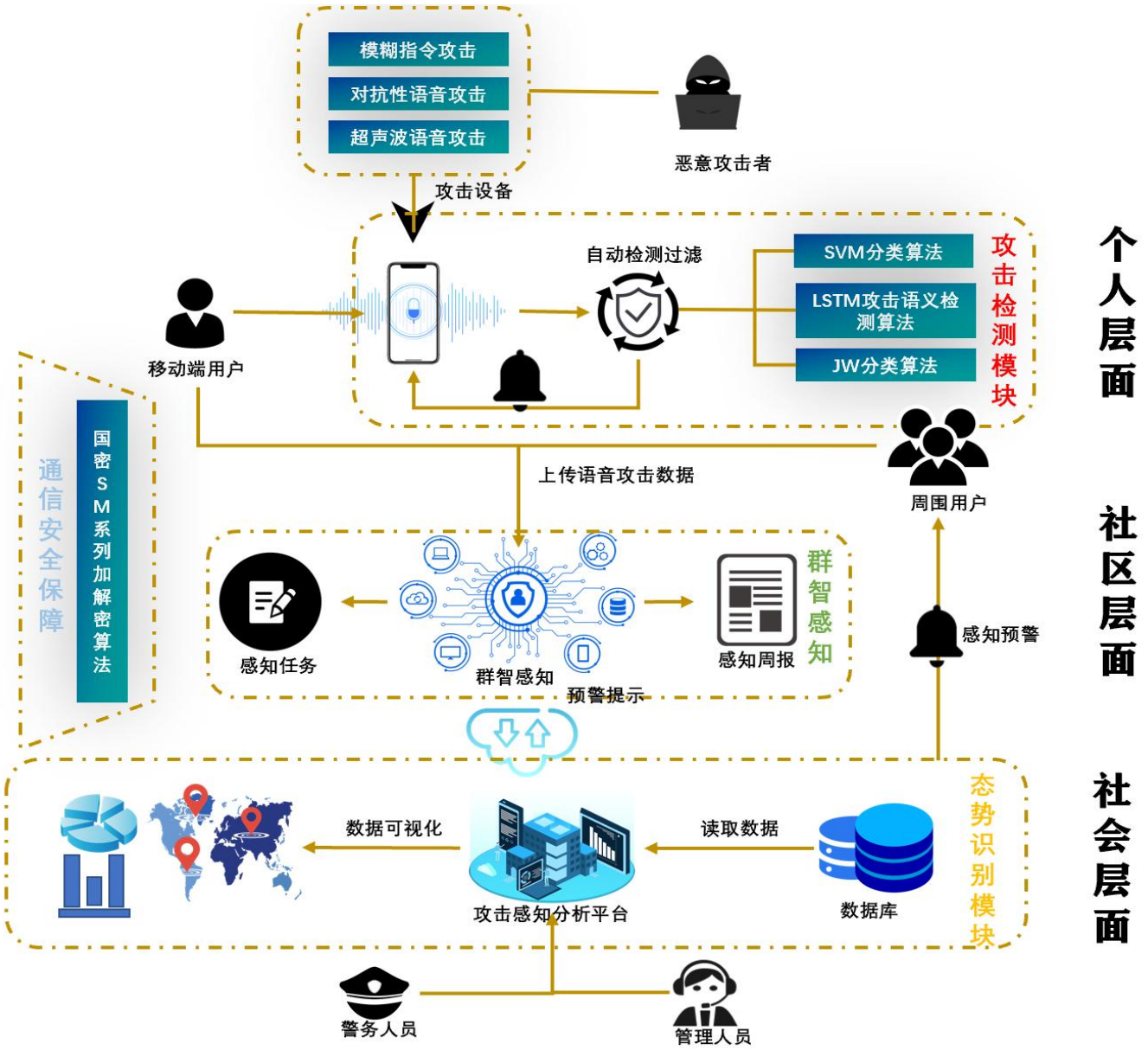
2024 | 中国高校计算机大赛
C4-NETWORK TECHNOLOGY CHALLENGE
网络技术挑战赛
适应新变·激励创新·甄选英才·助力产业

慧音 Guardian

Smart Sound Guardian

物联网语音安全领航者



作品设计文档

对应页码：P3 到 P85

目录

第 1 章 目标问题与意义价值

内容提要

- ❑ 背景分析
- ❑ 作品简介
- ❑ 特色描述
- ❑ 创新性说明

本章从作品背景分析、研究目标、特色表述和创新性说明四个方面对本系统进行介绍。

1.1 背景分析

近年来，人工智能给人们的日常生活带来了越来越多的便利，随着 5g 技术和边缘计算这两大技术的诞生，一个更加完美、更加便利的物联网时代即将到来。而智能语音系统作为人工智能的一个重要应用，伴随着嵌入式设备的日益普遍，在人类社会中扮演着越来越重要的角色，智能家居、智能办公、智能驾驶等物联网生态，无一不需要智能语音系统的支持。

随着人工智能的发展，使语音识别技术成为了人类社会中不可或缺的角色，根据头豹研究院发布的《2019 年中国智能语音行业研究报告》[17] 显示（如图 1-1 所示），预计未来几年中国智能语音市场规模将持续增长，并且根据预测显示 2023 年将超过 600 亿元。

图 1-1: 中国语音市场规模增长趋势

现有的智能语音控制系统主要以移动设备的语音助手，如 Siri、Cortana 等，以及家居智能语音系统，如小爱同学等为主。虽然此类智能语音系统对于语音识别的准确性上，已经有了较高的准确率。但随着智能语音系统在物联网生态中重要地位不断提升，现有的智能语音控制系统的安全隐患也逐渐显现出来，**无法对智能语音攻击进行有效识别与防御，缺乏及时的语音安全预警以及安全性低**成为了语音控制系统的一大痛点。

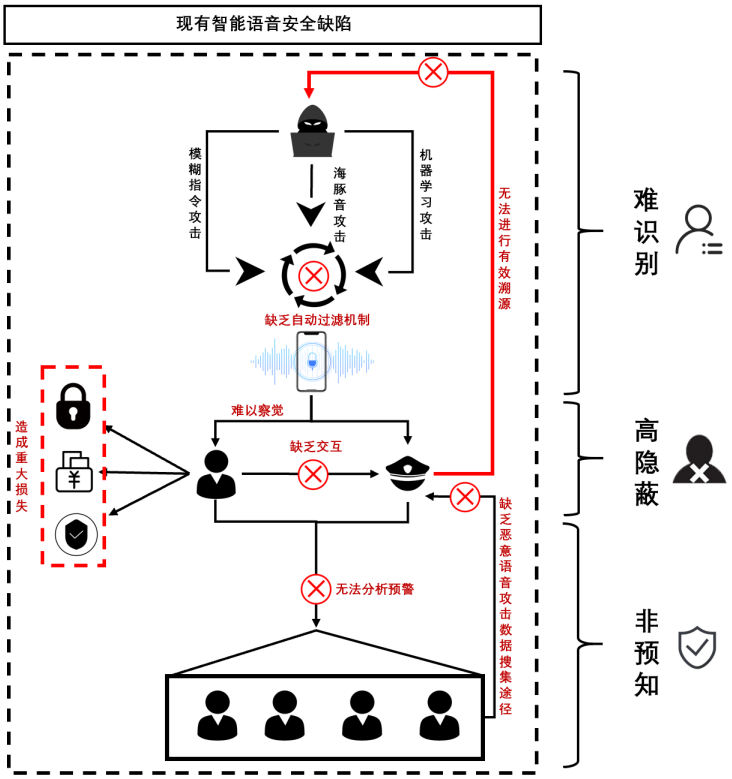
在这个对移动设备极度依赖的智能时代，用户的智能设备一般包含着用户的重要隐私信息、财产信息等重要信息。**而一旦智能语音控制系统被攻击，用户的设备相当于被攻击者直接控制，可能造成隐私、财产和安全等多方面的巨大损失**。

可感知的语音攻击层面，包括模糊指令攻击。香港知名企业汇丰银行，被孪生兄弟用相似声音通过身份验证发现的认证漏洞。日本电气通信大学和美国密歇根大学发现利用激光冒充人类的语音从而控制相连的设备。**而在不可感知的语音攻击层面**，浙江大学根据语音助手依赖的声音传感器成功攻破各大品牌语音助手，谷歌的 Google Assistant、苹果的 Siri、亚马逊的 Alexa、三星的 S Voice、微软的 Cortana 以及华为的 HiVoice 无一“幸免”。日本早稻田大学最近的一项研究将攻击命令转移为声波信息从而控制远距离的智能音箱。

可以说，语音识别漏洞所带来的威胁已经通过各种各样的攻击方式渗透到了用户的隐私、财产和安全的方方面面。

1.1.1 问题描述

如上所述，语音攻击已经威胁到用户的财产、隐私甚至生命安全，因此对于语音攻击的防范刻不容缓。智能语音攻击能够利用语音识别软件或者硬件系统的漏洞，在人完全意识不到的情况下攻击设备，并不定向地能够攻击多种语音助手。除此以外，现有的语音攻击已经不仅仅局限于可感知的方式，还包括了不可感知的方式。同时考虑到智慧社区的建设潮流，而现有的智能语音攻击数据无法得到有效收集与利用，这些都是现有智能语音攻击方案亟待解决的问题。



我们归纳总结了智能语音攻击难以防御的原因有以下五点：

（1）**智能语音攻击手段具有隐蔽性和非定向性**。现有的智能语音攻击能够在人完全意识不到的情况下攻击设备，同时不能够定向地攻击多种语音助手。

（2）**语音攻击方式具有发展性和多样性**。除了能感知的传统攻击方式以外，又诞生了难以感知的新型攻击方式，越来越多攻击方式使得设备对智能语音攻击的检测、防御系统要求越来越高。

（3）**现有语音防范系统缺乏自动检测和过滤能力**。现有产品仅能对某种特定攻击实现检测，但识别准确率较低，并且没有相应的后续补救措施。

（4）**市场急需语音设备物联网实现预警，形成社区预警防御机制**。现有的防御方案，仅仅只能从个体层面形成对于恶意语音攻击的识别与防御。

（5）**现有方案无法利用现有数据对区域语音攻击态势进行感知预测**。缺乏对于区域整体态势的感知与预测，无法根据总体态势及时制定合理措施。

1.1.2 解决方案

基于上述问题，我们创新性地提出了慧音 Guardian，一个专门为物联网设备提供语音安全保障的智能语音攻击防御系统。

我们主要从以下三个方面解决目标问题：

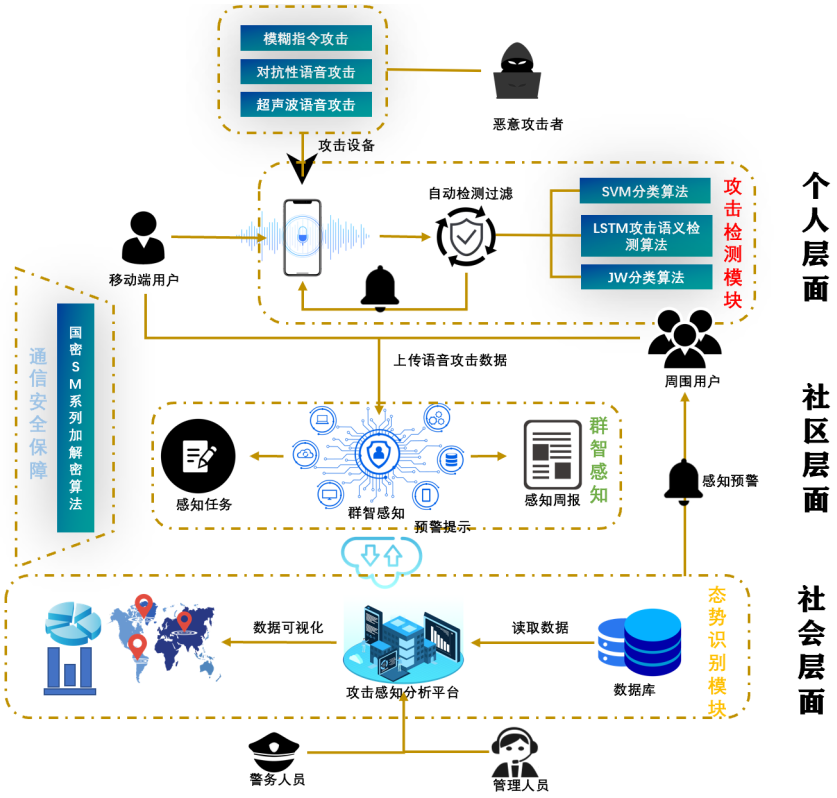
（1）**从个人层面**，本系统通过构建语音攻击模拟场景，使用 SVM 算法对正常语音命令与常见三类（模糊指令攻击、对抗性语音攻击、海豚音攻击）攻击语音进行分类并且过滤，再通过 JR 算法将攻击语音进行在分类，准确判断是哪一类语音攻击。

（2）**从社区层面**，通过构建基于群智感知的实时监测大数据分析平台，记录每一次攻击数据，警示被攻击用户周围其他用户形成辐射保护，并通过可视化的方式呈现攻击记录，提醒用户免受再次攻击，并通过攻击趋势形成溯源机制。

（3）**从社会层面**，引入态势感知技术，构建基于 GRU 和 CAM 的语音安全态势感知平台。利用的语音攻击记录，实现对于区域内的智能语音攻击的预测预警，便于管理人员或警务人员采取针对性的措施，调用社区力量。

1.2 作品简介

为解决恶意智能语音攻击问题，慧音 Guardian—智能语音攻击防御系统提供了一种多层次、多维度的智能语音防御方案。为了解决现有恶意语音攻击**难识别、非预知、高隐蔽**的痛点，慧音 Guardian 以基于 SVM 分类算法、LSTM 攻击语义检测算法、JW 分类算法的恶意智能语音攻击分类检测为核心，结合慧音 Guardian—语音安全 APP 形成用户设备语音安全保护屏障。通过慧音 Guardian—感知管理平台，引入群智感知技术与态势识别技术构建大数据实时监测中心，实现恶意语音攻击的预警与溯源。从个人层面、社区层面、社会层面形成智能语音安全生态三重保护。整体系统功能如图所示：



作品功能介绍：

（1）慧音 Guardian—语音安全 App，高效识别过滤恶意智能语音攻击，保障用户设备语音系统安全。

本系统针对三种具有较强威胁性的语音攻击，设计了一个识别过滤模型。当用户设备受到语音攻击时，该模型能够将攻击语音和正常语音加以区分，避免用户设备（主要为移动终端）被攻击者恶意操控。因为智能语音信息可能会包含大量用户个人隐私，所以慧音 Guardian App 采用发布任务的方式，鼓励用户在思考后上传恶意攻击语音，参与到语音安全生态保护中来。

（2）慧音 Guardian—感知管理平台，使用群智感知的方法实时收集大量终端用。

户的语音攻击信息并通过大数据分析平台检测样本数据。群智感知是结合众包思想和移动设备感知能力的一种新的数据获取模式，是物联网的一种表现形式。本系统构建了一个“中心—用户”的感知平台，中心通过发布感知任务、由终端用户执行任务、最终收集到有效的语音攻击信息；配合以适当的激励机制、鼓励和刺激参与者参与到感知任务中，并提高感知数据的质量和可靠性。

（3）构建了基于 NIN 网络和 CAM 的具有语音攻击预测态势识别，实现恶意语音攻击预警。

系统通过对搜集到的数据清理、分析、归纳，对区域进行智能预警，判定该区域发生潜在恶意语音攻击的风险，并利用深度学习模型实现了安全态势感知的功能，在更大范围的社会网络中为将来可能出现得恶意语音攻击事件提供更好的防范措施，从而做到区域语音生态保护。

（4）从音频数据传输角度构建可靠的保护系统信息安全体系。

本系统在客户端与服务端通信时采用国密 SM2/3/4 算法，保障传输音频信息和预警信息的安全性。

总体而言，本系统是一款识别精度高、预警能力强、能够有效防范多种语音攻击的语音安全保障系统。

1.3 特色描述

慧音 Guardian 构建了多维度语音安全保障系统，以**智能语音识别过滤**为核心，结合群智感知、态势识别等技术支持，有效的解决了智能语音攻击**难识别**、**非预知**、**高隐蔽**的痛点，同时具有恶意攻击预警、感知任务、智能周报等功能。

本作品相较于传统技术和产品具有以下几大创新点：

（1）**基于支持向量机算法设计了恶意语音音频分类算法**。作品可以从音频信号层面对三种最有威胁的恶意语音攻击进行分类过滤，平均识别准确率达到 88.7%，最高识别准确率达到 96.2%，实现恶意语音攻击的有效识别与过滤。

（2）**基于 LSTM 算法和 Jaro-Winkler 算法设计了智能攻击语义检测过滤算法**。本系统可以从语音内容层面对具有威胁性的语音音频进行识别过滤，从而更进一步的保障移动设备的语音安全。

（3）**创新性地将基于 NIN 网络和 CAM 的具有攻击预测感知功能应用于智能语音攻击领域**。针对潜在攻击，本作品能通过 CAM（类激活映射）态势感知以及走失预测与预警功能，更加具有针对性地提出防范措施以实现安全预警的功能。

（4）**构建了基于群智感知的实时数据收集与监测平台**。传统防范技术只停留在检测语音攻击是否存在的层面，本作品致力于将终端设备和数据监测平台“连接”起来，使用户既是被保护者也是参与者。采用群智感知的方法构建语音安全生态网络，将攻击源搜索效率至少提高 80%。同时该平台还使用了基于国密 SM2/SM3/SM4 算法的加解密 KPI 模型，极大地提升了数据通信安全水平。

1.4 创新性说明

慧音 Guardian 构建了智能语音攻击防御系统，以**智能语音识别过滤**为核心，结合群智感知、态势识别等技术支持，有效的解决了智能语音攻击**难识别**、**非预知**、**高隐蔽**的痛点，同时具有恶意攻击预警、感知任务、智能周报等功能。

本作品相较于传统技术和产品具有以下几大创新点：

（1）**基于支持向量机算法设计了恶意语音音频分类算法**。将梅尔频率倒谱系数（MFCC）、短时平均过零率（CER）、均分误差（RMSE）作为特征可以从音频信号层面对三种最有威胁的恶意语音攻击进行分类过滤，平均识别准确率达到 88.7%，最高识别准确率达到 96.2%，实现恶意语音攻击的有效识别与过滤。

（2）**基于 LSTM 算法和 Jaro-Winkler 算法设计了智能攻击语义检测过滤算法**。本系统可以从语音指令内容层面对具有威胁性的语音音频进行识别过滤，通过分词处理、构建词向量、提取关键词、距离计算等步骤对语音文本信息进行三分类处理，保证语音内容合法性，更进一步的保障移动设备的语音生态安全。

（3）**创新性地提出了基于 NIN 网络和 CAM 映射的语音攻击态势预测模型**，并应用于智能语音攻击防御领域。针对潜在攻击，本作品通过实现态势感知以及走失预测与预警功能，能够让用户提出更加具有针对性的防范措施，实现安全预警与攻击溯源。

（4）**构建了基于群智感知的实时数据收集与监测平台**。本作品致力于将终端设备和数据监测平台“连接”起来，使用户既是被保护者也是参与者。采用群智感知的方法构建语音安全生态网络，将攻击源搜索效率至少提高 80%。同时平台使用了基于国密 SM2/SM3/SM4 算法的加解密 KPI 模型，极大地提升了数据通信安全水平。

解决了传统智能语音攻击防御方案中**识别精度低、类别少，无法提供精确预警，无法有效溯源**的痛点。

1.5 小结

我们对本作品做了一个整体的介绍，包括作品的背景及研究意义、特色描述以及应用前景等。从当前社会存在的热点问题——语音攻击问题入手，进行了充分的调研，分析了本作品的主要研究内容和优势，为后期作品的设计与实现奠定基础。

第 2 章 需求分析

内容提要

- 智能语音攻击防御领域
- 现有智能语音攻击防御系统解决方案
- 群智能感知技术中激励机制

本章主要调研了目前在在音频攻击防御和群智感知领域已有的相关研究和产品，并对它们进行了分析和比对。

2.1 智能语音攻击防御领域的论文调研

1.2017 年论文《Mel Frequency Cepstral Coefficient (MFCC) tutorial》[12] 中讲述了多维度音频信息 MFCC 参数提取的方法。包括音频能力、频谱、过零率等信号参数的提取与分析。

2.2016 年论文《Hidden Voice Commands》[3]、2017 年《DolphinAttack: Inaudible Voice Commands》[16]、2018 年《Audio Adversarial Example: Targeted Attacks on Speech to-Text》[2] 三篇论文中讲述了现有语音助手的实现原理以及语音攻击音频的制备原理。包括三种主要的语音攻击——模糊指令攻击、海豚音（即超声波）攻击和对抗性语音攻击。这三种语音攻击隐蔽性极高，现有防御系统对上述三种攻击识别率较低。

3.2006 年论文《labelling unsegmented sequence data with recurrent neural networks》[7]、2014 年《Deep speech: Scaling up end-to-end speech recognition》[8] 介绍了基于 CTC 的语音识别系统的工作原理，并以一款基于端到端的语音识别系统 DeepSpeech 为例详细介绍了语音识别神经网络的搭建过程，使我们对其有了一个充分的了解。

4.2015 年论文《Understanding LSTM Networks》[13] 中讲述了 LSTM 神经网络的基本原理，并将其运用在文本情感分析领域，对后续构建语音语义分析提供了思路。2016 年论文《群智感知激励机制研究综述》[18] 中对群智感知激励机制进行了详细的分类讨论，我们从中选择了适用于本系统的两种激励模式——多属性拍卖和虚拟积分模式。

5.2012 年论文《Monetary incentives in participatory sensing using multi-attributive auctions》[9] 介绍了 Krontiris 等人采用逆向拍卖中的多属性拍卖机制，不仅考虑参与率问题，还考虑到感知数据的质量问题。服务器平台能够通过拍卖过程影响数据质量，同时，参与者能够通过拍卖结果的反馈提高自身感知数据的质量，从而提高竞标价格。

6.2014 年《a new paradigm for application simulations and measurements》[5] 该论文的核心思想是将群智感知应用中所面临的较大数据分配问题进行数据重组、分布计算。在此基础上，我们解决了感知参与者位置不确定性的问题，提出了多目标参与者选择算法。

2.2 群智感知技术中激励机制的调研

群智感知是指大规模的用户通过其携带具备感知、计算能力的移动终端采集并共享感知数据，对数据进行测量、分析、估计等处理后提取与公共利益相关现象或信息的技术。群智感知是物联网与众包思想的结合，在 Infocom、Ubicomp、Percom 和 Mobicom 等国际知名学术会议以及 UCLA、CENS、UCB 和 CSD 等著名研究单位都是新颖的热点问题。

群智感知任务的执行依赖于大量用户的参与，需要消耗用户的时间精力以及其智能终端设备的电量、存储与计算资源，并且存在泄露用户隐私的风险。用户应被给予相应的报酬以激励其参与感知任务，但用户都是自私的，可能会发起欺骗或合谋攻击来获取更多的奖励。因此，设计一种安全可信的激励机制就显得尤为重要。我们对常见的三种激励机制进行了调研。

信誉机制

信誉机制评价用户的信誉值，高信誉用户可获得更好的服务。可以通过采用信誉机制，隔离感知系统中低水平工作者以激励高水平工作者参与感知任务，从而获得高质量的任务解决方案。

互惠机制

互惠机制根据用户贡献度匹配等价的服务。通过研究在给定了社会信任机构的基础上，构建基于社会信任的互惠激励机制，并对该激励机制用户的响应效率进行了深入的研究。

激励机制

基于电子货币的激励机制使用电子货币激励用户参与群智感知任务。群智感知中基于质量的综合定价机制，可以根据感知质量水平得到工作者的客观排名。根据贡献程度作为支付标准来设计激励机制，可以提高了理性参与者上传高质量感知数据的积极性。

2.3 现有智能语音攻击防御系统解决方案的调研与思考

我们调查研究了目前现有智能语音攻击防御方案，并选出了其中的典型解决方安，进行简单介绍。

（1）针对运营重放攻击的方案

针对语音重放攻击，大量工作已经提出了相应的防御方案 [4]。虽然目前该类攻击已经得到有效防御，但攻击者也很少采用此类攻击方式。

（2）AuDroid

2015 年，G.Petracca 等人设计的 AuDroid[14] 就已经提出来管理音频通道权威。通过针对不同的音频通道使用模式，设置不同的安全级别，AuDroid 可以抵御使用设备内置扬声器的语音攻击。但是，AuDroid 使用说话者验证系统来防御外部重放攻击，显然这种防御效果不够好。

（3）VSButton

2017 年，X.Lei 等人设计的基于亚马逊的 Alexa 系统 [10]，提出了一个虚拟安全按钮（VSButton），它利用 Wi-Fi 技术来检测室内人体运动，并且只有在检测到人体运动时才接受语音命令。但是它也存在一定的局限性，因为语音命令并不一定都伴随着可检测的运动。

（4）VAuth

H.Feng 等提出了 VAuth[6]，它通过可穿戴设备收集用户的身体表面的振动，来保证语音命令来自用户。VAuth 的限制是用户必须穿戴一定的设备（即耳塞，眼镜和项链），这会给用户造成许多的不便。

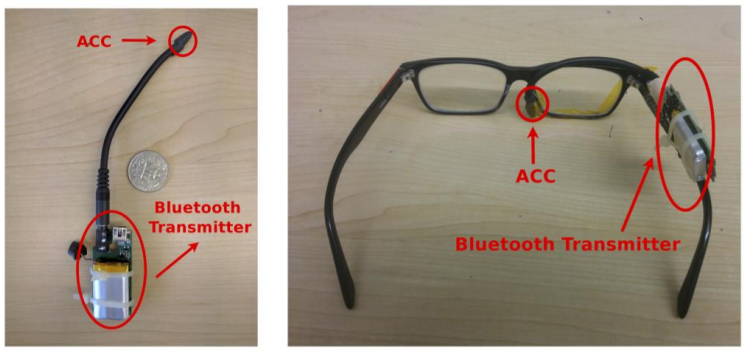


图 2-1: VAuth 可穿戴设备

（5）磁力仪防御系统

S.Chen 等 [4] 提出了一种防御系统，通过磁力仪确定语音命令的来源是否是扬声器。如果是，就拒绝这些命令。这项工作的缺陷在于它的防御距离只能达到 10 厘米，这比通常的人与设备之间的距离还要小。此外，这种方法不适用于非常规扬声器和恶意信号发生器。

分析以上结果我们可以得出，现有的解决思路都是在避免设备受到一次语音攻击，即在努力提高语音攻击的识别精度或检测用户行为特征来避免设备受到攻击。这些产品或方案没有从源头上彻底捣毁攻击源，也没有

实时监测模块来实时预警用户，用户仍可能再次受到攻击。本系统在已有工作的基础上不仅构造了针对多种语音攻击的防御模型，还利用群智感知的思想构建了一个语音设备物联网，绘制动态热力图呈现攻击源分布，为用户提供了一个安全、精确、可溯源的语音安全保障系统，避免受到二次攻击。

2.4 小结

在本章中，对智能语音攻击防御领域论文、群智能感知技术中激励机制以及现有智能语音攻击防御系统进行了调研，分析了现有方案的不足，并在此基础上进一步说明了我们的作品的改进点与优势。

第 3 章 设计思路与方案

内容提要

- 系统整体方案设计
- 架构设计
- 系统核心技术
- 系统功能模块设计
- 安全体系设计

本章主要对本系统的结构、算法和功能进行设计和阐述，通过规范的设计方法，明确系统的框架和流程，为后一章节的实现奠定基础。

3.1 系统整体方案设计

我们设计的多维度语音安全保障系统主要分为系统客户端、身份认证平台、大数据分析平台、感知任务平台和音频发生设备五个角色。本系统的总体流程设计如图 3- 1 所示。

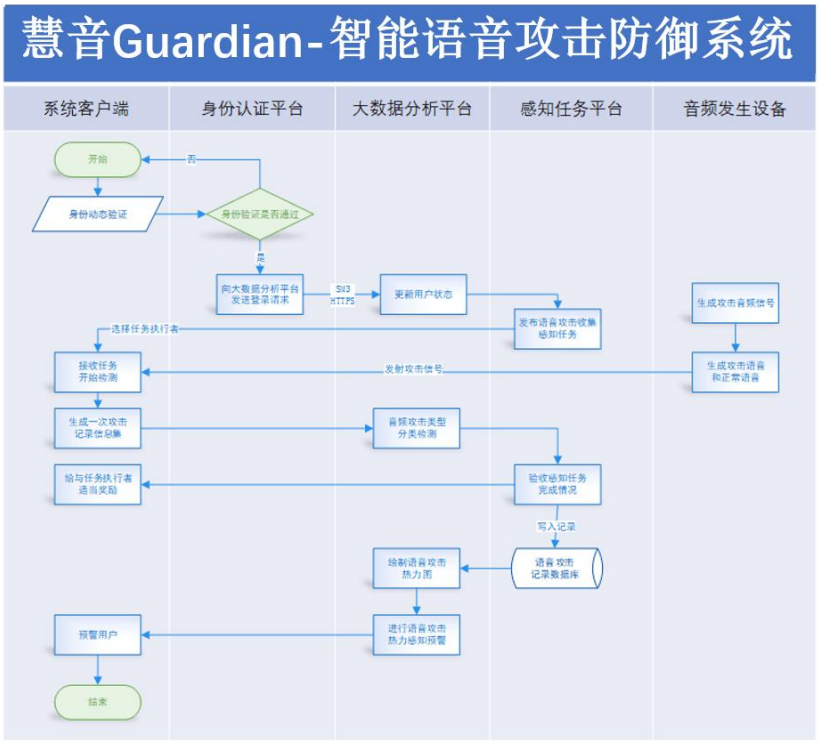


图 3-1: 智能语音攻击防御系统设计图

（1）**系统客户端**：用户与实时数据监测平台以及音频发生设备交互的工具，用户通过客户端发起一系列工作事项，包括身份验证、数据收集、记录生成、预警显示等。同时系统客户端还作为感知任务的执行者，帮助中心收集语音攻击数据。

（2）**身份认证平台**：用户在平台登录时首先会访问数据库检测该用户是否已经注册。若用户已注册则输入口令及动态验证码即可成功进入系统，或选择通过身份信息验证及动态验证码找回密码；若用户未注册则采集用户信息并要求用户设置强口令，完成注册。

（3）**大数据分析平台**：对上报音频进行特征提取、分类器识别、生成攻击记录等操作，如果判断为语音攻击，则向客户端返回警告信息，并将本次攻击添加攻击热力图。

（4）**感知任务平台**：负责感知任务的发布以及对用户反馈的感知数据的提炼归纳，并将其写入攻击案例库中。通过智能合约的方式进行预设任务内容、完成要求和审查任务完成方式，合约状态和合约值的设定决定了不同的任务内容。

（5）**音频发生设备**：生成用于模拟攻击的攻击音频，在不同地理位置、以不同攻击类型攻击用户设备，以测试系统功能。

3.2 架构设计

本系统的架构将从网络架构、软件架构、硬件架构、安全体系设计四个角度进行详细说明，为后续说明提供整体概念。

3.2.1 网络架构

整个系统采用“C/S + B/S”混合架构模式进行开发，前者主要用于与普通用户的交互，而后者用于与服务端管理员进行交互。服务端主要的业务功能就是实现对数据的收集和处理；客户移动端主要的功能就是实现对音频的采集和上传，并根据服务端返回的攻击种类实施对应的操作。本系统的网络架构设计如图 3-2 所示。

