

一种新的用于音频认证数字水印算法^{*}

申军伟¹, 潘峰^{1,2}, 张明书¹, 肖海燕¹

(1. 武警工程大学 电子技术系 网络与信息安全武警部队重点实验室, 西安 710086; 2. 西安电子科技大学 网络信息安全教育部重点实验室, 西安 710071)

摘要: 为确保数字音频数据感知意义上的完整性, 提出了一种新的音频认证水印算法。该算法把原始音频帧, 每帧分为两节, 在第一节时域中嵌入同步码, 将第二节进行 DCT 变换, 计算邻帧关系向量, 并将其作为水印信息嵌入其他帧第二节的 DCT 中频系数。认证时, 计算邻帧关系误差, 据事先设定的阈值来判断音频是否被篡改, 若遭受篡改, 可进行篡改定位。仿真实验表明, 该算法在无原始水印参与的情况下, 不仅可以实现音频媒体的完整性认证, 而且篡改定位准确。

关键词: 数字水印; 音频认证; 音频水印; 离散余弦变换; 篡改定位

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)10-3850-04

doi:10.3969/j.issn.1001-3695.2012.10.065

New digital watermarking algorithm for audio authentication

SHEN Jun-wei¹, PAN Feng^{1,2}, ZHANG Ming-shu¹, XIAO Hai-yan¹

(1. Key Laboratory of Network & Information Security of APF, Dept. of Electronic Technology, Engineering University of APF, Xi'an 710086, China; 2. Key Laboratory of Network & Information Security of Ministry of Education, Xidian University, Xi'an 710071, China)

Abstract: This paper proposed a new digital watermarking algorithm for audio authentication to protect the audio integrity in the perceptual sense. Firstly, it framed an original digital audio and then divided each frame into two parts. Secondly, with the spatial watermarking technique, embedded synchronization code into the first part. And then performed the DCT and calculated the AFRV (adjacent-frame relationship vector) used as the watermark information on the second part. Finally embedded the watermark bit into the DCT intermediate frequency coefficients of other frame's second part. When authenticated, the algorithm could determine whether the audio had been malicious tampering by computing AFRE (adjacent-frame relationship error) and according to the pre-set threshold without the help from the origin watermark. If the data were subjected to tampering, it could find out the data error localization. Simulation results show that the proposed algorithm can not only realize the audio media integrity authentication, but also accurately identify the tamper location.

Key words: digital watermarking; audio authentication; audio watermarking; discrete cosine transform (DCT); tamper localization

0 引言

随着数字化音频技术的不断发展, 音频数字水印已经成为一个非常热门的研究领域。数字音频认证技术可实现对音乐、语音等音频数据的完整性/真实性进行保护, 可确保接收到的音频数据在传送过程中未经第三方恶意篡改, 即在人类感知意义上与原始音频是完全相同的^[1]。

目前学术界关于数字音频认证的文献还很有限。Radhakrishnan 等人^[2]根据“两个听觉质量相似音频之间的掩蔽曲线几乎是一样的”原理, 首先计算音频掩蔽曲线的 hash 函数值, 然后采用已知的信息隐藏方法将其作为水印嵌入到音频信号中。Chen 等人^[3]提出一种使用小波包分解和最优树选取的基于质量的认证算法, 以音频分段的小波包分解系数作为与质量相关的鲁棒特征。小波包系数的选取采用一种在最小熵意义下的最优树算法, 该算法可有效识别随机剪切和局部区域篡改操作。张敏情等人^[4]提出一种基于局部保持映射的音频

数字签名算法, 在经过小波包变换和线性预测形成音频特征后, 利用 LPP 降维, 再通过混沌置乱来得到最终签名。以上算法的共同缺点是, 在认证失败的情况下, 作品数据作为一个整体被视为“无用”数据而被抛弃。但导致认证失败的原因很可能只是由于一小部分数据被篡改。如果能找到作品被篡改的位置, 无疑可保留绝大部分未被篡改部分的价值^[5], 进而加以利用。

本文提出了一种新的用于音频认证的数字水印算法。该算法采用分段分节处理, 通过时域频域变换, 将同步码及邻帧关系向量作为水印信息进行嵌入。认证时算法可对篡改进行准确定位。实验结果表明, 该算法科学有效。

