陈光科

Phone: (+86) 191-2170-2230 Email: chengk@shanghaitech.edu.cn
Website: guangkechen.site

教育经历

上海科技大学 2021 年 9 月 - 至今
计算机科学与技术 博士研究生 导师: 宋富

- GPA: 专业 3.91 / 4.0 总 3.8 / 4.0
- 部分课程: 凸优化 (A+) 深度学习 (A) 密码学 (A) 算法设计与分析 (A) 计算理论 (A+)

上海科技大学 2019 年 9 月 - 2021 年 7 月
计算机科学与技术 硕士研究生 导师: 宋富

- GPA: 专业 4.0 / 4.0 总 3.9 / 4.0

华南理工大学，广州 2015 年 9 月 - 2019 年 7 月
信息工程 学士；本科

- GPA: 3.77 / 4.0
- 毕业论文: 说话人识别系统对抗攻击安全研究

荣誉奖项

- 2020，2022 | 三好学生 | 上海科技大学
- 2020 | 硕士研究生国家奖学金 | 上海科技大学
- 2018 | 校级十大三好学生提名（共 20 名）| 华南理工大学
- 2018 | 本科生国家奖学金 | 华南理工大学
- 2018 | 国家级大学生创新创业训练项目优秀结题（项目负责人）| 华南理工大学
- 2017 | 国家励志奖学金 | 华南理工大学
- 2016 | 企业奖学金 | 华南理工大学
- 2016, 2017, 2018 | 三好学生 | 华南理工大学

研究兴趣

机器学习安全及隐私 (如声纹识别和语音识别系统)

多媒体安全及隐私 (如语音)

论文

第一作者:

1. Who is Real Bob? Adversarial Attacks on Speaker Recognition Systems
Guangke Chen, Sen Chen, Lingling Fan, Xiaoning Du, Fu Song, Yang Liu
In Proc. of the 42nd IEEE Symposium on Security and Privacy (Oakland, S& P), 55-72, 2021
(CCF-A, Accept rate: 115/952=12%, **citation>120**)

2. QFA2SR: Query-Free Adversarial Transfer to Speaker Recognition Systems
   Guanke Chen, Yedi Zhang, Zhe Zhao, Fu Song
   USENIX Security Symposium 2023 (CCF-A)

3. AS2T: Arbitrary source-to-target adversarial attack on speaker recognition systems
   Guangke Chen, Zhe Zhao, Fu Song, Sen Chen, Lingling Fan, Yang Liu
   IEEE Transactions on Dependable and Secure Computing (TDSC)
   (CCF-A, IF=6.791)

4. Towards Understanding and Mitigating Audio Adversarial Examples for Speaker Recognition
   Guangke Chen, Zhe Zhao, Fu Song, Sen Chen, Lingling Fan, Feng Wang, Jiashui Wang
   IEEE Transactions on Dependable and Secure Computing (TDSC)
   (CCF-A, IF=6.791)

其他：

1. Attack as Defense: Characterizing Adversarial Examples using Robustness
   Zhe Zhao, **Guangke Chen**, Jingyi Wang, Yiwei Yang, Fu Song, Jun Sun
   In Proc. of the 30th International Symposium on Software Testing and Analysis (ISSTA), 42-55, 2021
   (CCF-A, Accept rate: 51/219=23%)

2. Attack as Detection: Using Adversarial Attack Methods to Detect Abnormal Examples
   Zhe Zhao, **Guangke Chen**, Tong Liu, Taishan Li, Fu Song, Jingyi Wang, Jun Sun
   Under review

3. BDD4BNN: A BDD-based Quantitative Analysis Framework for Binarized Neural Networks
   Yedi Zhang, Zhe Zhao, **Guangke Chen**, Fu Song, Taolue Chen
   In Proc. of the 33rd International Conference on Computer-Aided Verification (CAV), 175-200, 2021
   (CCF-A, Accept rate: 79/290=27%)

4. QVIP: An ILP-based Formal Verification Approach for Quantized Neural Networks
   Yedi Zhang, Zhe Zhao, **Guangke Chen**, Fu Song, Min Zhang, Taulue Chen, Jun Sun
   To appear in the 37th IEEE/ACM International Conference on Automated Software Engineering (ASE) 2022. (CCF-A)

