






Shubham Tiwari

PhD student, University of Washington

 sbhtwr.github.io  tshubham@cs.washington.edu  github.com/sbhtwr  Google Scholar
 +1 (206) 853-0284

Research Interests

Broadly interested in the intersection of Systems and AI, currently working on improving the efficiency of serving models in Compound AI Systems. Previously, I have worked on a broad spectrum of problems, for example: memory management optimizations in hypervisors, network measurement, and congestion control (LEOScope, iBox).

Education

University of Washington, Seattle Ph.D. in Computer Science (<i>ongoing</i>) Advisors: Simon Peter, Ratul Mahajan	Sept 2023 - Present
Birla Institute of Technology and Science (BITS), Pilani B.E. Computer Science and M.Sc. Mathematics, CGPA: 8.52 Thesis: <i>Data-Driven Network Simulation with iBox</i>	Aug 2016 - July 2021

Experience

Microsoft Research, Redmond <i>Research Intern with Ishai Menache</i> <u>Project</u> : Improvements to Azure Compute's VM allocation service.	June 2024 - Sept 2024
Microsoft Research, Bangalore <i>Research Fellow with Debopam Bhattacharjee, Venkat Padmanabhan</i> <u>Projects</u> : LEO Satellite Networks (<u>LEOScope</u>), Data-Driven Network Simulation (<u>iBox</u>)	Aug 2021 - Aug 2023
Microsoft Research, Bangalore <i>Research Intern with Venkat Padmanabhan, Nagarajan Natarajan</i> <u>Project</u> : iBox	Jan 2021 - July 2021
VMware, Bangalore <i>Intern, xLabs with Jayneel Gandhi</i> <u>Project</u> : Page-table Replication (<u>Mitosis</u>)	Aug 2020 - Dec 2020
Samsung Research, Bangalore <i>Research Intern with Shital Bhatia</i> <u>Project</u> : Cellular Network Planning	May 2020 - July 2020
Software-Defined Networking Lab, BITS Pilani <i>Research Assistant with K. Hari Babu</i> <u>Project</u> : Passive Estimation of Link Latency (<u>qMon</u>)	Jan 2019 - Dec 2019

Publications

C=Conference, J=Journal, P=Preprint

- C.1 **Boosting Application Performance using Heterogeneous Virtual Channels: Challenges and Opportunities**
[link]
Talal Touseef, William Sentosa, Milind Kumar Vaddiraju, Debopam Bhattacharjee, Balakrishnan Chandrasekaran, Brighten Godfrey, Shubham Tiwari
22nd ACM Workshop on Hot Topics in Networks HotNets'23

- P.1 T3P: Demystifying Low-Earth Orbit Satellite Broadband** [link]
 Shubham Tiwari, Saksham Bhushan, Aryan Taneja, Mohamed Kassem, Cheng Luo, Cong Zhou, Zhiyuan He, Aravindh Raman, Nishanth Sastry, Lili Qiu, Debopam Bhattacharjee
Working paper arXiv'23
- C.2 Simulating Network Paths with Recurrent Buffering Units** [link]
 Divyam Anshuman, Sriram Balasubramanian, Shubham Tiwari, Nagarajan Natarajan, Sundararajan Sellamanickam, and Venkata N. Padmanabhan
 37th AAAI Conference on Artificial Intelligence AAAI'23
- C.3 Data-Driven Network Path Simulation with iBox** [link]
 Sachin Ashok, Shubham Tiwari, Nagarajan Natarajan, Venkata N. Padmanabhan, and Sundararajan Sellamanickam
 ACM SIGMETRICS / IFIP PERFORMANCE 2022 SIGMETRICS'22
- J.1 qMon: A method to monitor queueing delay in OpenFlow networks** [link]
 Sandhya Rathee, Shubham Tiwari, K Haribabu, and Ashutosh Bhatia
 Journal of Communications and Networks JCN'22

Projects

Retrify: Memory-Management Strategies for Efficient Serving of Models in Compound AI Systems April 2024 - Present

Advisors: Simon Peter, Ratul Mahajan

- Evaluated various memory-management techniques to improve the performance of long-context requests in Compound AI Systems.
- Developed a prototype of a serving-system interacting with an external environment to simulate performance of workloads under various traffic conditions (bursty, variable API call durations).

Optimizations to Azure Compute's VM Allocation Service June 2024 - Sept 2024

Advisors: Ishai Menache

- Evaluated the impact of enabling fine-grained VM placement strategies at scale.
- Proposed changes to resource allocation service with the potential of saving (> \$1M) in operational costs.

LEOScope: Enabling Experimentation Across Low-Earth Orbit (LEO) Satellite Networks July 2022 - Present

Advisors: Debopam Bhattacharjee, Venkat Padmanabhan [code]

- Lead an effort with **Azure Space**, **MSRA**, and academic collaborators to build a platform of a global scale for experimentation across Low-Earth Orbit Satellite networks.
- Drove the effort through engineering challenges such as platform architecture, implementation of experiment scheduler, executor, and the central orchestrator.
- Initiated large-scale measurements based on ping and iperf to characterize satellite network paths.

iBox: Internet in a Box Jan 2021 - June 2022

Advisors: Venkat Padmanabhan, Nagarajan Natarajan [website]

- Built a data-driven network simulator that uses data to recreate end-to-end behavior of a network path.
- Leveraged a combination of internet measurement data and ML models to capture the impact of complex network phenomena such as cross-traffic and packet reordering.
- Integrated iBox with ns-2, ns-3, netem and **Microsoft Teams's** in-house network simulator.
- Resulting papers published at **SIGMETRICS'22** and **AAAI'23**.

qMon: Passive Delay Monitoring in SDNs Jan 2019 - Dec 2019

Advisor: K. Hari Babu [paper] [code]

- Devised **qMon**, a scalable latency monitoring technique with zero data plane footprint.
- Developed an Open vSwitch based prototype to fetch queue length information using OpenFlow and passively estimate link latency.
- Evaluated qMon on a physical testbed under various traffic scenarios.
- Resulting paper published at **JCN 2022**.

Mitosis: Enabling Page-Table Replication in ESXi

Aug 2020 - Dec 2020

Advisor: Jayneel Gandhi

- › Implemented page-table replication (Mitosis) in VMware's core virtualization product – ESX.
- › Developed prototypes to disambiguate the design and code needed to support page table replication in the ESX kernel.
- › Conducted workload profiling to estimate the performance benefits of page-table replication; realized gains of upto **17%** in workload execution time.

Miscellaneous

- › Awarded a grant of 50,000 INR by AUGSD, BITS Pilani for developing a miniature autonomous driving vehicle.
- › Demonstrated iBox at TAB – MSR India's annual technical event.
- › Presented iBox at SIGMETRICS'22. [video]