

ZHUOFU CHEN

✉ aetiurf@gmail.com · 🌐 chenzhuofu · 🔗 <https://chenzhuofu.github.io/>

EDUCATION

Tongji University

Shanghai, China

B.S. in Computer Science and Technology (Elite Class)

Sept. 2021 - Present

- GPA: 4.89/5.00 Ranking: 1/20

RESEARCH INTERESTS

I have broad interests in building system infrastructures to systematically bring better *performance, resilience, and usability* to *real-world applications*. Specifically, I often contemplate how to *redesign next-generation datacenter/cloud operating systems* to bridge the gap between existing hardware and emerging needs of software, and to serve numerous applications such as AI inference/training and cloud computing.

EXPERIENCE

Institute of Parallel and Distributed Systems, Shanghai Jiao Tong University

Research Intern advised by [Xingda Wei](#) and [Haibo Chen](#)

Jul. 2023 - Apr. 2024

- Developed an optimal GPU disaggregation system for transparently and efficiently serving AI applications.
- Created a theoretical model characterizing the overhead of disaggregating applications, achieving an error margin within 10%.
- Studied the lower bound of datacenter network for supporting GPU disaggregation.
- One [paper](#) is pending review.

Key Laboratory of Embedded System and Service Computing, Tongji University

Research Intern advised by [Zhijun Ding](#)

Nov. 2022 - Oct. 2023

- Implemented a WebAssembly-based runtime with an OCI shim to bridge orchestration tools and runtime.
- Invented a dynamic-import mechanism that provides an easy-to-use, isolated but sharable resource allocation mechanism for WebAssembly modules.
- One paper is pending review.

SELECTED PROJECTS

XPURemoting, a performant GPU disaggregation system

- Hijacked the CUDA driver API to transparently redirect GPU calls.
- Created a concise and elegant network abstraction for GPU disaggregation.
- Designed an easy-to-use perf tool to model the overhead of disaggregation for arbitrary applications.
- 4k LOC in C++ (v1), 15k LOC in Rust (v2).

High-performance Wasm-based serverless workflow system

- Accelerates the performance of functions by pure thread-level shared-memory communication and WebAssembly's inherent near-native execution efficiency.
- Schedules function workflow based on affinity at the upper layer.

SELECTED AWARDS

National 1 st Prize (0.55%) in Contemporary Undergraduate Mathematical Contest in Modeling	2023
Regional 1 st Prize in Contemporary Undergraduate Mathematical Contest in Modeling	2023
China National Scholarship (top 0.2%)	2022
Regional 2 nd Prize in Contemporary Undergraduate Mathematical Contest in Modeling	2022
Bronze Medal of National Olympiad in Informatics (NOI)	2020
1 st Prize of National Olympiad in Informatics in Provinces	2019