

SARTHAK CHAKRABORTY

✉ Sarthak Chakraborty | ✉ sarthak.chakraborty@gmail.com | 🌐 sarthak-chakraborty | 🐙 sarthak-chakraborty.github.io

RESEARCH INTERESTS

My broad research interests lie in the area of applying machine learning for an optimized system design. More specifically, I am interested in building reliable and intelligent systems that are optimized in performance and can heal itself during any faults. Apart from this, I also have experience with data-driven systems, distributed systems and federated learning.

EDUCATION

- **Dual Degree (B. Tech + M. Tech) in Computer Science and Engineering** July 2016 - April 2021
Indian Institute of Technology Kharagpur, India
Cumulative GPA: 9.74/10.00 (Class Rank 2)

PUBLICATIONS AND PREPRINTS

- [1] Mohammad Azam Ikram, *Sarthak Chakraborty*, Subrata Mitra, Shiv Kumar Saini, Saurabh Bagchi, Murat Kocaoglu. **Root Cause Analysis of Failures in Microservices through Causal Discovery**. In *Advances in Neural Information Processing Systems* (NeurIPS '22), 2022. [Link to be updated] (Acpt. Rate: 25.6%)
- [2] *Sarthak Chakraborty*, Sandip Chakraborty. **Proof of Federated Training: Accountable Cross-Network Model Training and Inference**. In *2022 IEEE International Conference on Blockchain and Cryptocurrency (ICBC '22)*, pp. 1-9, 2022. [LINK] (Acpt. Rate: 18.6%)
- [3] Lovish Chopra*, *Sarthak Chakraborty**, Abhijit Mondal, and Sandip Chakraborty. **PARIMA: Viewport Adaptive 360-degree Video Streaming**. In *Proceedings of the Web Conference 2021 (WWW '21)*, pages 2379–2391, 2021. [LINK] (Acpt. Rate: 20.6%)

PATENTS

- [1] *Sarthak Chakraborty*, Sunav Choudhary, Atanu R. Sinha, Sapthotharan Nair, Manoj Ghuhan A, Yuvraj Gagneja, Atharva Anand Joshi, Atharv Tyagi, Shivi Gupta. **Generating Concise and Common User Representations for Edge Systems from Event Sequence Data stored on Hub Systems**. [Filed] (US Patent App. 17/849,320)

WORK EXPERIENCE

- **Research Associate - Adobe Inc. (BigData Intelligence Lab)** Jul 2021 - ongoing
Group: Data-driven Systems, Insights and Experience *Bangalore, India*
 - * Worked on multiple topics including performance diagnosis and system reliability for service systems, edge aware user profile compression, compute optimization and budget constrained media spend strategy
 - * Submitted papers, filed patents and involved in delivering research technologies to product teams
 - * Participated and selected as a speaker in a global internal technical conference at Adobe Tech Summit 2022
- **Research Intern - Adobe Inc. (BigData Intelligence Lab)** Apr 2020 - Jul 2020
Topic: Architecting Large-Scale Asynchronous Federated Learning *Sunav Choudhary, Manoj Ghuhan*
 - * Designed a scalable and flexible framework for federated learning to support synchronous and asynchronous model training, with on-device learning on heterogeneous target devices including android mobiles, web browsers, and desktops
 - * Devised an algorithmic strategy to effectively aggregate stale gradients and deployed the framework on over 100 clients to perform image classification boundary prediction task with real world production model
- **MITACS Globalink Research Intern - University of Waterloo** May 2019 - Aug 2019
Topic: Advanced Optimization Methods for Machine Learning | *Github Link* *Dr. Hans de Sterck*
 - * Designed a randomized ALS algorithm targeted for CP Decomposition and Completion of Sparse Tensors
 - * Computed leverage scores for the rows of factor matrices to sample non-zero data points using weighted reservoir sampling
 - * Performed several diagnostics and validated our method against benchmark algorithms for tensor completion like conventional ALS, SGD, CCD++ and RRALS algorithms
- **Undergraduate Research Intern - IIT Kharagpur (Funded by Shell India Pvt. Ltd.)** May 2018 - Oct 2018
Topic: Predictive Maintenance of a Sensors | *Github Link* *Dr. Swanand Khare*

