软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007181] [CSTR: 32375.14.jos.007181]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于声感知的移动终端身份认证综述*

周 满[1,2,3], 李向前[1,2,3], 王 骞[4], 李 琦[5], 沈 超[6], 周雨庭[1,2,3]

[1](分布式系统安全湖北省重点实验室 (华中科技大学), 湖北 武汉 430074)
[2](湖北省大数据安全工程技术研究中心 (华中科技大学), 湖北 武汉 430074)
[3](华中科技大学 网络空间安全学院, 湖北 武汉 430074)
[4](武汉大学 国家网络安全学院, 湖北 武汉 430072)
[5](清华大学 网络科学与网络空间研究院, 北京 100084)
[6](西安交通大学 网络空间安全学院, 陕西 西安 710049)
通信作者: 周满, E-mail: zhouman@hust.edu.cn

摘 要: 随着移动终端的普及和用户隐私数据保护需求的增强, 基于移动终端的身份认证研究引起了广泛关注. 近年来, 移动终端的音频传感器为设计性能优良的新颖身份认证方案提供了更大的灵活性和可拓展性. 在调研了大量相关科研文献的基础上, 首先按照依赖凭据和感知方法的不同将基于声感知的移动终端身份认证方案进行分类, 并描述相应的攻击模型; 然后梳理移动终端基于不同认证凭据和基于声感知的身份认证国内外研究进展, 并进行分析、总结和对比; 最后结合当前研究的困难和不足, 给出衡量身份认证系统性能的两大指标 (安全性和实用性), 对未来的研究方向进行展望.

关键词: 身份认证; 声感知; 移动终端; 安全性; 认证凭据
中图法分类号: TP309

中文引用格式: 周满, 李向前, 王骞, 李琦, 沈超, 周雨庭. 基于声感知的移动终端身份认证综述. 软件学报. http://www.jos.org.cn/1000-9825/7181.htm
英文引用格式: Zhou M, Li XQ, Wang Q, Li Q, Shen C, Zhou YT. Survey on Acoustic Sensing-based User Authentication on Mobile Devices. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7181.htm

Survey on Acoustic Sensing-based User Authentication on Mobile Devices

ZHOU Man[1,2,3], LI Xiang-Qian[1,2,3], WANG Qian[4], LI Qi[5], SHEN Chao[6], ZHOU Yu-Ting[1,2,3]

[1](Hubei Key Laboratory of Distributed System Security (Huazhong University of Science and Technology), Wuhan 430074, China)
[2](Hubei Engineering Research Center on Big Data Security (Huazhong University of Science and Technology), Wuhan 430074, China)
[3](School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)
[4](School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China)
[5](Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China)
[6](School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: With the popularity of mobile devices and the enhancement of users' requirements for privacy protection, studies of user authentication on mobile devices have attracted widespread attention. Recently, the audio infrastructures of mobile devices have provided greater flexibility and scalability for the design of novel user authentication schemes with excellent performance. After surveying a large number of related works, this study first classifies acoustic sensing-based user authentication schemes on mobile devices according to the difference in authentication metrics and sensing methods and describes the corresponding attack model. Then, it analyzes and compares single authentication metric-based and acoustic sensing-based user authentication schemes on mobile devices. Finally, combined with the

* 基金项目: 国家自然科学基金 (62202180, U20B2049, U21B2018, 62132011)
收稿时间: 2022-12-26; 修改时间: 2023-07-17; 采用时间: 2024-03-19; jos 在线出版时间: 2024-11-06

problems of existing works, this study gives two metrics (security and practicability) to measure the performance of the user authentication system and discuss future research directions.

Key words: user authentication; acoustic sensing; mobile device; security; authentication metric

1 引 言

近年来, 随着移动互联网和集成电路技术的快速发展, 以智能手机为代表的移动终端的普及率越来越高, 给人们的生产生活方式带来了巨大变革. 根据爱立信 2022 年移动终端调查报告显示, 截至 2021 年底, 移动终端总订阅量高达 82 亿, 其中智能手机的订阅量占比为 77%, 达到 63 亿[1]. 人们可以随时随地进行网络冲浪、电子商务、在线导航、社交平台、线上支付等丰富应用, 享受着移动终端带来的巨大便利. 然而, 在数字生活日益丰富和繁荣的同时, 信息安全问题也变得愈加严重. 近年来, 有关用户数据泄露的事件屡见不鲜, 严重损害用户的财物安全和个人隐私. 安全可靠的身份认证机制作为移动终端抵御非法访问最关键的防线, 在保护用户数据安全方面发挥着巨大的作用.

移动终端通常要求用户向身份认证系统提供指定类型的身份认证凭据, 验证通过后方可解锁设备或登录 APP, 获取相应权限. 而攻击者往往尝试采取各种手段直接攻击或绕开身份认证系统, 达到非法访问的目的. 例如, 在基于秘密知识的身份认证系统中, 最直接的攻击手段是肩窥合法用户的身份认证过程[2], 窃取 PIN 码、手势口令等不同类型的身份认证凭据. 除此之外, 还有更先进的攻击手段, 如攻击者在用户进行击键或按压触摸屏等操作时, 从运动传感器 (加速度计, 陀螺仪, 磁力计)[3], 无线设备 (WiFi, 蓝牙, NFC)[4], 音频设备 (扬声器, 麦克风)[5]等不同类型传感器中获取实时数据, 结合机器学习算法, 推断出用户的手部运动, 推断出一系列的候选 PIN 码或手势口令. 基于生物特征的身份认证系统也遭受着严重的安全威胁. 随着社交网络的普及和深度伪造技术的发展, 各种生物特征 (例如人脸和指纹) 的伪造成本越来越低. 而现有的人脸认证和指纹认证方法往往难以区分真实特征和深度伪造特征, 导致它们容易遭受演示攻击. 例如, 攻击者通过 3D 打印技术伪造用户人脸的硅胶面具, 可以欺骗大量人脸认证和活体检测系统[6], 攻击者利用高精度假手指会绕过大量商用移动终端指纹锁的防欺骗功能[7]. 面对日益严峻的安全威胁, 进一步提高身份认证安全性, 防止移动终端被越权访问, 是亟待解决的现实问题.

随着移动终端内嵌传感器精度的提升和种类的多样化, 身份认证系统的安全强度和可用性持续提升. 普遍存在于移动终端的传感器为身份认证系统的设计提供了更大的灵活性和可拓展性. 近年来, 各式各样新颖的移动终端身份认证方案层出不穷[8]. 值得关注的是, 几乎所有移动终端都配备扬声器和麦克风, 并且这些音频设备的制造工艺不断完善, 性能不断优化, 支持更高的采样率, 能够记录和播放高质量音频. 而基于声信号的感知技术能够为用户提供识别追踪、健康监护、安全与隐私保护、人机交互、定位导航等诸多功能, 具有很高的应用价值[9].

声信号具有高度普适性和低硬件成本[10], 利用声信号感知用户身份认证过程中的独特隐性特征, 实现认证实体与凭据的安全绑定, 将极大地提高攻击者伪造身份认证凭据的难度和成本, 可以有效提高移动终端身份认证系统的安全性. 目前, 基于声感知的身份认证安全性增强研究已经得到了科研人员的广泛关注, 研究成果逐渐丰富. 该技术对保护用户数据安全和隐私具有重要的意义, 所以将国内外相关研究成果进行系统的分类梳理十分必要. 然而, 根据调研发现, 目前没有基于声感知的移动终端身份认证研究综述. 因此, 本文从内容上涵盖当前最新的研究成果, 探索该领域在未来的发展趋势. 基于该认识, 本文首先根据所依赖的身份凭据类型的不同, 将移动终端身份认证方法分为基于秘密知识和生物特征的身份认证系统, 以及双/多因素身份认证系统. 在此基础上, 本文将基于声感知的移动终端身份认证细粒度地分为基于声感知的典型身份认证、双因素身份认证和认证活体检测, 然后给出了基于声感知的移动终端身份认证的常见攻击模型, 并分别阐述了这些攻击的实现方法和危害. 接着, 按照前文的分类, 本文对移动终端基于不同认证凭据的身份认证国内外研究进展进行分析、总结和对比, 并对移动终端基于声感知的身份认证国内外研究进展进行分析、总结和对比, 进一步在技术实现、误判率、抵抗攻击类型等关键指标上横向对比各研究成果, 突出展现其性能和优缺点. 最后, 本文就当前研究推进的难点和存在的不足, 总结了移动终端安全身份认证研究面临的 3 大挑战, 并从安全性和实用性这两个角度分析各类型身份认证系统的发展

趋势. 本文认为利用声信号提取身份认证凭据, 构建双/多因素身份认证系统, 将在未来成为该领域研究的主流趋势.

本文第 2 节介绍基于移动终端的身份认证分类和基于声感知的身份认证分类, 并给出针对基于声感知的身份认证攻击模型. 第 3 节和第 4 节分别按照前文的分类标准, 对移动终端身份认证和基于声感知的身份认证国内外研究进展进行分析、总结和对比. 第 5 节归纳当前面临的研究挑战和发展趋势. 第 6 节对本文内容进行简要总结.

2 研究概述

基于声感知的身份认证主要利用移动终端上广泛配备的扬声器麦克风捕获用户的独特身份信息进行认证. 本节首先根据认证的基本要素介绍移动终端身份认证方法的分类, 然后回顾了声感知技术, 对基于声感知的身份认证方法进行分类, 最后介绍了基于声感知的身份认证攻击模型.

2.1 移动终端身份认证分类

回顾发展历程, 根据依赖的身份凭据的不同, 可以将移动终端身份认证的基本要素分为以下 3 种: (1) 秘密知识: 即用户所知道的信息, 例如口令、PIN 码; (2) 生物特征: 即用户独特的身体特征, 例如人脸、指纹; (3) 信任器件: 即移动终端具有独特物理特性的自带器件, 例如加速度计、陀螺仪、扬声器、麦克风、摄像头、屏幕等. 目前, 移动终端自带信任器件的物理特征只能支持小规模用户认证, 很少单独用于移动终端身份认证, 通常与其他身份认证的基本要素结合进行双/多因素身份认证, 如图 1 所示.

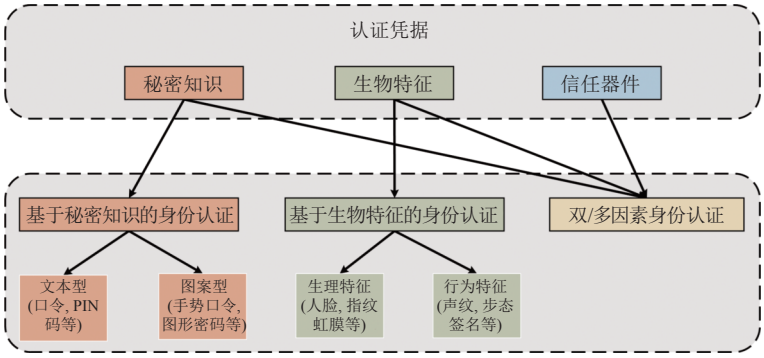


图 1 基于移动终端的身份认证

基于秘密知识的身份认证 (knowledge-based authentication, KBA) 长期以来被广泛地应用于各种移动终端, 拥有庞大的用户群, 主要依靠只在合法用户和身份认证系统之间共享的秘密知识进行身份认证. 传统的 KBA 系统按照秘密知识的类型, 可以分为以 PIN 码和口令为代表的文本型和以手势口令和图形口令为代表的图案型. 虽然 KBA 被广泛应用, 但是仍然存在一些问题. 根据调查数据显示, 大多数用户更喜欢使用生日, 电话, 名字等个人信息作为秘密知识的主体, 当攻击者掌握一些用户个人信息时, 破解秘密知识的成功率会大大提高[11]; 但是, 当用户使用随机性更强的秘密知识时, 又会增加记忆负担, 降低实用性.

生物特征分为生理特征和行为特征. 基于生理特征的身份认证利用人脸、指纹、虹膜等生理特征进行认证, 在安全性和实用性方面得到了一定程度地提升. 但是, 高精度生理特征数据的获取对硬件的要求更加严苛, 并且部分类型生理特征的采集需要额外的特殊硬件支持, 这无疑增加了成本, 降低了实用性. 此外, 由于生理特征不可改变, 一旦攻击者获取用户的生理特征, 将导致此类型身份认证方法不再安全. 基于行为特征的身份认证大多选取声纹、步态、触控手势等行为特征进行身份认证, 可以利用低值传感器进行特征采集. 但是, 移动终端内嵌的传感器对于除声纹外的其他行为特征的提取精度有限, 并且现有认证方案存在局限性, 导致准确度不够, 鲁棒性不强, 用户体验不够友好.

双/多因素身份认证方法采用两个或多个身份凭据来对用户进行认证, 相较于单一凭据的身份认证方法在安

全性上有所提高. 但是目前大多数双/多因素身份认证方法只是简单地对多个身份凭据逐一认证 (例如先要求用户输入 PIN 码, 然后再进行面部或指纹认证), 认证过程繁琐, 实用性降低. 多因素有机融合的身份认证方法已经成为研究趋势. 例如, 在提取语音声纹特征的同时捕获唇部运动特征[12]; 在输入手势口令的同时捕获手指在触摸屏上的压力、方向、速度等行为特征[13]. 只有多个认证因素同时满足要求时, 才可能通过认证. 多因素的有机融合使得安全性进一步提升, 并且用户不需要进行额外操作, 实用性高.

2.2 基于声感知的身份认证分类

声信号作为一种独特的物理信号, 由于它在短距离内具有良好的反弹特性和强抗干扰能力, 正在逐渐被关注和利用起来. 相比于其他物理信号, 声信号更容易获取, 且无需额外成本. 此外, 在声感知领域还有扎实的理论基础和相对成熟的技术支持, 能够充分提取声信号特征. 例如基于到达时间 (time-of-arrival, ToA) 的感知技术利用传输时间测量声信号在精准同步的发射器和接收器之间的距离; 在不同步的情况下, 基于调频连续波 (frequency modulated continuous wave, FMCW) 的感知技术利用啁啾信号的几何关系得出发射和接收信号之间的频率差, 从而量化传输时间; 基于信号强度 (signal strength)、相位变化 (phase change) 和多普勒频移 (Doppler shift) 的感知技术, 可以追踪物体 (例如用户手指) 的运动, 以用于各种传感应用. 经过长时间探索, 许多研究已经表明, 基于移动终端的声感知已在各种应用领域显示出优越性能. 如图 2 所示, 这些应用主要集中于安全与隐私、活动识别与跟踪、定位与导航、近距离通信等领域.

