Kailash Gogineni

Curriculum Vitae

5390 Science & Engineering Hall
GWU, Washington, DC 20052
(+1) 8049226099
kailashg26@gwu.edu
kailashgogineni.com
Kailash Gogineni

Research Interest

My current research focuses on systems and computer architecture, with technical expertise in workload characterization, performance modeling, analysis, and performance optimizations. Within this domain, I particularly specialize in high-performance single-agent/multi-agent reinforcement learning (MARL). Additionally, my interests extend to software security, where I have prior experience in malware detection and customizing communication protocols for continuous authentication.

Education

2019 - Present **Doctor of Philosophy in Computer Engineering**, *The George Washington University, DC, USA.*, Advisor: Dr. Guru Venkataramani,
CGPA: 3.85/4.0
Expected graduation: late 2024/early 2025

2015 - 2019 **Bachelors in Computer Science & Engineering**, *Vellore Institute of Technology*, Vellore, India,
CGPA: 9.2/10

Work Experience

May 2024 - Present **Research Intern Co-op**, *Advanced Micro Devices (AMD)*, Austin, Texas.
CPU Architecture Research Group - **Mentor: Onur Kayiran**

Jan 2022 - May 2024 **Graduate Research Assistant**, *The George Washington University, Washington DC, USA*,
Advisor: Dr. Guru Venkataramani, Lab for Intelligent Networking and Computing (LINC)
Project: High-Performance Single/Multi-Agent Reinforcement Learning
Data Caching/Reusing and Workload Characterization:

- We implemented RACER, a novel software caching approach proposed to address performance bottlenecks in accessing transition data from large experience replay buffers during Reinforcement Learning (RL) training. We developed a caching layer tailored to improve memory access times for critical transition batches within the experience replay memory. Leveraging multiple measures such as temporal difference error and advantage weighting, RACER dynamically inserts and evicts the transition data from the cache during training. Performance evaluation across three state-of-the-art offline RL algorithms, under various task environments, and on two different systems, demonstrates that RACER achieves significant training time improvements in the RL transition data sampling phase (a speedup of $6\times$) and end-to-end training time (up to $2\times$), with comparable or even higher rewards. (Submitted to MICRO 2024)
- Analyzed computational and scalability issues and identified performance bottlenecks in multi-agent systems.
  - The research presented in MLArchSys 2023, and FastPath 2023 introduced a neighbor sampling optimization to enhance cache locality, resulting in a significant reduction of up to 27.39% in sampling phase training time. The proposed optimizations target irregular memory access patterns and cache misses, utilizing techniques such as cache-aware sampling and importance sampling. We tailored the sampling phase to optimize address fetch patterns and guide the hardware prefetcher for enhanced efficiency while ensuring an algorithmic performance guarantee. An extended version of the work is currently under review.
- Developed AccMER to exploit high-performing trajectories to the fullest across multiple time steps and further improve cache locality, resulting in a training time reduction of 25.4% over the state-of-the-art baselines. (ASAP 2023)
- Key areas: Systems for ML, CPU-GPU, Run-time Performance Optimizations, Workload Characterization, Hardware/Software Prefetching, Multi-threading, Perf Analysis.

Feb 2023 - May 2024 **Project: High-Performance Single/Multi-Agent Reinforcement Learning Real Processing-In-Memory Systems (UPMEM PIM):**

- Developed SwiftRL, a framework leveraging Processing-In-Memory (PIM) architectures to enhance Reinforcement Learning (RL) training. It addresses RL challenges, such as memory-bound training and data transfer bottlenecks, by performing computations within memory devices. The study adapts Tabular Q-learning (also Multi-Agent version of Tabular Q-learning) and SARSA algorithms on UPMEM PIM systems, optimizing performance and demonstrating near-linear scaling of 15× with a 16× increase in PIM cores. Comparative evaluations against Intel CPU and NVIDIA GPU highlight the superior performance of PIMs in various implementations. (ISPASS 2024 - in collaboration with SAFARI Group ETH Zürich)
- Key areas: Systems for ML, CPU-GPU, Run-time Performance Optimizations, Workload Characterization, Quantization, Processing-In-Memory, Hardware/Software Prefetching, Multi-threading, Perf Analysis, Roofline Analysis, Memory Bottleneck.

