

基于合成语音的计算安全隐写方法

李梦涵^{1,2}, 陈可江^{1,2}, 张卫明^{1,2}, 俞能海^{1,2}

(1. 中国科学技术大学网络空间安全学院, 安徽 合肥 230001;

2. 中国科学院电磁空间信息重点实验室, 安徽 合肥 230001)

摘要: 计算安全的隐写理论很早就被提出, 但一直不能用于主流的以多媒体数据为载体的隐写术。原因在于计算安全隐写的前提是可以获得载体的精确分布或可以按照载体分布精确采样, 而自然采集的图像、音/视频无法满足这个前提条件。近几年, 随着深度学习的发展, 多媒体生成技术逐渐成熟且在互联网上的应用越来越普遍, 生成媒体成为合理的隐写载体, 隐写者可以用正常的生成媒体掩盖秘密通信, 即在媒体生成过程中隐写信息, 并与正常的生成媒体不可区分。一些生成模型学到的分布是可知或可控的, 这将为计算安全隐写推向实用提供契机。以当前广泛应用的合成语音模型为例, 设计并实现了计算安全的对称密钥隐写算法, 即在音频生成过程中, 根据样本点的条件概率, 按算术编码的译码过程将消息解压缩到合成音频中, 消息接收方拥有相同的生成模型, 通过复现音频合成过程完成消息提取。在该算法的基础上进一步设计了公钥隐写算法, 为实现包括隐蔽密钥交换在内的全流程隐蔽通信提供了算法支撑, 在保证隐写内容安全的同时, 还可以实现隐写行为安全。理论分析显示, 所提隐写算法的安全性由嵌入消息的随机性决定, 隐写分析实验进一步验证了当前技术下攻击者无法区分合成的载体音频与载密音频。

关键词: 音频隐写; 语音合成; 生成模型; 公钥隐写

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2022025

Computationally secure steganography based on speech synthesis

LI Menghan^{1,2}, CHEN Kejiang^{1,2}, ZHANG Weiming^{1,2}, YU Nenghai^{1,2}

1. School of Cyber Technology and Science, University of Science and Technology of China, Hefei 230001, China

2. CAS Key Laboratory of Electro-magnetic Space Information, Hefei 230001, China

Abstract: The steganography theory of computing security has been proposed for a long time, but it has not been widely adopted for mainstream steganography using multimedia data as a carrier. The reason is that the prerequisite

收稿日期: 2021-05-17; 修回日期: 2021-08-31

通信作者: 俞能海, ynh@ustc.edu.cn

基金项目: 国家自然科学基金(62102386, 62002334, 62072421, 62121002); 中国博士后基金面上项目(2021M693091); 安徽省自然科学基金项目(2008085QF296)

Foundation Items: The National Natural Science Foundation of China (62102386, 62002334, 62072421, 62121002), China Postdoctoral Science Foundation (2021M693091), The Nature Science Foundation of Anhui Province (2008085QF296)

引用格式: 李梦涵, 陈可江, 张卫明, 等. 基于合成语音的计算安全隐写方法[J]. 网络与信息安全学报, 2022, 8(3): 134-141.

Citation Format: LI M H, CHEN K J, ZHANG W M, et al. Computationally secure steganography based on speech synthesis[J]. Chinese Journal of Network and Information Security, 2022, 8(3): 134-141.

for calculating secure steganography is to obtain the accurate distribution of the carrier or to accurately sample according to the carrier distribution. However, naturally collected images and audio/video cannot meet this prerequisite. With the development of deep learning technology, various machine-generated media such as image generation and synthesized speech, have become more and more common on the Internet and then generated media has become a reasonable steganography carrier. Steganography can use normal generated media to cover up secret communications, and pursue in distinguishability from normal generated media. The distribution learned by some generative models is known or controllable, which provides an opportunity to push computational security steganography for practical use. Taking the widely used synthetic speech model as an example, a computationally secure symmetric key steganography algorithm was designed and implemented. The message was decompressed into the synthetic audio according to the decoding process of arithmetic coding based on the conditional probability of sample points, and the message receiver had the same generation model to complete the message extraction by reproducing the audio synthesis process. The public key steganography algorithm was additionally designed based on this algorithm, which provided algorithmic support for the realization of full-flow steganographic communication. Steganographic key exchange ensured the security of steganographic content and the security of steganographic behavior was also achieved. The theoretical analysis showed that the security of the proposed algorithm is determined by the randomness of the embedded message. And the steganography analysis experiment further verified that the attacker cannot distinguish the synthesized carrier audio from the encrypted audio.