1 理论基础

1.1 相关定义

将原始音频记为 A , 嵌入水印后的音频记为 A' 。首先将 A 分为 m 帧 (每帧长度为 l), 然后对每一帧 $A(i)$ ($i = 0, 1, \dots$,

收稿日期: 2012-03-07; 修回日期: 2012-04-18 基金项目: 国家自然科学基金资助项目 (60842006)

作者简介: 申军伟 (1984-), 男, 内蒙古丰镇人, 讲师, 硕士研究生, 主要研究方向为数字水印 (jhpklwfnjsjw@163.com); 潘峰 (1967-), 男, 教授, 硕士, 主要研究方向为密码学、多媒体安全; 张明书 (1978-), 男, 讲师, 硕士, 主要研究方向为数字水印; 肖海燕 (1985-), 女, 助教, 硕士, 主要研究方向为数字水印。

$m-1$) 进行离散余弦 DCT 变换,得到变换系数 $CA(i)(j)$ ($j=0,1,\dots,l-1$)。

定义 1 序列 $CA(i)(j)$ 的前 p 个数 $CA(i)(k)$ ($k=0,1,\dots,p-1, 0 \leq k < l$) 称为内容系数, $CA(i)_p = [CA(i)(0), CA(i)(1), \dots, CA(i)(p-1)]$ 称为内容向量。内容向量包含了音频帧 DCT 变换后的低频系数,因此它反映了音频帧数据的主要内容。

定义 2 $CA(i)_p$ 和 $CA(i+1)_p$ 分别为音频第 i 帧和第 $i+1$ 帧的内容向量,则映射函数 $f(CA(i)_p, CA(i+1)_p)$ 称为邻帧关系函数(具体描述见 1.2 节),映射结果 $R_{i,i+1} = f(CA(i)_p, CA(i+1)_p)$ 称为邻帧关系向量,其中的元素称为邻帧关系系数。

本文算法就是将邻帧关系向量 $R_{i,i+1}$ 作为水印信息嵌入到原始音频第 i' 帧 $A(i')$ 的 DCT 中频系数中,其中 $i' \neq i, i' \neq i \pm 1$ 。认证时,对于相邻两帧 $A'(i)$ 和 $A'(i+1)$,计算其邻帧关系向量得到 $R'_{i,i+1}$,从第 i' 帧中提取水印信息得到 $R''_{i,i+1}$,并有 $R''_{i,i+1} = R_{i,i+1} + \Delta R$ 。这里 ΔR 表示由误码、篡改等原因导致的误差。

定义 3 $\delta(A'(i), A'(i+1)) = |R'_{i,i+1} - R''_{i,i+1}|$ 为邻帧关系误差。

记 $\theta_1 = |R'_{i,i+1} - R_{i,i+1}|$, $\theta_2 = |\Delta R|$, 则 θ_1 反映了原始音频帧 $A(i)$ 和 $A(i+1)$ 在经历某些操作之后内容变化程度, θ_2 反映了原始水印信息在经历这些操作之后的失真程度。由于 $|R'_{i,i+1} - R''_{i,i+1}| \leq |R'_{i,i+1} - R_{i,i+1}| + |\Delta R| = \theta_1 + \theta_2$, 这样就可以事先设定一个误差阈值 $\theta \leq \theta_1 + \theta_2$, 用于检测音频是否遭遇了超过许可的操作或篡改。

1.2 构造邻帧关系向量

由于音频信号在时域上的一个短时间段内(一般为 20 ms 左右)被认为是相对稳定且前后相关的^[6],因此本文采用著名统计学家卡尔·皮尔逊设计的统计指标相关系数来构造邻帧关系向量。定义 2 中的邻帧关系函数 $f(x, y)$ 可分两个步骤:

a) 由式(1)计算邻帧相关系数

$$\rho = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (1)$$

由相关系数的性质可知: ρ 的值介于 -1 与 +1 之间,即 $|\rho| \leq 1$ 。

b) 将 ρ 的值转换为八位的二进制原码,即整数位为符号位,其余七位为小数位。这个八位的二进制序列就是本文构造的邻帧关系向量 $R_{i,i+1}$ 。

2 水印嵌入及认证

2.1 水印生成

将原始音频 A 按照长度 l 分帧,再将每帧分成 L 和 R 两小节,长度分别为 l_1 和 l_2 ,即 $l = l_1 + l_2$ 。对每帧数据 $A(i)$ 的 R 节进行 DCT 变换,计算邻帧关系向量 $R_{i,i+1}$ 作为水印信息。为提高安全性,将所有 $R_{i,i+1}$ 进行置乱加密,得到待嵌入的水印序列 $w_{i,i+1}$ 。

