张景尧

jzhan502@ucr.edu | 谷歌学术 | 个人主页 | GitHub | 领英 | Riverside, USA

研究兴趣：致力于提升**隐私保护系统**在**软件**和**硬件**层面的**性能**。

教育背景

加州大学河滨分校 (University of California, Riverside) Riverside, USA
博士研究生 计算机科学, GPA: 3.7/4.0 2021 年 9 月 - 至今
导师：*Elaheh Sadredini*

西安电子科技大学 (Xidian University) 西安, 中国
硕士 电子与通信工程, GPA: 3.7/4.0；优秀毕业论文奖 2018 年 9 月 - 2021 年 6 月
学士 通信工程, GPA: 3.7/4.0；教改班 (前 5% / 800+) 2014 年 9 月 - 2018 年 6 月

工作经历

英伟达 (Nvidia), 数据中心服务器平台安全部门 *Mentor: Lohith Rangappa*
暑期实习生 2024 年 6 月 - 2024 年 9 月

- 开发了 BMC 辅助证明功能，以支持基于 SPDM 的非 PLDM 设备远程证明。
- 通过将密钥清单 (Key Manifest) 集成到外部信任根 (External Root of Trust)，为非 PLDM 设备实现了安全固件更新。

研究与开源项目

阿里云, 操作系统实验室 *Mentor: 乾越*
外部开发者 2023 年 8 月 - 2024 年 11 月

- 开发了一个安全的 LLM 推理平台，其可信网络网关 (Trusted Network Gate) 支持 HTTP/WebSocket。
- 为 CoCo KBS 设计并开发了一个“组合证明”(Combined Attestation) 框架，以支持 GPU 证明。[repo] [demo]
- 对用于 Nvidia 机密计算的 nvTrust 进行了研究，包括验证和证明。
- 调研了现有的支持 GPU 机密计算的系统，例如 Azure Confidential AI。

中国科学院, 开源之夏 *Mentor: 马丁 & 张家乐*
项目开发者 2023 年 & 2024 年

- 构建了Confidential-AI框架，该框架基于Attestation Agent、CDH和可信网络网关 (Trusted Network Gate)，用于安全的 AI 工作流。
- 开发了Tee-RefValue工作流，该流程可在 AMD SEV-SNP 平台上自动为用户镜像、固件和内核生成参考度量值，并与机密容器兼容。
- 研究了 AMD SEV-SNP 平台上的证明过程，包括参考度量值的生成。
- 评估了各大云服务提供商（如谷歌云和 Azure）的证明工具。

发表论文

1. **Jingyao Zhang**, Jaewoo Park, Jongeun Lee and Elaheh Sadredini. “SAIL: SRAM-Accelerated LLM Inference System with Lookup-Table-based GEMV.” *In Submission.*
2. **Jingyao Zhang**, and Elaheh Sadredini. “A Near-Cache Architectural Framework for Cryptographic Computing.” *In Submission.*
3. **Jingyao Zhang**, and Elaheh Sadredini. “Unlocking Energy-Efficient and High-Throughput Secure Data Communication in IoT with Memory-Centric Computing.” *Poster in Proc. of the 39th International Parallel and Distributed Processing Symposium (IPDPS). June 2025.*
4. Sahar Ghoflsaz, **Jingyao Zhang**, and Elaheh Sadredini. “Enabling Low-Cost Secure Computing on Untrusted In-Memory Architectures.” *In Proc. of the 34th USENIX Security Symposium. August 2025.* (***CCF-A* 类**)

5. **Jingyao Zhang**, Mohsen Imani, and Elaheh Sadredini. "BP-NTT: Fast and Compact in-SRAM Number Theoretic Transform with Bit-Parallel Modular Multiplication." *In Proc. of the 60th Design Automation Conference (DAC). July 2023.* (***CCF-A* 类**)
6. **Jingyao Zhang**, and Elaheh Sadredini. "Inhale: Enabling High-Performance and Energy-Efficient In-SRAM Cryptographic Hash for IoT." *In Proc. of the 41th International Conference on Computer-Aided Design (ICCAD). November 2022.* (***CCF-B* 类**)
7. **Jingyao Zhang**, Hoda Naghibijouybari, and Elaheh Sadredini. "Sealer: In-SRAM AES for High-Performance and Low-Overhead Memory Encryption." *In Proc. of the 22th International Symposium on Low Power Electronics and Design (ISLPED). August 2022.* (***CCF-C* 类**)
8. **Jingyao Zhang**, Huaxi Gu, Li Zhang, Bing Li, and Ulf Schlichtmann. "Hardware-Software Codesign of Weight Reshaping and Systolic Array Multiplexing for Efficient CNNs." *In Proc. of the 24th Design, Automation and Test in Europe (DATE). February 2021.* (***CCF-B* 类**)

