一种基于差分进化的鲁棒音频隐写算法

苏兆品[1,2,3]，沈朝勇[1]，张国富[1,2,3]，岳峰[1,2]，胡东辉[1,2,3]

（1. 合肥工业大学计算机与信息学院，安徽 合肥 230601；
2. 合肥工业大学工业安全应急技术安徽省重点实验室，安徽 合肥 230601；
3. 合肥工业大学智能互联系统安徽省实验室，安徽 合肥 230009）

摘　要：音频隐写是将秘密信息隐藏到音频载体中，已成为信息隐藏领域的一个研究热点。已有研究大多聚焦最小化隐写失真，却以牺牲隐写容量为代价，且往往被一些常规信号攻击后难以正确提取秘密信息。为此，基于扩频技术，首先，分析了隐写参数（分段隐写强度和分段隐写容量）与不可感知性和鲁棒性的关系，并构建了一种以分段隐写强度、分段隐写容量为自变量，以不可感知性和隐写容量为优化目标，以信噪比为约束条件的音频隐写多目标优化模型；然后，提出了一种基于差分进化的鲁棒音频隐写算法，设计了相应的编码、适应度函数、交叉和变异算子。对比实验结果表明，所提隐写算法能够在保证不可感知性和抗隐写分析能力的前提下达到更好的鲁棒性，可以有效抵御一些常规信号处理攻击。

关键词：音频隐写；隐写参数；差分进化；鲁棒性

中图分类号：TP309.2

文献标识码：A

doi: 10.11959/j.issn.1000−0801.2021246

A robust audio steganography algorithm based on differential evolution

SU Zhaopin[1,2,3], SHEN Chaoyong[1], ZHANG Guofu[1,2,3], YUE Feng[1,2,], HU Donghui[1,2,3]

1. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China
2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology (Hefei University of Technology), Hefei 230601, China
3. Intelligent Interconnected Systems Laboratory of Anhui Province (Hefei University of Technology), Hefei 230009, China

Abstract: Audio steganography is to hide secret information into the audio carrier and has become a research hotspot in the field of information hiding. Most of the existing studies focus on minimizing distortion at the expense of steganography capacity, and it is often difficult for them to extract secret information correctly after some common sig-

收稿日期：2021−05−24；修回日期：2021−10−14

基金项目：安徽省重点研究与开发计划（No.202004d07020011，No.202104d07020001）；教育部人文社会科学研究青年基金项目（No.19YJC870021）；广东省类脑智能计算重点实验室开放课题（No.GBL202117）；中央高校基本科研业务费专项资金项目（No.PA2020GDKC0015，No.PA2021GDSK0073，No.PA2021GDSK0074）

Foundation Items: The Anhui Provincial Key Research and Development Program (No.202004d07020011, N0.202104d07020001), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (No.19YJC870021), Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation (No.GBL202117), Fundamental Research Funds for the Central Universities (No.PA2020GDKC0015, No.A2021GDSK0073, No.PA2021GDSK0074)

nal processing attacks. Therefore, based on the spread spectrum technology, firstly, the relationship between steganography parameters (i.e., segmented scaling parameters and steganography capacity) and imperceptibility as well as robustness was analyzed. Next, a multi-objective optimization model of audio steganography was presented, in which segmented scaling parameters and steganography capacity were decision variables, imperceptibility and steganography capacity were optimization objectives, and the signal-to-noise ratio was a constraint. Then, a robust audio steganography algorithm based on differential evolution was proposed, including the corresponding encoding, fitness function, crossover and mutation operators. Finally, comparative experimental results show that the proposed steganography algorithm can achieve better robustness against common signal processing attacks on the premise of ensuring imperceptibility and anti-detection.

Key words: audio steganography, steganography parameters, differential evolution, robustness

1　引言

隐写术是利用人的感知冗余和数字载体的统计冗余，将秘密信息隐藏于公开载体之中而不损坏载体的质量，以“隐匿秘密和通信存在”的方式实现秘密信息的传递[1]。与图像隐写[2]不同，在辨别微小失真方面，人的听觉系统要比视觉系统敏感得多。因此，音频隐写显得异常困难，发展较为缓慢[3]。

已有音频隐写方法大多利用人类听觉系统的掩蔽效应，在音频信号的时域或频域进行隐写。常见的有最低有效位（least significant bit，LSB）、相位编码和变换域等。例如，邹明光等[4]首先比较音频每个采样点分组中振幅值之间的关系，然后基于振幅值修改实现秘密信息的写入。Liu 等[5]根据加权能量自适应的选择宿主音频的小波包子带来嵌入秘密信息。Ahani 等[6]利用离散小波变换（discrete wavelet transformation，DWT）和稀疏分解将秘密信息嵌入信号的更高语义层。Meligy 等[7]基于提升小波变换系数修改和 LSB 对秘密信息进行嵌入。吴秋玲等[8]通过调节 DWT 的中高频系数来隐藏秘密信息。Kanhe 和 Aghila[9]利用语音信号的发声部分，通过离散余弦变换（discrete cosine transform，DCT）和奇异值分解（singular value decomposition，SVD）的级联实现隐写。Bharti 等[10]在嵌入之前对秘密音频的幅度位进行函数转换，然后将转换后的幅度嵌入载体音频的幅度中，将秘密音频信号嵌入载体音频的 LSB。Ali 等[11]基于分形编码和混沌 LSB 将秘密音频嵌入具有相同大小的封面音频中。张雪垣等[12-13]基于双层校验格码（syndrome-trellis codes，STC）提出一种音频分段隐写方法，通过对音频 LSB 载体流进行分段切割，利用最佳子校验矩阵嵌入秘密信息。