5. ACROBAT: Accelerating CEGAR-based Neural Network Verification via Adversarial Attacks
   Zhe Zhao, Yedi Zhang, **Guangke Chen**, Fu Song, Taolue Chen and Jiaxiang Liu
   To appear in the Proc. of the 29th Static Analysis Symposium (SAS) 2022 (CCF-B)

6. Precise Quantitative Analysis of Binarized Neural Networks: A BDD-based Approach
   Yedi Zhang, Zhe Zhao, **Guangke Chen**, Fu Song, Taolue Chen
   ACM Transactions on Software Engineering and Methodology (TOSEM, CCF-A)

专利

授权：

1. 一种基于深度学习的非常态语音区别方法
   奉小慧, **陈光科**, 贺前华, 巫小兰, 李艳雄
   发明专利授权，授权号：CN108766419B，授权日期：2020.10.27

2. 脉率变异性和睡眠质量融合的心理压力监测方法及装置
   邢晓芬, **陈光科**, 江士尧, 林立韬, 陈东华
   发明专利授权，授权号：CN107874750B，授权日期：2020.01.10

公开/受理：

1. 基于语音声学特征压缩的语音对抗样本防御方法及应用
   宋富，**陈光科**，赵哲
   发明专利实质审查，公开号：CN114242083A，公开日期：2022-03-25

2. 一种基于攻击成本的对抗样本检测方法
   宋富，赵哲，**陈光科**
   发明专利实质审查，公开号: CN112381152A，公开日期: 2021.02.19

3. 一种基于样本鲁棒性差异的对抗样本检测方法
   宋富，赵哲，**陈光科**
   发明专利实质审查，公开号: CN112381150A，公开日期: 2021.02.19

项目经历

声纹识别模型无查询黑盒对抗攻击 2022 年 5 月 - 2022 年 10 月

研究内容：提出了一种新的无需查询的黑盒攻击方法。

- 利用语音对抗样本的迁移性，并提出三种方法不断提高语音对抗样本的迁移性。所提出的攻击无需知道目标模型的任何信息，生成对抗样本时不需要与模型进行任何查询交互，大大地提高对声纹识别模型的语音对抗样本攻击在现实场景下的实用性。
- 对微软、天聪智能、科大讯飞和京东的声纹识别 API 分别取得了 82.8%-99.5%、27.4%-55%、39.5%-70% 和 66%-96% 的攻击成功率，比现有攻击在迁移攻击设置下的攻击成功率高 20.9%-70.7%。
- 对天猫精灵、Google Assistant 和 Apple Siri 这几个支持声纹识别的语音助手分别取得 70%，60% 和 46% 的攻击成功率。

研究成果：

- 相关科研成果被计算机安全四大顶级会议的 Usenix Security 2023 接收（第一作者）。
- 漏洞汇报获得厂商感谢，其中科大讯飞认定属于中等漏洞等级，并给予 1000 元现金奖励。

声纹识别模型的对抗鲁棒性系统全面评估和测试框架 2021 年 12 月 - 2022 年 7 月

研究内容：设计能对声纹识别模型的对抗鲁棒性进行系统和全面的评估和测试的框架。

- 评估和测试覆盖不同子任务以及不同类别组合（场景）。声纹识别涉及三种不同子任务，且它们有各自适用的场景，而这三种子任务的识别和决策机制存在差异；和图像分类等闭集任务不同，声纹识别部分子任务的开集属性使得攻击的原始说话人和目标说话人所属的类别组合众多。现实场景下，攻击者为了达到不同的攻击目的，会采用不同的类别组合。因此，安全评估框架有必要涵盖三种子任务和不同的类别组合。已有攻击在这方面存在局限性，即只能对某种子任务或某种类别的说话人进行测试。针对这一点，本课题为不同子任务以及原始说话人和目标说话人的类别组合设计了子任务依赖以及面向类别组合的目标函数，这些设计的目标函数不仅适用于对应场景，从而能有效产生语音对抗样本，还确保了设计的目标函数相比其他已有目标函数性能优异，从而能高效产生对抗样本，达到了对不同场景下声纹识别系统的安全性进行测试。已有方案只能对部分场景进行测试，而本方案能支持所有场景的测试，场景覆盖率提升到 100%。在某些场景下，已有攻击采用的目标函数攻击成功率基本为 0，远远低于本方案所设计的损失函数的成功率（接近 100%）。
- 大规模物理评估和测试。对于暴露给外部的接口类型是物理信道形式的待评估声纹识别系统，进行测试时，音频需要由播放器播放，在空气介质中传播，再被麦克风接收。该过程涉及播放和录制等繁琐的人工劳动，大大限制了评估和测试的效率和规模。本课题设计不同转换函数对物理信道中的各种干扰因素进行建模，构建模拟物理信道，避免了需要大量人力进行测试的缺点，达到了自动化测试评估以及便于开展大规模测试评估的效果。进行物理信道测试时，本方案对 1000 个语音样本进行测试耗时不到 10 分钟，而已有方案测试 100 个语音样本耗时超过 30 分钟。

