一种抗裁剪的半脆弱音频水印算法

赵 红 [1,2] 沈东升 [1,2] 朱元辉 [1,2]

摘 要 半脆弱水印不但可以容忍一定程度的常规信号处理操作, 而且可以检测出对多媒体数据的恶意篡改并定位篡改区域. 本文提出一种抗裁剪的半脆弱音频水印算法, 具有以下特点: 1) 充分利用多级置乱技术准确定位篡改区域, 最大限度改善被篡改后水印的视觉效果; 2) 在篡改定位过程中, 无需原始水印参与; 3) 利用水印的归一化相关系数, 可进行完整性认证并给出篡改程度的度量. 实验结果表明, 该算法计算简单, 有很好的抗裁剪性能, 同时对篡改的定位也很精确.

关键词 半脆弱水印, 数字水印, 裁剪攻击
中图分类号 TP309

A Semi-fragile Audio Watermarking Against Shearing

ZHAO Hong[1,2] SHEN Dong-Sheng[1,2] ZHU Yuan-Hui[1,2]

Abstract Semi-fragile watermarking can tolerate general signal processing, detect any spiteful tamper on marked multimedia data, and locate spatial regions within an altered image. A semi-fragile audio watermarking against shearing is proposed in this paper. The features of the proposed algorithm are as follows: 1) Full advantage of multilevel scrambling is taken to locate tampered region; 2) The algorithm can locate tampered regions without the help from the original watermark; 3) By using the normalized cross-correlation, the proposed scheme can not only authenticate the integrity of the audio signal but also evaluate the temper degree. The experimental results show that the proposed algorithm has the advantages such as simple computation complexity, good robustness against shearing attack, and accurate location for tamper.

Key words Semi-fragile watermarking, digital watermarking, shearing attack

随着多媒体技术的飞速发展和日益完善, 任何人都可以方便地更改数字媒体, 如音乐、视频或图像等, 这使得人们对数字媒体的完整性、媒体内容的真实性产生怀疑, 认证水印正是应这一需要而发展起来的. 近年来, 采用数字水印技术进行网络传输中的图像认证, 音频认证, 已经成为当前信息认证领域的研究热点. 依据对篡改敏感性的不同, 用于认证的数字水印分为脆弱性和半脆弱性两类, 由于任何对媒体信息的更改都会破坏其完整性, 使水印检测失败, 因此脆弱性水印主要用于完全级认证, 而半脆弱性水印则主要用于内容级认证[1]. 近年来, 图像水印和视频水印技术研究取得了很大进展, 但数字音频水印技术的研究却相对滞后, 而可用于认证的音频水印方案更是匮乏[2−8], 但随着数字音频技术的不断发展及数字音乐作品的日益普及, 音乐作品版权的保护变得愈加重要. 现代数字音频编辑和处理技术使得高质量伪造可以以极低的代价进行, 音频数据的语义可以通过简单地重排或去掉几个小的片断进行改变, 因此只依靠人的听觉测试来判断音频数据的完整性是完全不够的[1]. 由此为了得到安全的音频应用, 篡改检测变得愈发重要.

半脆弱水印技术是数字水印技术的一个重要分支, 其主要的优势在于: 能检测出多媒体数据是否被恶意篡改, 并可以定位篡改区域. 定位方式主要分为两大类: 1) 数字媒体认证时, 首先从测试对象中提取水印信息, 然后, 将其与原始水印信息相比较来确定测试对象是否被篡改, 若发生篡改, 应给出有关测试对象篡改的详细信息; 2) 将数字媒体的内容或特征作为水印信息嵌入到数字媒体中, 并确保水印的嵌入不会改变数字媒体的这些内容或特征, 在数字媒体认证时, 只需将提取的水印信息与测试对象的内容或特征进行比较, 而不必另外提供原始水印信息. 文献 [2−3] 均采用了第 1) 类方法, 将提取的水印图像与原始水印图像相比较来确定被篡改的位置, 根据两者的差图中稀疏点与稠密点数量的比值来判断攻击类别, 能达到很好的定位效果, 但在判断过程中, 检测方需设定一个非负的门限值, 此门限值的高低决定了判断篡改攻击的严格性及准确性; 同时稀疏点与稠密点的定义也直接影响了篡改定位的准确性, 目前绝大多数水印为有意义的二值水印图像, 而

