基于联合失真的 AAC 安全隐写算法

蔡 森[1], 任延珍[1], 王丽娜[1]

空天信息安全与可信计算教育部重点实验室, 武汉大学国家网络安全学院 武汉 中国 430072

摘要 随着互联网技术的普及, 越来越多的音视频通信应用融入到了人们的日常生活中。AAC(Advanced Audio Coding), 作为目前互联网应用中使用最广泛的音频压缩编码标准之一, 拥有优秀的压缩效果和出色的音频质量, 使得越来越多的音视频作品利用 AAC 进行编码传输, 这也为信息隐藏提供了新的、更多、更好的隐写空间。本文分别对现有的隐写算法生成的含密音频的 Huffman 码字进行统计分析和算法中基于掩蔽曲线理论设计的失真函数进行分析, 发现现有隐写算法会在不同程度上破坏音频的 Huffman 码字的统计安全性和音频的听觉掩蔽性。为了解决现有的隐写算法在统计安全性和听觉隐蔽性方面的不足, 本文提出了基于 Huffman 码字统计分布特性和实际听觉掩蔽阈值曲线相结合的 AAC 隐写联合失真代价函数。同时, 利用最小失真编码框架 STCs(Syndrome-Trellis Codes)实现了面向 AAC 压缩参数域的自适应安全隐写算法。实验结果表明, 本文提出的隐写方法的隐写容量在理论上最大可达 12 kbps。另外, 在统计安全性方面, 相较于现有的隐写方法, 最高可以提升 30%。在听觉质量方面也有进一步的提升。在可扩展性方面, 由于 MP3(Moving Picture Experts Group Audio Layer III)的编码流程和 AAC 的编码流程在大体上类似, 本文所提出的新方案可以经过微小的调整后, 便可直接应用在基于 MP3 的隐写算法中, 对现有基于 MP3 的隐写算法在安全性和音频质量方面也有一定的提高。

关键词 AAC; Huffman 编码; 隐写术; 听觉掩蔽效应

中图法分类号 TP391 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2022.03.01

An AAC Steganographic Algorithm Based on Joint Distortion

CAI Sen[1], REN Yanzhen[1], WANG Lina[1]

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China

Abstract With the popularization of Internet technology, more and more audio and video communication applications have been integrated into People's Daily life. AAC (Advanced Audio Coding), as one of the most widely used Audio compression Coding standards in Internet applications, has excellent compression effect and excellent Audio quality, making more and more Audio and video works use AAC for Coding transmission. This also provides new, more and better steganographic space for information hiding. In this paper, the statistical analysis of Huffman code words with dense audio generated by the existing steganography algorithm and the distortion function designed by the algorithm based on the masking curve theory are carried out respectively, and it is found that the existing steganography algorithm will destroy the statistical security of Huffman code words and audio masking to varying degrees. In order to solve the shortcomings of existing steganography algorithms in statistical security and auditory concealment, the AAC steganographic joint distortion cost function is proposed based on Huffman code word statistical distribution characteristics and actual auditory masking threshold curve. At the same time, an adaptive steganographic algorithm for AAC compression parameter domain is implemented by using the minimum distortion coding framework STCs (Syndrome-Trellis Codes). The experimental results show that the maximum steganographic capacity of the proposed steganographic method can reach 12 kbps. In addition, statistical security can be improved by up to 30% compared to existing steganographic methods. There was also a further improvement in auditory quality. In terms of scalability, because the encoding process of MP3 (Moving Picture Experts Group Audio Layer III) is basically similar to that of AAC, the new scheme proposed in this paper can be directly applied to steganographic algorithm based on MP3 after minor adjustment. It also improves the security and audio quality of the steganographic algorithm based on MP3.

Key words AAC; Huffman Code; steganography; masking effect of the psychoacoustic model

通讯作者: 蔡森, 硕士, Email: limingong@whu.edu.cn。

本课题得到国家自然科学基金(No. 61872275, No. U1836112, No. 61876134)资助。

收稿日期: 2020-12-10; 修改日期: 2021-03-02; 定稿日期: 2022-01-10

1 引言

随着互联网时代的高速发展, 越来越多的多媒体融入了人们的生活。其中, 在互联网中传输的大量音视频作品为信息隐藏提供了新的隐写空间。信息隐藏是信息安全中一个重要分支。隐写技术是一种将秘密信息嵌入到多媒体数据载体中的技术[1-3], 它利用了载体的感知冗余和统计冗余的特点。隐写分析技术是隐写术的一种对抗技术[4]。它在未知或者已知隐写嵌入算法的情况下, 通过对目标载体的部分统计特征和其先验统计特征进行比较等多种检测手段, 判断目标载体是否含有秘密信息、秘密信息的实际嵌入量以及隐写位置分布的情况。高级音频编码(Advanced Audio Code, AAC[5]), 作为主流的音频压缩标准, 和 MP3 在编码流程上有很多相类似的编码模块, 但是以其更低的压缩比和更好的音质, 已逐渐取代 MP3 音频编码, 并广泛地应用到各种音频和视频软件中, 如 QQ、WeChat、Twitter、 Facebook 等。因此, AAC 音频编码中的压缩参数域成为音频信息隐藏领域重要的隐写载体, 并可以在公开网络中实现保密通信。

目前, 针对 AAC 的隐写算法成果较少。由于 AAC 和 MP3 在编码原理上的相似性, 因此, MP3 的隐写思路可对 AAC 隐写有一定的参考和借鉴作用。根据 AAC 和 MP3 的编码原理, 隐写算法的嵌入域划分, 主要分为三种类型: MDCT (Modified Discrete Cosine Transform) 系数域[6-9], 比例因子域[10-11]和 Huffman 编码域[12-16]。这些方法主要通过修改 AAC 的编码参数来嵌入秘密信息。在 Huffman 编码域, Zhu 等[13]提出了一种利用 Huffman 码字的符号位进行嵌入的隐写算法。该方法对 Huffman 码字对应的 MDCT 系数的符号位进行编码, 实现秘密信息的嵌入。在 Yan 等[14]中, 将 Huffman 码表中的码字为了两组, 形成码字映射关系。对应的码字在码字空间中根据密钥进行选择, 通过码字替换实现隐写。以上的 AAC 隐写算法是非自适应隐写算法, 它们都存在一个共同的问题, 即没有或者较少考虑 AAC 编码过程中参数之间的相关性, 会降低统计安全性以及声音质量变化的不可感知性。为了解决这些问题, Yang 等[15]设计一种基于人类听觉的绝对阈值曲线的内容感知失真函数, 利用 STCs[17]实现自适应隐写, 提高了声音质量。Yi 等[16]在 Yang 等[15]的基础上提出了一种基于 Huffam 码字的自适应帧块失真优选隐写框架算法。将 Huffman 码表中的码字分组策略从一一对应改为一组内有多个相互替换码字的策略, 构造了多个码字动态映射的关系, 最后, 利用 STCs[17]框架实现自适应嵌入算法。Yi 等[16]相比 Yang 等[15]提出了多组码字映射的关系, 可以更好地根据实际失真计算选择合适码字进行替换, 提高了统计安全性。同时, 增加了帧间排序嵌入策略, 但不是针对全局进行每个码字地选择。Yang 等[9]提出了一种采用向前向后的联合失真代价隐写算法, 通过计算帧内相邻 QMDCT 系数的相关性, 构造失真函数, 利用 STCs[17]实现嵌入。这三种自适应隐写算法比传统的非自适应隐写算法[6-8,10-14]进一步提升了统计安全性和不可感知性, 但这些方法仅仅考虑了人耳掩蔽绝对阈值曲线的特性或者帧内相关性, 在 Huffman 码字的统计分布特征上未考虑, 降低了 Huffman 码字的统计安全性。

针对隐写所引入载体信号特征的变化, 目前 AAC 和 MP3 隐写分析技术也不断出现。现有的 AAC 隐写分析算法主要分为两类: 一类是通过手工提取特征并使用二分类器构造一个隐写分析方法[18-22]; 另一类是基于深度学习方法构造神经网络提取特征[23-26], 进行二分类器判别。其中, 文献[18-22,24]均是对帧内的 MDCT 或 QMDCT 系数的邻间相关性进行分析, 构造隐写分析的特征。Yu 等[27]是针对 MP3Stego[28]利用边信息实现隐写分析。Lin 等[23]提出一种改进的 CNN 网络的隐写分析方法。Wang 等[24]使用 QMDCT 系数作为输入进行构造 CNN 神经网络的隐写分析。Wang 等[25]提出了一种基于高滤波的全连接 CNN 网络的 MP3 隐写分析方法。

现有的 AAC 隐写算法并不能保持 Huffman 码字的统计分布特性, 导致了含密音频在 Huffman 码字统计分布特征方面的安全性较低。同时, 人耳绝对阈值曲线也无法完全表示编码过程中实际的音频能量变化。基于以上的分析, 本文提出了一种基于联合失真的 AAC 安全隐写算法。

本文工作的主要贡献包括以下 3 个方面:

1) 针对现有隐写算法所引入的 Huffman 码字直方图统计分布异常特征, 本文首次提出基于 Huffman 码字直方图统计特征的隐写失真代价, 提升了隐写算法的抗检测能力。

2) 针对现有隐写算法采用人耳听觉的绝对掩蔽阈值无法充分利用 AAC 编码的隐写修改空间, 提出基于实际听觉掩蔽阈值的隐写失真代价, 提升了隐写算法的听觉隐蔽性。

3) 结合以上两种隐写修改失真代价, 提出基于联合失真的 AAC 安全隐写算法框架, 该框架结合了 Huffman 码字直方图统计分布和实际掩蔽阈值曲线, 设计联合失真代价函数, 提高了隐写算法的