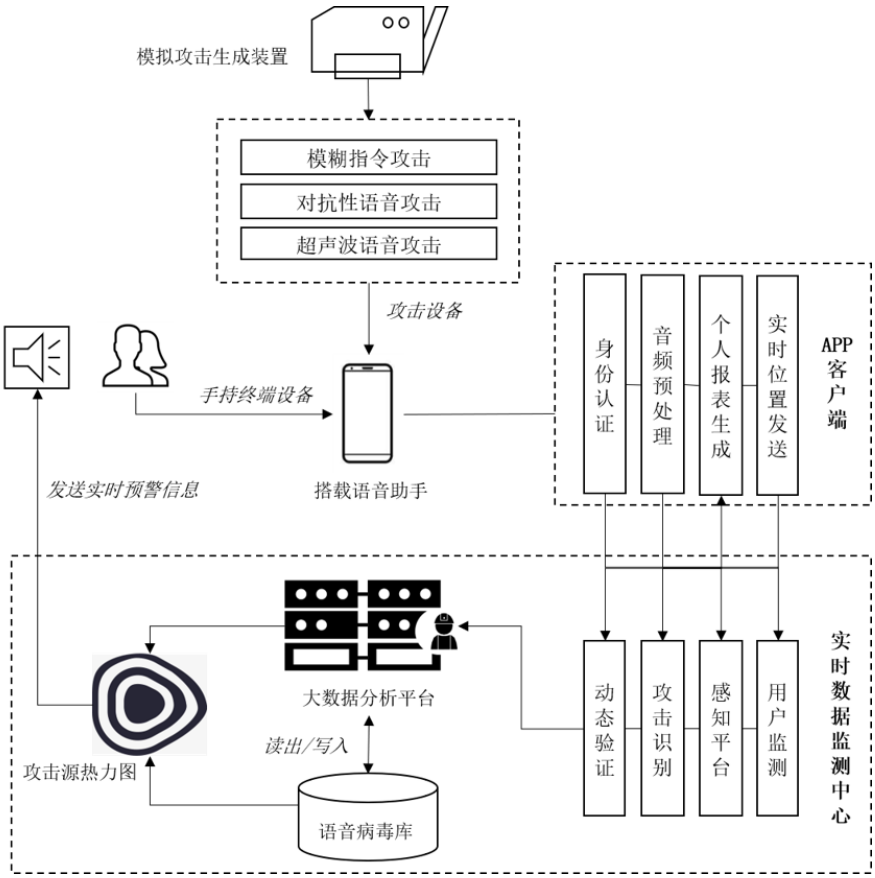


图 3-2: 智能语音攻击防御系统网络架构设计

（1）**C/S 架构的选取**：该产品的定位是常驻与用户后台进程中，以类似于一个移动管家的角色实时为用户的语音系统保驾护航，选取在 Android 平台上开发用户客户端系统。后端的体系架构虽然现在有很多语言都支持

基于 UWSGI 的 HTTP 服务器，但由于后端要能够在响应用户请求的基础之上实现对数据的分析处理，实现后台处理智能化，所以最终采用 Python 进行开发。

（2）**B/S 架构的选取**：实时监测平台的开发采用 B/S 架构模式，通过收集并分析用户上传来的数据，呈现出一个实时的、动态的音频攻击热力图，并以此协助管理员对音频攻击源的排查。管理员只需打开网页，就可实现对当前数据的可视化，保证产品简单易上手、功能全面，并方便操作管理。

（3）**客户端主要的功能就是实现对音频的采集和上传，并根据服务端返回的攻击种类实施对应的操作**。例如当检测到恶意攻击音频时，客户端会自动过滤掉该音频，防止语音助手响应该非法操作。另外，用户还能根据实时地理位置预防自身进入非法音频高发区。通过搜集用户设备上的攻击音频，上传至数据库供服务端进行使用。

（4）**服务端主要的业务功能就是实现对数据的收集和处理**。那么结合 Spark 大数据分析框架和深度学习算法就可实现诸如海豚音识别、机器学习级攻击识别和地理信息去杂、可视化等操作。

（5）**国密 SM2/3/4 算法应用到音频加密传输领域**。在客户端和数据中心交互过程中，利用国密算法 SM 系列完成密钥协商和数据加密，保证音频数据传输的安全，提高了系统的安全性。

3.2.2 软件架构

本系统为基于音频识别的语音安全保障系统，系统采用分层架构模式，从存储层 -网络层-逻辑层-接口层-应用层-表示层六层设计系统整体架构。下面将详细介绍分层架构图的设计，如图 3-3 为系统分层架构图：

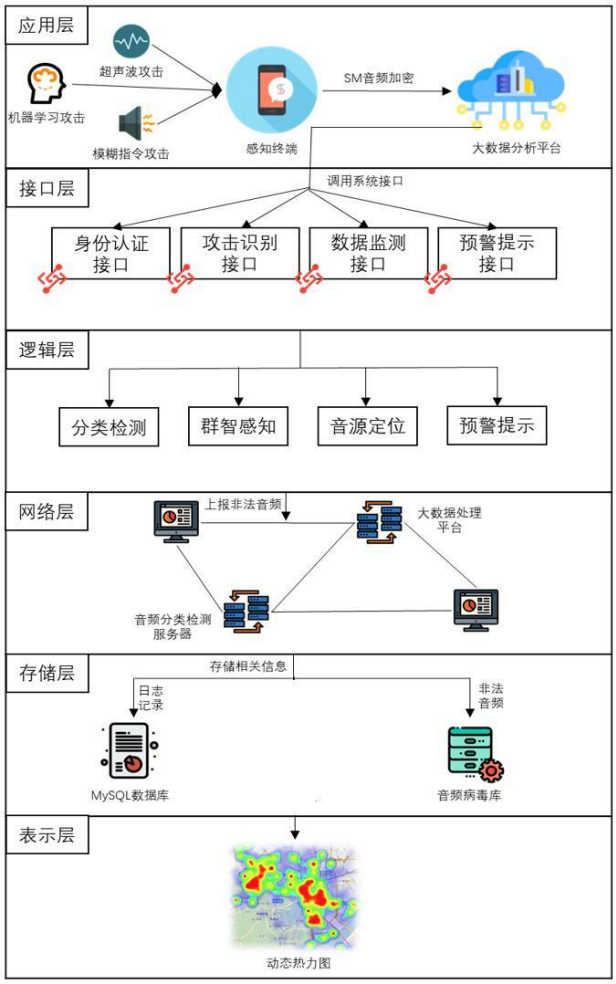


图 3-3: 智能语音攻击防御系统软件架构图

（1）应用层

应用层包含音源制备、移动终端收集音频、音频识别和加密传输等部分。超声波制备设备制备出和正常语

音指令同样内容的超声波，并在不同位置站点发射声波。移动终端接收到数据后生成 wav 音频文件，并上传至中心。这里采用国密加密传输算法，可以避免音频被窃取和篡改，增强系统的安全性。

（2）接口层

接口层包含了身份认证接口、攻击识别接口、数据监测接口和预警提醒接口，提供给上层用户调用接口来过滤识别非法音频，并将非法音频上报至音频病毒库。当上层用户进入已有攻击区范围时，会自动调用预警，实时控制中心会向终端设备发送警告，避免二次攻击。

（3）逻辑层

逻辑层在本系统中作为处理逻辑业务的模块，为上层接口提供了可调用的函数。应用层发起音频识别和分类请求，将通过非法音频过滤接口调用逻辑层函数实现音频识别和分类。群智感知作为系统调用函数，当系统想要收集用户终端非法音频信息时，系统发起感知任务，调用系统函数，向所有终端发起调用，从而帮助系统构建中心病毒库，实现预警和防范的功能。

（4）网络层

网络层采用端到端模式，由中央服务器和大数据中心服务器组成。中央服务器处理基本数据通信和交互，而大数据中心服务器依据 spark 大数据分析及可视化绘制实时热力图，防范预警。

（5）存储层

存储层包括两大数据中心——系统中央数据库和中心音频病毒库。其中系统中央数据库用来存储用户基本信息、移动设备信息、音频分类信息等系统基础模块信息；中心音频病毒库是我们为群智感知收集非法信息而专门设立的一个数据库，主要存储过滤识别出来的非法音频的信息、音源地点、时间等信息，为后续打击非法音源站点提供基础性帮助。

（5）表示层

表示层将搜集得到的数据通过热力图算法进行汇总，构建动态热力图和，为用户提供直观的可视化预警。

3.2.3 音频发生器硬件架构

本系统采用 RIGOL DG4062 作为任意波形信号发生器，U 盘搭载音频信号，输出端外接 2 个超声波发射探头，固定于发射挡板并对准手机麦克风，构建成超声波模拟攻击场景；本系统又利用音箱播放模糊指令，手机作为音频收集设备实时采集音频信号，构建成机器学习级别的模拟攻击场景。本系统的硬件架构图如图 3-4 所示：

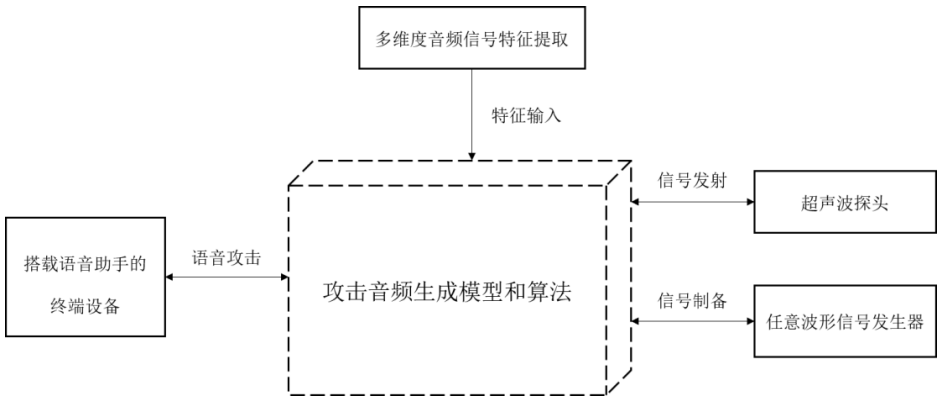


图 3-4: 智能语音攻击防御系统硬件架构图

3.3 系统核心技术

3.3.1 攻击语音检测过滤模型

3.3.1.1 多维度音频特征提取算法

本系统对语音攻击进行分类和防范，其首要工作就是要能够提取和把握攻击音频和正常音频的特征区别。尽管攻击音频是通过正常语音制备而来，但其本身仍与原始语音具备较大差异性，这为我们的分类提供了理论合理性。通过分析不同类型攻击音频的制备过程，我们认为可以通过能量、频率和波动性三个维度来对音频特征进行把握，原因如下:

①尽管利用了录音系统功率放大器非线性变换得到的音频和原始音频的频谱非常相似，但是由于谐波的自卷积，恶意音频在低频具有更高的能量比例，这体现了正常音频和攻击音频的频率差异；

②当音频的能量总体增强时，恶意音频在低频部分的能量会随之增加，但是合法音频的低频能量增加很少，这体现了正常音频和攻击音频的能量差异；

③在时域上，恶意音频在正方向的振幅偏移量高于负方向振幅偏移量，而在合法语音中这两部分偏移量几乎相等，这体现了正常音频和攻击音频的波动性差异。而能量、频率和波动性三个特征就对应了音频的 MFCC 系数、过零率和 RMSE。

(1) MFCC 特征参数提取

MFCC 是梅尔频率倒谱系数的缩写，它基于人耳听觉特性计算，能够较好地反映音频的能量特征。具体提取流程如图 3-5 所示:

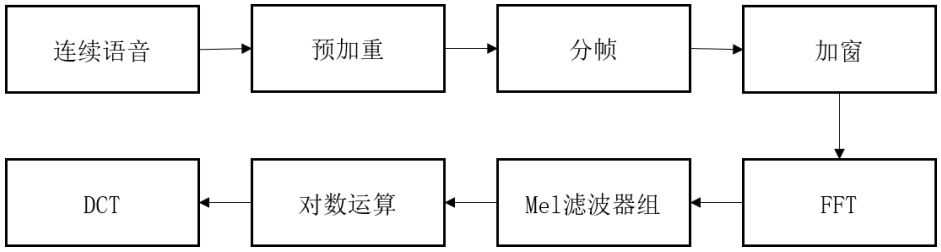


图 3-5: MFCC 特征提取流程

a) 我们首先将音频通过一个一阶有限激励响应高通滤波器，使信号的频谱变得平坦。然后我们将音频切割成帧长为 32ms、帧叠为 16ms 的片段，便于后序处理。

b) 接着我们对所有语音片段按照公式（3.1）进行快速傅里叶变换，将时域信号变为功率谱。

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{-f\frac{2\pi}{N}} \qquad k=0,1,\cdots,N-1 \tag{3.1}$$

c) 接着用 24 个三角窗滤波器对信号的功率谱滤波，并对输出值进行求对数操作。

d) 最后再进行离散余弦变换 DCT，取前 13 个系数即为我们所求的 MFCC 特征系数，我们用该系数表示音频的能量特征。

(2) 短时平均过零率 Cross_Zero_Rate

我们特征提取算法基于的第二项特征是音频的过零率，它良好地反映了受测音频的频率变化特性。短时平均过零率是指每帧内信号通过零值的次数，对有时间横轴的连续语音信号，可以观察到语音的时域波形通过横轴的情况。

在离散时间语音信号情况下，如果相邻的采样具有不同的代数符号就称为发生了过零，因此可以计算过零的次数。我们系统中所用的过零率是整个音频文件切分成的所有片段的平均过零率，它可以帮助我们衡量音频频率变化的快慢。

我们用公式（3.2）计算音频每个片段的过零率：

$$Z = |sgn[x(n)] - sgn[x(n-1)]| \times w(n) \tag{3.2}$$

其中 sgn[] 为符号函数，即

$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{3.3}$$

(3) 均方误差 RMSE

我们选取的区分音频的第三维特征是均方误差，它衡量了音频的波动性特征。我们依据音频的时间-振幅序列，计算音频振幅的均方误差如下：

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{3.4}$$

综上，我们获得了输入音频的 MFCC 参数、过零率和 RMSE，分别用来反映音频的能量、频率和波动性特征。我们将三个返回矩阵展开成一维然后进行拼接，便可以得到我们的目标特征向量，可以作为我们后序的系统输入。

3.3.1.2 基于支持向量机的音频分类过滤机制

本系统针对正常音频和超声波、机器学习等多种攻击音频的特征差异，基于 MFCC 算法提取音频的能量、频率和过零率等特征作为输入，为每一类攻击音频单独设置分类标签，训练 One-Versus-One 的 SVM 多分类模型，从而更好的把握语音特征，提高分类精度。

(1) 训练数据获取

本系统的训练数据主要包括如下类别：正常语音、超声波攻击语音、多种机器学习攻击语音等。训练音频数据集采集于多类经典语音攻击制备论文，参考文献 [1–3, 15, 16]。攻击语音制备原理分别基于混淆黑盒攻击、基因算法黑盒攻击、白盒攻击等。

我们基于 3.3.1.1 节阐述的 MFCC 特征提取算法，从训练音频中提取音频特征矩阵作为模型输入。

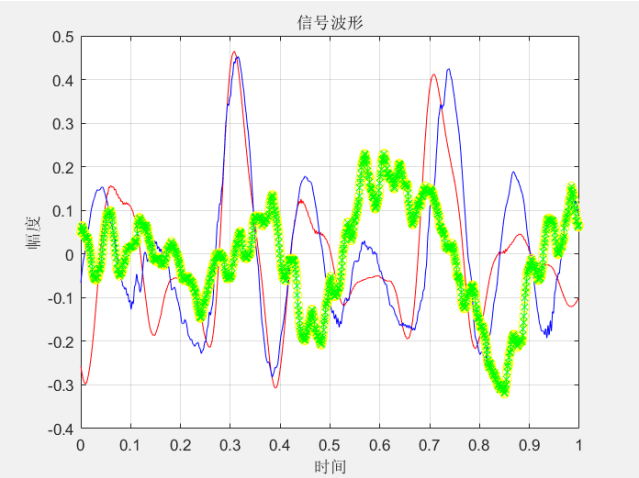


图 3-6: 多类别攻击音频时域特征对比图

如图 3-6 所示，我们绘制了多种类别攻击音频的时域特征对比图，音频内容均为“你好，Siri”中截取的相同片段, 图中蓝色线条为原始语音波形。

绿色线条为通过混淆黑盒攻击方式制备的攻击音频，因为这种攻击原理是扔掉部分音频频段，所以可以观察到波形整体振幅降低。

红色线条是基因算法黑盒攻击语音，因为其制备原理是在全音频范围随机添加噪音进行迭代，可以观察到整个红色线条就是围绕原语音进行上下波动。通过上述实例分析，可见不同攻击音频之间有着显著差异，通过音频特征训练模型具备合理性。

(2) 模型训练过程

为得到较好的泛化能力及非线性分类效果，我们使用基于高斯核的软间隔支持向量机，来解决多类别攻击语音的识别分类问题：

$$\begin{aligned} &\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \zeta_n \\ &\text{s.t. } y_n(w^T x_n + b) \geq 1 - \zeta_n \quad n = 1, 2, \cdots, N; \quad \zeta_n \geq 0 \end{aligned} \tag{3.5}$$

式中，$w^T x_n + b$ 为分类超平面，x_n 为通过 MFCC 算法提取的音频特征矩阵的列向量；y_n 为不同音频的分类标签；w 为垂直于音频分类模型的超平面法向量；b 为模型超平面相对于坐标原点的偏移；C 为惩罚系数；ζ_n 为松弛变量，用于一定程度上消除音频噪声的影响；N 为训练样本数；该分类超平面可表示为下式：

$$w^T x_n + b = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b \tag{3.6}$$

式中，$K(x_i, x)$ 即为高斯核函数，x_i 为第 i 个训练特征样本，x 为测试特征样本；α_i 为拉格朗日系数。为取得较好训练效果，本文采用网格搜索的方法对分类参数进行寻优。寻优参数分别为惩罚系数 C，以及核函数参数 γ。其中 γ 的搜索范围为 $[2^{-10}, 2^{8}]$，以步长为 2 指数递增；C 的范围为 $[2^{-10}, 2^{10}]$，步长也为 2 的指数递增。

由于支持向量机解决二分类问题，为建立音频分类模型，本系统采用 one-against-one 的方法进行多分类。假设有 k 个音频类别，训练阶段两两组合进行训练，最终得 k(k−1)/2 个分类模型。测试阶段采用投票的方式。先将各类得票数置为 0；再对测试数据使用训练所得分类模型进行分类，若结果为第 i 类，则该类得票数加 1，以此类推；选择得票最多的那个结果作为最终音频分类结果。

3.3.1.3 基于 LSTM 和 Jaro-Winkler similarity 的语音文本情感检测

针对在音频信号层面不具有攻击性的音频，有可能在指令内容层面具有威胁性。例如某段音频经过识别后被翻译为“转账给 xxx”，那么我们仍有必要对用户进行一定程度的提醒。这里我们采用两套方案对语音语义进行检验，以降低误报率，提高检验精确度。

(1) 基于 LSTM 进行三类文本情感分析

针对某段指令内容 x，将其划分为三种类别输出 y 来表示不同极性程度：

$$y = \begin{cases} +1 & x = \text{positive} \\ 0 & x = \text{neutral} \\ -1 & x = \text{negative} \end{cases} \tag{3.7}$$

其中 positve 代表不具有攻击性的指令内容，而 negative 则对应具有攻击性的指令内容。为提高模型的准确度，我们引入一个 neutral 来表示攻击性相对模糊的指令内容，这样从最大程度上避免用户设备受到攻击。

a) Word2Vec 算法

Word2Vec 是一种很好的词向量表征算法，它实质上是通过学习文本来用词向量的方式表征词的语义信息，通过一个嵌入空间使得语义上相似的单词在该空间内距离很近。这里我们使用 google 开源工具 Word2Vec 来用高维向量表示词语，并把相近的词语放在相近的位置。给定一个指令文本 s，通过 Word2Vec 可以得到一个表征语句含义的词向量集合：

$$\vec{s} = (\vec{x_1}, \vec{x_2}, \cdots, \vec{x_n}) \tag{3.8}$$

这里的 $\vec{x_i}$ 表示每个词语所对应的高维向量。

b) LSTM 网络模型

LSTM ——是一种 RNN 特殊的类型，可以学习长期依赖信息。LSTM 由 Hochreiter 和 Schmidhuber (1997) 提出，并在近期被 Alex Graves 进行了改良和推广，在语言模型领域依据上下文进行语义预测方面有非常优秀的应用效果。LSTM 网络结构如图 3-7 所示：

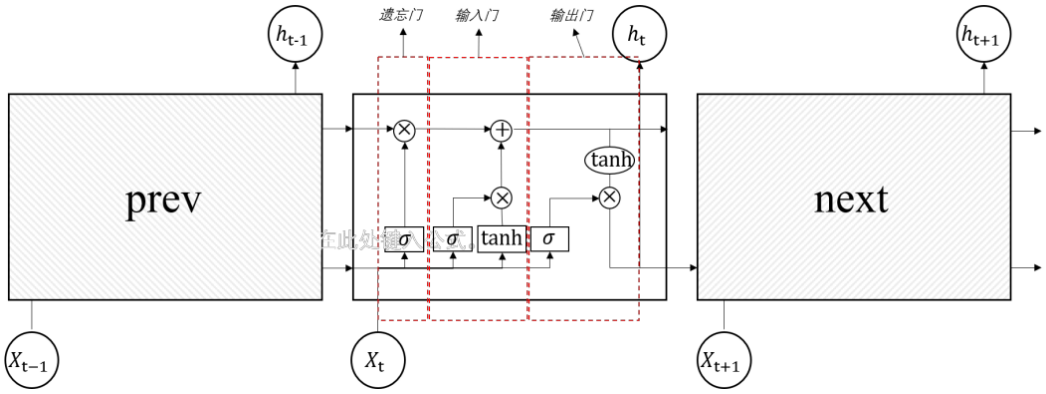


图 3-7: LSTM 网络结构

通过三个门来了解 LSTM 是如何将具有攻击性语义的文本和正常文本分类区分的。首先是“遗忘门”，在语言模型中，当前细胞状态（上图中绿色块部分）可能包含了当前主语的性别，如“男性”，那么针对这个主语的代词便可以被选择出来。然而当我们看到新的主语时，我们希望他能够忘记旧的主语。因此第一个门的输出可以表示为：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3.9}$$

其中 h_{t-1} 表示上一个细胞的输出，x_t 表示当前细胞的输入，σ 表示 sigmoid 函数。其次是“输入门”，当丢弃掉旧的主语后，需要加入新的主语信息，这里分两个步骤决定新加入细胞的信息：①经过一个 sigmoid 层决定哪些信息需要更新；②一个 tanh 层生成一个向量，也就是备选的用来更新的内容。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3.10}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{3.11}$$

接着对丢弃和新加的两部分进行叠加，得到当前状态：

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{3.12}$$

最终我们需要明确输出什么值，即“输出门”。语言模型中，当前的代词被做出正确分析后，紧接在其后的动词可能会受到这个代词的影响，因此我们需要将当前的状态输出给下一个细胞分析。

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3.13}$$

$$h_t = O_t \times \tanh C_t \tag{3.14}$$

c) 基于 Word2Vec 和 LSTM 网络的语音文本情感分类检测模型

考虑到音频指令文本攻击性分类与上述文本情感分类在解决思路上有异曲同工之妙，因此受到启发构造出基于 Word2Vec 和 LSTM 网络的语音文本情感分类检测模型，模型的网络结构图如图 3-8 所示：

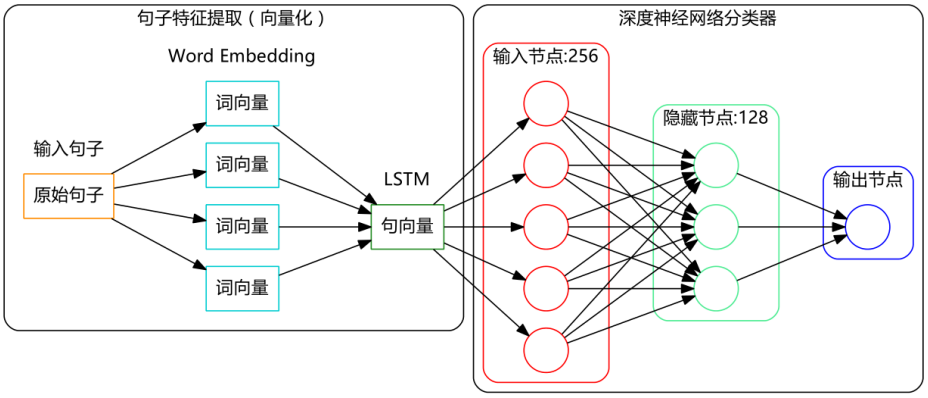


图 3-8: 语音文本情感分类检测模型

模型基于 Keras 框架(是一个开源的高层神经网络库)搭建，准备了三个类别基本语料库，包括 pos\neg\neutral，内含 1000 条已标记语料信息，用于模型训练与测试评估。通过对语料信息预处理、创建词向量及索引、定义 LSTM 网络结构等步骤，最终训练得到模型参数。

(2) 文本内容相似度分析

目前计算短文本的相似度更多使用的是编辑距离 (Levenshtein 距离)，但是编辑距离更适合计算纯文本的差异，不考虑文本的顺序和含义，所以在相似文本较多，或者我们希望得到的相似文本更符合人的理解时，编辑距离给出的答案就不是那么理想了。

The Jaro–Winkler distance (Winkler, 1990) 是 Jaro distance 算法的变种，最后得分越高说明相似度越大，是适合于串比如名字这样较短的字符之间计算相似度。0 分表示没有任何相似度，1 分则代表完全匹配。

由于不同语音识别框架如 DeepSpeech，Siri 等对语音的处理方式及原理不同，因此，当采集到音频样本时，同时递交两种语音识别框架进行识别，并对结果进行相似性评估，便可以检测是否存在语义层面的混淆攻击。

算法定义：

a) 局部相似度

对于两个字符串 s1 和 s2，它们的 Jaro 相似度由下面公式给出：

$$d_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m}\right) & \text{otherwise} \end{cases} \tag{3.15}$$

其中，$|s1|$ 和 $|s2|$ 表示字符串 $s1$ 和 $s2$ 的长度，d_j 是最后得分，m 是匹配的字符数，t 是换位的数目 transpositions 的一半。

Jaro 算法的字符之间的比较是限定在一个范围内的，如果在这个范围内两个字符相等，那么表示匹配成功，如果超出了这个范围，表示匹配失败。而这个范围就是匹配窗口 (Match Window)，在 Jaro 算法中，它被定义为不超过下面表达式的值：

$$MW = \frac{MAX(|s1|, |s2|)}{2} - 1 \tag{3.16}$$

匹配窗口是一个阈值，在这个阈值之内两个字符相等，可以认为是匹配的；超过了这个阈值，即使存在另一个字符与该字符相等，但由于它们的距离太远了，二者的相关性太低了，不能认为它们是匹配的。从上面的公式可以看出，该算法强调的是局部相似度。

b) Jaro–Winkler distance

Jaro-Winkler 算法给予了起始部分就相同的字符串更高的分数，它定义了一个前缀范围 P，对于要匹配的两个字符串，如果前缀部分有长度为 L 的部分字符串相同，则 Jaro-Winkler Distance 为:

$$dw = d_j + L \times P(1 - d_j)dw \tag{3.17}$$

其中，d_j 是 Jaro distance 最后得分，L 是前缀部分匹配的长度，最大不超过 4 个，P 是一个范围因子常量，用来调整前缀匹配的权值，P 越大，表示共同前缀权重越大，但是 P 的值不能超过 0.25，因为这样最后得分可能超过 1 分。Winkler 的标准默认设置值 $P = 0.1$。

为了排除噪声引起的语音框架识别的误差，通过设定相似度阈值，如 80% 来界定语音样本的语义层面是否被修改，若低于阈值，则认为受到攻击。

3.3.2 基于群智感知的激励算法

本系统基于群智感知的激励机制，构建了一个语音安全保障系统——由感知任务平台发布感知任务（语音攻击信息收集任务），用户作为任务执行者（基本感知单元）执行任务，大数据分析平台对收集的感知数据进行清洗、分析、归纳，再由感知任务平台评估任务完成情况，可以根据感知质量水平得到工作者的客观排名。最后再将感知数据写入语音攻击记录数据库中，得到某一区域的攻击热力分布，并对用户进行预警和对公安提供办案线索。如图 3-9 为语音安全保障系统的群智感知框架结构。

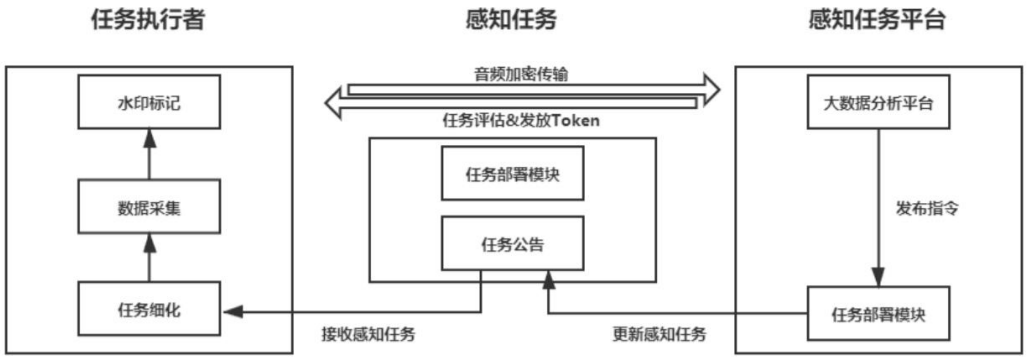


图 3-9: 智能语音防御系统的群智感知框架结构图

3.3.2.1 感知任务激励算法

感知任务由感知任务平台发布，平台给出任务名称、任务功能、任务要求和质量公告等信息。在质量公告里，具体说明了报酬标准以及质量评定指标。

考虑到任务执行者可能存在位置不确定性、所处环境噪音不可控性等因素，我们对参与者的分配上将会参考多种因素，例如数据质量、参与时间、地理位置等多方面因素。同时这样参与者也能通过感知反馈提高自身数据的质量以及竞争力，获取更多信任和激励积分。其效用函数可以用如下公式表示:

$$S(x) = \frac{\sum_{i=1}^{n} w_i S(x_i)}{\sum_{i=1}^{n} S(x_i)} \tag{3.18}$$

$$\text{Token} = \sum_{i=1}^{n} w_i S(x_i) \tag{3.19}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{3.20}$$

其中，w_i 是各属性权值，$S(x_i)$ 是各属性的效用函数，$S(x)$ 是对各属性的有效函数求和之后归一化的结果，用来表示任务完成的质量。

该群智感知系统还存在一个隐含的激励机制，即用户在完成感知任务的同时，在一定程度上也保证了自己的设备免受语音攻击的威胁，会实时受到数据中心的监测，保障设备安全；在进入攻击热力区域时还会受到实时的预警提醒；每周系统还会生成个性化周报，分析你的受攻击情况，给出一个详细的报表，让用户自身可以明确威胁防御攻击。这些都可以作为用户持续接收感知任务的回报激励。

3.3.2.2 多目标参与者选择策略

结合上述激励机制的分析结果，我们提出了多目标参与者选择算法。在该算法中及考虑到用户移动轨迹、感知能力对收集感知信息质量的影响，也考虑到单个移动节点的激励要求对感知参与者数量的征召与维护的影响。

首先初始化一组感知任务 Γ，其中 $\forall q \in \Gamma$。每个感知任务的激励成本 C_q，移动节点完成感知任务 q 的激励要求为 C_α^q，备选的移动节点集合为 Ω，移动节点 α 完成感知任务 q 的感知能力为 u_α^q，移动节点 α 的转移概率矩阵为 $Q = q_{ij}$；输出为要选择的移动节点集合 S。

3.3.3 基于态势感知的安全预警算法

态势感知是在大规模数据网络环境中，对能够引起态势发生变化的所有安全要素进行获取、理解、显示以及预测未来的发展趋势，并不拘泥于单一的安全要素。态势感知技术首先对各种影响系统安全性的要素进行检测获取，然后对安全信息采用分类、归并、建立数据模型、分析等手段进行融合，接着对融合的信息进行综合分析，得到数据网络的整体安全状况及其应对措施，并对网络安全状况的发展趋势进行预测，最后为智慧社区安全管理提供可靠的数据参考和决策支持。态势感知模块主要分为四个部分：态势察觉、态势理解、态势评估、态势预测四个子模块。安全感知平台收集攻击数据就是态势察觉；数据平台通过数据分析算法对态势进行理解和评估；最后平台基于理解和评估的结果进行态势预测。

3.3.3.1 基于 NIN 网络的态势察觉、理解和评估算法

NIN 网络全称为 Network in Network[11]，如图 3-10 所示整体网络结构由 Input、MLPConv、GAP（全局均值池化）和 softmax 组成。NIN 网络改进了传统的 CNN 网络结构，使算法参数减小的原来的十分之一，大大提高计算效率。