收稿日期 2006-12-15 收修改稿日期 2007-06-26
Received December 15, 2006; in revised form June 26, 2007
福建省自然科学基金 (2007J0194), 福建省教育厅科技项目 (JA06020, JA07159) 资助
Supported by Natural Science Foundation of Fujian Province (2007J0194) and Scientific Technology Project of Education Department of Fujian Province (JA06020, JA07159)
1. 漳州师范学院计算机科学与工程系 漳州 363000 2. 漳州师范学院信息安全技术福建省高等学校重点实验室 漳州 363000
1. Department of Computer Science and Engineering, Zhangzhou Normal College, Zhangzhou 363000 2. Key Laboratory of Information Security Technology, Zhangzhou Normal College, Zhangzhou 363000
DOI: 10.3724/SP.J.1004.2008.00647

水印图像中黑色点与白色点的分布各自较为集中, 对于不同的水印图像分布方式也各有区别, 所以找出一种适当的稀疏点定义难度较大.

本文针对上述半脆弱水印技术中篡改定位能力的缺陷与不足, 提出一种新的半脆弱数字音频水印算法, 其创新之处在于, 将置乱技术应用于篡改定位, 在无需原始水印参与的情况下, 能够保证准确定位篡改区域, 有效提高所提取水印的视觉效果; 同时该算法对于常规信号处理 (特别是 MP3 压缩) 具有很强的鲁棒性, 另外, 本文还有计算简单, 易于实现等特点.

1 置乱技术在定位篡改区域中的应用

置乱技术目前已经广泛应用于数字水印技术之中, 引入置乱技术虽然不能加强水印的鲁棒性, 但可以通过置乱技术有效地消除二值水印图像的像素空间相关性, 将错误信息扩散分布, 避免遭受攻击后的水印发生大面积信息丢失的情况, 提高数字水印的主观鲁棒性, 即视觉效果.

假设二值水印图像为: $W = \{w(i,j), 1 \le i \le M, 1 \le j \le N\}$, 其中, $w(i,j) \in \{0,1\}$ 代表二值水印图像的第 i 行、第 j 列像素灰度值, M, N 分别为水印图像 W 的行数与列数. 为了抵抗任意剪裁攻击, 数字水印在嵌入前要先进行预处理, 本文采用经典直观的 Arnold 变换图像置乱技术[9] 来对水印图像 W 进行加密调制, 得到置乱后的水印图案 W'.

接下来, 对水印进行分段处理, 并根据定位被篡改区域所要求的精度来确定水印的段长: 段长越短, 对音频分段的段数越多, 定位的精度越高, 但嵌入的附加信息量越大, 相应的, 算法执行时间越长; 而若段长越长, 定位的精度越低, 但嵌入的附加信息量越少, 算法执行时间越短. 设每一段含水印的像素数总数为 S_{total}, 则 $W = \{w(k,l), 1 \le k \le M \times N/S_{total}, 1 \le l \le S_{total}\}$, 其中 $w(k,l) \in \{0,1\}$ 代表分段后的二值水印图像的第 k 段的第 l 个像素灰度值, 设图像上的像素值为 1, 即表现为白色点, 反之像素值为 0, 表现为黑色点, 则可得到原始水印图像 W 中, 第 k 段中白色像素点总数为

$$S(k)_{white} = \sum_{l=1}^{S_{total}} w(k,l) \tag{1}$$

置乱后水印图像 W' 中, 第 k 段中白色像素点总数为

$$S'(k)_{white} = \sum_{l=1}^{S_{total}} w'(k,l) \tag{2}$$

继而, 分别计算出每一段水印 (置乱前后) 中白色点数占总点数的比例, $Prop(k) = S(k)_{white}/S_{total}$, $Prop'(k) = S'(k)_{white}/S_{total}$.