研究成果：相关科研成果被计算机安全顶级期刊 TDSC 接收（第一作者；CCF-A, SCI, EI, CAS-JCR-Q1）。

声纹识别模型语音对抗样本防御 2020 年 4 月 - 2022 年 11 月

研究内容：研究如何防御语音对抗样本，提高声纹识别模型的对抗鲁棒性。

- 现有防御评估。对现有防御在自适应攻击下的防御性能进行了测试，得出若干有助于防御选取和部署的发现和结论。

- 基于语音声学特征压缩的防御方法。针对声纹识别仍依赖于人工设计的语音声学特征这一特点（和视觉系统的重大差异），提出了一种基于语音声学特征压缩的防御算法。所提出防御的主要特征在于在语音声学特征层级进行变换，区别于已有防御直接在语音波形进行变换。通过与对抗训练结合，所提出的防御算法能将对抗鲁棒性提高 13% 以上，优于已有防御 7% 以上，且将攻击开销增加两个数量级。
- 开源平台。开发和开源了基于 pytorch 的声纹识别安全性评估平台 SpeakerGuard, 集成了主流声纹识别模型、数据集、白盒以及黑盒算法、防御方法、自适应攻击技术以及客观听觉测试指标等。

研究成果：

- 相关科研成果被计算机安全顶级期刊 TDSC 接收（第一作者；CCF-A, SCI, EI, CAS-JCR-Q1）。
- 所提出防御方法已申请发明专利（导师外第一发明人；实质审查中）。

声纹识别模型基于查询的黑盒对抗攻击 2019 年 1 月 - 2020 年 4 月

研究内容：针对声纹识别模型，提出了一种符合现实攻击场景的基于查询的黑盒语音对抗样本攻击。

- 阈值估计算法。声纹识别模型决策时依赖于预设的阈值，但黑盒场景下攻击者不知道阈值信息，而阈值和攻击成功与否紧密相关，为此，提出了一种能精确估计阈值的算法。算法得到的估计阈值严格大于真实阈值，且和真实阈值差距小于 0.03。
- 梯度估计算法。黑盒场景下攻击者无法访问模型内部信息获取梯度，为此，使用效率和准确性较高的基于遗传算法的梯度估计算法。所提出的攻击能攻破商用的声纹识别系统，如在天聪智能（声纹识别领头公司）上取得了 100% 攻击成功率，平均查询次数仅为 2500 次。

研究成果：相关科研成果被计算机安全四大顶级会议的 IEEE S&P 2021 接收（共同一作排序第一；非同等贡献），目前论文引用大于 120 次。IEEE S&P 2021 国内只有 8 所机构的 9 篇文章入选。

学术服务

- Program Committee Member:
  - the 23rd International Conference on Information and Communications Security (ICICS 2021)
  - the 24th International Conference on Information and Communications Security (ICICS 2022)
- Artifact Evaluation Committee Member:
  - Usenix Security 2023
- Session Chair:
  - Session 8 (Attack and Vulnerability Analysis II) of ICICS 2022
- Reviewer:
  - the 24th ISCA INTERSPEECH Conference (InterSpeech 2023)
  - IEEE Transactions on Dependable and Secure Computing (TDSC) (x2)
  - Springer Cybersecurity 2023
  - ACM Transactions on Privacy and Security
- Sub-reviewer:
  - The 33rd IEEE International Symposium on Software Reliability Engineering (ISSRE 2022)
  - IEEE Transactions on Reliability (TR) 2022
- Teaching Assistant:
  - CS240 Algorithm Design and Analysis, ShanghaiTech University, 2020-2021, Spring Semester

技能及证书

- 语言: 大学英语四级（621）；大学英语六级（520）
- 编程语言: Python, Matlab, Java, C/C++, LaTeX
- 技术:
  机器学习 (NumPy, Scipy, Scikit-learn)

深度学习 (PyTorch, TensorFlow)
语音信号处理 (Kaldi, torchaudio, SpeechBrain)