Evaluation of SDN-based Conflict Avoidance between Inter-Node Communication and Staging Communication based on Packet Monitoring

Arata Endo[1] Chunghan Lee[2] Susumu Date[3] Yasuhiro Watashiba[3] Yoshiyuki Kido[3] Shinji Shimojo[3]

Graduate School of Information Science and Technology[1], Osaka University, Japan

Fujitsu Laboratories Ltd.[2], Japan

Cybermedia Center[3], Osaka University, Japan

1. High-Performance Computing

Today's HPC systems generally adopt cluster architecture consisting of many computing nodes connected on an interconnect. These systems generally allow multiple users to run their parallel computation simultaneously.

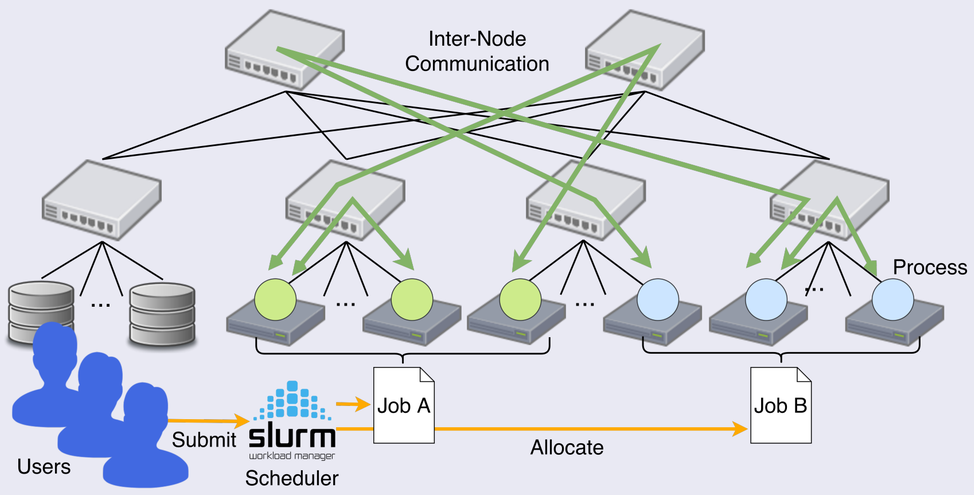

Figure 1: An Example of the HPC System

2. Two-Layered File System Structure

In many cases, HPC clusters adopt a two-layered file system structure, which is composed of global file system (GFS) providing large storage capability and local file system (LFS) providing high performance of storage I/O. In the two-layered file system structure, input and output data of an application stays on LFS while the application is running and it stays on GFS while the application is not executed. This data arrangement is realized through the use of staging, or data transfer between GFS and LFS before and after application execution.

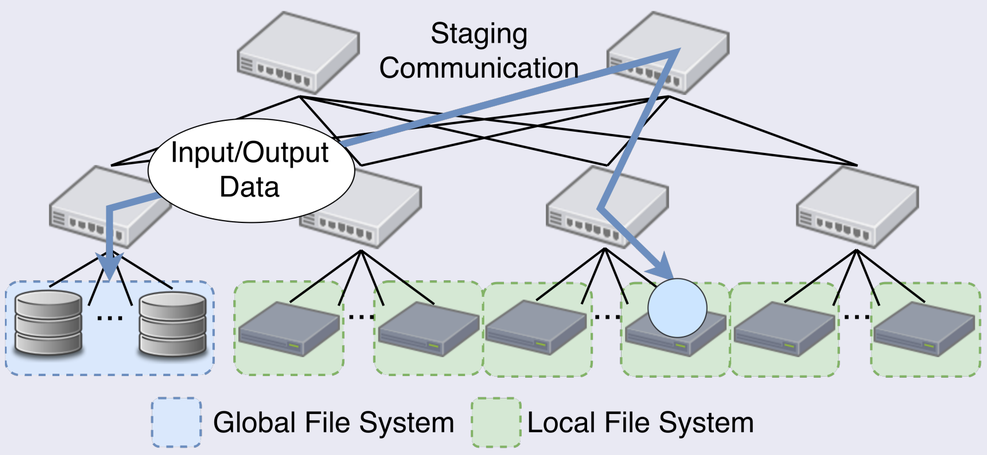

Figure 2: Two-Layered File System Structure

3. The Problem of Staging

In the case that the interconnect of the HPC cluster is not independent of the storage network between GFS and LFS, inter-node communication and staging communication share the interconnect and the traffic conflict between both type of communication may decrease the performance in each type of communication. In our research, we focus on this problem to accelerate the staging execution using the network programmability of Software-Defined Networking (SDN).

