Tuhinangshu Choudhury

PHD STUDENT, CARNEGIE MELLON UNIVERSITY

CIC, 4th floor, CMU
Pittsburgh, PA, USA
tuhinanc@andrew.cmu.edu | Website | Linkedin
Contact: (+1) 412 708 7547

INTERESTS

Performance Modeling, Scheduling, Machine Learning, Applied probability, Multi-armed Bandit

---

EDUCATION

Carnegie Mellon University, Pittsburgh, PA
PhD student, Electrical and Computer Engineering   *2019 - Present*
Advisor: Gauri Joshi and Weina Wang
Thesis: Optimizing the cost-aware performance of large-scale computing systems by accounting for and exploiting traffic patterns that mimic real-world demand. Developed *near-optimal* scheduling algorithms for *ML inference system* that operate within an accuracy-cost trade-off. Proposed novel resource allocation algorithms for cloud storage while accounting for time-varying and correlated demand for files, thereby achieving *order-wise* improvements in latency. Utilizing and advancing theoretical techniques in scheduling using erasure codes, optimization, and bandits to solve real-world resource allocation problems in large-scale computing systems.

Indian Institute of Technology Bombay, Mumbai, India
B.Tech. and M.Tech., Electrical Engineering   *2014 - 2019*
Advisor: Nikhil Karamchandani
Thesis: Mode Estimation and Clustering

PREPRINT

Optimizing Latency in Inference Systems with Accuracy Constraints [Paper]
***T.Choudhury**, G.Joshi, W.Wang,*

PUBLICATIONS

Tackling Heterogeneous Traffic in Multi-access Systems via Erasure Coded Servers [Poster][Paper]
***T.Choudhury**, W.Wang, G.Joshi*, ACM Mobihoc 2022
Best Paper Award, Poster presented at ACM Sigmetrics'22

Job Dispatching Policies for Queueing Systems with Unknown Service Rates [Paper]
***T.Choudhury**, G.Joshi, W.Wang, and S.Shakkottai*, ACM Mobihoc 2021

Sequential Mode Estimation with Oracle Queries [Paper]
*D.Shah, **T.Choudhury**, N.Karamchandani and A.Gopalan*, AAAI 2020

Top-m Clustering with a Noisy Oracle [Paper]
***T. Choudhury**, D. Shah and N. Karamchandani*, National Conference on Communications 2019

INTERNSHIP

Cheap Distributed Inference for Large ML models
Guide: Yu Cheng, Anand P. Iyer, Microsoft Research at Redmond   *Summer'22*

Description: Experimentation with ensemble techniques and gating mechanisms on models of different sizes to identify the fastest model for inference without incurring any impact on loss. Worked on various task-specific systems (Bert→Speech/ResNet→Image) and combined models from different classes within the same task (ensemble of ResNet and MobileNet for image classification) to improve performance.

Micro-magnetic simulation for Spintronics biosensors
Guide: Prof. Niladri Banerjee, Loughborough University   *Summer'17*

**AWARDS &
ACHIEVEMENTS**

Best Paper Award at ACM Mobihoc'22
Jack and Mildred Bowers Scholarship in Engineering 2022-23
CIT Dean's fellowship 2019
Excellence in Teaching Assistant twice at IIT-Bombay
All India Rank of 779 in JEE Advanced 2014
KVPY fellowship by the DST, Govt. of India 2013
Top 300 (top 1%) in Indian National Physics Olympiad and Astronomy Olympiad

RELAVANT COURSES	**Machine Learning**: Machine Learning, Deep Learning, Convex Optimization, Stat and ML methods, Distributed and Federated Learning **Statistics**: Probability, Intermediate Statistics, Alg. and Ana. for Scalable Computing Systems, Markov Chains and Queuing Systems, Random Graphs, Internet Economics	

PROGRAMMING	Python, Matlab, C++, PyTorch, LaTeX	

REVIEWER	International Symposium on Information Theory (ISIT)	2021
	IEEE/ACM Transactions on Networking	2022, 2024

TEACHING ASSISTANT	Intro to ML for Engineers, *CMU*	Fall'21
	Intro to ML for Engineers, *CMU*	Fall'22
	Communication systems, *IIT Bombay and IIT Dharwad*	Spring'19
	Statistical signal processing, *IIT Bombay*	Fall'18

ONGOING RESEARCH PROJECTS	**Scheduling mechanism for Distributed Machine learning** *Guide : Prof. Weina Wang, Prof. Gauri Joshi* **Understanding Priority Algorithms using Routing Tuples** *Guide : Prof. Weina Wang, Prof. Gauri Joshi*	

SELECT COURSE PROJECTS	**Load Balancing for Parallel Jobs**, *Alg. and Ana. for scalable computer system*	*Spring'21*
	Predicting Taxi Trip Duration in New York City, *Intro to ML*	*Fall'19*
	Biased walk on random graphs,	*2018*
	Information Spread, *SRE*	*Spring'18*
	De-reverberation of speech signal, *Advanced Signal Processing*	*Fall'17*