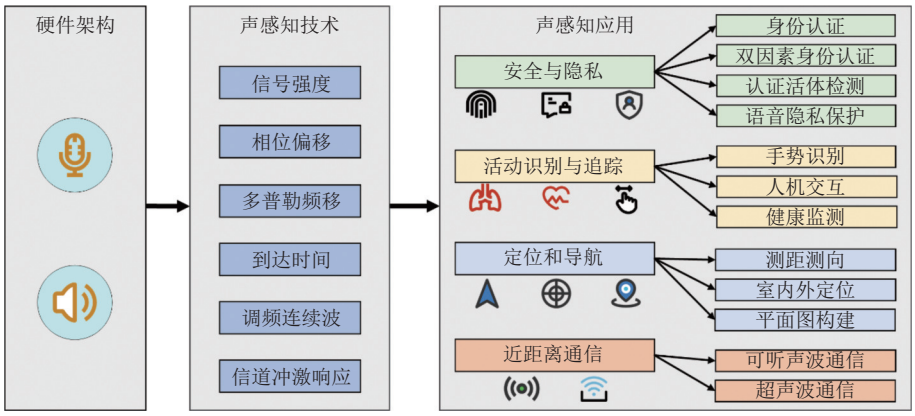


图 2 基于声感知技术的应用

基于声感知的移动终端身份认证方法利用声信号感知秘密知识、生物特征等认证凭据进行身份认证, 主要包括典型身份认证、双因素身份认证和认证活体检测, 如图 3 所示. 基于声感知的典型身份认证系统主要分为两类: 第一种类型是被动声感知身份认证, 即认证实体发射特征声信号. 例如, BreathPrint[14]从 3 个级别的人类呼吸声中提取声学特征, 即嗅探、正常呼吸和深呼吸, 用于用户身份验证; 第 2 种类型是主动声感知身份认证, 即认证系统发射提取用户身份特征的声信号. 例如, EchoPrint[15]利用内嵌的音频设备发射探测超声波, 经面部反射后, 提取回声的声学特征能够反馈面部形状, 为身份认证系统提供生理认证凭据. 基于声感知身份认证的另一大研究分支是基于声感知的双因素身份认证 (two-factor authentication, 2FA). 基于声感知的部分提取双因素身份认证方法就是向认证系统额外添加声感知的认证指标 (每种指标可以是秘密知识、生物特征, 或者信任器件). 例如, Soundproof[16]使用秘密知识作为第 1 认证因子, 声感知的注册设备和登录设备接近程度作为第 2 认证因子. 它利用设备麦克风录制环境噪音, 分别提取注册设备和登录设备所录制音频的声学特征, 计算相似度, 提供接近程度凭据. 基于声感知的一体化提取双因素身份认证方法利用声信号能够同时提取用户的多个身份认证因素. 与传统的多因素身份验证方法相比, 这种方法没有额外硬件成本, 更加方便, 用户不需要进行额外操作, 体验更好. 基于声感知的认证活体检测主要分为被动声感知活体因素和主动声感知活体因素. 第 1 种类型利用真实用户发射的声信号和扬声器发射的声信号之间的区别鉴别认证对象是否为活体. 例如, VoicePop[17]利用用户在靠近麦克风说话时呼吸产生

的气爆音鉴别认证对象是否为真实活体. 第 2 种类型利用真实用户反射的声信号鉴别认证对象是否为活体. 例如, VoiceGesture[18]通过测量反射声信号的多普勒频移来识别认证实体的发声姿势, 从而区分真实用户的声音和扬声器发出的声音.

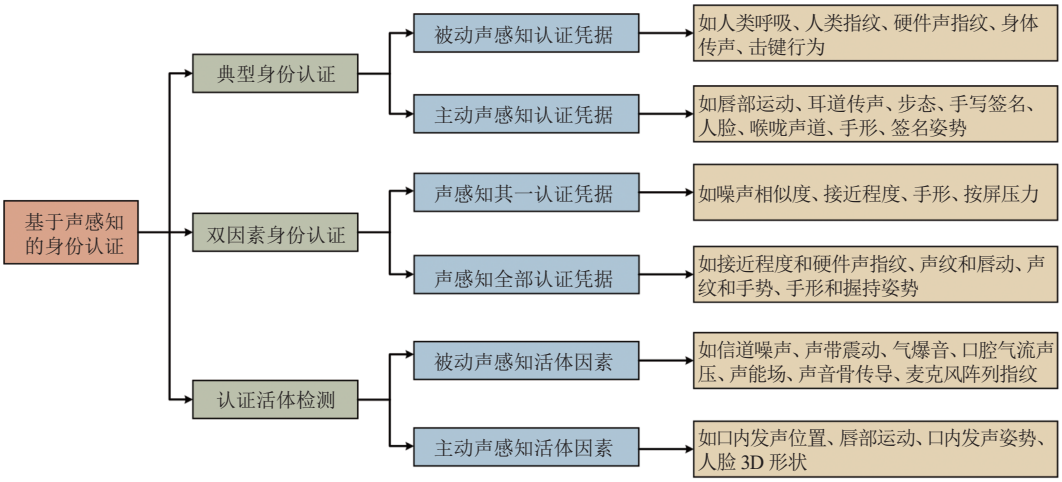


图 3　基于声感知的身份认证

相较于传统的身份认证方法, 基于声感知的身份认证方法有着独特的优势. 基于秘密知识的身份认证容易受到肩窥攻击、智能口令猜测等攻击. 而基于声感知的身份认证可以提取秘密知识输入过程中的独特隐性特征来增大秘密空间, 提高认证系统的安全性和鲁棒性, 并且不需要额外操作. 基于生物特征的身份认证一般需要特定传感器, 成本较高. 声学传感器具有普适低成本特性, 绝大多数移动终端自带声学传感器, 这为基于生物特征的身份认证设计提供了更大的灵活性. 此外, 基于生物特征的身份认证容易受到深度伪造攻击的威胁, 可以通过声感知的方法提供多种活体检测方法. 由于声学信道的开放性, 基于声感知的身份认证系统也面临着许多安全问题. 例如, 攻击者可以发动重放攻击, 以欺骗身份认证系统. 对于语音识别控制系统, 攻击者可能会发动对抗样本攻击, 例如将恶意的语音命令嵌入正常播放的音频中, 在受害者无法感知的情况下攻击语音助手系统[19], 这样类型的攻击是极易传播, 危害巨大; 攻击者还可以利用音频设备监听键盘击键声音, 以定位击键区域[20], 或推测设备解锁模式[21]. 目前基于声感知的移动终端身份认证安全性增强已成为一大研究方向.

2.3　基于声感知的认证攻击模型

由于声学信道的开放性和音频设备的易得性, 一般基于声感知的身份认证极易遭受攻击. 目前, 部署最为广泛的声感知认证方案是声纹认证. Shirvanian 等人[22]对攻击者模仿目标用户说话的模仿攻击进行了深入研究, 对部署在几个最先进的 Android 和 iOS 应用程序中的声纹认证系统分别进行了语音重放攻击和合成攻击, 均展现出较高的攻击成功率. 随着对抗神经网络的发展, 语音对抗样本攻击作为一种新型攻击手段, 被证明能够以隐蔽的方式重构语音并欺骗系统做出错误行为, 对现有的语音识别控制系统具有极大的危害[23]. Web 登录场景下 2FA 容易遭受中间人攻击 (man-in-the-middle attack, MiM) 和同位攻击 (co-located attack)[16]. 本节将对不同攻击类型逐个进行阐述.

重放攻击是最常见的攻击方式. 攻击者首先偷录合法用户在身份认证过程中释放的声信号, 然后进行重放, 以此达到欺骗身份认证系统, 获取权限的目的. 例如, EchoPrint[15]利用音频设备发射 FMCW 声信号, 并收集面部回声, 提供面部生理特征认证凭据. 攻击者同样可以在合法用户身份认证期间, 录制回声音频, 然后重放录制的音频对系统进行攻击; 基于声纹的身份认证系统, 一般依赖用户语音中独特的声轨形状作为身份凭据, 完成对用户的认证. 而攻击者可以录制受害者的语音并通过扬声器将录制的声音播放, 以达到欺骗认证系统的目的.

合成攻击的主要思路是从用户语音中获取声纹特征, 并生成具有相同声纹特征的其他语音[24]. 具体做法是攻

击者输入大量语音数据, 通过语音合成算法, 训练语音合成模型 model. 然后将受害者的语音数据按照一定的格式进行标注, 生成一个新的数据集. 使用新的数据集对模型 model 进行微调训练, 之后得到新的模型 model′, 使用模型 model′生成声纹认证系统所需的语音数据, 并进行播放. 合成攻击对声纹认证系统具有较大的威胁.

语音对抗样本攻击是通过在语音数据集中添加细微扰动形成有害语音样本, 该样本会导致声纹认证系统以高置信度给出一个错误的输出[25]. 机器学习模型对于语音对抗样本攻击十分敏感, 攻击者只对原始语音样本进行轻微改动, 即可导致系统出错. 通过认证后, 攻击者可以播放一段特别制作的恶意语音命令, 人耳无法觉察, 但是语音控制系统会执行相关恶意语音命令操作.

中间人攻击是针对基于声感知的 2FA 的有效攻击方式. 如图 4 所示, 攻击者在空间位置上远离受害者, 并冒充受害者身份向 Web 服务器发送登录请求, 然后 Web 服务器分别向攻击者的登录设备和与受害者账号绑定的注册手机发送信息. 它们做出响应后, 双方所处环境下的声信号会通过隐蔽的高速传输通道实时传递给对方, 继而提供有效的接近证明.

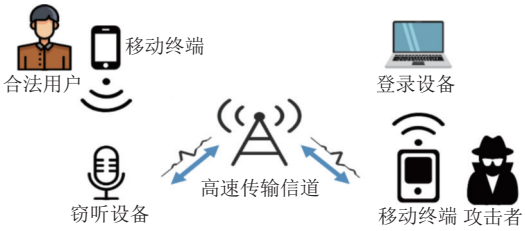


图 4　中间人攻击模型

同位攻击与中间人攻击不同, 它要求攻击者在空间位置上接近受害者. 如图 5 所示, 当攻击者尝试登录时, 会触发注册手机自动响应, 生成对应的声信号. 攻击者的登录设备可以偷录到该声信号响应. 但是, 一般情况下, 同位攻击需要攻击者与受害者很接近时才能攻击成功, 攻击的隐蔽性差, 容易被察觉.

图 5　同位攻击模型

3　移动终端身份认证进展

本节主要对移动终端基于不同认证凭据 (秘密知识或生物特征) 的身份认证国内外研究进展进行分析、总结和对比, 基于信任器件的身份认证方法很少单独使用, 我们在此不展开讨论. 双/多因素身份认证通常结合两个或多个身份认证指标来提供比使用单一验证指标更安全的身份认证. 多个身份验证指标结合的科研工作较为繁杂, 我们在下一节专注讨论基于声感知的双因素身份认证科研进展.

3.1　基于秘密知识的身份认证

基于秘密知识的身份认证是一种长期以来被广泛应用在移动终端上的身份认证方式, 主要依靠只在合法用户和身份认证系统之间共享的秘密知识进行身份认证. 如图 6 所示, KBA 系统基本可以依照秘密知识的类型分为以 PIN 码和口令为代表的文本型和以手势口令[26]和图形口令为代表的图案型.

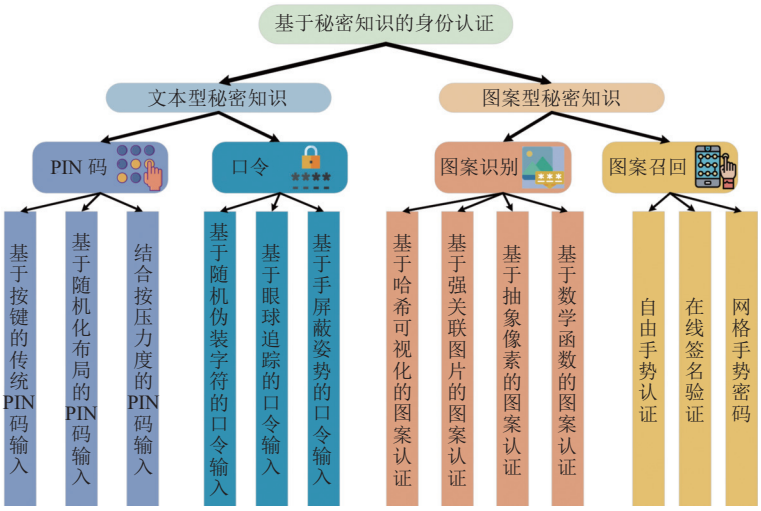


图 6　KBA 系统分类

3.1.1　文本型秘密知识身份认证

文本型秘密知识可以是数字字符、字母字符、数字和字母的组合. 用户简单地在移动终端触摸屏上输入 PIN 码或口令以进行身份验证. 到目前为止, 这些基于文本的身份认证方法在验证用户身份、保护用户数据安全、移动终端访问控制的应用场景下, 仍然应用最为广泛[27]. 大量网站的口令强度评价方法为基于规则的方法, 即口令强度依据所包含的字符长度和类型, 字符长度越长, 类型越复杂, 被随机猜测或暴力破解的可能性就越低[28]. 近年来, 深度学习技术的发展为改进口令猜测技术、提高口令破解效率提供了潜在的新途径, 智能口令猜测逐渐形成了口令破解的最新方法[29]. 为确保用户的 PIN 码或口令不被其他人猜到, 用户必须记住超过一定长度的字符串. 例如, 与移动终端上的应用程序 (例如移动支付) 关联的 PIN 码长度通常为 4 位或 6 位数字. 在相同长度下, 口令相比 PIN 码安全级别更高, 因为字母数字组合比纯数字组合可能性更多. 然而, 基于文本的身份认证面临极其严峻的安全问题. Bonneau 等人[11]根据对 1 100 多名银行客户的调查, 分析了用户选择 4 位 PIN 的偏好, 发现人们倾向选择容易使用的 PIN 或口令. Lee[30]发现 PIN 码认证容易受到肩窥攻击, 并提出用于 PIN 输入定量安全性分析的新概念. 当合法用户输入 PIN 码时, 攻击者可以通过直接观察、借助镜子或隐蔽的针孔摄像头轻松窥探用户输入的 PIN 码[31]. Khan 等人[32]通过实验表明, 平均而言, 两次观察就可以很容易地正确猜测 PIN 码, 并且一般的屏幕倾斜防御对于肩窥攻击的抵抗力有限. 王平等人[28]提出, 一方面用户重用口令的趋势越来越严重, 另一方面相当比例的用户口令曾被泄露. 攻击者必然会利用用户曾经泄露的口令来攻击用户当前的口令. 可以发现, 传统 PIN 和口令认证方法无法抵抗肩窥攻击, 而且在实际使用过程中由于用户的自身原因, 组成字符的选择存在一定倾向性, 造成了严重的安全隐患.

面对上述的问题, 研究人员提出一系列的方法来增强 PIN 码和口令的安全性, 使攻击者难以破解. von Zezschwitz 等人[33]提出了基于简单触摸手势的安全 PIN 输入方案. 该方案在每个 PIN 数字输入后, 布局再次随机化. Krombholz 等人[34]使用 iPhone 6s 的 3D Touch 功能来获取压力等级, 并结合压力和 PIN 码作为口令输入. 但是 Khan 等人[32]通过实验表明基于压力信号的 PIN 认证安全性增强方案具有固有的定时侧信道信号, 如果攻击者进一步获取这些信号, 该方案对于肩窥攻击几乎无效. 为了抵御肩窥攻击, Yan 等人[35]要求用户在口令输入期间利用手屏蔽的姿势主动构建相对安全环境, 只有在该环境下, 用户才可以获得隐藏的口令位的随机映射. Li 等人[36]开发了一种眼球追踪系统, 用户通过眼睛注视相应的按键, 在手机上输入口令. Alsuhibany 等人[37]重新考虑了典型的口令输入机制, 在真实口令两段添加任意长度的随机伪装字符, 并通过激活主键和失效主键界定真实口令在整个口令中的位置, 将简单的口令映射到任意复杂的口令, 可以有效抵抗肩窥攻击. Mayer 等人[38]为了解决在游戏手柄输入秘密知识容易被肩窥的问题, 提出了基于 16 花瓣形菜单结构的防肩窥认证方案, 经过多轮选择完成文本口

令的输入. 为了确保口令的复杂性, Castelluccia 等人[39]开发了一种基于马尔可夫模型的自适应口令强度计. 该方案通过推导组成口令的 *n*-gram 概率来估计口令强度. 此外, Kelley 等人[40]使用多个口令猜测算法评估在各种口令组合策略下创建的基于文本秘密的强度. 他们发现, 尽管系统管理员强制用户创建符合某些策略的口令 (例如, 必须有大写、小写字母和特殊字符) 以使其口令更难猜测, 但攻击者如果根据策略创建足够的预训口令, 仍然可以成功猜测口令. 由此可见, 文本型秘密知识的弱点主要在于用户记忆的方便性和秘密知识强度之间的平衡, 这是现有 PIN 码和口令方案可能被破解的根源性问题.