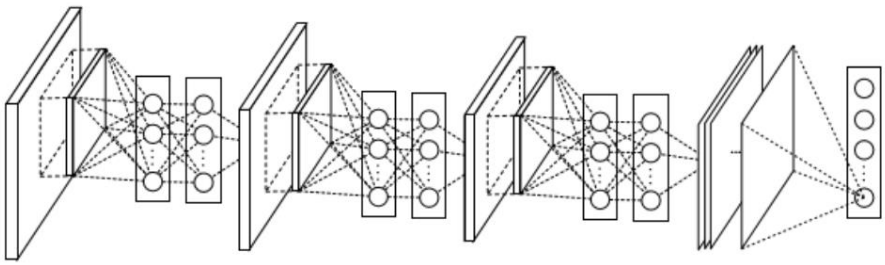


图 3-10: Network in Network 网络结构图

MLPConv 层可以看成是每个卷积的局部感受野中还包含了一个微型的多层网络。

对于 MLPConv 层每张特征图的计算公式如下：

$$f_{i,j,k_i}^1 = \max\left(w_{k1}^1 x_{i,j} + b_{k1}, 0\right) \tag{3.21}$$

$$f_{i,j,k_n}^n = \max\left(w_{k1}^n T_{i,j}^{n-1} + b_{kn}, 0\right) \tag{3.22}$$

相对于传统的卷积层，MLPConv 对每个局部感受野的神经元进行了更加复杂的运算。提高了模型的非线性，同时也没有引入太复杂的计算。

利用 NIN 网络结构进行特征提取进行态势理解并使用 softmax 进行分类，将走事件态势分为：安全、警告、危险三个级别。

3.3.3.2 基于类激活映射的热力图生成算法

我们使用类激活映射算法生成走失热力图，类激活图是指对输入图像生成类激活的热力图，表示每个位置对该类别的重要程度，有助于了解一张图片的哪个部分使得卷积神经网络做出最终的决策，可以定位图像中特点的目标。把类激活图利用在事件上，我们可以知道是哪些区域的走失信息是分类器做出判断的重要依据。因此对于不安全的评估结果，这些区域的安全风险非常高，可以视作走失高风险区域。

类激活映射就是把图片中的重要区域用输出层权重映射回卷积层特征的方式标记出来。具体来说，对于一个给定的图，用 $f_k(x,y)$ 代表最后一个卷积层在空间坐标 (x,y) 中单元 k 的激活值。然后，对于每个单元 k，通过 GAP 后的结果 F_k 为 $\sum_{x,y} f_x(x,y)$。则，对于每个类 c，输入 softmax 的 S_c 为 $\sum_k w_k^c F_k$，w_k^c 代表单元 k 对应的类 c 的权重。把 $F_k = \sum_{x,y} f_k(x,y)$ 带入 S_c，得：

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y) \tag{3.23}$$

我们用 M_c 定义类别 c 的 CAM，则空间每个元素为：

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \tag{3.24}$$

则 $S_c = \sum_{x,y} M_c(x,y)$, 所 $M_c(x,y)$ 直接表明了把空间网络 (x,y) 激活对图片划分为类别 c 的重要性。最后生成的 CAM 图就是：

$$CAM = \sum_{x,y} M_c(x,y) \tag{3.25}$$

这个 CAM 图就可以作为该时段的走失热力图。这种热力图生成算法不同于传统样方法和核密度法，使用类激活映射方法生成热力图是从整体感知的并且是基于安全态势的理解和侧重于安全态势的评估热力图生成算法，因此这种热力图生成算法更适合态势感知分析。

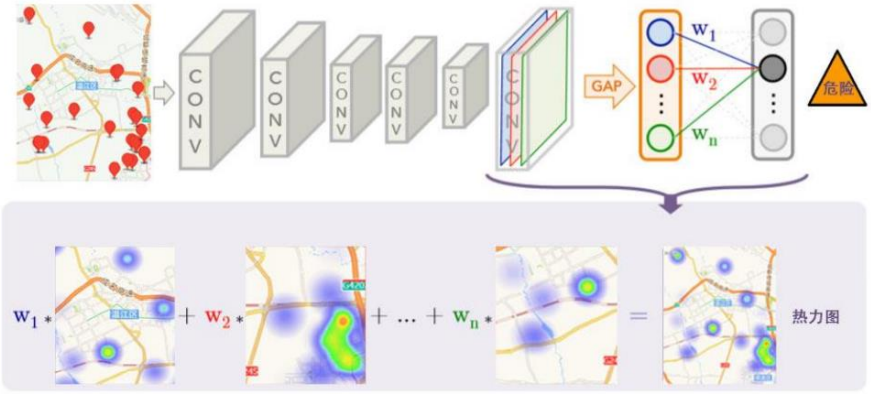


图 3-11: CAM 图

3.3.3.3 基于 CNN-GRU 的热力图态势预测算法

我们从倪铮等人的工作中得到启发，不同于倪铮等人使用 LSTM 做时序建模，我们使用 GRU 对热力图进行时序建模。LSTM 在大数据集情况下表达性能较好，在普通数据集下和 GRU 性能差不多，但由于 GRU 参数比 LSTM 少，更容易收敛的特性，我们选择了使用 GRU 进行时序建模。

GRU 有两个控制门，即更新门和重置门。更新门用于控制前一时刻的状态信息被代入到当前状态的程，更新门的值越大说明前一时刻的状态信息带入越多。重置门用于控制忽略前一时刻的状态信息的程度，重置门的值越小说明忽略得越多。

主体结构为：

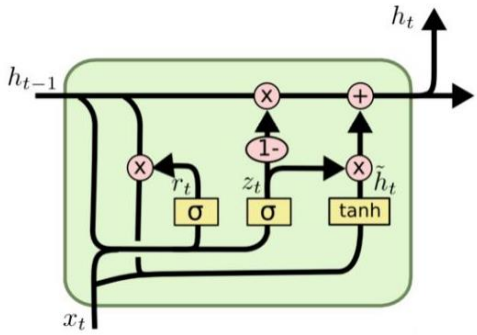


图 3-12: GRU 主体结构图

卷积神经网络 (CNN) 是一种多层神经网络，擅长处理图像特别是大图像的相关机器学习问题。它通过一系列方法，成功将数据量庞大的图像识别问题不断降维，最终使其能够被训练。一个典型的卷积神经网络由卷积层、池化层、全链接层共同组成，这其中最重要的就是卷积层。在我们的模型中，丢弃了全链接层，只保留了卷积层和池化层。模型通过卷积提取每一个时间的热力图的高维特征，并把他们作为时序数据传入 GRU 进行时序建模。模型结构如下：

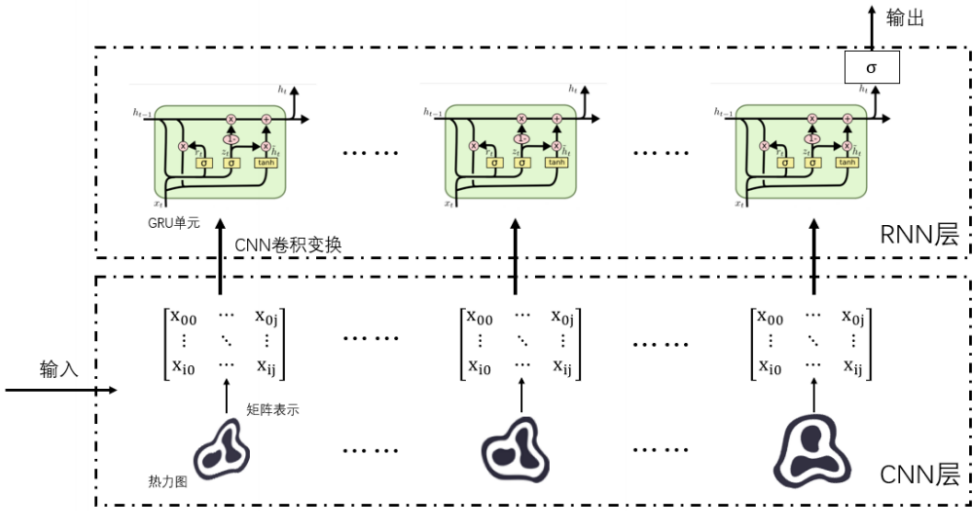


图 3-13: 模型结构图

GRU 层最后一个单元输出下一时刻的预测特征，通过输出特征我们可以预测热力图的扩张趋势，进行态势预测。

3.4 系统功能模块设计

本系统功能模块可分为五个子模块，分别为“身份认证“、“攻击识别”、“实时监测”、“预警提醒”和“大数据分析”五大模块。

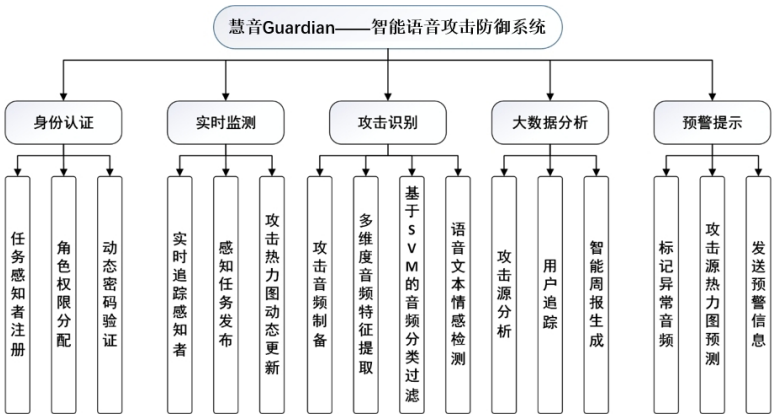


图 3-14: 智能语音攻击防御系统功能结构图

3.4.1 身份认证模块

本模块包括感知任务者的声纹注册、动态密码验证、权限管理三个功能点。当用户想要注册成为系统使用者时需要提供一段自己的声音音频，经系统处理得到声音特征信息，以便登录时进行声纹认证。声纹认证失败时，可通过动态密码验证方式找回密码。给用户、管理员、公安办案人员赋予不同权限。

下图为用户在 App 端登录的流程图：

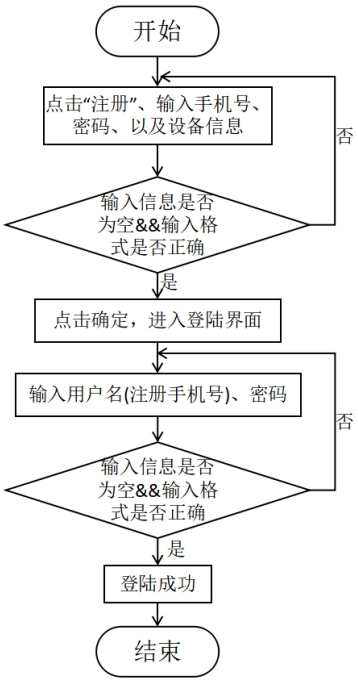


图 3-15: App 端登录流程图

3.4.2 实时监测模块

本模块主要使用了群智感知的思想，保障了整个系统的实时性、安全性和预警及时性。而本模块主要包括了实时追踪感知者、感知任务发布和热力图动态更新三个功能点。我们的系统是一个全天候实时追踪用户设备

安全的系统，对于已登记设备我们会实时追踪位置和状态信息，当有异常信息发生时系统会立即做出响应。同时系统会定期发布感知任务，更大范围内获取攻击源的信息，并动态更新系统备份的攻击信息，做到攻击最新化、识别最准化、预警最及时。该模块的整体流程图如下所示：

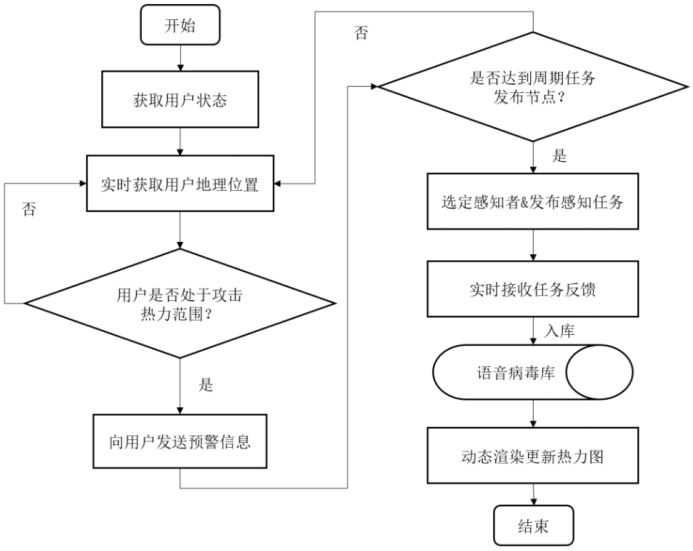


图 3-16: 群智感知任务模块整体流程图

这里主要介绍一下感知任务发布和热力图动态更新这两个子模块。

3.4.2.1 感知任务发布子模块

系统将用户位置与历史攻击热力图比较，通过距离计算，当用户进入相应攻击范围时，及时预警提醒用户离开所在区域或做出相应防范措施。当用户因为某些原因需要留在高风险区域时，管理员可以通过平台在一定时间节点内发布周期性感知任务，选定感知者利用 APP 客户端检测所在区域是否存在持续攻击，这样可以保证攻击热力图的最新化，实现了安全保障的实时性。感知任务发布基本字段如下所示：

表 3-1: 感知任务发布基本信息字段表

字段信息	举例
管理员 ID	155xxxxxxxx
任务进行时间段	13:00-14:00
任务持续日期	未来 7 天
任务地点	
任务执行手机型号	小米
任务针对语音助手	小爱同学

用户如果进入感知任务指定的区域，除了会受到预警提示外，也可以接收感知任务，搜集恶意语音攻击情况。因为群智感知可能会搜集部分用户的隐私信息，我们采用了前面提到的门限环签名技术，用以保护用户隐私。在用户搜集的音频信息上传时，我们使用了音频水印加密与检测算法对其进行保护。当系统使用提到的群智感知激励机制后，判定用户上传为有效信息，用户则会收到对应激励积分，加强用户置信度。流程图如下所示。

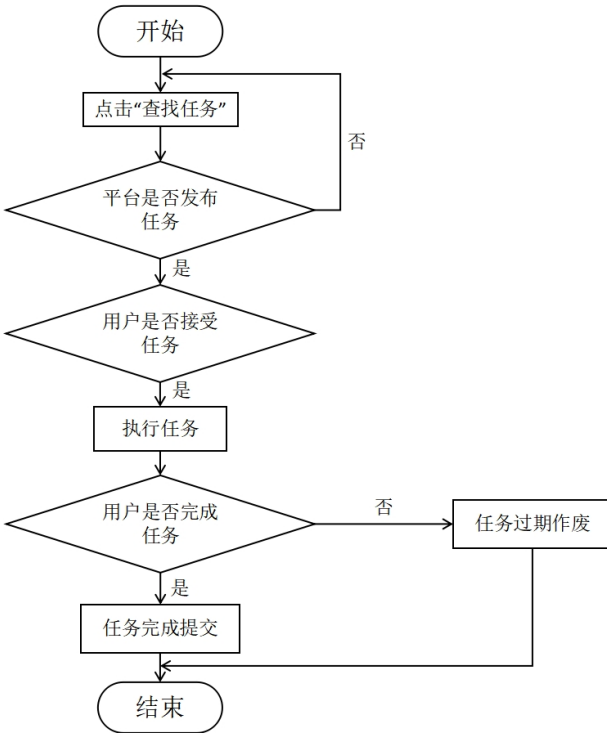


图 3-17: 感知任务反馈子模块流程图

3.4.2.2 攻击热力图动态更新

由感知任务所反馈的信息会上传至系统数据库，并通过热点图、热力图的形式进行可视化展示，便于管理人员能够更直观的查看区域的智能语音攻击情况，实现流程图如下所示。

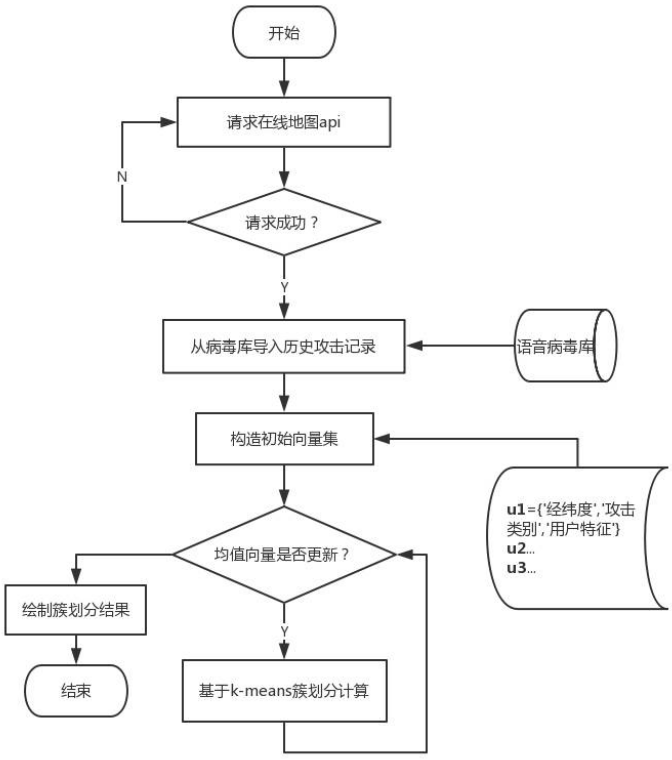


图 3-18: 攻击热力图动态更新流程图

3.4.3 攻击识别模块

模拟攻击包括攻击音频制备、多维度音频特征提取、基于 SVM 的音频分类过滤和语音文本情感检测四个功能点构成。首先我们需要制备几种语音攻击用于模型训练，然后提取分析音频的特征，包括频域分析、时域分析、能量、过零点等。通过构建基于 SVM 的音频分类器将截获的音频进行分类过滤，然后再从语义的角度对语音文本的情感进行分析检测，从而进一步过掉具有攻击语义的语音。最后向用户发出提示，提醒用户正受到攻击。这里主要介绍一下攻击音频制备模块，其他子模块在 3.3 均有介绍。

3.4.3.1 模糊指令攻击制备

模糊指令攻击根据论文《Hidden Voice Commands》中提出的黑盒攻击方式进行实现。攻击者的目标是产生受混淆的命令，受害人的语音识别系统会接受该命令，但听众无法理解。

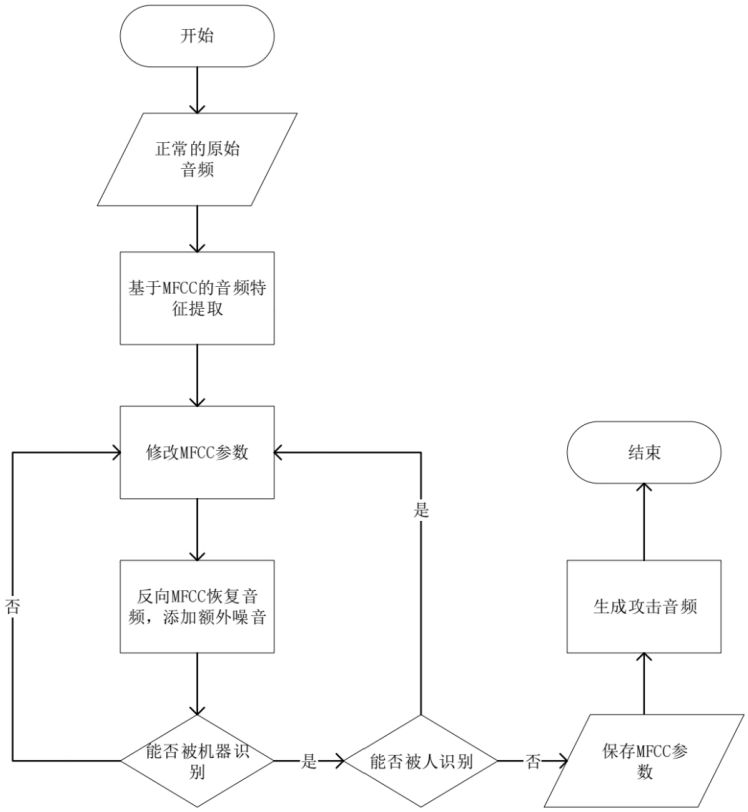


图 3-19: 基于模糊指令攻击制备流程图

其实现流程如上图所示，首先准备一段包含正常指令的原始音频。初始化一组参数，进行特征提取。然后修改参数（主要包括 wintime：信号被视为恒定的时间，hoptime：相邻窗口之间的时间步长，numcep：倒频谱系数的数量，nbands：聚合能级的扭曲光谱带数），反向 MFCC 将音频特征转化为音频样本，此时的音频会由于参数的修改增加一定的噪声。将音频首先用于语音识别系统进行识别，再进行人识别。当语音识别系统正常识别音频内容，但人无法立即音频内容，完成参数确定，可进行攻击音频生成。其他情况均需要重新设置参数，再次测试。

实质上是在尝试删除语音识别系统中未使用但人类听众可能会用来理解的所有音频特征。以达到人无法察觉音频指令，而机器收到音频指令的效果，进行相应的攻击。

3.4.3.2 机器学习攻击制备

机器学习攻击根据《Audio Adversarial Example: Targeted Attacks on Speech-to-Text》中的方法实现。通过向原始音频添加一个几乎不可见的扰动，使新音频被识别为任何期望的内容。

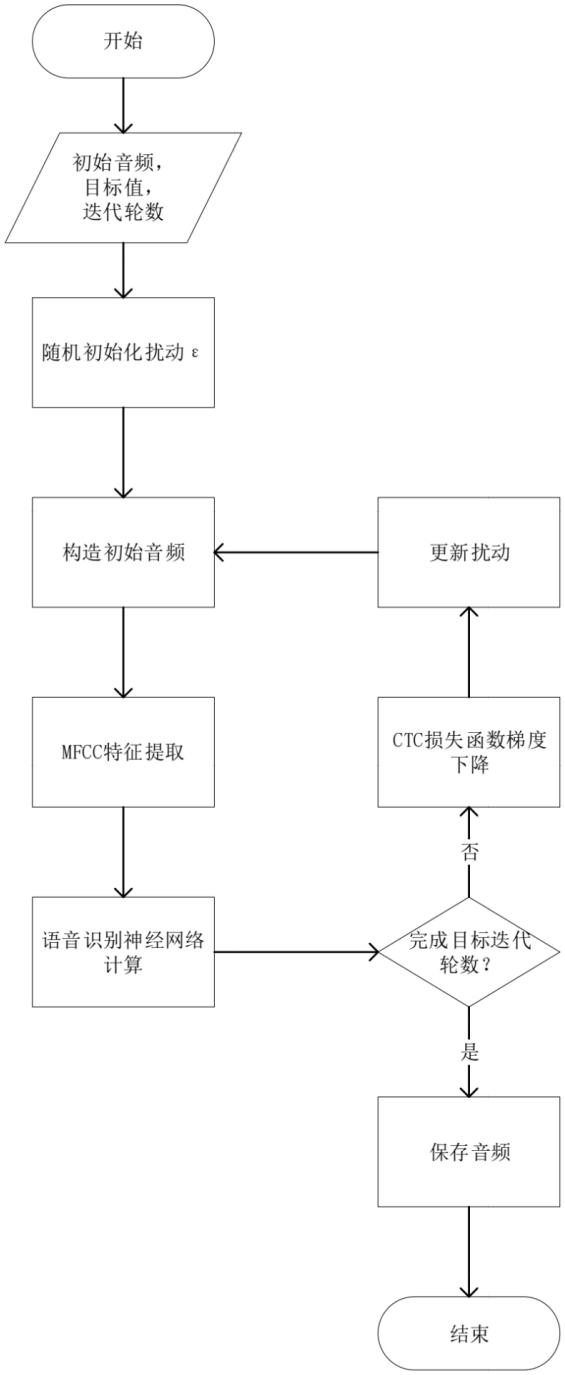


图 3-20: 基于机器学习级别攻击制备流程图

处理方式如上图所示，随机初始化扰动，将扰动添加到原始音频上。利用音频特征提取算法计算出 MFCC 特征值，并作为语音识别神经网络模型的输入。得到的结果进行梯度下降，直到完成目标迭代轮数，最后将训练出的攻击音频保存。

3.4.3.3 海豚音攻击制备

海豚音攻击通过 AM 调制，将低频原始音频搭载到高频载波上，形成具有攻击内容的高频音频指令。

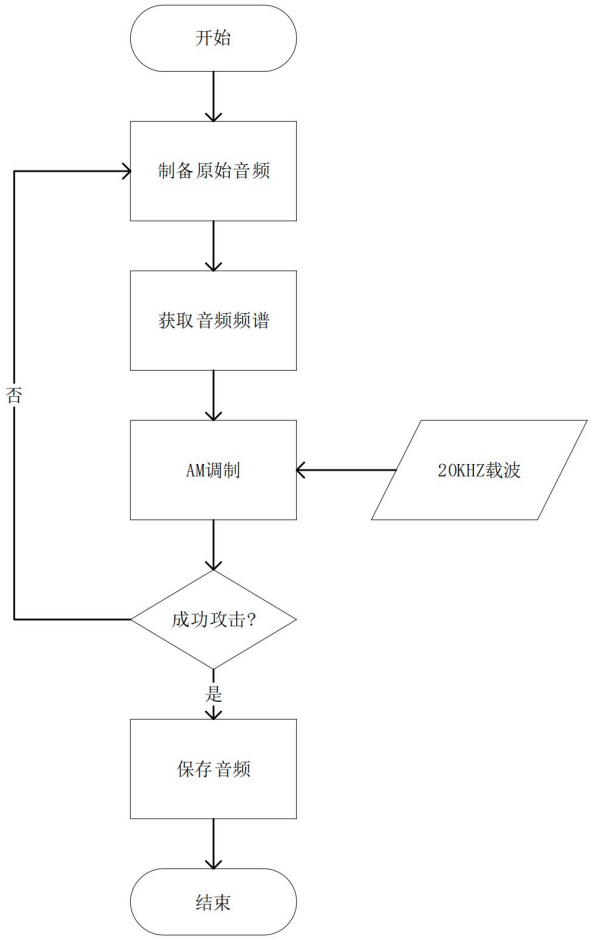


图 3-21: 海豚音攻击制备流程图

流程如上图所示，超声波攻击首先需要制备出一个基带信号——即原始正常音频信号，然后通过调制基带信号使它在经过麦克风放大器时能够有效地被解调为可以被语音助手识别的音频信号。AM 调制：在 AM 中，载波的幅值与基带信号成比例变化，调幅产生的信号功率集中在载频和两个相邻的边带。调制的逆过程叫解调，调制是一个频谱搬移过程，它是将低频信号的频谱搬到载频位置。从已调信号的频谱中，将位于载频的信号频谱搬移回来。调制和解调都完成频谱搬移，各种调幅都是利用乘法器实现的。

3.4.4 大数据分析模块

本模块主要包括攻击源分析、用户追踪和智能周报生成三个子功能模块构成。收到音频信息后，系统对攻击源进行分析，确定是何种攻击方式，进行用户追踪，得到用户的详细地址，然后经过大数据平台的分析智能生成周报。

3.4.4.1 攻击源分析与用户追踪子模块

系统再得到用户权限许可后，根据 JW 算法对之前过滤的恶意语音攻击音频信息进行再分类，将其细分为三种常见攻击语音中的具体一类，并将用户的位置信息、时间信息、攻击类型信息上传至个人攻击记录数据库

以及恶意语音攻击数据库，同时将音频信息使用水印加密的方式上传至恶意语音攻击资源数据库，并将分类完成后的攻击类型信息反馈给用户。攻击最终地段表如下所示

表 3-2: 基于用户的攻击追踪字段表

字段信息	举例
用户 ID	155xxxxxxxx
监测时间段	
语音攻击总次数	21
语音攻击分类检测	机器学习 12 次、超声波 5 次
攻击位置分布	xx 教学楼，xx 科研楼
历史攻击路线	xx 地->xx 地->…

3.4.4.2 智能周报子模块

系统通过用户上传的攻击记录，生成用户周报。包括一定时间范围内收到的恶意语音攻击的次数、类型，并以可视化的方式展现给用户。

3.4.5 预警提示模块

本模块主要包含异常音频标记、攻击源热力图预测和发送预警信息三个功能点。本模块接收分类过滤模块的输出结果作为输入，并判断是否存在异常攻击。若存在异常攻击，则将攻击地理位置、攻击信息动态添加到数据库中，并实时将预警信息发送给周围用户，告知用户尽快离开攻击区域或者停止支付等危险操作。对于不在攻击音频库中出现过的音频，我们将其做一个特殊标记，由人工对其进行分析和识别。

实时预警流程图如下所示。

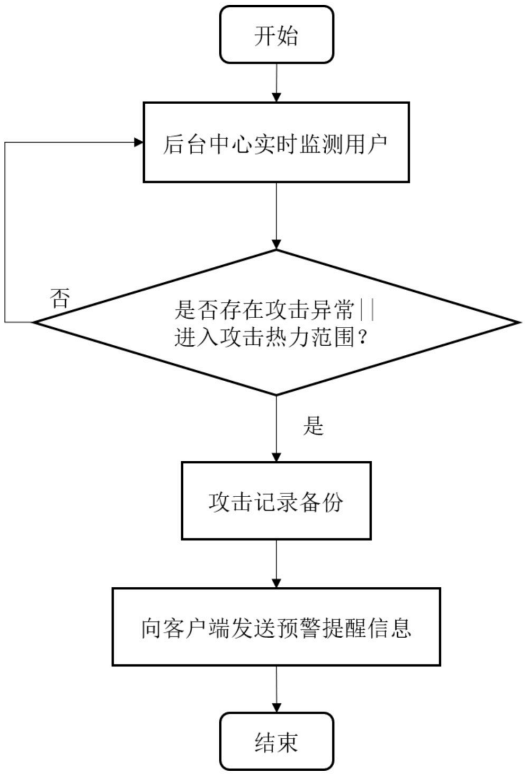


图 3-22: 实时预警流程图

其中攻击源热力图预测为本模块的核心部分，其核心思想为语音攻击态势感知，主要分为四个部分：态势察觉、态势理解、态势评估、态势预测四个子模块。安全感知平台收集攻击数据就是态势察觉；数据平台通过数据分析算对态势进行理解和评估；最后平台基于理解和评估的结果进行态势预测，生成预警报告。流程图如下所示

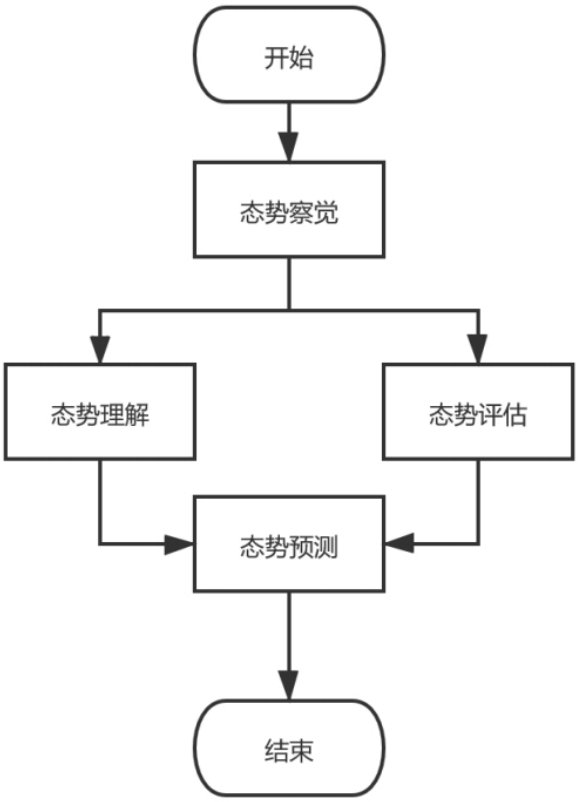


图 3-23: 攻击态势感知模块流程图

3.4.5.1 攻击态势察觉子模块

用户设备上传攻击数据到安全感知平台。该过程已经由感知任务发布子模块进行实现。攻击热点图如下所示。

图 3-24: 攻击热点图

3.4.5.2 攻击态势理解和评估子模块

攻击态势理解和评估模块，我们利用了 NIN 网络进行态势理解并把 softmax 分类结果作为态势评估结果。首先，模块对于分区域块的攻击点图进行矩阵化表示转化，根据攻击点相对位置确定相应坐标，攻击时间、攻击次数、攻击频率划分三通道，生成三通道二维表。然后我们使用了 NIN 网络结构进行特征提取再通过 softmax 进行分类，分为：安全、警告、危险三个安全风险级别。我们使用类激活图进行攻击热力图的更新，类激活图是指对输入图像生成类激活的热力图，表示每个位置对该类别的重要程度，有助于了解一张图片的哪个部分使得卷积神经网络做出最终的决策，可以定位图像中特点的目标。把类激活图利用在攻击点图上，我们可以知道是哪些区域的攻击信息是分类器做出判断的重要依据。对于不安全的评估结果，这些区域的安全风险非常高，因此平台可以通过热力图实时预警危险范围。