Keywords: audio steganography, speech synthesis, generative model, public key steganography

0 引言

隐写术是隐蔽通信的一种重要形式，它将机密信息隐藏在内容可公开的载体（如文本、音频、图像、视频等）中，使非授权方难以识别该通信形式的存在。近年来，隐写术研究从非自适应隐写，如最低有效位（LSB, least significant bit）替换^[1]、LSB 匹配^[2]等发展到基于最小化失真的自适应隐写，目前的研究致力于在 STC 编码^[3]的基础上设计合理的失真函数，并实现了良好的安全性，典型的算法如 HUGO^[4]、WOW^[5]、HILL^[6]、JUNIWARD^[7]等。

传统隐写术的安全性表现为隐写分析实验验证得到的经验安全，并不能从理论上验证其安全性。1998 年，Cachin^[8]建立了信息论模型研究隐写术的安全性，使用相对熵定义无条件安全隐写。Hopper 等^[9]首次提出使用计算复杂度理论来研究隐写术，并在理论上证明了单向函数的存在意味着安全隐写协议的存在。在此基础上，研究者对可证明安全隐写进行了进一步的探究。Dedic 等^[10]研究了黑盒隐写术可以实现的最佳效率，进一步降低了对隐藏信道的要求，并基于拒绝采样编码

器给出了两个改进算法。文献[11]提出了公钥隐写术。文献[12]基于算术编码提出了 P-codes，建立了新的高效的公钥隐写体系。

然而，以上研究都基于一个共同的前提假设，即隐写者拥有完备的载体概率分布或可以对载体精确采样。而主流的隐写载体是自然采集的多媒体数据，隐写者无法获得载体的精确分布，也无法进行精确采样，故可证明安全隐写理论长期被搁置，无法付诸实践。

近年来，随着机器学习的发展，多媒体生成技术逐渐成熟，基于生成对抗网络^[13]（GAN, generative adversarial nets）或变分自编码器^[14]（VAE, variational auto-encoder）的生成模型走进人们的日常生活中，典型的应用有图像风格变换、图像生成、语音合成等，促使以机器生成媒体作为载体的生成式隐写成为新的研究热点。

语音合成技术的应用越来越广泛，如有声小说、语音应答服务等，在这种场景下，检测到合成语音并不能作为隐写存在的证据，因此合成语音是合理的隐写载体。已有的生成式音频隐写研究如 AAG-Staga^[15]在音频自动生成过程中编码每

个采样点的条件概率分布,并根据比特流选择相应的信号输出,实现秘密信息的嵌入。Shi 等^[16]将秘密信息映射到噪声向量中,并使用生成对抗网络基于噪声向量生成载体图像。基于显式概率分布的生成模型在生成过程中可以得到每个样本点的条件概率分布情况,这完全符合可证安全隐写理论对载体的要求。以生成模型为基础,已有的研究可证明安全隐写方案可以由理论转为现实。

另外,传统隐写术的密钥交换等过程仍然依赖于密码学的相关协议,也就是说,通信双方在通信之前需要先通过一个不隐蔽的渠道交换共享的密钥,才可以在隐蔽信道进行隐蔽通信,然而密码学协议的通信过程恰恰已经影响了隐写的隐蔽性。在隐写分析技术愈来愈成熟,隐写者的行为安全显得尤为重要。为了实现更加完备的隐写行为安全,本文以可证明安全隐写为理论基础,以显式概率音频生成模型为实现方案,设计了计算安全的对称密钥隐写方案和公钥隐写方法。

1 预备知识

1.1 隐写系统

Simmons^[17]提出的囚犯模型是经典的隐写模型。此模型中, Alice 和 Bob 被囚禁在不同的牢房中,准备秘密地制定逃跑计划,他们被允许交流信息,但通信过程将受到看守员 Wendy 的监视,如果 Wendy 发现他们有传递有害信息的行为,就会切断二人的通信。