2.2 水印嵌入

水印嵌入算法分为以下五个步骤:

a) 为抵抗同步攻击,在每帧的 L 节时域中嵌入 13 位的巴克码,形成 L' 。

b) 将 $w_{i,i+1}$ 嵌入到第 i' 帧的 R 节的 DCT 中频系数中。 i' 由 i 来确定,即 $i' = \varphi(i)$, φ 作为密钥函数,以提高系统安全性。

c) 具体的嵌入方法采用均值量化^[7]的方法:

(a) 计算待修改的 DCT 系数 ca 的统计平均值 μ 。

(b) 修改 ca 为 ca' 。修改规则为: $ca' = ca + [\mu' - \mu]$ 。其中:当水印比特与 $\lfloor \mu/\Delta \rfloor$ 的奇偶性相同时, $\mu' = \lfloor \mu/\Delta \rfloor \times \Delta + \Delta/2$; 当水印比特与 $\lfloor \mu/\Delta \rfloor$ 的奇偶性不同时:

$$\mu' = \begin{cases} (\lfloor \mu/\Delta \rfloor - 1) \times \Delta + \Delta/2, & \mu \in [\lfloor \mu/\Delta \rfloor \times \Delta, \lfloor \mu/\Delta \rfloor \times \Delta + \Delta/2) \\ (\lfloor \mu/\Delta \rfloor + 1) \times \Delta + \Delta/2, & \mu \in [\lfloor \mu/\Delta \rfloor \times \Delta + \Delta/2, \lfloor \mu/\Delta \rfloor \times \Delta + \Delta) \end{cases}$$

Δ 表示均值量化的步长, $\lfloor \cdot \rfloor$ 表示取下整数。

d) 嵌入完成后,对 ca' 做 DCT 逆变换到时域生成 R' ,并连接到 L' 之后,形成一帧的处理结果。

e) 对每一帧做步骤 b) ~ d)。最后经合并形成完整的带水印的时域音频信号。

2.3 认证过程

认证时,计算邻帧关系向量 $R'_{i,i+1}$,并提取水印信息 $R''_{i,i+1}$ 。其步骤如下:

a) 根据巴克码确定各帧起始位置,对音频信号进行分帧,并找到每帧的 R 节数据。

b) 对每帧的 R 节数据进行 DCT 变换,计算其邻帧关系向量 $R'_{i,i+1}$,同时按式(2)提取水印信息 $w'_{i,i+1}$,并对其反置乱解密得到 $R''_{i,i+1}$ 。

$$w_b = \begin{cases} 0 & \lfloor ca' * \Delta \rfloor \bmod 2 = 0 \\ 1 & \lfloor ca' * \Delta \rfloor \bmod 2 = 1 \end{cases} \quad (2)$$

其中: w_b 表示提取的水印比特。

c) 对每帧计算邻帧关系误差 $\delta(A'(i), A'(i+1))$,并将其结果转换为十进制小数 $d(i)$,同时按式(3)记入向量 M 。

$$M(i) = \begin{cases} 0 & d(i) < \theta \\ 1 & d(i) \geq \theta \end{cases} \quad (3)$$

若 M 中所有的元素均为 0,则认证通过;否则认证失败。导致认证失败的情况主要有三种:(a) 待检音频作品与原始音频没有关系;(b) 数据传输处理中出现误码;(c) 待检音频作品被篡改,且篡改程度超出了事先设定的阈值。

2.4 篡改定位

在认证失败的情况下,可通过下面的规则进行篡改定位:

据上述算法过程,若第 i 帧被篡改,有 $M(i-1) = 1$, $M(i) = 1$ 和 $M(\varphi'(i)) = 1$ (这里 φ' 表示 φ 的逆)。反过来,若 $M(i-1)$ 、 $M(i)$ 和 $M(\varphi'(i))$ 的值不同时为 1,则第 i 帧一定未被篡改,此时重置 $M(i) = 0$ 。按照此规则对向量 M 进行多次遍历,便可定位所有被篡改的帧,即 M 中所有非 0 分量对应的帧。