- Designed an algorithm to model sensor data and generated threshold conditions to detect anomaly
- Modelled data under GMM distribution and estimated its parameters by a randomized version of EM algorithm

SELECTED PROJECTS

- Outage Prediction from Alerts generated in Production** Feb 2022 - ongoing
Adobe Research
 - Monitored alerts fired from a microservice-based system and performed domain-specific feature engineering
 - Designed and implemented an outage prediction model and leveraged Shapley value based explainability model to localize the alerts responsible for generating outage predictions
 - Implemented the inference and training pipeline and successfully incorporated the model into the engineering stack
- Runtime Prediction for Spark Jobs in Multi-Tenant System** Feb 2022 - ongoing
Adobe Research
 - Designed a pipeline to predict latencies of spark jobs having workload descriptions in a multi-tenant streaming system
 - Engineered features to capture system state and used random forest based model to forecast latency and detect anomalies
 - Worked with the engineering team to successfully transfer the model to a production system at Adobe
- Root Cause Analysis via Intervention Modelling of Faults** Nov 2021 - Sep 2022
Adobe Research
 - Designed an algorithm to detect root causes of faults in a microservice-based system by modelling faults as interventions
 - Leveraged causal modelling of metrics of various microservices with the non-faulty observational data and the faulty interventional data to identify the intervention target in the causal graph
 - Significantly reduced computation time against popular baselines and evaluated against real production data
- Edge Aware User Profile Compression** Aug 2021 - Jan 2022
Adobe Research
 - Developed an algorithm to create concise and multi-purpose user profile embedding based on the behavioral attributes
 - Involved techniques of sequence modelling and multi-task learning to generate the representations that aimed to bring user profile from the hub (cloud) to the edge servers
 - Representations were faithful to the storage constraints of the edge server, with various evaluations performed around compression and updatability of the embeddings
- Cross-Chain Training of Learning Models via Blockchain Interoperability** Aug 2020 - Apr 2021
Advisor: Dr. Sandip Chakraborty | [Paper Link](#) *Master's Thesis Project*
 - Developed an end-to-end system to train a model in a cross-silo setting over multiple smart contract enabled federated networks via the concept of blockchain interoperability
 - Incorporated permissioned blockchain networks to store auditable model states learned by the federated system
 - Constructed a relay-based cross-chain transfer mechanism to transfer the model state from one network to the other via HTTP channel. Signatures ensured verifiability and authenticity of the data transferred.
- Distributed Collaborative Editor** Jan 2021 - Apr 2021
Supervisor: Dr. Aurobinda Gupta | [Github Link](#) *Term Project - Distributed Systems*
 - Designed a distributed system where multiple users can collaborate to edit a single document at once
 - Used Operational Transformation concept to maintain consistency in the documents at the server side and the client side
 - Implemented distributed system concepts like passive replication scheme, crash detection and recovery handling in case of faults. The system used a master-worker architecture of servers.
- PARIMA: Viewport Adaptive 360-degree Video Streaming** Jul 2019 - May 2020
Advisor: Dr. Sandip Chakraborty | [Github Link](#) | [Paper Link](#) *Bachelor's Thesis Project*
 - Designed an online viewport adaptive video streaming algorithm along with a client-server streaming platform
 - Developed a novel PARIMA algorithm: an augmented Passive-Aggressive(PA) model and time series(ARIMA) model for viewport detection using video content as well as personalized head movement tracking.
 - Employed a pyramidal adaptive bitrate allocation scheme to maximize the Quality of Experience
 - Used HEVC video encoding, GPAC for segmenting video chunks and 'MP4Client' for client streaming of video
- Scalable Method for Representing Large Scale Graphs** Aug 2019 - Jan 2020
Dr. Sourangshu Bhattacharya *Term Project - Scalable Data Mining*
 - Developed a hierarchical community-detection based algorithm for network embedding of large scale graphs