研究经历

AREA Lab, 加州大学河滨分校

导师: *Elaheh Sadredini*

研究生助理研究员 2021 年 9 月 - 至今

- 正在设计一个专为存内计算 (PIM) 架构定制的可信执行环境 (TEE) 框架，其设计受 sNPU 等安全 NPU 启发；集成了轻量级的基于 tile 的权限检查、无需虚拟内存的 SRAM 级 bank 隔离以及感知 NoC 的安全数据路由，以最小的 TCB 实现安全执行并支持缓存行感知的内存映射。
- 使用基于 MLIR 的 Polyhedral 模型和基于 E-Graph 的搜索，为领域特定加速器开发了一个通用编译器，特别针对涉及向量和标量内核以及混合精度的工作负载。
- 为加速量化语言模型设计了一种片上解决方案，包括动态数据精度自适应和高效的运行时去/量化。
- 开发了一个框架，将存内 SRAM 计算无缝集成到现有计算机系统中，以实现高效、安全的片上（前）后量子密码学处理。
- 为在 SRAM 内高效、安全地计算 NTT，开发了一种采用隐式移位技术的位并行模乘算法，在低开销 SRAM 阵列上优化了性能。
- 利用行/通道级数据对齐，为 AES/SHA-3 算法的片上加速设计了一个安全的存内 SRAM 架构，实现了高能效、高吞吐量和高面积效率。

先进网络技术实验室, 西安电子科技大学

导师: 顾华玺

硕士研究生 2018 年 9 月 - 2021 年 6 月

- 开发了一种面向 CNN 的软硬协同设计框架，利用权重重塑和脉动阵列复用技术，并结合遗传算法优化硬件性能。
- 构建了一个基于 FPGA 的分布式推理系统，采用 HLS 描述硬件，利用脉动阵列加速 CNN，并通过 Aurora/Ethernet 协议进行板间通信。
- 设计了一种灵活紧凑的 N×N 等离子体开关拓扑，并提出专用配置算法确保其可重构无阻塞，适用于管理数据中心的混合流量。
- 为光片上网络中的 Mesh 网络设计了一种低损耗、紧凑的等离子体路由器，相比其他结构具有更低的插入损耗和更小的尺寸。

教学经历

CS 010C 数据结构与算法导论

Instructor: Patrick Miller

助教 2024 年春季

- 每周主持三次实验课，以辅助讲座和作业。
- 每周设立办公时间以解答学生问题。

CS 161 计算机系统设计与体系结构

Instructor: Elaheh Sadredini

助教 2023 年秋季, 2024 年冬季

- 每周主持三次讨论课，以辅助讲座、作业和实验。
- 每周设立办公时间以解答学生问题。

CS 213 多核处理器架构和编程 *Instructor: Elaheh Sadredini*

助教 2022 年秋季

- 主持了两场学生报告的讨论会。
- 每周设立办公时间以解答学生问题。
- 批改了作业和编程项目。

其他经历

gem5 训练营 Davis, USA

营员 2022 年 7 月

- 模拟并分析计算机架构性能，研究了不同工作负载和基准测试套件在各种计算机架构上的行为。
- 评估了不同的设计选择（如变化的缓存大小或使用不同的互连拓扑）对系统性能的影响，并探索了不同微架构特性的效果。

Xilinx 夏令营 线上

营员 & 队长 2020 年 7 月 - 2020 年 8 月

- 开发了一种基于 FPGA 的分布式加速平台，母板通过以太网将文件发送到子板的监控文件夹，以触发程序执行。[代码]