3.1.2 图案型秘密知识身份认证

图案型秘密知识主要包括手势口令、图形口令等. 与基于文本的身份认证相比, 基于图案的身份认证减轻了字母数字内容繁琐的负担. 根据用户是否需要识别或复制秘密内容, 图案型秘密知识可分为两类: 基于识别的秘密知识 (要求用户识别秘密图案) 和基于召回的秘密知识 (要求用户重新输入秘密图案). 基于识别的图案认证系统要求用户从一个大集合中选择图形内容的一个子集作为秘密知识来注册身份信息. 在身份验证过程中, 用户需要正确识别所有预先选择的秘密图案内容, 以证明其身份. 这类方法主要研究如何从图案内容 (如图片、图标和符号) 中提取秘密信息, 并将其转换为图案口令. Dhamija 等人[41]应用哈希可视化技术生成随机图片, 供用户选择口令, 并在身份验证期间基于识别正确的预选图案对用户进行验证. 后来, 基于识别的秘密图案因其简单性和便于记忆而被移动终端广泛采用. 移动终端中可用的各种图形内容已被探索用于身份验证, 例如具有强关联性的图片[42]、数学函数[43]生成的抽象像素以及已安装的 APPs 图标[44]. 这些认证方法丰富了基于识别的图案认证系统, 并为用户提供易于识别的便利, 然而现有图案认证系统秘密空间小, 被破解可能性大.

与基于识别的图案认证系统不同, 基于召回的图案认证系统要求用户输入秘密图案内容, 而不是简单地识别它们. 它可以进一步分为基于回忆的技术和基于提示回忆的技术. 基于回忆的技术要求用户在没有任何提示的情况下重现秘密的图像内容. Jermyn 等人[45]让用户在触摸屏上绘制一个独特的图案模式 (例如, 一个字母“a”) 来进行身份认证. 如果图案与用户先前注册过程中的图案相匹配, 则用户通过身份认证. 此外, 研究表明, 自由形式的手势[46]和签名[47]都是移动终端上使用效果良好的图案秘密. 另一种在移动终端上广泛使用的基于召回的图案秘密是手势口令, 它允许用户按照特定的规则在给定网格上的点绘制特定图案模式. 手势口令更容易记忆, 同时保持与 PIN 码相当的安全强度. 例如, 一个 3×3 网格映射有 389 112 种不同的图案模式, 与 6 位 PIN 码 (即 1 000 000 个可能的组合) 相比, 攻击者进行破解需要花费同一级别的时间. Cho 等人[48]提出一种系统引导用户设置手势口令的方案, 该方案要求用户选择的图案模式必须包含系统随机生成的几个点, 以此确保用户设置手势口令没有习惯性偏好. 类似的工作还包括姚沐言等人[49]提出一种基于上采样单分类的智能手机手势口令隐式身份认证机制, 融合用户使用手势口令的行为特征抵抗肩窥攻击. Double Patterns[50]允许用户依次输入两种图案并叠加, 来完成身份验证过程. 相比于传统的单图案手势口令和 4–6 位 PIN 码, 安全性得到提升. Munyendo 等人[51]通过使用图案黑名单, 即禁止用户选用常见的图案来提高安全性, 并建议图案黑名单的容量为 100 个, 可以有效地权衡安全性和实用性. 基于提示回忆的技术允许用户使用提示再现秘密图形内容. 口令点[52]要求用户在任意图像上选择 5 个有序点/像素序列对用户进行身份验证.

后文表 1 总结和对比了各种基于秘密知识的典型认证方案. 相比文本型秘密知识, 图案型秘密知识更加易于记忆, 一定程度上缓解了记忆性和秘密强度上的冲突, 但是图案型秘密知识也因其技术特征有自己的弱点: 现有图案型秘密知识空间较小, 特别是对于基于识别的秘密, 受预定义图像池的限制. 同时还受倾向性的影响, 不同的用户倾向于选择相似的点或图像作为其秘密的一部分, 这使得对手能够基于常用的图案秘密发起字典攻击. 图案秘密的安全强度还受制于生成秘密的规则, 这可能导致一些图形秘密比文本秘密弱. 例如, 复杂手势口令 (例如, 8 条线构成的图案) 可能显示出更小的组合空间, 并且滑动过程中可能比简单手势口令 (例如, 4 条线构成的图案) 更容易中断.

3.2 基于生物特征的身份认证

随着配备多种传感器的移动终端的普及, 越来越多生物特征识别技术被应用到移动终端中. 基于生物特征的

身份认证主要分为两种: 基于生理特征的认证和基于行为特征的认证. 如图 7 所示, 生理特征通常是指用户独特的身体特征 (如指纹、人脸) 或生命体征 (如心电信号), 行为特征主要反映某些习惯动作 (如语音、步态) 中的动作模式.

表 1　基于秘密知识的典型认证方案对比与总结

秘密类型	秘密	认证方式	抵抗攻击类型	文献	误判率 (%)
文本类	PIN码	基于召回	—	[11,27,30–32]	<3.5
	口令	基于召回	—	[28]	1–5
	增强PIN码	基于召回	肩窥攻击	[33,34]	>10
	强制口令	基于召回	肩窥攻击	[35−40]	>5
图案类	图片	基于识别	肩窥攻击	[41,42]	<6
	抽象像素	基于识别	肩窥攻击	[43]	5
	应用图标	基于识别	肩窥攻击	[44]	4.5
	秘密绘画	基于召回	—	[45]	—
	自由手势	基于召回	肩窥攻击	[46]	10–40
	在线签名	基于召回	肩窥攻击	[47]	5–30
	手势口令	基于召回	统计猜测攻击	[48−51]	<12
	像素序列	基于召回	字典攻击	[52]	21

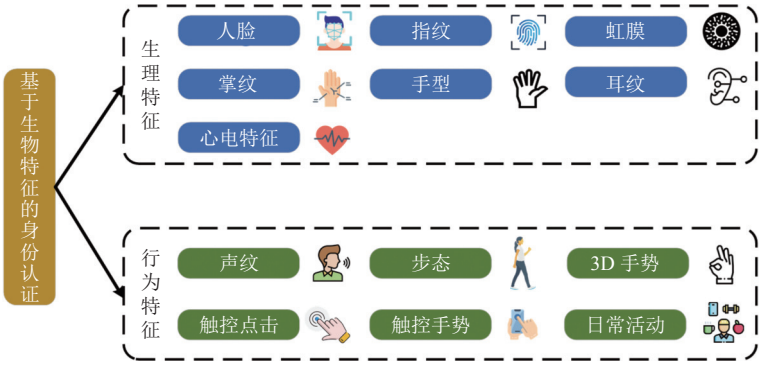


图 7　基于生物特征的身份认证系统

3.2.1　基于生理特征的身份认证

基于生理特征的主流认证方法包括指纹认证, 人脸认证等. 指纹认证通过识别用户手指末端皮肤上凸凹不平的纹路所蕴涵的大量信息来验证用户的身份, 关键想法是捕捉用户指纹摩擦脊的显著特征 (如毛孔、初期脊、折痕). 传统方法利用移动终端上现成的传感器 (如摄像头) 来提取用户指纹特征. Raghavendra 等人[53]利用智能手机摄像头以非接触式方式提取指纹特征, 提出的方案包括手指图像分割、预处理和缩放、细节提取和比较 3 个环节. 近年来, 屏下光学指纹传感器和电容指纹传感器被智能手机等移动终端广泛用于身份认证. 尽管指纹是目前在移动终端上应用最广泛的生物特征, 但是现有指纹认证方法容易在采集传感器上留下印痕, 从而被用来复制用户指纹. 因此在实际使用中, 指纹认证易受到演示攻击 (如假手指) 的威胁. 对此, 许多研究学者提出了活体检测的方法来防御演示攻击, 即检测输入的指纹是否有生命特征. Komeili 等人[54]提出融合心电图和指纹特征, 从指尖同时捕获的心电图不仅用于身份验证, 还用于指纹的活体检测. Park 等人[55]提出一种使用小型全卷积网络检测演示攻击的方法, 减少处理时间和内存需求, 能够将现有算法集成到智能手机中. 另外, Rathore 等人[7]利用指纹认证时指尖表皮产生的 3D 触觉反应效果与触觉表面的相互作用, 反映活人手指和假手指不同的解剖结构. 但是, 指纹活体检测无法防御指纹认证中的傀儡攻击, 即攻击者将失去意识 (例如睡着或者昏厥) 的受害者的真实指纹放在指纹传感器上以绕过指纹认证系统. 对此, Wu 等人[56]设计了一种基于用户行为生物特征的指纹认证增强方案, 通过分

析指尖触摸的行为特征判断当前输入指纹的用户是否为合法用户.

人脸认证从数字图像或视频帧中提取用户的面部特征用于身份验证, 如眼睛、鼻子、颧骨和下巴的相对位置、大小和形状. 由于图像的旋转和缩放等变形会降低传统基于网格的人脸认证的精度, Taigman 等人[57]通过 CNN 神经网络提高人脸认证的性能, 准确率达到了 97.35%, 首次超过人类眼睛对人脸图像的认证水平. 在新冠疫情的影响下, 人们外出经常戴着口罩, 这给公共场所人脸认证带来了巨大的挑战. Aswal 等人[58]通过重新训练带有口罩的人脸模型, 生成面部特征向量以进行有效的戴口罩人脸认证. 近年来, 随着准确率的不断提高, 人脸认证系统已经广泛部署在移动终端上, 其安全性引起广泛关注. 在认证过程中, 系统容易受到演示攻击 (如通过利用受害者的人脸照片、视频和硅胶面具欺骗人脸认证系统). 对此, Li 等人[59]要求用户认证时将手机摄像头围绕用户面部旋转, 通过测量由面部视频和手机惯性传感器算出的头部运动姿势之间的一致性, 判断认证人脸的立体性, 显然这种方法无法抵御 3D 硅胶面具攻击. 为了克服这个困难, 田野等人[60]提出了一种基于局部二值模式和多层离散余弦变换的人脸活体检测算法. Liu 等人[61]提出了一种新颖的上下文对比感知学习框架, 从而有效地利用真实人脸和面具人脸之间的细粒度特征区别进行区分. 近年来, 3D 人脸认证技术也已应用于一些高端移动终端, 它比 2D 媒体捕捉更多的面部特征, 并提供更高的安全级别. 随着苹果 FaceID 的推出, 人脸识别由于其高准确性已经广泛部署在移动终端上, 然而用户人脸图像的易得性严重威胁现有人脸认证系统的安全.

基于生理特征的认证还包括以下几种: 掌纹认证通过摄像头捕捉手掌图像, 提取掌纹的主线、皱纹、表皮纹路等用于认证. 为了使得不同设备捕获的非接触式掌纹样本识别精度高, 同时既不需要类别标签进行训练, 也不需要使用预训练过滤器, Genovese 等人[62]引入了一种新颖的卷积神经网络 (CNN) 分类器, 通过无监督程序调整掌纹特定过滤器, 在训练期间不需要类别标签. 手形认证提取手指和手掌的几何形状尺寸用于认证. Song 等人[63]结合手的几何形状和多点触控的行为特征进行认证, 能够抵抗常见的几种攻击方法. 虹膜认证从视频或图像中提取用户眼睛虹膜的纹理模式用于认证. Ma 等人[64]结合虹膜识别和眼球运动追踪进行远程用户认证, 能够防止使用虹膜图像的攻击方法. 耳纹认证, 通过耳朵的特征 (如外耳形状、耳道形状与长度) 来进行用户认证. Fahmi 等人[65]利用用户耳朵的形状和纹理信息来表示人耳特征. 作者首先对所有局部二值模式 (LBP) 进行组合提取并连接到单个直方图中, 然后使用人耳定位的思想得到人耳的几何特征. 心电认证, 利用连接到智能手机的心电硬件提取心电测量实现用户认证. 肖剑等人[66]结合判别相关分析最大化两个特征集的相关性的特点, 提出了一种心电与光电容积脉搏波多模态生物识别模型. 表 2 总结和对比了各种基于生理特征的典型认证方案. 基于手形、耳纹和心电图的认证系统错误拒绝率较高, 也需要相对较长的时间. 另外, 基于虹膜、掌纹和心电图的认证系统需要额外的硬件, 它们目前在移动终端上的可用性较低, 并且很难广泛部署.

表 2　基于生理特征的典型认证方案对比与总结

特征	传感器	衡量标准	分类器	文献	误判率 (%)
指纹	光学、电容、指纹传感器	脊线细节	SVM、LR、CNN、KNN	[7,53–56]	<4
人脸	相机	五官位置和形状	CNN	[57–61]	<4
掌纹	相机	表皮纹路, 主线	CNN	[62]	3
手形	相机	手指/手掌长/宽	k近邻法、SVM	[63]	<10
虹膜	相机	虹膜纹理	SVM	[64]	<1
耳纹	相机	耳朵形状和纹理	最近邻分类器	[65]	7.5
心电	导联电极	心电与脉搏信号	SVM、KNN	[66]	2–25

3.2.2　基于行为特征的身份认证

基于行为特征的认证主要有声纹认证, 步态模式认证等. 声纹认证提取每个用户独有的声纹特征 (例如频率倒谱系数) 来进行认证. 它既可以是文本独立的 (接受任意内容的语音), 也可以是文本依赖的 (只接受特定内容的语音). 文本独立的声纹认证更灵活, 但是需要长时间的语音才能达到良好的表现. Zhang 等人[67]设计了一种基于深度网络和三重态损失的文本独立说话人验证框架. 与传统的 i-Vector/PLDA 方法相比, 该方法大大简化了声纹认证系统. Wan 等人[68]提出了一套采用单通道数据训练的端到端神经网络的声纹认证方案, 并且利用测试背景已知

的数据增强策略进一步提升性能. 余玲飞等人[69]提出了一种基于卷积神经网络 (CNN) 和循环神经网络 (RNN) 的声纹认证方案来提高文本独立声纹认证的精度. 该方案结合了 CNN 和 RNN 的优势, 可用于移动终端声纹认证. 目前, 文本依赖的声纹认证应用更广泛, 因为它不需要长语音进行验证, 用户只需要说出指定文本 (如“Hey Siri”), 并且识别精度更高. 现有声纹认证系统由于成本低廉, 使用简单, 在移动终端上部署方便, 但是声音传播信道的开放性使得其容易受到重放攻击和环境噪声的影响. 为了减少环境噪声对声纹认证的影响, Li 等人[70]利用毫米波感知用户发声时的喉结震动, 能够将复杂的环境噪音从毫米波信号中分离出去, 然后提取与文本无关的声道和声源特征进行用户认证.