以囚犯问题为基础,隐写通信过程可被形式化地定义,如图 1 所示。Alice 和 Bob 共享隐写密钥 k , 两人的正常通信内容为载体 C , Alice 将秘密消息 m 利用密钥 k 加密,再利用与 Bob 约定好的隐写嵌入算法将加密后的信息嵌入载体 C 中,得到载密文件 S 。然后经过公共信道将 S 传输给 Bob, 此过程受 Wendy 监视:

$$\text{Emb}(C, m, k) = S \quad (1)$$

Bob 收到 S 后将通过隐写提取算法 $\text{Ext}()$ 和密钥 k 得到消息 m , 完成隐蔽通信过程:

$$\text{Ext}(S, k) = m \quad (2)$$

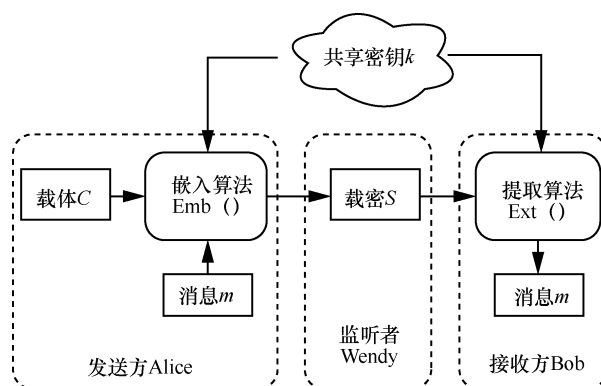


图 1 隐写模型

Figure 1 Steganography model

1.2 显式概率生成模型

生成模型指的是可以生成新数据的概率模型,基本思想是将训练集 $D = \{x\}^N$ 中的真实样本数据视为符合某未知分布的样本,定义其参数化的分布为 $P_\theta(x)$, 参数 θ 为深度神经网络的参数。通过深度学习训练参数 θ 以近似训练数据的经验分布 $P_D(x)$ 。然后利用训练完成的模型 $P_\theta(x)$ 来生成新的样本,生成的样本和真实的样本应尽可能地相似。

根据能否得到数据的显式概率分布,生成模型可以分为显式概率生成模型和隐式概率生成模型。显式概率生成模型如 PixelCNN^[18]、MADE^[19]、NADE^[20]、WaveNet^[21]等,都是利用概率的链式法则,将生成样本的概率分布分解为一系列条件概率分布的联乘积形式,即 N 个随机变量的概率分布为

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i | x_1, x_2, \dots, x_{i-1}) \quad (3)$$

依据链式法则原理,可参数化建模所有条件概率分布 $p(x_i | x_1, \dots, x_{i-1})$, 从而完整建模样本 $p(x_1, x_2, \dots, x_N)$ 的分布情况。这种模型是自回归的,因为它将数据视为序列,并使用参数化模型 $p_\theta(x_i | x_1, \dots, x_{i-1})$ 迭代地依据历史序列 $x_1, \dots, x_{i-1}$ 预测下一随机变量 x_i 的分布。显式概率生成模型如图 2 所示。

1.3 语音合成模型

语音合成领域的主流方法是基于统计参数的语音合成模型。首先利用声码器从语音数据库中提取语音波形的特征表示,得到基频、倒谱等参

数,然后将文本特征作为输入,声学参数作为输出,利用统计建模的方法训练声学模型。生成过程则根据训练好的声学模型,通过特征表示对语音波形进行重构,将声学特征转化为合成语音。

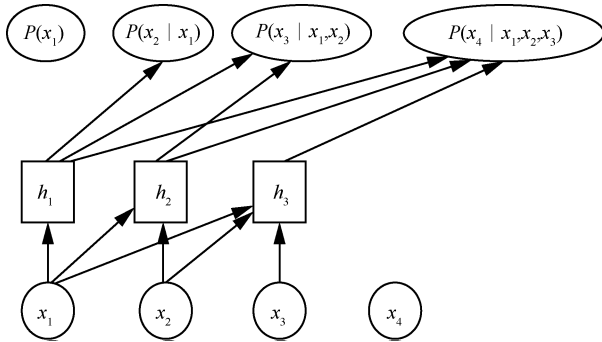


图2 显式概率生成模型
Figure 2 Explicit probabilistic generative model

本文采用的 WaveNet 声码器是一个基于 PixelCNN 架构的生成模型。该模型可以根据梅尔倒谱图合成逼真的人类声音。WaveNet 在生成过程中对数据流 $x_1, x_2, \dots$ 的联合概率分布建立模型,按照时间序列得到数据流的每一个元素的条件概率,每一个样本受限于所有的已生成的样本。以 8 bit 语音合成为例, WaveNet 在 softmax 层输出 256 个样本点对应的概率。在正常生成过程中, WaveNet 根据当前样本点的概率分布进行随机采样,得到最终的合成语音,本文将采样过程记为 G ,如图 3 所示。

