

Sri Pramodh Rachuri

———— Educational Qualifications

PhD Candidate in Computer Science, Stony Brook University, New York, USA

Advisor – *Prof. Anshul Gandhi*

Courses – Operating Systems, Network Security, Fundamentals of Computer Networks

2020 – Present

GPA – **4.0/4.0**

B.Tech (Hons) in Electrical Engineering, IIT Bhilai, Chhattisgarh, India

Advisor – *Prof. Arzad Alam Kherani*

Courses – Operating Systems, Distributed Systems, Mobile Communication Systems, Wireless Communication, Wireless MAC Modeling, Computer Networks, Information Theory, Embedded Programming, Wireless Security

Director's Gold Medal recipient for outstanding performance in academic and extracurricular activities.

R P Singh Memorial Award recipient for leadership and social welfare initiatives.

2016 – 2020

GPA – **9.03/10.00**

———— Work Experience

Research Intern, AT&T Labs (continuing collaboration)

June'22-Aug'22

- Evaluated the requirements for Virtualized RAN (vRAN) pooling and discerned differences in radio resource consumption patterns between LTE macro and microcells, as well as NR (5G) macro and mmWave cells.
- Currently involved in estimating the benefits of resource pooling and employing statistical methods for placement decisions by using anonymized production data traces from real-world radio equipment deployments.

Research Intern, IIT Kharagpur

May'19-July'19

- Used the non-stationary Multi-armed Bandit reinforcement learning method to design a solution for task distribution challenges in Fog computing.
- Demonstrated, through simulation, that the design scheme outperforms existing cloud and fog solutions, reducing overall latency by 59%. Further details are available in the publication [5].

———— Research Projects

Sustainable Deep Learning at the Edge

Aug'22-Present

- Conducted power consumption benchmarking on edge computing devices such as Nvidia's Jetson Nano and Xavier NX. Explored variations in hardware and workload parameters, including CPU and GPU frequency, core count, batch size, and neural network model complexity.
- Compared to DVFS (Dynamic Voltage and Frequency Scaling), optimizing CPU and GPU frequencies resulted in a 19% reduction in energy consumption without compromising accuracy. Simultaneous tuning of workload and hardware parameters yielded a 42% reduction in energy consumption while slightly enhancing accuracy. A portion of this research is detailed in the publication [1].
- These findings have been instrumental in assessing trade-offs between energy consumption and latency, and are currently being utilized to develop energy-efficient system policies.

Orchestrating Live Video Analytics on Disaggregated Systems at the Edge

Feb'21-Present

- Engineered an innovative microservice architecture for deploying independent components of a video analytics pipeline at the edge. Used Python-Flask, Gunicorn, Docker containers, and TensorRT on Nvidia Jetson Nano for the implementation. Experimentation showed that our deployment strategy incurred 60% less overhead compared to a naive policy. Detailed insights available in the publication [3].
- Modularized the Video Analytics functions for scalability and orchestration using Kubernetes. Designed novel algorithms for scheduling and load-balancing, resulting in a 1.93x increase in throughput and a 90% reduction in frame drop rates in comparison to a naive policy. A publication on this part of the project is currently under peer review.

Detecting Network Utilization Patterns in MapReduce jobs (in collaboration with TU Berlin)

Aug'22-Present

- Investigating network transfer patterns to identify connections that can be optimized to enhance the overall performance and reduce completion times of MapReduce jobs in Apache Spark. This research involves simulating various network topologies that can be deployed to mitigate network bottlenecks.

Near Data Processing (NDP) Optimization for Apache Spark Jobs

May'21-Jan'22

- Collaborated on Open Infras Lab's customized version of Spark and HDFS to implement intelligent push-down operations from Spark to HDFS, optimizing job execution time, data transfer latency, and bandwidth utilization

- o Developed a Discrete Event Simulator using the Simpy library in Python to replicate this setup. The simulator emulated clusters of various sizes, configurations and specifications, demonstrating that our policy could theoretically reduce query execution time by 71% compared to standard Apache Spark and Hadoop systems.
- o In empirical experiments conducted on real hardware, our policy achieved an average query execution time reduction of 42% compared to the default strategy of not implementing NDP. Additional details can be found in the publication [2].

Accurate Modeling of Request Life Cycle in Two-Tier Storage Systems

May'22-Present

- o Analyzing request arrival patterns in a two-tier storage system consisting of HDDs with SSD caching and studying the associated service, response, and wait times.
- o Currently focused on modeling various distributions that closely mimic request behavior, aiming to enhance simulations and potentially inform future scheduling policies.