表 1 置乱前后的水印图像

Table 1 The watermarks and the scrambled images

Flower	Scrambled flower	Protect	Scrambled protect
			

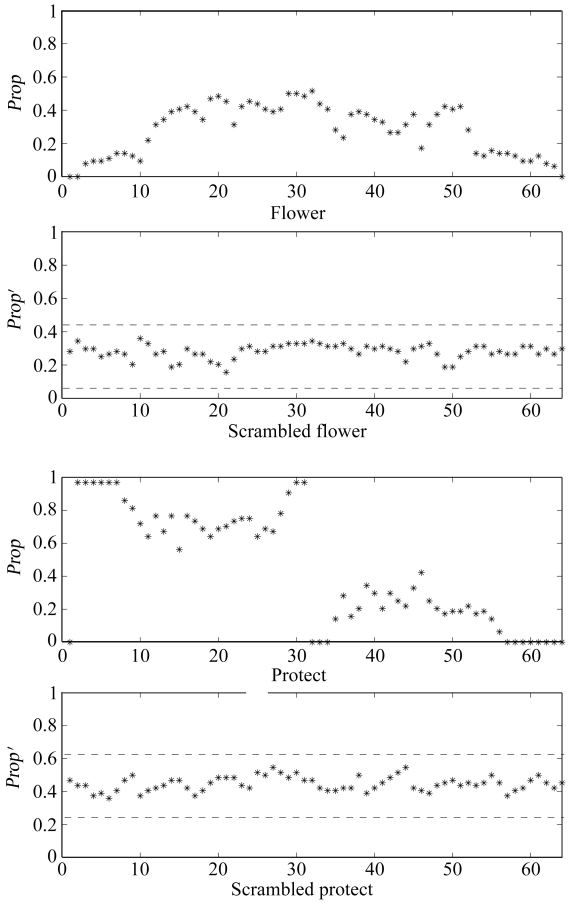

图 1 置乱前后每一行水印中白色点所占比例

Fig. 1 The proportions of the white pixels and the total pixels

从表 1 中, 不难看出, 有意义的二值水印图像中黑色点与白色点的分布各自较为集中, 对于不同的水印图像分布方式也各有区别, 置乱技术消除了水印图像的空间相关性, 使得水印图像中的黑色点及白色点近似均匀分布. 从图 1 可进一步得出此结论, 对于任意一种水印图像, 黑色点与白色点的分布随水印图像的不同而不同, 即白 (黑) 色点所占比例 $Prop$ 值是无规律可寻的; 而对于其置乱图像, 每一段的白色点与黑色点个数所占比例相差不大, 即白

(黑) 色点所占比例 $Prop'$ 落在一个比较窄的 “安全区域” 上, 而若 $Prop'$ 异常, 即超出此 “安全区域” 则认为有恶意篡改发生, 其定位过程无需原始水印图像. “安全区域” $\Re$ 的定义如下

$$X - \triangle \le \Re \le X + \triangle \tag{3}$$

$$X = S_{total}/(M \times N) \sum_{k=1}^{M \times N/S_{total}} Prop'(k) \tag{4}$$

其中, $\triangle$ 为安全区域调节因子, 可通过 $\triangle$ 调节 “安全区域” 的大小. 当 $Prop'(k) \notin \Re$, 认为与之对应的第 k 段音频有恶意篡改发生.

2　水印嵌入与检测的基本原理

2.1　水印的组成

第 k 段水印 $W^*(k)$ 由水印段号 k 的二进制形式 $P(k)$ 与置乱后的水印图像信息 $W'(k)$ 两部分组成, 其中 $P(k)$ 的长度定义为 P_{total}, $W'(k)$ 的长度定义为 S_{total}, 为了保证段号的稳定性, 采用段号重复嵌入的策略, 即将 $P(k)$ 重复 x 次嵌入到音频中. 则第 k 段水印为

$$W^*(k) = P(k) + W'(k) = \{w^*(k,l)\} \tag{5}$$

其中, $1 \le l \le S_{total} + P_{total} \times x, 1 \le k \le M \times N/S_{total}$.

2.2　数字水印嵌入的基本原理

假设原始数字音频信号为 $A = \{a(i), 1 \le i \le Length\}$. 其中, $Length$ 为音频数据的个数, $a(i) \in \{0, 1, 2, \cdots, (2^p - 1)\}$ 是第 i 个音频数据的幅度值, p 为表示每个数据所使用的比特数.