基于步态模式的认证方案旨在实现用户携带移动终端时对用户进行连续认证, 具有特征难以伪装、远距离识别、用户可接受等优势. Sprager 等人[71]提出利用智能手机的加速度计捕捉用户独特行走模式下的加速度信号, 通过特征空间中加速度数据变换实现了步态模式的统计分析用于身份认证. 为了提高智能手机步态认证的鲁棒性, Zou 等人[72]提出了一种混合深度神经网络用于稳健步态特征表示, 由卷积神经网络和循环神经网络依次连接空间域和时域中的特征. 施沫寒等人[73]提出了一种兼具准确性和可解释性的步态识别方法, 使用一种基于 Shapelet 的时间序列分类方法进行步态的识别与认证. von Hamme 等人[74]对基于加速度计和陀螺仪组合的 IMU 传感器的步态认证系统进行了深入的安全性分析, 评估了针对无保护和受保护步态认证系统攻击的有效性. 步态识别不容易伪装, 适合远距离的身份识别, 但是现有步态识别系统准确率和鲁棒性有待提升.

基于行为特征的认证还包括以下几种: 触屏点击认证使用智能手机传感器提取的敲击行为特征来进行连续的用户认证. Bo 等人[75]利用用户的触屏点击行为特征以及引起的手机微小运动以静默和透明的方式进行身份认证. 触摸手势认证捕捉移动终端多点触控屏幕上用户多个手指输入的独特模式. Sae-Bae 等人[76]定义了一套全面的五指触摸手势, 在对动作特征进行分类的基础上, 利用多点触控表面将手势输入和生物识别技术相结合. 3D 手势认证获取用户在空中的 3D 手势行为特征进行身份认证. Shahzad 等人[77]允许用户在空中输入预定义的手势, 并根据 WiFi 信号的信道状态信息 (CSI) 从信号中提取 3D 手势下的频率变换特征. 日常活动认证利用嵌入在移动终端中的多个传感器来捕获用户独特的行为和环境特征进行连续的用户认证. Shi 等人[78]利用 WiFi 信号捕获用户日常活动中包含的独特生理和行为特征, 并开发基于深度学习的模型来提取不同用户的独特 WiFi 指纹. 表 3 总结和对比了各种基于行为特征的典型认证方案. 行为特征难以被伪造, 因此对暴力攻击和观察攻击具有较强的鲁棒性. 但是除了声纹认证之外, 几乎所有基于行为生物特征的认证系统都存在较高的错误拒绝率, 这使得合法用户体验差.

表 3　基于行为特征的典型认证方案对比与总结

特征	传感器	衡量标准	分类器	文献	误判率 (%)
声纹	麦克风	光谱特征、到达时间差	CNN、RNN、SVM、HMM	[67–70]	<5
步态	加速计、陀螺仪	步态特征	CNN、LSTM、最近邻分类器	[71–74]	<12
触屏点击	触摸屏	按压压力	SVM	[75]	7–20
触摸手势	触摸屏	几何特征、手势时间	相似度分数	[76]	10
3D手势	3D相机、WiFi	频域特征、统计学特征	高斯混合模型	[77]	5–15
日常活动	WiFi	行为特征、统计学特征	DNN	[78]	9

4　基于声感知的身份认证进展

第 3 节分别介绍了移动终端不同认证凭据的身份认证国内外研究进展, 基于声感知的身份认证利用声信号感知这些凭据进行身份认证. 本节主要对移动终端基于声感知的身份认证国内外研究进展进行分析、总结和对比, 包括基于声感知的典型身份认证、双因素身份认证和认证活体检测.

4.1　基于声感知的典型身份认证

近年来, 基于声感知的身份认证因其低成本的特性以及扬声器和麦克风在移动终端中的广泛部署而得到了很好的探索. 基于声感知的典型身份认证主要包括被动声感知身份认证和主动声感知身份认证, 如图 8 所示.

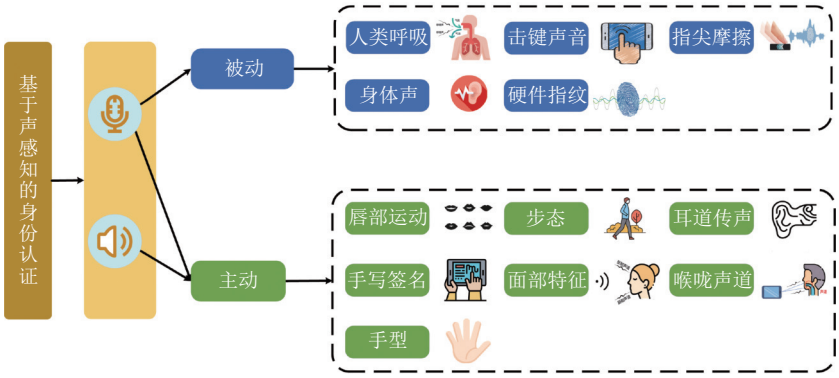


图 8 基于声感知的典型身份认证

被动声感知身份认证利用认证实体发射的特征声信号以进行身份认证. 例如, 在用户产生的声信号中提取行为特征或者生理特征, 在移动终端产生的声信号中提取硬件指纹. BreathPrint[14]从 3 个级别的人类呼吸声中提取嗅探、正常呼吸和深呼吸中的声学特征 (如伽马通频率倒谱系数 (GFCC)) 用于用户身份认证. 在此基础上, Chauhan 等人[79]利用智能手机、智能手表和树莓派芯片 3 种移动终端进行用户呼吸声认证, 并使用浅层分类器 (如 SVM、GMM、逻辑回归) 和深度学习分类器 (如 LSTM、MLP) 进行分类. 结果表明 LSTM 模型尺寸最小, 认证时间最短, 并且比其他分类器准确率高. Zhou 等人[80]利用来自击键的声音验证用户的合法性, 提取了几种声学特征 (如声信号的信号强度、梅尔频谱系数 (MFCC)), 并进一步应用 SVM 进行身份验证. SonicPrint[81]发现不同用户手指在物体表面滑动产生的摩擦声具有独特性, 而这种独特性取决于手指的表面纹理, 即指纹脊纹, 因此可以将摩擦声作为指纹特征进行用户身份认证. EarID[82]和 EarPrint[83]作为新颖的身份认证方法, 利用低成本嵌入式麦克风感知通过用户耳道传输的身体声音, 并根据不同用户耳道的唯一性提取出对应的生物特征. Zhou 等人[84]和 Das 等人[85]首次提出利用智能手机扬声器和麦克风的频率响应的独特性和稳定性为智能手机生成设备指纹. 其中 Zhou 等人[84]精心挑选了提取声指纹的音频信号的频率范围和模式, 从而减少非线性效应和环境噪声的影响, 并保证用户察觉不到信号特征的提取过程. Luo 等人[86]为声指纹提出了一个新的特征集, 范围能量差 (BED) 描述符, 用于数字语音记录的来源归属, 证明了声信号的频率响应曲线可以作为一个表征录音设备的稳健指纹. 由于击键声、呼吸声、摩擦声、耳内发声都在可听到的频率范围内, 很容易受到周围环境噪声的干扰. 而且, 这些研究都是基于提取的声发射特征, 因此可能受到重放攻击.

而主动声感知身份认证则主要发出人耳难以察觉的声信号来感知人类的行为特征或者生理特征, 以进行用户认证. SilentKey[87]利用智能手机发出超声波信号, 并分析唇部运动对反射声信号的细粒度影响, 提取出不同用户的独特特征进行身份认证. LipPass[88,89]同样利用智能手机发出超声波信号, 感知用户唇部运动, 认证用户身份, 同时还能对用户进行活体检测. 进一步将声道、舌头的个体独特性考虑进来, 从反射声信号中提取特征. Wang 等人[90]利用声学传感器感知用户步态数据, 通过深度神经网络 (DNN) 验证用户身份, 但是和传统步态认证方法相比, 认证距离受限. ASSV[91]提出一种基于声感知的在线手写签名验证方案, 使用了一种基于弦的方法来估计由微小动作引起的声信号相位变化, 利用声信号的变化来实现手写签名验证. 进一步, Zhao 等人[92]利用手机内置的扬声器和麦克风传输专门设计的训练序列, 并记录相应的回声用于信道脉冲响应 (CIR) 估计. EarEcho[93]基于人耳耳道独特的物理和几何特征, 利用耳塞扬声器发出 chirp 声信号, 然后使用麦克风记录通过用户耳道传播的声信号, 提取离散傅里叶 (DFT) 特征进行认证. Zhou 等人[15]利用智能手机中的扬声器向用户面部发出高频声信号, 从面部反射的声信号中, 使用 CNN 提取声学特征进行人脸认证, 并允许手机握持姿势的变化. 最近, VocalLock[94]提出利用调频连续波 (FMCW) 感知用户声道特征用于用户身份认证. 值得注意的是, VocalLock 不依赖固定的口令, 但有效认证距离很短 (<10 cm). Huang 等人[95]利用媒体声感知验证用户持握智能手机的手部形状, 提出基于 CNN 的用户身份验证方法. 由于反射的声信号感知精度有限, 主动声感知身份认证目前只能对小数据集用户进行认证. 表 4 总结和对比了各种基于声感知的典型安全身份认证方案.

表 4 基于声感知的典型安全身份认证方案对比与总结

分类	特征类型	特征	技术	频带	文献	误判率 (%)
被动	行为特征	人类呼吸	GFCC	可听范围	[14,79]	6–20
		击键声音	MFCC	可听范围	[80]	11
		指尖摩擦	MFCC、LPCC	可听范围	[81]	3
	生理特征	耳道	FFT、MFCC	可听范围	[82,83]	<5
	信任器件	硬件声指纹	MFCC、高斯超向量	—	[84–86]	<4
主动	行为特征	唇部运动	多普勒频偏	17.5 kHz[87]/20 kHz[88,89]	[87–89]	<10
		步态	多普勒频偏	38–42 kHz	[90]	<10
		手写签名	相位变化、CIR	大于17 kHz	[91,92]	5.8
	生理特征	耳道	DFT	小于6 kHz	[93]	5.8
		面部特征	FMCW	不可听范围	[15]	6.2
		喉咙声道	FMCW	17–20 kHz	[94]	8.9
		手形	声谱特征	小于20 kHz	[95]	<5

4.2 基于声感知的双因素身份认证

基于声感知的双因素身份认证 (2FA) 只有在系统成功验证两种认证因素后, 用户才会被授予访问权限. 这样攻击者伪造认证凭据的难度更大, 可以明显提高用户身份认证的安全强度.

直接的方法是向身份认证系统额外添加声感知的认证指标 (每种指标可以是秘密知识、生物特征或者是信任器件), 即声感知提取其中一种认证因子. Sound-proof[16]将环境噪声相似度作为口令认证的第 2 个因素, 以提高移动终端的安全性和私密性. 它可以测量用户智能手机接收到的环境噪音与浏览器之间的相互关系. 此系统可能存在安全漏洞, 智能手机上的应用程序所触发的通知或警报声音都可能被当成环境噪音, 从而破坏 Sound-proof 中的第 2 个认证因素[96]. 随后的研究, 如 Home alone[97]和 Listen watch[98], 提出使用随机选择的声音信号进行认证. Home alone 使用智能手机主动发出的通知声音来衡量智能手机与浏览器的接近程度, 而 Listen watch 则使用人类语音作为声音因素来检测智能手表和浏览器之间的接近程度. 与 Sound-proof 类似, 他们使用相互关系来衡量用户的设备记录的声音与登录时使用的声音的相似性. 在此研究基础上, Proximity-echo[99]利用用户注册的电话和登录设备的接近度作为 2FA 的第 2 个认证因子, 该系统通过两个设备的扬声器交替发出声学信号, 然后用麦克风感应反射回声来提取位置特征, 并进行接近检测, 可以很好防御中间人攻击, 并且在距离大于 0.8 m 时可以有效抵抗同位攻击. EchoPrint[100]利用智能手机中的音频设备和前置摄像头同时提取面部轮廓, 进行用户身份验证. 它将声学特征和面部特征作为联合特征输入到 SVM 分类器进行训练和分类. TouchPrint[101]提出了一种基于声信号感知用户手部姿势的双因素身份认证方法, 当用户输入 PIN 码或手势口令时, 利用声信号捕获不同用户独特的手形特征作为第 2 认证因子. 类似地, 当用户输入手势口令时, SwipePass[102]利用声信号捕获不同用户的握持手形特征作为第 2 认证因子, 并进一步提高了准确率. PressPIN[103]提出了基于结构声衰减的触屏压力感知方法, 用户在输入 PIN 码的同时, 可以考虑轻按、重按、由轻到重按压不同的 PIN 码位, 通过提取 PIN 码输入过程中手指的压力信号组成压力码, 作为第 2 认证因子. Wu 等人[104]提出使用智能手机摄像头进行手形认证时, 利用主动声感知技术同时感知另一只手的握持手形作为第 2 认证因子, 从而抵御针对手形认证的呈现攻击. 然而, 这种双因子身份认证方法需要多个独立的输入, 大多需要用户进行额外操作, 增加了硬件和计算成本, 降低了它们在实践中的可用性.

另一种基于声感知的双因素身份认证方法是利用声信号一体化感知提取用户的全部认证因子. 与声感知提取其一认证因子的双因素身份认证方法相比, 这种方法没有额外硬件成本, 更加方便, 用户不需要进行额外操作, 体验更好. Proximity-proof[105]通过 OFDM 调制的声学信号将双因素身份验证响应传输到浏览器, 可以防止中间人攻击和同位攻击. 在传输过程中, 提取音频的独特指纹, 验证发送信号的智能手机. 这一步作为第 1 个接近检查, 可以抵抗中间人攻击. LVID[106]利用嘴唇反射的高频声信号捕获不同用户说话时的独特唇部运动的细微差别, 然后和语音信号进行有机融合组成多生物特征, 以增强智能手机声纹认证的安全性与鲁棒性, 如图 9 所示. EchoLock[107]

通过研究手掌挤压智能手机对以手机本身为介质传播的声信号的影响来感知用户手形, 然后利用声信号感知的手形生理和行为特征可以用于识别不同用户的身份. HandLock[108]使用智能音箱内置的扬声器和麦克风来发出和感知手掌反射的超声波信号以检测特定手势动作. 该方法可以作为语音助手执行特定敏感命令前声纹认证的第 2 认证因素. SoundID[109]提出了一种基于上下文共存检测和动态声学指纹决策的 2FA 系统. 其中, 动态声学指纹采用了声学硬件频率响应 (扬声器和麦克风) 的稳定唯一性和声信号传播的不可预测性, 可以有效抵抗中间人攻击和同位攻击. 表 5 总结和对比了基于声感知的双因素身份认证方案.

