MRITYUNJAY KUMAR
MRIT-yen-jay koo-MARR

CS Grad'20 | Stony Brook University

✉ mrkumar@cs.stonybrook.edu

🔗 blog

📍 California, US

I work at vSAN team at VMware. I have keen interest in distributed systems and system engineering. I have 5 years of experience in software development roles. I worked in multimedia team at Mediatek for two years. In 2016, I joined Data Science team at Talentica Software, as Senior Software Engineer and worked for 3 years.

CURRENT EXPERIENCE

Member Of Technical Staff   **VMware** | July 2020-till

vSAN Team

RESEARCH EXPERIENCE

Graduate Research Assistant   Advisor : Prof. Shuai Mu | June 2019 - May 2020

Distributed Multi-core Transactional Database

- Implementing asynchronous replication with multi-process paxos to remove contention

- Implemented fast header-only/compiled, C++ logging library facilitating aligned memory allocators with features like Rotating log and auto-flush in approx. 1200 LOC.

- Implemented optimistic replay protocol to guarantee serializibility and wrote verification pipeline for logged streams.

- Wrote various evaluation micro-benchmarks to inspect contention points and indenfied flows help of mutrace and gprof tool

PROJECTS

Raft   CSE 535 Asynchronous Lab

Implemented Sharded and replicated fault-tolerance key-value store based on raft protocol.

Map-Reduce library   CSE 535 Asynchronous Lab

Implemented distributed map reduce library and worker failures from paper

Backup File System in Linux Kernel   Stony Brook FSL Lab

Implemented stackable file system which takes created backup on every successful write/delete/rename having queue based version & retention policy.

Encryption based System tool for Linux Kernel   Stony Brook FSL Lab

Implemented system call for encrypt/decrypt file using AES provided by kernel crypto API.

Music Recommendation System   Data Visualization Lab

Implemented recommendation system for music based on mood and activities. Project Link

PREVIOUS EXPERIENCE

Teaching Assistant   **Stony Brook University | June 2019**

Taught Introduction to Data Science (CSE-351) with Prof. Martin Radfar. Helped to set quizes and midterms. Wrote test module to auto generate scores for homework assignments and quizes

Senior Software Engineer   **Talentica Software | April 2016 - Jan 2019**

I joined data science team as a software engineer and worked almost three years in Applications ranging from search engine, File Sync Application, NLP Pipelines, Chat bots and applying machine learning algorithms in network based solutions. We developed a novel algorithm which coins the answerability of a question.

- **Network Traffic Estimation**
  Implemented a ensembled regression model for network bandwidth prediction using SLA metrics and real feedback from native Speedtest Tool.

- **ML Model Realtime Deployment**
  Implemented an ingestion pipeline for model deployment using Storm, Kafka, Python, Java

- **Indoor Location Positioning for BLE assets**
  95% classifying the region for Static Assets and 68% accuracy in Regression based on region triangulation using RSSI and the interference correction

- **Network Traffic Identification**
  Developed a multi-modal Machine Learning model for network traffic classification for Audio and Video streaming using Ensemble of regression and auto-encoders.

- **Data dump tool for AWS Athena**
  Added an ad-hoc client to Storm topology to support collection of Real-time data to Disk & Batch pipeline using Java, Spark.

- **Live BLE asset view**
  Developed a live Bluetooth Low Energy asset view portal to support data collection team, adding 70% more correctness to site data calibration using JavaScript, Kafka, Python

- **Machine Learning Algorithm object storage service**
  Designed continuous delivery pipeline for Machine Learning models using gRPC, protobuf, Redis, RabbitMQ, AWS S3 improving deployment frequency by 40%. Wrote multiple clients in Java, python, Go, C++ to facilitate object serialization, improving frequency of model update as weekly release.

- **Single cell Identity Classification**
  Auto-encoder based neural network to identify the single cells with 83% accuracy for cell based print technology.

- **File Sync Application**
  Lead development of file sync tool developement; Implemented delta file sync Service in c++ to reduce load in sync server by 200%.; Wrote stateless software update tool to push new releases without killing of running tool; Developed a client application update framework to support release notification for multi-talent architecture using Java, C++, Python; Designed rate limiter service to prevent throttling of indexing service, improving sync frequency by 30%.

- **Financial Document Search Engine**
  Improved search relevancy by 45%. Improved Online Ontology enhancer by 30%, NLP Pipeline for Document Clustering, Keyword Extraction, Text Classification.

Software Engineer   **MediaTek India | Aug 2014 - April 2016**

Joined multimedia team and helped with development of Audio Player in feature phones. Also, I lead development of in-house tools for stress based testing for Wifi/BT/GPS.

- **Audio Player Module**
  Improved Audio Player sub-modules Playlist, Integrated BT Stack in MMI Layer

- **Internal Tool Development**
  Wrote SW Layer code for Stress Test based Combo(GPS-WiFi-BT) Tool; Implemented event driven asynchronous architecture for packet data interface

RESEARCH PAPERS

Conference Papers

1. R. Guntur and Mrityunjay Kumar. Learning to fingerprint the latent structure in question articulation. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 73–80, Dec 2018.

EDUCATION

MS in Computer Science
Stony Brook University, NY
Jan 2019 - May 2020

B.Tech in Computer Science
Maulana Azad National Institute of technology, Bhopal, India
2010 - 2014

SKILLS

Relevant Courses

Distributed Systems; Operating System

Computer Vision; Data Visualization;

Data Science Fundamentals

Technical

C++/ Python/ C/ Java/ Golang

Spark, Storm, Keras, Scikit-learn

protobuf, gRPC,SimpleRPC, Neo4j

Professional

Team player

Strong problem solver

Good time management

Workflow

Agile Development

Scrum

AWARDS

Winner for Talentica Semis Codefest
Talentica Software
Feb 2018

Runner Up Talentica prelims Codefest
Talentica Software
Jan 2018

Mediatek On-Spot VAward
Mediatek
April 2016

Mediatek On-Spot VAward
Mediatek
Jan 2016

Mediatek On-Spot VAward
Mediatek
Nov 2015

LANGUAGES

English (Fluent)

Hindi (Native)