3.4.5.3 态势预测子模块

通过热力图的时序变化来预测态势，基于 CNN-GRU 的网络结构提取时序的热力图特征，并通过 GRU 进行时序的建模和预测。平台会显示预测下一时段的热力图作为态势预测图，给予预警。攻击热力图如下所示。

图 3-25: 攻击热力图

3.5 安全体系设计

3.5.1 基于 HTTPS 的安全通信体系

HTTPS，是 HTTP 协议的改进版，末尾字母的 S 代表它新增了安全方面（security）的设计，HTTPS 通信的基础是在原有的 5 层通信模型中的传输层之上新增了 SSL 层，即安全套接字层，其中 SSL 层内置了两个协议，上层为 SSL 握手协议，用于数据传输前双方的身份验证、加密算法和加密密钥的协商；底层为 SSL 记录协议，用于实际的数据加密、封装、压缩等。

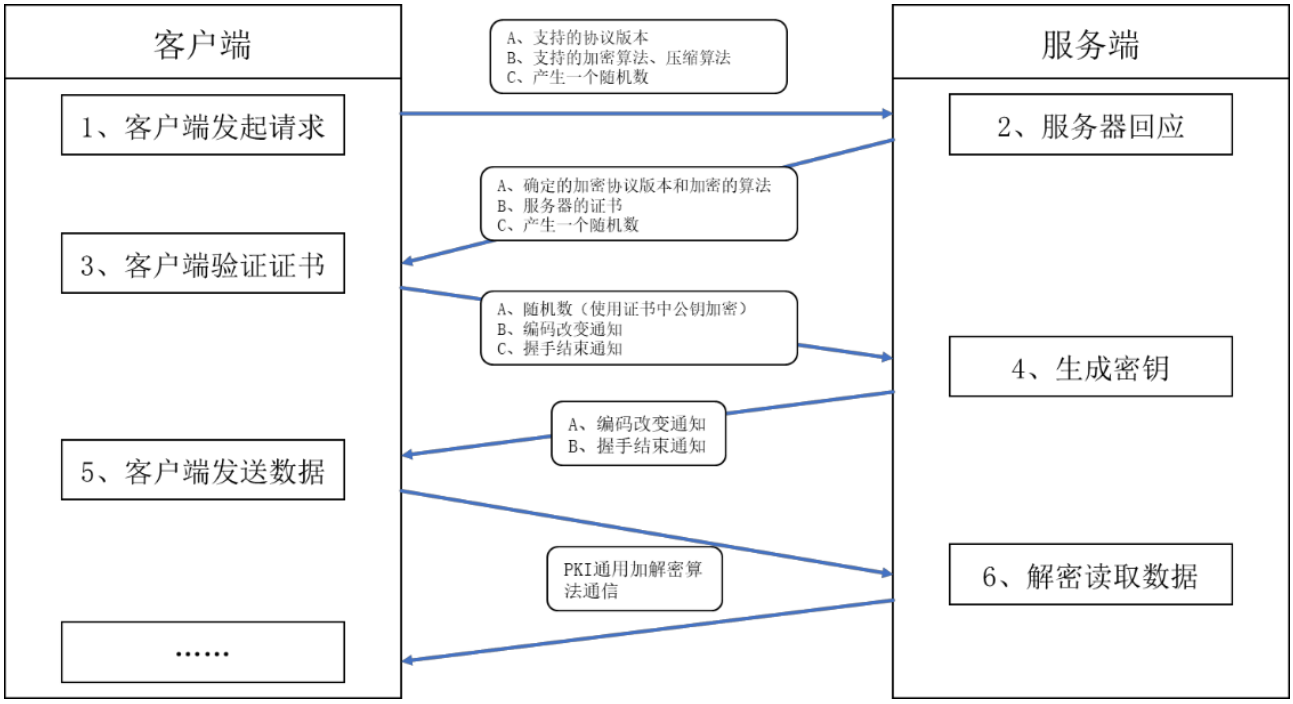


图 3-26: 基于 HTTPS 的安全通信握手协议流程图

上方图中客户端首先发出请求，提供自己支持的协议版本以及支持的加密方法，由于目前所使用的 SSL 通信协议方法复杂多样，因此，为了确保通信双方能够正确的加解密数据，就必须要协商加解密算法。

3.5.2 基于国密 SM2/SM3/SM4 加解密算法的 PKI 模型

(1) SM2 非对称加解密算法

SM2 是中国国家密码管理局 2010 年公布的椭圆曲线公钥密码算法，这是一种非对称体系算法，其计算复杂度为完全指数级，同时用于其处理速度快，机器性能消耗小，使得它比传统的非对称加密算法速度快百倍以上。

(2) SM3 单向散列算法

SM3 是中国国家密码管理局 2010 年公布的中国商用密码杂凑算法标准，该算法适用于商用密码应用中的数字签名和验证，是在 SHA-256 基础上改进实现的一种算法。SM3 算法采用 Merkle-Damgard 结构，消息分组长度为 512 位，摘要值长度为 256 位。该散列算法经过了暴力碰撞测试，至少现如今暂未提出能够进行逆推解密的方法。

(3) SM4 信息明文的对称加密算法

SM4 是中国国家密码管理局于 2012 年公布的一种基于轮函数的对称加解密算法，加解密过程中密钥相同，长度与分组明文一致，解密流程使用加密轮密钥的逆序。SM4 的优点在于加解密过程类似，只是使用顺序相反，因此其效率高，资源占用低，并且有较高的安全防护性。

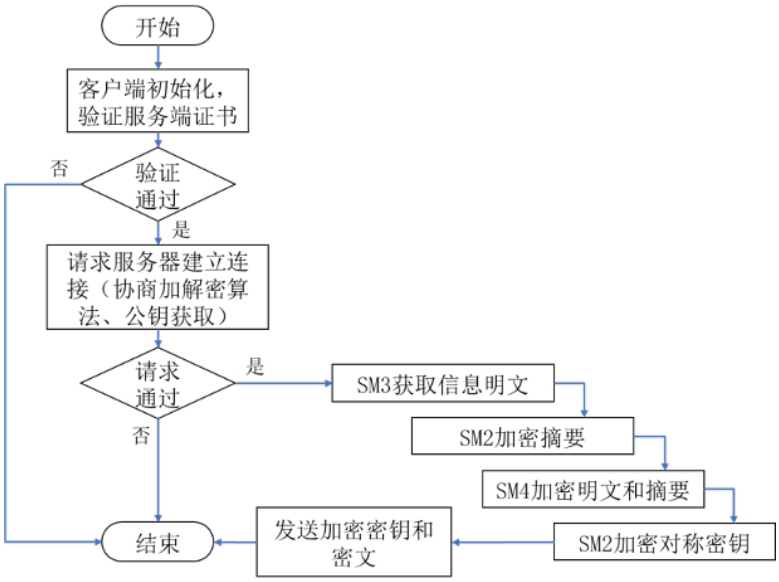


图 3-27: 基于 SM2:SM3:SM4 的 PKI 加密流程图

将上述三种国密算法整合至 PKI 通信过程当中，即用 SM2 作为非对称加解密算法，用于加解密对称密钥；用 SM3 作为明文和证书摘要算法；用 SM4 作为对称加解密算法，直接对信息明文进行加解密。如图 2-7 所示是基于 SM2/SM3/SM4 的数据传输流程，图中只展示了数据发送方的流程。

3.5.3 数据存储安全

慧音 Guardian 系统的存储数据库总共有三个，分别是攻击记录数据库、用户数据库以及本地用户音频池。其中，攻击记录数据库和用户数据库保存在云端，而本地用户音频池则保存由用户搜集到所有恶意语音攻击音频信息，进一步保护用户隐私。数据库展示如下图所示。

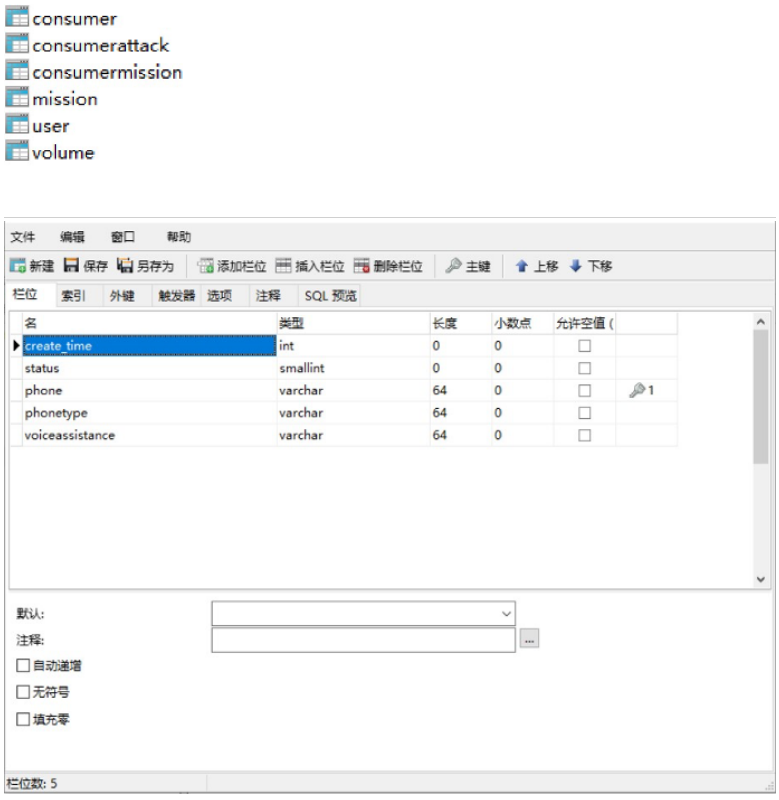


图 3-28: 数据库展示图

第 4 章 方案实现

内容提要

- 系统环境搭建
- 语音安全 App
- 感知管理平台
- 音频分类算法
- 语音文本情感分析算法
- 攻击源热力图算法
- 群智感知激励机制算法
- 群智感知多目标参与者选取算法
- 音频信号发生装置
- 安全体系

本章详细阐述作品的实现，主要包括系统环境的搭建、核心算法实现、实时数据监测分析平台实现、APP 客户端实现、安全体系实现五个部分。

4.1 系统环境搭建

在我们的作品中，服务器采用阿里云服务器，搭建在 python 运行平台，并利用 Python flask+jinjia 框架。系统部署在 Centos 操作系统上，通过 Mysql 数据库进行数据管理，所以需要在系统上进行相应安装和配置。

4.2 慧音 Guardian—语音安全 App 实现

客户端通过 Android Studio 平台进行开发，主要为用户提供用户注册认证、语音监听、预警返回、周围攻击源提醒、感知任务接收以及周报查看功能，不仅为用户提供了语音攻击的预警、保护，更通过感知任务的接受与完成丰富了管理平台的数据，是平台得以更好发展的基础。

4.2.1 用户注册登录

本模块包含了用户的注册、登陆操作。用户进入注册界面，需要填写用户名、密码、密码确认、手机型号、语音助手。用户名栏输入手机号对用户进行唯一标识。当检测到输入内容的格式、或者用户名已在服务器注册时，将会提醒用户重新输入，注册成功时，界面跳转至登陆界面，输入刚注册的用户名以及密码即可进入 APP 功能界面。注册登录界面与主页面如下所示:

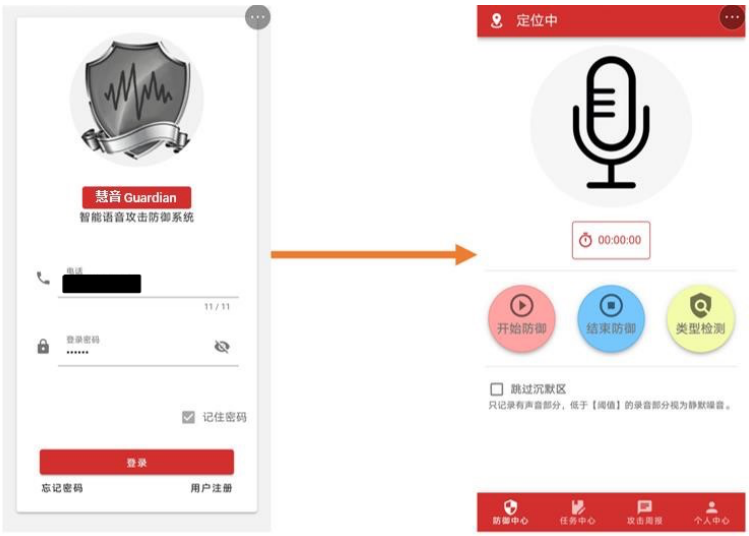


图 4-1: 注册界面与主界面展示图

4.2.2 语音攻击识别

用户登陆后进入第一个功能界面，即语音攻击的防御界面，客户端会提醒用户 App 会获得手机的麦克风以及存储的权限。点击“开始防御”按钮时，客户端即可获取到手机的麦克风权限，并实时监听用户周围的音频信息。

图 4-2: 语音攻击识别展示图

4.2.3 攻击类型分析

点击“检测”按钮，通过音频的上传处理，获得从服务器的返回值即可判断出在防御期间内，用户是否受到机器学习、海豚音等语音攻击。当服务器检测到攻击音频时，会记录受到攻击的音频的地理位置信息并在平台的攻击热力图上进行可视化标注。当用户在较短时间内，遭到多次语音攻击，则会生成一份该用户的语音攻击追踪图，并对该用户进行预警。恶意语音攻击识别与分析界面如下所示

图 4-3: 攻击类型分析展示图

4.2.4 语音攻击警示

用户第一次登陆成功进入功能界面时，客户端会提醒用户获取该用户的地理位置权限，该定位服务通过百度 SDK 实现。客户端每 10 秒获取一次用户的地理位置，并将其上传。服务器通过客户端上传的地理位置信息，即经纬度 (精确至小数点后十万位)，与服务器数据库中已存的语音攻击源位置进行聚类对比，将用户周围的攻击源个数返回给客户端，达到对用户的警示效果。

4.2.5 感知任务接收

完善管理平台数据覆盖率通过向客户端进行感知任务的发布与完成来实现。当平台向区域特定用户 (携带语音助手) 发布带有任务日期、任务期限以及任务地点的感知任务时，用户会查找到该感知任务，并在任务栏进行任务的接受或者拒绝。当用户在指定任务地点完成任务时，点击“完成任务”按钮，即可上传搜集到的音频信息。该功能不仅有利于用户检测自身生活环境的语音攻击隐患，同时也是平台得以发展的基础。感知任务接收界面如下所示

图 4-4: 感知任务接接收展示图

4.2.6 智能周报查看

客户端以一周作为一个周期为用户生成一份量身定做的周报，周报由两个部分组成，第一个部分是通过柱状图进行显示，在一周内每一天用户受到不同语音攻击的次数，通过对比可以直观看到用户自身生活环境安全与否; 第二个部分是对一周内分析出的不同种类的语音攻击类型进行统计显示。智能周报界面如下所示

图 4-5: 智能周报查看展示图

4.2.7 个人信息管理

用户可以通过个人中心界面查看、修改自己的个人信息、查看自己的积分以及其他操作，个人中心界面如下所示

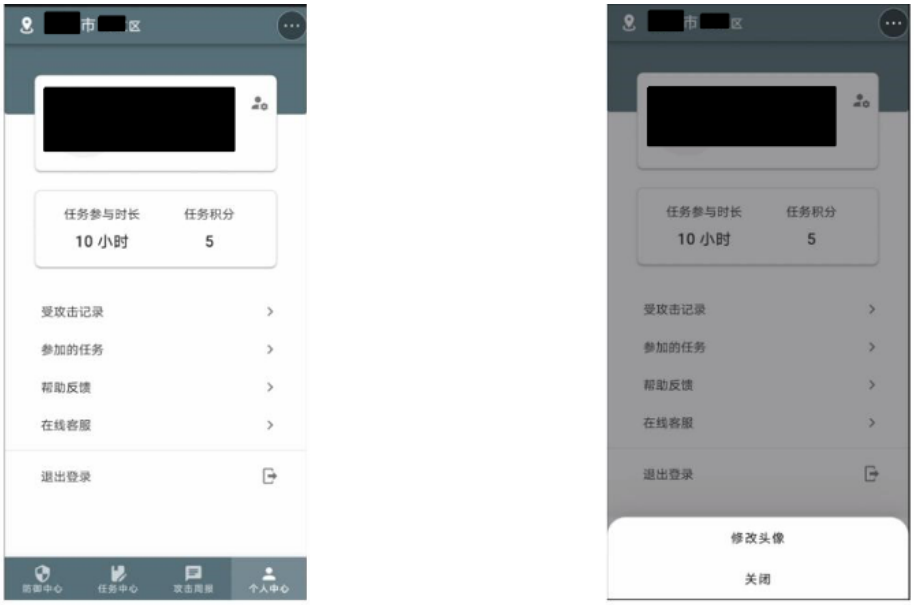


图 4-6: 个人中心界面展示图

4.3 慧音 Guardian—感知管理平台

后台管理系统旨在实现对用户被攻击事件的收集、分析汇总，并相应的生成周报，辅助作业人员进行决策。整个系统包含身份认证与权限管理系统，实时监测更新系统，感知任务发放与反馈，周报生成系统，态势预测热力图绘制系统，服务端实现采用 Python flask+jinjia 的组合框架实现。

4.3.1 管理员注册登陆

慧音 Guardian—感知管理平台的登录页面，管理员输入相应的账号密码后，点击登录，若信息正确则成功进入主界面。用户的密码在数据库中以加盐哈希的方式存储，有效地避免了针对用户数据库泄露的攻击，同时使用 session 与 cookie 设置并保存用户信息，使用设置短时会话时间的方式避免用户无意之间陷入的跨站脚本攻击。界面如下图所示

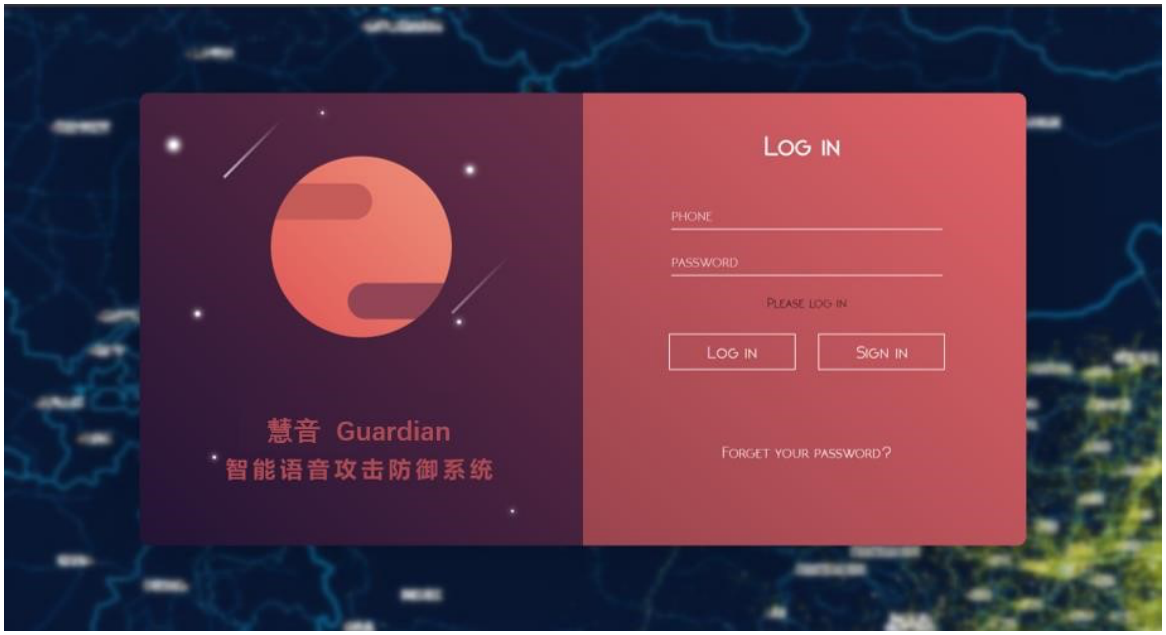


图 4-7: 管理员注册登录界面展示图

4.3.2 智能语音攻击分布

智能语音攻击分布图主要展示区域内恶意语音攻击事件发生的位置，更加直观地给管理员反映出恶意语音攻击事件的分布情况。页面中可以通过地图的拉伸查看具体的事件位置，所有事件位置都由蓝色标志标出。智能语音攻击分布如下所示

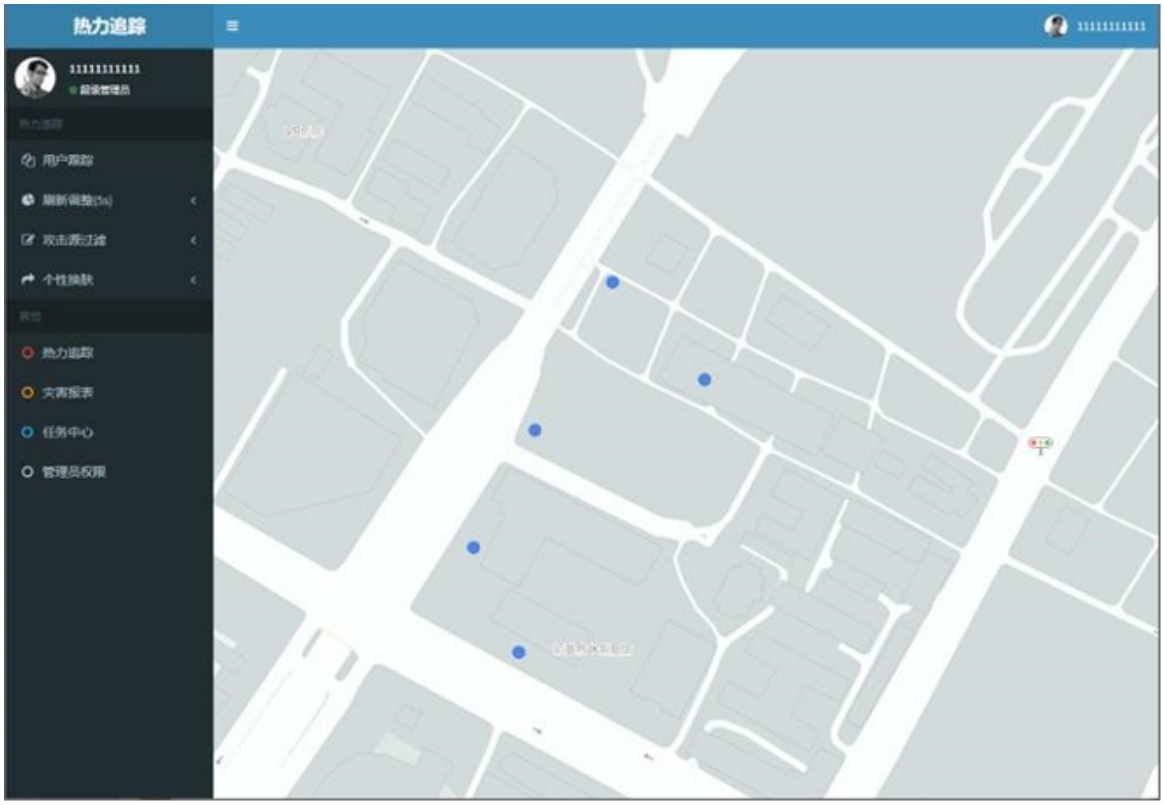


图 4-8: 语音攻击分布展示图

4.3.3 感知任务发布

如果有用户遭受到恶意语音攻击并上传，系统会自动发布基于该用户时间位置的感知任务。管理员也可以自行设定感知任务进行发布。实现界面如下所示。由感知任务搜集到的攻击位置类型信息通过会在智能语音攻击分布图进行标记。

图 4-9: 感知任务发布展示图

4.3.4 态势感知预测

通过语音攻击感知热力图可以查看成都市整个管辖范围的走失热力图，即每一个受到走失信息的地点周围都会形成一个环形带，多个地点标注从而形成走失老人感知热力图。通过该图我们可以清楚地观察到成都市哪个地区走失老人事事件比较严重，从而采取有针对性地预防措施，感知热力图如下所示。界面里面，我们也可以看到成都市的预测热力图，根据 GRU 时序预测算法和类激活映射 CAM 算法，也可以将计算的语音攻击热力预测图描绘出来，形成区域走失监测、预警的完整系统。

图 4-10: 态势感知预测展示图

4.3.5 智能攻击报表

系统会根据感知任务搜集的数据形成智能攻击报表，包括事件类型数目统计，历史跟踪人数、全国攻击情况、用户评论和历史记录等，智能攻击报表界面如下所示。

图 4-11: 智能攻击报表展示图

4.3.6 权限设置

超级管理员还可以对平台其他管理员的权限进行管理。权限管理界面如下所示：

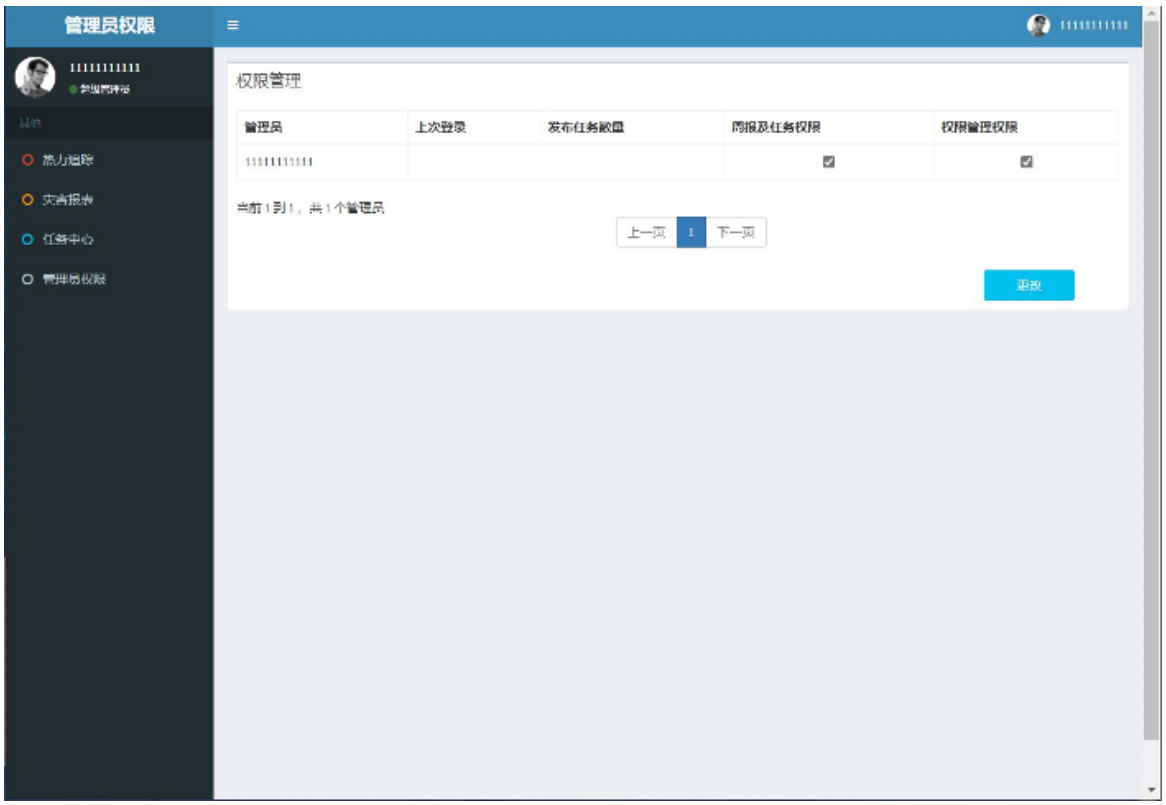


图 4-12: 权限管理界面展示图

4.4 算法实现

本节的算法原理在上文 3.3 节系统核心技术已经简述过，所以在本节中将算法实现的具体逻辑以伪代码的形式展示，原理细节不再赘述。

4.4.1 音频分类算法实现

分类过滤算法的原理如 3.3.1 所示，接下来对其具体实现过程进行简要概述。输入一段音频信号，将其特征向量加入 SVM 分类器，得到分类结果——包括三种主要的语音攻击类型。音频分类核心算法如下图所示：

```
Input：语音波形文件 AudioWave
Output：预测结果 PredictResult
Function AudioClassify(AudioWave)
1：    model <- Load_Model(SVM.h5)
2：    data_raw   <- input_transform(XXX.wav)
3：    data_processed <- MFCC(data_raw)
```

```
4:      if model(data) == 0 then
5:          result <- 合法指令
6:      else if model(data) == 1 then
7:          result <- 超声波攻击
8:      else
9:          result <- 机器学习攻击
10:     end if
11:     return result
12:end function
```

4.4.2 语音文本情感分析算法实现

本模块采用 python 语音搭建，并调用 LSTM 算法模型参数，语音文本合法性检测具体实现代码如下图所示：

```
Input：语音文本内容 AudioPhrase
Output：预测结果 PredictResult
1：function AudioTestPredict(phrase)
2：         result ← 0
3：         model ← load_model(lstm.h5)
4：         data ← input_transfer(phrase)
5：         if model(data) == 1 then
6：             result ← 合法指令
7：         else
8：             if model(data) == 1 then
9：                 result ← 模糊待识别
10：            else
11：                result ← 非法指令
12：            end if
13：        end if
14：        return result
15：end function
```

4.4.3 攻击源热力图算法实现

首先我们持续调用在线地图的 api 直到成功，然后从语音病毒库中导出历史攻击数据，并进行向量化以适应算法模型的输入。设置过滤条件如下。之后便按照 3.3.3 中基于 k-means 的算法、最终计算出攻击源的簇划分，最后将其动态绘制到地图中。过滤条件如下：

$$sift = \{'dolphin_{attack} : Boolean';' machine_{attack} : Boolean';' price_{search} : Boolean'.\} \tag{32}$$

攻击热力图的实现过程如图 4-13 所示：

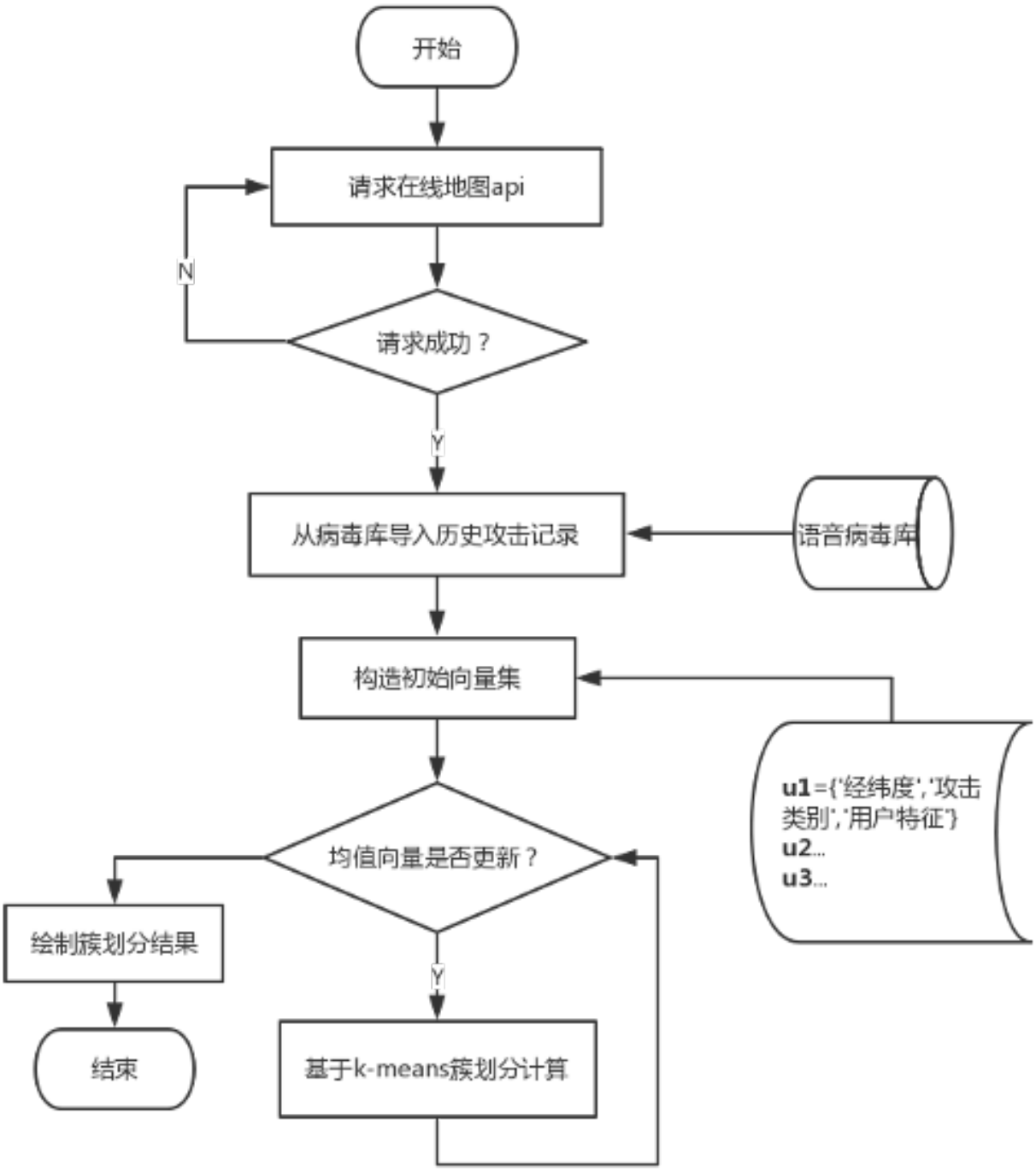


图 4-13: 攻击热力图实现流程图

4.4.4 群智感知激励机制算法

在 $In-script$ 中，$Sig_{SK_{platform}}(Task-Claim)$ 表示感知平台对所发布的任务的签名，根据感知任务的用户 u_i 的感知数据 $Data_{u_i}$ 判断任务完成的质量 $S(x)$，当符合 $Tocken$ 给予条件的时候给予任务完成者 $Tocken_{u_i}$。N 为参与感知任务的用户的数量，$Tocken$ 为用户的最佳报酬，验证加密签名过的用户数量 $Data_{u_i}$ 来验证用户身份。

Input: $Sig_{SK_{Platform}}$ (Task-Claim), $S(x)_{u_i}$, n, $Token_{u_i}$, $Token^*$, Sig_{u_i}(Hash($Data_{u_i}$)), $Data_{u_i}$;
Output: Verify $Data_{u_i}$;
1: if ($S(x)_{u_i} \geq 0.5 \&\& S(x)_{u_i} \leq 1$)
2: $Token_{u_i} = Token^* \times \lg n + S(x)_{u_i} \lg S(x)_{u_i} + (1 - S(x)_{u_i}) \lg(1 - S(x)_{u_i}) / (n - 1)$;
3: else
4: $Token_{u_i} = 0$;
5: end if
6: Verify $Data_{u_i}$
7: $Data_{sign} = Sig_{u_i}\left(\text{Hash}\left(Data_{u_i}\right)\right)$;
8: if ($Data_{sign} = Hash\left(Design\left(Data_{u_i}\right)\right)$
9: return true;
10: else
11: return false;
12: end if
Time-lock: Deadline.