需要指出的是，上述已有工作大多只考虑如何使隐写失真达到最小化，牺牲了隐写容量，且忽略了鲁棒性。隐写是为了在不被发觉的情况下传输秘密信息，在很多网络音频应用中，发送者上传的隐写后的携密音频文件很可能会经过常规信号处理或有损转码、滤波、加噪等操作。而且，如果网络攻击者高度怀疑音频文件中包含秘密信息，但利用隐写分析器又检测失败，这时很可能采取“得不到就毁掉”极端措施，对音频文件进行改动、信号处理技术的加工或环境噪声攻击等。而已有工作，特别是基于压缩域的音频隐写，很难适应上述的信号处理和有损转码等情形，从而可能导致接收者无法从携密音频中正确提取出秘密信息，造成信息传递的失败。可见，音频隐写的鲁棒性也是实现秘密信息安全传递的一个重要指标，尤其是在商业机密等领域，理应受到足够的重视。

基于上述背景，本文基于扩频（spread spectrum，SS）[14]通信技术，通过分析隐写强度、隐写容量与不可感知性、鲁棒性之间的关系构建了

一种音频隐写多目标优化模型，然后基于差分进化（differential evolution，DE）[15]进行求解，提出一种鲁棒音频隐写算法，最后通过实验验证了所提隐写算法的有效性。

2 基于扩频的音频隐写优化模型

2.1 基于扩频的音频隐写方案

已有研究表明，基于 SS 的信息写入方案通常具有较好的鲁棒性[14]，但通常是直接对整个音频分段后进行信息写入，而没有考虑静音段和有声段对信息写入的影响不同，无法发挥音频载体隐写的最大性能。为了综合考虑鲁棒性和不可感知性，本文采用如图 1 所示的音频隐写方案，只在音频的有声段进行隐写。

为此，首先利用语音激活检测（voice activity detection，VAD）[16]技术划分出原始音频 S 的若干有声段 M 和静音段 N，再将每个有声段 M 分为 MA 和 MB 两部分。MA 通过量化平均值法嵌入同步码后得到 MA′；MB 经 DWT、DCT 和 SVD 后，利用扩频通信方法基于式（1）写入秘密信息，再经相应的逆变换得到 MB′。最后整合 MA′和 MB′得到 M'，再与 N 合并得到载密音频 S'。

$$\mathrm{MB}_i^{'} = \mathrm{MB}_i + \lambda_i \cdot C_i \tag{1}$$

其中，λ_i 和 C_i 均为隐写参数，$\lambda_i \in (0,1)$ 表示第 i 个有声段 MB_i 的分段隐写强度，$C_i \geqslant 0$ 表示第 i 个有声段 MB_i 的分段隐写容量。

2.2 隐写参数与不可感知性和鲁棒性之间的关系

为了分析隐写参数与不可感知性和鲁棒性之间的关系，本文以信噪比（signal-to-noise ratio，SNR）和客观差异等级（objective difference grade，ODG）作为音频隐写不可感知性的评价指标，以隐写音频经过如表 1 所示的 8 种攻击后的秘密信息提取的误码率（bit error rate，BER）的平均值（mean bit error rate，MBER）作为鲁棒性的评价指标。测试音频选取采样率为 48 kHz、分辨率为 16 bit/s 的 5 条不同时长的单声道音频见表 2。所有有声段 MB_i 的隐写强度 $\lambda_i \in (0.05,1)$，总隐写容量 C 在 400 bit 和 4 000 bit 之间。

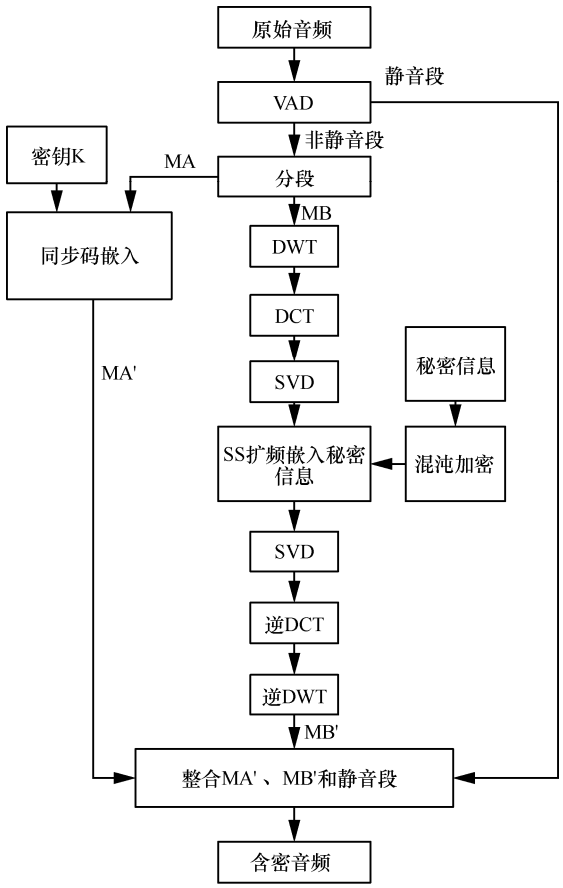


图 1 基于 SS 的音频隐写方案

隐写参数包括分段隐写强度和分段隐写容量两个变量。为了分析的方便，当分析不可感知性和鲁棒性与分段隐写强度的关系时，将总隐写容量固定为 2 500 bit，平均分到各有声段；当分析不可感知性和鲁棒性与分段隐写容量的关系时，将每个有声段 MB_i 的 λ_i 固定为 0.28。

