SC22 Network Research Exhibition: Demonstration Abstract

N-DISE: NDN for Data Intensive Science Experiments

Submitted by the N-DISE Team

Abstract

The NDN for Data Intensive Science Experiments (N-DISE) project aims to accelerate the pace of breakthroughs and innovations in data-intensive science fields such as the Large Hadron Collider (LHC) high energy physics program and the BioGenome and human genome projects. Based on Named Data Networking (NDN), a data-centric future Internet architecture, N-DISE will deploy and commission a highly efficient and field-tested petascale data distribution, caching, access and analysis system serving major science programs. The N-DISE project will build on recently developed high-throughput NDN caching and forwarding methods, containerization techniques, leverage the integration of NDN and SDN systems concepts and algorithms with the mainstream data distribution, processing, and management systems of CMS, as well as the integration with Field Programmable Gate Arrays (FPGA) acceleration subsystems, to produce a system capable of delivering LHC and genomic data over a wide area network at throughputs approaching 100 Gbits per second, while dramatically decreasing download times. N-DISE will leverage existing infrastructure and build an enhanced testbed with high performance NDN data cache servers at participating institutions.

Goals

This demonstration is designed to exhibit improved performance of the N-DISE system for workflow acceleration within large-scale data-intensive programs such as the LHC high energy physics, BioGenome and human genome programs. To achieve high performance, the demonstration will leverage the following key components: (1) the transparent integration of NDN with the current CMS software stack via an NDN based XRootD Open Storage System plugin, (2) joint caching and multipath forwarding capabilities of the VIP algorithm, (3) integration with FPGA acceleration subsystems, (4) SDN support for NDN through the work of the Global Network Advancement Group (GNA-G) and its AutoGOLE/SENSE and Data Intensive Sciences Working Group. The demonstration activities will take place over an upgraded N-DISE WAN testbed connecting participating institutions (Northeastern/MGHPCC, Caltech, UCLA, Tennessee Tech, and StarLight Chicago) including several 100G links, and 4 X 100G connectivity between Caltech and the CENIC PoP in Los Angeles.

Integration of NDN with CMS Software Stack

We will demonstrate a transparent integration of NDN with the CMS Software components (CMSSW) via an NDN based XRootD Open Storage System plugin [1]. This includes developing and NDN consumer embedded into the OSS plugin, capable of translating all related file system calls needed for file transfers into NDN Interest packets and send them to the local high performance NDN-DPDK forwarder [2]. We will also demonstrate the NDN producer which has access to HEP event files on its local storage, present in the NDN network, and capable of responding with NDN Data to the Interest queries made by one or many consumers. Both applications will make use of the NDNc library developed by the N-DISE team, which offers an integration of the ndn-cxx (NDN C++) library with the DPDK based NDN forwarder using the CISCO-developed shared memory interface (memif). The demonstration will showcase transfers of files or directory of files using the NDN consumer application as well as CMS job completion using the NDN XRootD plugin.

Joint Caching and Multipath Forwarding

We will demonstrate the enhanced performance of the adaptive, distributed VIP joint caching and forwarding algorithm [3], which has been implemented with the NDN-DPDK forwarder over the N-DISE WAN testbed. First, we will demonstrate a multi-threaded version of the VIP algorithm, which implements split-VIP queues into multiple forwarding threads. This is expected to remove bottlenecks and significantly increase total throughput over the 100 Gbps high performance WAN testbed. Second, we will demonstrated an updated version of the VIP algorithm, which jointly optimizes caching and forwarding for real CMS data of varying size. Finally, we will demonstrate the joint multipath forwarding and caching capabilities of the VIP algorithm, over complex network scenarios enabled by the N-DISE WAN testbed. Specifically, path diversity and multipath forwarding are expected to yield more caching opportunities, leading to improved performance in terms of throughput, delay, and cache hit rates.

FPGA Acceleration Subsystems

We will demonstrate the performance acceleration a FPGA can offer when offloading NDN-DPDK workloads. The demonstration will show that the original slowdowns in NDN-DPDK forwarding have been minimized. The slowdowns originally identified in the NDN-DPDK forwarder are due to the hash computation and dispatch thread table look up. Although an ideal integrated solution would have FPGA acting as an accelerated network interface card, our demonstration instead has the CPU communicating to the FPGA, causing some potential performance penalties.