统计安全性和不可感知性。

本文剩余内容的组织结构如下: 第2节简要介绍了 AAC 音频编码原理; 第3节介绍目前最新的研究工作, 并分析现有的隐写算法所存在的问题; 第4节描述了本文所提的隐写算法、失真函数的设计以及实现细节; 第5节给出了实验结果和结果分析, 分别从安全性、不可感知性和隐写容量3个方面对所提的算法和已有的算法进行比较和分析。最后, 在第6节对全文进行总结。

2 AAC 编码基本原理

AAC 是一种基于心理声学模型Ⅱ的有损感知音频压缩标准。AAC 的主要编码过程如图 1 所示, 其中主要包括 MDCT 变换、量化过程和 Huffman 编码。

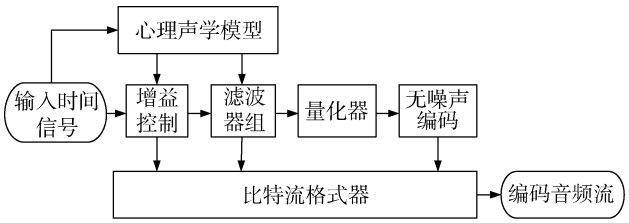


图 1 AAC 的主要编码流程

Figure 1 The main coding process of AAC

在音频编码过程中, 首先, AAC 对输入的脉冲信号进行时域和频域的转换, 根据心理声学模型Ⅱ和相应的感知熵, 计算最大允许失真, 并在时频转换过程中进行滤波。然后对 MDCT 变换得到的 MDCT 系数值进行三层循环量化和 Huffman 编码, 最后, 将编码后的信息和一些边信息按照特定的格式组合成 AAC 音频编码流。

MDCT 变换主要是利用心理声学模型Ⅱ计算输入信号能量所允许的最大失真能量, 计算一组信号掩码比和阈值, 标记出信号的块类型, 分别为长块、开始块、结束块、短块。量化编码是 AAC 实现音频编码压缩的主要过程。原始 MDCT 系数通过三层循环量化, 获得了最佳的量化系数 QMDCT, QMDCT 能够满足 AAC 音频码率和感知质量的要求。

在 Huffman 编码过程中, QMDCT 系数通过 Huffman 无损编码。AAC 按 12 个固定码表进行码字选择并完成编码。如表 1 所示。

在 Huffman 编码过程中, m 个 QMDCT 系数用码本中的码字表示, 其中 m 可以是 2 或者 4, 取决于码表。在实际编码过程中, 有两个码表同时进行预编码, 最后选择编码效率最高的码表进行编码。Huffman 编码是可变长度编码(variable-length code, VLC)。

表 1 Huffman 码表

Table 1 Huffman Code Table

码书索引	量化系数的个数 m	最大值的绝对值	符号位
0	—	0	—
1	4	1	Yes
2	4	1	Yes
3	4	2	No
4	4	2	No
5	2	4	Yes
6	2	4	Yes
7	2	7	No
8	2	7	No
9	2	12	No
10	2	12	No
11	2	16(ESC)*	

3 现有隐写算法及问题分析

目前, 基于 Huffman 编码域的隐写算法主要是 Yan 等[14]、Yang 等[15]、Yi 等[16]。 Yan 等[14]设计了一种基于 Huffman 码字替换的隐写算法。其中, 码字替换需要满足 3 个条件, 分别是替换的码字长度相同、替换的码字符号位相同、替换的码字对应的 QMDCT 系数之间的距离绝对值为 1, 如公式(1~3)所示。该方法通过替换 Huffman 码字进行嵌入秘密信息。该方法解决了其他隐写方法造成的帧偏移问题, 且隐藏容量大, 但是只利用了码字分组, 映射到二进制序列, 通过码字替换实现秘密消息的嵌入, 没有通过设计失真代价函数并利用 STC 框架实现自适应隐写算法, 因此, 在统计安全性和不可感知性上有所不足。

$$code\_len(vlc_i) = code\_len(vlc_j) \tag{1}$$

$$Sign(vlc_i) = Sign(vlc_j) \tag{2}$$

$$\left|\left(w_i + x_i + y_i + z_i\right) - \left(w_j + x_j + y_j + z_j\right)\right| = 1 \tag{3}$$

其中, i, j 代表 Huffman 码字的索引, $code\_len(vlc)$ 表示 Huffman 码字的长度, $Sign(vlc)$ 表示 Huffman 码字的符号位, w, x, y, z 分别代表 Huffman 码字对应的 QMDCT 系数。

Yang 等[15]和 Yi 等[16]在 Yan 等[14]的基础上, 将互相替换的 Huffman 码字进行分组, 利用人类听觉的绝对阈值曲线计算每个 Huffman 码字的失真代价, 最后, 利用 STCs[17]框架嵌入秘密信息。Yang 等[15]和 Yi 等[16]是目前最新的自适应隐写算法。与传统的非自适应隐写算法相比, 在安全性上有进一步的提高。

现有 AAC 或 MP3 隐写分析算法主要都是基于 MDCT 或 QMDCT 系数的统计分布特征, 构造一个二分类器进行隐写分析。其中, QMDCT 系数是 MDCT 系数的量化值。Jin 等[18]提出了一种基于 QMDCT 系数的统计特征的隐写分析方法。该方法提取了广义高斯函数的参数和 QMDCT 系数的直方图分布模型, 得到频域的马尔可夫转移矩阵部分统计特征, 以及 QMDCT 帧内和帧间的相关性特征, 利用支持向量机(Support Vector Machine, SVM)构造一个二分类器的隐写分析算法。Qiao 等[21]基于同一子带的 QMDCT 系数一阶差分, 从水平方向和垂直方向提取马尔可夫转移特征构建隐写分析器。Yi 等[16]提出了一种对同一子带的 QMDCT 系数提取二阶差分, 构建四组隐写分析特征: 基于频域的子带矩统计特征、累积马尔科夫转移特征、累积邻近密度特征和广义高斯分布特征形状参数, 设计一种联合特征提高隐写分析器的性能。Kuriakose 等[20]结合帧间 QMDCT 系数的二阶差分的马尔可夫转移特征和累积相邻联合概率密度来提高性能。Ren 等[22]提出了一个基于 MDCT 系数矩阵的富特征模型, 分别计算了 MDCT 帧与帧之间的一阶差分和二阶差分的残差系数矩阵, 然后基于残差矩阵计算马尔可夫转移概率特征, 最后构造一个联合概率特征的隐写分析器。这些传统的隐写分析框架都是通用的。以及目前使用深度学习来构建隐写分析框架, 如 Lin 等[23]、Wang 等[24]、Wang 等[25]均是提出基于 CNN 的隐写分析网络, 利用 QMDCT 系数作为输入, 提取 QMDCT 系数中包含的特征进行分析检测。

现有隐写算法在统计安全性和听觉掩蔽性上存在以下两种问题, 提出了两个方面的分析:

3.1 Huffman 码字统计分布特征分析

本文对原始音频和对应的含密音频的 Huffman 码字进行直方图统计, 这两种隐写算法分别是 Yang 等[15]和 Yi 等[16]。

统计实验的结果如图 2 所示, 横坐标表示 Huffman 码字的序号, 纵坐标表示在原始音频与含密音频中相对应的每个 Huffman 码字出现频率的差值。图 2 中的统计实验使用的样本为编码码率为 128 kbps、时长为10 s 的4000个 AAC 音频数据, 数据是4000个样本的平均值。由图2可知, 大部分 Huffman 码字统计频率的差值的绝对值超过了20。这表明了 Yang 等[15] 和 Yi 等[16]的隐写算法在嵌入过程中并没考虑对 Huffman 码字统计分布的影响, 直接影响了原始音频中 Huffman 码字正常地统计分布, 最终会降低这方面的统计安全性, 含密音频将会很容易被检测到。

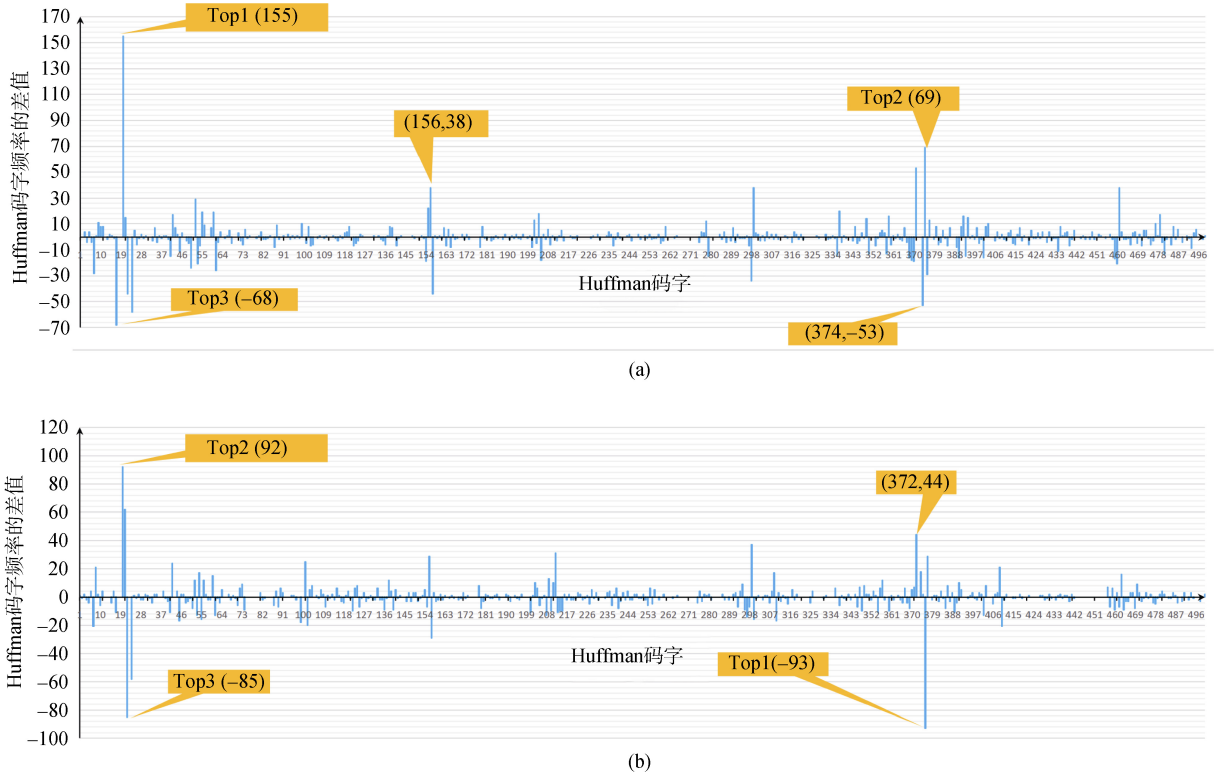


图 2　原始音频和含密音频的 Huffman 码字直方图统计频率的差值分布图, (b)为 Yang 等[15], (b)为 Yi 等[16]

