Pragya Sharma

https://www.linkedin.com/in/pragyasharma29 | https://sharma-pragya.github.io/ | pragyasharma@ucla.edu

EDUCATION

University of California Los Angeles, Los Angeles, CA **Sept 2021 - Present**
Doctor of Philosophy in Electrical and Computer Engineering

Carnegie Mellon University, Mountain View, CA **Dec 2018**
Master of Science in Electrical and Computer Engineering

VIT University, Vellore, India **May 2017**
Bachelor of Technology in Electronics and Communication Engineering

WORK EXPERIENCE

University of California Los Angeles, Los Angeles, CA

GRADUATE RESEARCH ASSISTANT **Sept 2021 - Present**

PROJECT: CADET (Cooperative Autonomy through Distributed Experimentation Toolkit)

- Architected CADET, an open-source distributed autonomy platform integrating CARLA and Mininet across heterogeneous platforms (Jetson Orin, A6000), enabling reproducible multi-vehicle simulation with <10ms synchronization overhead. (ICRA, under review)
- Enabled V2V and V2I evaluation with unified telemetry framework capturing metrics across model, system, and task layers, enabling rigorous analysis of perception-to-action pipelines with synchronized time stamping.
- Demonstrated through CADET that V2V intent packets eliminated collisions under 100 ms tail delays, while RSU-assisted perception preserved safety margins until compute saturation under concurrent load

PROJECT: FMaaS & FMTK (Foundation Models-as-a-Service & Time-Series Platform)

- Pioneered FMaaS, the first unified framework for deploying LLMs, VLMs, and TSFMs across GPU clusters and edge devices, achieving throughput improvement via task and architecture-level sharing strategies. (MLSys)
- Developed FMTK, a lightweight toolkit which reduced TSFM deployment from 500 lines to 10 lines of configuration code while maintaining full pipeline flexibility across encoding, adaptation, and decoding stages, a 95% code reduction. (NeurIPS, accepted)

PROJECT: Distributed Perception & Control in Autonomous Vehicles

- Characterized device–edge–cloud placement of perception (YOLOv8/YOLO11 variants) and control (PID vs. MPC) models in AV sense-to-act workloads. (Published at ICCCN, HotMobile, MILCOM)
- Analyzed latency–accuracy–energy trade-offs in closed-loop vs. open-loop pipelines, demonstrating that on-device control minimized error/jerk, while cloud perception extended detection range by >2.5×. Findings challenged the assumption that cloud deployments are unsuitable for real-time inference tasks.
- Proposed Adaptive Control Placement (ACP) strategy achieving 30% TTC improvement by dynamically switching between local PID and cloud-based MPC based on network conditions.

Amazon.com (Amazon Lab126), Sunnyvale, CA

EMBEDDED SOFTWARE ENGINEER **Feb 2019 - Aug 2021**

- Designed and developed Amazon Sidewalk IoT network for Amazon, Ring and 3P devices. Built 900 MHz based communication protocol and embedded systems firmware optimized for low-power devices.
- Increased flash memory efficiency by designing an end-to-end protocol configuration management system. Created an abstraction layer and associated APIs to moderate flash access by application layer.
- Led low-latency firmware development for hardware-software integration, independently debugging critical issues and coordinating cross-functional teams to meet aggressive product timelines.

Bosch LLC, R&D Department, Sunnyvale, CA

WIRELESS SYSTEMS INTERN **May 2018 - Aug 2018**

- Collaborated with the Wireless Connectivity Group on Perfectly Keyless passive entry system, engineering seamless car locking/unlocking experience.
- Engineered the hardware for the localization of the user using Ultra-Wide Band nodes on the car and modeled new localization algorithms to further the accuracy (~30cm) and security of the system.