不可感知性和隐写参数的关系如图 2 所示，当隐写容量固定时，不可感知性均随着隐写强度 λ_i 的增加而逐渐降低；当隐写强度固定时，不可感知性随着总隐写容量的增加而降低。因此，隐写后音频的不可感知性与隐写强度和隐写容量均成典型的反比关系。

表 1　常规信号处理攻击方式

序号	攻击方式
（1）	MP3 压缩（128 bit/s）
（2）	重量化（8 bit/s）
（3）	幅值修改（90%）
（4）	重采样：先以原始采样率的一半进行下采样，再以原始采样率进行上采样
（5）	加高斯白噪声
（6）	低通滤波
（7）	中值滤波
（8）	裁剪：每 2 000 个样本点随机选择 5 个位置进行裁剪

表 2　测试音频

音频	test1	test 2	test 3	test 4	test 5
音频长度/s	10	20	30	40	50

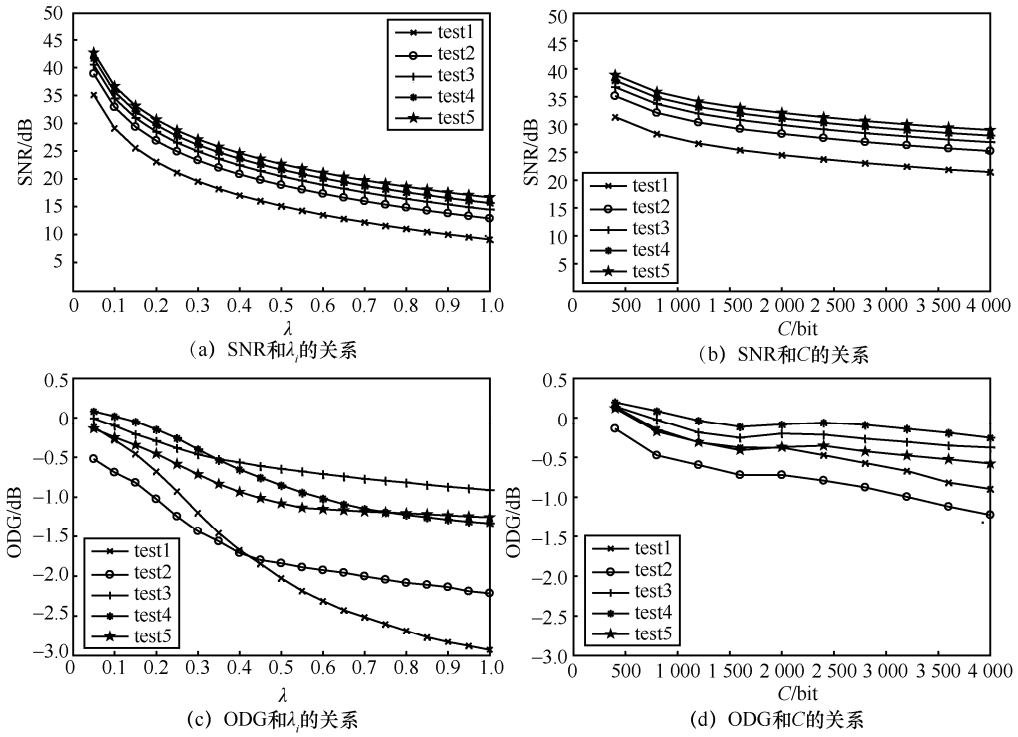
鲁棒性与隐写参数的关系如图 3 所示，当隐写容量固定时，鲁棒性会随着隐写强度增加而逐渐增加；当隐写强度固定时，鲁棒性随着隐写容量增加基本保持平稳。但从总体看，MBER 值波动平缓，验证了基于 SS 的音频隐写方案本身就具有很好的鲁棒性。

2.3　音频隐写优化模型的构建

基于 SS 的音频隐写方案具有很好的鲁棒性，其不可感知性会随着隐写容量或隐写强度的增加而逐渐降低。因此，本文构建如下的以隐写强度 λ_i、隐写容量 C_i 为自变量，以 ODG 和总隐写容量 C 为优化目标，以 SNR 为约束条件的音频隐写多目标优化模型：

$$\begin{aligned}&\max f_1 = \mathrm{ODG}(\lambda_i; C_i)\\&\max f_2 = C = \sum_i C_i\\&\text{s.t.}\quad \mathrm{SNR} > T_{\mathrm{thred}}\end{aligned} \tag{2}$$