Figure 2　The difference of the histogram of the Huffman Codeword between Cover and Stego in different steganographic algorithms: (a) Yang et al. [15], (b) Yi et al. [16]

3.2 AAC 中实际掩蔽曲线分析

在AAC编码过程中, 输入的PCM音频数据由长帧或短帧通过子带滤波器得到 1024 或 128 个MDCT。在这里, 本文对长帧进行分析, 短帧情况类似。其中, 长帧谱分布范围为 0~1024, 1024 条谱线量化为 1024 个 QMDCT 系数。因此, 人类听觉系统对每个 QMDCT 系数的灵敏度是不同的。如图 3 所示, 显示的是一帧的音频数据所表示的频谱数据。其中, 绿色曲线代表的是 PCM 音频数据帧内的原始频谱曲线, 橙色曲线代表的是人类听觉的绝对阈值曲线。蓝色曲线代表的是 AAC 在编码过程中利用心理声学模型 II 将音频信号转化成频谱系数(MDCT), 它代表了 AAC 编码过程中实际使用的听觉掩蔽曲线。

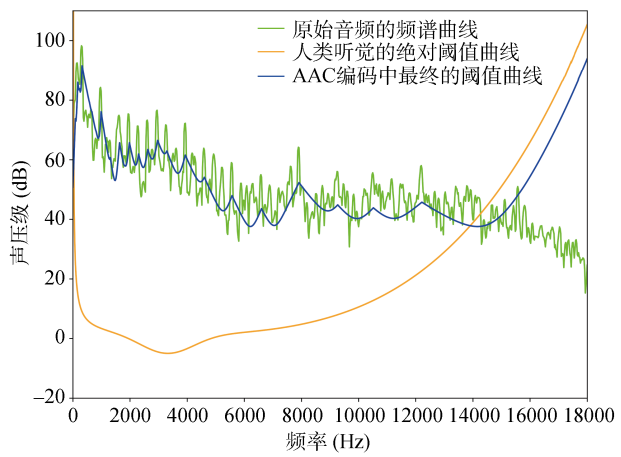


图 3 音频信号的实际阈值曲线样例

Figure 3 the masking threshold of signal

Yang 等[15] 和 Yi 等[16]是利用人类听觉的绝对阈值曲线来计算嵌入操作对人耳听觉感知造成的失真代价。事实上, 人类听觉的绝对阈值曲线是表示人在安静的环境中听到不同频率的纯音所需要的能量。图 3 中三条曲线的关系表明, 人类听觉的绝对阈值曲线(黄色曲线)仅仅是最终实际阈值曲线(蓝色曲线)的部分计算, 并不能完全表示 AAC 编码中音频信号的实际变化。所以, 本文采用实际阈值曲线去计算嵌入过程对人耳感知的影响, 可以更好地感知在实际编码过程中音频信号的能量变化。

根据 3.1 和 3.2 的分析, 可以得出以下结论: 首先, 由于现有的隐写算法没有考虑 Huffman 码字的统计分布特性, 导致了 Stego 在这方面统计安全性较低。其次, 现有的隐写算法所利用的人类听觉的绝对阈值曲线并不能完全表示 AAC 编码过程中音频能量的实际变化。为了提高含密音频的统计安全性和不可感知性, 本文提出了一种结合 Huffman 码字的统计分布特征和实际听觉阈值曲线特征的自适应安全隐写算法。

4 基于联合失真的 AAC 安全隐写算法

基于以上分析, 本文提出一种基于联合失真的 AAC 安全隐写算法, 该框架基于多方面的联合统计失真, 包括 Huffman 码字统计分布特征和实际听觉阈值曲线特征, 最后, 利用 STCs[17]实现自适应隐写。

首先, 本文在设计隐写算法时, 需要考虑嵌入操作对 Huffman 码字的统计分布特征的影响, 即改变了 Huffman 码字的直方图分布情况。为了尽可能地降低对这方面安全性地影响, 本文基于 Huffman 码字的统计分布特征设计了失真代价 D_{cost1}。

其次, Yang 等[15]和 Yi 等[16]在设计失真代价函数时, 所采用的人类听觉的绝对阈值曲线并不能完全表示编码过程中音频信号的实际变化, 所以, 设计的失真代价函数并不能很好地计算嵌入操作对含密音频的不可感知性造成的影响。为了进一步提高含密音频的不可感知性, 本文利用心理声学模型 II 中实际的听觉掩蔽曲线特征设计了失真代价 D_{cost2}。

最后, 通过权重值 α 将 D_{cost1} 和 D_{cost2} 结合成新的联合失真函数, 这种失真计算同样可以直接应用到其他基于 AAC、MP3 自适应隐写算法, 包括 Yang 等[15]和 Yi 等[16]。

4.1 基于 Huffman 码字统计分布的失真代价 D_{cost1}

如图 2 所示, 对于同一个 AAC 音频片段, 原始音频和含密音频的 Huffman 码字频率直方图的分布是有明显的不同。因此, 为了尽可能地避免嵌入操作对 Huffman 码字频率分布造成地影响, 本文对每一种码字进行频率统计, 并得到出现概率 p。本文可以得出, 相互可以替换的 Huffman 码字的出现概率 p 越接近, 则进行 Huffman 码字替换操作对 Huffman 的频率分布的影响就越小, 即含密音频和原始音频在 Huffman 码字频率分布上就越接近, 对统计安全性的影响就越低。

$$D_{cost1}: a(H_{cj}, H'_{cj}) = \frac{1}{2}\left(\frac{\left|p_{cj}-p'_{cj}\right|}{p_{cj}}+\frac{\left|p_{cj}-p'_{cj}\right|}{p'_{cj}}\right)\left(-\log_2\left(p_{cj}+p'_{cj}\right)\right) \tag{4}$$

D_{cost1} 是针对 AAC 音频编码中 Huffman 码字的出现概率进行计算的, 如公式(4)所示。

其中, p_{cj} 和 p'_{cj} 之差的绝对值分别与 p_{cj} 和 p'_{cj} 的比值, 如果比值越小, 说明同一样本集中 H_{cj} 和 H'_{cj} 出

现的频数很接近, 利用差值的比值计算避免了因为码字总数过大, 导致部分低频率出现的码字带来的计算误差的问题。对 p_{cj} 和 p'_{cj} 之和进行对数计算, 为了调节高频的码字和低频的码字在概率计算上的数量级差距, 保证最终的失真代价结果在合适的取值区间内。

其中, 实验使用的样本为编码码率为 128 kbps、时长为 10 s 的 4000 个 AAC 音频数据。

4.2 基于实际掩蔽阈值曲线的失真代价 D_{cost2}

基于 3.3 的分析, AAC 编码标准采用的是心理声学模型Ⅱ, 将音频信号的掩蔽阈值与人类听觉的绝对阈值相结合, 将曲线以下的噪声剔除, 获得最终的音频能量掩蔽阈值曲线。因此, 根据心理声学模型Ⅱ, 本文采用最终的掩蔽阈值曲线来设计人类听觉敏感失真函数。AAC 中计算掩蔽阈值的流程如下:

步骤 1: 声压级标准化和频谱分析:

对音频信号逐帧地进行快速傅里叶变换(Fast Fourier Transform, FFT), 计算声压级得到标准化的功率谱密度 $P(k)$。PN 表示为每一帧频谱中最大的声压级和标准声压级 96 dB 的差值。如公式 5 所示:

$$P(k)=PN+20\log_{10}\left|\sum_{n=0}^{N-1}w(n)x(n)\exp\left(-\frac{j2\pi kn}{N}\right)\right| \quad (5)$$

步骤 2: 识别音调和非音调成分:

在功率谱密度值中, 需要找出局部最大值, 然后计算出音调分量 S_T, 如公式(6)、(7)所示:

$$S_T=\begin{cases}P(k)\Big|P(k)>P(k\pm1)\\ P(k)\Big|P(k)>P(k\pm\Delta_k)+7(\text{dB})\end{cases} \quad (6)$$

$$\Delta_k\in\begin{cases}2 & 2<k<63 \quad (0.17\sim5.5\text{ kHz})\\ [2,3] & 63\leqslant k<127 \quad (5.5\sim11\text{ kHz})\\ [2,\cdots,3] & 127\leqslant k\leqslant156 \quad (11\sim20\text{ kHz})\end{cases} \quad (7)$$

从上一步得到的音调成分中计算三个连续的频谱值, 作为音调成分的掩蔽音的声压级 $P_{TM}(k)$, 如公式(8)所示:

$$P_{TM}(k)=10\log_{10}\left(10^{\frac{p(k-1)}{10}}+10^{\frac{p(k)}{10}}+10^{\frac{p(k+1)}{10}}\right)(\text{dB}) \quad (8)$$