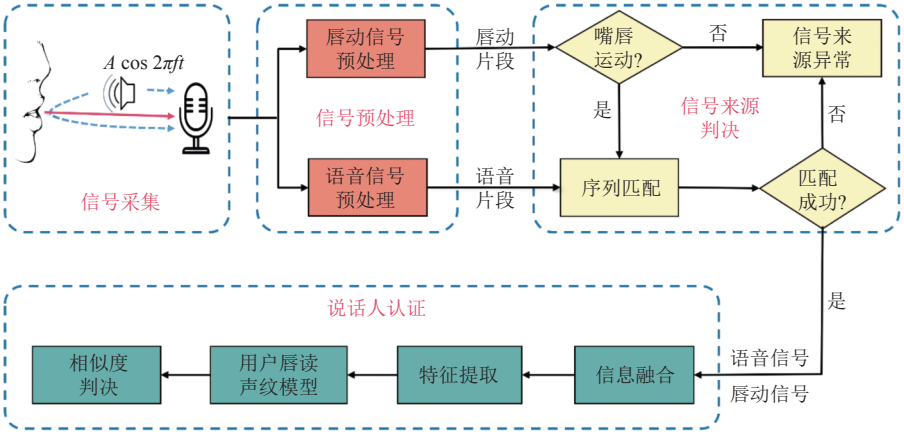


图 9 基于声感知的语音和唇动信号双因素身份认证

表 5 基于声感知的双因素身份认证方案对比与总结

分类	特征	技术	抵抗攻击类型	频带	文献	误判率 (%)
部分提取	噪声相似度	1/3倍频程	相似环境攻击	可听范围	[16]	1
	接近程度	1/3倍频程	—	50–4000 Hz	[97]	<30
	接近程度	相关性分析	环境猜测攻击、同位攻击	可听范围	[98]	5
	接近程度	PCCs	中间人攻击、同位攻击	14–15 kHz	[99]	<12
	面部特征	FMCW	重放攻击	不可听范围	[100]	2
	手形	ZC序列	模仿攻击	14–20 kHz	[101]	8
	握持手形	ZC序列	模仿攻击、重放攻击	17–23 kHz	[102]	3.2
	按屏压力	固体声提取	已知PIN码攻击、肩窥攻击	18–22 kHz	[103]	3.3
	握持手形	ZC序列	呈现攻击	17.46–22.54 kHz	[104]	2.45
一体化提取	接近程度和硬件指纹	OFDM	中间人攻击、同位攻击	18.1–20 kHz	[105]	—
	声纹和唇动	MFCC、GMM	重放攻击、模仿攻击	18–20 kHz	[106]	5
	手形和握持姿势	MFCC、FTT	重放攻击、模仿攻击	18–22 kHz	[107]	6
	声纹和手势	正交相位提取	重放攻击、模仿攻击	>12 kHz	[108]	3.5
	接近程度和硬件指纹	1/3倍频程	中间人攻击、同位攻击	18–20 kHz	[109]	3.4

4.3 基于声感知的认证活体检测

目前, 许多生物特征认证方案都难以抵抗特征深度伪造的攻击, 导致它们的认证安全性大大减弱. 例如, 人脸认证容易受到静态图像、视频、高精度硅胶面具等的欺骗; 声纹认证容易遭受重放攻击、合成攻击以及语音对抗样本攻击. 此现象最根本的原因是它们普遍缺乏对认证实体进行活体检测的环节. 鉴于有生命的活体和无生命的物体产生或反馈的声信号具有较大差异的事实, 基于声感知的认证活体检测利用声信号判断认证生物特征是否来自有生命的活体. 基于声感知的认证活体检测主要分为被动声感知活体因素和主动声感知活体因素, 如图 10 所示.

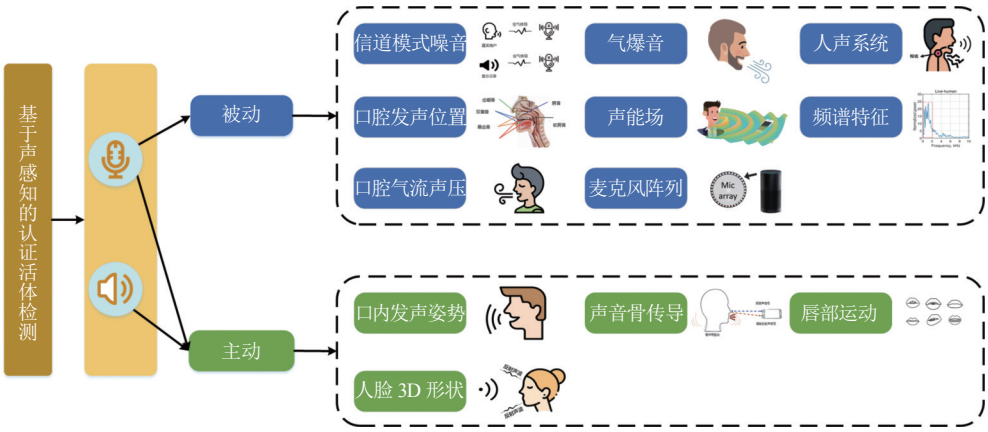
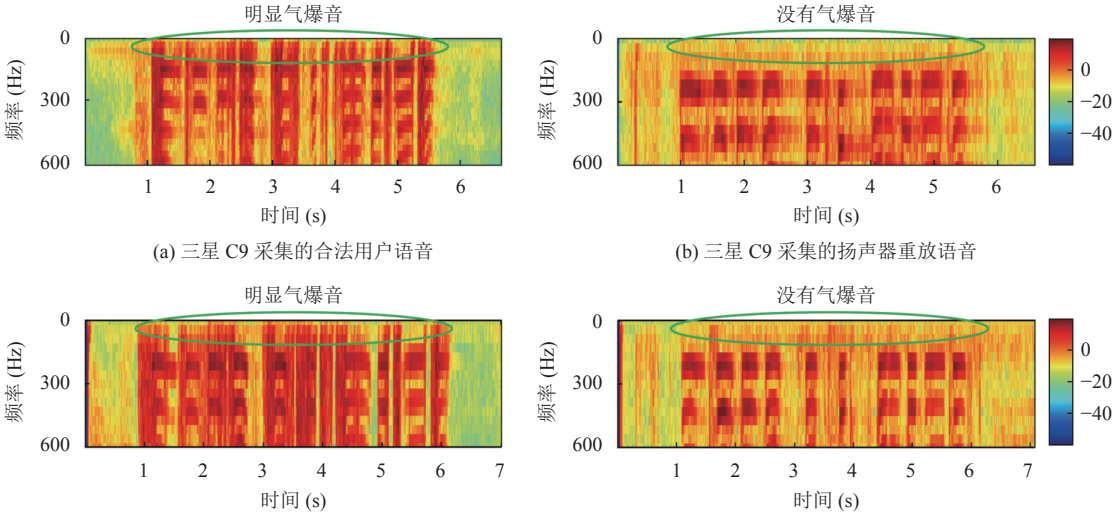


图 10　基于声感知的认证活体检测

被动声感知活体检测利用真实用户发射的声信号和扬声器发射的声信号特征之间的区别鉴别认证对象是否为活体, 主要是对声纹认证进行活体检测. Wang 等人[110]提出了基于信道模式噪声的重放攻击检测方法, 该方法基于真实用户语音与重放语音所经过信道存在差异的事实, 选择提取 Legendre 系数和统计系数共 12 个长期特征来描述信道指纹, 以此构建重放攻击检测器. Voicelive[111]通过测量接收到的不同音素到达时间差 (TDoA) 序列的动态特征来区分真实用户的声音和扬声器发出的声音. 然而该方案要求用户以固定姿势将嘴部靠近智能手机, 对用户嘴和手机的相对位置要求很严格. Shang 等人[112]设计一种鲁棒的声感知活体检测系统软件, 提供了 3 种有效活体检测方法: 基于频谱的活体检测方法利用重放音频难以刻画嘴部和喉咙部位的频谱差异的事实; 基于运动的活体检测方法利用内嵌加速度计提取人声系统运动的加速度序列; 第 3 种方法利用内嵌的振动电机在音频录制过程中随机添加振动效果, 区分真实语音和重放音频. VoicePop[17]利用用户在靠近麦克风说话时口腔呼气产生的气爆音鉴别认证对象是否为真实活体, 如图 11 所示, 重放语音由于偷录距离无法非常靠近用户口腔通常不含气爆音. 该工作基于气爆音在低频具有高能量以及持续时间范围为 20–100 ms 的事实, 设计气爆音检测算法来判断语音样本中是否含有气爆音, 并利用 GFCC 特征来排除环境噪声和硬件噪声导致的错误检测结果, 从而达到区分真实用户和重放设备的目的. VoicePop+[113]对先前的工作进行了优化, 通过在音素水平上定位气爆音, 并结合声压与口腔气压的一致特性, 进一步提高了判别音频中气爆音的准确率. Wang 等人[114]通过构建理论模型描述口腔气流压力与音素之间的关系, 然后根据模型估算语音中压力信号, 并结合辅助气流传感器捕获的真实压力信号进行一致性判别, 从而区分真实用户和重放设备. CaField[115]提出了基于声能场的活体检测方案, 借助人体声源和音频重放语音存在明显的物理结构差异, 导致它们的声能场在方向性存在可量化差异的基本事实, 通过提取长期平均场纹 (long-time average fieldprint, LTAF) 来刻画满足文本独立条件下的声能场一致性特征, 从而达到区分真实用户和重放设备的目的. VOID[116]利用从单声道音频中提取的频谱特征进行快速轻便的语音活体检测, 能够抵御常见的声纹认证攻击. 但是它没有使用多声道音频空间信息, 仍然容易受到更高级的攻击, 例如调制重放攻击. Li 等人[117]提出了一个针对机器音频攻击的整体化解决方案. 该工作利用智能音频系统上配备的多通道麦克风阵列来获取幅度和相位等信息, 并通过构建深度学习模型来实现对机器音频攻击的精准检测. ArrayID[118]提出了一种鲁棒的具有活体特征的麦克风阵列指纹, 并严格证明了阵列指纹主要与音源有关, 即真实用户与重放设备所反映的阵列指纹存在差异, 可以用于被动活体检测. 由于真实用户发射的声信号通常在可听到的低频率范围内, 很容易受到周围环境噪声的干扰, 而且很多工作需要额外的传感器, 对用户嘴部的认证距离和方向有比较严格的限制.

主动声感知活体检测通过主动发射探测信号, 并接收认证实体反射的声信号特征鉴别认证对象是否为活体. 主要是对声纹认证和人脸认证进行活体检测. Zhang 等人[18]鉴于人体发声与扬声器发声机理具有显著差异的基本事实, 提出了主动活体检测方案 VoiceGesture. 它通过测量反射声信号的多普勒频移来识别认证实体的发声姿势, 从而区分真实用户和重放设备. 之后, Zhang 等人[119]对 VoiceGesture 进一步优化, 使其具备文本无关的活体检测

能力. VibLive[120]利用 IoT 设备内置扬声器和麦克风对, 主动感应声音的骨传导振动, 并与空气传导的声音进行特征对比, 从而区分真实用户和重放设备. Echoface[121]利用声感知技术为人脸认证提供了一种活体检测方案. 它通过主动声学感应来区分面部的不均匀立体结构和平坦的伪造介质, 从而区分真实人脸和静态图片或视频. FaceLip[122]根据随机挑战动态生成声信号来捕获和分析说话人的嘴唇运动模式, 并消除环境中的噪声信号来实现远距离的人脸认证活体检测. Echo-FAS[123]提出一种双分支的架构, 巧妙融合了全局和局部频率特征, 以准确捕捉面部活性, 并且可以方便地与基于 RGB 的 FAS 系统结合, 以执行更加安全强大的面部反欺骗. 大多数主动声感知认证活体检测方案将发射的声信号限定在次超声波频段, 能够避免环境噪声的干扰, 并且用户难以察觉, 不影响用户体验, 但是认证距离和方向也有比较严格的限制. 表 6 总结和对比了各种基于声感知的认证活性检测方案.



(a) 三星 C9 采集的合法用户语音　(b) 三星 C9 采集的扬声器重放语音

(c) 三星 Note4 采集的合法用户语音　(d) 三星 Note4 采集的扬声器重放语音

图 11　基于气爆音的声纹认证活体检测

表 6　基于声感知的认证活体检测方案对比与总结

分类	特征	技术	抵抗攻击类型	频带	文献	误判率 (%)
被动	信道模式噪声	勒让德多项式系数	重放攻击	可听频带	[110]	2.8
	口腔发声位置	TDOA	重放攻击、替换攻击	可听频带	[111]	1
	人声系统、手机振动	STFT、频谱减法	重放攻击	17–20 kHz	[112]	<7
	气爆音	GFCC、STFT	重放攻击、模仿攻击	0–170 Hz	[17,113]	7.1
	口腔气流声压	HMM	重放攻击、合成攻击	可听频带	[114]	2.08
	声能场	STFT、GMM	重放攻击、合成攻击	可听频带	[115]	0.85
	频谱特征	STFT、LPCC	重放攻击、合成攻击	0–15 kHz	[116]	8.7
	麦克风阵列	LPCC、STFT	重放攻击、对抗样本攻击	0–5 kHz	[117,118]	<9
主动	口内发声姿势	STFT、DWT	重放攻击、模仿攻击	20 kHz	[18,119]	1
	声音骨传导	LPC、RSS	重放攻击、合成攻击	0–3 kHz	[120]	<3
	人脸3D形状	DTW、汉明窗	媒体伪造攻击	12–21 kHz	[121]	4
	唇部运动	STFT、LSTM-CNN	媒体伪造攻击、投影攻击、对抗样本攻击	18–21 kHz	[122]	5
	人脸3D形状	FMCW、STFT	呈现攻击	12–21 kHz	[123]	4.34

综合第 4.1–4.3 节发现, 现有基于声感知的安全身份认证方案主要对小数据集用户进行认证, 感知效果依赖用户的距离和方向, 大多需要用户进行初始化配置和适应性学习. 如何在大数据集用户上实现高准确率身份认证, 提高用户认证灵活性, 减少初始化配置和学习成本, 提高用户体验, 需要进一步深入研究.

5 研究挑战与展望

基于声感知的移动终端身份认证研究已经引起科研人员的广泛关注, 成本低廉且功能强大, 具有高度易得性和广泛的应用范围, 但是仍然存在一些研究限制和未来研究的开放问题. 本节结合当前研究的困难和挑战, 给出了衡量身份认证系统性能的两大指标 (安全性和实用性), 对未来的研究方向进行了展望.

5.1 研究挑战

目前, 基于声感知的移动终端身份认证研究面临的挑战主要集中在以下几个方面.

第一, 面对各种攻击实现认证实体与凭据间的可信绑定. 基于秘密知识的身份认证利用仅在合法用户和身份认证系统之间共享的知识凭据 (例如 PIN 码、手势口令等) 鉴定用户身份. 然而用户设置的秘密知识具有一定的倾向性, 容易受到智能口令猜测攻击. 因此, 如何利用声感知技术增大真实秘密空间, 构建强制执行高安全强度秘密知识输入方案, 权衡安全性和实用性需要进一步研究. 此外, 知识凭据容易通过肩窥攻击和各种侧信道信息推断出来. 如何利用声感知技术提取人眼无法观测、难以通过侧信道推断的隐性认证因子需要深入研究. 基于生物特征的身份认证技术利用人体固有生理特征 (例如人脸、指纹等) 或行为特征 (语音、步态等) 鉴定用户身份. 然而重放攻击、合成攻击、对抗样本攻击和特征深度伪造等多种攻击手段已经严重威胁基于生物特征认证的可靠性. 现有研究主要针对一部分攻击提高认证安全性, 例如 EchoFace[121]通过主动声学感应来区分面部的不均匀立体结构和平坦的伪造介质, 能够抵御人脸媒体伪造攻击, 但是无法抵御声信号重放攻击. 因此, 如何面对各种攻击, 在人机物三元空间中利用声感知技术实现认证实体与凭据之间的安全可信绑定是研究的难点.

