后通过 **IDFT** 得到伪装信号；在回复秘密信息之前，必须采用同步技术，找到信号的分段。已知序号哦的长度，接收者就能计算DFT，并检测出相位。该算法对窄体信息的再取样有鲁棒性，但对 **大多数音频压缩算法敏感** 由于仅在第一个信号片段进行编码 **数据传输率很低**

第五章 图像信息隐藏

1. 图像位平面中 **高位是图像信息 低位是随机噪声** （对于256级的灰度图像，第一个（连同）第二个位平面对图像能量贡献较小，可用于秘密信息隐藏）
2. 图像的LSB方法 优点：简单 易实现 容量大 缺点：安全性不高 不能抵抗叠加噪声 有损压缩等破坏
3. 提高LSB安全性的措施：**对秘密信息先加密后隐藏 多次重复嵌入 引入纠错编码技术，先进行纠错编码在进行隐藏**
4. 图像的奇偶校验和方法：如果奇偶校验位和mi不匹配，则翻转该区域中所有像素的最低比特位
5. 调色板图像的信息隐藏包括：**修改调色板颜色向量LSB** 灰度图像不明显，RGB图像可能会产生偏差； **修改图像索引LSB** 相邻索引值代表的颜色差别很大 解决方法：可以先将颜色按照某种顺序进行排序，使得相邻的颜色比较接近 **调色板的排序方式进行编码**:不具有鲁棒性
6. **基于量化编码的信息隐藏** 优点：巧妙 缺点：稳健性不强

7. 二值图像的信息隐藏 zhao-koch方案：稳健性参数,该参数越大，鲁棒性越强，注意还有标识无效块 利用游程编码
   Q: 已知某图像轮廓的游程编码为<a0， 3><a1， 4><a2，4><a3，7>。现需修改游程长度以隐藏秘密信息，约定隐藏0时游程长度为偶数(约定长度在2i和2i+1之间翻转，例如2-3，4-5，...)，则隐藏秘密信息1100后，游程编码变为( )。
   A.<a0，3><a1，5><a2+1，2><a3 -1，8>
   B. <a0，3><a1，5><a2，2><a3，8>
   C.<a0，5><a1 +2，5><a2+2，4><a3+2，8>
   D. <a0，5><a1 +2，3><a2+1，4><a3+1，8>
   答案：C
8. 图像隐藏的变换域技术： DCT DWT DFT，在DCT域中的信息隐藏，能够有效抵抗 **JPEG有损压缩**
9. 进行二维DCT变换，左上角的那个系数为直流和低频系数，右下角部分为高频系数，中间部分为中频部分 **中低频系数**包括图像大部分能量，一般选择 **中频系数** 隐藏信息
   Q:现接收到一使用DCT系数相对关系(隐藏1时，令B(u1,v1)>B(u3,v3)+D，且， B (u2， v2) > B (u3，V3)+D)隐藏秘密信息的图像，已知D=0.5，对该图像作DCT变换后， 得到约定位置((u1，v1)u2,v2) (u3,v3))的系数值为(1. 6，2.1，1. 0)，(0. 7，1. 2，1.8)，(0.9，1.8，1.2)，则可从中提取的秘密信息是( )。
   A.0，1，1 B. 1，0，0 C. 1，0，无效 D. 0，1，无效
   答案：C
10. 一级小波分解得到四个部分：**左上低频近似部分 右上水平方向细节部分** 左下：**垂直方向细节部分** 右下：**对角线方向细节部分** 图像能量主要集中在低频近似部分 修改系数法 系数比较法
11. 文件格式隐藏法：bmp文件头 位图信息头 调色板 位图数据 保存文档可能造成隐藏数据的丢失 因此不安全
12. Q:简单描述BMP图像格式位图文件的两个有效的数据结构之间的信息隐藏方法;
    A：BMP图像由BMP文件头 位图信息头 调色板 和位图数据组成，可以在BMP调色板与位图数据之间隐藏秘密
13. 统计隐藏技术：对某些统计特性进行明显的修改，表现嵌入信息1，在不知道原始载体的情况下，根据统计行的改编，提取信息 （**伪装密钥法**）
14. 变形技术：对载体进行某种修改，修改方式与需要嵌入的秘密信息比特相关联。通过比较修改后的载体和原始载体的差别提取信息 （对载体的修改不易察觉）
15. Q:请你简述行移位编码 字移位编码 和特征编码
    A:行移位编码就是在文本的每一页中，每隔一行轮流地嵌入水印信息，但嵌入信息的行的相邻上下两行的位置不动。作为参考，需嵌入信息的行根据水印信息的比特流进行轻微上移或下移。在移动过的一行编码一个比特信息，如果上移编码为1，下移编码为0