然后, 计算非音调成分(类噪声) 的掩蔽音的声压级 $P_{NM}(k)$, 如公式(9)所示。

$$P_{NM}(k)=10\log_{10}\sum_j10^{0.1P(j)}\ (\text{dB}) \quad \left(j\notin(k,k\pm1,k\pm\Delta_k)\right) \quad (9)$$

步骤 3: 和绝对阈值曲线比较, 并计算最终的掩蔽阈值曲线:

将步骤 2 计算的掩蔽音的声压级和人耳的绝对阈值曲线比较, 将小于绝对阈值的掩蔽音剔除。然后, 用带宽为临界带宽二分之一的滑动窗口抽取窗口内的掩蔽音, 如公式(10)所示。

$$\begin{cases}P_{TM,NM}(i)=P_{TM,NM}(k)\\ P_{TM,NM}(k)=0\\ P_{TM,NM}(k)\geqslant T_q(k)\end{cases} \quad (10)$$

$$i=\begin{cases}k & 1\leqslant k\leqslant48\\ k+(k\bmod2) & 49\leqslant k\leqslant96\\ k+3-((k-1)\bmod4) & 97\leqslant k\leqslant232\end{cases}$$

人类听觉的绝对阈值曲线是在静音环境下, 人耳对不同的声音频率能量感知的最小声压级。低于该阈值曲线的声音不能被人耳听见。人类听觉的绝对阈值曲线是固定不变的, 其非线性曲线接近如公式(11)所示。

$$T_q(f)=3.64\times\left(\frac{f}{1000}\right)^{-0.8}-6.5\times e^{-0.6\times\left(\frac{f}{1000}-3.3\right)^2}+10^{-3}\times\left(\frac{f}{1000}\right)^4 \quad (11)$$

其中, $T_q(f)$ 表示人耳在声音频率为 f 时, 听到的绝对阈值。

步骤 4: 分别计算单个音调和噪声的掩蔽阈值:

1) 计算单个音调的掩蔽阈值, 如公式(12)所示:

$$T_{TM}(i,j)=P_{TM}(j)-0.275z(j)+SF(i,j)-6.025 \quad (12)$$

其中, $T_{TM}(i,j)$ 代表单个音调成分频率 i 对频率 j 的掩蔽阈值, $SF(i,j)$ 代表频率对频率的掩蔽函数。

2) 计算单个非音调的掩蔽阈值, 如公式(13)所示:

$$T_{NM}(i,j)=P_{TM}(j)-0.175z(j)+SF(i,j)-2.025 \quad (13)$$

其中, $T_{NM}(i,j)$ 代表单个非音调成分频率 i 对频率 j 的掩蔽阈值。

步骤 5: 计算最终的全局掩蔽阈值, 把单个音调成分、非音调成分和绝对阈值曲线相加就可以求出最终的全局阈值曲线。如公式(14)所示。

$$T_g(i)=10\log_{10}\left(10^{0.1T_q(f)}+\sum_{l=1}^{L}10^{0.1T_{TM}(i,j)}+\sum_{m=1}^{M}10^{0.1T_{NM}(i,m)}\right) \quad (14)$$

基于 Huffman 编码域的隐写算法是通过替换码字实现秘密信息的嵌入。Huffman 码字的替换最终会

改变对应的 QMDCT 系数的值。而 QMDCT 系数的值对应的音频时域声音能量的大小。所以, 当相互替换的 Huffman 码字 H_{cj} 和 H'_{cj} 间接对应的时域声音能量的差值越大时, 则原始音频和含密音频在听觉感知型的差距也越大, 即含密音频的不可感知性就越差。通过上文的分析, 为了尽可能地减小嵌入操作对听觉感知的影响, 本文利用公式(14)来设计有关不可感知性的失真函数, 相比 Yang 等[15]、Yi 等[16]中使用人类听觉绝对阈值曲线更能准确地计算嵌入操作对音频编码过程中声音能量变化地影响。失真函数 D_{cost2} 如公式(15)所示:

$$D_{cost2}:b(H_{ci},H'_{ci})=\frac{\left|x_i-x'_i\right|-\left|y_i-y'_i\right|}{\log_{10}\left(\dfrac{T_g(i)+T_g(i+1)}{2}+\theta\right)} \tag{15}$$

其中, $i\in\{1,2,3,\cdots,1024\}$ 表示着 QMDCT 系数的索引值, x,y 表示对应 QMDCT 系数。为了保证 $\dfrac{T_g(i)+T_g(i+1)}{2}+\theta>1$ 以及一些极端值的影响, θ 根据实际的音频数据设定为一个合适的偏移量。

4.3 基于联合失真代价的 AAC 安全隐写算法

4.3.1 基于联合失真代价函数设计

为了同时保持统计安全性和不可感知性的平衡, 本文使用一个权重值 α, 将 D_{cost1} 和 D_{cost2} 组合成一个新的联合失真代价函数 $D_{cost-union}$。如公式(16)所示:

$$D_{cost-union}:\rho(H_{ci},H'_{ci})=\alpha\left(a(H_{ci},H'_{ci})\right)+(1-\alpha)\left(b(H_{ci},H'_{ci})\right) \tag{16}$$

其中, 权重值 $\alpha\in\{0.1,0.2,\cdots,0.9\}$。在后续的实验结果分析和图表中, 采用的是 α=0.4 情况下的实验结果。另外, 不同权重的实验情况会在后面的实验中展开。

4.3.2 隐写算法框架的整体设计

为了提高现有隐写算法的统计安全性和不可感知性, 本文提出了一种基于联合失真的 AAC 安全隐写算法框架, 将上文所设计的联合失真代价函数结合自适应嵌入策略 STCs[17], 实现一个完整可行的隐写算法, 其主要流程框架如图 4 所示, 分为以下四个部分:

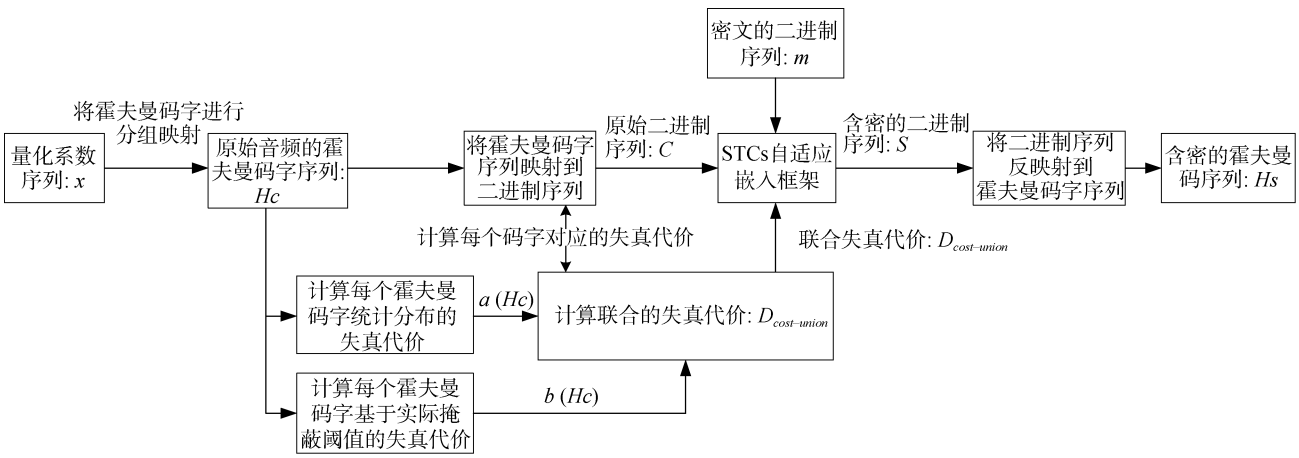


图 4 本文所提的隐写框架

Figure 4 The flowchart of the proposed scheme

1) 将 QMDCT 进行 Huffman 编码, 得到 Huffman 码字, 然后, 根据公式(1)、(2)、(3)进行分组配对, 其中公式(3)的距离值设置为 2。这样, 就可以得到多个 Huffman 码字替换的分组。

2) 将 Huffman 码字进行分组后, 根据联合失真函数计算每一个 Huffman 码字的整体代价, 并映射到二进制比特流序列中。

3) 计算基于整个音频文件的全局最优的嵌入路径, 利用 STCs[17]实现秘密信息的嵌入, 并输出含秘的二进制比特流。

4) 将含秘的二进制比特流逆映射到 Huffman 码字上, 进行码字替换, 最终得到含密音频。

其中, 本文所提算法的嵌入流程和 Yang 等[15]、Yi 等[16] 的区别如下:

1) 相比于 Yang 等[15], Yang 等[15]采用的是 Yan 等[14]一样分组策略, 即一对一替换, 在嵌入前已经是固定替换对象, 这样就会导致在嵌入时不能动态地选择最优的两个 Huffman 码字进行替换。在本文所提的算法中, 本文采用的是多个 Huffman 码字相互替换的策略, 这样可以动态选择最优的码字间替换;

2) 相比于 Yi 等[16], Yi 等[16]采用以帧为单位的选帧策略, 选择最优的帧嵌入顺序, 将秘密信息自适应地嵌入到整个音频文件中合适的帧和帧内合适的位置。而本文采用的是以码字为单位的嵌入策略, 通过 STCs[17]将秘密信息自适应地嵌入到全局最优的位置, 将嵌入修改的影响尽可能地降低到全局最小, 相对更均匀地分布在整个音频样本中, 尽可能地避免嵌入修改带来的损失集中在某些局部区域的情况, 提高了统计安全性。

4.3.3 嵌入流程

如表 2 所示, 算法 1 的具体嵌入操作如下:

