Contribution ID: **103** Type: **Talk**

Understanding and Mitigating Interference between MPI and I/O Traffic on Fat-tree Networks

Monday, 3 December 2018 14:00 (30 minutes)

Many important HPC applications are communication-bound and/or I/O-bound. These applications depend on efficient inter-process communication and I/O operations, hence, network interference can cause significant performance degradation. Unfortunately, most modern HPC systems use the same network infrastructure for both MPI and I/O traffic, with multiple jobs sharing the system concurrently. The scarcity of studies that investigate the interference between MPI and I/O jobs leaves us with only a vague understanding of how these types of traffic interact with each other; the interference characteristics are not well explored and neither are the strategies for avoiding this interference in order to improve performance.

In this talk, we discuss the important characteristics of the interference between I/O and MPI traffic on fat-tree networks, exposing the impact of factors such as message size, job size, and communication frequency on the resulting interference. We show the extent to which MPI traffic is more sensitive to interference than I/O traffic on a fully provisioned fat-tree network, and we categorize configurations that can cause even an I/O job to be slowed by 1.9X due to interference from MPI traffic. This work has pinpointed the most significant aspect of the performance trends: the I/O-congestion threshold. This threshold refers to the frequency of sending I/O requests when MPI jobs start experiencing detrimental performance degradation due to I/O interference while, simultaneously, I/O traffic becomes relatively insensitive to MPI interference.

The insights gained from the interference characterization can be used with knowledge of the network topology to mitigate the effects of this inter-job interference on application performance. Our work shows how careful placement of jobs and I/O servers can, independently, mitigate interference. Additionally, I/O throttling can be guided by the I/O-congestion threshold to improve MPI performance by up to 200% while incurring only a 18% slowdown in the I/O performance.

Presenter Biography

Kevin Brown received his PhD in Mathematical and Computational Science from the Tokyo Institute of Technology Tokyo Tech) in 2018 under the supervision of Prof. Satoshi Matsuoka. His thesis title was "Resource Contention on HPC System" and covered novel methodologies for measuring and analyzing communication performance on HPC systems with fa-tree networks. He received his MSc. from Tokyo Tech in 2014 and his BSc. from the University of Technology, Jamaica in 2008. Kevin's areas of interests include performance analysis, MPI libraries, I/O subsystems, and resource contention.

Between 2008 and 2012, Kevin worked as a Unix Systems Administrator at Digicel Jamaica, Ltd., where he managed the organisations Linux servers and clusters. His responsibilities included managing the company's hierarchical backup infrastructure and ensuring disaster recovery strategies are in place for business critical systems. Throughout his professional and research life, Kevin has garnered experience across many areas: from implementing software and hardware solutions to management and procuring.

Primary authors: Dr BROWN, Kevin A. (Tokyo Institute of Technology); Dr JAIN, Nikhil (Lawrence Livermore National Laboratory); Dr MATSUOKA, Satoshi (RIKEN Center for Computational Sciences); Prof. SCHULZ, Martin (Technical University of Munich); Dr BHATELE, Abhinav (Lawrence Livermore National Laboratory)

Presenter: Dr BROWN, Kevin A. (Tokyo Institute of Technology)

Session Classification: HPC Technologies

Track Classification: Storage and I/O