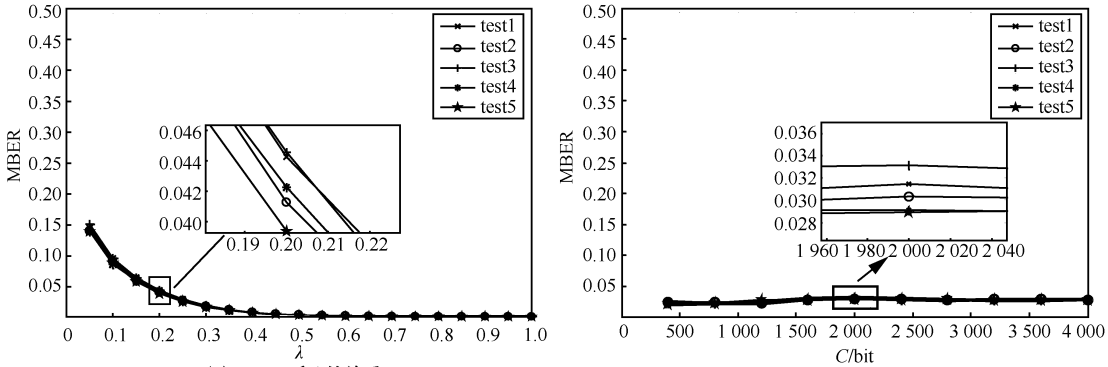
其中，f_1 是根据整个音频隐写前后计算的 ODG 值；T_{thred} 表示 SNR 的阈值，根据日本信息隐藏标准委员会的建议，$T_{\mathrm{thred}} = 20$ dB 时不可感知性较好[14]。由于 $\mathrm{ODG} \in [-4, 0]$，且 $\mathrm{ODG} \ll C$，为了消除不同数量级可能造成的误差，需要对式（2）进行归一化处理：



（a）SNR和λ_i的关系　（b）SNR和C的关系

（c）ODG和λ_i的关系　（d）ODG和C的关系

图 2　不可感知性与隐写参数的关系



(a) MBER和λ_i的关系　(b) MBER和C的关系

图 3　MBER 与隐写参数的关系

$$\begin{aligned}&\max f_1 = \frac{\mathrm{ODG}(\lambda_i; C_i)+4}{4}\\&\max f_2 = \frac{\sum_i C_i}{C^*}\\&\text{s.t.}\quad \mathrm{SNR} > T_{\text{thred}}\end{aligned} \tag{3}$$

其中，C^*为一个较大的归一化常数，通常设为所有有声段在经过小波变换后的系数位数之和。为了方便求解，将式（3）转化为如下的一个加权单目标优化模型：

$$\begin{aligned}&\max f = \alpha \cdot f_1 + \beta \cdot f_2\\&\text{s.t.}\quad \mathrm{SNR} > T_{\text{thred}}\end{aligned} \tag{4}$$

其中，$\alpha, \beta \in [0,1]$分别表示不可感知性和总隐写容量的权重，满足$\alpha + \beta = 1$。

3　基于 DE 的鲁棒音频隐写算法

对于式（4）的音频隐写优化模型，其包含两类自变量：分段隐写强度λ_i和分段隐写容量C_i。对于不同的音频，基于 VAD 分割出的有声段数目不同，其自变量个数也不同，且既有整数变量又有实数变量，再加上约束条件的限制，导致求解难度大，传统的数学规划方法已经难以应对，而以生物进化为启发来认知和模拟智能的进化算法非常适合求解这类问题，并已在音频水印中得到成功应用[17]。为此，本文引入 DE[15]算法求解式（4），通过设计个体编码方式、适应度函数和交叉、变异算子，得出基于式（4）的最优隐写参数。

3.1　个体编码与初始化

本文采用二维实数编码表示 DE 中的一个个体，如图 4 所示，第一行为分段隐写强度λ_i，第二行为分段隐写容量C_i，$i = 1,2,\cdots,n$，n为音频经过 VAD 分割出的有声段数量。

λ_1	λ_2	…	λ_{n-1}	λ_n
C_1	C_2	…	C_{n-1}	C_n

图 4　个体编码结构

对于初始种群中的每个个体，初始化时，$\lambda_i \leftarrow \mathrm{rand}(0,1)$为 [0,1] 范围的随机数；$C_i \leftarrow \mathrm{rand}(0, l_i)$为$(0, l_i)$区间的随机数，其中$l_i$为有声段$\mathrm{MB}_i$经 DWT、DCT 和 SVD 后获取系数的位数，满足$C^* = \sum_{i=1}^{n} l_i$。

3.2　适应度值计算

以式（4）的目标函数作为个体的适应度函数。对于给定的个体，首先对其进行解码，获取每个有声段的分段隐写强度λ_i和分段隐写容量C_i，并根据λ_i和C_i按照图 1 对原始音频进行隐写，获得载密音频，再利用原始音频和载密音频计算ODG值和总隐写容量C，最后基于式（4）计算该个体的适应度值。显然，个体的适应度值越大，表示该个体的隐写参数越优，即对应的隐写容量越大，不可感知性越好。

3.3　变异操作

变异操作采用 DE/rand/1/bin 方式[18]：

$$v_t = x_{r1} + F \cdot (x_{r2} - x_{r2}),\ t \neq r1 \neq r2 \neq r3 \quad (5)$$

其中，t 表示当前个体在种群中的序号；$r1,r2,r3 \in [1,N]$ 为 3 个不同的随机整数，N 表示种群规模；$F \in [0.4,1]$ 表示缩放因子；v_t 表示当前种群中的一个变异个体，x_{r1}、x_{r2}、x_{r3} 表示从当前种群中随机选取的另外 3 个不同初始个体。

为了避免算法的早熟和进化后期较优解易被破坏，本文选择文献[19]的自适应变异算子：

$$F = F_0 \cdot 2^{\theta} \quad (6)$$

$$\theta = \mathrm{e}^{1-\frac{G_m}{G_m+1-G}} \quad (7)$$

其中，$F_0 \in [0,1]$ 表示初始缩放因子，G 表示当前迭代次数，G_m 表示最大迭代次数。在 DE 进化的初期，缩放因子处于 F_0到$2F_0$ 之间，具有较大值，可保持初期个体的多样性，避免早熟。随着 DE 的进化，缩放因子的值逐渐降低，到后期接近 F_0，从而可有效保留优良个体，避免较优解被破坏，增加搜索到全局最优解的概率[19]。