水印的嵌入过程如图 2 所示, 首先将音频 A 根据图像水印的段数 k 进行分段, 然后再将每个音频段划分成前后两个部分以进行信息嵌入, 其中, 前面部分用于空域嵌入同步信号, 为了保证半脆弱水印对于常规信号处理的鲁棒性, 采用文献 [4] 的方法, 将同步码嵌入到多个采样值的统计平均值上. 后面部分经 H 层小波变换后取其部分低频系数 $a(k,l)^H$, 将水印信息 $W^*(k)$ 量化嵌入其中 (图 3 给出了嵌入信息的结构), 得到 $a'(k,l)^H$, 并最终经过反变换得到嵌入水印后的音频段. 嵌入公式如下

$$a'(k,l)^H = \begin{cases} Q(k,l) \times d + d/2, \\ \quad 若\ \mathrm{mod}(Q(k,l),2) = w^*(k,l) \\ Q(k,l) \times d - d/2, \\ \quad 若\ \mathrm{mod}(Q(k,l),2) \ne w^*(k,l) \end{cases} \tag{6}$$

$$Q(k,l) = \lfloor a(k,l)^H/d \rfloor \tag{7}$$

其中 d 为量化步长, $1 \le l \le S_{total} + P_{total} \times x, 1 \le k \le M \times N/S_{total}$.

2.3　数字水印的检测

水印的检测及篡改数据的定位是本文的水印检测算法完成的两个主要任务. 本算法属于盲水印算法, 即检测数字水印信号时不需要原始音频载体的参与. 数字水印的检测过程如下:

1) 采用文献 [4] 中的帧同步码逐位比较方式查找同步码, 定位检测器的起始位置 B.

2) 对 B 后的音频数据段 $A^*(k)$ 进行 H 级 DWT, 得到小波系数, 取其中部分低频系数 $A^{*H} = a^*(k,l)^H$.

3) 提取行号及水印信息 $W''(k)$, 提取公式如下

$$W''(k) = \{w''(k,l)\} = \lfloor a^*(k,l)^H/d \rfloor \mathrm{mod}\ 2 \tag{8}$$

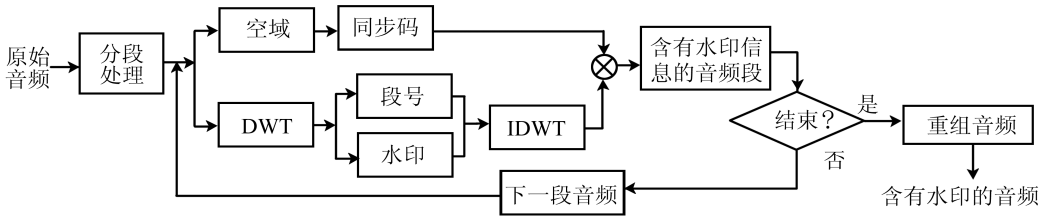

图 2　水印的嵌入原理

Fig. 2　Watermark embedding scheme

	音频段 k-1			音频段 k			音频段 k+1			
… …	空域	频域		空域	频域		空域	频域		… …
… …	同步码	段号	水印信息	同步码	段号	水印信息	同步码	段号	水印信息	… …

图 3　嵌入信息的结构

Fig. 3　Construction of embedding information

表 2 数字水印对恶意篡改的定位能力

Table 2 The watermark detection results for malicious attack

裁剪	未攻击	第一秒	中间 0.5 秒	最后一秒	0.1 秒	0.01 秒 (2 段)	0.01 秒 (1 段)	插入一段音频
反置乱前的水印图像								
恢复后的水印图像								
NC	1.0000	0.9307	0.9061	0.9137	0.9623	0.9650	0.9810	0.9786

4) 将行号恢复为十进制, 根据行号重构水印.

5) 若行号没能正确检测, 则水印值保留其初始值 0.

3 仿真实验

3.1 算法对恶意篡改的定位能力

如第 2 节所述, 可根据未经反置乱的水印图像定位篡改区域, 本文为了兼顾定位精度及算法的时间复杂度, 将二维水印的行号作为水印段号, 应用此方法的另一个优势是将二维水印嵌入到一维音频信号中无需将水印转换为一维信号, 降低了算法的时间复杂度.

实验中均选用采样频率为 44.1 kHz, 分辨率为 16 比特的单声道数字音频信号作为原始音频载体, 数字水印采用了 64×64 的二值图像 Flower. 小波变换采用了常见的 Daubechies–1 小波基. 小波变换级数 H 选取为 3.