第二, 利用声感知一体化提取具有唯一性的多特征认证因子. 移动终端上的某种身份认证技术 (例如 PIN 码、手势口令、人脸和指纹) 大多依赖单一认证因子, 无法在认证过程中提取多特征认证因子. 单一因子的认证技术容易受到环境因素干扰, 被假冒和窃取的风险高. 利用多重安全因子可以提高系统的安全性和鲁棒性, 同时做到多种认证手段的无缝接入, 因此多因子身份认证有着单因子身份认证无法比拟的优势. 然而现有大多数多因子身份认证技术对多因子特征只能采用分开提取方式, 无法满足多因子一体化提取认证需求. 例如, SwipePass[102]利用触屏感知输入的手势口令作为第 1 认证因子, 利用声信号捕获用户另一只手的握持手形作为第 2 认证因子. 基于移动终端的声感知技术已在各种应用领域显示出其优越性能, 在身份认证应用中, 可以利用声感知技术提取秘密知识、生理特征、行为特征和硬件指纹等多种类型认证因子. 因此, 如何设计安全可控的声信号感知方法, 支持多源异构特征的同源感知, 一体化提取具有唯一性的多特征认证因子, 实现多种生物特征、秘密知识、信任器件的有机融合, 使得安全性进一步提升, 并且不需要额外硬件成本和用户进行额外操作是第 2 个研究难点.

第三, 解决身份认证安全性与实用性之间的矛盾对立问题. 传统的身份认证方法由于认证因子容易被伪造和窃取, 对环境干扰的鲁棒性不强, 难以满足不同场景和安全级别需要. 多因子身份认证和活体检测方法提取多种认证因子特征, 并检测每种认证因子的真实性, 使得认证和识别过程更加安全可靠, 例如银行 APP 为了保证交易的安全性, 要求用户同时输入手机交易码和动态口令. 然而这些方法通常需要用户执行多种认证操作, 利用多种传感器采集认证数据, 移动终端需要消耗大量计算资源融合多模态异构数据. 此外, 认证系统的识别准确率受限于传感器精度和数据采集的各种限制. 例如, 现有基于声感知的身份认证安全性增强方法只能对小数据集用户进行认证, 大多需要用户进行复杂初始化配置和适应性学习, 严重影响用户体验. 几乎所有基于行为特征的身份认证系统都表现出较高的错误拒绝率, 给用户带来了不好的体验, 导致了此类身份认证方案难以广泛部署. 因此, 如何降低用户操作复杂度、硬件成本和异构数据融合难度, 提高身份认证系统识别准确性, 解决身份认证安全性与实用性和成本之间的矛盾对立是研究的第 3 个难点.

5.2 研究展望

随着智能终端内置硬件精度的提升和传感器技术的进步, 身份认证系统所能依赖的认证凭据趋于多样化. 从被广泛部署于移动终端的秘密知识, 到充分利用各类传感器提取的生理和行为特征. 基于声感知的移动终端身份

认证系统的研究已经取得了许多优秀的成果, 但是仍然存在很大的发展空间. 用于衡量身份认证系统性能的指标主要有两个, 即安全性和实用性, 一个性能优良的身份认证系统应该同时兼顾两者. 本文将从安全性和实用性两个角度对身份认证系统未来的发展趋势进行展望.

5.2.1 安全性提升

身份认证系统的主要功能是抵御非法访问, 保护用户隐私数据. 所以安全性对于身份认证系统至关重要. 身份认证系统安全性体现在与攻击者的博弈中是否展示出较强的防御能力. 而安全性的高低主要与其内部的设计实现有关, 例如身份凭据类型、身份凭据的提取、识别算法的实现等方面. 不同的身份凭据类型具有不同的安全强度, 例如虹膜被认为是最安全可靠的生物特征之一; 身份凭据的提取方法也会影响原始数据的可靠性, 例如指纹特征可以利用光学式、电容式和超声波式等方法获取, 其中电容式指纹识别最重要的优点是能够进行指纹活体检测, 抵抗假手指攻击; 识别算法通过对原始数据进行特征提取来进行身份决策, 将原始数据映射到特征矩阵或将特征矩阵映射到用户标签的算法实现, 都直接影响了认证系统的异常检测能力. 如今, 移动终端内嵌的多种传感器硬件会在合法用户进行身份认证的过程中收集丰富的侧信道信息数据, 攻击者可能利用它们实施秘密知识推断, 提取生理行为特征. 以下将针对几类不同身份认证系统分别给出在安全性方面的研究展望.

基于秘密知识的身份认证系统拥有庞大的用户群, 依靠合法用户与身份认证系统共享的秘密知识进行身份认证. 但该类型身份认证普遍存在秘密知识泄露的风险. 一般情况下, 用户设置的秘密知识具有一定的倾向性, 有助于攻击者通过智能口令猜测来获取秘密知识. 因此, 需要进一步研究高安全强度的秘密知识生成规则, 设计构建安全环境下的秘密知识输入方案. 除此之外, 一旦攻击者通过肩窥攻击、油污攻击、眼球追踪等侧信道攻击手段获取秘密知识, 那么身份认证系统将形同虚设. 因此, 如何提取无法被观测, 难以通过侧信道攻击手段推测的认证因子, 应该被进一步研究. 此前, 已经有部分研究是针对这一方向的, 例如 PressPIN[103]通过提取 PIN 码输入过程中手指与触摸屏的压力信号作为隐性认证因子, 增强 PIN 码的安全性. 对于基于图案的身份认证系统, 存在秘密空间小, 被破解可能性大的问题. 因此, 应该进一步研究如何从图案中提取更丰富的秘密信息, 并与生物特征、信任器件等因素进行有机融合.

基于生理特征的身份认证系统通过提取用户独特的生理特征或生命体征来认证用户身份. 目前应用最为广泛的是人脸认证和指纹认证, 它们的安全性也受到广泛的关注. 单一地对每次认证采集的人脸和指纹信息与注册收集的相关信息进行比对被证明并不安全, 两者都容易受到演示攻击, 3D 打印出来的人脸硅胶面具和带指纹的树脂指套都会对身份认证系统造成严重威胁. 例如, 在 2020 年的 DEFCON 网络安全会议上, 研究人员仅用一台基于 UV 树脂的 SLA 3D 打印机和价值 10 美元的材料就能破解三星 S10 智能手机中使用的最新超声波指纹认证技术[124]. 因此, 需要研究如何提取认证过程中其他难以被冒充伪造的活体检测和认证因子来增强安全性. 目前已经有一些关于这方面的研究, 例如 Komeili 等人[54]从指尖同时捕获的心电图不仅可以作为第 2 认证因子, 还可以用于指纹认证的活体检测.

信道的开放性使得基于声感知的身份认证系统在认证过程中产生的声信号很容易被偷录, 从而发起重放攻击. 如何辨别声源的真实性, 抵抗重放攻击需要进一步研究. 目前, 基于人工智能技术的高级攻击, 如语音合成攻击和语音对抗样本攻击等, 都会对声纹认证系统造成严重威胁. 如何抵御这些高级攻击, 也会成为此领域未来研究的热点. 除此之外, 鉴于声信号良好的感知能力和在身份认证安全性增强方面巨大的应用潜力, 研究者还应重点关注如何利用声信号巧妙地获取多种具有唯一性的独特隐性认证因子, 实现多种生物特征、秘密知识、信任器件的有机融合, 从而增强身份认证系统的安全性.

5.2.2 实用性提升

实用性也是衡量身份认证系统的重要指标. 一个缺乏实用性的身份认证系统即使拥有很高的安全性, 也难以得到用户的认可和广泛的部署. 实用性主要反映在以下几个方面: 1) 认证系统初始化配置复杂度和学习成本大小; 2) 用户的操作复杂度; 3) 身份认证系统错误拒绝率; 4) 认证系统硬件成本的高低. 身份认证系统应该同时兼顾这些方面, 才能带来良好的用户体验, 并收获广泛的用户群体. 以下将从实用性的角度提出现有身份认证系统的研究展望.

对于基于秘密知识的身份认证, 要始终关注用户记忆成本与秘密知识强度之间的平衡, 减少用户选择的倾向性, 研究和设计同时兼顾实用性和安全性的文本型和图案型身份认证部署方案. 在基于生理特征的身份认证系统中, 某些生理特征 (如虹膜、掌纹、心电图等) 需要额外的特殊硬件支持, 而这些硬件往往是中低端移动终端所不具备的, 导致对应的身份认证系统难以广泛部署和推广. 研究人员应该关注如何利用移动终端广泛配备的各类传感器采集可以用于身份认证的生理特征. 在基于行为特征的身份认证系统中, 除声纹认证之外, 其他几乎所有的行为特征认证系统存在准确率不够高, 系统鲁棒性不强的问题, 从而带来了不友好的用户体验. 因此, 需要进一步研究提高基于行为特征的身份认证准确率的方法, 这对于该类型身份认证系统的推广具有重要意义. 基于声感知的身份认证系统大都需要用户进行复杂初始化配置和适应性学习, 容易遭受多径传播和环境噪音干扰, 严重影响用户体验. 如何简化初始化配置和学习成本, 加快系统的学习过程, 降低操作复杂度, 增强系统鲁棒性和跨设备适应性, 是未来研究需要解决的问题.

5.2.3 兼顾安全性和实用性的双/多因素身份认证

双/多因素身份认证系统要求用户提供符合要求的多个身份认证凭据, 当且仅当每个身份凭据都通过认证时, 用户才能进入系统获得相应权限, 在一定程度上提高了身份认证系统的安全性. 正因其在安全性方面的表现明显优于单因素身份认证, 双/多因素身份认证正在逐渐成为身份认证系统的发展趋势. 但是目前的双/多因素身份认证系统大多只是对多个认证因子进行简单的逐一认证, 各认证因子之间关联性差, 攻击者可以利用现有的攻击手段逐个击破, 导致安全性的提升并不显著, 还会使得认证过程变得繁琐, 增加用户操作复杂度. 因此, 双/多因素身份认证系统无论在安全性还是实用性方面都有很大的提升空间. 目前, 基于声感知的双/多因素身份认证已经取得了初步的成果. 例如, LVID[106]利用语音提取声纹特征作为第 1 认证因子, 利用高频声信号捕获用户说话时嘴唇运动特征作为第 2 身份认证因子. 这对认证因子具有很强的关联性, 很容易进行有机融合组成多生物特征, 从而提高智能手机语音认证的安全性与鲁棒性. 研究人员应该以兼顾安全性和实用性为核心, 着重探索发现强关联的身份认证因子, 利用现有的硬件设备实现多模态异构数据的同源感知, 一体化提取多特征认证因子, 建立多认证因子有机融合的身份认证系统.

6 总 结

面对日益严峻的安全威胁, 实现安全可靠的移动终端身份认证是亟待解决的现实问题. 基于声感知的移动终端身份认证因其高度普适性和低硬件成本, 可以有效提高移动终端身份认证系统的安全性. 本文对移动终端身份认证和基于声感知的身份认证国内外研究进展进行了分类梳理, 提出了当前研究工作面临的挑战, 探讨了未来基于声感知的安全身份认证系统的发展趋势. 基于声感知的移动终端身份认证解决方案越来越多样化, 未来的研究重心将始终以提升安全性和实用性为目标, 逐渐向多因子有机融合的身份认证系统转移.

References:

[1] Ericsson. Ericsson mobility report. 2022. https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf

[2] Ye GX, Tang ZY, Fang DY, Chen XJ, Wolff W, Aviv AJ, Wang Z. A video-based attack for Android pattern lock. ACM Trans. on Privacy and Security, 2018, 21(4): 19. [doi: 10.1145/3230740]

[3] Chen DJ, Zhao ZH, Qin X, Luo YH, Cao MS, Xu H, Liu AF. MagLeak: A learning-based side-channel attack for password recognition with multiple sensors in IIoT environment. IEEE Trans. on Industrial Informatics, 2022, 18(1): 467–476. [doi: 10.1109/TII.2020.3045161]

[4] Yang E, Fang S, Markwood I, Liu Y, Zhao SQ, Lu Z, Zhu HJ. Wireless training-free keystroke inference attack and defense. IEEE/ACM Trans. on Networking, 2022, 30(4): 1733–1748. [doi: 10.1109/TNET.2022.3147721]

[5] Zhou M, Wang Q, Yang JX, Li Q, Jiang PP, Chen YJ, Wang ZB. Stealing your Android patterns via acoustic signals. IEEE Trans. on Mobile Computing, 2021, 20(4): 1656–1671. [doi: 10.1109/TMC.2019.2960778]

[6] Qin L, Peng F, Long M, Ramachandra R, Busch C. Vulnerabilities of unattended face verification systems to facial components-based presentation attacks: An empirical study. ACM Trans. on Privacy and Security, 2022, 25(1): 4. [doi: 10.1145/3491199]

[7] Rathore AS, Shen YJ, Xu CH, Snyderman J, Han JS, Zhang F, Li ZX, Lin F, Xu WY, Ren K. FakeGuard: Exploring haptic response to mitigate the vulnerability in commercial fingerprint anti-spoofing. In: Proc. of the 29th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2022. 1–17.

[8] Wang C, Wang Y, Chen YY, Liu HB, Liu J. User authentication on mobile devices: Approaches, threats and trends. Computer Networks, 2020, 170: 107118. [doi: 10.1016/j.comnet.2020.107118]

[9] Bai Y, Lu L, Cheng J, Liu J, Chen YY, Yu JD. Acoustic-based sensing and applications: A survey. Computer Networks, 2020, 181: 107447. [doi: 10.1016/j.comnet.2020.107447]

[10] Lu L, Yu JD, Li ML. Towards a real-time anti-theft method for mobile devices leveraging acoustic sensing. Chinese Journal of Computers, 2020, 43(10): 2002–2018 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2020.02002]

[11] Bonneau J, Preibusch S, Anderson R. A birthday present every eleven wallets? The security of customer-chosen banking pins. In: Proc. of the 16th Int'l Conf. on Financial Cryptography and Data Security. Kralendijk: Springer, 2012. 25–40. [doi: 10.1007/978-3-642-32946-3_3]

[12] Zhang Q, Wang D, Zhao R, Yu YG, Shen JJ. Sensing to hear: Speech enhancement for mobile devices using acoustic signals. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, 5(3): 137. [doi: 10.1145/3478093]

[13] Shi D, Tao D, Wang JT, Yao MY, Wang ZB, Chen HJ, Helal S. Fine-grained and context-aware behavioral biometrics for pattern lock on smartphones. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, 5(1): 33. [doi: 10.1145/3448080]

[14] Chauhan J, Hu YN, Seneviratne S, Misra A, Seneviratne A, Lee Y. BreathPrint: Breathing acoustics-based user authentication. In: Proc. of the 15th Annual Int'l Conf. on Mobile Systems, Applications, and Services. Niagara Falls: ACM, 2017. 278–291. [doi: 10.1145/3081333.3081355]

[15] Zhou B, Xie ZX, Zhang YN, Lohokare J, Gao RP, Ye F. Robust human face authentication leveraging acoustic sensing on smartphones. IEEE Trans. on Mobile Computing, 2022, 21(8): 3009–3023. [doi: 10.1109/TMC.2020.3048659]

[16] Karapanos N, Marforio C, Soriente C, Capkun S. Sound-proof: Usable two-factor authentication based on ambient sound. In: Proc. of the 24th USENIX Security Symp. Washington: USENIX Association, 2015. 483–498.