3.4　交叉操作

交叉操作采用对初始个体 x_t 和变异个体 v_t 进行列交叉：

$$u_{t,j} = \begin{cases} v_{t,j}, \text{ rand}(0,1) \leqslant \text{CR} \\ x_{t,j}, \text{ 其他} \end{cases} \quad (8)$$

其中，$x_{t,j}$、$v_{t,j}$ 和 $u_{t,j}$ 分别表示原始个体 x_t、变异个体 v_t 和交叉个体 u_t 的第 j 列基因值；$\text{CR} \in [0,1]$ 为交叉概率。也就是说，对于 u_t 中的每一列 j，在(0,1)之间生成一个随机数，如果这个随机数小于或等于交叉概率，则复制原始个体 x_t 中同位置的基因值，否则复制变异个体 v_t 中同位置的基因值。

3.5　边界处理

经过变异和交叉操作后，种群中的个体很可能存在 $\lambda_i < 0$、$\lambda_i > 1$、$C_i < 0$ 或 $C_i > l_i$ 的情形。当出现上述任意一种情况时，该个体即非法个体，无法评估其适应度值，意味着这个个体的进化是无用的，这对于优化嵌入参数极为不利。因此，需要对其进行边界处理以保持个体的可行性。具体来说，当 $\lambda_i < 0$ 时，$\lambda_i \leftarrow 0$；当 $\lambda_i > 1$ 时，$\lambda_i \leftarrow 1$；当 $C_i < 0$ 时，$C_i \leftarrow 0$；当 $C_i > l_i$ 时，$C_i \leftarrow l_i$。

3.6　选择操作

将初始种群与交叉后的进化种群组合在一起，按照适应度值由大到小进行排序，选取前 N 个最佳个体组成新的初始种群继续进化。这样，每一代都保留了种群中的优良个体，促使种群持续探索更好的解。

3.7　算法描述

本文提出的基于 DE 的鲁棒音频隐写算法流程如图 5 所示，具体步骤如下。

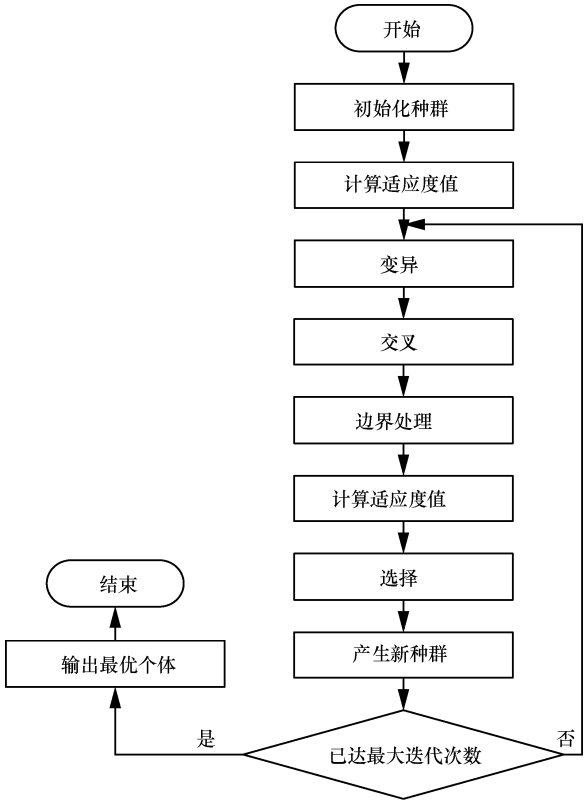


图 5　基于 DE 的鲁棒音频隐写算法流程

步骤 1　利用 VAD 将原始音频 S 分为若干有声段 M 和静音段 N，再将每个有声段 M 分为 MA 和 MB 两部分。其中，MA 通过量化平均值法嵌入同步码后得到 MA′，MB 包含 n 个有声段 MB_i。

步骤 2 根据 MB_i 的数目 n 和 l_i 对种群按照个体编码与初始化方法进行初始化，组成初始种群。

步骤 3 根据适应度值计算方法计算初始种群中每个个体的适应度值。

步骤 4 对初始种群进行变异和交叉操作组成进化种群。按照边界处理方法对进化种群中的每个个体进行边界检查和处理。

步骤 5 根据适应度值计算方法计算进化种群中每个个体的适应度值。

步骤 6 对初始种群和进化种群进行选择操作，生成新的初始种群。

步骤 7 如果已达最大迭代次数，则结束算法，输出种群中的最优个体；否则，转步骤 4 继续进化。

4 实验结果和分析

为了验证本文基于 DE 的鲁棒音频隐写算法（后称 DE 算法）的有效性，将 DE 算法与文献[12-13]提出的分段 STC 方法（简称 STC）和文献[6,8]提出的 DWT 方法进行对比实验。测试音频数据集为 50 条采样率为 48 kHz、分辨率为 16 bit/s、长度为 55 s 的 wav 格式音频。为了直观地展示对比方法的效果，秘密信息采用 64×64 的二值图像，根据载体本身的特性，重复进行隐写。

