SC21 Network Research Exhibition: Demonstration Abstract

High Performance Data Transfer Nodes for Petascale Science with NVMe-over-Fabrics as Microservice

Se Young Yu, Jim Chen, Fei Yeh, Joe Mambretti
International Center for Advanced Internet Research - Northwestern University, young.yu, jim-chen, fyeh, j-mambretti@northwestern.edu

Abstract

The PetaTrans with NVMe-over-Fabrics as microservice is a research project aimed at improving large-scale WAN microservices for streaming and transferring large data among high-performance Data Transfer Nodes (DTNs). We are designing, implementing, and experimenting with NVMe-over-Fabrics on 100 Gbps Data Transfer Nodes (DTNs) over large-scale, long-distance networks with direct NVMe-to-NVMe service connections. NVMe-over-Fabrics microservice connects remote NVMe devices without userspace applications, which will reduce overhead in high-performance transfer. The primary advantage of NVMe-over-Fabrics microservice is that it can be deployed in multiple DTNs as a container.

Goals

1. The PetaTrans 100 Gbps Data Transfer Node (DTN) with NVMe-over-Fabrics research project is aimed to improve large scale WAN microservices for high-performance data streaming and transfer using a novel NVMe-over-Fabrics technique.

2. We are designing, developing, and experimenting with 100 Gbps Data Transfer Nodes (DTNs) with NVMe-over-Fabrics as microservice over 100 Gbps Wide Area Networks (WANs) to demonstrate the capabilities of NVMe-to-NVMe direct connections in high-performance long-distance networks.

3. NVMe-over-Fabrics microservice is being designed specifically to optimize capabilities for supporting E2E (e.g., edge servers with 100 Gbps NICs and multiple NVMes) large scale, high capacity, high performance, reliable, high quality, sustained individual data streams for science research.

4. Related research includes DTN components, such as NUMA and NVMe, multiple flows of transport protocols, system and network monitoring, orchestration of multiple systems, pipelined workflows, NVMe-over-Fabrics as microservice, and other considerations.

5. Enhancements include additional capabilities for controlling NVMe-over-Fabrics using microservice architecture to establish direct a mapping between remote NVMe devices and transparent data streaming and transfer with low overhead over high-performance long-distance networks.

Resources

Required resources from SCinet are use of some portion of the 1.2 Tbps path SCinet has been asked to provision from the StarLight facility in Chicago to the StarLight booth on the SC21 showfloor.

Involved Parties

- Se-Young Yu, iCAIR, young.yu@northwestern.edu
- Jim Chen, iCAIR, jim-chen@northwestern.edu
- Fei Yeh, iCAIR,fyeh@northwestern.edu
- Joe Mambretti, iCAIR, j-mambretti@northwestern.edu

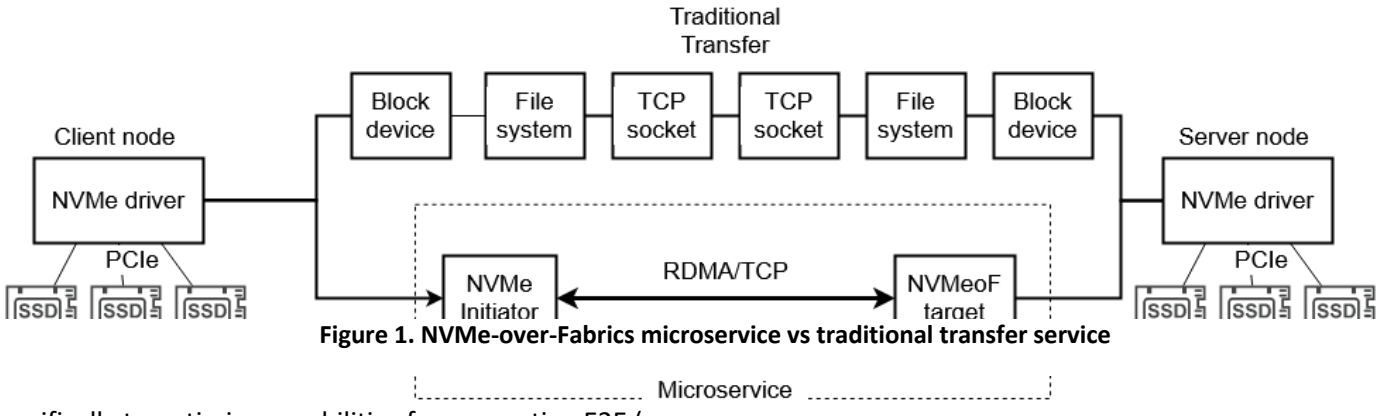

Figure 1. NVMe-over-Fabrics microservice vs traditional transfer service