赵 哲

上海市浦东新区中科路 1 号上海科技大学

zhaozhe1@shanghaitech.edu.cn, Google Scholar, Homepage, Tel:+86 18816535610, +65 83448534

教育经历

上海科技大学 **上海**

博士研究生（硕博连读）：计算机软件与理论 2018 年 9 月 至今

- 个人研究领域为 AI 安全，包括： 导师：宋富教授
  - 神经网络测试与验证、对抗样本生成与防御；
  - 构造更为鲁棒、更易于验证的人工智能系统。
- GPA: 3.96/4.0
- 就读期间获**国家奖学金**，校级三好学生（前 5%），卓越助教等。

Singapore Management University 新加坡管理大学 **新加坡**

Research Assistant（访问学生）： 2022 年 7 月 至今

- RISE Lab 智能软件工程实验室 PI：孙军教授

上海科技大学 **上海**

访问学生 2018 年 3 月 - 2018 年 9 月

中国海洋大学 (985，211) **山东，青岛**

工学学士，国家保密学院：计算机科学与技术 2012 年 8 月至 2016 年 6 月

- 曾获**国家励志奖学金**，市级优秀志愿者，校级社会实践奖学金、优秀学生、优秀学生干部、优秀毕业生等

工作经历

Hewlett-Packard（HP，惠普） 2016 年 7 月 - 2017 年 11 月

测试开发工程师，负责自动化测试、性能测试等 上海

论文发表

博士期间已发表八篇高质量学术论文，其中三篇为第一作者（含共同一作），另有一篇一作论文在投。

一作文章如下（按照时间排序）：

3. **[CCF-B] Zhe Zhao**, Yedi Zhang, Guangke Chen, Fu Song, Taolue Chen and Jiaxiang Liu. *Accelerating CEGAR-based Neural Network Verification via Adversarial Attacks.* SAS 2022. 该论文旨在使用对抗样本生成技术加速神经网络验证，选取 PGD attack 和基于反例引导的抽象精化方法为实例，利用测试方法大幅减少了对于验证工具的调用次数，从而加速整个验证过程。

2. **[CCF-A] Zhe Zhao**, Guangke Chen, Jingyi Wang, Yiwei Yang, Fu Song, Jun Sun. *Attack as Defense: Characterizing Adversarial Examples using Robustness.* ISSTA 2021. 该论文聚焦于对抗样本检测，使用对抗样本攻击成本表征正常样本与异常样本中的鲁棒性差异并区分两者。该方法对决策边界附近的对抗样本检测效果优异，在与大扰动检测方法、对抗训练等方法结合后，可以抵抗 adaptive attack。该论文发表于软件工程顶级会议 ISSTA 21。

1. **[CCF-A]** Lei Bu[†], **Zhe Zhao**[†], Yuchao Duan, Fu Song. *Taking Care of The Discretization Problem: A Comprehensive Study of the Discretization Problem and A Black-Box Adversarial Attack in Discrete Integer Domain.* TDSC, [†]co-first author. 该论文首先对对抗样本的离散化问题进行了系统的分析和研究，并提出了几种可能的缓解办法，随后提出了一种基于搜索优化的黑盒攻击方法，该方法可以在自定义的搜索空间内高效寻找对抗样本，从而完全避免了离散化问题，且性能与其他实数域的对抗攻击方法相当。该论文发表于计算机安全领域顶刊 TDSC。

其他文章（按照时间排序）：

5. **[CCF-A]** Yedi Zhang, **Zhe Zhao**, Guangke Chen, Fu Song, Min Zhang, Taolue Chen. *Precise Quantitative Analysis of Binarized Neural Networks: A BDD-based Approach.* TOSEM 2022

4. [**CCF-A**] Yedi Zhang, **Zhe Zhao**, Guangke Chen, Fu Song, Min Zhang, Taolue Chen, Jun Sun. *QVIP: An ILP-based Formal Verification Approach for Quantized Neural Networks.* ASE 2022

3. [**CCF-A**] Guangke Chen, **Zhe Zhao**, Fu Song, Sen Chen, Lingling Fan, Yang Liu. *AS2T: Arbitrary Source-To-Target Adversarial Attack on Speaker Recognition Systems.* TDSC, early-access

2. [**CCF-A**] Yedi Zhang, **Zhe Zhao**, Guangke Chen, Fu Song, Taolue Chen. *BDD4BNN: A BDD-based Quantitative Analysis Framework for Binarized Neural Networks.* CAV 2021

1. [**CCF-A**] Guangke Chen, Sen Chen, Lingling Fan, Xiaoning Du, **Zhe Zhao**, Fu Song, Yang Liu. *Who is Real Bob? Adversarial Attacks on Speaker Recognition Systems.* S&P Oakland 2021

科研竞赛与荣誉称号

- CVPR 2022：真实世界对抗样本检测竞赛 第五名 — 2022 年 6 月
  竞赛主要内容为对抗样本检测，需要在黑盒条件下依靠图片本身信息完成图片分类。
- **国家奖学金** — 2021 年 12 月
- 上海科技大学 三好学生 — 2021 年 12 月
- OPPO 安全 AI 挑战赛 优胜奖 — 2021 年 12 月
  竞赛主要内容为黑盒人脸识别系统攻击，在 2000 多支队伍中排名前 10。
- ACM MM 2021：针对电商标识检测的鲁棒性防御比赛 **第三名** — 2021 年 7 月
  竞赛主要内容为鲁棒神经网络训练，复赛第一名，总分第三名。
- CVPR 2021：防御模型的白盒对抗攻击竞赛 **第三名** — 2021 年 3 月
  竞赛主要内容为白盒对抗样本攻击，需要在有限的时间成本内针对对抗防御后的模型进行攻击。
- 百度 AI 安全对抗赛 **第一名** — 2019 年 12 月
  竞赛主要内容为黑盒对抗样本攻击，考察对抗样本的隐蔽性和迁移性。
- 优秀毕业生 — 2015 年 - 2016 学年度
- **国家励志奖学金**、学习优秀奖学金、优秀学生 — 2014 年 - 2015 学年度
- 优秀学生干部 — 2013 年 - 2014 学年度
- 社会实践奖学金 — 2012 年 - 2013 学年度

志愿服务

程序委员会 (AEC)	OSDI 2022, USENIX ATC 2022, ISSTA 2022
论文审稿人	ISSRE 2021, ICICS 2021, CAV 2020, ICECCS 2022 2020 2019
学生志愿者	ISSTA 2019
助教	曾任 CS132（软件工程）课程助教，获评信息学院卓越助教奖