微软创新中心 西安, 中国

实习生 2017 年 7 月 - 2017 年 8 月

- 探索了从早期的模拟系统到 5G 技术的蜂窝网络演进过程中的进步和挑战。

学术报告

1. **Jingyao Zhang**. "In-SRAM Computing for Cryptography." 在计算所夏培肃论坛上报告. [slides] *Dec 2023*
2. **Jingyao Zhang**. "Combined Attestation." *In Confidential Containers Community Meeting.* [slides] [video] [demo] *Nov 2023*
3. **Jingyao Zhang**. "Safeguard Your Cloud Workloads: An In-depth Look at Confidential Computing." *In Proc. of IEEE Inland Empire Data Science Workshop (IEDSW).* [link] *Nov 2023*
4. **Jingyao Zhang**. "Safeguard Your Cloud Workloads and Then Accelerate: An In-depth Look at CPU and GPU Confidential Computing." 受邀于字节跳动进行线上报告. ***Live Attendance: ~300***. [slides] [video] *Oct 2023*
5. **Jingyao Zhang**. "BP-NTT: Fast and Compact in-SRAM Number Theoretic Transform with Bit-Parallel Modular Multiplication." *In Proc. of the 60th Design Automation Conference (DAC). San Francisco, CA.* [slides] [video] *Jul 2023*
6. **Jingyao Zhang**. "Inhale: Enabling High-Performance and Energy-Efficient In-SRAM Cryptographic Hash for IoT." *In Proc. of the 41th International Conference on Computer-Aided Design (ICCAD). San Diego, CA.* [slides] [video] *Nov 2022*
7. **Jingyao Zhang**. "Sealer: In-SRAM AES for High-Performance and Low-Overhead Memory Encryption." *In Proc. of the 22th International Symposium on Low Power Electronics and Design (ISLPED). Online.* [slides] [video] *Aug 2022*
8. **Jingyao Zhang**. "Hardware-Software Codesign of Weight Reshaping and Systolic Array Multiplexing for Efficient CNNs." *In Proc. of the 24th Design, Automation and Test in Europe (DATE). Online.* [slides] [video] *Feb 2021*

获奖情况

DAC 青年学者奖, 设计自动化会议 2023 年
院长杰出奖学金, 加州大学河滨分校 2021 年
优秀毕业论文奖, 西安电子科技大学 2021 年
一等奖学金, 西安电子科技大学 (前 14% / 560+) 2018 年, 2019 年
优秀学生奖, 西安电子科技大学 2018 年, 2019 年

所获经费

Conference Travel Grant, University of California, Riverside 2023
Student Travel Grant, gem5 Boot Camp 2022

学术服务

审稿论文: 16 篇

- **期刊论文审稿:**

  Elsevier Integration (1) *2025*
  IEEE Computer Architecture Letters (CAL) (13) *2023, 2024, 2025*
  IEEE Transactions on Dependable and Secure Computing (TDSC) (1) *2024*

- **会议论文审稿:**

  IEEE International Conference on Communication (ICC) (1) *2024*

成果评估: 17 项

- **成果评估委员会 (期刊):**

  Journal of Systems Research (JSys) (2) *2023*

- **成果评估委员会 (会议):**

  IEEE International Symposium on High-Performance Computer Architecture (HPCA) (1) *2024*
  USENIX Security Symposium (2) *2024*
  ACM European Conference on Computer Systems (EuroSys) (3) *2024*
  Annual Network and Distributed System Security Symposium (NDSS) (2) *2024*
  ACM Symposium on Operating Systems Principles (SOSP) (2) *2023*
  ACM Conference on Computer and Communications Security (CCS) (1) *2023*
  ACM International Conference On Mobile Computing And Networking (MobiCom) (1) *2023*
  USENIX Annual Technical Conference (ATC) (1) *2023*
  USENIX Symposium on Operating Systems Design and Implementation (OSDI) (2) *2023, 2024*

技能

- **编程语言:** C, C++, Python, Verilog, Rust, MLIR
- **技术工具:** Gem5, Sniper, HSpice, PyTorch, LLVM
- **语言:** 中文 (母语), 英文 (专业)