表 2 嵌入算法流程

Table 2 The process of embedding procedure

Algorithm 1
Input: H_c , m , k
Output: H_s
1: for i = 1 to the length of H_c do
2: $C[i]=f\ ctb(H_c[i])$
3: end for
4: Calculate the $\rho(H_{ci},H'_{ci})$
5: Scramble the message m with a same key k to m'
6: $S=STCs(C,m',\rho)$
7: Inverse scramble the S
8: for i = 1 to the length of H_c do
9: $S[i]=f\ btc(S[i],H_c[i])$
10: end for
11: return H_s

1) 将 wav 音频输入到 AAC 编码器中进行编码, 经过第一次完整编码后, 在 Huffman 编码环节, 获得整个原始音频文件对应的 Huffman 码字序列 $H:\{H_1,H_2,\cdots,H_n\}$;

2) 同时, 在第一次完成编码过程中, 将 Huffman 码字进行分组, 并将 Huffman 码字序列 $H_c:\{H_1,H_2,\cdots,H_n\}$ 映射到二进制比特流序列 $C:\{C_1,C_2,\cdots,C_n\}$;

3) 按照设计的联合失真函数, 计算每一个 Huffman 码字对应的实际失真代价, 并得到失真代价序列 $\rho:\{\rho_1,\rho_2,\cdots,\rho_n\}$;

4) 用密钥 k 将秘密信息 m 二进制比特流序列进行置乱处理, 得到 $m^{'}$, 以保证安全性;

5) 然后, 将载体序列 C、失真代价序列 ρ、$m^{'}$ 一起输入 STCs[17]中, 计算最优的嵌入路径, 得到含秘的载体序列 S;

6) 然后, 将含秘的载体序列 S 进行逆映射操作, 得到对应的 Huffman 码字序列 H_s;

7) 最后, 经过第二次的 AAC 编码, 在 Huffman 编码环节中, 进行码字替换, 输出含密音频, 隐写嵌入结束。

4.3.4 提取流程

如表 3 所示, 算法 2 的具体提取操作如下:

表 3 提取算法流程

Table 3 The process of extracting procedure

Algorithm 2
Input: H_s , k
Output: m
1: for i = 1 to the length of H_s do
2: $S[i]=f\ ctb(H_s[i])$
3: end for
4: Extract m' from *STCs*
5: Scramble the message m' with a same key k to m
6: return m

1) 将含密音频输入到 AAC 解码器进行解码, 经过完整地解码后, 在 Huffman 解码环节, 获得码字序列 $H_s:\{H_1,H_2,\cdots,H_n\}$;

2) 同时, 在第一次解码过程中, 将 Huffman 码字进行分组, 并将 Huffman 码字序列 $H_s:\{H_1,H_2,\cdots,H_n\}$ 映射到二进制比特流序列 S;

3) 将 S 输入到 STCs[17]中, 得到置乱后地秘密消息 m';

4) 用密钥 k 将置乱后地秘密信息 m' 二进制比特流序列进行逆处理, 得到秘密信息 m, 提取过程结束。

5 实验结果及结果分析

本章实验将统计安全性、不可感知性和嵌入容量三个方面分别设计实验一、二、三。同时, 在每个实验中对所提方法的有效性进行结果展示和分析。其中, 统计安全性通过检测准确率进行评价, 听觉掩蔽性通过 PEAQ[29](Perceptual Evaluation of Audio Quality)进行评价, 隐写容量通过每秒音频能嵌入的最大比特数进行评价。

5.1 实验设置

5.1.1 音频数据集

音频数据库由 4000 个采样率 44.1kHz、时长为 10s 的 wav 音频文件组成, 来自于 Yang 等[15]、Yi 等[16]中所提供的数据集。为了和 Yang 等[15]、Yi 等[16]保

持一致实验设置, 本实验采用 128kbps 的比特率进行编码。

5.1.2 对比的隐写算法

为了对比本文所提算法和已有隐写算法的实验结果, 设计了 6 组对比隐写算法: Yang 等[15], Yi 等[16], Yang 等[15]- D_{cost1} , Yi 等[16]- D_{cost1} , Yang 等[15] - D_{cost2} , Yi 等[16]- D_{cost2} 。其中, Yang 等[15] - D_{cost1} 表示用 D_{cost1} 的失真函数替代 Yang 等[15]中的失真函数进行实验, Yang 等[15]- D_{cost2} 表示用 D_{cost2} 的失真函数替代 Yang 等[15]中的失真函数进行实验, 嵌入流程保持不变。Yi 等[16]- D_{cost1} 、Yi 等[16]- D_{cost2} 组合类似。

为了确保实验环境是一致的, 这 7 组实验都是在 Huffman 参数域实现秘密消息的嵌入。另外, 本文使用有效负载 *Payload* 表示隐写嵌入率, 即绝对嵌入率, 后续实验中的 5 种嵌入率也和 Yang 等[15] 、Yi 等[16]保持一致。定义如公式(17)所示:

$$Payload = \frac{n}{T} \tag{17}$$

其中, n 是消息 m 的比特数, T 是音频的时长, 这里是 10s。

5.1.3 隐写分析算法

为了更好地对本文所提算法的统计安全性进行检测和分析, 本文分别采用了传统手工提取特征的隐写分析方法和利用深度学习神经网络构建的隐写分析方法, 即 Yang 等[15], Yi 等[16]中使用的传统隐写分析算法 Ren 等[22]和现有基于 CNN 的 Wang 等[24]进行实验。其中, Ren 等[22]是传统隐写分析的代表, 而 Wang 等[24]是基于 CNN 隐写分析方案的代表。

另外, 在本文中, 本文还根据 Huffman 码字的统计分布特征, 利用 SVM 构建了一个新的隐写分析器, 针对 Huffman 参数域的分布进行检测, 本文中用 *Huffman – SVM* 表示。

5.2 统计安全性实验结果和分析

第一个实验部分是评估所提方法的统计安全性, 并和已有的隐写算法进行对比和分析。

实验用了 Ren 等[22]、Wang 等[24]和 *Huffman – SVM* 来评估以上 7 组实验的统计安全性。在 Ren 等[22]的实验中, 实验测试了 4000 对原始音频和含密音频, 其中 2400 对原始音频和含密音频用作训练集, 1600 对原始音频和含密音频作为测试集。在 Wang 等[24]的实验中, 其中, 2400 对原始音频和含密音频用作训练集, 1600 对原始音频和含密音频作为测试集, 提取每一条音频数据的前 96 帧的特征, 相比 Yang 等[15]和 Yi 等[16]每条音频增加了 46 帧的特征数据, 采集了更多的特征数据, 保证了更高的实验精度。在 *Huffman – SVM* 的实验中, 使用 2400 对原始音频和含密音频作为训练集, 采用 1600 对原始音频和含密音频作为测试集, 提取 Huffma 码字的直方图分布特征作为输入, 构建 SVM 分类器进行检测分析。在实验中, 使用隐写分析算法的检测准确度 *ACC* 评估隐写算法的统计安全性, 定义如公式(18)所示。

$$ACC = \frac{TNR + TPR}{2} \tag{18}$$

其中, *TNR* (True Negative Rate)和 *TPR* (True Positive Rate)分别表示对含密音频预测正确和对原始音频预测正确。

5.2.1 Wang 等[24]方案的隐写分析结果和分析

首先, 本文利用 Wang 等[24]的隐写分析算法对 7 组隐写算法, 分别按照不同的有效载荷进行检测分析。实验结果如表 4 和图 5 所示, 可以看到本文所设计的失真函数 D_{cost1} 和 D_{cost2} , 仅仅是替换已有隐写算法的失真部分, 如: Yang 等[15]- D_{cost1} , Yi 等[16]- D_{cost1} , Yang 等[15]- D_{cost2} , Yi 等[16]- D_{cost2} , Wang 等[24]的检测率出现明显地下降。本文所提出的联合失真将两个失真函数相结合, 设计本文的隐写方法, 在安全性上能达到最高。相比 Yang 等[15], 在低绝对嵌入率 1.6 kbps 时, 安全性提高了 11%, 在高绝对嵌入率 6 kbps 时, 安全性提高的更明显——约 16%。相比 Yi 等[16], 安全

表 4 Wang 等[24]的隐写分析检测下的实验结果

Table 4 The decetion accuracy of Wang et al.[24] to detect schemes

隐写算法	*Payload*				
	1.6	2.1	3.0	4.0	6.0
Yang 等[15]	0.6452	0.6748	0.6845	0.7064	0.8426
Yang 等[15] - D_{cost1}	0.5571	0.5903	0.6245	0.6842	0.7831
Yang 等[15] - D_{cost2}	0.6147	0.6358	0.6513	0.6972	0.8287
Yi 等[16]	0.665	0.6975	0.7315	0.7464	0.8837
Yi 等[16]- D_{cost1}	0.5495	0.5635	0.5925	0.6355	0.7085
Yi 等[16]- D_{cost2}	0.6389	0.6842	0.7134	0.7218	0.8576
本文算法	**0.5326**	**0.5647**	**0.5898**	**0.6246**	**0.6821**

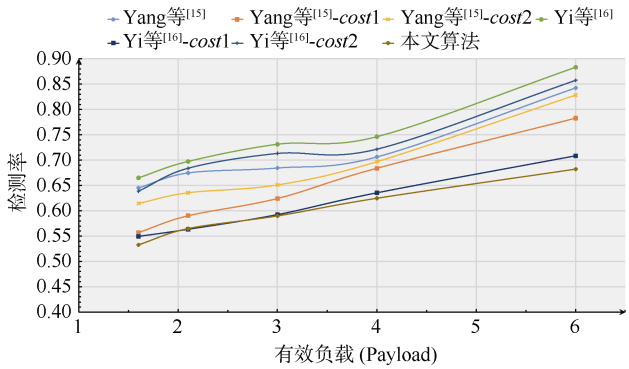


图 5　Wang 等[24]的检测结果折线图

Figure 5　The line chart of the decetion accuracy of Wang et al.[24]

