Parth Thakkar

+1-217-200-2828 | thakkar.parth.d@gmail.com | thakkarparth007.github.io

EDUCATION

University of Illinois, Urbana-Champaign **Aug 2021 - May 2023**
MS in Computer Science | CGPA: 4.00
Graduate Research Assistant | Prof. Tianyin Xu
Graduate Teaching Assistant | CS 411 Database Systems

National Institute of Technology, Trichy **Aug 2014 - May 2018**
B.Tech (Hons.) in Computer Science and Engineering

PUBLICATIONS

1. *Optimizing Network Provisioning through Cooperation*. **NSDI 2022**.
   H. Sharma*, **P. Thakkar***, S. Bharadwaj*, R. Bhagwan, V. Padmanabhan, et. al. (*Equal contributors)
2. *AutoSens: Inferring Latency Sensitivity of Users through Natural Experiments*. **ACM IMC 2021**.
   P. Thakkar, R. Saxena, V. Padmanabhan
3. *Scaling Hyperledger Fabric using Sparse Peers and Pipelined Execution*. **ACM SoCC 2021**.
   P. Thakkar & S. Nathan
4. *Performance Benchmarking and Optimizing Hyperledger Fabric Blockchain Platform*. IEEE MASCOTS 2018.
   P. Thakkar, S. Nathan, B. Vishwanathan (**Best Paper Award**)

EXPERIENCE

Meta *May 2022 - Aug 2022*
SWE Intern | AI Infra

- Worked on automatically analyzing IO behavior of AI pipelines to improve privacy and transparency.
- Developed a scalable static analysis pipeline for Python packages using Pyre and Buck.
- Trained an LLM-based classifier to identify if a function performs IO, obtained 80% accuracy.
- The project gained significant interest, and will be continued along my suggested extensions.

Microsoft Research *Jul 2019 - Jul 2021*
Research Fellow | Systems & Networking Group

- Worked on optimizing WAN bandwidth costs by leveraging first party setting.
- Proposed and implemented a mathematical framework for network provisioning and tested on Microsoft's production WAN.
- Estimated savings are in the order of tens of millions of dollars.
- Also worked on modelling the effect of latency on user engagement in online applications.

IBM Research *Jul 2018 - Jul 2019*
Research Engineer | Blockchain

- Improved performance, scalability & cost-efficiency of Hyperledger Fabric, IBM's Blockchain Platform.
- Developed a parallelization technique to improve CPU utilization by 2× and throughput by 40%.
- Developed a sharding technique to make the system horizontally scalable.
- Developed an auto-scaling mechanism to help reduce provisioning costs.
- Improved overall throughput by 3.7× and sped up scaling up time by 12-26×.

Research Intern | Blockchain *May 2017 - Jul 2018*

- Performed the first ever rigorous performance study of Hyperledger Fabric.
- Wrote a generic, highly configurable & reusable load generator, which was used for further studies.
- Made 3 key optimizations that improved performance 16× (from 140tps to 2250tps).

PROJECTS

Reverse Engineering Github Copilot *Nov 2022 - Dec 2022*

- Reverse engineered the obfuscated Copilot VSCode extension to extract key insights.
- Extracted modules and deobfuscated using hand-written AST transformations.
- Named and categorized **400 modules** automatically using OpenAI's Codex model via few-shot prompting.
- Dissected **prompt construction logic**, **model post-processing** and **telemetry collection** code among others.
- Distilled key findings into a blogpost coupled with a custom code-browser tool.

AI Assistant for Kubernetes Configurations *Sep 2022 - Ongoing*

Advisors: Prof. Tianyin Xu | UIUC, Dr. Mandana Vaziri | IBM Research

- Developing an "edit" model for Kubernetes Configurations.
- Working on synthesizing edit data from Helm charts templates.
- Developed a technique to **fuzz** Helm chart templates to produce edit diffs, and automatically label them.
- Built Kurator, a tool to collect human labelled data for evaluation.
- Working towards fine-tuning Salesforce's Codegen models.

Learning configuration validators for distributed systems *Sep 2021 - May 2022*

Advisors: Prof. Tianyin Xu, Prof. Madhusudan Parthasarathy | UIUC

- Given Configuration tests $T(Conf) \rightarrow Bool$ that check if a configuration is valid, we aim to generate validator formulas $V(Conf) \rightarrow Bool$ that mimic the test behaviour but are much cheaper to run.
- Working in a team to develop strategies to use dynamic analysis along with program synthesis techniques to learn the validators.

NL2CMD: Converting natural language instructions to Bash *Oct 2021 - Dec 2021*

Term Project | Advisor: Prof. Heng Ji | UIUC

- Given natural language instructions, we generate bash commands that satisfy the user's intent.
- Attempting to model command *execution* for both verification and training of models.
- Working on a model to incorporate command descriptions in synthesizing commands.

KEY SKILLS

Languages	C++, Java, Python, GoLang, Typescript, Javascript, C, Bash, SQL
Tools & Tech	Pytorch, Slurm, MySQL, Azure Cosmos, Pyre, Soot, Z3, NodeJS, Git

AWARDS AND ACHIEVEMENTS

- ACM SoCC Student Scholarship (2021)
- JN Tata Endowment Scholarship (2020)
- **Best Paper Award** for "Performance Benchmarking & Optimizing Hyperledger Fabric Blockchain Platform" at MASCOTS 2018
- **2nd** in In-Out Hackathon, one of India's largest student-run hackathons (2016)

Talks

- Invited industry talk on "Optimizing the performance of Hyperledger Fabric Platform" at ICDCN 2019
- Conducted workshop on Introduction to Blockchain Systems as a part of Vortex 2018, NIT-T CSE Symposium