为了保证算法的不可听性, 在实验中, 音频的前 0.5 秒不进行水印嵌入. 为了验证检测算法对不同恶意篡改的定位能力, 仿真实验分别对含有水印音频信号的不同位置进行了不同大小的裁剪攻击与插入其他音频的攻击. 表 2 描述了本文算法对恶意篡改的定位能力, 通过反置乱前的水印图像可以粗略定位被篡改的音频段, 另外, 由于置乱技术的使用, 即使含水印的音频遭到恶意篡改, 最终仍可得到较好视觉效果水印图像, 同时, 通过采用归一化相关系数 (Normalized cross-correlation, NC) 对提取出的水印和原始水印的相似性进行定量评价, 可得出篡改程度的度量.

图 4 从客观角度精确定位篡改位置. 可根据图中安全区域外的异常 $Proc'$ 来确定对应音频段有篡改发生. 其中图 4 (b) 和 (c) 均为裁剪任意 0.01 秒音频, 但由于裁剪位置不同, 影响的音频段落数量也有所不同, 通过图 4 (b) 可定位第 31, 32 段音频被篡改, 通过图 4 (c) 定位第 37 段音频被篡改, 实验中的段长大于 0.01 秒, 所以, 当裁剪任意 0.01 秒时, 最多影响两个音频段.

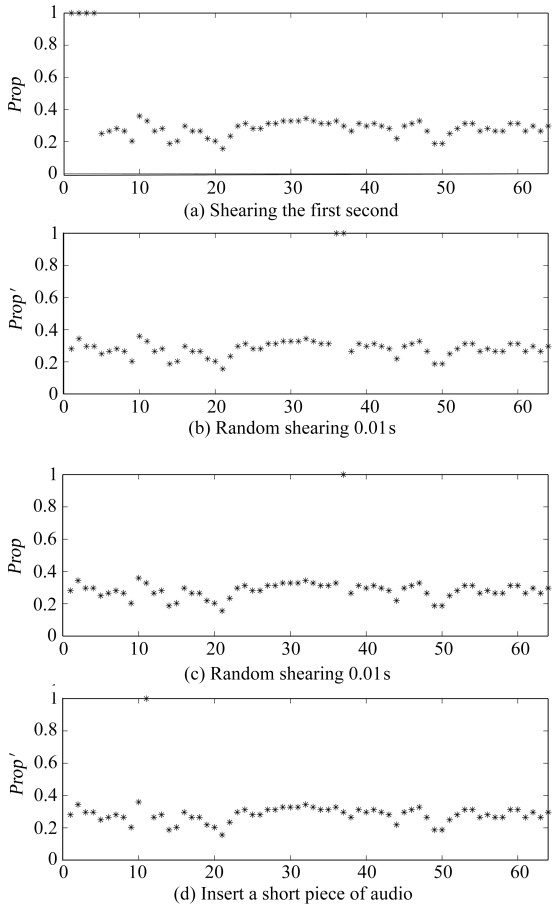

图 4 精确定位篡改位置

Fig. 4 The precise positions

3.2 算法抵抗常规信号处理的能力

当无篡改发生时, 为了检测算法的鲁棒性, 仿真实验对含水印音频信号进行了一系列攻击, 包

括: 重新采样, 重新量化, 高斯噪声干扰, MP3 压缩. 其中: 1) 重新采样, 将音频信号采样频率下降为 22.05 kHz、11.025 kHz, 再还原为原采样频率 44.1 kHz; 2) 重新量化, 先将音频信号从 16 比特量化为 8 比特, 再量化为 16 比特; 3) 叠加噪声, 对数字音频信号在时域中加入均值为 0、均方差为 0.01 的高斯白噪声; 4) 有损压缩, 先对音频信号进行压缩 (截至频率 56 kbits/s ∼ 32 kbits/s), 再解压缩.

表 3 给出了本文算法对于抵抗常规信号处理能力的实验结果, 同时采用归一化相关系数对提取出的水印和原始水印的相似性进行定量评价, 进一步证明了本算法的有效性.