DE 算法实验参数 $N=30$，$G_m=100$，$F_0=0.5$，$CR=0.9$，$\alpha=0.75$，$\beta=0.25$。STC 方法中的子校验矩阵的规格为 6×6。DWT 方法中的通带截止频率为 7.8 kHz，阻带截止频率为 7 kHz，通带衰减为 3 dB，阻带衰减为 40 dB，分段长度为 20 ms。

所有对比方法的代码均基于 MATLAB 编写，并均在 AMD Ryzen CPU 3.59 GHz、16 GB 内存、Windows 10 操作系统的个人计算机上进行测试。

4.1 鲁棒性测试

3 种不同方法在测试数据集上的 MBER 结果如图 6 所示，STC 方法的 MBER 值最大为 0.43，最小为 0.38，平均值在 0.4 左右，DWT 方法的 MBER 值最大为 0.33，最小为 22，平均值在 0.28 左右，而本文 DE 算法的 MBER 值最低为 0.02，最高为 0.29，平均在 0.14 左右，明显低于 STC 方法和 DWT 方法。上述实验结果表明，在无攻击环境下，本文 DE 算法的误码率要更低。

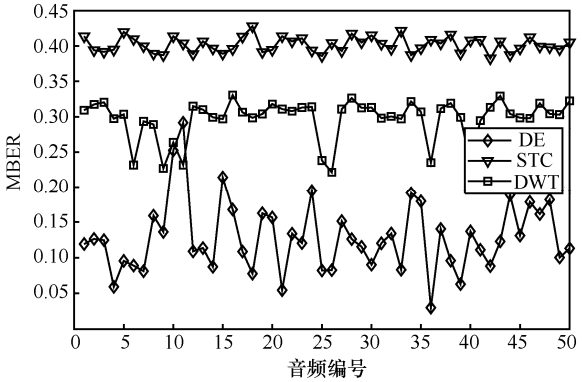


图 6 不同方法在测试数据集上的 MBER 结果

为进一步对比 3 种方法的抗攻击能力，随机选取了 test17 和 test47 两条音频，将二值图像完全写入音频，对隐写后的音频进行表 1 所示的 8 种常规攻击。表 3 和表 4 给出了这两条音频在每种攻击之后提取秘密信息与原始密信之间的 BER。可以看出，本文 DE 算法在 MP3 压缩、重量化、幅值修改、重采样和中值滤波后的误码率几乎接近于 0；在经加高斯白噪声和低通滤波后的误码率在 0.1 左右。STC 方法只有在裁剪攻击下误码率约为 0.04，在其他 7 种情况下均很难恢复出秘密信息。DWT 方法在幅值修改、重采样、加高斯白噪声和裁剪后的误码率在 0.15 左右，在其他 4 种情况下也很难恢复出秘密信息。可能的原因是，本文测试音频相对较长，对 20 ms 经过 DWT 变换之后，高频系数的前半部分能量或后半部分能量可能会为 0，即使在无攻击情况下也会出现提取错误。上述实验结果表明，在常规信号处理攻击环境下，本文 DE 算法的误码率更低，可以提取出全部或者大部分秘密信息。

表 3　test17 在 8 种攻击下的 BER

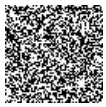
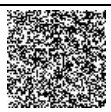
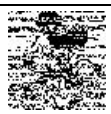
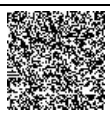
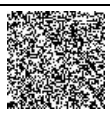
攻击	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
STC								
BER	0.348 1	0.492 1	0.299 1	0.385 5	0.517 8	0.496 1	0.493 9	0.043 0
DWT								
BER	0.488 0	0.480 0	0.136	0.177 7	0.438 2	0.465 1	0.136 5	0.157 7
DE								
BER	0.016 4	0.001 5	0.004 6	0	0.040 3	0.119 1	0	0.007 3

表 4　test47 在 8 种攻击下的 BER

攻击	（1）	（2）	（3）	（4）	（5）	（6）	（7）	（8）
STC								
BER	0.363 4	0.508 5	0.293 4	0.411 1	0.489 3	0.504 4	0.479 0	0.042 7
DWT								
BER	0.485 1	0.509 0	0.144 3	0.183 1	0.436 3	0.461 7	0.145 0	0.173 1
DE								
BER	0	0.002 6	0.003 4	0	0.115 7	0.194 6	0.007 0	0.021 0

综上，本文 DE 算法所给隐写方案的鲁棒性更强，即使在被信号处理攻击后也能保持较低的误码率。

4.2　不可感知性测试

为进一步量化分析不同方法的不可感知性，3 种方法在测试数据集上的 SNR 和 ODG 值如图 7 所示，STC 的不可感知性最好，DWT 的不可感知性最差，本文 DE 算法不可感知性居中。这是因为，在优化目标上，STC 只追求失真度最小，而 DWT 同时考虑 SNR 和 BER，本文 DE 算法兼顾隐写容量和 ODG。具体来说，本文 DE 算法在 50 条测试音频的 SNR 值均大于 20 dB；除了 test17，其余测试音频的 ODG 值均大于−2，人耳已经很难感知载体音频的细微变化。上述实验结果表明，本文 DE 算法给出的隐写方案保持了很好的不可感知性。此外，test17 的 VAD 结果如图 8 所示，在第 37 s 和 40 s 之间，分割出的 3 个有声段长度过小，且包含了静音段，导致此段的 ODG 的值较低。基于这一发现，未来可以考虑不在时长过短的有声段中嵌入秘密信息，以确保更好的不可感知性。