4.4.5 群智感知多目标参与者选取算法

Input：一组感知任务Γ，其中∀q∈Γ。每个感知任务的激励成本 Cq，移动节点完成感知任务 q 就要求 Caq，备选的移动节点集合Ω，移动节点 a 完成感知任务 q 的感知能力为 uaq，移动节点 a 的转移概率矩阵 Q = {qij}
Output：要选择的移动节点集合 S
1：S ← null
2：for each q in Γ do
3： Ω tmp ← Ω
4： Cleft ← Cq
5： for each a in Ω tmp do
6： 计算 a 距离理想点的贴近点 Ta
7： end for
8： while Cleft ≥ Caq do
9： β ← argmina∈Ω tmp {Ta}
10： Cleft = Cq - Caq

```
11:          if Cleft ⩾ 0 then
12:              Ωtmp ← Ωtmp - {β}
13:              S ← S + {β}
14:          end if
15: end for
```

4.5 音频信号发生装置实现

4.5.1 任意信号发生器选择

RIGOL DG4000 系列函数/任意波形发生器曾在 2012 年荣获“美国 R&D100 大奖”，在全球市场得到了媒体、专家、业内以及用户的广泛认可。DG4202 函数/任意波形发生器继承了 DG4000 系列函数/任意波形发生器的诸多优秀功能，标配等性能双通道，具有 500MSa/s 采样率，14bits 垂直分辨率，集函数发生器、任意波形发生器、脉冲发生器和谐波发生器功能于一身，是一款高性能、多功能、性价比突出的实力产品。

4.5.2 音频信号发射模块

超声波探头是在超声波检测过程中发射和接收超声波的装置。探头的性能直接影响超声波的特性，影响超声波的检测性能。在超声检测中使用的探头，是利用材料的压电效应实现电能、声能转换的换能器。探头中的关键部件是晶片，晶片是一个具有压电效应的单晶或者多晶体薄片，它的作用是将电能和声能互相转换。

4.5.3 终端语音助手选择

本系统能够为多种终端语音助手提供语音攻击检测和防范的功能，表 4-1 列举了终端设备及其搭载的语音助手。

表 4-1: 终端语音识别助手选择表

序号	终端设备	操作系统	语音识别助手
1	iPhone 6	iOS 10.2.1	Siri
2	iPhone 7	iOS 10.2.1	Siri
3	iPad mini 4	iOS 10.2.1	Siri
4	MacBook	macOS Sierra	Siri
5	Mi 9	Android 9.0	小爱同学
6	Mi PAD 4	Android 8.1	小爱同学
7	Honor 7	Android 6.0	HiVoice
8	Galaxy note 7	Android 6.0	S Voice
9	Lenovo 小新潮 7000	Windows 10	Cortana

4.5.4 攻击音频制备模块

超声波攻击的方法是利用听不见的声音指令去控制语音助手，其最关键的部分是制备出具有攻击指令的超声波音频信号。详细来说，超声波攻击首先需要制备出一个基带信号——即原始正常音频信号，然后通过调制基带信号使它在经过麦克风放大器时能够有效地被解调为可以被语音助手识别的音频信号。

AM 调制：在 AM 中，载波的幅值与基带信号成比例变化，调幅产生的信号功率集中在载频和两个相邻的边带。调制的逆过程叫解调，调制是一个频谱搬移过程，它是将低频信号的频谱搬到载频位置。从已调信号的频谱中，将位于载频的信号频谱搬移回来。调制和解调都完成频谱搬移，各种调幅都是利用乘法器实现的。

我们采集了多组原始音频，将其分别标签为具有合法或非法内容，用以攻击系统和防范系统的输入：

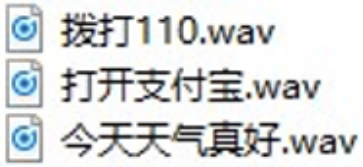


图 4-14: 原始音频图

利用 matlab 数字信号处理，我们实现了将原始语音搭载到高频载波上，制备出超声波攻击用例。图 4-15、16 为音频波形信号图：

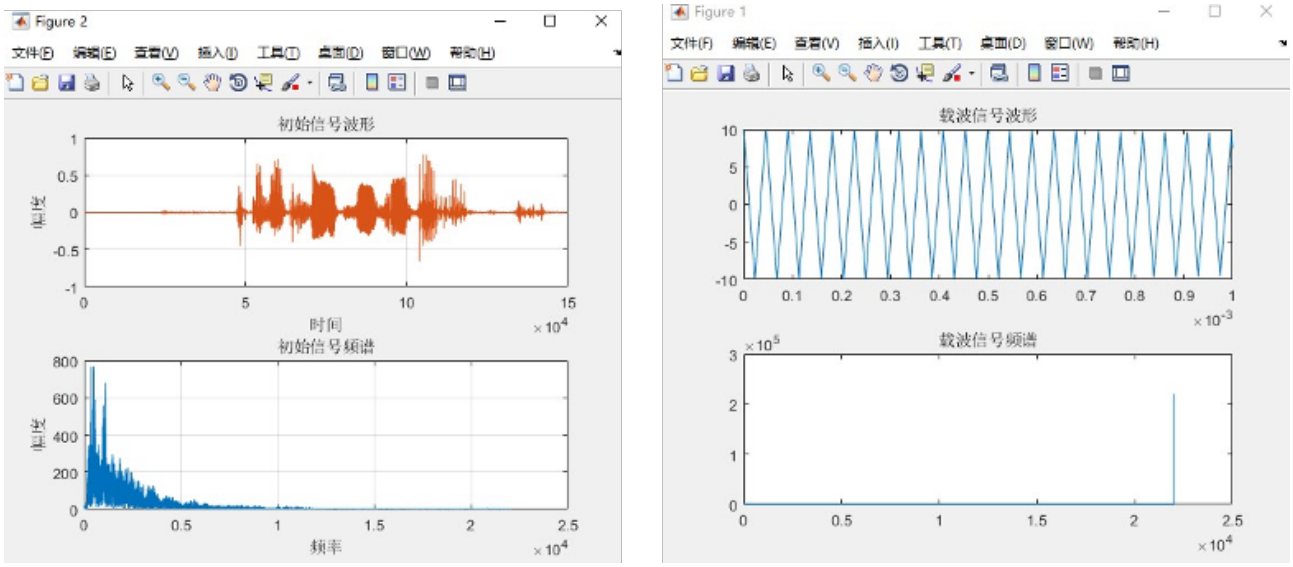


图 4-15: 原始音频信号波形和频谱: 高频载波波形和频谱图

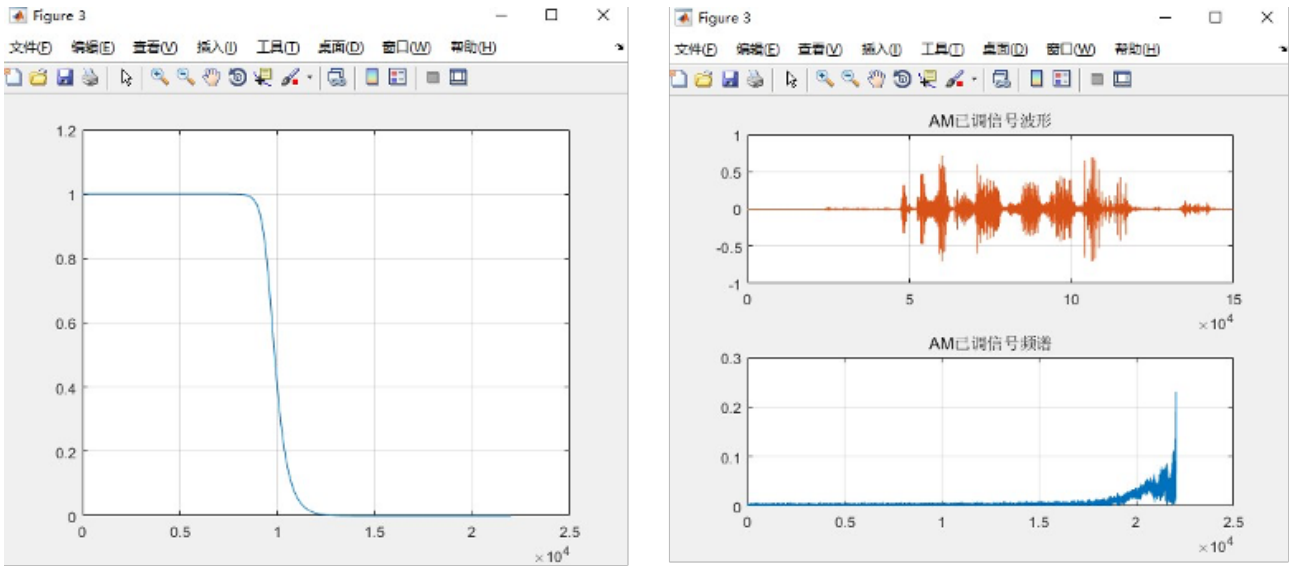


图 4-16: 滤波器波形图: 调制音频信号波形和频谱图

4.6 安全体系实现

本系统使用 HTTPS 通信体系，内置三种国密加解密算法，用于保证系统的实时性、可靠性和低资源消耗，其中国密 SM3 算法用于实现摘要信息散列计算，包括证书的签名以及信息明文的单向加密，保证数据来源的未被篡改；国密 SM4 算法用于对信息明文进行单向加密，此过程可保证用户信息不会暴漏在公网环境中；国密 SM2 算法用于对对称密钥进行加密，保证数据传输过程中不被破解，同时由于该算法只对对称密钥进行加解密，因此速度比一般情况要快。

本系统采用 Openssl 实现 SM2 及其他国密算法，openssl 是一个开源的软件库，主要用来实现安全通信的，其基本组件包括命令、协议、模块，一般上层应用使用函数调用相关加解密方法，来实现安全通信，不过需要注意的是，需要使用的加解密算法需要被 openssl 里的协议支持，同时有相关模块包含该部分算法，否则需要自己实现。以下，就该系统中用到的这些加解密方法进行详细阐述，并在系统中实现这些安全传输相关算法。

(1) SM2 算法

SM2 算法用于实现数据的加解密，通过公钥进行数据加密，接收方用自己的私钥进行解密，保证数据不被非法窃听，同时发送方使用自己的私钥加密一段数据，接收方用对方公钥解密这段数据，用于实现数据发送方的不可否认性，这一步类似于数字签名的过程，不过数字签名还要多一个散列算法。HTTP 请求和响应起到了至关重要的作用，它们帮助我们在互联网上获取所需信息。

如表 4-2 为 SM2 算法实现接口：

表 4-2: SM2 算法实现接口

方法名称	返回类型	方法描述
ec_param_new(ec_param *)	Void	生成椭圆参数结构体
ec_param_init(ec_param *, char**,int,int)	Int	椭圆曲线参数初始化
ec_param_free(BYTE*)	Void	释放结构体空间
set_z_p_r(sm2_dh_st *,BYTE*)	Void	设置 Z_P_R，用于数据传输
SM_gene_R(sm2_dh_st*, BYTE*, BYTE*,ec_param *)	Void	密钥协商函数生成 R 值，用于协商密钥的参数
SM_gene_K(sm2_dh_st*,sm2_dh_st*, ec_param *,int)	Void	密钥生产函数

(2) SM3 算法

SM3 算法用于对信息进行摘要计算，确保数据传输过程中没有被篡改，摘要算法具有不可碰撞性和不可逆性，即很难通过摘要信息得到原文信息，也很难仿造一段不一样的内容使得两者摘要相同。因此该算法可以检验数据中途是否被人恶意篡改或伪造。

如表 4-3 为 SM3 算法实现接口：

表 4-3: SM3 算法实现接口

方法名称	返回类型	方法描述
SM3_Init (void)	Void	初始化参数和环境
SM3_Update(BYTE*, DWORD)	Void	对 message 进行摘要计算
SM3_Final_byte(BYTE*)	Void	以字节方式存储摘要结果

(3) SM4 算法

SM4 直接处理对象为用户传输的明文信息，通过对明文进行加密传输，防止其暴漏在公网环境中，之所以 HTTP 协议不再安全，是因为数据的传输是以明文形式传输的，而非密文。

如表 4-4 为 SM4 算法实现接口：

表 4-4: SM4 算法实现接口

方法名称	返回类型	方法描述
sm4Sbox(unsigned char)	unsigned char	S 盒子置换函数，对输入数据进行置换
sm4Lt(char*, char*, char*, int)	unsigned char	线性变换 L 函数
sm4F(unsigned long, unsigned long, unsigned long, unsigned long,unsigned long)	Void	轮函数，对数据进行多次迭代运算
sm4_one_round(unsigned long*, unsigned char*, unsigned char)	Void	SM4 中置换函数，进行一轮迭代计算

第 5 章 运行效果

内容提要

- 测试方案
- 测试环境
- 算法测试
- 语音模拟攻击测试
- 音频分类过滤测试
- 功能测试
- 实时监测后台系统测试
- APP 客户端测试
- 性能测试
- 安全性测试
- 安全性分析
- 系统运行流程
- 服务器认证
- 非法访问控制
- 音频加密传输
- 小结

5.1 测试方案

本章对本系统的实现进行详细的测试，已验证功能的完善性和运行的正确性。测试主要分为三个部分：算法模块测试、功能测试以及性能测试。

算法测试，本章主要针对攻击音频制备、音频分类过滤算法、语音文本合法性分析、攻击源定位与追踪以及感知任务分配算法进行测试。

功能测试，本章主要为针对了本系统的五大模块进行测试——“身份认证”、“攻击识别”、“实时监测”、“预警提醒”、“大数据分析”进行模块测试。

性能测试，本章分别从实时性、准确性、可移植性的角度进行了测试。

本作品的测试均采用黑盒测试方法，将预先设置的正确结果和系统运行的结果进行对比，判定系统是否正确运行、是否达到预期目标。

5.2 测试环境

系统测试主要包括对指定设备的语音攻击测试、实时数据监测后台测试、客户端预警与反馈测试，因此在测试环境方面我们充分考虑到环境的复杂性、多样性和容错性。针对市面上比较流行的智能手机和语音识别助手，我们都分别对其进行模拟攻击；为保证后台系统在不同机型、操作系统下都能够正常运行，我们也选用了不同种类的设备分别进行测试；针对语音攻击场景可能存在的因素我们也一一纳入考虑范围之内。本章我们还选择了十名健康（尤其是听力正常）的成年男女性来测试我们的音频是否具有有效的攻击性，具体的环境配置和测试准备如表 5-1 所示：

表 5-1: 基于群智感知的语音攻击监测系统测试配置表

<table>
<tr><td rowspan="2">服务器端配置</td><td colspan="2">硬件配置</td><td colspan="2">软件配置</td></tr>
<tr><td colspan="2">CPU：Intel 酷睿 i7 8 代
系统内存：16G
硬盘容量：1T</td><td colspan="2">操作系统：CentOS 7
Restful api: Flask
数据库：MySql Community Server 5.7
环境：python 3.6</td></tr>
<tr><td rowspan="3">客户端配置</td><td>测试机型</td><td>操作系统</td><td>语音识别助手</td><td>运行内存</td></tr>
<tr><td>iPhone 6</td><td>iOS 10.2.1</td><td>Siri</td><td>4G</td></tr>
<tr><td>iPhone 7</td><td>iOS 10.3.1</td><td>Siri</td><td>4G</td></tr>
</table>

	iPad mini 4	iOS 10.2.1	Siri	8G
	Mi 9	Android 9.0	小爱同学	4G
	Mi PAD 4	Android 8.1	小爱同学	8G
	Honor 7	Android 6.0	HiVoice	4G
	Galaxy note 7	Android 6.0	S Voice	4G
	Vivo x7	Android 6.0	ViVoice	4G
实时数据监测系统	**计算机型号（CPU/运行内存）**	**操作系统**	**浏览器版本**	**分辨率**
	MacBook Air (core i5/8G)	macOS Sierra	Chrome	1920X1080
	MacBook Air (core i5/8G)	macOS Sierra	Firefox	1920X1080
	MacBook Air (core i5/8G)	macOS Sierra	Safari	1920X1080
	Lenovo 小新潮 7000 (i5-8250U/8G)	Win10	Chrome	1920X1080
	Lenovo 小新潮 7000 (i5-8250U/8G)	Win10	Firefox	1920X1080
	Lenovo 小新潮 7000 (i5-8250U/8G)	Win10	ie	1920X1080
	小米笔记本 Air 13 (i5-8250U/8G)	Win7	Chrome	1920X1080
	小米笔记本 Air 13 (i5-8250U/8G)	Win7	Firefox	1920X1080
	小米笔记本 Air 13 (i5-8250U/8G)	Win7	ie	1920X1080
攻击音频有效性检测	测试人员（编号）	性别	年龄	听力健康（kHZ/LdB/RdB）
	1	男	21	3/30/35
	2	女	22	4/35/40
	3	男	21	3/40/60
	4	男	18	3.5/30/30
	5	男	23	5/50/45
	6	女	37	4/40/35
	7	男	35	3/30/45
	8	男	32	3/40/60
	9	男	14	6/50/40
	10	女	13	6/35/40
攻击场景	环境噪音	攻击方式	被攻击者状态	被攻击设备状态
	较强	超声波攻击	携带设备	移动
	中等	超声波攻击	离开设备	静止
	较弱	超声波攻击	离开设备	静止
	较强	模糊指令攻击	携带设备	移动
	中等	模糊指令攻击	离开设备	静止
	较弱	模糊指令攻击	离开设备	静止
	较强	对抗性语音攻击	携带设备	移动
	中等	对抗性语音攻击	携带设备	移动
	较弱	对抗性语音攻击	离开设备	静止

5.3 算法测试

算法测试部分主要对模拟攻击场景和音频分类过滤的测试。

5.3.1 语音模拟攻击测试

本节测试主要根据已存在的三种主要语音攻击 [2][3][5] 进行攻击有效性测试。我们使用黑盒测试的方法，抽取样本，分别让机器和人来识别音频，从而确保音频能够有效地在人无法意识到的情况下攻击设备。

表 5-2: 语音模拟攻击测试用例

用例编号	NO.01		
模块名称	攻击音频	测试方法	黑盒测试
测试说明	输入攻击音频和正常音频，分别让语音识别系统和人理解		
预置条件	无		
判断准则	能被语音识别系统成功识别，但无法被人察觉		
测试输入	多种攻击音频： 模糊指令攻击 [2] 对抗性语音攻击 [5] 超声波攻击 [3] 正常音频		
测试输出	语音识别系统输出、志愿者反馈		
测试评价	超过 94.6% 的攻击音频　符合攻击标准 超过 88.4% 的攻击音频　符合攻击标准 超过 81.2% 的攻击音频　符合攻击标准		

（1）模糊指令测试

模糊指令是指一段能够被机器识别却不能够被人理解的音频 [2]。对于模糊指令攻击情况，我们测试了三种攻击语音“发邮件”、“打开支付宝”和“拨打电话”，分别输入识别系统以及给志愿者听。系统测试样本包括 50 个正常音频和 50 个攻击音频，考虑到人具有主观性所以对于人的测试样本增加到 100 个。测试情况如表 5-3 所示，其中“正”代表正常音频，“攻”代表攻击音频，对应给出音频识别准确率及样本数情况。

表 5-3: 模糊指令测试情况表

	发邮件给 xxx		打开支付宝		拨打电话给 xxx	
	识别系统	人	识别系统	人	识别系统	人
正	92%(46/50)	98%(98/100)	90%(45/50)	96%(96/100)	92%(46/50)	94%(94/100)
攻	84%(42/50)	22%(22/100)	80%(40/50)	18%(18/100)	88%(44/50)	24%(24/100)

考虑到周围环境噪音可能对测试产生干扰，我们将周围噪音的影响因素也考虑在测试范围内，即用信噪比（信号功率/噪声功率）（dB）来表示影响因子。上表测试结果是在 15dB 的信噪比环境下进行的测试结果。如图 5-1 所示为在不同信噪比（从 5dB—35dB）下、对于不同系统（包括 Android 和 IOS）的识别精度的变化趋势图：

（2）对抗性语音攻击测试

现有语音识别系统主要依靠 CTC+RNN 神经网络搭建，对抗性语音攻击正是利用这种神经网络存在的漏洞设计出的一种攻击，主要表现为人所听到的语音和机器识别出来的含义完全不一样 [5]。具体来说如图 5-2 所示：

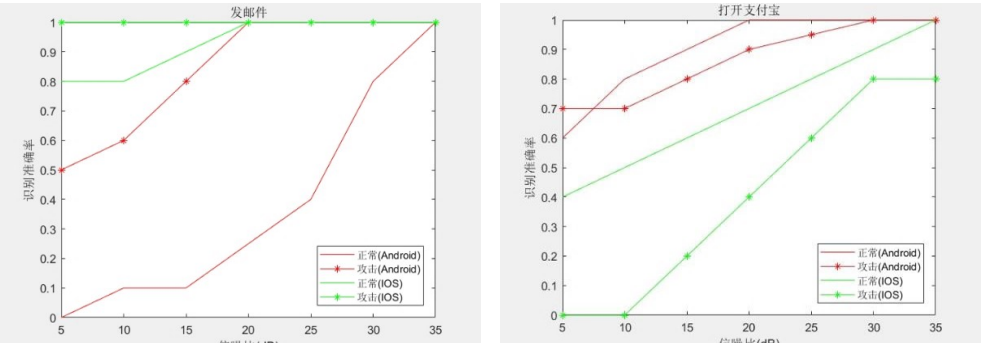


图 5-1: 不同信噪比及不同识别系统下识别精度变化图

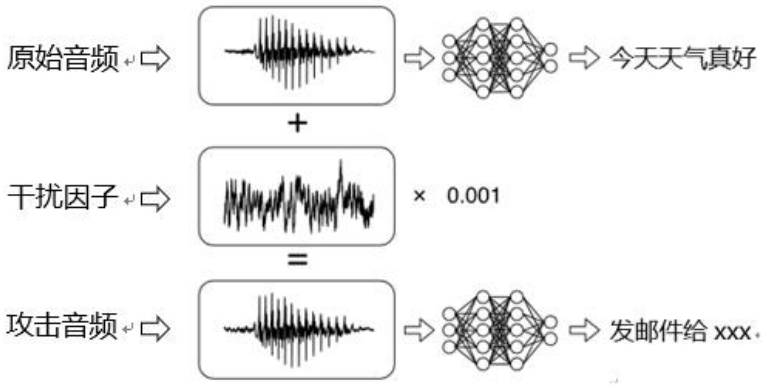


图 5-2: 对抗性语音攻击原理说明图

同样我们基于三种指令进行测试，其中原始音频内容是不具有攻击性的，经过模型训练，我们将训练后的音频输入语音识别系统，期望识别出的语义是具有攻击性的。通过测试得出 实际输出满足期望输出的准确率；原始音频与攻击音频在音频信号层面的差别小于 5% 的样本数（图 5-3 中橙色代表原始音频波形，蓝色代表攻击音频波形，可以看到两者仅在某些细小的地方存在差别，总体拟合度非常好）。

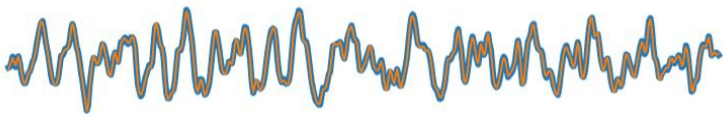


图 5-3: 原始音频与攻击音频拟合情况

针对每条指令内容，我们分别进行了样本数为 30 的测试。最终的测试结果如表 5-4 所示，可以看出攻击音频满足预定攻击标准，能够有效攻击设备。

表 5-4: 对抗性语音攻击测试情况表

	期望输出	原始音频	实际输出准确率	音频波形差别 <5% 的样本数
Siri	发邮件	你好呀	82.1%	26/30
	打开支付宝	今天天气真好	85.4%	24/30
	打电话	谢谢你	88.6%	27/3
小爱同学	发邮件	你好呀	92.1%	22/30
	打开支付宝	今天天气真好	89.4%	21/30
	打电话	谢谢你	88.1%	24/30
HiVoice	发邮件	你好呀	87.1%	28/30
	打开支付宝	今天天气真好	89.4%	29/30

	打电话	谢谢你	84.6%	29/30

（3）超声波攻击指令测试

超声波攻击是利用手机麦克风非线性放大器的漏洞、设计出的一种人耳完全无法察觉的语音攻击 [3]。

a) 声波传输距离测试

在超声波攻击中，隐蔽性是一个重要的特性。超声波攻击源距离被攻击设备的距离越远，则意味着攻击者有更加广阔的攻击范围，攻击效果也越好。下面我们测试了超声波搭载语音指令“发邮件给 xxx”，分别攻击 Siri\ 小爱同学 \ HiVoice\ ViVoice\ SVoice 语音识别助手，记录了识别率随声波传输距离的变化关系：

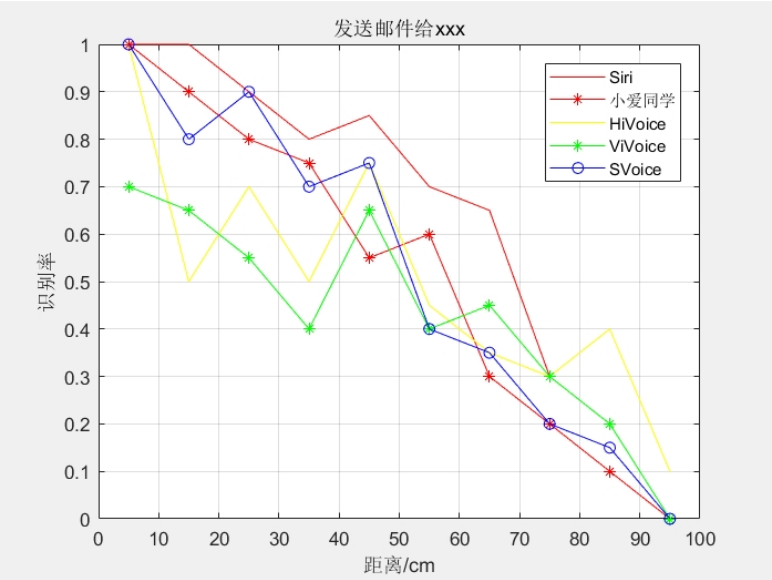


图 5-4: 各种语音助手识别率与超声波发射源的距离关系曲线

b) 声波发射角测试

实际测试过程中，我们发现超声波攻击的有效性与其发射角的方向有很强的关系。声波发射角具体来说就是超声波探头垂直发射方向与接收设备之间的夹角（如图 5-5 所示），试验装置使用了两个探头，垂直发射方向选取两探头的中垂线方向。我们发现越偏离发射器所朝方向，感受到声波的能量就越低，识别效果可能就会越差。

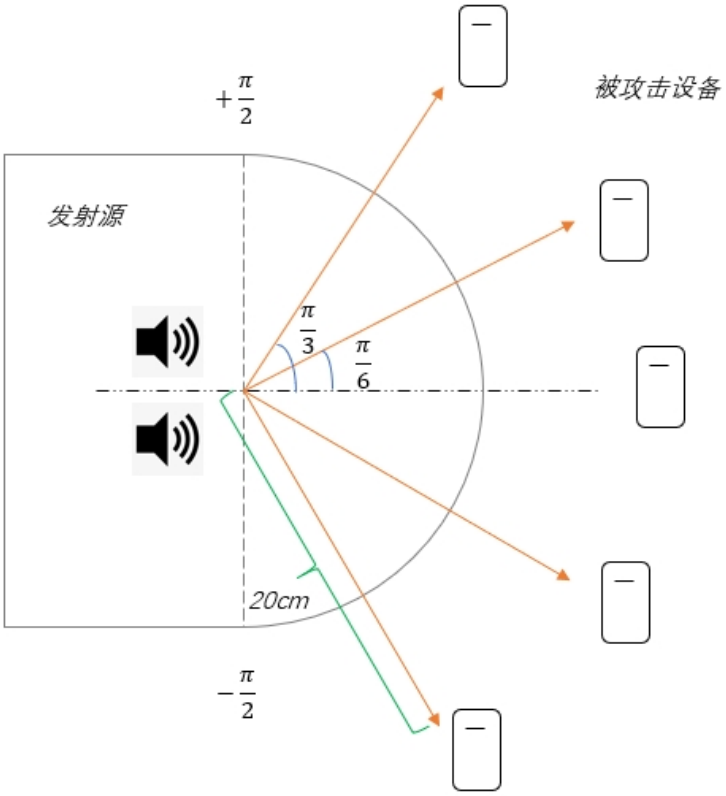


图 5-5: 声波发射角示意图

依据这个重要的发现，我们设定了发射角在 $[-\frac{\pi}{2},+\frac{\pi}{2}]$ 范围内的测试规约，控制声波传输距离为 20cm，超声波指令选取“发邮件”。测试结果如表 5-5 所示（针对五种语音助手，在不同发射角情况下的识别率，单位为%）