[17] Wang Q, Lin X, Zhou M, Chen YJ, Wang C, Li Q, Luo XY. VoicePop: A pop noise based anti-spoofing system for voice authentication on smartphones. In: Proc. of the 2019 IEEE Conf. on Computer Communications. Paris: IEEE, 2019. 2062–2070. [doi: 10.1109/INFOCOM.2019.8737422]

[18] Zhang LH, Tan S, Yang J. Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Dallas: ACM, 2017. 57–71. [doi: 10.1145/3133956.3133962]

[19] Zheng BL, Jiang PP, Wang Q, Li Q, Shen C, Wang C, Ge YJ, Teng QY, Zhang SY. Black-box adversarial attacks on commercial speech platforms with minimal information. In: Proc. of the 2021 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2021. 86–107. [doi: 10.1145/3460120.3485383]

[20] Yu JD, Lu L, Chen YY, Zhu YM, Kong LH. An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing. IEEE Trans. on Mobile Computing, 2021, 20(2): 337–351. [doi: 10.1109/TMC.2019.2947468]

[21] Zhou M, Wang Q, Yang JX, Li Q, Xiao F, Wang ZB, Chen XF. PatternListener: Cracking Android pattern lock using acoustic signals. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 1775–1787. [doi: 10.1145/3243734.3243777]

[22] Shirvanian M, Vo S, Saxena N. Quantifying the breakability of voice assistants. In: Proc. of the 2019 IEEE Int'l Conf. on Pervasive Computing and Communications. Kyoto: IEEE, 2019. 1–11. [doi: 10.1109/PERCOM.2019.8767399]

[23] Zhou M, Qin Z, Lin X, Hu SS, Wang Q, Ren K. Hidden voice commands: Attacks and defenses on the VCS of autonomous driving cars. IEEE Wireless Communications, 2019, 26(5): 128–133. [doi: 10.1109/MWC.2019.1800477]

[24] Wenger E, Bronckers M, Cianfarani C, Cryan J, Sha A, Zheng HT, Zhao BY. "Hello, it's me": Deep learning-based speech synthesis attacks in the real world. In: Proc. of the 2021 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2021. 235–251. [doi: 10.1145/3460120.3484742]

[25] Chen GK, Chenb S, Fan LL, Du XN, Zhao Z, Song F, Liu Y. Who is real bob? Adversarial attacks on speaker recognition systems. In: Proc. of the 2021 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2021. 694–711. [doi: 10.1109/SP40001.2021.00004]

[26] Uellenbeck S, Dürmuth M, Wolf C, Holz T. Quantifying the security of graphical passwords: The case of Android unlock patterns. In: Proc. of the 2013 ACM SIGSAC Conf. on Computer & Communications Security. Berlin: ACM, 2013. 161–172. [doi: 10.1145/2508859.2516700]

[27] Markert P, Bailey DV, Golla M, Dürmuth M, Aviv AJ. This PIN can be easily guessed: Analyzing the security of smartphone unlock PINs. In: Proc. of the 2020 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2020. 286–303. [doi: 10.1109/SP40000.2020.00100]

[28] Wang P, Wang D, Huang XY. Advances in password security. Journal of Computer Research and Development, 2016, 53(10): 2172–2188 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2016.20160483]

[29] Wang D, Zou YK, Tao Y, Wang B. Password guessing based on recurrent neural networks and generative adversarial networks. Chinese Journal of Computers, 2021, 44(8): 1519–1534. (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2021.01519]

[30] Lee MK. Security notions and advanced method for human shoulder-surfing resistant PIN-entry. IEEE Trans. on Information Forensics and Security, 2014, 9(4): 695–708. [doi: 10.1109/TIFS.2014.2307671]

[31] Shukla D, Kumar R, Serwadda A, Phoha VV. Beware, your hands reveal your secrets! In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 904–917. [doi: 10.1145/2660267.2660360]

[32] Khan H, Hengartner U, Vogel D. Evaluating attack and defense strategies for smartphone PIN shoulder surfing. In: Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems. Montreal: ACM, 2018. 164. [doi: 10.1145/3173574.3173738]

[33] von Zezschwitz E, De Luca A, Brunkow B, Hussmann H. SwiPIN: Fast and secure PIN-entry on smartphones. In: Proc. of the 33rd Annual ACM Conf. on Human Factors in Computing Systems. Seoul: ACM, 2015. 1403–1406. [doi: 10.1145/2702123.2702212]

[34] Krombholz K, Hupperich T, Holz T. Use the force: Evaluating force-sensitive authentication for mobile devices. In: Proc. of the 12th USENIX Conf. on Usable Privacy and Security. Denver: USENIX Association, 2016. 207–219. [doi: 10.5555/3235895.3235913]

[35] Yan Q, Han J, Li YJ, Zhou JY, Deng RH. Designing leakage-resilient password entry on touchscreen mobile devices. In: Proc. of the 8th ACM SIGSAC Symp. on Information, Computer and Communications Security. Hangzhou: ACM, 2013. 37–48. [doi: 10.1145/2484313.2484318]

[36] Li ZJ, Li M, Mohapatra P, Han JS, Chen SY. iType: Using eye gaze to enhance typing privacy. In: Proc. of the 2017 IEEE Conf. on Computer Communications. Atlanta: IEEE, 2017. 1–9. [doi: 10.1109/INFOCOM.2017.8057233]

[37] Alsuhibany SA. A camouflage text-based password approach for mobile devices against shoulder-surfing attack. Security and Communication Networks, 2021, 2021: 6653076. [doi: 10.1155/2021/6653076]

[38] Mayer P, Gerber N, Reinheimer B, Rack P, Braun K, Volkamer M. I (don't) see what you typed there! Shoulder-surfing resistant password entry on gamepads. In: Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems. Glasgow: ACM, 2019. 549. [doi: 10.1145/3290605.3300779]

[39] Castelluccia C, Dürmuth M, Perito D. Adaptive password-strength meters from Markov models. In: Proc. of the 19th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2012.

[40] Kelley PG, Komanduri S, Mazurek ML, Shay R, Vidas T, Bauer L, Christin N, Cranor LF, Lopez J. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In: Proc. of the 2012 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2012. 523–537. [doi: 10.1109/SP.2012.38]

[41] Dhamija R, Perrig A. Déjà VU—A user study using images for authentication. In: Proc. of the 9th USENIX Security Symp. Denver: USENIX Association, 2000. 4–4.

[42] Davis D, Monrose F, Reiter MK. On user choice in graphical password schemes. In: Proc. of the 13th USENIX Security Symp. San Diego: USENIX, 2004. 151–164.

[43] De Angeli A, Coutts M, Coventry L, Johnson GI, Cameron D, Fischer MH. VIP: A visual approach to user authentication. In: Proc. of the 2002 Working Conf. on Advanced Visual Interfaces. Trento: ACM, 2002. 316–323. [doi: 10.1145/1556262.1556312]

[44] Sun HP, Wang K, Li X, Qin N, Chen Z. PassApp: My APP is my password! In: Proc. of the 17th Int'l Conf. on Human-computer Interaction with Mobile Devices and Services. Copenhagen: ACM, 2015. 306–315. [doi: 10.1145/2785830.2785880]

[45] Jermyn I, Mayer A, Monrose F, Reiter MK, Rubin AD. The design and analysis of graphical passwords. In: Proc. of the 8th Conf. on USENIX Security Symp. Washington: USENIX Association, 1999. 1. [doi: 10.5555/1251421.1251422]

[46] Sherman M, Clark G, Yang YL, Sugrim S, Modig A, Lindqvist J, Oulasvirta A, Roos T. User-generated free-form gestures for authentication: Security and memorability. In: Proc. of the 12th Annual Int'l Conf. on Mobile Systems, Applications, and Services. Bretton Woods: ACM, 2014. 176–189. [doi: 10.1145/2594368.2594375]

[47] Sae-Bae N, Memon N. Online signature verification on mobile devices. IEEE Trans. on Information Forensics and Security, 2014, 9(6): 933–947. [doi: 10.1109/TIFS.2014.2316472]

[48] Cho G, Huh JH, Cho J, Oh S, Song Y, Kim H. SysPal: System-guided pattern locks for Android. In: Proc. of the 2017 IEEE Symp. on Security and Privacy. San Jose: IEEE, 2017. 338–356. [doi: 10.1109/SP.2017.61]

[49] Yao MY, Tao D. Implicit authentication mechanism of pattern unlock based on over-sampling and one-class classification for

smartphones. Computer Science, 2020, 47(11): 19–24 (in Chinese with English abstract). [doi: 10.11896/jsjkx.200600004]

[50] Forman T, Aviv A. Double patterns: A usable solution to increase the security of Android unlock patterns. In: Proc. of the 36th Annual Computer Security Applications Conf. Austin: ACM, 2020. 219–233. [doi: 10.1145/3427228.3427252]

[51] Munyendo CW, Grant M, Markert P, Forman TJ, Aviv AJ. Using a blocklist to improve the security of user selection of Android patterns. In: Proc. of the 17th USENIX Conf. on Usable Privacy and Security. 2021. 3. [doi: 10.5555/3563572.3563575]

[52] Dirik AE, Memon N, Birget JC. Modeling user choice in the PassPoints graphical password scheme. In: Proc. of the 3rd Symp. on Usable Privacy and Security. Pittsburgh: ACM, 2007. 20–28. [doi: 10.1145/1280680.1280684]

[53] Raghavendra R, Busch C, Yang B. Scaling-robust fingerprint verification with smartphone camera in real-life scenarios. In: Proc. of the 6th IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems. Arlington: IEEE, 2013. 1–8. [doi: 10.1109/BTAS.2013.6712736]

[54] Komeili M, Armanfard N, Hatzinakos D. Liveness detection and automatic template updating using fusion of ECG and fingerprint. IEEE Trans. on Information Forensics and Security, 2018, 13(7): 1810–1822. [doi: 10.1109/TIFS.2018.2804890]

[55] Park E, Cui XN, Nguyen THB, Kim H. Presentation attack detection using a tiny fully convolutional network. IEEE Trans. on Information Forensics and Security, 2019, 14(11): 3016–3025. [doi: 10.1109/TIFS.2019.2907184]

[56] Wu C, He K, Chen J, Zhao ZM, Du RY. Liveness is not enough: Enhancing fingerprint authentication with behavioral biometrics to defeat puppet attacks. In: Proc. of the 29th USENIX Security Symp. USENIX Association, 2020. 2219–2236.

[57] Taigman Y, Yang M, Ranzato MA, Wolf L. DeepFace: Closing the gap to human-level performance in face verification. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 1701–1708. [doi: 10.1109/CVPR.2014.220]

[58] Aswal V, Tupe O, Shaikh S, Charniya NN. Single camera masked face identification. In: Proc. of the 19th IEEE Int'l Conf. on Machine Learning and Applications. Miami: IEEE, 2020. 57–60. [doi: 10.1109/ICMLA51294.2020.00018]

[59] Li Y, Li YJ, Yan Q, Kong HC, Deng RH. Seeing your face is not enough: An inertial sensor-based liveness detection for face authentication. In: Proc. of the 22nd ACM SIGSAC Conf. on Computer and Communications Security. Denver: ACM, 2015. 1558–1569. [doi: 10.1145/2810103.2813612]

[60] Tian Y, Xiang SJ. LBP and multilayer DCT based anti-spoofing countermeasure in face liveness detection. Journal of Computer Research and Development, 2018, 55(3): 643–650 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2018.20160417]

[61] Liu AJ, Zhao CX, Yu ZT, Wan J, Su AY, Liu X, Tan ZC, Escalera S, Xing JL, Liang YY, Guo GD, Lei Z, Li SZ, Zhang D. Contrastive context-aware learning for 3D high-fidelity mask face presentation attack detection. IEEE Trans. on Information Forensics and Security, 2022, 17: 2497–2507. [doi: 10.1109/TIFS.2022.3188149]

[62] Genovese A, Piuri V, Plataniotis KN, Scotti F. PalmNet: Gabor-PCA convolutional networks for touchless palmprint recognition. IEEE Trans. on Information Forensics and Security, 2019, 14(12): 3160–3174. [doi: 10.1109/TIFS.2019.2911165]

[63] Song YP, Cai ZM, Zhang ZL. Multi-touch authentication using hand geometry and behavioral information. In: Proc. of the 2017 IEEE Symp. on Security and Privacy. San Jose: IEEE, 2017. 357–372. [doi: 10.1109/SP.2017.54]

[64] Ma Z, Yang YL, Liu XM, Liu Y, Ma SQ, Ren K, Yao C. EmIr-Auth: Eye movement and iris-based portable remote authentication for smart grid. IEEE Trans. on Industrial Informatics, 2020, 16(10): 6597–6606. [doi: 10.1109/TII.2019.2946047]

[65] Fahmi PA, Kodirov E, Choi DJ, Lee GS, Azli AMF, Sayeed S. Implicit authentication based on ear shape biometrics using smartphone camera during a call. In: Proc. of the 2012 IEEE Int'l Conf. on Systems, Man, and Cybernetics. Seoul: IEEE, 2012. 2272–2276. [doi: 10.1109/ICSMC.2012.6378079]

[66] Xiao J, Li SZ, Dong W, Li QF Hu F. An identity recognition method based on electrocardiograph and photoplethysmograph feature fusion. Journal of Electronics & Information Technology, 2021, 43(10): 3010–3017 (in Chinese with English abstract). [doi: 10.11999/JEIT200904]

[67] Zhang CL, Koishida K, Hansen JHL. Text-independent speaker verification based on triplet convolutional neural network embeddings. IEEE/ACM Trans. on Audio, Speech, and Language Processing, 2018, 26(9): 1633–1644. [doi: 10.1109/TASLP.2018.2831456]

[68] Wan L, Wang Q, Papir A, Moreno IL. Generalized end-to-end loss for speaker verification. In: Proc. of the 2018 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Calgary: IEEE, 2018. 4879–4883. [doi: 10.1109/ICASSP.2018.8462665]

[69] Yu LF, Liu Q. Research and application of deep recurrent neural networks based voiceprint recognition. Application Research of Computers, 2019, 36(1): 153–158 (in Chinese with English abstract). [doi: 10.19734/j.issn.1001-3695.2017.07.0661]

[70] Li HN, Xu CH, Rathore AS, Li ZX, Zhang HB, Song C, Wang K, Su L, Lin F, Ren K, Xu WY. VocalPrint: Exploring a resilient and secure voice authentication via mmWave biometric interrogation. In: Proc. of the 18th Conf. on Embedded Networked Sensor Systems. ACM, 2020. 312–325. [doi: 10.1145/3384419.3430779]

[71] Sprager S, Juric MB. An efficient HOS-based gait authentication of accelerometer data. IEEE Trans. on Information Forensics and

Security, 2015, 10(7): 1486–1498. [doi: 10.1109/TIFS.2015.2415753]

[72] Zou Q, Wang YL, Wang Q, Zhao Y, Li QQ. Deep learning-based gait recognition using smartphones in the wild. IEEE Trans. on Information Forensics and Security, 2020, 15: 3197–3212. [doi: 10.1109/TIFS.2020.2985628]

[73] Shi MH, Wang ZH. An interpretable gait recognition method based on time series features. Scientia Sinica Informationis, 2020, 50(3): 438–460 (in Chinese with English abstract). [doi: 10.1360/N112018-00326]

[74] van Hamme T, Rúa EA, Preuveneers D, Joosen W. On the security of biometrics and fuzzy commitment cryptosystems: A study on gait authentication. IEEE Trans. on Information Forensics and Security, 2021, 16: 5211–5224. [doi: 10.1109/TIFS.2021.3124735]

[75] Bo C, Zhang L, Li XY, Huang QY, Wang Y. SilentSense: Silent user identification via touch and movement behavioral biometrics. In: Proc. of the 19th Annual Int'l Conf. on Mobile Computing & Networking. Miami: ACM, 2013. 187–190. [doi: 10.1145/2500423.2504572]

[76] Sae-Bae N, Ahmed K, Isbister K, Memon N. Biometric-rich gestures: A novel approach to authentication on multi-touch devices. In: Proc. of the 2012 SIGCHI Conf. on Human Factors in Computing Systems. Austin: ACM, 2012. 977–986. [doi: 10.1145/2207676.2208543]

