Jiuchen Shi

Research interests: cloud native, microservice, regionless

■ Homepage: shijiuchen.github.io ■ Tel: (+86) 198-2128-8336 ■ E-mail: shijiuchen@sjtu.edu.cn

Education

2019/09-now **Shanghai Jiao Tong University** **Computer Science** **PhD Candidate**

- **4 published papers (1st author; CCF-Ax1; CCF-Bx3), co-authored papers (CCF-Ax2;CCF-Bx1)**
- Responsible for 2 research projects, including: cross-region VM scheduling and network optimization, etc.
- GPA 3.86 (4.0); TA of Advanced Computer Architecture

2015/09-2019/06 **Dalian University of Technology** **Software Engineering** **Undergraduate**

- National Scholarship, First Class Scholarship for Learning, Social Practice Scholarship, etc.
- Responsible for 1 innovation project for college students, 1 Google collaboration project, 2 published papers
- GPA 4.07 (5.0); Ranking: 8/284 (2.8%)

Papers

Nodens: Enabling Resource Efficient and Fast QoS Recovery of Dynamic Microservice Applications in Datacenters **1st author Published USENIX ATC 2023**

- This work considers the load and call graph dynamics in microservices. Based on the load blocking relationships, Nodens is proposed. Utilizing network monitoring, load prediction, load blocking updates, and queue draining, Nodens achieves fast QoS recovery of microservices and high resource efficiency. Compared to state-of-the-art works, Nodens reduces the QoS recovery time by 10X while ensuring high resource efficiency.

Characterizing and Orchestrating VM Reservation in Geo-distributed Clouds to Improve the Resource Efficiency **1st author Published SoCC 2022**

- This work analyzes the VM request patterns of the top 20 tenants in public cloud. We propose a resource orchestration and VM scheduling system called ROS for the Geo-distributed DCs. ROS consists of a resource predictor, a multi-tenant multi-region orchestrator, and a scheduling compensator. ROS can meet the SLAs of different tenants while reducing the total costs of resource reservation by over 50%.

QoS-awareness of Microservices with Excessive Loads via Inter-Datacenter Scheduling **1st author Published IPDPS 2022**

- This work focuses on peak load scenario for microservices and utilizes remote DCs for scaling. Considering both compute and network performance, we propose an online microservice deployment system called ELIS. ELIS includes a resource manager and a microservice deployer. At peak loads, ELIS can ensure the QoS of microservices and reduce the overall and remote computing resource usage by over 20% and 50%, respectively.

Reliability and Incentive of Performance Assessment for Decentralized Clouds **1st author Published JCST 2022**

- This work focuses on decentralized clouds and utilizes TEEs to perform reliable performance assessment of cloud providers, incentivizing them to provide better computing performance.

Adaptive QoS-aware Microservice Deployment with Excessive Loads via Intra- and Inter-Datacenter Scheduling **1st author Under-review TPDS 2023**

- This work considers the popular disaggregated storage and compute architecture in datacenters and efficiently deploys microservices between the two clusters. Compared to prior works, this work can reduce network bandwidth usage by more than 40% and increase peak throughput by 30%.。

Projects

Optimization of Compute/Network Costs in Regionless **Project leader** **2023/02-now**

- Collaborate with Huawei Cloud. Considering the network cost caused by the different positions between data and compute, we decide multi-tenants' VM request scheduling and the data placement among geo-distributed DCs.

Resource Reservation under Ultimate Elasticity **Project leader** **2021/09-2022/09**

- Collaborate with Huawei Cloud. Under different VM request patterns of large tenants in public cloud, this project orchestrates computing resources among geo-distributed datacenters to reduce deployment costs.

Skills

- Kubernetes, Container Runtime, Cgroups, RPC
- CET-6 538, TOEFL iBT 85, good writing and communication skills