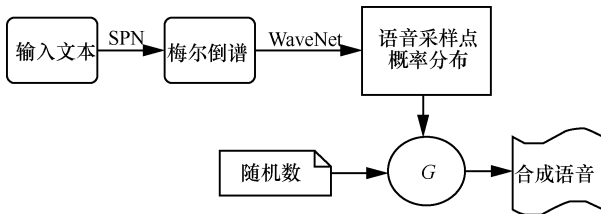


图3 基于 WaveNet 的语音合成过程
Figure 3 Speech synthesis process based on WaveNet

2 全流程隐蔽通信体系

安全有效的隐蔽通信既要考虑通信信道中传输信息的内容安全性,也要保证通信过程的行为安全性。大多数隐写者重点关注的是隐写算法的抗检测性能,而近年来,攻击者开始关注通信双

方的异常行为,如文献[22]从侧信道攻击的角度考察社交网络用户所分享的图片,并根据图像序列的相关性来识别隐写者。

因为传统隐写算法是对称密钥算法,通信双方的密钥交换、身份认证等过程依赖于公钥密码或密码协议。但是执行密码协议这一行为本身就可能暴露隐蔽通信的存在。隐写分析者可以通过识别异常通信行为定位到隐写者。为了解决这个问题,实现兼具行为安全和内容安全的隐蔽通信系统,本文希望建立一个能够实现公钥隐写的全流程的隐蔽通信体系。

本文依据文献[12]提出的可证明安全隐写模型 P-codes,以显式概率的音频生成模型 WaveNet 为实现方案,提出了计算安全的对称密钥隐写算法,如图 4 所示。并在此基础上,构建公钥隐写算法,这使得通信的全流程可在隐蔽信道中实现。

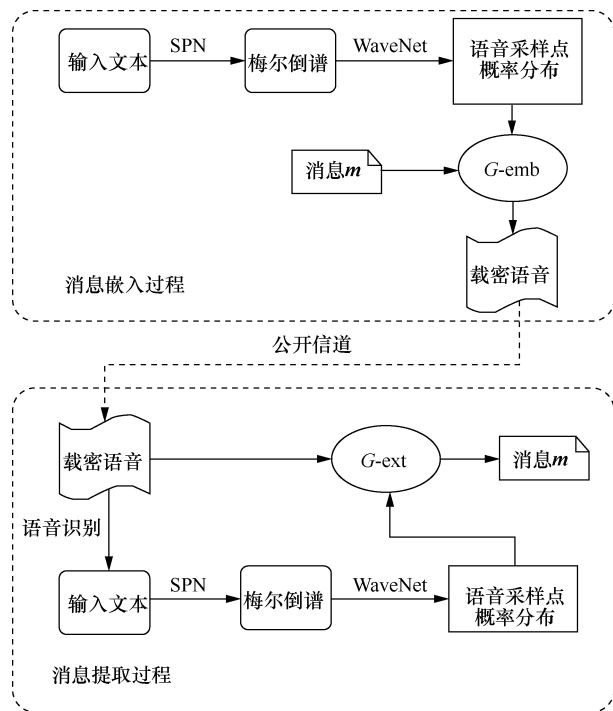


图4 基于音频生成的可证安全隐写框架
Figure 4 Computationally secure steganography framework based on speech synthesis

2.1 对称密钥隐写算法

本文设计的隐写算法修改 WaveNet 语音生成模型的采样算法 G ,在音频生成过程中,根据嵌入信息,依照概率分布,选择最终生成

的样本点。

2.1.1 算法描述

WaveNet 生成音频的过程即为嵌入过程。生成模型的输入为声学特征，在生成每一个样本点 x_i 时，经过 WaveNet 语音生成模型得到样本点的概率分布，其形式为声音信号取 -128 到 127 时所对应的 256 个条件概率值，记为 p_h ， h 表示当前已生成的样本片段 $x_1, x_2, \dots, x_{i-1}$ 。正常的生成过程将按照概率随机选择样本点的值。而本文的隐写嵌入算法 $G\text{-emb}$ 会依据嵌入信息和概率 p_h 共同决定样本点的取值。

假设收发双方已共享了一个密钥 k ，发送方首先利用密钥 k 和一个安全的对称密码算法，如 SM4^[23] 或 AES^[24]，对消息 m 加密得到的 $m' = (m_1, m_2, \dots, m_n) \in \mathcal{M}^n$ ， $\mathcal{M} = \{0, 1\}$ 。