Aug 2021 - July 2023 **Cyber Deception: Hardware-assisted Self Tuning Cyber Deception**

- Developed Foreseer for forecasting malware presence, reducing detection time by 40%. (SEED 2022)
- Explored hardware support for transparently countering malware with near-native execution.
- Introduced MAYALOK for inserting or skipping malware instructions during runtime. (ASAP 2023)
- Key areas: Hardware/Software Security, Proactive Defense/Deception.

Aug 2019 - Sep 2022 **Cyber Security: Communication Protocol Customization via Cross-grafting, and Trimming - [More details - DIALECT]**

- Created Verify-Pro for using protocol dialects as fingerprints for continuous authentication. (MILCOM 2022)
- Developed a reliable application-layer moving target defense model (MPD) via customized communication protocols. (SecureComm 2021)
- Key areas: Cyber Security, Machine Learning for Security.

June 2018 - Aug 2018 **Software Developer Intern**, *gnani.ai, Bangalore, India*, Mentor: Anant Nagaraj

- Developed a robust architecture for a speech decoding application, catering to a user base exceeding a million users daily.
- Led data cleaning efforts to ensure high-quality data inputs for the application.
- Spearheaded web scraping activities utilizing the Scrapy web crawler tool, gathering data from diverse websites to enrich the application's content and functionality.

Research Publications

IEEE ISPASS'24 **Kailash Gogineni, Sai Santhosh, Juan Gómez-Luna, Karthik Gogineni, Peng Wei, Tian Lan, Mohammad Sadrosadati, Onur Mutlu, and Guru Venkataramani** , *SwiftRL: Towards Efficient Reinforcement Learning on Real Processing-In-Memory Systems* , IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2024, Indianapolis, Indiana.

IEEE ASAP'23 **Kailash Gogineni, Yongsheng Mei, Peng Wei, Tian Lan, and Guru Venkataramani** , *AccMER: Accelerating Multi-Agent Experience Replay with Cache Locality-aware Prioritization* , IEEE International Conference on Application-specific Systems, Architectures, and Processors (ASAP), 2023, Porto, Portugal. [PDF]

IEEE ASAP'23 **Preet Derasari, Kailash Gogineni, Guru Venkataramani** , *MAYALOK: A Cyber-Deception Hardware Using Runtime Instruction Infusion* , IEEE International Conference on Application-specific Systems, Architectures, and Processors (ASAP), 2023, Porto, Portugal , Also presented at HoTSoS'24.

MLArchSys'23 (ISCA'23) **Kailash Gogineni, Peng Wei, Tian Lan, and Guru Venkataramani** , *Towards Efficient Multi-Agent Learning Systems* , ML for Computer Architecture and Systems (MLArchSys, ISCA), 2023, Orlando, Florida. [PDF]

ACM GLSVLSI'23 **Preet Derasari, Kailash Gogineni, Guru Venkataramani** , *MAYAVI: A Cyber-Deception Hardware for Memory Load-Stores* , Proceedings of the Great Lakes Symposium on VLSI (GLSVLSI), June 2023. [PDF]

FastPath'23 (ISPASS'23) **Kailash Gogineni, Peng Wei, Tian Lan, Guru Venkataramani** , *Scalability Bottlenecks in Multi-Agent Learning Systems* , 10^{th} International Workshop on Performance Analysis of Machine Learning Systems (In conjunction with ISPASS 2023). [PDF]

IEEE SEED'22 **Kailash Gogineni, Preet Derasari, Guru Venkataramani** , *Foreseer: Efficiently Forecasting Malware Event Series with Long Short-Term Memory* , Proceedings of IEEE International Symposium on Secure and Private Execution Environment Design (SEED), September 2022. [PDF]

IEEE MILCOM'22 **Kailash Gogineni, Yongsheng Mei, Guru Venkataramani, Tian Lan** , *Verify-Pro: A Framework for Server Authentication using Communication Protocol Dialects* , Proceedings of IEEE Military Communications Conference (MILCOM), December 2022. [PDF]

Springer SecureComm'21 **Yongsheng Mei, Kailash Gogineni, Tian Lan, Guru Venkataramani** , *MPD: Moving Target Defense through Communication Protocol Dialects* , Proceedings of the International Conference on Security and Privacy in Communication Networks (SecureComm), September 2021. [PDF]

