



Guest Editorial: Special issue on Network and Parallel Computing for Emerging Architectures and Applications

Guangming Tan¹ · Guang R. Gao²

Accepted: 9 June 2021 / Published online: 13 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

The aim of this special issue is to provide an introduction to the burgeoning topic of computer system research for high-performance computers and big-data systems. It consists of 9 papers that are briefly discussed as follows:

The first article, by Jiansong Li on “Compiler-assisted Operator Template Library for DNN Accelerators” addresses an important research problem that dedicated DNN accelerators lack suitable tool chains, and thus force the programmers to write low-level assembly.

In “Compiler-assisted Operator Template Library for DNN Accelerators” the authors present TOPLib, which allows the programmers to write high-level primitives without considering many low-level details to utilize the accelerators.

The article, by Tianba Chen on “M-DRL: Deep Reinforcement Learning based Coflow Traffic Scheduler with MLFQ threshold adaption” proposes M-DRL, a deep reinforcement learning approach to optimize coflow scheduling for data-parallel clusters.

In “M-DRL: Deep Reinforcement Learning based Coflow Traffic Scheduler with MLFQ threshold adaption” the authors present DRL model. The DRL model dynamically adapts to coflow traffic characteristics by adjusting MLFQ thresholds, thus achieving more appropriate threshold granularity and lower average coflow completion time.

The article, by Zhanyuan Di on “High-performance migration tool for live container in a workflow” propose a tool for container migration, which reduces the startup time of docker containers by bypassing the process of redundant read/write file operations.

✉ Guang R. Gao
ggao.capsl@gmail.com

Guangming Tan
tgm@ict.ac.cn

¹ Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

² University of Delaware, Newark, USA

In “High-performance migration tool for live container in a workflow” the authors present a multi-container migration strategy based on pipeline. The migration time is reduced by 30%, compared to sequential migration.

The article, by Ziyu Zhang on “RDMA-based Apache Storm for High-performance Stream Data Processing” propose two new implementations for Apache Storm communication components with the help of RDMA technology. In “RDMA-based Apache Storm for High-performance Stream Data Processing” the authors present both the Netty component’s limitations on Storm performance and the increased CPU load in IPoIB communication mode. The experimental results show that the optimized Storm performance has been significantly improved.

The article, by Yang Bai on “CCRP: Converging Credit-based and Reactive Protocols in Datacenters” addresses the problems when reactive CC (i.e. DCQCN) and proactive CC (i.e. Expresspass) are mixed together. In “CCRP: Converging Credit-based and Reactive Protocols in Datacenters” the authors present a new congestion control protocol called CCRP, aiming for converging credit-based and reactive protocols in datacenters.

The article, by Jianxi Fan on “Fault-Tolerant and Unicast Performances of the Data Center Network HSDC” prove that the connectivity and tightly super connectivity of the logical graph of HSDC are both n , and give an $O(n)$ routing algorithm to find a shortest path for any two nodes in it. In “Fault-Tolerant and Unicast Performances of the Data Center Network HSDC” the authors present theoretically analyzes the connectivity of HSDC and designs a routing algorithm.

The article, by Mengshan Yu on “Location-based and time-aware service recommendation in mobile edge computing” addresses multidimensional inverse user similarity recommendation algorithm to solve the cold start problem for service recommendation system in edge computing. In “Location-based and time-aware service recommendation in mobile edge computing” the authors propose a service recommendation method based on collaborative filtering (CF) and location, by comprehensively considering the characteristic of services at the edge, mobility and demands of users at different timeperiods.

The article, by Zhou Jin on “Segmented Merge: A New Primitive for Parallel Sparse Matrix Computations” undertakes a theoretical analysis on a new technique called segment merge for accelerating the parallel sparse matrix computation. In “Segmented Merge: A New Primitive for Parallel Sparse Matrix Computations” the authors present the technique of merging the segments is implemented on GPU and the experiment results show that the performance improvement can be up to 109.15x.

The article, by Xiao Hu on “A Configurable Hardware Architecture for Runtime Application of Network Calculus” undertakes a theoretical analysis on a configurable hardware architecture to enable runtime application of network calculus. In “A Configurable Hardware Architecture for Runtime Application of Network Calculus” the authors present a prototype NoC system incorporating this hardware for dynamic flow regulation to effectively achieve QoS at runtime.

The special issue was preceded by the 17th IFIP international conference on Network and Parallel Computing, held in September 2020 in Zheng Zhou, China.

The articles have undergone rigorous peer-review according to the journal's high standards.

These 9 contributions encompass a wide range of research, thereby appealing to both the experts in the field and those who want a snapshot of the current breadth of computer system research, as well as the future directions.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.