性也有明显的提升, 在 13%~20%。实验中有效载荷的设计和 Yang 等[15], Yi 等[16]是保持一致的。

实验结果及原因分析如下:

1) 使用失真函数 D_{cost1} 设计失真, 在安全性上有提升, 是因为已有的隐写算法在 Huffman 参数域进行隐写嵌入, 却没有考虑对 Huffman 码字的统计分布的影响, 通过 D_{cost1} 的引用, 尽可能降低在这方面的影响, 所以安全性提升。

2) 使用失真函数 D_{cost2} 设计失真, 在安全性上有提升, 是因为已有的隐写算法使用人耳绝对阈值曲线去计算嵌入操作带来的失真代价。从 3.2 的分析和图 3 可以知道, 人耳绝对阈值曲线并不是完全模拟编码中音频实际变化, 仅仅是其中一部分, 所以, 本文采用实际掩蔽阈值曲线, 相对人耳绝对阈值曲线进行失真代价的计算更能贴近实际的失真情况, 从而尽可能地减少含密音频和原始音频的差距, 提高了安全性。

5.2.2　Ren 等[22]方案的隐写分析结果和分析

第二个算法安全性评价实验, 本文使用 Ren 等[22]的隐写分析方法对 7 组隐写算法进行隐写检测, Ren 等[22]使用的是人工提取特征的方法, 相比 Wang 等[24]的检测精度会更高一些。如表 5 和图 6 所示, 其实验结论和图 5 是一致, D_{cost1} 和 D_{cost2} 单独进行实验, 在安全性上具有不同的提升, 而本文方法将二者相结合, 提升效果也有一定提升, 相比 Yang 等[15], 提升在 4%~10%左右。相比 Yi 等[16], 提升在 8%~11%。

表 5　在 Ren[22]的隐写分析检测下的实验结果

Table 5　The decetion accuracy of Ren[22]to detect schemes

隐写算法	*Payload*				
	1.6	2.1	3.0	4.0	6.0
Yang 等[15]	0.6123	0.6489	0.6963	0.7654	0.8829
Yang 等[15] - D_{cost1}	0.5855	0.6189	0.6432	0.7016	0.8288
Yang 等[15] - D_{cost2}	0.5987	0.6312	0.6743	0.7411	0.8621
Yi 等[16]	0.6857	0.7194	0.7488	0.7794	0.8647
Yi 等[16]- D_{cost1}	0.6694	0.6988	0.7189	0.7413	0.7976
Yi 等[16]- D_{cost2}	0.6734	0.7132	0.7358	0.7669	0.8433
本文算法	**0.5693**	**0.5998**	**0.6408**	**0.6943**	**0.7814**

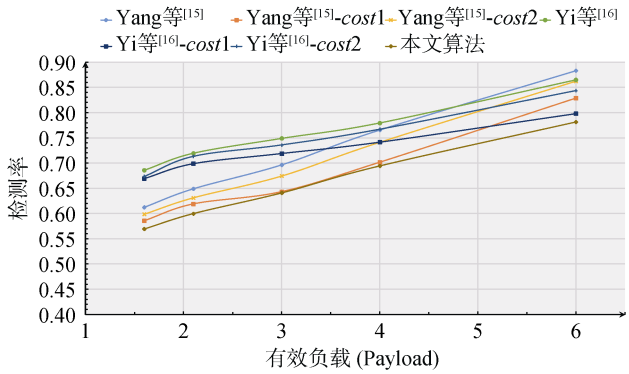


图 6　Ren 等[22]的检测结果折线图

Figure 6　The line chart of the decetion accuracy of Ren et al.[22]

5.2.3　基于 *Huffman – SVM* 方案的隐写分析结果和分析

第三个算法安全性评价实验是本文提出的, 利用 Huffman 码字的统计分布特征作为隐写分析特征, 最后利用 SVM 进行二分类判决。因为, Yang 等[15], Yi 等[16]以及本文所提的方法都是在 Huffman 参数域进行嵌入操作。设计这种专门针对特定隐写域的隐写分析算法, 可以很容易检测隐写算法的安全性。表 6 和图 7 的实验结果表明 *Huffman – SVM* 的检测性能是要明显优于 Ren 等[22]和 Wang 等[24]的。同时, 也看到了 Yang 等[15]和 Yi 等[16]对 Huffman 码字的统计分布特性有明显地影响, 降低了这方面的统计安全性。因为是特殊

表 6 在 *Huffman – SVM* 的隐写分析检测下的实验结果

Table 6 The decetion accuracy of *Huffman – SVM* to detect schemes

隐写算法	*Payload*				
	1.6	2.1	3.0	4.0	6.0
Yang 等[15]	0.8872	0.9548	0.9796	0.9924	0.9993
Yang 等[15] - D_{cost1}	**0.5314**	**0.6032**	**0.6793**	**0.7516**	**0.8136**
Yang 等[15] - D_{cost2}	0.8327	0.8692	0.9142	0.9431	0.9782
Yi 等[16]	0.8886	0.9515	0.9795	0.9952	0.9997
Yi 等[16]- D_{cost1}	**0.5295**	**0.6054**	**0.6806**	**0.7929**	**0.8216**
Yi 等[16]- D_{cost2}	0.8103	0.8632	0.9174	0.9517	0.9912
本文算法	0.5824	0.6836	0.7416	0.8631	0.9245

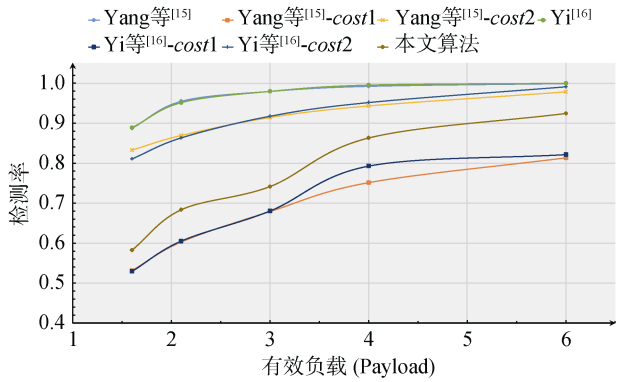


图 7 *Huffman – SVM* 的检测结果折线图

Figure 7 The line chart of the detection accuracy of *Huffman – SVM*

的专用型隐写分析方法，所以在提出的 D_{cost1}、D_{cost2} 和联合失真中，安全性最优是 D_{cost1}，相比 Yang 等[15]，安全性提升约在 18%~35%左右。相比 Yi 等[16]，安全性提升约在 17%~36%。这正是因为 *cost*1 也是专门针对原始音频和含密音频之间的 Huffman 码字的统计分类特征的差距而计算的失真代价，在专门的隐写分析下，提升的效果比较明显。

同时，本文也统计了本文算法的含密音频和原始音频音频样本中相对应的每个 Huffman 码字出现频率的差值。如图 8 所示，同 3.1 的图 2 是在相同的实验条件下统计得到的，实验结果表明，频率的差值有明显下降，整体趋势较图 2 更加平缓。本文所提的隐写算法相较 Yang 等[15]和 Yi 等[16]更能保持了这方面的统计分布特征，提高了统计安全性。

以上 3 组实验结果表明，本文所提的基于联合失真的隐写算法相较 Yang 等[15]和 Yi 等[16]在统计安全性上有显著的提升。

5.2.4 本文算法的 D_{cost1}、D_{cost2} 和 $D_{cost-union}$ 的对照分析

本小节实验主要是针对 D_{cost1}、D_{cost2} 和 $D_{cost-union}$ 在本文算法中设计的对照分析实验，分别利用 Ren 等[22]、Wang 等[24]和 *Huffman – SVM* 对 3 种情况进行检测分析。实验结果如图 9 所示:

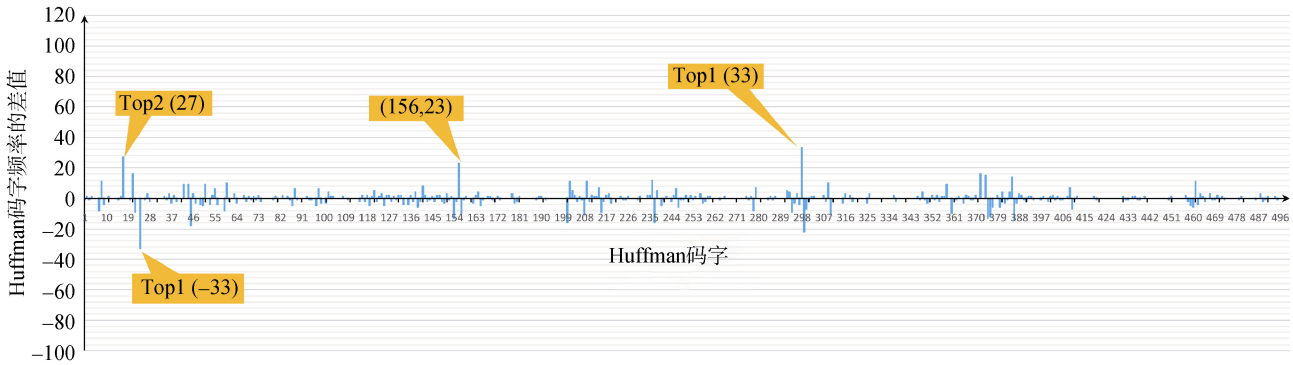


图 8 原始音频和所提方法生成的含密音频的 Huffman 码字直方图统计频率的差值分布图

Figure 8 The difference of the histogram of the Huffman Codeword between Cover and Stego in the proposed steganographic algorithms