[77] Shahzad M, Zhang SH. Augmenting user identification with WiFi based gesture recognition. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(3): 134. [doi: 10.1145/3264944]

[78] Shi C, Liu J, Liu HB, Chen YY. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In: Proc. of the 18th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing. Chennai: ACM, 2017. 5. [doi: 10.1145/3084041.3084061]

[79] Chauhan J, Rajasegaran J, Seneviratne S, Misra A, Seneviratne A, Lee Y. Performance characterization of deep learning models for breathing-based authentication on resource-constrained devices. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(4): 158. [doi: 10.1145/3287036]

[80] Zhou QQ, Yang YN, Hong F, Feng Y, Guo ZW. User identification and authentication using keystroke dynamics with acoustic signal. In: Proc. of the 12th Int'l Conf. on Mobile Ad-hoc and Sensor Networks. Hefei: IEEE, 2016. 445–449. [doi: 10.1109/MSN.2016.082]

[81] Rathore AS, Zhu WJ, Daiyan A, Xu CH, Wang K, Lin F, Ren K, Xu WY. SonicPrint: A generally adoptable and secure fingerprint biometrics in smart devices. In: Proc. of the 18th Int'l Conf. on Mobile Systems, Applications, and Services. Toronto: ACM, 2020. 121–134. [doi: 10.1145/3386901.3388939]

[82] Zou YP, Lei HB, Wu KS. Beyond legitimacy, also with identity: Your smart earphones know who you are quietly. IEEE Trans. on Mobile Computing, 2023, 22(6): 3179–3192. [doi: 10.1109/TMC.2021.3134654]

[83] Gao Y, Jin YC, Chauhan J, Choi S, Li JY, Jin ZP. Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, 5(1): 12. [doi: 10.1145/3448113]

[84] Zhou Z, Diao WR, Liu XY, Zhang KH. Acoustic fingerprinting revisited: Generate stable device ID stealthily with inaudible sound. In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 429–440. [doi: 10.1145/2660267.2660300]

[85] Das A, Borisov N, Caesar M. Do you hear what I hear? Fingerprinting smart devices through embedded acoustic components. In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 441–452. [doi: 10.1145/2660267.2660325]

[86] Luo D, Korus P, Huang JW. Band energy difference for source attribution in audio forensics. IEEE Trans. on Information Forensics and Security, 2018, 13(9): 2179–2189. [doi: 10.1109/TIFS.2018.2812185]

[87] Tan JY, Wang XL, Nguyen CT, Shi Y. SilentKey: A new authentication framework through ultrasonic-based lip reading. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 36. [doi: 10.1145/3191768]

[88] Lu L, Yu JD, Chen YY, Liu HB, Zhu YM, Liu YF, Li ML. LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals. In: Proc. of the 2018 IEEE Conf. on Computer Communications. Honolulu: IEEE, 2018. 1466–1474. [doi: 10.1109/INFOCOM.2018.8486283]

[89] Lu L, Yu JD, Chen YY, Liu HB, Zhu YM, Kong LH, Li ML. Lip reading-based user authentication through acoustic sensing on smartphones. IEEE/ACM Trans. on Networking, 2019, 27(1): 447–460. [doi: 10.1109/TNET.2019.2891733]

[90] Wang YX, Chen YN, Bhuiyan ZA, Han Y, Zhao SH, Li JX. Gait-based human identification using acoustic sensor and deep neural network. Future Generation Computer Systems, 2018, 86: 1228–1237. [doi: 10.1016/j.future.2017.07.012]

[91] Ding F, Wang D, Zhang Q, Zhao R. ASSV: Handwritten signature verification using acoustic signals. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2019, 3(3): 80. [doi: 10.1145/3351238]

[92] Zhao R, Wang D, Zhang Q, Jin XY, Liu K. Smartphone-based handwritten signature verification using acoustic signals. Proc. of the

ACM on Human-computer Interaction, 2021, 5(ISS): 499. [doi: 10.1145/3488544]

[93] Gao Y, Wang W, Phoha VV, Sun W, Jin ZP. EarEcho: Using ear canal echo for wearable authentication. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2019, 3(3): 81. [doi: 10.1145/3351239]

[94] Lu L, Yu JD, Chen YY, Wang Y. VocalLock: Sensing vocal tract for passphrase-independent user authentication leveraging acoustic signals on smartphones. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(2): 51. [doi: 10.1145/3397320]

[95] Huang L, Wang C. Notification privacy protection via unobtrusive gripping hand verification using media sounds. In: Proc. of the 27th Annual Int'l Conf. on Mobile Computing and Networking. New Orleans: ACM, 2021. 491–504. [doi: 10.1145/3447993.3483277]

[96] Shrestha B, Shirvanian M, Shrestha P, Saxena N. The sounds of the phones: Dangers of zero-effort second factor login based on ambient audio. In: Proc. of the 2016 ACM SIGSAC Conf. on Computer and Communications Security. Vienna: ACM, 2016. 908–919. [doi: 10.1145/2976749.2978328]

[97] Shrestha P, Shrestha B, Saxena N. Home alone: The insider threat of unattended wearables and a defense using audio proximity. In: Proc. of the 2018 IEEE Conf. on Communications and Network Security. Beijing: IEEE, 2018. 1–9. [doi: 10.1109/CNS.2018.8433216]

[98] Shrestha P, Saxena N. Listening watch: Wearable two-factor authentication using speech signals resilient to near-far attacks. In: Proc. of the 11th ACM Conf. on Security & Privacy in Wireless and Mobile Networks. Stockholm: ACM, 2018. 99–110. [doi: 10.1145/3212480.3212501]

[99] Ren YZ, Wen P, Liu HB, Zheng ZR, Chen YY, Huang PC, Li HW. Proximity-echo: Secure two factor authentication using active sound sensing. In: Proc. of the 2021 IEEE Conf. on Computer Communications. Vancouver: IEEE, 2021. 1–10. [doi: 10.1109/INFOCOM42981.2021.9488866]

[100] Zhou B, Lohokare J, Gao RP, Ye F. EchoPrint: Two-factor authentication using acoustics and vision on smartphones. In: Proc. of the 24th Annual Int'l Conf. on Mobile Computing and Networking. New Delhi: ACM, 2018. 321–336. [doi: 10.1145/3241539.3241575]

[101] Chen HJ, Li F, Du W, Yang S, Conn M, Wang Y. Listen to your fingers: User authentication based on geometry biometrics of touch gesture. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(3): 75. [doi: 10.1145/3411809]

[102] Chen YL, Ni T, Xu WT, Gu T. SwipePass: Acoustic-based second-factor user authentication for smartphones. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2022, 6(3): 106. [doi: 10.1145/3550292]

[103] Zhou M, Wang Q, Lin X, Zhao Y, Jiang PP, Li Q, Shen C, Wang C. PressPIN: Enabling secure PIN authentication on mobile devices via structure-borne sounds. IEEE Trans. on Dependable and Secure Computing, 2023, 20(2): 1228–1242. [doi: 10.1109/TDSC.2022.3151889]

[104] Wu C, Chen J, He K, Zhao ZM, Du RY, Zhang C. EchoHand: High accuracy and presentation attack resistant hand authentication on commodity mobile devices. In: Proc. of the 2022 ACM SIGSAC Conf. on Computer and Communications Security. Los Angeles: ACM, 2022. 2931–2945. [doi: 10.1145/3548606.3560553]

[105] Han DQ, Chen YM, Li T, Zhang R, Zhang YC, Hedgpeth T. Proximity-proof: Secure and usable mobile two-factor authentication. In: Proc. of the 24th Annual Int'l Conf. on Mobile Computing and Networking. New Delhi: ACM, 2018. 401–415. [doi: 10.1145/3241539.3241574]

[106] Wu LB, Yang JX, Zhou M, Chen YJ, Wang Q. LVID: A multimodal biometrics authentication system on smartphones. IEEE Trans. on Information Forensics and Security, 2020, 15: 1572–1585. [doi: 10.1109/TIFS.2019.2944058]

[107] Yang YL, Wang Y, Chen YY, Wang C. EchoLock: Towards low-effort mobile user identification leveraging structure-borne echos. In: Proc. of the 15th ACM Asia Conf. on Computer and Communications Security. Taipei: ACM, 2020. 772–783. [doi: 10.1145/3320269.3384741]

[108] Zhang SH, Das A. HandLock: Enabling 2-FA for smart home voice assistants using inaudible acoustic signal. In: Proc. of the 24th Int'l Symp. on Research in Attacks, Intrusions and Defenses. San Sebastian: ACM, 2021. 251–265. [doi: 10.1145/3471621.3471866]

[109] Liu D, Wang Q, Zhou M, Jiang PP, Li Q, Shen C, Wang C. SoundID: Securing mobile two-factor authentication via acoustic signals. IEEE Trans. on Dependable and Secure Computing, 2023, 20(2): 1687–1701. [doi: 10.1109/TDSC.2022.3162718]

[110] Wang ZF, Wei G, He QH. Channel pattern noise based playback attack detection algorithm for speaker recognition. In: Proc. of the 2011 Int'l Conf. on Machine Learning and Cybernetics. Guilin: IEEE, 2011. 1708–1713. [doi: 10.1109/ICMLC.2011.6016982]

[111] Zhang LH, Tan S, Yang J, Chen YY. Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones. In: Proc. of the 2016 ACM SIGSAC Conf. on Computer and Communications Security. Vienna: ACM, 2016. 1080–1091. [doi: 10.1145/2976749.2978296]

[112] Shang JC, Chen S, Wu J. Defending against voice spoofing: A robust software-based liveness detection system. In: Proc. of the 15th IEEE Int'l Conf. on Mobile Ad Hoc and Sensor Systems. Chengdu: IEEE, 2018. 28–36. [doi: 10.1109/MASS.2018.00016]

[113] Jiang PP, Wang Q, Lin X, Zhou M, Ding WB, Wang C, Shen C, Li Q. Securing liveness detection for voice authentication via pop noises. IEEE Trans. on Dependable and Secure Computing, 2023, 20(2): 1702–1718. [doi: 10.1109/TDSC.2022.3163024]

[114] Wang Y, Cai WD, Gu T, Shao W, Li YN, Yu Y. Secure your voice: An oral airflow-based continuous liveness detection for voice assistants. Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 3(4): 157. [doi: 10.1145/3369811]

[115] Yan C, Long Y, Ji XY, Xu WY. The catcher in the field: A fieldprint based spoofing detection for text-independent speaker verification. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: ACM, 2019. 1215–1229. [doi: 10.1145/3319535.3354248]

[116] Ahmed ME, Kwak IY, Huh JH, Kim I, Oh T, Kim H. VOID: A fast and light voice liveness detection system. In: Proc. of the 29th USENIX Security Symp. USENIX Association, 2020. 2685–2702.

[117] Li ZH, Shi C, Zhang TF, Xie Y, Liu J, Yuan B, Chen YY. Robust detection of machine-induced audio attacks in intelligent audio systems with microphone array. In: Proc. of the 2021 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2021. 1884–1899. [doi: 10.1145/3460120.3484755]

[118] Meng Y, Li JC, Pillari M, Deopujari A, Brennan L, Shamsie H, Zhu HJ, Tian Y. Your microphone array retains your identity: A robust voice liveness detection system for smart speakers. In: Proc. of the 31st USENIX Security Symp. Boston: USENIX Association, 2022. 1077–1094.

[119] Zhang LH, Tan S, Chen YY, Yang J. A continuous articulatory-gesture-based liveness detection for voice authentication on smart devices. IEEE Internet of Things Journal, 2022, 9(23): 23320–23331. [doi: 10.1109/JIOT.2022.3199995]

[120] Zhang LH, Tan S, Wang Z, Ren YL, Wang Z, Yang J. VibLive: A continuous liveness detection for secure voice user interface in IoT environment. In: Proc. of the 36th Annual Computer Security Applications Conf. Austin: ACM, 2020. 884–896. [doi: 10.1145/3427228.3427281]

[121] Chen HX, Wang W, Zhang J, Zhang Q. EchoFace: Acoustic sensor-based media attack detection for face authentication. IEEE Internet of Things Journal, 2019, 7(3): 2152–2159. [doi: 10.1109/JIOT.2019.2959203]

[122] Zhou M, Wang Q, Li Q, Zhou WY, Yang JX, Shen C. Securing face liveness detection on mobile devices using unforgeable lip motion patterns. IEEE Trans. on Mobile Computing, 2024, 23(10): 9772–9788. [doi: 10.1109/TMC.2024.3367781]

[123] Kong CQ, Zheng KX, Wang SQ, Rocha A, Li HL. Beyond the pixel world: A novel acoustic-based face anti-spoofing system for smartphones. IEEE Trans. on Information Forensics and Security, 2022, 17: 3238–3253. [doi: 10.1109/TIFS.2022.3202115]

[124] Levalle Y. Bypassing biometric systems with 3D printing. 2020. https://www.youtube.com/watch?v=hJ35ApLKpN4

附中文参考文献:

[10] 卢立, 俞嘉地, 李明禄. 基于声波感知的移动设备实时防窃方法研究. 计算机学报, 2020, 43(10): 2002–2018. [doi: 10.11897/SP.J.1016.2020.02002]

[28] 王平, 汪定, 黄欣沂. 口令安全研究进展. 计算机研究与发展, 2016, 53(10): 2172–2188. [doi: 10.7544/issn1000-1239.2016.20160483]

[29] 汪定, 邹云开, 陶义, 王彬. 基于循环神经网络和生成式对抗网络的口令猜测模型研究. 计算机学报, 2021, 44(8): 1519–1534. [doi: 10.11897/SP.J.1016.2021.01519]

[49] 姚沐言, 陶丹. 基于上采样单分类的智能手机手势密码隐式身份认证机制. 计算机科学, 2020, 47(11): 19–24. [doi: 10.11896/jsjkx.200600004]

[60] 田野, 项世军. 基于 LBP 和多层 DCT 的人脸活体检测算法. 计算机研究与发展, 2018, 55(3): 643–650. [doi: 10.7544/issn1000-1239.2018.20160417]

[66] 肖剑, 李思卓, 董威, 李清华, 胡芳. 基于心电与光电容积脉搏波特征层融合的身份识别方法. 电子与信息学报, 2021, 43(10): 3010–3017. [doi: 10.11999/JEIT200904]

[69] 余玲飞, 刘强. 基于深度循环网络的声纹识别方法研究及应用. 计算机应用研究, 2019, 36(1): 153–158. [doi: 10.19734/j.issn.1001-3695.2017.07.0661]

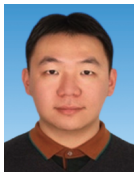
[73] 施沫寒, 王志海. 一种基于时间序列特征的可解释步态识别方法. 中国科学: 信息科学, 2020, 50(3): 438–460. [doi: 10.1360/N112018-00326]

周满(1993－), 男, 博士, 副研究员, CCF 专业会员, 主要研究领域为身份认证安全, 移动终端安全, 智能系统安全.

李向前(2002－), 男, 本科生, CCF 专业会员, 主要研究领域为语音增强.

王骞(1980－), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为人工智能安全, 数据存储, 查询及计算外包安全, 隐私保护, 无线系统安全, 应用密码学.

李琦(1979－), 男, 博士, 副教授, 博士生导师, CCF 高级会员, 主要研究领域为互联网安全, 人工智能安全.



沈超(1985－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为信息物理融合系统优化与安全, 网络和系统安全, 人工智能安全.

周雨庭(2000－), 男, 硕士生, 主要研究领域为身份认证安全.