Carnegie Mellon University, Mountain View, CA

GRADUATE RESEARCH ASSISTANT **Aug 2018 - Dec 2018**

- Co-designed a novel admission control algorithm to control congestion and allocate sufficient resources for interactive video sessions while accommodating background traffic.
- Acting as wireless lead, designed the testbed for experiments over 802.11g (WiFi) to implement this algorithm. The system showed 47% improvement with respect to network throughput over traditional routers. Findings have been published in IEEE JSAC 2019 - Special Issue on Multimedia Economics.

GRADUATE RESEARCH ASSISTANT **Jan 2018 - May 2018**

- Research lead for Passive Entry - Passive Start (PEPS) Car Security project sponsored by Ford R&D.
- Designed and implemented the infrastructural firmware and hardware using Software-Defined Radios to demonstrate a man-in-the-middle attack on passive keyless entry systems over a line-of-sight range of 300ft.

GRADUATE RESEARCH ASSISTANT **Jan 2018 - May 2018**

- Sponsored by Argonne National Labs, Chicago with the aim to develop an open source implementation of vehicle-to-vehicle communications along with analyzing its security vulnerabilities.
- Engineered the DSRC/WAVE stack on BladeRF (SDR) to emulate vehicles exchanging information.

RELEVANT PUBLICATIONS

- *FMaaS: Foundation Models as a Service*
  P. Sharma*, H. H. Shastri*, P. Shenoy, M. B. Srivastava [MLSys '26, under review]
- *CADET: Enabling Cooperative Autonomy via Distributed Experimentation Toolkit*
  P. Sharma, B. Wang, M. B. Srivastava [ICRA '26, under review]
- *FMTK: A Modular Toolkit for Composable Time Series Foundation Model Pipelines*
  H. H. Shastri, **P. Sharma**, P. Shenoy, M. B. Srivastava [NeurIPS '25, accepted]
- *Cloud Is Closer Than It Appears: Revisiting the Tradeoffs of Distributed Real-Time Inference*
  P. Sharma, H. Qui, M. B. Srivastava [ICCCN '25]
- *Towards a Performance-Driven Device-Edge-Cloud Relationship*
  P. Sharma, B. Wang, X. Ouyang, R. Nanayakkara, B. Balaji, P. Tabuada, M.B. Srivastava [HotMobile 2025]
- *Impact of Delays and Computation Placement on Sense-Act Application Performance in IoT*
  P. Sharma, M. B. Srivastava [MILCOM '23]
- *Procuring Spontaneous Session-Level Resource Guarantees for Real-Time Applications: An Auction Approach*
  M. Harishankar, S. Pilaka*, **P. Sharma***, N. Srinivasan, C. Joe-Wong, P. Tague [IEEE-JSAC, 2019]

LEADERSHIP EXPERIENCE

- Mentored over 50 undergraduate and graduate students for research projects
- Mentor, The Women of Color Project
- Graduate Instructor, Los Angeles Computing Circle (CS Bootcamp 2022, 2023, 2024)
- Software Engineering Mentor, Amazon Mentorship Program
- Head, Women in ECE (WinECE), Carnegie Mellon University, Silicon Valley
- Campus Ambassador, Carnegie Mellon University, Silicon Valley
- Program Representative, Student Council, ECE batch of 2017, VIT University

TECHNICAL SKILLS

- **Programming Languages**: C, Python, C++, HTML/CSS/JavaScript, Assembly, Verilog/VHDL
- **Technologies**: LoRa, FSK, WiFi, Bluetooth, BLE, DSRC, Matter, Free RTOS, Matter, ROS, UWB
- **Hardware Platforms**: Arduino, AWS DeepRacer, Jetson, Nordic, NVIDIA GPUs, Raspberry Pi, Semtech, TI
- **Software Platforms**: AWS Suite, Docker, GNURadio, Iperf, MATLAB, Mininet, Wireshark
- **AI/ML**: CUDA, PyTorch, TensorFlow, TensorRT, OpenCV, NCNN, LLMs, VLMs, Quantization