表 3 数字水印对常规信号处理的鲁棒性

Table 3 The watermark detection results for general signal processing

信号处理	重量化	重采样 22050	重采样 11025	高斯噪声
NC	1.0000	0.9418	0.8578	0.9234
MP3 压缩	56 kbits/s	48 kbits/s	40 kbits/s	32 kbits/s
NC	0.9918	0.9857	0.9023	0.9445

4 结论

本文提出了一种抗裁剪的半脆弱音频水印算法, 其基本思想为: 二值水印图像中白色点 (像素值为 1) 与黑色点 (像素值为 0) 的分布随图像的不同而不同, 其白色点的点数与黑白总点数之比没有一定的规律. 经过置乱变换后的二值水印图像中白色点与黑白总点数之比是在一个较窄的安全区域上, 如果在某一段上的白色点数目与总点数之比出现异常, 不是落在较窄的安全区域上, 则认为是恶意攻击, 即可定位于该音频段上被恶意攻击; 该算法充分利用了多级置乱技术, 不但对于一定程度的常规信号处理具有较好的鲁棒性, 而且最大限度地恢复水印的视觉效果; 此外, 在对篡改定位的过程中, 与文献 [2−3] 比较, 无需原始水印参与, 便可精确定位被篡改区域 (尤其受到裁剪攻击时); 最后, 该算法具有易于实现, 执行效率高等优点, 具有很高的理论与应用价值.

References

1 Li Wei, Yuan Yi-Qun, Li Xiao-Qiang, Xue Xiang-Yang, Lu Pei-Zhong. Overview of digital audio watermarking. *Journal on Communications*, 2005, **26**(2): 100−111
(李伟, 袁一群, 李晓强, 薛向阳, 陆佩忠. 数字音频水印技术综述. 通信学报, 2005, **26**(2): 100−111)

2 Zhang Jing, Zhang Chun-Tian. Semi-fragile watermarking for JPEG2000 image authentication. *Acta Electronica Sinica*, 2004, **32**(1): 157−160
(张静, 张春田. 用于 JPEG2000 图像认证的半脆弱性数字水印算法. 电子学报, 2004, **32**(1): 157−160)

3 Li Chun, Huang Ji-Wu. A semi-fragile image watermarking resisting to JPEG. *Journal of Software*, 2006, **17**(2): 315−324
(李春, 黄继武. 一种抗 JPEG 压缩的半脆弱图像水印算法. 软件学报, 2006, **17**(2): 315−324)

4 Wang X Y, Zhao H. A blind audio watermarking algorithm robust against desynchronization attack. In: Proceedings of International Conference on Computational Intelligence and Security. Xi′an, China: 2005. 617−622

5 Licks V, Jordan R. Geometric attacks on image watermarking system. *IEEE Multimedia*, 2005, **12**(3): 68−78

6 Wu C P, Kuo C C J. Fragile speech watermarking based on exponential scale quantization for tamper detection. In: Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing. Florida, USA: IEEE, 2002. 3305−3308

7 Wu C P, Kuo C C J. Fragile speech watermarking for content integrity verification. In: Proceedings of IEEE International Symposium on Circuits and Systems. Phoenix-Scottsdale, USA: IEEE, 2002. 436−439

8 Quan Xiao-Mei, Zhang Hong-Bin. Fragile audio watermarking algorithm for telltale tamper proofing and authentication. *Journal of Electronics and Information Technology*, 2005, **27**(8): 1187−1192
(全笑梅, 张鸿宾. 用于篡改检测及认证的脆弱音频水印算法. 电子与信息学报, 2005, **27**(8): 1187−1192)

9 Ding Wei, Yan Wei-Qi, Qi Dong-Xu. Digital image scrambling technology based on Arnold transformation. *Journal of Computer-Aided Design and Computer Graphics*, 2001, **13**(4): 338−341
(丁玮, 闫伟齐, 齐东旭. 基于 Arnold 变换的数字水印图像置乱技术. 计算机辅助设计与图形学学报, 2001, **13**(4): 338−341)

赵 红 漳州师范学院计算机科学与工程系助教. 主要研究方向为信息安全与音频水印. 本文通信作者.
E-mail: fhq_xa@163.com
(**ZHAO Hong** Assistant in Department of Computer Science and Engineering at Zhangzhou Normal University. Her research interest covers information security and audio watermarking. Corresponding author of this paper.)

沈东升 漳州师范学院计算机科学与工程系副教授. 主要研究方向为信息安全与机器学习. E-mail: ds_shen@126.com
(**SHEN Dong-Sheng** Associate professor in Department of Computer Science and Engineering at Zhangzhou Normal University. His research interest covers information security and machine learning.)

朱元辉 漳州师范学院计算机科学与工程系讲师. 主要研究方向为信息安全.
E-mail: zyhui1118@gmail.com
(**ZHU Yuan-Hui** Lecturer in Department of Computer Science and Engineering at Zhangzhou Normal University. His main research interest is information security.)