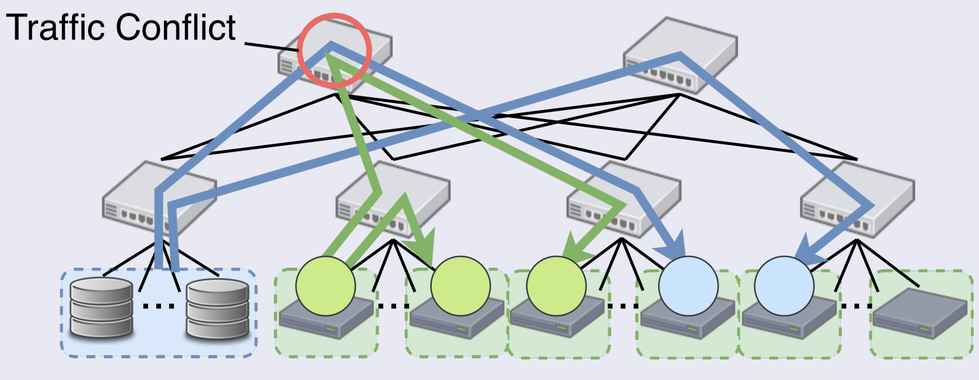

Figure 3: Traffic Conflict between Inter-Node Communication and Staging Communication

4. A Prototype of SDN-Based Conflict Avoidance

We propose link separation conflict avoidance method. We have implemented a prototype of SDN-based conflict avoidance on fat-tree topology based on our proposed method. The proposed method allocates a dedicated communication path for each of inter-node communication and staging communication to keep link capacity for the staging communication. If the path allocation is always activated, the links allocated to the staging communication are not used by any communication while the staging is not executed. Therefore, the path allocation is activated only while staging is executed in our proposed method. The proposed method uses SDN to activate the path allocation based on staging execution information sent by the scheduler of the HPC cluster.

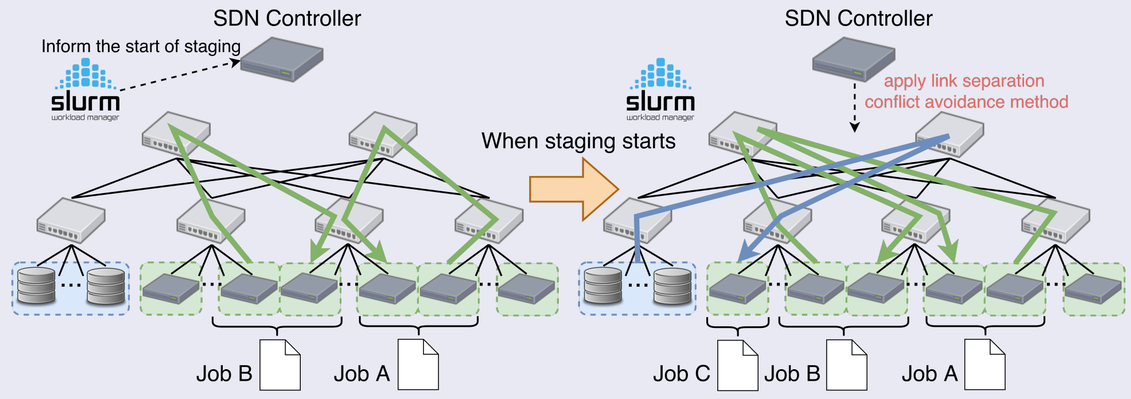

Figure 4: The Proposed Method

5. Evaluation Plan

For evaluating how the traffic conflict affects on each performance of inter-node communication and staging communication and how the proposed method can alleviate the effect of the traffic conflict, we plan to execute a packet monitoring experiment on the actual HPC cluster shown in Figure 5. In the experiment, we submit a job set to the HPC cluster and capture packets of inter-node communication and staging communication on monitoring nodes by configuring port mirroring on ofs1 and ofs2.

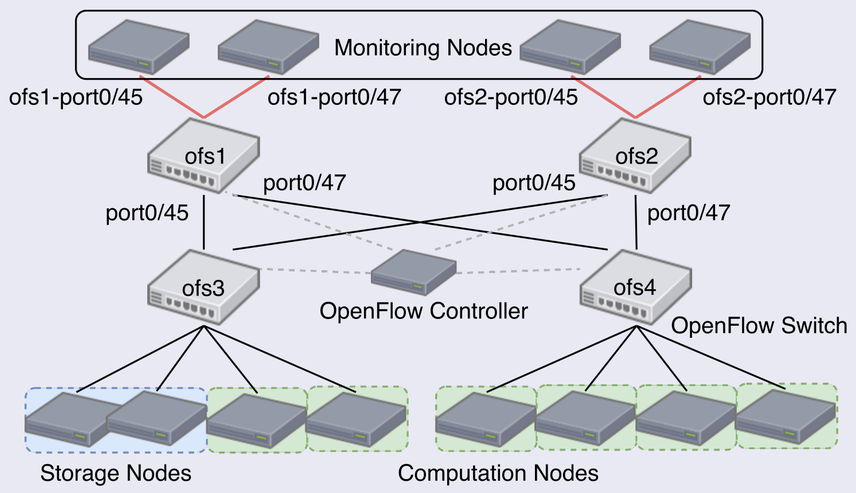

Figure 5: The Evaluation Environment

6. Conclusion

We propose link separation conflict avoidance method as a prototype method to alleviate the effect of the traffic conflict between inter-node communication and staging communication. We plan a packet monitoring experiment to evaluate how the traffic conflict between inter-node communication and staging communication affects on each performance. Also how the proposed method can alleviate the effect of the traffic conflict is investigated through the experiment.

Acknowledgements

This work is supported by JSPS KAKENHI Grant Numbers JP16H02802, JP16K12419 and JP26330145. We thank Fujitsu Laboratories Ltd. for technical advice.

Contact: endou.arata@ais.cmc.osaka-u.ac.jp