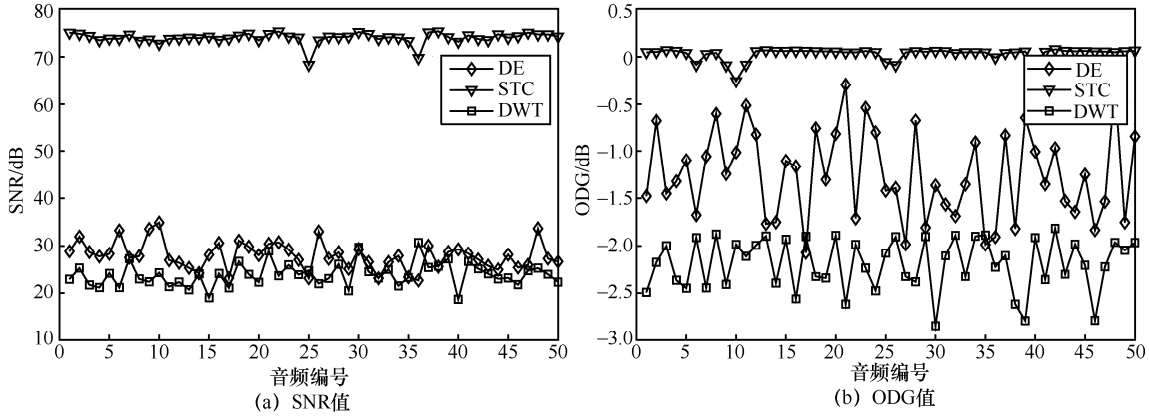


图 7　不同方法的不可感知性测试结果

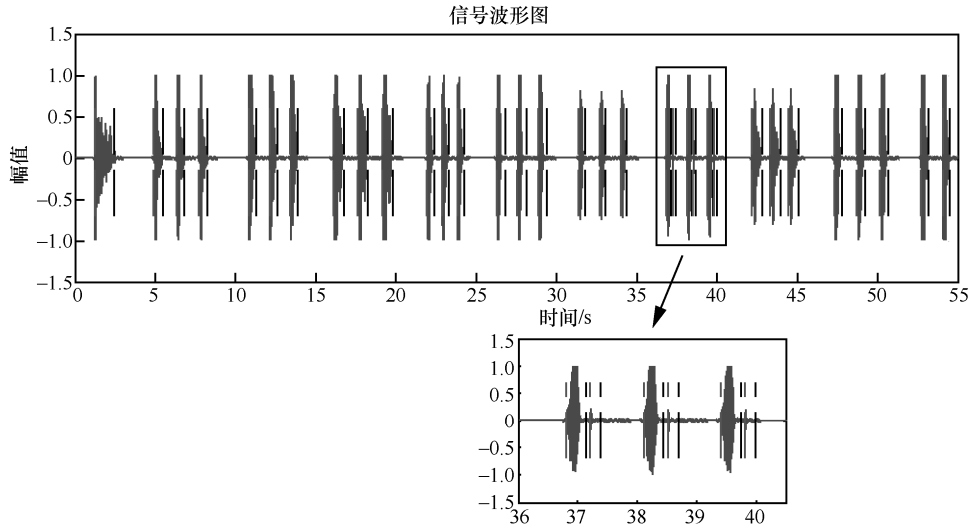


图 8　音频 test 17 的 VAD 结果

4.3　抗隐写分析性能测试

采用文献[20]的卷积神经网络通用隐写分析器检测 DE、STC 和 DWT 3 种隐写方法的抗隐写分析能力。首先，从开源平台“喜马拉雅”下载采样率为 48 kHz、分辨率为 16 bit/s 的 wav 格式音频，利用 CoolEdit 将其切分为每条时长 55 s、共 1 300 条语音文件；然后，分别利用 DE、STC 和 DWT 对其进行隐写，得到各自的 1 300 条载密音频；并分别用 1 300 条含密语音与 1 300 条原始音频训练隐写分析器；最后，利用训练好的隐写分析器检测前面实验采用的50条测试音频和对应的 50 条含密音频。

当隐写分析器同时检测出原始音频和对应的隐写音频时，认为检测正确。3 种方法的隐写分析检测结果见表 5，本文 DE 算法可以检测出 10 组，正确率为 20%；STC 方法检测出 11 组，正确率为 22%；DWT 方法检测出 9 组，正确率为 18%。上述实验结果表明，本文 DE 算法在抗隐写分析性能上与 STC 和 DWT 方法不相上下。

表 5　隐写分析检测结果

隐写方法	音频数	检测出的数目	正确率
DE	100	20	20%
STC	100	22	22%
DWT	100	18	18%

5　结束语

音频隐写是信息隐藏领域中的一个研究热点，主要致力于将秘密信息写入音频信号，同时在鲁棒性、不可感知性、隐写容量和抗隐写分析性能之间达到一个理想的均衡。为此，本文基于扩频技术研究音频隐写，通过分析隐写参数与不可感知性和鲁棒性的关系，构建了一种音频隐写多目标优化模型，并基于 DE 算法进行求解，设计了相应的编码和初始化方法，变异、交叉、选择和边界处理方法。实验结果表明，本文所提隐写算法在保持不可感知性和抗隐写分析能力的基础上，具有更好的鲁棒性，可以有效抵御一些常规信号处理的攻击。

