High-Throughput Direct Data Transfer between PCIe SSDs

NEC

Jun Suzuki, Masato Yasuda, Masahiko Takahashi, Yoichi Hidaka, Junichi Higuchi, Yoshikazu Watanabe, and Takashi Yoshikawa

Motivation and Goals

- PCIe over Ethernet provides scalable connections of I/O devices to host server
- Lots of data transfers between I/O devices without any processing in host server
  - Data reallocation(Disk-to-Disk), Writing data to NAS (NIC-to-Disk)
- All transferred data sent to main memory of host server
- Congestion of data flows between I/O devices on server link becomes bottleneck of throughput in data transfer

*Why not transfer data directly between I/O devices?*

Proposal: Direct Connect
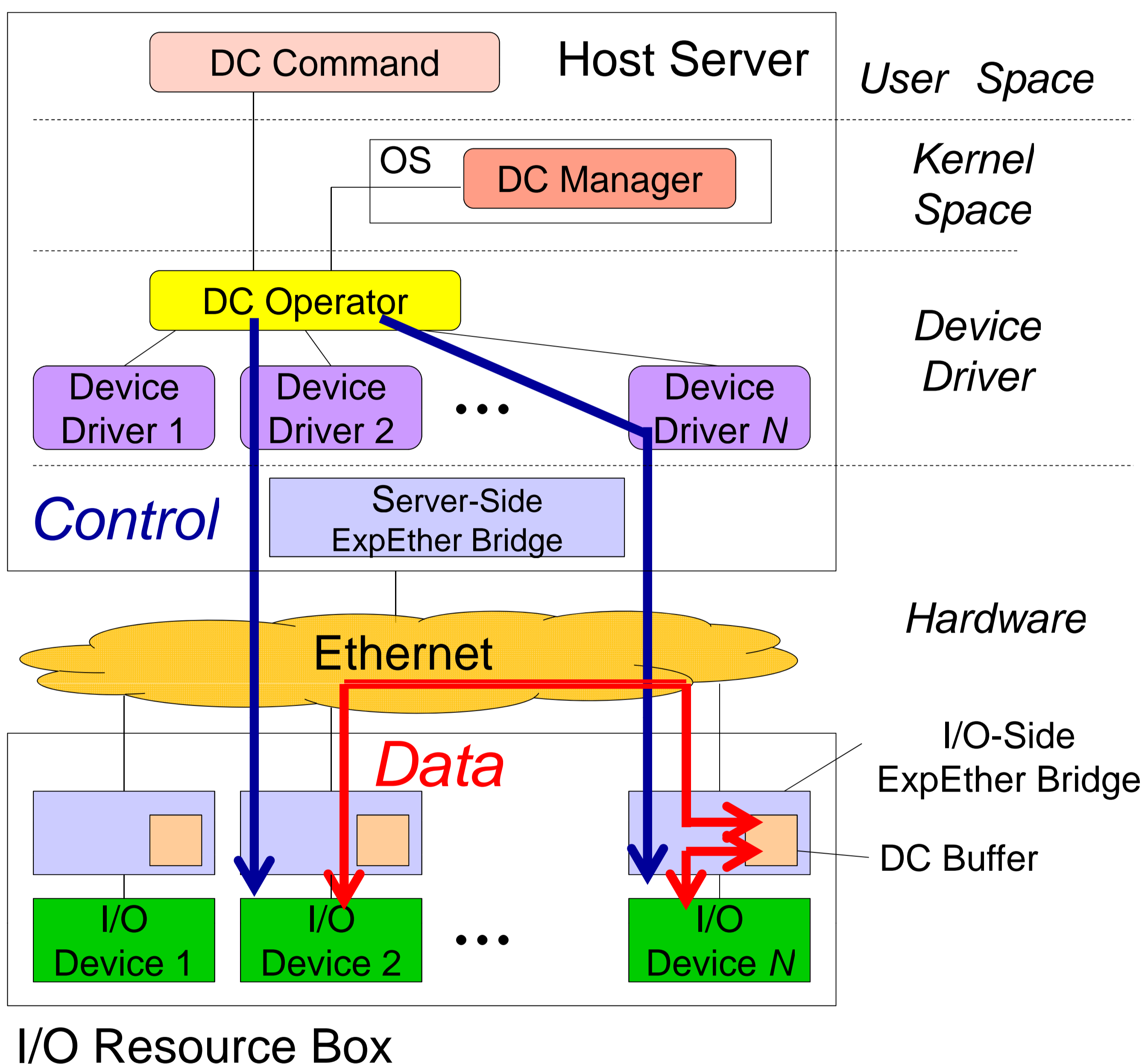
- I/O devices cannot transfer data directly with each other