3 仿真实验及分析

用 MATLAB 2010b 进行仿真实验,采用 16 位量化精度、44.1 kHz 采样率的单声道数字音频信号。将该算法应用到不同种类的音乐片断进行实验,包括流行歌曲(菊花台)、摇滚(海阔天空)、萨克斯独奏(回家)、钢琴独奏(秋日私语)、小提琴协奏曲(梁祝)和吉他独奏(爱的罗曼斯),每个音频片断长度为 10 s。萨克斯独奏(回家)的原始波形、加入水印后的波形及两者差的波形如图 1 所示。从图 1(a)(b)中看不出原始音频波形和加入水印波形的差别;(c)中波形差也反映了两者的

的区别很小。

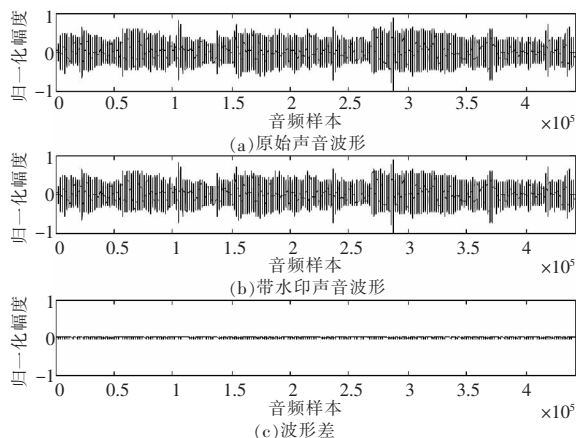


图1 原始波形、带水印波形及两者的波形差

3.1 主客观听觉测试

采用 subjective difference grades (SDG) 进行非正式主观音频质量测试。SDG 的含义如表1所示。将原始音频和带水印的音频分别提供给10个测试者,让他们对每一种音频按照SDG打分,然后取平均值,结果如表2所示。

表1 主观听觉测试区分度

等级	描述
0.0	不可感觉
-1.0	可感觉但不刺耳
-2.0	轻微刺耳
-3.0	刺耳
-4.0	非常刺耳

表2 主观听觉测试结果

种类	流行	摇滚	萨克斯	钢琴	小提琴	吉他
SDG	-0.1	0.0	0.0	-0.1	0.0	0.0

从表2中可以看出,测试结果基本上一样都接近于0,即原始音频和带水印的音频在听觉质量上是十分相似的。嵌入水印后的音频信号的峰值信噪比如表3所示。

表3 不同种类音乐的峰值信噪比

种类	流行	摇滚	萨克斯	钢琴	小提琴	吉他
PSNR	40.79	48.33	49.96	39.68	44.53	45.81

从实验结果可以看出,该方案无论在主观上还是客观上都具有很好的不可感知性。

3.2 攻击测试

为检测算法对常规信号处理具有鲁棒性,对恶意操作具有敏感性,将含水印不同种类音频分别进行了如下处理:

a) 常规信号处理

(a) 叠加噪声。在时域中加入高斯白噪声,使信噪比为低于20 dB。

(b) 低通滤波。截止频率为4 kHz。

(c) 重量化。先将信号由16位量化成8位,再由8位量化成16位。

(d) 重采样。将音频信号采样率下降为22.05 kHz,再还原为原来的44.1 kHz。

(e) MP3 压缩/解压缩。将音频信号分别进行128、64和48 kbps的MP3压缩,再将压缩后的文件解压成wav文件。

b) 恶意操作

(f) 回声。对音频信号进行回声攻击。

(g) 剪切。随机剪切20帧(约0.46 s)替换为白噪声。

(h) 变速。采用Cool Edit中的Fast Talker将音频按接合频率为52 Hz,叠加34%、比率为220进行变速处理。

(i) 变调。采用Cool Edit中的Down a Whole Step将波形长度由10.0 s改变为11 s。

(j) 调幅。对音频进行120%的调幅处理。

经过上述常规信号处理后,算得归一化相关系数NC如表4所示。

表4 常规信号处理后的归一化相关系数

攻击种类	流行(NC)	摇滚(NC)	萨克斯(NC)	钢琴(NC)	小提琴(NC)	吉他(NC)
无攻击	1	1	1	1	1	1
(1)	0.980 0	0.983 5	0.984 0	0.982 0	0.977 9	0.981 1
(2)	0.967 8	0.974 1	0.976 7	0.965 8	0.972 6	0.972 8
(3)	0.980 6	0.988 9	0.997 6	0.990 0	0.987 8	0.986 0
(4)	0.991 0	0.991 2	0.994 9	0.988 5	0.990 5	0.988 1
(5-1)	0.976 3	0.970 0	0.973 8	0.965 6	0.976 5	0.984 7
(5-2)	0.945 2	0.950 4	0.957 9	0.948 9	0.945 2	0.952 7
(5-3)	0.915 8	0.921 9	0.931 0	0.922 2	0.915 3	0.912 1
(6)	0.525 6	0.656 3	0.680 6	0.535 8	0.548 3	0.621 9
(7)	0.645 8	0.676 3	0.685 2	0.657 5	0.640 0	0.630 5
(8)	0.343 2	0.380 5	0.360 7	0.326 8	0.298 6	0.309 7
(9)	0.286 3	0.301 0	0.338 2	0.351 2	0.318 5	0.327 4
(10)	0.386 2	0.452 7	0.440 9	0.428 7	0.436 2	0.409 9

从表4中可以看出,含水印音频文件在经过保持内容操作(即常规信号处理如重采样、重量化、低通滤波、MP3压缩)后,归一化相关系数都在0.9以上,因此反映水印失真程度的 θ_2 的值比较小;而对于经过改变音频内容操作(即恶意操作如剪切、变调、变速处理)后,归一化相关系数相对较小,这样就使得 θ_2 的值相对较大,从而使得邻帧关系误差 θ 的值相对较大。

3.3 邻帧关系误差阈值 θ 的估计

由定义3中邻帧关系误差阈值 θ 主要取决于下面两个因素:a)允许内容相对变化的程度 θ_1 ;b)允许的作为水印嵌入的邻帧关系系数的失真程度 θ_2 。在这两个因素中,b)可以通过大量的实验获得,比较容易确定,而a)较难确定,因为涉及到界定多大量的内容改变是恶意和非恶意的。本文中通过实验的方法,对含水印音频进行如3.2节中的攻击测试。首先来看嵌入水印对 θ_1 的影响。因 $\theta_1 = |R'_{i,i+1} - R_{i,i+1}|$,将计算结果转换为十进制小数。实际上 θ_1 相当于 $|\rho' - \rho|$,其中 ρ' 为嵌入水印后的邻帧相关系数。图2为嵌入水印对相关系数的影响(为了显示清楚,只取了100帧)。

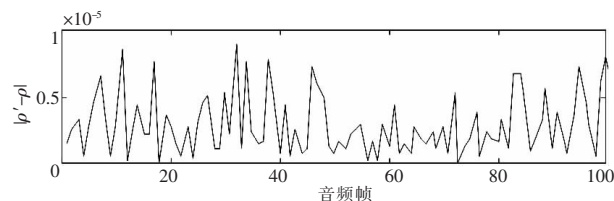


图2 嵌入水印对相关系数的影响

从图2中可以看出嵌入水印对 θ_1 的影响小于 10^{-5} ,基本上可以忽略不计。接下来分别来看常规信号处理和恶意操作对 θ_1 的影响,如图3所示(为了显示清楚,只取了50帧,对于MP3压缩/解压缩对相关系数的影响结果与之类似,不再画图)。

从图3(a)中可以看出常规信号处理对相关系数的影响小于0.02,而从(b)中可以看出恶意操作对相关系数的影响都大于0.02,因此通过实验可以取允许内容相对变化的程度 θ_1 为0.02。而 $\theta_2 = |R''_{i,i+1} - R_{i,i+1}|$,用同样的实验方法获得其取值为0.17。因此,可以取邻帧关系误差阈值 θ 的参考值为0.19,在实际应用中可以调节 θ ,以获得更好的效果。