但是，本文只是对基于搜索的音频隐写的一个初步探索，未来仍有许多问题需要进一步的深入研究。首先，需要在目标函数中考虑抗检测性，以期进一步提升隐写方案的抗隐写分析能力；其次，需要基于 VAD 分辨出时长过短的有声段，以避免在这些有声段进行隐写造成的过度失真；此外，还将考虑基于多目标 DE 算法求解音频隐写模型，以期在隐写容量、鲁棒性、不可感知性和抗检测性之间达到一个理想的均衡。

参考文献：

[1] 李凤华, 李晖, 贾焰, 等. 隐私计算研究范畴及发展趋势[J]. 通信学报, 2016, 37(4): 1-11.
LI F H, LI H, JIA Y, et al. Privacy computing: concept, connotation and its research trend[J]. Journal on Communications, 2016, 37(4): 1-11

[2] 孙树亮. 基于非下采样 Contourlet 变换和 Hill 编码的图像隐写[J]. 电信科学, 2016, 32(6): 124-128.
SUN S L. Image steganography based on nonsubsampled Contourlet transform and Hill cipher[J]. Telecommunications Science, 2016, 32(6): 124-128.

[3] DUTTA H, DAS R K, NANDI S, et al. An overview of digital audio steganography[J]. IETE Technical Review, 2020, 37(6): 632-650.

[4] 邹明光, 李芝棠. 基于振幅值修改的 wav 音频隐写算法[J]. 通信学报, 2014(z1): 36-40.
ZOU M G, LI Z T. Wav-audio steganography algorithm based on amplitude modifying [J]. Journal on Communications, 2014(z1): 36-40.

[5] LIU H L, LIU J J, HU R G, et al. Adaptive audio steganography scheme based on wavelet packet energy[C]//2017 IEEE 3rd international conference on big data security on cloud (bigdatasecurity), IEEE international conference on high performance and smart computing (hpsc), and IEEE international conference on intelligent data and security (ids). Piscataway: IEEE Press, 2017: 26-31.

[6] AHANI S, GHAEMMAGHAMI S, WANG Z J. A sparse representation-based wavelet domain speech steganography method[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(1): 80-91.

[7] MELIGY A M, NASEF M M, EID F T. An efficient method to audio steganography based on modification of least significant bit technique using random keys[J]. International Journal of Computer Network and Information Security, 2015, 7(7): 24-29.

[8] 吴秋玲, 吴蒙. 基于小波变换的语音信息隐藏新方法[J]. 电子与信息学报, 2016, 38(4): 834-840.
WU Q L, WU M. Novel audio information hiding algorithm based on wavelet transform[J]. Journal of Electronics & Information Technology, 2016, 38(4): 834-840.

[9] KANHE A, AGHILA G. A DCT-SVD-based speech steganography in voiced frames[J]. Circuits, Systems, and Signal Processing, 2018, 37(11): 5049-5068.

[10] BHARTI S S, GUPTA M, AGARWAL S. A novel approach for audio steganography by processing of amplitudes and signs of secret audio separately[J]. Multimedia Tools and Applications, 2019, 78(16): 23179-23201.

[11] ALI A H, GEORGE L E, ZAIDAN A, et al. High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain[J]. Multimedia Tools and Applications, 2018, 77(23): 31487-31516.

[12] 张雪垣, 王让定, 严迪群, 等. 基于分段 STC 的音频隐写算法[J]. 电信科学, 2019, 35(7): 115-123.
ZHANG X Y, WANG R D, YAN D Q, et al. An audio steganography algorithm based on segment-STC[J]. Telecommunications Science, 2019, 35(7): 115-123.

[13] ZHANG X Y, WANG R D, YAN D Q, et al. Selecting optimal submatrix for syndrome-trellis codes (STCs)-based steganography with segmentation[J]. IEEE Access, 2020, 8: 61754-61766.

[14] SU Z P, ZHANG G F, YUE F, et al. SNR-constrained heuristics for optimizing the scaling parameter of robust audio watermarking[J]. IEEE Transactions on Multimedia, 2018, 20(10): 2631-2644.

[15] STORN R, PRICE K. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359.

[16] CHANG J H, KIM N S, MITRA S K. Voice activity detection based on multiple statistical models[J]. IEEE Transactions on Signal Processing, 2006, 54(6): 1965-1976.

[17] LEI B Y, SOON I Y, TAN E L. Robust SVD-based audio watermarking scheme with differential evolution optimization[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(11): 2368-2378.

[18] DAS S, SUGANTHAN P N. Differential evolution: a survey of the state-of-the-art[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 4-31.

[19] YANG Z Y, TANG K, YAO X. Self-adaptive differential evolution with neighborhood search[C]//2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). Piscataway: IEEE Press, 2008: 1110-1116.

[20] JIANG S Z, YE D P, HUANG J Q, et al. SmartSteganogaphy: Light-weight generative audio steganography model for smart embedding application[J]. Journal of Network and Computer Applications, 2020, 165: 102689.

[作者简介]

苏兆品（1983− ），女，博士，合肥工业大学副教授，主要研究方向为语音安全、进化算法。

沈朝勇（1996− ），男，合肥工业大学硕士生，主要研究方向为音频隐写、进化算法。

张国富（1979− ），男，博士，合肥工业大学教授，主要研究方向为语音安全、软件工程。

岳峰（1981− ），男，博士，合肥工业大学副研究员，主要研究方向为软件工程、信息安全。

胡东辉（1973− ），男，博士，合肥工业大学教授，主要研究方向为信息安全、机器学习。