表 5-5: 声波发射角测试结果表

	Siri	小爱同学	HiVoice	ViVoice	SVoice
-π/2	0	0	0	0	0
-π/3	10	15	20	15	10
-π/4	35	30	35	25	25
-π/6	55	50	45	35	50
0	80	85	85	75	80
+π/6	55	55	50	40	40
+π/4	35	40	30	25	30
+π/3	15	20	15	10	15
+π/2	0	0	0	0	0

5.3.2 音频分类过滤测试

本节测试主要用来验证系统对于三种主要语音攻击 [2][3][5] 的防范有效性。我们使用控制变量的方法，在多种攻击音频和环境设置条件下，记录软件防范攻击的有效性。根据本部分测试内容的特性，选取如下三个属性作为控制变量：发射角、发射距离、攻击内容。我们每次仅测试其中一个控制变量，控制另外两个相同，观察这一个控制变量的改变对实验结果的影响。

表 5-6: 音频分类过滤测试用例

用例编号	NO.02	模块名称	音频分类过滤
测试方法	黑盒测试	测试日期	2023.4.15
测试说明	分别在关闭和开启软件的条件下， 控制其他环境变量相同并进行多组交叉验证，观测音频攻击防范结果		
判断准则	软件是否对语音攻击分类准确		
测试输入	多种攻击音频： ①模糊指令攻击 [2] ②对抗性语音攻击 [5] ③超声波攻击 [3] 正常音频		
测试输出	分类标签		
测试评价	超过 98.2% 的攻击音频　可以被正确分类和拦截 超过 88.2% 的攻击音频　可以被正确分类和拦截 超过 93.5% 的攻击音频　可以被正确分类和拦截		

（1）控制环境变量一：发射角

本部分测试旨在验证不同发射角情况下，软件对三种语音攻击的防范有效性。其他两种环境变量保持相同，发射距离均为 50cm，攻击内容均为“打开支付宝”。观测信息记录为 (AAA/BBB) 形式，AAA 表示不开启软件的情况下 50 次攻击成功的次数；BBB 表示开启软件的情况下 50 次攻击被正确分类防范的次数。

表 5-7: 不同发射角下语音防范测试情况

发射角	正常语音	超声波攻击	模糊指令攻击	对抗性语音攻击
0	50/50	48/46	50/50	50/48
$+\pi/6$	50/50	44/45	50/49	50/44
$+\pi/3$	50/50	42/46	48/50	48/46

（2）控制环境变量二：发射距离

本部分测试旨在验证不同发射距离情况下，软件对三种语音攻击的防范有效性。其他两种环境变量保持相同，发射角均为 0，攻击内容均为“打开支付宝”。观测信息记录为 (AAA/BBB) 的形式，AAA 表示不开启软件的情况下 50 次攻击成功的次数；BBB 表示开启软件的情况下 50 次攻击被正确分类防范的次数。

表 5-8: 不同发射距离下语音防范测试情况

发射距离	正常语音	超声波攻击	模糊指令攻击	对抗性语音攻击
0.5m	50/50	48/46	50/50	50/46
1.0m	50/50	44/45	50/49	50/44
3.0m	42/44	42/46	44/47	42/45

（3）控制环境变量三：攻击内容

本部分测试旨在验证不同攻击内容情况下，软件对三种语音攻击的防范有效性。其他两种环境变量保持相同，发射角均为 0，发射距离均为 50cm。观测信息记录为 (AAA/BBB) 的形式，AAA 表示不开启软件的情况下，50 次攻击成功的次数；BBB 表示开启软件的情况下，50 次攻击被正确分类防范的次数。

表 5-9: 不同攻击内容下语音防范测试情况

攻击内容	正常语音	超声波攻击	模糊指令攻击	对抗性语音攻击
打开支付宝	50/50	48/46	50/50	50/48
发邮件	50/50	46/45	50/49	50/44
打电话	50/50	48/46	48/50	48/46

以上测试结果如图 5-6 所示：

图 5-6: 声波发射角示意图

5.4 功能测试

5.4.1 实时监测后台系统测试

后台管理系统设计之初就是为了实现集收集、分析、决策以及群智感知任务发布与接收于一体的实时数据检测平台。考虑到系统业务逻辑复杂，数据流量较大，因此后台系统进行以下四个测试，以确保系统的实时性、安全性、可用性和可靠性。通过系统权限测试，确保后台系统能够正确的识别管理者，以分配给不同管理者不同的管理权限，方便作业人员对系统的操控，也方便对管理员的管理；任务发布与接收测试，旨在系统能够正确的处理管理员布置任务的入库与发放操作，能够将任务精准的推送给相应用户，即做到用户细分，而当用户对任务进行反馈操作后，后台系统能够正确的检测出用户的行为，以及对用户操作给予奖励或默认操作；周报生成与决策分析测试，当系统完成一个周期的数据收集行为后，就要做出正确的周报生成和决策分析操作，以辅助作业人员更好的对所在辖区的恶意音频源进行排查处理；热力图动态更新测试，这一部分主要测试系统能否准确、美观、简洁的将数据呈现给作业人员。

(1) 管理权限测试 后台系统管理人员权限等级分为：可查看，可部署和可查看和高级管理员。其中，可查看管理员只能够获取热力追踪模块和灾害报表模块，而可部署和可查看管理员除热力追踪和灾害报表模块，还可进行感知任务的发放与接收操作，高级管理员可以进行所以业务逻辑功能，即在前面的基础上，新增一项对管理员权限的管理。测试用例如表所示，测试结果如表 5-10 所示。

表 5-10: 权限管理模块测试用例

用例编号	NO.03		
模块名称	权限管理模块	测试方法	黑盒测试

测试说明	不同管理员登录系统后，系统能够呈现出相应的系统接口供管理员使用，管理员不得出现跨权限操作
预置条件	以获取三种不同管理员账号
判断准则	可查看管理员只能拥有热力追踪和灾害报表模块， 可部署和合查看管理员拥有热力追踪、灾害报表和任务中心模块， 高级管理员还可以进行管理员权限设置操作
测试数据	登入可查看管理员账号；登入可部署和卡查看管理员账号登入高级
测试输出	系统展现出对应的管理用功能接口，功能使用正常
测试评价	系统权限模块测试通过，系统可根据不同管理员权限呈现不同功能

图 5-7 为可查看、可部署和可查看两种管理员对应系统的功能接口展现，可以看到可查看权限管理员的系统界面只有热力追踪和灾害报表接口，而可部署和可查看权限的管理员还拥有任务中午中心接口，以进行群智感知任务的发放与接收操作。

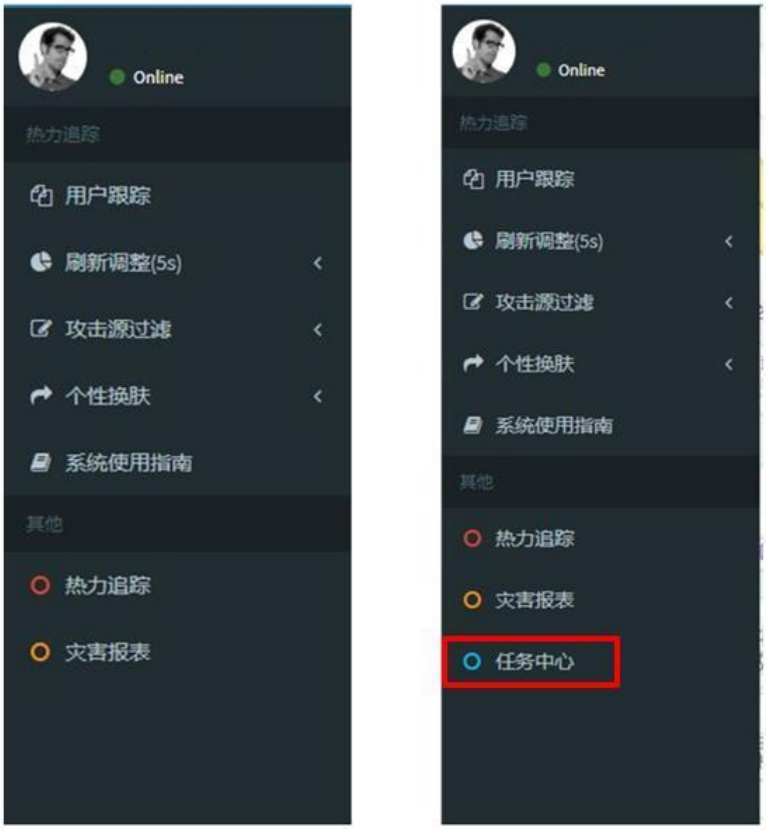


图 5-7: 可查看、可部署和可查看管理员对应系统呈现

（2）任务发布与接收测试

A. 发布测试

后台管理人员除了能够对数据进行收集和查看外，还能够进行群智感知任务发放与接收，这是后台管理系统一大核心功能点，旨在通过大量用户群体的数据量上传结合云端算法，以帮助分类算法更好的优化改进，给手机用户更“安全”的体验。因此任务发布与接收测试就是测试系统能否正常的进行这一部分操作，包括任务数据的入库，手机细分用户的接收以及系统对用户感知任务反馈的响应。这一部分由于业务逻辑复杂，因此分为两个测试用例进行测试。测试用例如表 5-11 所示，测试结果如图 5- 8、9 所示。

表 5-11: 任务发布模块测试用例

用例编号	NO.04		
模块名称	任务发布模块	测试方法	黑盒测试
测试说明	管理员发放任务以后，系统能够正确的将任务数据入库， 而且能够做到细分用户，即只将任务推送给相应的用户		
预置条件	可部署和可查看的登入		
判断准则	管理员布置任务以后，系统正确的进行了任务数据入库， 且只有满足任务要求的用户才能收到任务的推送		
测试数据	感知任务数据（地点：XXXX 大学（XX 校区）二教， 工作时：11：00-13：00，推荐设备：iphone6，语音助手：siri）		
测试输出	系统任务数据库进行加 1 操作，展示界面已发布数量加 1， 满足任务要求的用户收到任务推送		
测试评价	任务发布模块测试通过，任务可正常布置，且推送操作合理人性		

以下两张图说明了系统能够正确的进行已发放任务的入库操作，展示结果体现出了可用性、可靠性和界面友好。

图 5-8: 任务操作

图 5-9: 任务发布数据库入库

如图 5-10 所示，使终端用户的任务推送结果显示，用户手持设备为 Android 系统，并且其位置处在x区，所以可以收到发布在其附件的任务信息。

图 5-10: 用户细分推送任务

B. 接收测试

任务发布后，用户会陆续完成任务并返回，这里我们对任务反馈接收模块进行测试，测试用例如下表 5-12 所示：

表 5-12: 任务反馈接收模块测试用例

用例编号	NO.05		
模块名称	任务反馈接收模块	测试方法	黑盒测试
测试说明	终端用户查看推送任务后，可进行任务的接收或拒绝操作， 那么相应的也要进行数据库入库操作，和系统数据展示， 通过查看数据展示部分以测试模块功能是否正常		
预置条件	任务已发布，多个用户对任务进行反馈操作		
判断准则	点击确认接受操作后，已接收数量加 1，查看人数加 1， 拒绝操作，查看人数加 1。用户达成任务条件后， 以回收数量加 1，待回收数量减 1		
测试数据	甲用户接收任务，并完成任务，乙用户拒绝接收任务。		
测试输出	甲用户操作完成以后，各类指标数量加 1，乙用户操作完成以后， 除查看人次外，其余指标减 1		
测试评价	任务反馈接收模块测试通过，系统能够正确处理布置任务的后续操作		

如图 5-11 所示是任务反馈接收模块的测试结果，可以看到，当用户甲用户点击确认接收任务操作以后，查看人次数量加 1，已接收人次数量加 1；而当用户完成该任务以后，已回收数量加 1。

图 5-11: 任务反馈系统展示

（3）周报生成

后台管理系统一大核心功能就是实现对数据的收集、分析汇总操作，那么测试这一模块，就间接的测试了后台管理系统的数据库，业务逻辑操作以及大数据量操作的鲁棒性。通过测试人员事先完成数据的模拟采集任务和数据上传，以确保数据库中含有大量接近真实数据，测试过程中，随机进行数据的上传，以确保系统的实时性。测试用例如表 5-13 所示，测试结果如图 5-12 所示。

表 5-13: 周报生成测试用例

用例编号	NO.06		
模块名称	周报生成	测试方法	黑盒测试
测试说明	后台管理系统能够满足数据的收集和处理操作，正确的将历史数据呈现在管理界面中。同时系统须达到一定实时性和可靠性，在大量数据并发上传的过程中，不会出现系统崩溃、延时响应等状态。		
预置条件	数据库中已含有多条“音频数据”，测试过程中，测试人员进行数据的随机上传。		
判断准则	系统能够完全的、正确的将数据库中数据呈现在系统当中，并且不会出现数据冗余，确保每一条展示数据的价值性、可靠性和真实性。实时上传过程中，数据会呈现出动态更新，但响应延时要控制在一定范围内。		
测试数据	测试前上传的大量数据，测试过程中，随机上传的数据。		
测试输出	后台管理系统输出历史所有数据，动态更新当前上传的数据		
测试评价	周报生成模块测试通过，历史数据呈现正确，当前数据实时更新。		

测试前测试人员已经上传的数据（海豚音攻击 66 起，模糊指令攻击 108 起，对抗性语音攻击 235 起）。数据的上传保证了模拟真实性、多样性和客观性。可以看到，测试人员实现上传了 66 条海豚音攻击事件数，108 条模糊指令攻击数，235 对抗性语音攻击数，后台管理系统的如数将上述数据呈现出来。

图 5-12: 周报动态更新展示

上图所示是测试环节中，测试人员随机上传的数据，发现，与测试前对比，各项数据都发生了改变，测试结果动态更新周期为 10s 一更新，响应延迟为 107ms，满足了实时数据的动态更新。

（4）热力图动态更新测试

热力图的呈现是以终端用户为单位，进行数据的点对点呈现，而如何保证大量用户点的数据呈现高精度和高识别性就是模块的关键部分了，由于这部分数据往往呈现出流量大、高动态性，因此测试部分应尽量模拟真实数据分布，由于数据包含经纬度部分，自行构造可能有失精度且费时费力，所以，测试数据从百度 api 中下载，模拟某地区一段时间内活跃用户分布，测试数据已经过检验，可以模拟被音频攻击者上传的数据。通过热力追踪模块的查看，作业人员应当清楚、方便的识别出某地区对应的攻击事件，并且要做到数据动态更新。测试用例如表 5-14 所示，测试结果如图 5-13 所示。

表 5-14: 热力追踪模块测试用例

用例编号	NO.07		
模块名称	热力追踪模块	测试方法	黑盒测试
测试说明	管理系统应当正确呈现出被攻击用户的数据，这部分数据应包含攻击种类、地点分布，且攻击密集区域，热力分布呈红色，稀疏区域呈黄色。还应保证数据动态更新，低延迟性		
预置条件	大量用户被攻击数据，且已标注了对应的攻击种类		
判断准则	在筛选条件为精确搜索下，所有数据精准呈现，即点对点；筛选条件为海豚音攻击下，屏蔽所有机器学习攻击点；筛选条件为模糊搜索下，区域显示攻击点，及点数据相近的情况下，合并为一个较大的热力点		
测试数据	大量用户被攻击数据，且已标注了对应的攻击种类，筛选条件：精确搜索，模糊搜索，仅海豚音攻击，仅机器学习攻击		
测试输出	精确搜索：所有数据精确呈现；模糊搜索，所有相近数据合并为较大热力区域；仅海豚音、精确搜索：海豚音攻击数据精确呈现；仅机器学习、模糊搜索：机器学习攻击数据模糊呈现。		
测试评价	热力追踪模块测试通过，满足可用性、界面友好性		

下方左图所示是精确搜索条件下的结果显示，可以看到，所有的点，包括海豚音攻击事件和机器学习攻击事件对应的点都展示了出来，而当过滤条件为仅海豚音攻击时，图中去除了大量的对应机器学习攻击的点，这样一种呈现方式可以更作业人员迅速的查找的对应区域的攻击事件。在动态更新测试中，更新周期为 5s（可自定义调节为 5s、10s、20s），响应延迟为 120ms，满足了实时性。

图 5-13: 热力追踪模块系统展示

5.4.2 APP 客户端测试

(1) 用户与设备信息认证测试

用户登录测试需要用户输入一定的个人信息 (手机号)、以及登录密码进行验证登录，设计的测试用例如表 5-15 所示。

表 5-15: 用户注册认证测试用例

用例编号	NO.08		
模块名称	用户登录测试	模块名称	用户登录测试
测试说明	验证测试人员是否能顺利登录		
预置条件	测试人员未登录		
判断准则	用户输入格式以及内容是否正确		
测试输入	用户手机号　132 密码输入：123456		
测试输出	用户登录成功进入系统界面		
测试评价	测试通过，用户登录功能正常		

测试过程以及结果如图 5-14 所示，用户点击“登录”按钮后即可进入 App，进行登陆操作。

图 5-14: 用户注册客户端信息

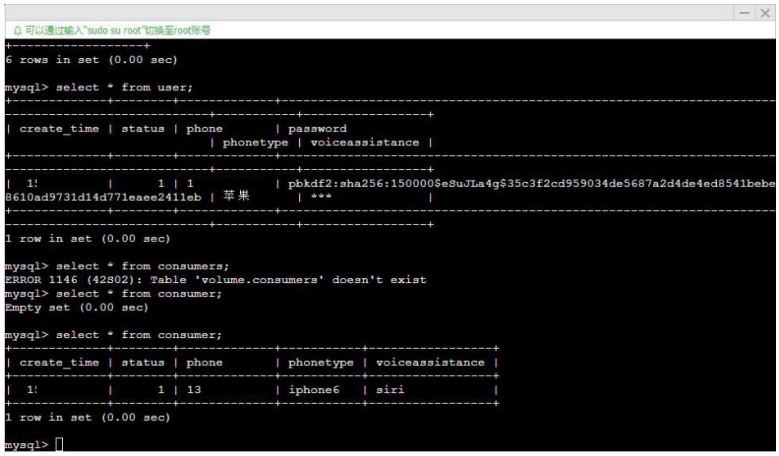


图 5-15: 平台登录信息

（2）音频接收与传输测试

用户在防御界面点击“开始防御”按钮时，客户端获取到麦克风权限，并对用户周围语音情况进行实时监听，点击“开始检测”按钮即可上传语音信息进行处理，并返回处理结果，设计的测试用例如表 5-16 所示。

表 5-16: 语音传输测试用例

用例编号	NO.09		
模块名称	语音传输测试	测试方法	黑盒测试
测试说明	验证测试人员是否能成功检测音频信息		
预置条件	测试人员对周围环境进行一定时间的监听		
判断准则	能否判断出防御期间内是否遭到攻击		

测试输入	已知的攻击音频
测试输出	成功防御并提醒用户已在防御期间受到攻击
测试评价	测试通过，音频传输接收功能正常

测试过程以及结果如图 5-16、17、18 所示，测试人员用制备好的攻击语音对测试人员手机进行攻击，进行检测后发现了攻击音频信号，客户端对用户进行提醒，并建议用户继续防御，为用户提供了有效地语音攻击防御。

图 5-16: 语音监听阶段

图 5-17: 语音传输检测结果

图 5-18: 音频分析结果

(3) 预警测试

客户端实时获取到用户手机的地理位置信息，并上传与服务器中已标记的危险攻击源进行聚类分析，将得到的结果返回至客户端为用户提供一定的危险预警功能。

表 5-17: 预警测试用例

用例编号	NO.10		
模块名称	预警传输测试	测试方法	黑盒测试
测试说明	验证测试人员是否能收到预警信息		
预置条件	测试人员手机允许客户端获取地理位置权限		
判断准则	能否检测出用户周围已知攻击源的数量并预警		
测试输入	用户地理位置信息 (经纬度信息)		

测试输出	接收到周围攻击源个数信息
测试评价	测试通过，攻击源预警功能正常

测试结果如图 5-19、20 所示，服务器接收到位置信息后，进行分析后将结果返回至客户端，以通知栏的形式对用户进行预警。

图 5-19: 攻击源提醒

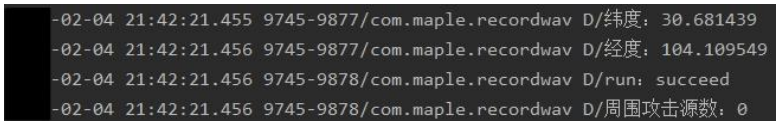


图 5-20: 服务器返回结果

（4）感知任务测试

感知任务模块用于接受完成平台上所发布的感知任务，该测试用于检测客户端能否接收到平台所发布的感知任务，并进行操作，设计的测试用例如下表 5-18 所示。

表 5-18: 感知任务测试

用例编号	NO.11		
模块名称	预警传输测试	测试方法	黑盒测试
测试说明	验证测试人员能否接收到任务、能否接受、拒绝任务		
预置条件	测试人员登陆客户端查找任务		
判断准则	在任务发布期间能否接收到任务、并接受或拒绝		
测试输入	用户唯一标识符		
测试输出	接收到已发布的任务、并接受或拒绝		
测试评价	测试通过，感知任务功能正常		

测试的结果如图 5-21、22、23 所示，测试人员在任务发布期间查找到任务后，点击“接受任务”按钮，会在平台上进行记录，在规定时间内完成任务，即可得到发布的任务地点在任务时间内的大致情况。

图 5-21: 平台发布任务

图 5-22: 接收任务

图 5-23: 接受发布任务

（5）个性化报表生成测试

客户端以一周作为周期为用户定制一份量身定做的个性化报表，通过柱状图等可视化形式为用户提供更为直观的用户安全信息，设计的测试用例如下表 5-19 所示：

表 5-19: 个性化报表生成测试用例

用例编号	NO.12		
模块名称	预警传输测试	测试方法	黑盒测试
测试说明	验证测试人员能否接收到周报		
预置条件	测试人员登陆客户端申请周报		
判断准则	在任务发布期间能否接收到周报		
测试输入	用户唯一标识符		
测试输出	接收到个性化周报		
测试评价	测试通过，个性化报表生成功能正常		

测试结果如图 5-24 所示，获取周报后用户可以看到一周内自身生活环境的音频安全状况。

图 5-24: 个性化周报

5.5 性能测试

性能测试主要考量系统在满足可用性的情况下进而测试其实时性和准确性，实时性部分分为用户端的 app 实时性和管理系统的响应实时性，测量点体现在单次请求下的响应延迟，和多用户并发下的请求响应延迟；准确性测试主要测试后端的几大接口，由于各类接口都属于 restful 类型，因此测试接口返回的 json 数据格式以及内容是否符合要求即可判断系统的准确性。而可移植性测试是性能测试的一个附加考察点，旨在测试客户端和管理端在不同设备、不同操作系统版本号和不同浏览器下的适配程度，主要的测试指标为各类功能响应正常与否和可用率（可用的功能模块与总功能模块数量比值）。

实时性测试：

1、手机移动 app

本测试环节通过测试 app 对各类操作的数据收发响应延迟以及数据渲染，判断客户端的总体响应时间，并由此给出用户关于操作延迟的体验评估。测试部分使用 Android Studio 自带的测试工具——Memory Profiler。测试实例如表 5-20 所示，结果如图 5-25 所示。

表 5-20: 客户端实时性测试

用例编号	NO.13		
模块名称	客户端实时性测试	测试方法	黑盒测试
测试工具	Android Studio 自带测试工 Memory Profiler。		
环境参数	关闭除本客户端外的所有应用，以排除其他应用的影响		
判断准则	客户端不同功能按钮测试响应时间的标准不一，一般来说低于 300ms 为低延迟，300-600ms 为中等延迟，高于 1000ms 为高延迟。		
测试输入	在短时间内连续向服务器发送多个数据、观察每次数据传输的波动情况		
测试输出			

功能	数据平均 发送速率	数据平均 接受速率	数据传输接受 平均响应时间	用户界面 平均响应时间
注册	5.1KB/s	3.1KB/s	200ms	340ms
攻击源预警	2KB/s	1.2KB/s	187ms	242ms
音频防御检测	740KB/s	8.1KB/s	980ms	1.4s
任务中心	10KB/s	7.3KB/s	100ms	150ms
测试评价	在短时间内连续点击按钮进行数据的传输接受，在不排除网络波动的情况下，因为音频防御检测功能传输的数据与其他功能相比较而言更大，客户端的响应时间相对来讲较长，但也在可控范围之内，用户各个功能的使用上不会因前后端数据的交互而感到不适。因此本客户端实时性强，可以满足用户需求。			

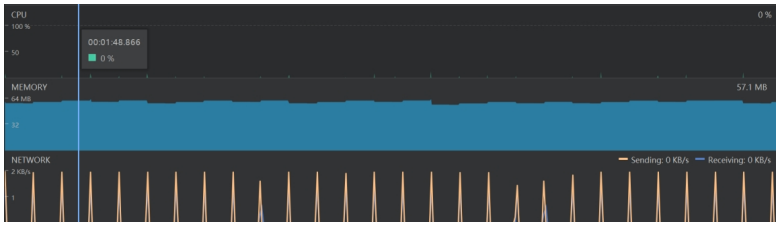


图 5-25: 客户端测试

2、管理系统

管理系统测试，由于管理系统的使用特点多为并发量小，连接时间长，因此测试环境的并发量要求就显得不为那么苛刻了，而是在于系统能否保持长连接状态下的响应时间能力。测试用例及结果如表 5-21 所示，测试性能分析报告如图 5-26 所示。

表 5-21: 管理系统实时性测试

用例编号	NO.14					
模块名称	后端管理系统实时性测试	测试方法	半自动化测试			
测试工具	chrome 浏览器开发者工具，在 network 一栏中进行响应延迟测试					
环境参数	模拟 n 个管理员同时对相同模块进行访问，n=50；开启网页渲染监测器，禁用缓存, 关闭日志保留，关闭域外数据测试，以排除域外文件对系统的延迟影响					
判断准则	单文件响应延迟测试，低于 100ms 为低延迟，100-300ms 为中等延迟，高于 300ms 为高延迟。总接收渲染响应测试，低于 2s 为低延迟，2-5s 为中等延迟，高于 5s 为高延迟					
测试输入	50 个用例线程同时测试，针对每个测试线程，测试对象为单个页面的各个文件加载速度和最终渲染速度					
测试结果						
页面模块	单文件低延迟数量百分比	单文件中延迟数量百分比	单文件高延迟数量百分比	页面最终渲染低延迟占测试线程个数百分比	页面最终渲染中延迟占测试线程个数百分比	页面最终渲染高延迟占测试线程个数百分比
注册	92%	8%	0%	100%	0%	0%
热力追踪	87%	6%	7%	10%	42%	48%

灾害报表	85%	7%	8%	92%	8%	0%
任务中心	87%	8%	5%	96%	4%	0%
权限管理	94%	6%	0%	100%	0%	0%
测试评价	在 50 个并发访问情况下，除热力追踪部分，其余功能模块均满足用户实时性要求，无论是单个文件的响应延时，还是总页面的接收渲染速度，都在低延迟等级中。而关于热力追踪部分，由于地图的渲染引用的是百度 API 接口，因此总渲染延迟取决于百度地图的数据上传能力，经多次测试，页面渲染的最终时间很大一部分比重在地图的获取上，且该部分时间与官方地图渲染时间相差无几，另外，由于用户缓存原因，在热启动时渲染速度会明显加快。因此本系统实时性强，可以满足用户需求。					

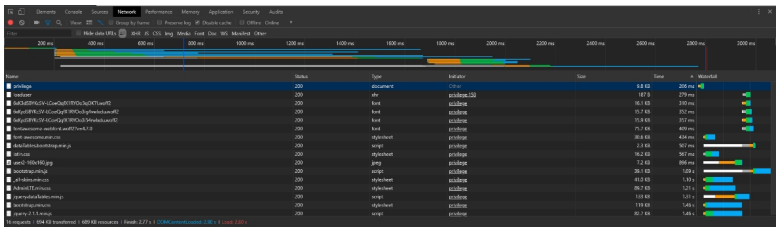


图 5-26: 后端系统权限管理界面测试结果

准确性测试：

准确性测试体现在针对不同请求，其对应的接口能否返回正确数据，由于客户端和管理系统访问的是同一个服务器下的若干接口，因此，直接才用接口测试，并以返回状态和内容作为测试指标点即可完成系统的准确性测试。此外，关于前端数据接收和渲染的准确性测试已经在功能性测试部分里测试，因此这里主要考察数据服务和传输环节。测试用例及结果如表 5-20 所示。

表 5-22: 数据接口服务及传输准确性测试

用例编号	NO.15			
测试内容	接口准确性测试	测试方法	接口单调	
测试工具	Postman native 4.0.8			
测试说明	使用 postman 针对每一个 restful 接口进行 get/post 请求，查看返回状态和内容是否与预期相符			
预置条件	单文件响应延迟测试，低于 100ms 为低延迟，100-300ms 为中等延迟，高于 300ms 为高延迟。总接收渲染响应测试，低于 2s 为低延迟，2-5s 为中等延迟，高于 5s 为高延迟			
测试方法	分别针对每一个测试接口，发送不同的请求数据，并与预期逻辑结果相比较，统计正确响应与错误响应情况的数量和比值			
测试内容	注册/登录模块下接口：2 个，热力追踪模块接口：4 个，灾害报表模块接口：4 个，任务中心模块接口：4 个，权限管理模块接口：2 个			
测试结果				
测试模块	测试接口	正确数量	错误/异常数量	正确率
注册/登录	登录接口	10	0	100%
	注册接口	10	0	100%
热力追踪	热力源数据获取接口	50	0	100%
	刷新速率调整接口	10	0	100%

	攻击源过滤接口	10	0	100%
	个性换肤接口	8	2	80%
灾害报表	数据源获取接口	50	0	100%
	攻击类型过滤接口	10	0	100%
	省市统计接口	10	0	100%
	查获数据统计接口	10	0	100%
任务中心	任务发布接口	10	0	100%
	任务调度反馈接口	10	0	100%
	回收任务查看接口	10	0	100%
	实时上报人次接口	10	0	100%
权限管理	管理员名单获取	10	0	100%
	权限修改接口	10	0	100%
测试评价	大部分接口响应正确，保证了系统功能的可用性和可靠性，唯一会出错的接口在于个性换肤上面，由于地图的皮肤功能同样来源于百度地图 API 接口，且不同样式配置需要兼容的地图 api 版本不同，因此多个样式就需要多个 Bmap 的版本，而百度官方由于没有考虑到这一点，就导致我的系统会出错，目前已知的改进措施是引用同一兼容版本的地图样式，但这样容易忽略界面美感。			

可移植性测试：

1、手机移动 app

客户端采用 Android Studio 进行开发，即发布后的 APP 只能为 Android 系统的手机能够下载安装。目前主流 Android 系统的手机品牌有：(1) 国内：HUAWEI、小米、OPPO、vivo、魅族（2）国外：三星、索尼等。该测试主要针对安装在不同品牌的下的客户端的界面是否正常以及能否实现正常功能，以此来确定客户端的可移植性与兼容性，测试用例及结果如表 5-23 所示。

表 5-23: 客户端可移植性测试用例

用例编号	NO.16			
测试内容	管理系统兼容性测试	测试方法	黑盒测试	
测试说明	针对不同品牌的手机，测试用户与不同模块的交互情况，统计各个品牌的手机各个模块的正常使用率，以及错误/异常情况			
预置条件	手机无损坏			
测试准则	查看每个界面是否显示正常，针对每个功能模块，判断其能够正常响应。			
测试结果				
浏览器	功能模块	功能使用情况	界面显示情况	错误分析
HUAWEI 荣耀系列	注册/登录	正常	正常	
	攻击源提醒	正常	正常	
	音频防御	正常	正常	
	任务中心	正常	正常	
	个人周报	正常	正常	
小米 note 系列	注册/登录	正常	正常	
	攻击源提醒	正常	正常	
	音频防御	正常	正常	
	任务中心	正常	正常	