IEEE IWSC'20 **Hongfa Xue, Yongsheng Mei, Kailash Gogineni, Guru Venkataramani, Tian Lan** , *Twin-Finder: Integrated Reasoning Engine for Pointer-Related Code Clone Detection* , Proceedings of IEEE International Workshop on Software Clones (IWSC), February 2020. [PDF]

Awards

- Received a Student Travel Grant for ISPASS 2024.
- Received a Student Travel Grant for ISCA 2023.
- Won $5,000 Finalist Prizes in the TCO 2022 Innovation Competition for "A Hardware-based Cyber Deception Framework Against Malware," with co-authors Guru Venkataramani, Preet Derasari, and Kailash Gogineni.
- Recipient of University Fellowships in June 2020, 2021, and 2022.
- Achieved the Audience Choice Award in the Research Blitz, Fall 2022, at GWU.

Technical Skills and Toolsets

Languages	Python, R, Bash scripting, C/C++
Tools	Weights & Biases, Tableau, RAPL, Perf, Anaconda, Jupyter Notebook, Google Colab, Numba, Microsoft Excel, Google Analytics, Vim, Git, GDB, PyCharm, LaTex, Intel VTune, NVIDIA Nsight, OpenMP, Intel Advisor
Databases	MySQL, Microsoft SQL
Cloud tech.	Google Cloud Platform, Amazon Web Services
Frameworks	Scikit-learn, TensorFlow/Keras, PyTorch, NLTK, Pandas, NumPy, Matplotlib
Microarchitecture	UPMEM Processing-In-Memory (PIM) Simulator (Real Hardware)
Statistics/ML	Linear & Logistic regression, Random Forest, SVM, Neural Networks, CNN, LSTMs, XGBoost, Naive Bayes classifier, Reinforcement Learning, Multi-Agent systems
Platforms	Windows, Linux, Pop-OS

Relevant Coursework

- Machine Learning
- Microcomputer Systems Architecture
- Stochastic Processes in Engineering
- Reinforcement Learning
- Linear Systems Theory
- Advanced Microarchitecture
- Introduction to Computer Networks
- Secure Computer Architecture
- Wireless Networks

Research Talks

- **Presenter:** "SwiftRL: Towards Efficient Reinforcement Learning on Real Processing-In-Memory Systems" at IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS'24)
- **Presenter:** "Scalability and Computational Bottlenecks in Multi-Agent RL Systems" at FastPath'23. [Talk Video]
- **Presenter:** "Towards Efficient Multi-Agent Learning Systems" at MLArchSys 2023 during ISCA-50, FCRC.
- **Poster Presenter:** Showcased "Verify-Pro" at the SEAS Student Research & Development Showcase 2022.
- **Poster Presenter:** Displayed "Verify-Pro" as a poster at the cybersecurity & privacy@gw day event. [More details]
- **Young Scholar Workshop Presenter:** Presented "Verify-Pro" at the Young Scholar Workshop during MILCOM 2022. [More details]
- **Presenter:** Presented "Foreseer" at the International Symposium on Secure and Private Execution Environment Design (SEED 2022).
- **Primary Presenter:** Delivered a presentation on "Communication Protocol Customization and Fuzzing" at the Software Security Summer School (SSSS'20, Hosted by: Purdue University and Sponsored by: ONR). [Talk Video]
- **Co-Presenter:** Presented "Twin-Finder" at the International Workshop on Software Clones (SANER 2020).

Professional Services

Student Volunteer	NSF CISE CAREER Workshop 2024
Student Volunteer	IEEE International Conference on Computer Design (ICCD 2023)
Sub. Chair	IEEE International Symposium on Workload Characterization (IIWSC), 2023, Ghent, Belgium
Reviewer	IEEE Micro (MICRO)
Reviewer	Journal of Hardware and Systems Security

References

- **Dr. Guru Venkatarmani**
  Professor, Department of Electrical and Computer Engineering
  George Washington University
  Contact: guruv@gwu.edu (Available upon request)
- **Dr. Tian Lan**
  Professor, Department of Electrical and Computer Engineering
  George Washington University
  Contact: tlan@gwu.edu (Available upon request)
- **Dr. Peng Wei**
  Associate Professor, Department of Mechanical and Aerospace Engineering
  George Washington University
  Contact: pwei@gwu.edu (Available upon request)