TIANJIAN HUANG

thuang2@andrew.cmu.edu | (412) 933-9229 | https://www.linkedin.com/in/tianjian-huang/

EDUCATION

Carnegie Mellon University - *M.Sc. in Computer Networking | GPA: 3.88/4.00, Ranking: Top 3%* 09/2021 – 05/2023

Major Coursework: Introduction to Computer Systems (15-513), Storage Systems (15-746), Deep Learning Systems (10-714), Database Systems (15-645), Advanced Cloud Computing (15-719), Principles of Software Construction (17-514), C++ Game Programming (15-666)

The Chinese University of Hong Kong, Shenzhen - *B.E. in Computer Science | GPA: 3.65/4.00, Ranking: Top 10%* 09/2017 – 07/2021

Major Coursework: Operating Systems, Data Structure & Algorithm, Distributed & Parallel Computing, Software Engineering, Database Systems, Robotics Systems, Cloud Computing

SKILLS AND PROFICIENCIES

Programming Languages	C/C++, Python, Go, Java, JavaScript, MATLAB, HTML, CSS, MySQL
Frameworks and Tools	**System**: Linux, FUSE, Bash, MPI, OpenGL, Git, GCC, GDB, gRPC, BusTub; **Cloud & DevOps**: AWS EC2/S3, Terraform, Apache Spark, Docker, Kubernetes, VMware Cloud Disaster Recovery (VCDR); **Robotics**: Robot Operating System (ROS), Gazebo; **Machine Learning**: Hugging Face, Needle, CUDA;

PROFESSIONAL EXPERIENCE

Carnegie Mellon University - Research Assistant Pittsburgh, PA | 09/2022 – 03/2023

- Introduced a novel, lightweight method that allows mobile application developers to elicit scenarios and privacy risk scores from users.
- Built a labeled 300-user-scenario corpus, and trained a BERT-based Named Entity Recognition (NER) model to identify information types in the scenarios. The NER model predicts information types with a weighted average precision of 0.70 and recall of 0.72.
- Paper under review: *Mobile Application Privacy Risk Assessments from User-authored Scenarios* in IEEE RE 2023.

VMware - Software Engineer Intern Palo Alto, CA | 05/2022 - 08/2022

- Using gRPC framework with multiple languages (C, java, etc.), implemented a task-reporting service that recorded key status of backup tasks; improved the scale of Site Recovery task status tracking from protection-group-level(multiple VMs) to VM-level.
- Improved the CLI tools for VMware Cloud Disaster Recovery (VCDR) to display backup status, and to present data with access control; the new CLI tools provides detailed views of tasks and enhanced customer interaction and debugging experience.
- Refactored source code for Docker-managed backup and restore service; improved readability and consistency of source code.

Carnegie Mellon University - Robotics Institute Summer Scholar Pittsburgh, PA | 06/2020 – 08/2020

- Wrote Python scripts to extract & format GB-level VERBOSE log data of RoboTutor (an Android tablet tutoring APP).
- Used Disco (a process mining tool) to create process models for RoboTutor's log data, performed educational process mining (EPM), and summarized children-tutor interaction behavior patterns.
- Published on *RISS Journal 2020*: *Using Process Mining to Analyze Children's Interactions with RoboTutor.*

COURSE PROJECTS

BusTub DBMS 01/2023 – 05/2023

- Developed key components for a C++ relational DBMS, encompassing:
  - An LRU-K Buffer Pool Manager for overseeing in-memory database pages, bolstered by Page Guards that facilitate read/write lock controls and scope management.
  - A B+Tree index, designed to expedite access to data pages while supporting leaf scans and concurrent operations such as search, insertion, and deletion.
  - Fundamental SQL query executors, encompassing concurrent sequential scan, CRUD operations, aggregations, joins, and sort & limit functionalities with top-N optimizations.
  - Implemented concurrency control, incorporating a 2PL lock manager operative under various isolation levels, and a system for deadlock detection and recovery.

Deep Learning Framework 09/2022 – 12/2022

- Designed and implemented Needle: a non-trivial Python and C++ deep learning framework.
- Developed a Tensor library: implemented the Tensorflow Computational Graph schema, supports common matrix operations with automatic differentiation, and supports two types of C++ NDArray backend (CPU/CUDA).
- Developed a Neural Network framework with Needle, which supports sequence modeling, SGD and Adam optimizer.
- Implemented modules that support CNN and sparse matrix; tested on Penn Treebank Dataset and Cora Dataset, and achieves loss and accuracy performance comparable to Pytorch library.

Cloud File System 10/2021 – 12/2021

- Built a simple hybrid cloud file system (CloudFS) that stored small files and metadata on local SSD and large file contents in cloud via FUSE and the Amazon S3 C API.
- Performed file-content-based deduplication using Rabin Fingerprinting algorithm, achieving reduction of cloud storage cost.
- Developed file system snapshot functionality, supporting backup/restore of the entire CloudFS and snapshot previewing.
- Implemented local caching for cloud file contents, which reduced total cloud cost by 43.6% compared with the benchmark.