J Supercomput (2016) 72:4415–4417
DOI 10.1007/s11227-016-1893-6

EDITORIAL

High-performance interconnection networks in the Exascale and Big-Data Era

Jesús Escudero-Sahuquillo[1] · **Pedro Javier Garcia**[2]

The Exascale and Big-Data challenges are driving the technological revolution of this decade, motivating significant research and development efforts from both industry and academia. In this context, the interconnection network plays an essential role in the architecture of HPC systems and datacenters, as the number of processing or storage nodes to interconnect in these systems is very likely to grow significantly to meet the higher computing and storage demands. Therefore, the interconnection network should provide high communication bandwidth and low latency, otherwise the network becoming the bottleneck of the entire system. In that regard, many design aspects should be considered to improve interconnection network performance, such as topology, routing algorithm, power consumption, reliability and fault tolerance, congestion control, programming models, control software, etc. This Special Issue of the Journal of Supercomputing is intended to offer an overview of the latest and most prominent efforts in the design and development of scalable high-performance interconnection networks, especially those oriented towards meeting the Exascale challenge and Big-Data demands.

In response to the call-for-papers of this Special Issue, we received very interesting submissions, from both industry and academia. Specifically, 15 papers were submitted, which have been assessed by 23 reviewers. As can be seen in the list included in this editorial paper, all the reviewers are experts of the highest level, from both industry

✉ Jesús Escudero-Sahuquillo
jesus.escudero@uclm.es

Pedro Javier Garcia
pedrojavier.garcia@uclm.es

1 Albacete Research Institute of Informatics, University of Castilla-La Mancha, Campus universitario s/n, 02071 Albacete, Spain

2 Faculty of Computer Science Engineering, Edificio Infante Don Juan Manuel, University of Castilla-La Mancha, Avda. de España s/n, 02071 Albacete, Spain

and academia, and their collaboration has been essential for the success of this Special Issue. We would like to thank their assessments, which have provided not only detailed evaluations of the submissions, but also valuable suggestions to enhance them.

As a result of the review process, we have selected eight papers, which reflect prominent efforts and advances in the design and development of scalable high-performance interconnection networks for High-Performance Computing (HPC) and Datacenter systems. Vignéras et al. present the routing architecture in the BXI interconnect technology developed by Atos, which has been designed to deal with faults transparently, so that applications remain unaffected by them. This paper also proposes several offline and online routing algorithms and analyzes their actual performance when computing routing tables. The paper by Zahid et al. proposes a novel routing algorithm for InfiniBand-based fat-tree topologies, called SlimUpdate, which employs path-preservation techniques to reduce the total amount of path modifications, compared to the out-of-the-box OpenSM fat-tree routing algorithm. The paper by Fuentes et al. reviews the main causes behind network unfairness in dragonfly topologies (including adversarial traffic patterns which can appear in real systems), and proposes a new solution using age-based arbitration to mitigate fairness issues, especially when using in-transit adaptive routing. The paper by Yebenes et al. analyzes the dynamics of congestion in different fully connected dragonfly patterns, and proposes a new queuing scheme called Hierarchical Two-Level Queuing (H2LQ), specially designed to reduce the HoL-blocking effect in fully connected dragonflies using minimal-path routing. The paper by Reaño et al. focuses on how to tune a remote GPU virtualization framework, whose communication module is implemented using InfiniBand Verbs. The experiments show great bandwidth improvements and significant reduction in execution time of applications using the remote GPU virtualization framework. The paper by Colombo et al. explains the modeling of the data-acquisition system of the ATLAS detector (which records particle-collision "events" delivered by the Large Hadron Colider at CERN, Switzerland) into a simulator environment. This paper addresses the main performance issues brought about the ATLAS software running onto a 2000-node farm communicating via TCP/IP through an Ethernet network. The paper by Shankar et al. proposes a microbenchmark suite that can be used to evaluate the performance of stand-alone Hadoop MapReduce, and demonstrates its ease-of-use with different network protocols, Hadoop distributions, and storage architectures. Finally, the paper by Andujar et al. proposes a family of tools for modeling realistic workloads that capture the behavior of MPI applications into self-related traces called VEF traces, so that simulation tools could replay MPI traffic (e.g., collective operations).

In our opinion, both the scope and the high technical quality of these papers make them very relevant for anyone involved, or just interested, in the Exascale and Big-data challenges, especially from the point-of-view of high-performance interconnection networks.

List of reviewers

Eitan Zahavi, Mellanox Technologies, Israel
Enrique Vallejo, University of Cantabria, Spain
Ernst Gunnar Gran, Simula Labs, Norway