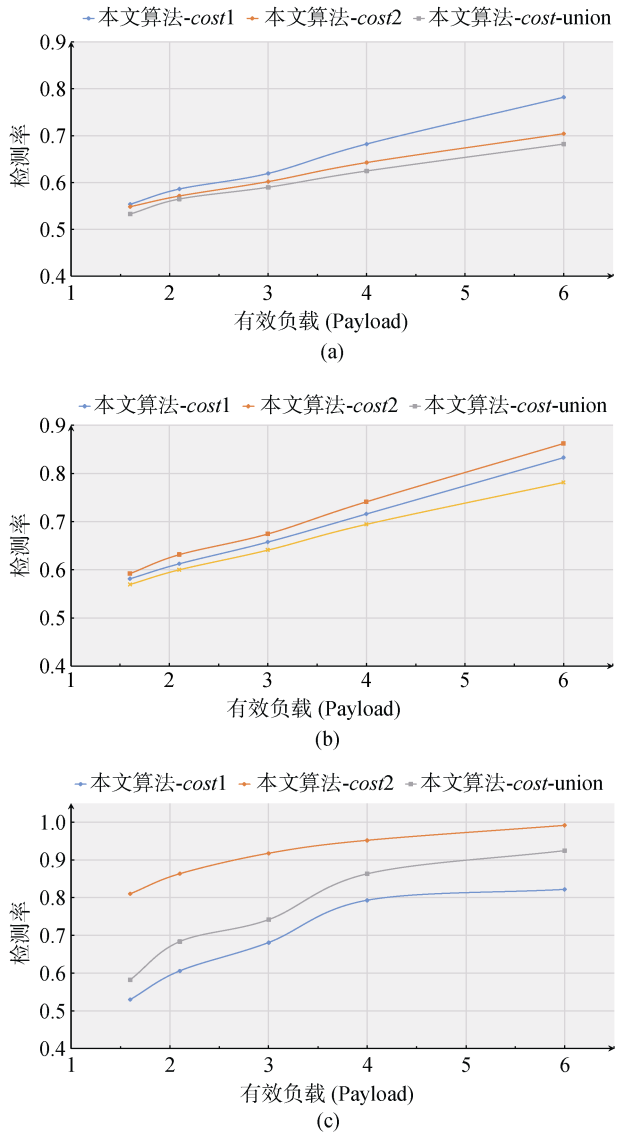


图 9　本文算法在 D_{cost1}、D_{cost2} 和 $D_{cost-union}$ 下的对比实验, (a)为 Wang 等[24], (b)Ren 等[22], (c)为 *Huffman – SVM*

Figure 9　The proposed scheme is tested under D_{cost1}, D_{cost2} and $D_{cost-union}$, (a) Wang et al.[24],(b) Ren et al.[22],(c) *Huffman – SVM*

如图 9 所示, 在 Ren 等[22]和 Wang 等[24]的隐写分析算法的检测情况下, 使用 $D_{cost-union}$ 相比单独使用 D_{cost1}、D_{cost2} 具有更好安全性, 因为使用 $D_{cost-union}$ 能够更好地平衡两个部分的失真影响; 在 *Huffman – SVM* 的隐写分析算法的检测情况下, 单独使用 D_{cost1} 相比 $D_{cost-union}$ 和 D_{cost2} 具有更好的安全性。在不同的隐写分析检测情况下, 利用不同失真代价函数的特性, 平衡整体的失真代价, 达到最优的效果。

5.2.5　联合失真权重α对安全性的影响

本节主要是为了分析在不同的隐写分析算法下, 联合失真函数权重 α 的变化情况。采用的和上面 3 组实验相同的样本集, 分别利用 Wang 等[24]、Ren 等[22]和 *Huffman – SVM* 对本文所提算法在 *Payload* 为 1.6, 3.0, 6.0 3 种有效载荷的情况下进行实验。实验结果如图 10 所示:

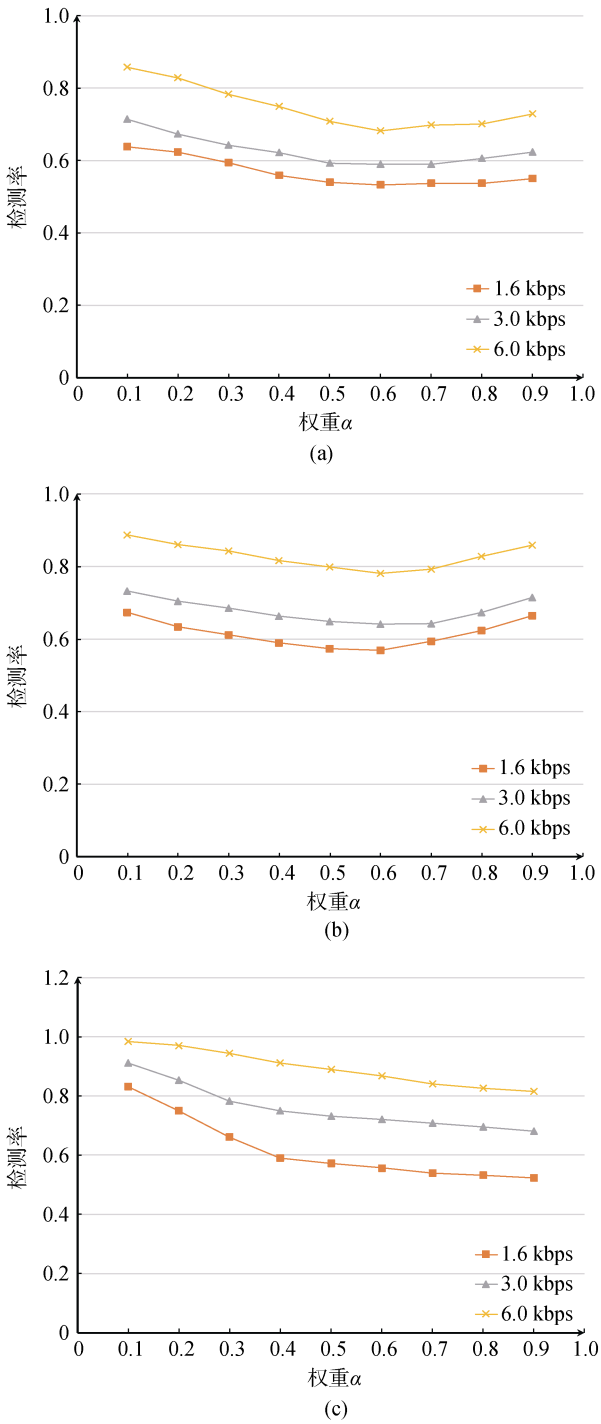


图 10　不同的权重α的检测曲线, (a)为 Wang 等[24], (b)Ren 等[22], (c)为 *Huffman – SVM*

Figure 10　The line chart of the detection accuracy of different α,(a) Wang et al.[24],(b) Ren et al.[22],(c) *Huffman – SVM*

图 10 的实验结果表明, 在不同的隐写分析算法下, 联合失真函数权重α的变化情况是不同的。在受到不同的隐写分析算法攻击时, 选择合适的权重α, 使得隐写算法的安全性能够在不同的隐写分析情况下保持平衡, 达到整体的最优效果。

5.3 听觉感知实验结果和分析

本实验使用 ITU(International Telecommunication Union)标准[29]中的音频质量感知评估(Perceptual Evaluation of Audio Quality, PEAQ)客观地测量感知到的音频质量。音频质量感知评估(PEAQ)是该评价方法利用人耳主观感知特性计算出信号的掩蔽阈值和失真阈值, 然后采用人工神经网络融合出一个评价参数 ODG(Object Difference Grade)。它常用于音频隐写算法的不可感知性测量。ODG 值反映了音频质量的不可感知性。根据定义, ODG 的值通常为[–4,0]。ODG 值越接近 0, 表明含密音频和原始音频之间的听觉相似性越高。另外, 如果隐写算法造成的失真非常小, 并且含密音频和原始音频非常相似, ODG 值很可能是这样大于零值。

在实验中, 使用 PEAQ 对 7 组隐写算法的 5 种绝对嵌入率情况进行评估。从表 7 和图 11 可以看到, 尽管D_{cost2}使用心理声学模型 II 中的掩蔽阈值曲线设计失真函数, 但是, D_{cost2}在安全性上面并不是最优的, 这意味着它在某些方面的修改量可能多一些, 直接造成含密音频和原始音频的数据特征的差距相对更大一些。ODG 反映声音质量。在某种程度上, 对于含密音频来说, 修改量越小, 和原始音频的差距越小, 相似度越高, 则安全性越高, 含密音频的音质也越接近原始音频。表 7 的实验结果表明, 整体上讲, 本文提出的隐写算法相较 Yang 等[15]和 Yi 等[16]的不可感知性是有进一步提高的。

表 7 不同隐写算法的 ODG 值

Table 7 The ODG values of different schemes

隐写算法	*Payload*				
	1.6	2.1	3.0	4.0	6.0
Yang 等[15]	–0.087	–0.142	–0.179	–0.275	–0.342
Yang 等[15] - D_{cost1}	–0.023	–0.036	–0.093	–0.116	–0.268
Yang 等[15] - D_{cost2}	**–0.021**	**–0.027**	**–0.087**	**–0.197**	**–0.124**
Yi 等[16]	–0.095	–0.163	–0.186	–0.293	–0.351
Yi 等[16]- D_{cost1}	–0.043	–0.096	–0.123	–0.146	–0.273
Yi 等[16]- D_{cost2}	–0.038	–0.087	–0.105	–0.131	–0.211
本文算法	**–0.026**	**–0.041**	**–0.094**	**–0.127**	**–0.164**