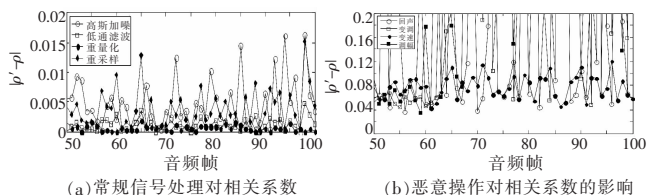


图 3 常规信号处理及恶意操作对相关系数的影响

3.4 恶意篡改定位能力

为验证算法对恶意篡改的定位能力,对图 1(b)所示的含水印音频实施了三种不同的恶意篡改攻击。攻击 1 为替换部分音频帧为高斯白噪声;攻击 2 为用部分不含水印的音频帧替换含水印的相应的音频帧;攻击 3 为替换部分音频帧为静音。图 4(a)为图 1(b)所示的含水印音频的前 40 000 个样点;(b)是实施攻击 1 后的波形图,将第 5~9 帧替换;(c)是实施攻击 2 后的波形图,第 10~14 帧替换;(d)是实施攻击 3 后的波形图,第 15~19 帧替换。

图 5(a)~(c)分别为图 4(b)~(d)对应的音频帧篡改定位结果。向量 M 中元素为 1 对应的帧为被恶意篡改的帧,为 0 对应的帧为未被恶意篡改的帧。

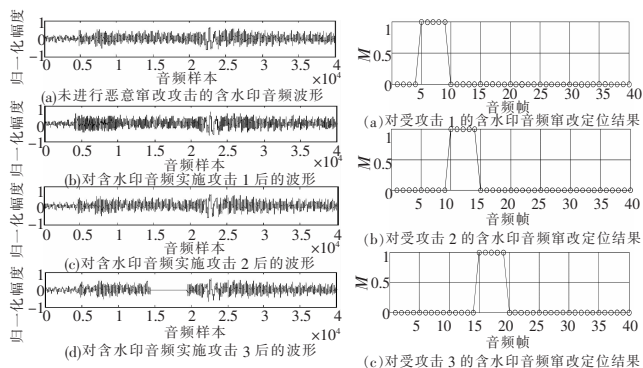


图 4 对含水印的音频实施不同攻击后的结果

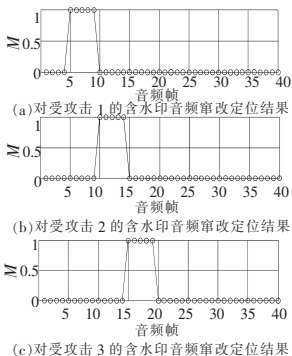


图 5 对不同攻击的含水印音频篡改定位结果

由图 5 可知,算法具有准确的篡改定位能力。

4 结束语

本文利用音频短时平稳的特征,将相邻两帧的皮尔逊相关系数经过变换后嵌入到选定帧的 DCT 中频系数中,认证时计算邻帧关系误差,看是否超过事先设定的阈值,以判定音频文件是否被篡改。若已遭受篡改,还能进行篡改定位。仿真实验表明,该算法可以实现音频媒体的完整性认证,在认证和篡改定位的过程中,均不需要原始水印参与,便可准确定位被篡改区域(尤其受到裁剪攻击时)。最后,尽管该算法易于实现,具有较高的理论与应用价值,但针对算法中的邻帧关系误差阈值的选取有待于进一步研究。

参考文献:

- [1] 李伟,汪花蓉,李晓强,等. 数字音频认证研究综述[J]. 计算机科学, 2009,36(10):21-24.
- [2] RADHAKRISHNAN R, MEMON N. Audio content authentication based on psycho-acoustic model[C]//Proc of SPEC Security and Watermarking of Multimedia Contents. Bellingham, WA:SPIE,2002:110-117.
- [3] CHEN Fan, LI Wei, LI Xiao-qiang. Audio quality-based authentication using wavelet packet decomposition and best tree selection[C]//Proc of Intelligent Information Hiding and Multimedia Signal Processing. Washington DC:IEEE Computer Society,2008:1265-1268.
- [4] 张敏情,杨晓元,吴翔. 基于局部保持映射的音频数字签名算法[J]. 解放军理工大学学报:自然科学版,2009,10(3):248-252.
- [5] 唐胜. 多媒体数字签名技术研究[D]. 北京:中国科学院,2006.
- [6] 汝学民. 音频密写与分析技术[D]. 杭州:浙江大学,2006.
- [7] 马天笑,王向阳,杨红颖. 一种基于均值量化的抗去同步攻击数字水印算法[J]. 计算机科学,2009,36(4):258-259.