    字移位编码通过将文本中的某一行单词进行水平移位，将某一个单词进行左移和右移，而与其相邻单词并不移动，这些单词作为解码的参考位置

    特征编码是通过改变文档中某个字母的特征来嵌入标记。在这种编码中，水印信息作为可见的噪声叠加到字母笔画的边缘或文本中图像的边界上，对噪声图像进行而知编码，从而达到水印嵌入的目的。

第六章 数字水印

1. **数字水印** 是永久镶嵌在其他数据（宿主数据）中具有 **可鉴别性**的 **数字信号** 或 **模式**，并且不影响宿主数据的 **可用性**
2. 数字水印应具有 **安全性 可证明性 不可感知性 健壮性** 等特点
3. 数字水印的三要素包括：**水印本身的结构 水印的加载过程 水印的检测过程**
4. 第一类错误，实际不存在水印却检测到水印，**虚警率** 第二类错误，实际有水印却没有检测到水印 **漏检率**

5. Q:信息隐藏（隐写术）和数字水印的区别
   A:**用途**：信息隐藏用于保密通信 数字水印用于版权标识

   前提：一般不知有信息隐藏 可以公布水印存在

   精确恢复：信息隐藏可精确恢复 数字水印不可精确恢复

   主要攻击：信息隐藏为隐写分析 数字水印为水印擦除
6. 数字水印的分类：
   - 从载体上分类：图像水印 视频水印 音频水印 **软件水印** （又可以被分为 **静态水印**和 **动态水印**） 文档水印
   - 从外观上分类：**可见水印**（如电视台标） **不可见水印**
   - 加载方式上：空间域水印（LSB，拼凑） 变换域水印（DCT变换等）
   - 检测方法上：**私有水印**（非盲水印 需要原始载体）**公开水印**（盲水印，无需公开载体）或者是 私钥水印（加载检测使用同一密钥）公钥水印（加载检测使用不同密钥）
   - 水印特性：**健壮性数字水印**（水印能够经受各种常用的操作，包括恶意或无意的处理） **脆弱性数字水印** （用于完整性保护，对载体变化很敏感，经过微小处理后水印就会被改变或毁掉）（**半脆弱水印技术主要用于内容篡改检测**，因为对半脆弱水印图像进行普通信号处理。例如，JPEG压缩、去噪等，不会影响水印的提取，但对图像内容的篡改将导致水印信息丢失。）
   - 使用目的：版权标识水印 数字指纹水印
     （会识别就可以 不用背）
7. Q:请你简述半脆弱和脆弱数字水印的区别
   A:脆弱数字水印对各种图像处理信号都很敏感，当载体数据发生微小变化时，水印信息就丢失了。有些场合要求图像内容没有变化就应该能够检测水印，因此就产生了半脆弱水印算法，这种算法能抵抗普通的信号操作（如压缩去噪），但是对内容篡改敏感

第七章 数字水印技术

1. 数字水印在表现上可分为几大类，一类是 **一串有意义的字符** 一类是 **一串伪随机序列** 一类是一个 **可视的图片**
2. 水印的稳健性体现在两个方面意识选择水印是应该考虑水印本身能欧容忍一定的误码；另一方面还要考虑水印的抗攻击能力以及检测方式。
3. 水印嵌入的位置需要考虑 **安全性问题**（嵌入的水印不能被轻易的提取或擦除） **对载体质量影响的问题**（嵌入的水印不能影响数字载体的使用，引起的失真应该对人眼不可察觉）
4. **脆弱性水印** 是指在保证多媒体信息一定感知质量的前提下，将数字 序列号 文字 图像 标志 等作为数字水印嵌入到多媒体数据中 当多媒体内容受到怀疑时 可将改水印提取出来 用于多媒体内容的真伪识别 并指出篡改的位置 甚至攻击类型等 具有特征：**检测篡改 稳健性和脆弱性 不可感知性 可靠性**
5. 脆弱性水印可分为 **完全脆弱水印 半脆弱水印**（允许一定的改变） **图像内容可鉴别**（更稳健） **自嵌入水印**

第八章 信息隐藏分析

1. 隐写分析分类：
   - 适用性分类：专用隐写分析和通用隐写分析
   - 根据已知消息分类：唯隐文攻击 已知载体攻击 已知消息攻击 选择隐文攻击 选择消息攻击 已知隐文攻击
   - 采用的分析方法进行分类：感官分析 统计分析 特征分析
   - 按照最终的效果进行分析：被动隐写分析 主动隐写分析
2. Q:主动隐写分析和被动隐写分析的特点：