SDN Support for N-DISE

SDN support for NDN in a multi-domain testbed is advancing rapidly through the work of the Global Network Advancement Group (GNA-G) and its AutoGOLE/SENSE and Data Intensive Sciences Working Groups. We will use a set of SENSE-enabled virtual circuits with bandwidth guarantees spanning multiple network domains, with nodes at the N-DISE collaborating sites (Caltech, Northeastern, Tennessee Tech, UCLA) and others including StarLight and SCinet, and an intelligent data plane and control plane using P4 and the Reservoir Labs gradient graph (G2) software. This installation will support decisions on impactful flow (group) identification, gradient graph calculations providing recommendations for congestion resolution or avoidance by moving selected flows to alternative paths and/or adjusting virtual circuit capacities. The P4 user-defined label parsing and matching capabilities will be used to enable flow-group identification and express flow handling requirements and decisions, and in-band telemetry (INT) will be used to track the end-to-end state and progress of flow-group transfers.

Demonstration Testbed

The N-DISE WAN testbed builds on the testbed of the SANDIE (SDN-Assisted NDN for Data Intensive Experiments) project, and connects N-DISE nodes located at Northeastern University (hosted at the Massachusetts Green High-Performance Computing Center (MGHPCC)), Caltech, UCLA, Tennessee Tech, and StarLight (Chicago). The nodes are all equipped with high-end Intel Xeon or AMD EPYC processors, large pools of DRAM memory and NVMe SSD storage, as well as high-performance intelligent NICs, including Mellanox ConnectX-5 and ConnectX-6. The path from Caltech to MGHPCC supports rates of up to 100 Gbps and will be controlled by SENSE, which

can provide virtual circuits with a 100 Gbps bandwidth guarantee. The layer 2 demo topology is provided by SCinet in collaboration with Internet2, ESnet, CENIC and other regional providers, to the nodes at NEU/MGHPCC, Caltech, UCLA, Tennessee Tech and StarLight. Each node is connected to all other nodes and to the booth at SC22 in layer 2 through the use of VLANs. The NDN-DPDK forwarder will be used for the demo, which runs on top of this layer 2 topology as the layer 3 forwarder.

Involved Parties

- Edmund Yeh, Northeastern Univ., eyeh@ece.neu.edu
- Harvey Newman, Caltech, newman@hep.caltech.edu
- Lixia Zhang, UCLA, lixia@cs.ucla.edu
- Jason Cong, UCLA, cong@cs.ucla.edu
- Susmit Shannigrahi, Tennessee Tech, sshannigrahi@tntech.edu
- Yuanhao Wu, Northeastern, wu.yuanh@husky.neu.edu
- Volkan Mutlu, Northeastern, fvmutlu@ece.neu.edu
- Yuezhou Liu, Northeastern, liu.yuez@husky.neu.edu
- Catalin Iordache, Caltech, catalinn.iordache@gmail.com
- Justas Balcas, Caltech, jbalcas@caltech.edu
- Raimondas Sirvinskas, Caltech, raimis.sirvis@gmail.com
- Sichen Song, UCLA, songsichen123@gmail.com
- Michael Lo, UCLA, milo168@g.ucla.edu
- Sankalpa Timilsina, Tennessee Tech, sankalpatimilsina12@gmail.com
- Davide Pesavento, davide.pesavento@nist.gov
- Chengyu Fan, Colorado State University, chengy.fan@gmail.com

References

[1] HEP Applications: Iordache, Cătălin, et al. "Named Data Networking based File Access for XRootD." *EPJ Web of Conferences*. Vol. 245. EDP Sciences, 2020.

[2] NDN-DPDK: Shi, Junxiao, Davide Pesavento, and Lotfi Benmohamed. "NDN-DPDK: NDN Forwarding at 100 Gbps on Commodity Hardware." *Proceedings of the 7th ACM Conference on Information-Centric Networking*. 2020.

[3] Yeh, et al. "VIP: A framework for joint dynamic forwarding and caching in named data networks." *Proceedings of the 1st ACM Conference on Information-Centric Networking*. 2014.