	个人周报	正常	正常	
OPPO R11	注册/登录	正常	正常	
	攻击源提醒	正常	正常	
	音频防御	正常	正常	
	任务中心	正常	正常	
	个人周报	正常	正常	
vivo X27	注册/登录	正常	正常	
	攻击源提醒	正常	正常	
	音频防御	正常	正常	
	任务中心	正常	正常	
	个人周报	正常	正常	
三星 note 系列	注册/登录	正常	正常	
	攻击源提醒	正常	正常	
	音频防御	正常	画面一定程度挤压	手机分辨率与客户端开发时相差较大
	任务中心	正常	画面一定程度挤压	手机分辨率与客户端开发时相差较大
	个人周报	正常	画面一定程度挤压	手机分辨率与客户端开发时相差较大
测试评价	除了三星 note 系列手机在移植上客户端后在画面上与预期设计上有一定程度上的不符，其他品牌的手机在画面与功能上都能很好的兼容。在针对不同品牌手机的分辨率问题时用 Android Studio 平台开发的客户端能够尽可能的符合主流手机品牌。			

2、管理系统

管理系统使用的 B 端架构，目前主流的浏览器有 Chrome、Safari、Firefox、Internet Explorer 和 Microsoft Edge，该测试环节主要测试系统不同界面在不同浏览器的正常响应情况，以此来优化管理系统的可移植性和兼容性。测试用例及结果如表 5-24 所示。

表 5-24: 管理系统可移植性测试用例

用例编号	NO.17		
测试内容	管理系统兼容性测试	测试方法	黑盒测试
测试说明	针对每一个，测试用户与不同模块的交互情况，统计各个品牌的手机各个模块的正常使用率，以及错误/异常情况		
预置条件	无		
测试准则	针对每个功能模块，判断其能够正常响应，包括数据的获取、渲染以及用户的交互。		
测试结果			
浏览器	功能模块	响应情况	错误分析
Chrome	注册/登录	正常	
	热力追踪	正常	
	灾害报表	正常	
	任务中心	正常	
	权限管理	正常	

Safari	注册/登录	正常	
	热力追踪	正常	
	灾害报表	图片无法加载	前端图表库该版本还未得到支持
	任务中心	图片无法加载	前端图表库该版本还未得到支持
	权限管理	正常	
Firefox	注册/登录	正常	
	热力追踪	正常	
	灾害报表	正常	
	任务中心	正常	
	权限管理	正常	
Internet Explorer(9.0)	注册/登录	正常	
	热力追踪	无法动态更新数据	较早版本无法支持 ajax 异步加载
	灾害报表	无法动态更新数据	较早版本无法支持 ajax 异步加载
	任务中心	无法动态更新数据	较早版本无法支持 ajax 异步加载
	权限管理	正常	
Microsoft Edge	注册/登录	动画缺失	flash 插件关闭
	热力追踪	动画缺失	flash 插件关闭
	灾害报表	动画缺失	flash 插件关闭
	任务中心	动画缺失	flash 插件关闭
	权限管理	正常	
测试评价	目前仅有 Chrome 和 Firefox 浏览器能够完全正常运行系统，其余错误/异常信息分析如下：灾害报表和任务中心使用的是 AdminLTE 前端图表代码库，分析是由于版本才进行过一次大更新，暂无法得到 Safari 该版本下的支持；Internet Explorer 浏览器版本过低，且更新速度极为缓慢，ajax 异步加载技术暂无法使用，不过 10 版本以上的 IE 支持，考虑到使用 IE9 及以下的用户不多，暂未采取修复计划；网页 flash 是否会导致安全问题一直还未下定论，就连 google 都未给出明确回答，而微软旗下的 edge 浏览器目前设置 flash 默认关闭，导致一些基于 flash 插件的动画无法展现，因此，考虑将前端动画采用原生 CSS 及 JS 重构。		

5.6 安全性测试

本测试部分旨在测试系统在采用国密 SM2/SM3/SM4 加解密算法后，对数据传输的正确性、有效性进行测试，并评估该算法能否满足系统的安全性、可靠性以及资源低占用性。**SM3 算法测试：**

SM3 算法属于摘要算法，摘要算法最核心的特点就是无碰撞性、不可逆性，因此测试的内容也即是判断大量不同明文信息进行 SM3 摘要算法后，相同散列值占所有散列值的比重，比重为 0 说明算法安全。另外关于逆推测试由于机器性能有限，所以不再测试，但理论上已证实目前还未找到有效方法能够逆推出明文信息。测试样例如表 5-25 所示，测试单一样本结果如图 5-27 所示。

表 5-25: SM3 算法抗碰撞性测试

用例编号	NO.019		
测试名称	SM3 碰撞性测试	测试方法	黑盒测试

测试说明	通过脚本进行大量不同信息明文的摘要算法加密，并将原始信息和摘要保存为 CSV 格式，统计相同摘要占所有摘要的比重
预置条件	脚本预先生成 10 万行不同信息明文，长度为 4 字节
判断准则	测试 10 万种不同明文的散列值是否有相同样例，没有则代表算法安全，有说明算法不安全
测试输入	10 万行不同原始信息明文
测试输出	10 万行原始信息的摘要值
测试结果	循环测试 5 次，每次相同摘要值信息数量所占总摘要值数量比重均为 0
测试评价	测试用例通过，SM3 算法能够作为安全的摘要算法

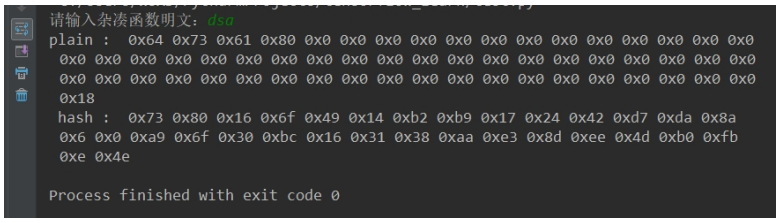


图 5-27: SM3 算法单一样本摘要计算结果

SM4 算法测试：

SM4 算法属于对称加密算法，该算法设计之初就是为了将用户的明文信息以密文的形式在公网上传输，防止非法用户窃听破解信息。该测试部分旨在测试用一个对称密钥加密信息，然后通过暴力破解的方式，看能否解密出原始用户信息。测试样例如表 5-26 所示，测试单一样本结果如图 5-28 所示。

表 5-26: SM4 算法暴力破解测试

用例编号	NO.19		
测试名称	SM4 暴力破解测试	测试方法	黑盒测试
测试说明	对信息明文 A 进行加密得到 B，采用 10 万组不同密钥进行暴力破解测试，看能否通过 B 得到 A		
预置条件	信息明文‘hello world’，SM4 加密后得到‘U2FsdGVkX1/d3AbA4dltAEbJJ8D+P5CyMnzLzczejns=’		
判断准则	10 种不同密钥进行解密，如果得到一条明文与‘hello world’相同，则算法不安全，否则，算法安全。		
测试输入	10 万行不同密钥		
测试输出	10 万行解密明文		
测试结果	10 行密钥对该加密信息进行解密，测试明文结果中有无与‘hello world’相同的字符串。		
测试评价	测试用例通过，SM4 算法能够作为安全的对称加密算法		

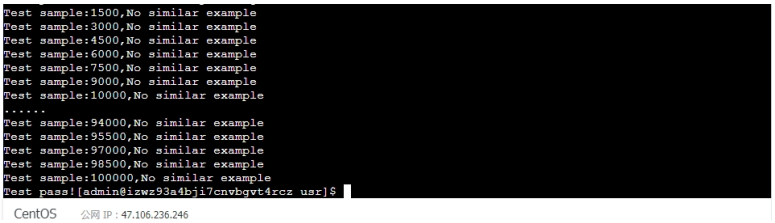


图 5-28: SM4 算法暴力破解测试

SM2 算法测试：

SM2 算法属于非对称密钥加解密算法，在安全通信体系中，有两个作用，一是用于验证发送方的身份唯一性，确保了信息来源的不可否认性，二是只有持有密钥的人才能解密与之对应的公钥加密的信息，保证信息无法被非法破解。因此该测试用力旨在测试算法的正确性，关于安全性测试，前面的 SM4 算法测试已经通过了暴力破解测试，且非对称解密难度高于对称解密难度好几个数量级，因此安全方面本环节不在测试。

表 5-27: SM2 算法准确性测试

用例编号	NO.19		
测试名称	SM2 准确性测试	测试方法	黑盒测试
测试说明	发送方为 A，接收方为 B，A 使用 A 的私钥加密一段数据，测试 B 能否正常解开，A 使用 B 的公钥加密一段数据，测试 B 能否正常解开。		
预置条件	两组非对称密钥		
判断准则	如果 B 能够解开两段信息，并与明文信息相同，则说明 SM2 算法满足正确性，否则不满足，且不安全。		
测试输入	两段信息明文，其中私钥加密“hello”，公钥加密“world”		
测试输出	两组输出，公钥解密信息为“hello”，私钥解密信息为“world”		
测试结果	两组测试均通过，能够正常解密出非对称算法加密的信息		
测试评价	测试用例通过，SM2 算法能够作为安全的非对称加密算法		

5.7 安全性分析

5.7.1 系统运行流程

系统运行流程如图 5-29 所示，针对普通用户而言，用户上传当前音频信息、地理位置等信息，服务器判断当前音频是否为攻击音频，如果是，返回警告消息，否则让用户手机按照正常流程执行语音指令。对于公安管理员，想服务器提交表单查询请求，包括全区域攻击音频热力分布，历史攻击、查获数据，群智感知任务的发布与查收，然后服务器返回相应结果。

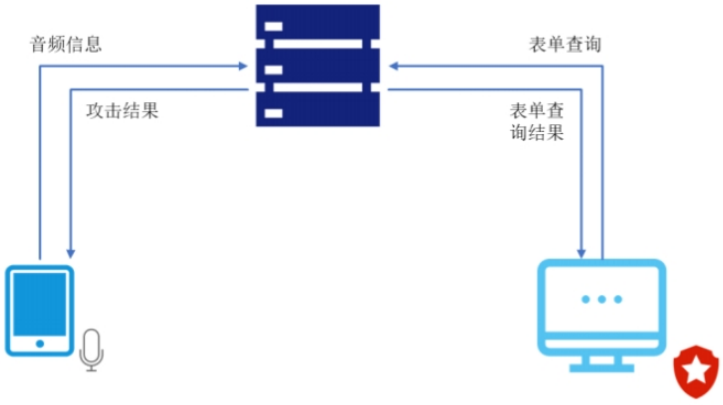


图 5-29: 系统运行交互流程

5.7.2 服务器认证

目前大多数用户浏览的网站都是经过 CA 认证的，因此在浏览百度、淘宝等其他大型网站时，都能看到地址栏前面有 https，在 chrome 浏览器可以查看该网站的证书。证书的作用就是确保用户目前所访问的网站是安全的，且身份已经过核实。一般来说，这些网站的证书是国际知名的 CA 认证公司颁布的，浏览器也内置了这些

公司颁布的证书解析器，用于核实证书是否安全。目前一些不法分子会利用伪造网站，和路由重定向这些互联网的漏洞，来实现劫持用户流量，当用户访问了这些网站，自身的信息安全也就得不到保障。而该系统实现了HTTPS 协议，因此任何被导向其余网站的访问都会提示连接不安全，用户得到这些提示后关闭页面即可。具体的关于服务器认证以及客户端核验证书流程如图 5-30 所示。

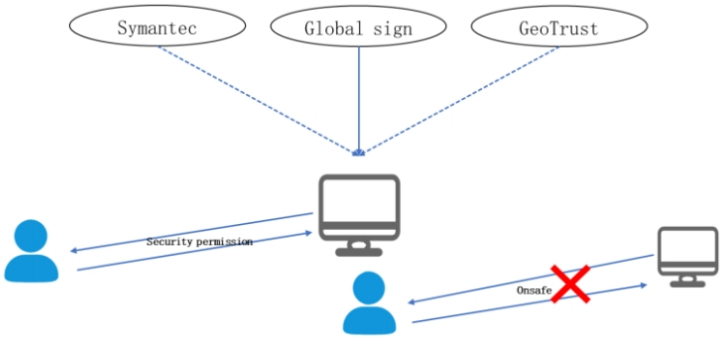


图 5-30: CA 证书认证即核查

图中经过了 CA 公司认证的网站即可被认为是安全，连接有保障的，未申请认证公司认证的，或者自己私自颁发证书的网站会认为是不安全的。另外，如何保障证书的安全也是该系统的关键，具体来讲，如果有人篡改证书，或者冒名顶替 CA 公司颁布证书，其实这些都会被客户端查获的。证书的颁布首先会经过摘要算法得到一个摘要值，然后对该明文信息和摘要值进行私钥加密，这样如果有人伪造证书，客户端用公钥就会解不开明文和摘要值，或者得到一个错误的摘要值，抑或当人非法篡改证书信息，此时当用明文再进行摘要计算，就会得到不一样的摘要值，如此，就能核实证书是否安全。

5.7.3 非法访问控制

非法访问是指通过扫描仪、黑客程序、隐蔽通道、远端操纵、密码攻击等窃取或截获用户名、口令，寻找网络安全性弱点，窃取超级用户权限，破解密码，窃取文件数据等。具体来讲，非法访问攻击切入点主要有：用户口令传输劫持，伪用户访问，数据库用户口令窃取。针对以上三种攻击形式，相应的系统也部署了应对措施，以防止非法用户进行非法访问操作。

用户口令传输劫持，通过使用国密安全通信加解密套件，使得正常用户在发出用户名和密码给远端服务器前就完成了用户名、口令的加密封装，使用 SM4 算法加密这部分信息，同时使用 SM2 算法为对称加密密钥再上一道锁，形成了二重保护，目前 SM2、SM4 是通过了国家密码安全局的认证，因此，该算法能够保证用户信息不被窃取、破解。

伪用户访问，具体的，客户端（手机 app、浏览器）部署人脸验证模块，注册环节采用手机验证码，登录环节采用动态口令，如此，可以确保用户端身份的真实有效性，为后续的数据传输安全奠定了基础。

数据库用户口令窃取，不仅在用户口令传输阶段得到了国密算法的加密，且在口令入库过程中，也进行了一次加密，使用到的是 AES 对称加密，并将密钥信息保存在另一台服务器，这样，即使黑客获取了服务器的用户信息，但因没有对称密钥，因此得到的也仅是一份乱码数据。

5.7.4 音频加密传输

音频数据的传输同样使用的是与用户口令传输安全等级最高的国密加密算法通信技术实现的。由于该系统中语音信息需要来回在客户端与服务端之间传输，因此，这里边需要关注两个问题，一是系统的实时性能否保障，二是用户语音隐私安全性。安全性之前已经提到过，使用 SM2 和 SM4 双重加密，能够保障用户隐私不被泄露。另一方面，使用 SM2/SM4 算法加密对时间效率的影响，首先 SM4 加密对象为音频数据，这一部分由于采用的是分组加密，因此使用采用并行分组加密，可以使时间复杂度降低到 O(n)，在对 5Mb 的音频进行加密，实

测所用时间为 0.47ms，相比传输所消耗的 100ms 级可忽略不及；同样，SM2 算法加密对象为对称密钥，由于密钥长度固定，因此加密所耗时间也为 10-1ms 级，可忽略不计。这样一来，既保证了音频数据的加密传输，又保证了时间效率。

5.8 小结

在本章，我们对作品进行了算法、功能、性能和安全测试与分析，采用了黑盒测试技术和规范的方法对本系统进行了测试，确保了本系统的各项功能正常运行。

第 6 章 创新与特色

内容提要

- 个人安全
- 企业保密
- 语音安全保障
- 图像识别、情感分析

随着嵌入式设备的应用越来越广泛，智能手机、智能家居与车载设备等一系列设备已经逐渐融入人们的生活，5g 技术与边缘计算技术的诞生，可以预见一个更加完善的物联网时代即将带来。物联网应用的信息技术——语音控制系统在各个领域都给日常生活带来了极大的便利，产生了一系列越来越好的经济、社会效益。然而，层出不穷的攻击也带给语音控制系统的安全问题新的考验。一旦系统被攻击者控制，带来的后果将是灾难性的。

攻击者可能对用户造成的损失和威胁包括不限于：①访问恶意网站；②实施人身监控；③泄露个人隐私和数据；④车载语音控制。本作品有效防止上述情况的发生，无论从个人安全还是企业保密方面都提供了极大地帮助，后续工作可以将其嵌入到现有语音识别助手中，有效降低因语音攻击而产生的损失。

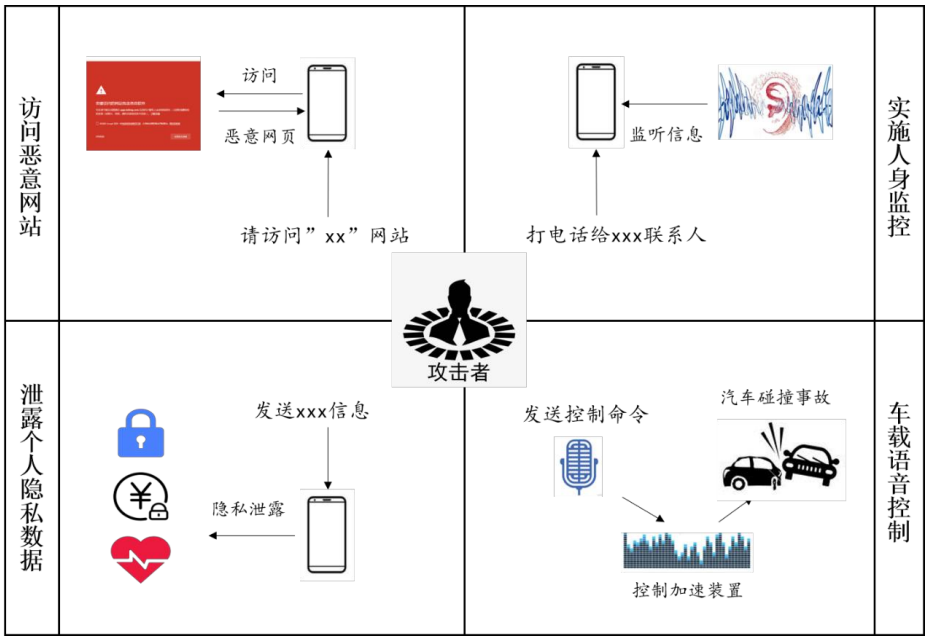


图 6-1: 语音攻击造成的损失和威胁

本系统提出了提出的恶意语音攻击识别分类算法，实现对于用户设备的语音控制系统的强力保护，并在此基础上提出的使用群智感知和态势识别的方法，借终端用户之手收集大量语音攻击信息，进而构造出一个语音攻击防御网，在源头上遏制语音攻击的蔓延。用热力图的形式可以直观地展现攻击源的分布情况，借助数据分析平台提供终端设备收到恶意语音攻击的轨迹。从一般用户使用层面，具有便捷性与低成本的特点。从管理员或者警务人员使用层面，能够对智能语音安全情况进行实时掌控，对智能语音安全风险进行及时管控。

同时，本作品中提出的分类过滤算法和群智感知获取大数据等技术不仅在语音安全保障领域大放光彩，而且也为图像识别、情感分析等领域提出了新的解决问题的思路。因此本作品具有广泛的应用前景。

本作品已入选国家级大学生创新创业计划，并正申请一项国家发明专利（队员为第一发明人）：

图 6-2: 本作品相关成果支撑

本作品已进行落地试点，并与相关公司建立合作关系，具有良好的实际应用价值：

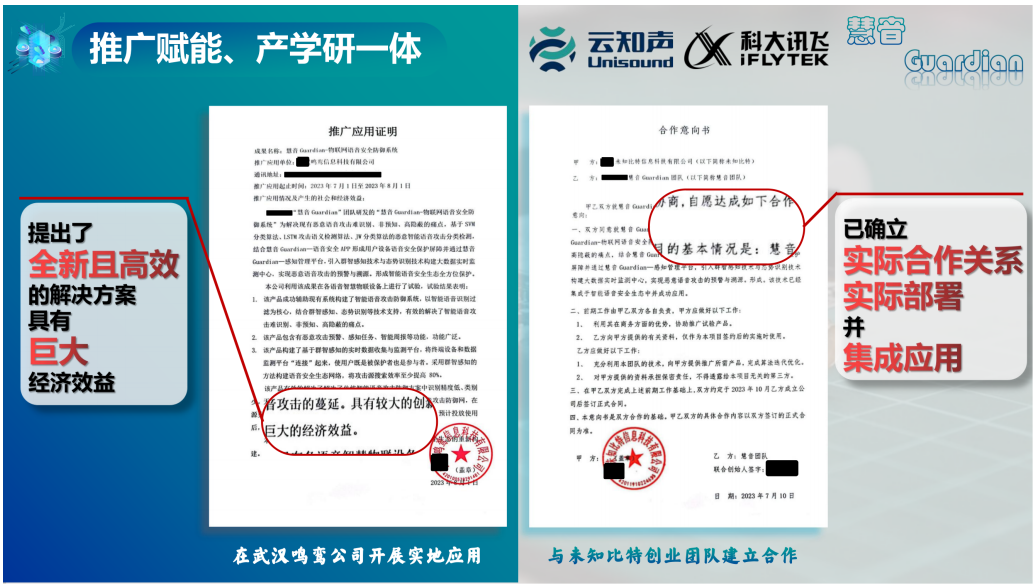


图 6-3: 本作品相关合作应用证明

参考文献

[1] Moustafa Alzantot, Bharathan Balaji, and Mani Srivastava. “Did you hear that? adversarial examples against automatic speech recognition”. In: *arXiv preprint arXiv:1801.00554* (2018).

[2] Nicholas Carlini and David Wagner. “Audio adversarial examples: Targeted attacks on speech-to-text”. In: *2018 IEEE security and privacy workshops (SPW)*. IEEE. 2018, pp. 1–7.

[3] Nicholas Carlini et al. “Hidden voice commands.” In: *Usenix security symposium*. 2016, pp. 513–530.

[4] Si Chen et al. “You can hear but you cannot steal: Defending against voice impersonation attacks on smartphones”. In: *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE. 2017, pp. 183–195.

[5] Frederica Darema. “Dynamic data driven applications systems: A new paradigm for application simulations and measurements”. In: *Computational Science-ICCS 2004: 4th International Conference, Kraków, Poland, June 6-9, 2004, Proceedings, Part III 4*. Springer. 2004, pp. 662–669.

[6] Huan Feng, Kassem Fawaz, and Kang G Shin. “Continuous authentication for voice assistants”. In: *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. 2017, pp. 343–355.

[7] Alex Graves et al. “Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks”. In: *Proceedings of the 23rd international conference on Machine learning*. 2006, pp. 369–376.

[8] Awni Hannun et al. “Deep speech: Scaling up end-to-end speech recognition”. In: *arXiv preprint arXiv:1412.5567* (2014).

[9] Ioannis Krontiris and Andreas Albers. “Monetary incentives in participatory sensing using multi-attributive auctions”. In: *International Journal of Parallel, Emergent and Distributed Systems* 27.4 (2012), pp. 317–336.

[10] Xinyu Lei et al. “The insecurity of home digital voice assistants–amazon alexa as a case study”. In: *arXiv preprint arXiv:1712.03327* (2017).

[11] Min Lin, Qiang Chen, and Shuicheng Yan. “Network in network”. In: *arXiv preprint arXiv:1312.4400* (2013).

[12] *Mel Frequency Cepstral Coefficient (MFCC) tutorial*. 2023. URL: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency%20cepstral-coefficients-mfccs/.

[13] Christopher Olah. “Understanding lstm networks”. In: (2015).

[14] Giuseppe Petracca et al. “Audroid: Preventing attacks on audio channels in mobile devices”. In: *Proceedings of the 31st Annual Computer Security Applications Conference*. 2015, pp. 181–190.

[15] Tavish Vaidya et al. “Cocaine noodles: exploiting the gap between human and machine speech recognition”. In: *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15)*. 2015.

[16] Guoming Zhang et al. “Dolphinattack: Inaudible voice commands”. In: *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*. 2017, pp. 103–117.

[17] 中商产业研究院数据库. 2023. URL: http://www.chnci.com/.

[18] 吴🄵 et al. “群智感知激励机制研究综述”. In: *Journal of Software* 27.8 (2016), pp. 2025–2047.

演示文档 PPT

对应页码：P87到 P103

2024 | 中国高校计算机大赛
C4 NETWORK TECHNOLOGY CHALLENGE
网络技术挑战赛
适应新变·激励创新·甄选英才·助力产业

慧音 Guardian

Smart Sound Guardian

物联网语音安全领航者

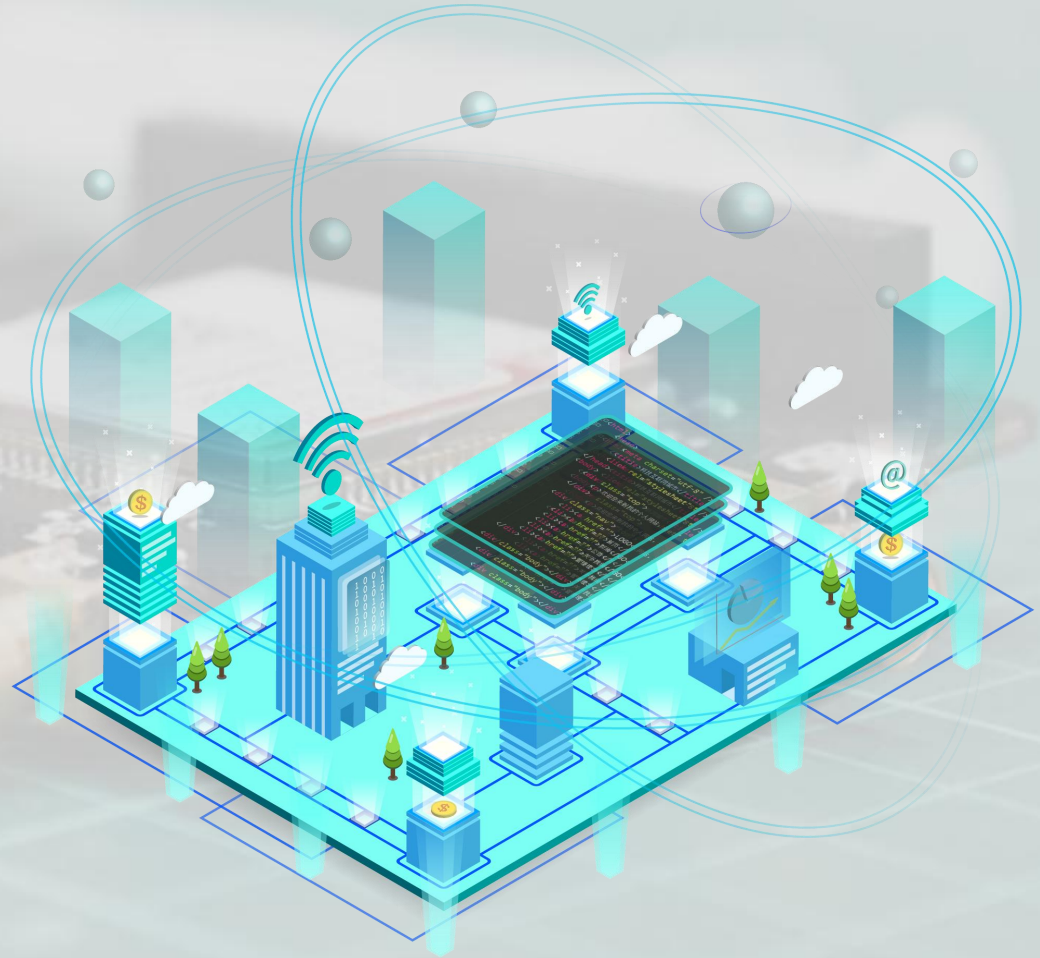
引言

5g通信技术

+

边缘计算技术

=



更完美的全新物联网时代

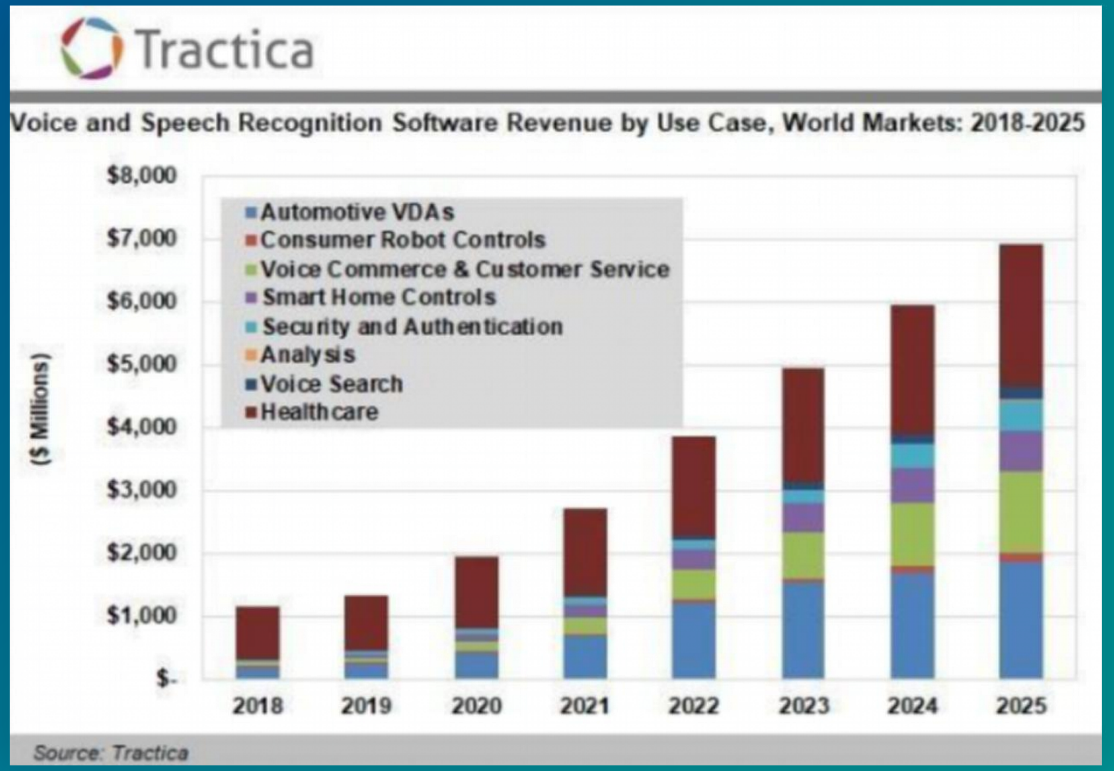
社会背景

智能语音市场规模逐年递增

[1] 中国智能语音市场行业报告

慧盲 Guardian

语音控制
在多个领域得到广泛应用



[2] *Voice/Speech Recognition Software Market Turning Up The Volume*

智能语音攻击

恶意攻击者

无声远程操控

超声波、光命令等

智能设备

语音控制系统

窃取个人隐私

通过远程操控移动智能设备，获取用户的**隐私文件**、**隐私记录**或其他高级个人隐私

造成财产损失

通过远程操控移动智能设备，使用**支付功能**或通讯功能，造成财产损失

威胁生命安全

通过远程操控**智能家居**、**智能车载系统**，威胁用户的生命安全

智能语音攻击的隐患

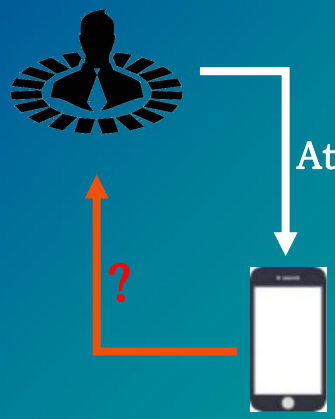
研究人员证明现有智能语音助手面对智能语音攻击测试

通过率为0%，无一幸免!