基于 G 的嵌入算法记为 $G\text{-emb}$ ，输入为密文序列 m' ；输出为长度为 l 的声音信号值 $x = (x_1, x_2, \dots, x_l) \in \mathcal{X}^l$ ，其中， $\mathcal{X} = \{-128, -127, \dots, 126, 127\}$ 。 $G\text{-emb}$ 如算法 1 所示。其中初始 $h = \epsilon$ 为空集， $\bar{m}'$ 表示二进制形式为 m' 的非负整数。

算法 1 嵌入算法 $G\text{-emb}$

输入 $m' = (m_1, m_2, \dots, m_n) \in \mathcal{M}^n$

输出 $x = (x_1, x_2, \dots, x_l) \in \mathcal{X}^l$

1) 令 $a = 0, b = 2^{2^n}, h = \epsilon$

2) **While** $\lceil a / 2^n \rceil < \lfloor b / 2^n \rfloor$, **do**

(a) 计算 $f_i = \sum_{0 \leq k \leq i} p_h(v_k)$ for all $v_i \in \mathcal{X}$

(b) 确定整数 j 使式 $f_j \leq (2^n \bar{m}' - a) / (b - a) < f_{j+1}$ 成立

(c) 令 $a' = a + (b - a)f_j, b' = a + (b - a)f_{j+1}$

(d) 令 $(a, b) = (a', b')$

(e) 使 $h = h \cup v_j$

(f) 更新 p_h

3) 输出 $x = h$

$G\text{-emb}$ 的输出结果为声音信号，再经过处理得到 wav 格式的音频作为生成的 stego。该算法的本质是算术编码的解码过程。

收发双方共享生成模型，如使用网上开源的同一个语音合成 WaveNet 模型，或者将训练好的 WaveNet 模型在秘密通信之前传输给接

收方。接收方利用语音识别工具，将载密语音识别成文本。收发双方如果拥有相同的生成模型，那么接收方可以得到与发送方一致的采样点概率分布。提取算法 $G\text{-ext}$ 则相当于算术编码的编码过程，将待生成的音频信号按概率编码得到嵌入信息。该算法的输入为载密音频 $x = (x_1, x_2, \dots, x_l) \in \mathcal{X}^l$ ，输出为提取到的密文信息 $m' = (m_1, m_2, \dots, m_n) \in \mathcal{M}^n$ ，再利用密钥 k 解密得到消息 m 。 $G\text{-ext}$ 如算法 2 所示，其中 v 表示数字 v 的二进制向量。

算法 2 提取过程 $G\text{-ext}$

输入 $x = (x_1, x_2, \dots, x_l) \in \mathcal{X}^l$

输出 $m' = (m_1, m_2, \dots, m_n) \in \mathcal{M}^n$

1) 令 $a = 0, b = 2^{2^n}, h = \epsilon$

2) **for** i **from** 1 **to** l **do**

(a) 计算 $f_i = \sum_{0 \leq k \leq i} p_h(v_k)$ for all $v_i \in \mathcal{X}$

(b) 确定整数 j 使得 $v_j = x_i$

(c) 令 $a' = a + (b - a)f_j, b' = a + (b - a)f_{j+1}$

(d) 令 $(a, b) = (a', b')$

(e) $h = h \cup v_j$

(f) 更新 p_h

3) $v = a / 2^n$

4) 输出 $m' = v$

2.1.2 安全性证明

本文所提算法以文献[12]中的安全隐写模型为基础，根据文献中的安全性证明可知，只要嵌入过程输入的消息是密码学安全的随机序列，那么该算法就可以保证 $G\text{-emb}$ 输出的载密序列 x 与正常语音合成器 G 生成的序列计算不可区分。本文算法中输入 m' 是由消息 m 经过安全的密码算法生成，因而满足上述条件，主要原理如下。

在嵌入算法 $G\text{-emb}$ 中， $v_i \in \mathcal{X}$ 表示信道 $\mathcal{X} = \{-128, -127, \dots, 126, 127\}$ 中的所有可能的取值，其中 $i = 0, 1, \dots, n$ ， v_i 的概率分布是由生成模型决定的，与嵌入信息序列 m' 是独立的， v_i 按一定标准排序，其顺序也与 m' 独立。假设 a, b 是实数，那么每次迭代过程中，如果 m' 是随机的，那么 $\bar{m}'$ 在给定历史条件 $h = x_1 \parallel x_2 \parallel \dots \parallel x_j$ 下在区

