

# Chengzhi Lu

---

## CONTACT INFORMATION

Phone: 86 18565796826  
Email: chengzhilu@um.edu.mo  
University of Macau

## RESEARCH INTERESTS

Cloud Computing  
Machine Learning for Cloud Computing, System for ML&DL  
Resource Management and Task Scheduling

## EDUCATION

**University of Chinese Academy of Sciences**, China, (Sep. 2019 – Dec. 2024 (Expected))

D.Eng, Computer Applied Technology  
Advisor: Prof. **Cheng-Zhong Xu**

Zhejiang University, China, (Sep. 2016 - Sep. 2018)

M.S., Software Engineering, Sep. 2018,

Wuhan University, China, (Sep. 2012 - June, 2016)

B.S., Information Security, June 2016

## RECENT RESEARCH PROJECTS

### **System optimization for AI Applications**

- Improve performance of AI applications from perspective of the system optimization.
- Combine task scheduling and speculative execution to maximize system resource utilization.

### **Cloud Computing Resource Management, Performance Optimization**

- Design scheduling algorithms with optimistic fault tolerance mechanism to improve throughput and cluster resource utilization.

### **Resource Allocation Algorithms for Large-Scale Clusters with Performance Guarantees**

- ML-based resource configuration for co-located cluster.

## PUBLICATIONS

[**EuroSys'25**] Chen W, **Lu C**, Xu H, et al. Multiplexing Dynamic Deep Learning Workloads with SLO-awareness in GPU Clusters.

[**SC'24**] **Lu C**, Xu H, Li Y, et al. SMiless: Serving DAG-based Inference with Dynamic Invocations under Serverless Computing.

[**EuroSys'23**] **Lu C**, Xu H, Ye K, et al. Understanding and Optimizing Workloads for Unified Resource Management in Large Cloud Platforms.

[**TPDS,2022**] Luo S, Xu H, **Lu C**, et al. An in-depth study of microservice call graph and runtime performance.

[**SoCC'21**] Luo S, Xu H, **Lu C**, et al. Characterizing Microservice Dependency and Performance: Alibaba Trace Analysis (Best Paper Award).

[**CLUSTER'21**] Chen W, **Lu C**, Ye K, et al. RPTCN: Resource Prediction for High-dynamic Workloads in Clouds based on Deep Learning.

[**ICPADS'19**] **Lu C**, Ye K, Chen W, et al. ADGS: Anomaly Detection and Localization based on Graph Similarity in Container-based Clouds.

[ICPADS'18] Ye K, Kou Y, **Lu C**, et al. Modeling Application Performance in Docker Containers Using Machine Learning Techniques.

[Big Data'17] **Lu C**, Ye K, Xu G, et al. Imbalance in the cloud: an analysis on Alibaba cluster trace (**Cited 211**).

## ACADEMIC EXPERIENCE

### University of Macau

Research Assitant

Feb. 2022 - Now

- Resource allocation algorithms for Large-Scale clusters with performance guarantees.
- Resource assignment for Serverless applications
- Request scheduling for LLM applications.

### Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Visting Student

July 2017 - Sep. 2019

- Distributed System resources management, including Microservices and DL Distributed training.
- Resource allocation algorithms for Large-Scale clusters with performance guarantees.
- Anomaly detection Algorithm for Large-Scale clusters.
- Server cluster administrator. Manage more than 140 physical servers, and provide experimental environment and services including Docker, GPU, Spark, Kubernetes, etc.

## WORK EXPERIENCE

### Alibaba Group, Zhejiang, China

Academic Interns

June 2020 - Feb. 2021

- Participate in the functional development of Alibaba infrastructure platform.

## HONORS AND AWARDS

Dean's award for academic performance, SIAT, CAS

Jan. 2021, 2024

Outstanding student, UCAS,

Jan. 2022

Outstanding student, SIAT, CAS

Feb. 2019

Outstanding graduates, ZJU

July 2018

## SERVICES

Teaching Assistant of Computer Network, SIAT, CAS

2021

## SKILLS

- Java, Python, GO, Git, Latex
- Microservice, Kubernetes, Docker