- Constructed hierarchy tree using Louvain community detection algorithm and studied the community structure of the graph to establish relevant inter-community links at each hierarchy level
- Generated embedding using Node2vec/Deepwalk at each hierarchy level and combined the individual node embeddings to get the network embedding

TEACHING AND SERVICES

- **Undergraduate Teaching Assistant:** *Database Management Systems (CS43002), Theory of Computation (CS41001)*
- **Reviewer:** Reviewed papers for DSN 2022, SRDS 2022, TSNM 2022, EuroSys 2023

MISCELLANEOUS PROJECTS

- Developed an image and content based search engine to query similar images from news article given a keyword and an image. It filters relevant documents using the keyword and then runs KNN over R-Tree to retrieve images similar to the query image
- Implemented max and min heap using a Loadable Kernel Module by maintaining concurrency control so that multiple processes can access it and have individual heap structures.
- Implemented Multimodal Emotion Classifier for conversations in MELD datasets. Used visual, textual, and acoustic as individual modes and trained a trimodal classifier to combine the activations from these modes
- MRP: Implemented a reliable message-oriented communication protocol over an unreliable User Datagram protocol
- Developed APIs of memory-resident file systems for linked-list based FAT and indexed based inode implementations
- TinyC: Implemented a compiler for a subset of C functionalities to translate the C code to x86 Assembly Language
- KGP-RISC: Designed a 32-bit single cycle CPU(RISC based architecture) in Verilog VHDL and tested it on FPGA
- Developed a software in Java following Waterfall SDLC model and UML diagrams to automate the operations of a library

SKILLS

- | | |
|----------------------------------|---|
| • Languages | Python, C, C++, Java, SQL, Golang, Verilog, MIPS, Scala |
| • Packages and Frameworks | scikit-learn, PyTorch, Keras, TensorFlow, DGL, Tensorflow-Federated, Git, Hyperledger Fabric, MongoDB, Docker, Kafka, Spark |

SCHOLASTIC ACHIEVEMENTS

- | | |
|--|-----------|
| • Graduated with a Department Rank and an Institute Rank of 2 among all the Dual degree students | 2020 |
| • Recipient of the Goralal Syngal Memorial Scholarship awarded by the Institute for academic excellence. | 2020, '19 |
| • Acknowledged by the Department of Computer Science and Engineering for performance par excellence. | 2018 |

RELEVANT COURSES

- **Core Courses:** Algorithms, Discrete Structures, Formal Language and Automata Theory, Compilers, Computer Organization and Architecture, Operating Systems, Computer Networks, Database Management Systems, High Performance Computer Architecture, Machine Learning, Artificial Intelligence, Scalable Data Mining, Reinforcement Learning, Deep Learning, Advances in Operating Systems Design, Advanced Database Systems, Distributed Systems
- **Other Courses:** Probability and Statistics, Linear Algebra, Regression and Time Series Models, Operations Research

EXTRA-CURRICULAR

- | | |
|--|-----------|
| • Presented my work on Federated Learning at the global internal technical summit at Adobe, Adobe Tech Summit | 2022 |
| • Successfully completed the three-day long summer school on AI organized by Google Research India | 2020 |
| • Presented a talk on 'Predictive Maintenance' at the Energy Transition Technology Summit (ETTS), 2018, held at Shell Technology Center, Bangalore | 2018 |
| • Had been a member of Team AUV, IIT Kharagpur and worked on design changes to bring stability in the existing model of underwater vehicle, Kraken 3.0 | 2017-2018 |
| • Successfully completed a week-long IEEE certified workshop on Image Processing at IIT Kharagpur | 2016 |