间 $[2^{-n}a, 2^{-n}b]$ 上的条件概率分布是均匀随机分布的, 可以认为 $u = (2^n \bar{m}' - a) / (b - a)$ 在 $(0, 1)$ 是均匀随机分布的, 与 $v_1, v_2, \dots, v_n$ 独立, 则由 u 确定的 v_j 与 $v_1, v_2, \dots, v_n$ 独立同分布, 也就是每一步选择的 v_j 可以认为是与正常的生成内容统计上无法区分。

在本文算法的实际应用中, a 和 b 取整数, 迭代完成时, 由于取整的影响, $\bar{m}'$ 在 $[2^{-n}a, 2^{-n}b]$ 的条件概率分布并非完全均匀随机的, 其与均匀随机的误差最大为 $4/(b-a)$, 又因为 $(b-a) > 2^n$, 所以误差小于 2^{2-n} 。那么当 n 足够大时, 误差趋于零, 这个隐写方案是安全的。

另外, WaveNet 语音生成过程本身就是由随机数驱动的, 对正常生成的载体进行消息提取也会得到随机数, 那么嵌入信息被计算安全的密码算法加密后, 攻击者无法区分提取出的秘密消息与随机序列。

综上, 该隐写算法的安全性由 m' 的随机性决定, 如果消息 m 被计算安全的对称密钥密码算法加密为伪随机信息再作为嵌入信息, 该隐写算法就是安全的。

2.2 公钥隐写算法

在上述隐写算法中, 将加密消息 m 的对称加密算法换成公钥加密算法, 如 SM9^[25] 或 RSA^[26], 就可以得到对应的公钥隐写算法。

公钥隐写算法可以用于密钥的隐蔽分发。如图 5 所示, 发送方用接收方的公钥 d 加密随机密钥 k , 所得密文 c 作为算法 1 的输入生成载密语音, 将载密语音通过公开信道发送给接收方, 接收方利用算法 2 从载密语音中提取出 c , 然后利用自己的私钥解密得到 k 。若 c 由安全的密码算法产生, 则对应的载密语音与正常的合成语音计算不可区分, 所以这个密钥分发过程是隐蔽的。之后, 利用共享的 k , 通信双方就可以用 2.1 节的对称密钥隐写算法进行通信, 由此实现全过程的隐蔽通信。

3 实验

本节对第 2 节中所提的隐写算法进行安全性验证实验。

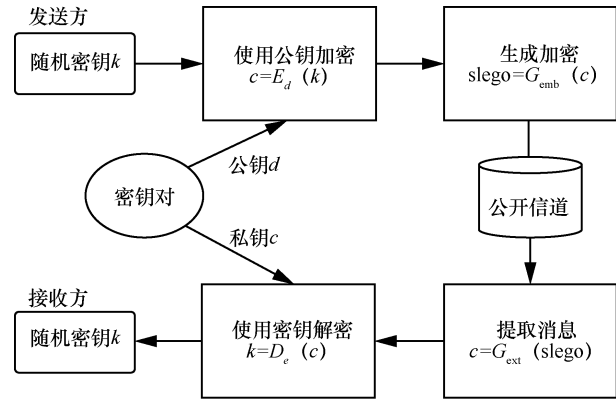


图5 基于公钥隐写算法的密钥分发过程
Figure 5 Key distribution process based on public key steganography algorithm

3.1 实验设置

本实验使用的数据集为英文朗读语音数据库 CMU_ARCTIC^[27], 其中包含 4 位美式英语说话人以及 3 位带有口音的说话人。本实验用该数据集训练 WaveNet 生成模型。

隐写分析实验采用隐写分析特征与集成分类器结合的隐写分析器对这两种语音进行分类。10 次检测的平均错误率 $\bar{P}_E$ 作为评价指标, 其中错误率 P_E 定义为检测器的虚警率 P_{FA} 和漏警率 P_{MD} 的均值。

$$P_E \triangleq \min_{P_{FA}} \frac{1}{2} (P_{FA} + P_{MD}) \quad (4)$$

本文采用的隐写分析特征 CTM^[28] 是一种基于时域和频域的联合音频隐写分析特征, 包括音频二阶差分的马尔可夫矩阵以及梅尔倒谱频率系数的均值方差统计量。

为了验证隐写分析算法在合成音频数据集上的有效性, 本文引入了基于修改的隐写方法进行安全性比较, 包括 LSB 匹配和导数滤波残差 (DFR, derivative filter residual)^[29]。在 LSB 匹配的嵌入过程中, 如果最低有效位与秘密数据位相匹配, 则该元素保持不变, 否则就进行随机 ± 1 修改。DFR 是一种自适应音频隐写方法, 它给不同的元素定义不同的失真, 然后采用 STC 编码^[3], 在最小化总体失真的情况下, 实现信息嵌入。STC 编码的性能可以逼近率失真的下界, 本实验利用理想的模拟嵌入代替 STC 编码进行实验。