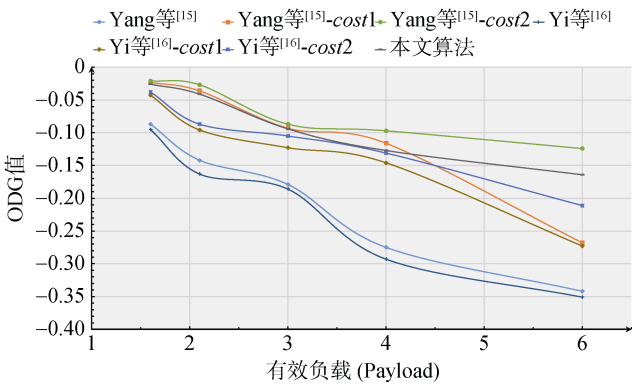


图 11 ODG 的折线图

Figure 11 The line chart of the ODG values

5.4 隐写容量

第三个实验部分是评估每种隐写算法的隐写容量。对 Yang 等[15]、Yi 等[16]以及本文提出的隐写算法, STCs[17]编码器中校验矩阵的高度和宽度等参数与 Yang 等[15]、Yi 等[16]中保持一致, 即阈值$T=3$。如表 8 所示, 本文所提的隐写算法的最大隐写容量约是 12 kbps。相比 Yang 等[15]、Yi 等[16], 本文所提方法的最大隐写容量并没有降低, 但是, 在相同嵌入容量下, 本文所提的方法在统计安全性和不可感知性均有提高。

表 8 隐写容量

Table 8 hiding capacity

隐写算法	平均嵌入容量 (kbps)	最大嵌入容量 (kbps)
本文算法	11.96	**12.08**
Yang 等[15]	11.69	12.10
Yi 等[16]	12.04	12.18

6 结论

本文提出了一种基于联合失真的 AAC 安全隐写算法, 该框架基于多方面的联合失真设计, 结合了 Huffman 码字统计分布特征和实际听觉阈值曲线特

征, 最后利用 STCs[17]实现自适应隐写, 能够解决现有隐写算法在统计安全性和不可感知性上的不足, 具有更好的提升效果。本文所设计的框架具有极好地扩展性, 可以直接应用到其他基于 AAC、MP3 的自适应隐写算法。本文算法的不足之处在于嵌入域是压缩域, 无法抵抗重压缩等恶意攻击。

后续工作将会结合其他的特征设计, 如帧内帧间系数的一阶、二阶的相关性等, 以及结合深度学习的方法探索更为有效地多重联合失真计算, 以及探索在不同的隐写分析情况下, 不同失真部分之间权重的关系和变化, 实现更安全、更高效的隐写方案。

致　谢　感谢易小伟博士分享 Yang 等[15]、Yi 等[16]实验中的音频数据库。在此向本文成文中给予指导的老师、提供帮助的同学和给本文提出建议的评审专家表示感谢。

参考文献

[1] Hussain M, Wahab A W A, Idris Y I B, et al. Image Steganography in Spatial Domain: A Survey[J]. *Signal Processing*: *Image Communication*, 2018, 65: 46-66.

[2] Ker A D, Bas P, Böhme R, et al. Moving Steganography and Steganalysis from the Laboratory into the Real World[C]. *The first ACM workshop on Information hiding and multimedia security*, 2013: 45-58.

[3] Li B, He J, Huang J, et al. A survey on image steganography and steganalysis[J]. *Journal of Information Hiding and Multimedia Signal Processing*, 2011, 2(2): 142-172.

[4] Shi Y Q, Chen C H, Chen W. A Markov Process Based Approach to Effective Attacking JPEG Steganography[C]. *Information Hiding*, 2007: 249-264.

[5] Bosi M, Brandenburg K, Quackenbush S, et al. ISO/IEC MPEG-2 advanced audio coding[S]. Journal of the Audio engineering society, 1997, 45(10): 789-814.

[6] Pinel J, Girin L, Baras C, et al. A High-Capacity Watermarking Technique for Audio Signals Based on MDCT-Domain Quantization[C]. *In Int. Congress on Acoustics*, 2010, 23.

[7] Wang Y J, Guo L, Wei Y F, et al. A Steganography Method for AAC Audio Based on Escape Sequences[C]. *2010 International Conference on Multimedia Information Networking and Security*, 2010: 841-845.

[8] Wang Y J, Guo L, Wang C P. Steganography Method for Advanced Audio Coding[J]. *Journal of Chinese Computer Systems*, 2011, 32(7): 1465-1468.
(王昱洁, 郭立, 王翠平. 一种以 AAC 压缩音频为载体的隐写方法[J]. *小型微型计算机系统*, 2011, 32(7): 1465-1468.)

[9] Yang Y Z, Wang Y T, Yi X W, et al. Defining Joint Embedding Distortion for Adaptive MP3 Steganography[C]. *The ACM Workshop on Information Hiding and Multimedia Security*, 2019: 14-24.

[10] Wei Y F, Guo L, Wang Y J. Controlling Bitrate Steganography on AAC Audio[C]. *2010 3rd International Congress on Image and Signal Processing*, 2010: 4373-4375.

[11] Xu S, Zhang P, Wang P, et al. Performance analysis of data hiding in MPEG-4 AAC audio[J]. *Tsinghua Science and Technology*, 2009, 14(1): 55-61.

[12] Zhu J, Wang R D, Li J, et al. A Huffman Coding Section-Based Steganography for AAC Audio[J]. *Information Technology Journal*, 2011, 10(10): 1983-1988.

[13] Zhu J, Wang R D, Yan D Q. The Sign Bits of Huffman Codeword-Based Steganography for AAC Audio[C]. *2010 International Conference on Multimedia Technology*, 2010: 1-4.

[14] Yan D Q, Wang R D, Zhang L G. A High Capacity MP3 Steganography Based on Huffman Coding[J]. *Journal of Sichuan University* (*Natural Science Edition*), 2011, 48(6): 1281-1286.
(严迪群, 王让定, 张力光. 基于 Huffman 编码的大容量 MP3 隐写算法[J]. *四川大学学报(自然科学版)*, 2011, 48(6): 1281-1286.)

[15] Yang K, Yi X W, Zhao X F, et al. Adaptive MP3 Steganography Using Equal Length Entropy Codes Substitution[C]. *Digital Forensics and Watermarking*, 2017: 202-216.

[16] Yi X W, Yang K, Zhao X F, et al. AHCM: Adaptive Huffman Code Mapping for Audio Steganography Based on Psychoacoustic Model[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(8): 2217-2231.

[17] Filler T, Judas J, Fridrich J. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes[J]. *IEEE Transactions on Information Forensics and Security*, 2011, 6(3): 920-935.

[18] Jin C, Wang R D, Yan D Q. Steganalysis of MP3Stego with Low Embedding-Rate Using Markov Feature[J]. *Multimedia Tools and Applications*, 2017, 76(5): 6143-6158.

[19] Jin C, Wang R D, Yan D Q, et al. A Novel Detection Scheme for MP3Stego with Low Payload[C]. *2014 IEEE China Summit & International Conference on Signal and Information Processing*, 2014: 602-606.

[20] Kuriakose R, Premalatha P. A Novel Method for MP3 Steganalysis[M]. Advances in Intelligent Systems and Computing. New Delhi: Springer India, 2014: 605-611.

[21] Qiao M Y, Sung A H, Liu Q Z. MP3 Audio Steganalysis[J]. *Information Sciences*, 2013, 231: 123-134.

[22] Ren Y Z, Xiong Q C, Wang L N. A Steganalysis Scheme for AAC Audio Based on MDCT Difference between Intra and Inter Frame[C]. *Digital Forensics and Watermarking*, 2017: 217-231.

[23] Lin Y Z, Wang R D, Yan D Q, et al. Audio Steganalysis with Improved Convolutional Neural Network[C]. *The ACM Workshop on Information Hiding and Multimedia Security*, 2019: 210-215.

[24] Wang Y T, Yang K, Yi X W, et al. CNN-Based Steganalysis of MP3 Steganography in the Entropy Code Domain[C]. *The 6th ACM Workshop on Information Hiding and Multimedia Security*, 2018: 55-65.

[25] Wang Y T, Yi X W, Zhao X F, et al. RHFCN: : Fully CNN-Based Steganalysis of MP3 with Rich High-Pass Filtering[C]. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019: 2627-2631.

[26] Ren Y Z, Liu D K, Xiong Q C, et al. Spec-ResNet: A General Au-

dio Steganalysis Scheme Based on Deep Residual Network of Spectrogram[EB/OL]. 2019: ArXiv Preprint ArXiv:1901.06838.

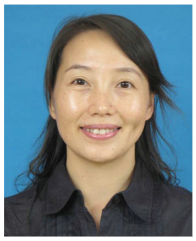
[27] Yu X M, Wang R D, Yan D Q. Detecting MP3Stego Using Calibrated Side Information Features[J]. *Journal of Software*, 2013, 8(10): 2628-2636.

[28] Petitcolas, F . Mp3stego. http://www.petitcolas.net/fabien/ steganography/mp3stego/. 1998.

[29] Thiede T, Treurniet WC, Bitto R, et al. PEAQ-The ITU standard for objective measurement of perceived audio quality[S]. Journal of the Audio Engineering Society, 2000, 48(1/2): 3-29.

蔡森 于 2018 年在武汉大学信息安全专业获得学士学位。现在武汉大学网络空间安全专业攻读硕士学位。研究领域为多媒体内容安全。研究兴趣包括: 音视频编码、多媒体内容安全。Email: limingong@whu.edu.cn

王丽娜 于 2001 年在东北大学获得博士学位。现任武汉大学国家网络安全学院教授。研究领域为多媒体安全、云计算安全、网络安全。研究兴趣包括: 隐写术、信隐写分析、 虚拟化、 数字信号处理与识别等。 Email: lnwang@whu.edu.cn



任延珍 于 2009 年在武汉大学通信专业获得伟士学位。现任武汉大学国家网络安全学院教授。研究领域为多媒体安全、信息隐藏、大数据安全。研究兴趣包括: 隐写术、信隐写分析、 大数据分析、 数字信号处理与识别等。 Email: renyz@whu.edu.cn