Multi-Access and Network Coding for Video Transmission (B.Tech Hons Thesis)

Jan'19-May'20

- o Built an SD-WAN that can be used to serve any UDP stream with a high data rate and less delay jitter using multiple LTE dongles. In all experiments conducted, packet losses remained below 1% while live streaming a 4K video content.
- o Used Open Air Interface to build an LTE testbed using Software Defined Radios (SDRs) and used its simulator mode to make a scalable testbench for this project. More details in publications [4,6,7,8].

Publications

1. *Evaluating the energy impact of device parameters for DNN inference on edge* Preprint
Anurag Dutt*, Sri Pramodh Rachuri*, Ashley Lobo, Nazeer Shaik, Anshul Gandhi, Z. Liu
Oct 2023, Work-in-progress track, 14th International Green and Sustainable Computing Conference (IGSC 2023)
2. *Optimizing Near-Data Processing for Spark* IEEEExplore
Sri Pramodh Rachuri, Arun Gantasala, Prajeeth Emanuel, Anshul Gandhi, Robert Foley, Peter Puhov, Theodoros Gkountouvas, Hui Lei
July 2022, 42nd International Conference on Distributed Computing Systems (ICDCS 2022)
3. *Decentralized Modular Architecture for Live Video Analytics at the Edge* ACM Digital Library
Sri Pramodh Rachuri, Francesco Bronzino, Shubham Jain
Oct 2021, 3rd ACM Workshop on Hot Topics in Video Analytics and Intelligent Edges (located in MobiCom'21)
4. *Traffic Splitting for Delay Jitter Control in Multi-access Systems* Springer Link
Megha Sahu, Sri Pramodh Rachuri, Ahtisham Ali Ansari, Arzad Alam Kherani
March 2022, Springer Telecommunication Systems - Volume: 80, Issue: 4
5. *Multi-Armed Bandit-based Decentralized Computation Offloading in Fog-Enabled IoT* IEEEExplore
Sudip Misra, Sri Pramodh Rachuri, Pallav Kumar Deb, Anandarup Mukherjee
June 2021, IEEE Internet of Things Journal - Volume: 8, Issue: 12
6. *On limiting Delay and Jitter characteristics at application-layer of Multi-connected Systems* IEEEExplore
Megha Sahu, Sri Pramodh Rachuri, Ahtisham Ali Ansari, Deepaknath Tandur, Arzad A. Kherani
Sept 2020, 3rd IEEE 5G World Forum (5GWF)
7. *An SD-WAN Controller for Delay Jitter Minimization in Coded Multi-access Systems* IEEEExplore
Ahtisham Ali Ansari, Sri Pramodh Rachuri, Arzad A. Kherani, Deepaknath Tandur
Dec 2019, 4th International Workshop on 5G and Future Wireless Technology (located in IEEE ANTS'19)
8. *Network-Coded SD-WAN in Multi-Access Systems for Delay Jitter Control* IEEEExplore
Sri Pramodh Rachuri, Ahtisham Ali Ansari, Deepaknath Tandur, Arzad A. Kherani, Sameer Chouksey
Dec 2019, 4th International Conference on Contemporary Computing and Informatics (iC3I 2019)

*Joint first-authorship

Technical Skills

- o Languages – Python, Java, Go Lang, C, C++, MATLAB, Verilog, Shell Scripting.
- o ML Frameworks – PyTorch, Keras, Scipy, SKLearn.
- o Miscellaneous – Docker, Kubernetes, Android Studio, Git, L^AT_EX.

Academic and Extra Curricular Activities

- o Reviewer for IEEE IoT Journal 2022 and subreviewer for ACM/IFIP Middleware 2023; ACM Sigmetrics 2023, 2024.
- o Invited as a speaker at CryptEng lab, TU Graz to present our publication on *Optimizing Near-Data Processing for Spark*.
- o Received travel grants from Sigmetrics 2023, ICDCS 2022 and MobiCom 2021.
- o Student Representative in Grad Committee of CS Dept at Stony Brook University since Oct 2021.
- o Serving as a Board Member of PanIIT USA since 2023.
- o Founding executive member and Vice President (UG) of Alumni Association of IIT Bhilai from 2020 to 2023.
- o Founding Secretary of Science and Technology Division in Council of Student Affairs, IIT Bhilai from 2017 to 2018.
- o Student Representative in Undergrad Committee of EECS Dept at IIT Bhilai in 2018.
- o Founded Electronics Club in 2016 at IIT Bhilai.
- o Coordinator for IIT Bhilai - Institute's Innovation Cell (IIC) in 2018.
- o Selection committee member for Young Researcher Awards, IIT Bhilai in 2022.