3.2 隐写算法安全性分析

本文挑选了 1 000 个原始音频生成两组载体

文件和 3 组载密文件,其中载体文件不嵌入信息,分别由不同的随机数驱动生成音频,3 组载密文件中分别嵌入 5 000 bit、10 000 bit、15 000 bit 信息,嵌入信息为 AES 加密算法对消息加密得到的随机序列。

隐写分析实验分别对第 1 组载体与 3 组载密进行分类,平均测试错误率约为 0.5,如表 1

所示。由图 6 可见,随着嵌入信息的增加,本文所提隐写算法的隐写分析错误率并没有明显变化,始终维持在 0.5 左右,而其他隐写算法的检测错误率显著下降,当嵌入 15 000 bit 时,LSBM 可被有效检测出。说明本文提出的生成式隐写算法可以抵抗隐写分析结果,与理论证明吻合。

表 1 隐写分析测试错误率
Table 1 Test error rate of steganalysis

嵌入比特/bit	测试错误率		
	G-emb	DFR	LSBM
5 000	0.496 4(+/-0.009 3)	0.470 2(+/- 0.005 3)	0.310 4(+/-0.012 7)
10 000	0.500 5 (+/- 0.013 0)	0.430 8(+/-0.006 4)	0.161 7(+/-0.014 6)
15 000	0.500 7(+/- 0.016 0)	0.386 6(+/-0.006 3)	0.085 7(+/-0.004 5)

此外,对由不同随机数驱动生成的两组载体进行分类,测试错误率为 0.509 1 (+/- 0.021 3),这说明隐写分析者得到的特征中,载体-载密与载体-载体并无区别,可以认为该隐写方案是安全的。

明验证了其安全性。该隐写算法对抗隐写分析的实验结果显示,攻击者无法区分载体和载密。

参考文献:

- [1] BENDER W, GRUHL D, MORIMOTO N, et al. Techniques for data hiding[J]. IBM Systems Journal, 1996, 35: 313-336.
- [2] MIELIKAINEN J. LSB matching revisited[J]. IEEE Signal Processing Letters, 2006, 13(5): 285-287.
- [3] 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.
- [4] PEVNÝ T, FILLER T, BAS P. Using high-dimensional image models to perform highly undetectable steganography[M]//Information Hiding. Berlin, Heidelberg: Springer, 2010: 161-177.
- [5] HOLUB V, FRIDRICH J. Designing steganographic distortion using directional filters[C]//Proceedings of 2012 IEEE International Workshop on Information Forensics and Security (WIFS). 2012: 234-239.
- [6] LI B, WANG M, HUANG J W, et al. A new cost function for spatial image steganography[C]//Proceedings of 2014 IEEE International Conference on Image Processing (ICIP). 2014: 4206-4210.
- [7] HOLUB V, FRIDRICH J, DENEMARK T. Universal distortion function for steganography in an arbitrary domain[J]. EURASIP Journal on Information Security, 2014, 2014: 1-13.
- [8] CACHIN C. An information-theoretic model for steganography[M]//Information Hiding. Berlin, Heidelberg: Springer, 1998: 306-318.
- [9] HOPPER N, AHN L V, LANGFORD J. Provably secure steganography[C]//Proceedings of IEEE Transactions on Computers. 2009: 662-676.
- [10] DEDIĆ N, ITKIS G, REYZIN L, et al. Upper and lower bounds on black-box steganography[J]. Journal of Cryptology, 2009, 22(3): 365-394.

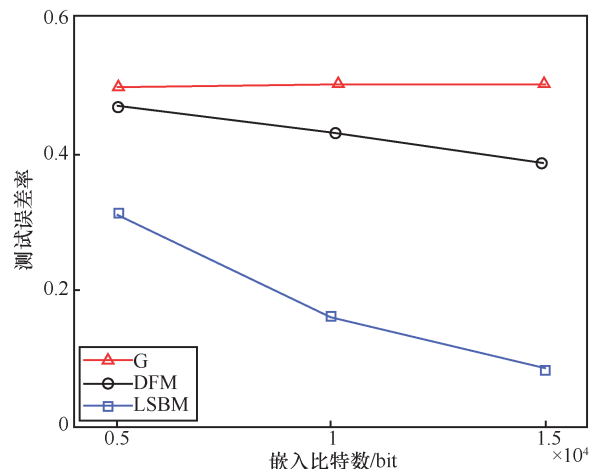


图 6 隐写分析测试错误率
Figure 6 Test error rate of steganalysis

4 结束语

深度学习的发展为隐写带来了新的机遇和挑战。生成式媒体为可证明安全隐写理论提供了实际应用载体。本文提出了一个计算可证安全的全流程隐蔽通信体系。本文基于音频合成模型 WaveNet 提出生成式隐写算法,在此基础上实现了对称密钥隐写和公钥隐写,并从实验和理论证