- 被动隐写分析:判断是否隐写及隐写使用的算法。特点:**判断。**

- 主动隐写分析:判断是否隐写，估计隐藏秘密信息的位置与数量，推算出所使用的密钥，并提取出秘密信息。特点:**通过分析判断并提取信息**

3. Q：隐写分析的目标是什么?
   A:隐写分析技术是对表面正常的图像、音频等载体进行检测，以判断载体中是否隐藏有秘密信息，甚至只是指出媒体中存在秘密信息的可能性。另外，隐写分析还可以对看似可疑的载体实施主动攻击，即删除或者破坏嵌入的秘密信息以达到阻止隐蔽通信的目的。
   目的：**防止隐写术的滥用 寻找隐藏算法的漏洞 促进研制安全性更高的隐藏算法**
4. Q：根据攻击者掌握信息的不同，隐写分析可分为哪五类，请简单介绍
   A：
   - (1)仅知掩蔽载体攻击:分析者仅持有可能有隐藏信息的媒体对象，对可能使用的隐写 算法和隐写内容等均全然不知，是完全的盲分析。
   - (2)已知载体攻击:将不含密的已知原始媒体与分析对象比较，检测其中是否存在差异。
   - (3)已知隐藏消息:分析者知道隐蔽的信息或者它的某种派生形式。
   - (4)可选隐藏对象:在已知对方所用隐写工具和掩蔽载体的基础上提取信息。
   - (5)可选消息:分析者可使用某种隐写工具嵌入选择的消息产生含密对象，以确定其中可能涉及某一隐写工具或算法的相应模式。
5. 隐写分析的层次包括： **发现 提取 破坏**
6. 隐写分析的评价指标包括：**准确性 适用性 实用性 复杂度** 各个性能指标相互制约 准确性高 适用性差；适用性好，准确性差；采用更高阶特征，复杂度高，准确性高，实时性差
7. **隐写率**（用于隐藏信息的样点数/载体样本点数） 和 **嵌入效率**（隐藏的秘密信息总数/载体样点总数）
8. Q:简述嵌入效率和载体数据利用率的含义，嵌入效率高意味着什么？ (从透明度和容量两方面分析。)
   A：嵌入效率(嵌入比特数/平均修改长度)指平均每修改1个样点可以嵌入多少比特秘密信息，载体数据利用率(秘密信息总数/样点总数)指平均每个样点可以隐藏多少比特秘密信息。嵌入效率高意味着同样嵌入量，对图像的修改少，失真小。但与此同时，载体数据利用率下降，隐藏相同的秘密信息需要更多的像素。
9. Q:简述 **卡方分析**、**RS分析** 和 **GPC分析** 的原理。
   答:
   (1)卡方分析原理:LSB隐写会使值对出现次数趋于相等，据此采用大数定理可以构造服从卡方分布统计量，计算待检测图像的该统计量可以判定图像是否经过LSB隐写。
   (2) RS分析原理:对自然图像，非负和非正翻转同等程度地增加图像的混乱程度。而对隐写图像，采用非负翻转后，规则图像块比例和不规则图像块比例的差值随隐写率的增大而减小，采用非正翻转却不会出现上述情况。
   (3) GPC分析原理:GPC分析也利用图像空间相关性进行隐写分析。对于自然图像，N0（图像的三维曲面穿越平面簇z=1.5,3.5，...，255.5的次数）近似等于N1（图像的三维曲面穿越平面簇z=0.5,2.5，...，254.5的次数）;而对于隐写图像，N1与N0的比值随隐写率增大而增加。
10. 现有一幅纹理丰富的待检测图像有可能经过了LSB隐写，则下面说法不正确的是（）。
    A.若秘密信息不是连续隐藏的，则卡方分析可能失效，而RS和GPC分析则不受该因素影响。
    B. 图像纹理丰富时，自然图像的N1和N0很大，LSB隐写引起的变化不明显，因此GPC分析可能失效。
    C.若隐写时使用的不是普通LSB算法，而是预留了部分像素用于平衡由隐写带来的直方图的变化，那么RS分析可能失效。
    D. 若隐写时使用的不是普通LSB算法，像素不是在值对2i和2i+1间翻转，2i可能变为2i-1，2i+1可能变为2i+2，那么GPC分析可能失效。
11. 下列关于改进算法的描述，不正确的是( )。
    A.最小直方图失真隐写算法在尽量保持F1和F-1翻转平衡的情况下，使直方图在隐写前后变化量尽可能小，可以抵抗卡方分析。
    B. 直方图补偿隐写算法确保隐写后，直方图中2i和2i十1的频度不再趋于相等，因此可以抵抗RS

分析。
C. 改进LSB隐写算法翻转像素灰度时，2i不仅可以变为2i+1，也可以变为2i-1。
D.　改进LSB隐写算法可以抵抗卡方、RS和GPC分析。
答案：B

第九章 数字水印的攻击

1. 数字水印的攻击分为：**去除攻击**（有损压缩 信号处理） **表达攻击**（几何攻击 翻转裁剪） **解释攻击**（使同一水印存在多个解释） **法律攻击**（利用法律漏洞）
2. 水印攻击软件 用于评价和测试水印算法的性能 （Unzign,Checkmark,Optimark,StirMark）