Concept and Design of SDN-enhanced MPI Framework

Keichi Takahashi¹, Dashdavaa Khureltulga¹, Baatarsuren Munkhdorj¹, Yoshiyuki Kido¹, Susumu Date¹, Hiroaki Yamanaka², Eiji Kawai² and Shinji Shimojo²

Osaka University¹, National Institute of Information Communications Technology²
Interconnect of HPC systems

- Modern **HPC (High Performance Computing)** systems are built on a computer cluster architecture.

- Due to the recent rapid scale-out in node numbers, the **cost and complexity of HPC interconnects** are getting increasingly high.

- **MPI (Message Passing Interface)** is the de facto standard library to develop parallel distributed applications on clusters.
Towards SDN-enhanced MPI

Leverage SDN architecture to realize tight integration between MPI application and network interconnect; MPI application-aware dynamic interconnect control.

Several PoCs:

• Offloading MPI broadcast to network

• Dynamic load balancing of traffic load using MPI layer
Proposed Architecture of the Framework

Overall Architecture

Tagging Mechanism

(1) Pre-install rules

(2) Packets are tagged with application level info

(3) Packet control based on tags

MPI Packet control

Tagging Module

MPI Application

MPI Library

Tagging Module

MPI Packet

User MPI Application

System Call

Network Stack

Socket API

User Space

Tagging Module

MPI Packet

To NIC

MPI Packet

Untagged MPI Packet

Tagged MPI Packet

SDN Controller

SDN Switch