- [11] VON AHN L, HOPPER N J. Public-key steganography[M]//Advances in Cryptology-EUROCRYPT 2004. Berlin, Heidelberg: Springer, 2004: 323-341.
- [12] VAN LE T. Efficient provably secure public key steganography[J]. IACR Cryptology Arch, 2003: 156.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. 2014: 2672-2680.
- [14] KIPF T N, WELING M. Variational graph auto-encoders[J]. arXiv preprint arXiv:1611.07308, 2016.
- [15] YANG Z, DU X, TAN Y, et al. Aag-stega: Automatic audio generation-based steganography[J]. arXiv preprint arXiv:1809.03463, 2018.
- [16] SHI H C, DONG J, WANG W, et al. SSGAN: secure steganography based on generative adversarial networks[M]//Advances in Multimedia Information Processing – PCM 2017. Cham: Springer, 2018: 534-544.
- [17] SIMMONS G J. The prisoners' problem and the subliminal channel[M]//Advances in Cryptology. Boston, MA: Springer, 1984: 51-67.
- [18] VAN DEN OORD A, KALCHBRENNER N, VINYALS O, et al. Conditional image generation with pixel CNN decoders[J]. CoRR, 2016, abs/1606.05328: null.
- [19] GERMAIN M, GREGOR K, MURRAY I, et al. Made: masked autoencoder for distribution estimation[C]//International Conference on Machine Learning. 2015: 881-889.
- [20] URIA B, CÔTÉ M A, GREGOR K, et al. Neural autoregressive distribution estimation[J]. The Journal of Machine Learning Research, 2016, 17(1): 7184-7220.
- [21] OORD A, DIELEMAN S, ZEN H, et al. Wavenet: A generative model for raw audio[J]. arXiv preprint arXiv:1609.03499, 2016.
- [22] LI L, ZHANG W M, CHEN K J, et al. Steganographic security analysis from side channel steganalysis and its complementary attacks[J]. IEEE Transactions on Multimedia, 2020, 22(10): 2526-2536.
- [23] 国家密码管理局. SM4 分组密码算法 GM/T 0002—2012[S]. 北京: 中国标准出版社, 2012.
State Cryptography Administration. SM4 block cipher algorithm. GM/T 0002—2012[S]. Beijing: Standards Press of China, 2012.
- [24] DAEMEN J, RIJMEN V. Reijndael: the advanced encryption standard[J]. Dr. Dobbs Journal: Software Tools for the Professional Programmer, 2001, 26(3): 137-139.
- [25] 国家密码管理局. SM9 基于身份的公钥密码算法[S]. (2016:3). GM/T 0044.2016.
State Cryptography Administration. SM9 Identity-based public key cryptography algorithm[S]. (2016:3). GM/T 0044.2016.
- [26] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems[J]. Communications of the ACM, 1978, 21(2): 120-126.
- [27] KOMINEK J, BLACK A. The CMU Arctic speech databases[J]. Fifth ISCA Workshop on Speech Synthesis, 2004: 223-224.
- [28] LUO W Q, LI H D, YAN Q, et al. Improved audio steganalytic feature and its applications in audio forensics[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(2): 1-14.
- [29] CHEN K J, ZHOU H, LI W X, et al. Derivative-based steganographic distortion and its non-additive extensions for audio[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 2027-2032.

[作者简介]



李梦涵（1998—），女，河北保定人，中国科学技术大学硕士生，主要研究方向为信息隐藏。



陈可江（1994—），男，浙江温州人，博士（后），主要研究方向为信息隐藏、人工智能安全。



张卫明（1976—），男，河北定州人，中国科学技术大学教授、博士生导师，主要研究方向为信息隐藏、多媒体内容安全、人工智能安全。



俞能海（1964—），男，安徽无为，中国科学技术大学教授、博士生导师，主要研究方向为多媒体信息检索、图像处理与视频通信、数字媒体内容安全。