- Relaying DMA of I/O devices using DC buffer in PCIe-to-Ethernet bridge

  1. DC operator makes source device write data to DC buffer

  2. DC operator makes destination device read data from DC buffer

- I/O devices are controlled by host server, while data are directly transferred between I/O devices

Results

Direct Data Transfer between PCIe SSDs
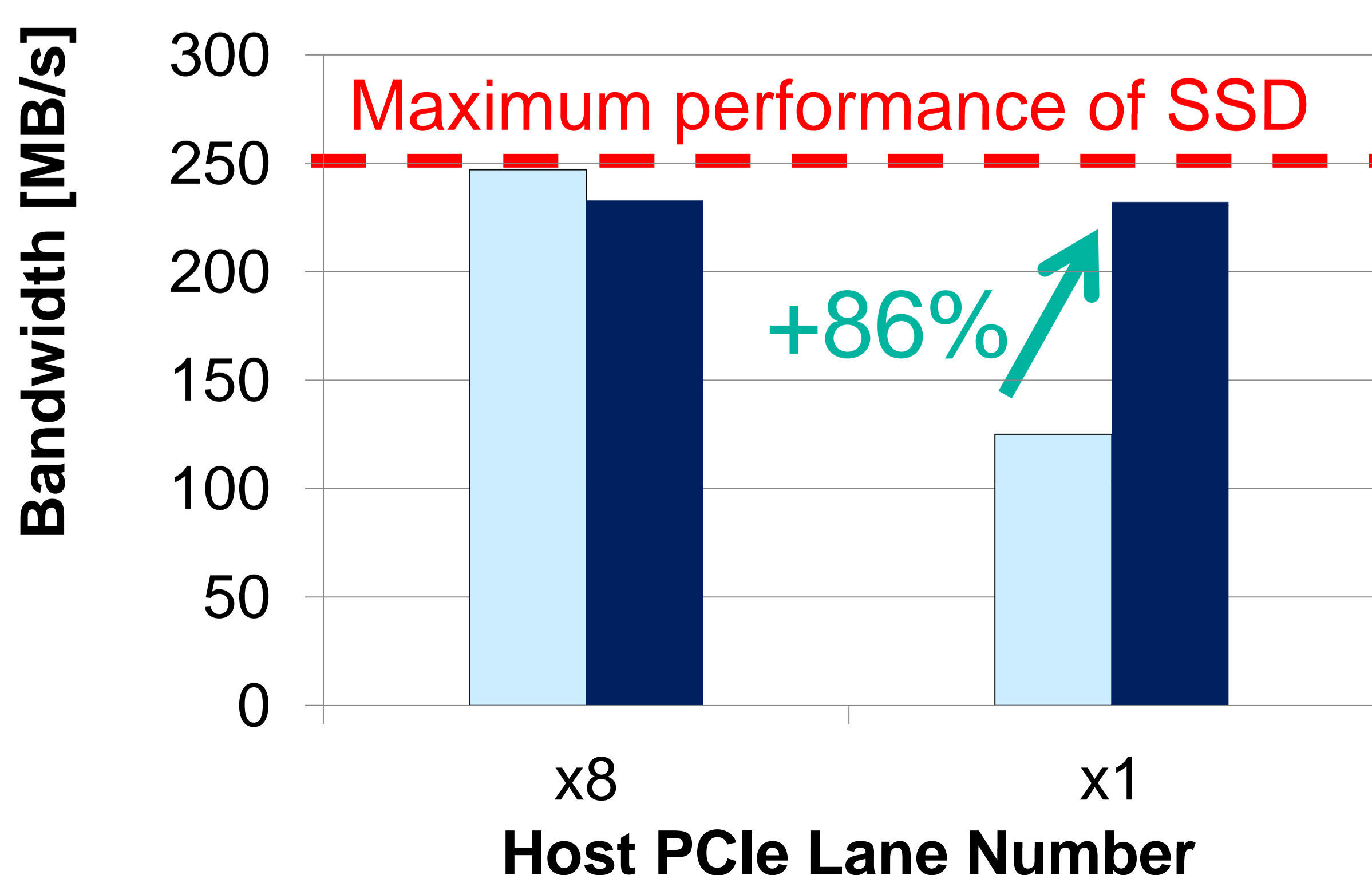
- Direct Connect enables high-throughput transfer even when server bandwidth is narrow (PCIe x1)

- ☐ Conventional Method: Data are sent to server memory

- ■ Direct Connect

Maximum performance of SSD

+86%

Bandwidth [MB/s]

Host PCIe Lane Number: x8, x1