痛点分析

难识别

现有方案仅能对一种类型的攻击进行防御



非预知

现有方案无法对智能语音攻击进行及时预警

高隐蔽

现有方案无法对智能语音攻击态势进行预测溯源

解决方案 ——改进终端设备语音安全物联网（创新）

慧盾 Guardian

传统语音攻击防御系统

缺陷

- 识别精度低
- 无法提供预警
- 难溯攻击源

智能语音攻击检测模型检测分析

基于MCFF的多维语音特征提取算法，提取音频特征

基于软间隔支持向量机的音频分类算法，多重攻击音频精确分类

基于时序LSTM的攻击语义检测算法，指令内容层面实现保护

构建群智感知搜集监测平台

以用户为主体，参与感知任务，为语音安全生态持续升级

精准感知预警，区域攻击情况可视化

攻击态势识别预警提示

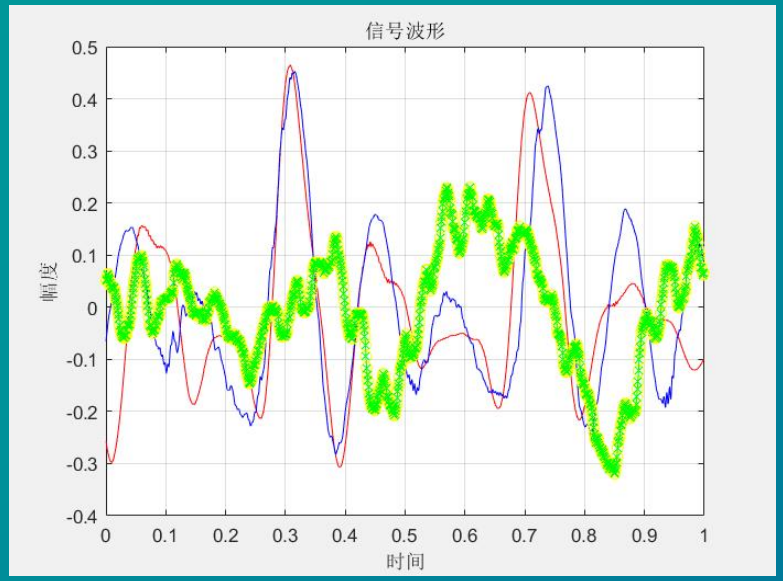
基于CAM映射和GRU的态势识别模型

智能语音态势可视化，助力警方精准攻击溯源

时域攻击数据一键掌控，开展合理措施

改进 终端设备语音安全物联网

攻击语音



① 由于谐波的自卷积，恶意音频在低频具有更高的能量比例——**频率**

② 当音频的能量总体增强时，恶意音频在低频部分的能量会随之增加，但是合法音频的低频能量增加很少——**能量**

③ 在时域上，恶意音频在正方向的振幅偏移量高于负方向振幅偏移量，而在合法语音中这两部分偏移量几乎相等——**波动性**

MFCC参数提取：

连续语音 → 预加重 → 分帧 → 加窗 → FFT → Mel滤波器组 → 对数运算 → DCT

短时平均过零率：

$$CZR = |sgn[x(n)] - sgn[x(n-1)]| * w(n)$$

用户音频

多维语音

常音频

声波攻击

指令攻击

器学习攻击

基于高

$s.t.\ y_n(w^T x_n$

$x) + b$

攻击语音检测过滤模型 ——模型效果

模型名称 $iLjF$	GPU上 运行时间$\bar{t}_G$	CPU上 运行时间$\bar{t}_C$	准确率 Acc
1L1F	32.9ms	79.3ms	58.3%
1L2F	35.8ms	82.6ms	74.1%
1L3F	36.5ms	88.9ms	72.5%
2L1F	48.0ms	106.6ms	80.8%
2L2F	49.0ms	116.0ms	84.2%
2L3F	**54.0ms**	126.4ms	**88.7%**

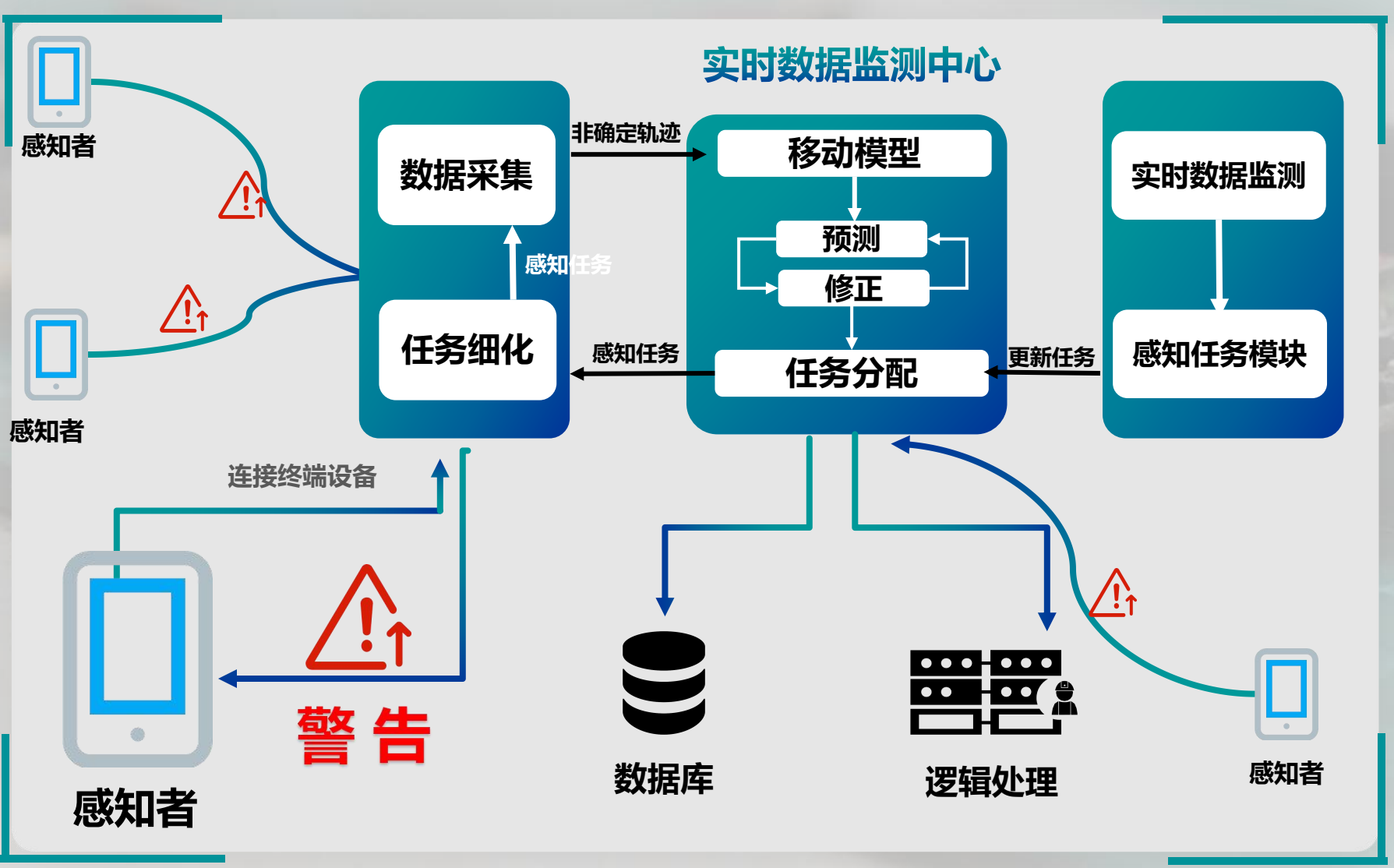
现阶段针对所有攻击的检测和防御

平均准确率为88.7%

准确率最高可达96.2%

慧音 Guardian

群智感知预警模式



语音攻击数据获取痛点：

数据不易搜集

涉及用户高级个人隐私

慧音系统采取的方案：

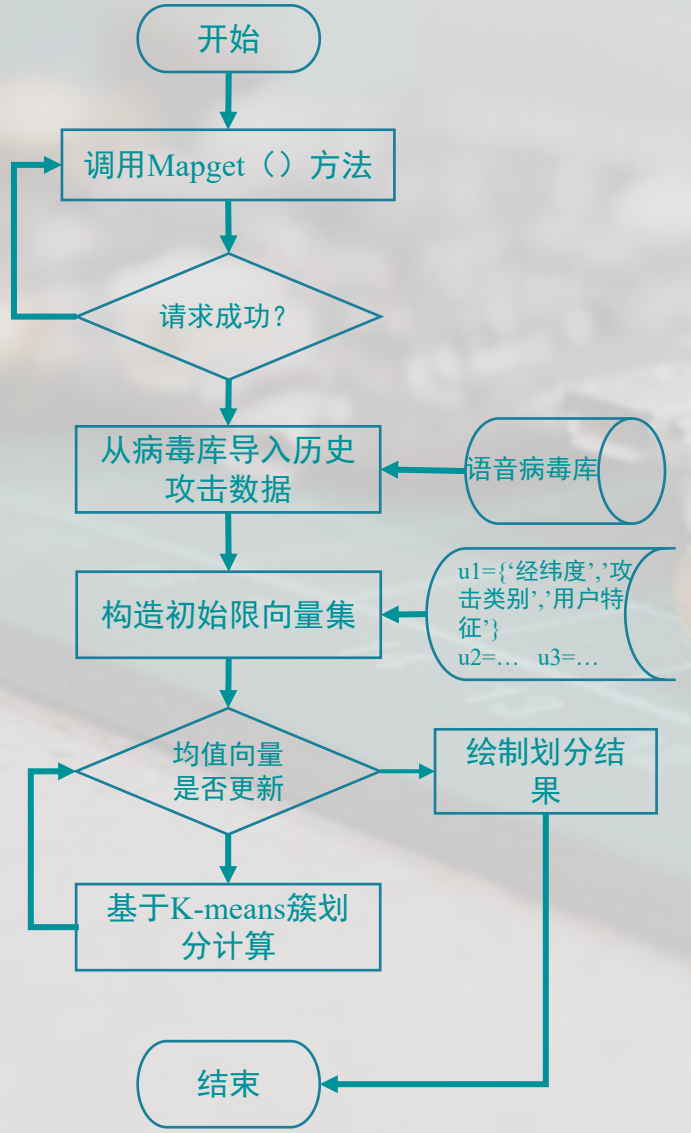
用户攻击数据本地储存
群智感知任务数据搜集
数据监测中心实时监测

智能语音安全生态全面升级！

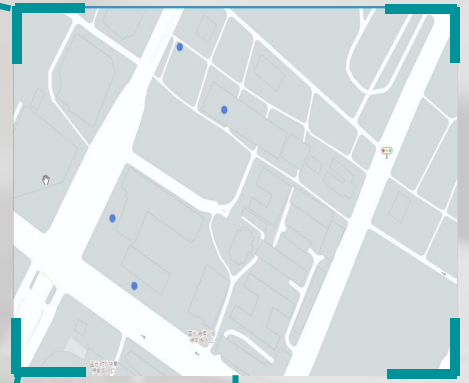
群智感知预警——感知预警流程（核心二）

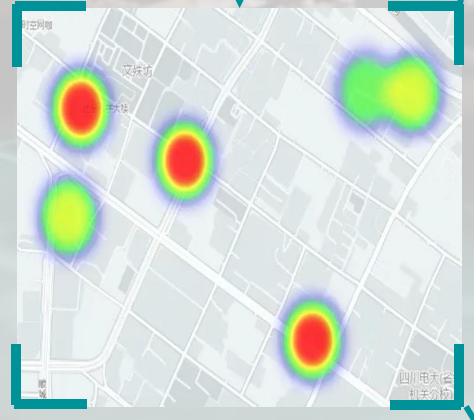
慧管 Guardian

攻击热力图实现流程图



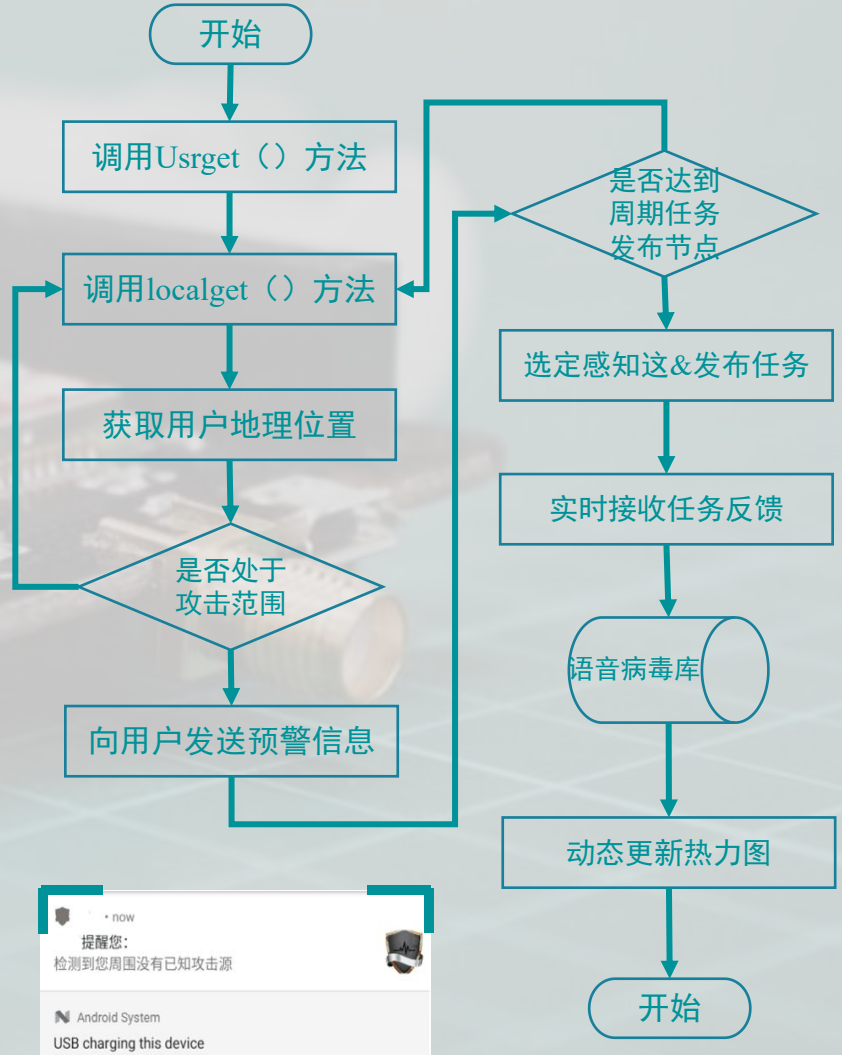
语音攻击热点图





语音攻击热力图

感知预警与任务分配流程图



实时提醒图

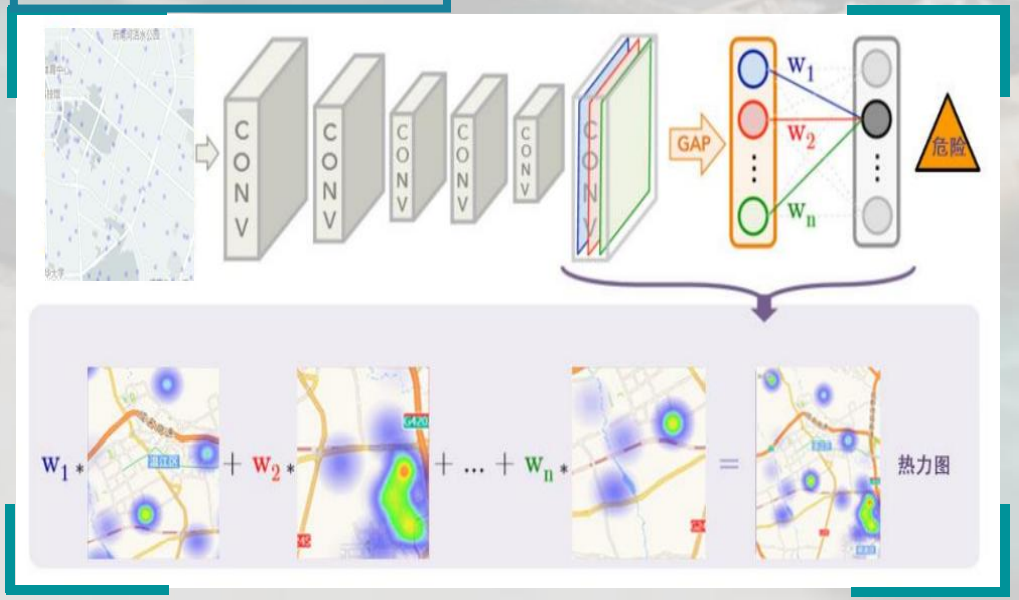
攻击态势识别预测模型——基于CAM映射与时序GRU（核心三）

慧眸 Guardian

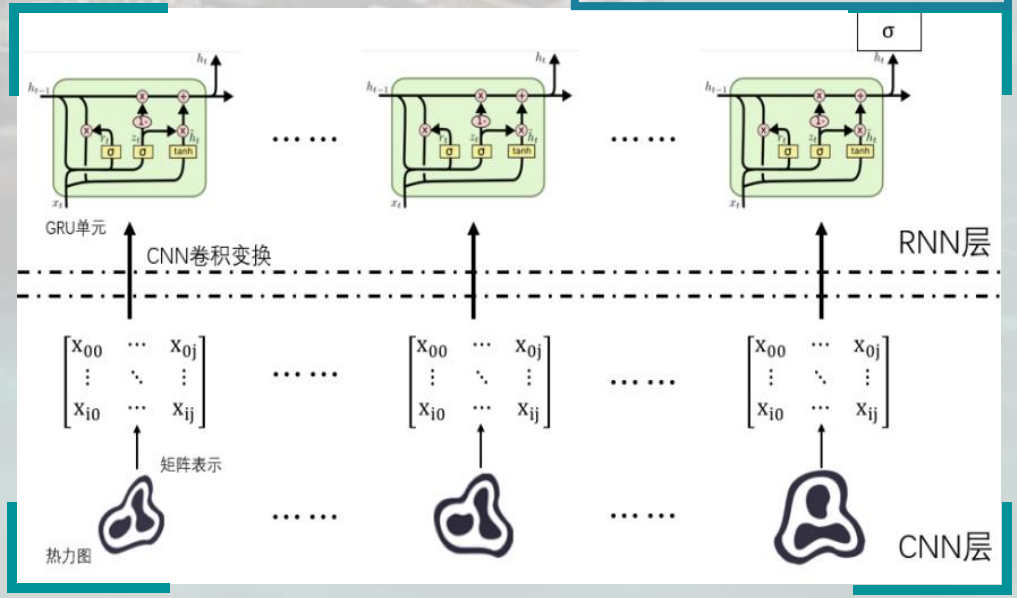
应用CAM映射将攻击关键区域标记，找到攻击的高危范围

使用基于GRU的时序模型，进行高危范围预测

类激活映射



GRU神经网络

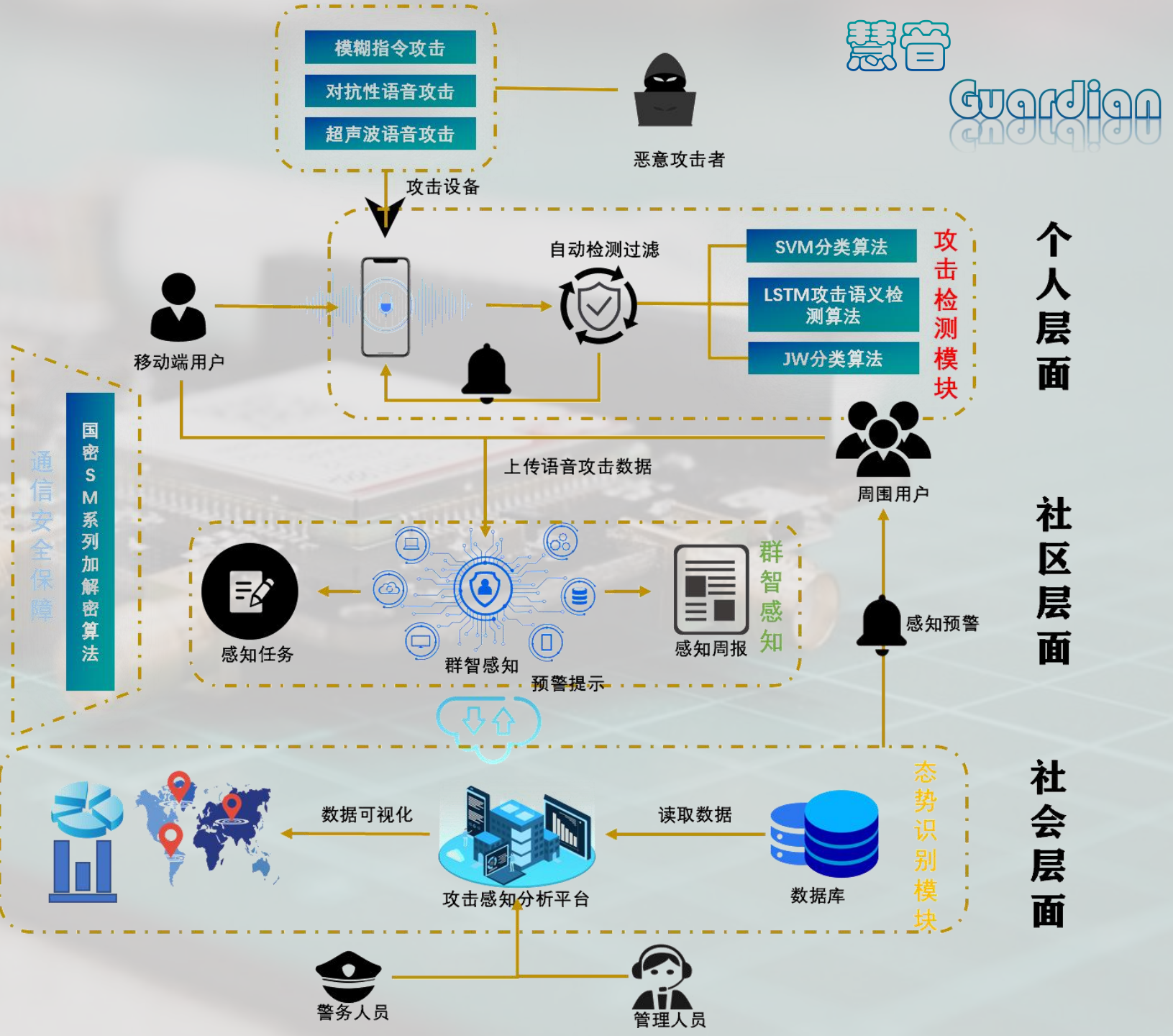


核心总结

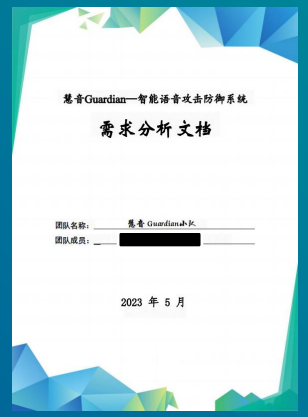
多层着手

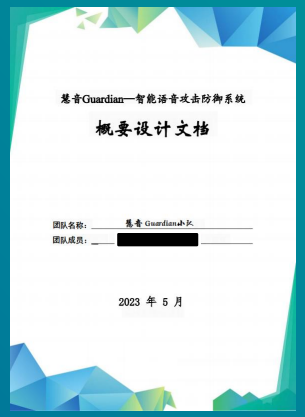
个人、社区、社会层面全覆盖

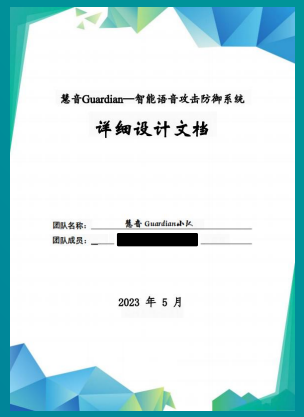
- **个人层面** 构建**语音攻击模拟场景**，对正常语音进行**两次分类并且过滤**，**准确判断**是哪一类语音攻击。
- **社区层面** 构建基于**群智感知**的**实时监测**大数据分析平台，形成**辐射保护**，并通过攻击趋势形成**溯源机制**
- **社会层面** 针对引入**态势感知技术**，构建基于GRU和CAM的**语音安全态势感知平台**，便于管理人员采取针对性措施。

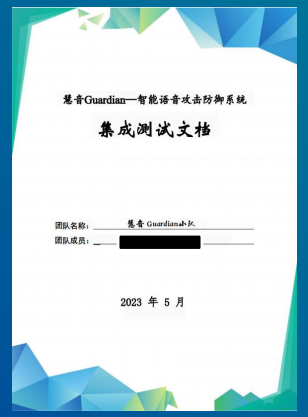


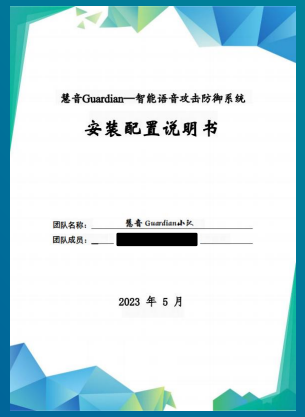
科教融合、成果斐然

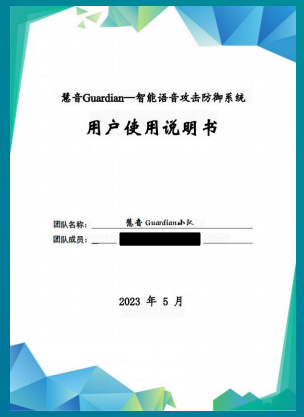












软件工程文档矩阵

入选**国家级**大学生创新创业计划

一项**国家发明专利**（准备申请，队员为第一发明人）

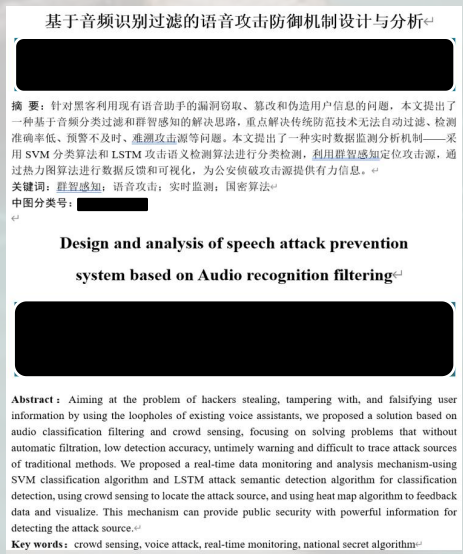
《智语安全中心app》

《智语感知检测平台》

CPCC 中国版权保护中心

两项软件著作權（队员为著作权人）

《基于音频识别过滤的语音攻击防御机制设计与分析》论文首页



基于音频识别过滤的语音攻击防御机制设计与分析

[redacted]

摘 要：针对黑客利用现有语音助手的漏洞窃取、篡改和伪造用户信息的问题，本文提出了一种基于音频分类过滤和群智感知的解决思路，重点解决传统防范技术无法自动过滤、检测准确率低、预警不及时、难溯攻击源等问题。本文提出了一种实时数据监测分析机制——采用 SVM 分类算法和 LSTM 攻击语义检测算法进行分类检测，利用群智感知定位攻击源，通过热力图算法进行数据反馈和可视化，为公安侦破攻击源提供有力信息。

关键词：群智感知；语音攻击；实时监测；国密算法

中图分类号：[redacted]

Design and analysis of speech attack prevention system based on Audio recognition filtering

[redacted]

Abstract：Aiming at the problem of hackers stealing, tampering with, and falsifying user information by using the loopholes of existing voice assistants, we proposed a solution based on audio classification filtering and crowd sensing, focusing on solving problems that without automatic filtration, low detection accuracy, untimely warning and difficult to trace attack sources of traditional methods. We proposed a real-time data monitoring and analysis mechanism-using SVM classification algorithm and LSTM attack semantic detection algorithm for classification detection, using crowd sensing to locate the attack source, and using heat map algorithm to feedback data and visualize. This mechanism can provide public security with powerful information for detecting the attack source.

Key words：crowd sensing, voice attack, real-time monitoring, national secret algorithm

CPCC 中国版权保护中心

一篇中文核心論文（在投，队员为论文一作）

推广赋能、产学研一体

云知声 Unisound　科大讯飞 iFLYTEK　慧音 Guardian

提出了
全新且高效
的解决方案
具有
巨大
经济效益

推广应用证明

成果名称：慧音 Guardian-物联网语音安全防御系统

推广应用单位：[redacted]鸣鸾信息科技有限公司

通讯地址：[redacted]

推广应用起止时间：2023 年 7 月 1 日至 2023 年 8 月 1 日

推广应用情况及产生的社会和经济效益：

[redacted]"慧音 Guardian"团队研发的"慧音 Guardian-物联网语音安全防御系统"为解决现有恶意语音攻击难识别、非预知、高隐蔽的痛点，基于 SVM 分类算法、LSTM 攻击语义检测算法、JW 分类算法的恶意智能语音攻击分类检测，结合慧音 Guardian—语音安全 APP 形成用户设备语音安全保护屏障并通过慧音 Guardian—感知管理平台，引入群智感知技术与态势识别技术构建大数据实时监测中心，实现恶意语音攻击的预警与溯源。形成智能语音安全生态全方位保护。

本公司利用该成果在各语音智慧物联设备上进行了试验，试验结果表明：

1. 该产品成功辅助现有系统构建了智能语音攻击防御系统，以智能语音识别过滤为核心，结合群智感知、态势识别等技术支持，有效的解决了智能语音攻击难识别、非预知、高隐蔽的痛点。
2. 该产品包含有恶意攻击预警、感知任务、智能周报等功能，功能广泛。
3. 该产品构建了基于群智感知的实时数据收集与监测平台，将终端设备和数据监测平台"连接"起来，使用户既是被保护者也是参与者。采用群智感知的方法构建语音安全生态网络，将攻击源搜索效率至少提高 80%。

该产品有效的解决了解决了传统智能语音攻击防御方案中识别精度低、类别少、[redacted]音攻击的蔓延。具有较大的创新[redacted]攻击防御网，在源[redacted]预计投放使用后，[redacted]巨大的经济效益。

[redacted]的重新构建。

[redacted]（盖章）

2023 年 [redacted] 日

在武汉鸣鸾公司开展实地应用

合作意向书

甲　方：[redacted]未知比特信息科技有限公司（以下简称未知比特）

乙　方：[redacted]慧音 Guardian 团队（以下简称慧音团队）

甲乙双方就慧音 Guardi[redacted]协商，自愿达成如下合作意向：

一、双方同意就慧音 Guar[redacted]Guardian-物联网语音安全[redacted]目的基本情况是：慧音[redacted]护屏障并通过慧音 Guardian—感知管理平台，引入群智感知技术与态势识别技术构建大数据实时监测中心，实现恶意语音攻击的预警与溯源。形成。该技术已经集成于智能语音安全生态中并成功应用。

二、前期工作由甲乙双方各自负责，甲方应做好以下工作：

1、　利用其在商务方面的优势，协助推广试验产品。

2、　乙方向甲方提供的有关资料，仅作为本项目签约后的实施时使用。

乙方应做好以下工作：

1、　充分利用本团队的技术，向甲方提供推广所需产品，完成算法迭代优化。

2、　对甲方提供的资料承担保密责任，不得透露给本项目无关的第三方。

三、在甲乙双方完成上述前期工作基础上，双方约定于 2023 年 10 月乙方成立公司后签订正式合同。

四、本意向书是双方合作的基础。甲乙双方的具体合作内容以双方签订的正式合同为准。

甲　方：[redacted]

乙　方：慧音团队

联合创始人签字：[redacted]

日　期：2023 年 7 月 10 日

与未知比特创业团队建立合作

已确立
实际合作关系
实际部署
并
集成应用

亮点总结

慧眸 Guardian

多维防护

精度高、层面多、范围广

- 智能语音攻击音频分类算法
  多特征分析音频信号层面
  后台自动持续识别过滤
- 智能语音攻击语义检测算法
  深度检测音频指令内容层面
  双层面保障设备语音安全
- 智能语音攻击检测过滤模型
  针对主流智能语音攻击实现
  精确防御，识别准确率平均
  可达88.7%，最高可达96.2%

实时高效

群智联合、预警快速、溯源高效

- 设计语音安全群智感知模式
  终端与检测中心紧密链接
  加快攻击音频数据获取效率
- 用户攻击数据反馈分析
  区域攻击精确实时预警
  二次恶意攻击成功率降低90%
- 实现攻击热力用户追踪
  攻击源搜索效率提高80%
  短时间精准锁定攻击源头

智能安全

预测准确、数据可视、传输可靠

- 语音攻击态势识别协助
  态势时序深度学习模型刻画
  态势预测准确率达到78%
- 可视化大数据分析与检测平台
  实时显示区域总体攻击形势
  大幅提升有效决策制定效率
- 国密算法保障数据通信安全
  提高音频数据可靠传输效率
  安全体系持续保护

2024 | 中国高校计算机大赛
C4-NETWORK TECHNOLOGY CHALLENGE
网络技术挑战赛
适应新变·激励创新·甄选英才·助力产业

慧眸 Guardian

物联网语音安全领航者

为您的设备语音安全保驾护航