Pre-Proposal Cover Page

Project Title: Integrated Active and Passive Performance Monitoring within Advanced High-Bandwidth Networks

Applicant/Institution: Philip DeMar
Fermi National Accelerator Laboratory
Wilson & Kirk Roads
Batavia, Illinois 60510-5011

Postal Address: P.O. Box 500; M.S. 368
Batavia, Illinois 60510-5011

Principal Investigator: Philip DeMar
Email: demar@fnal.gov
Tele.: 630-840-3678

Other Key Senior Personnel: Dr. Wenji Wu
Email: wenji@fnal.gov
Tele.: 630-840-4541

Administrative Point of Contact: Philip DeMar
Email: demar@fnal.gov
Tele.: 630-840-3678

Estimated Funding Request: $250k/yr for two years

Funding Opportunity FOA Number: LAB 14-1003

DOE/Office of Science Program Office: Office of Advanced Scientific Computing Research

DOE/Office of Science Program Office Technical Contact: Dr. Steven Lee
DOE/SC/ASCR
Phone: (301) 903-5710

Research Area(s): Next Generation Networking for Science

Integrated Active and Passive Performance Monitoring within Advanced High-Bandwidth Networks

I. Problem Space

Today, 10GE-connected systems are becoming the defacto standard for computing centers involved in big data movement. The network infrastructure is correspondingly evolving toward deployment of 40GE and 100GE technologies. Projecting this trend out to a five year time horizon, one can envision 40GE-connected host systems being supported over network infrastructure based on 400GE in the network core and out into the WAN, with Nx100GE becoming the base-level "commodity" connection technology out to the access layer.

Software-defined networking (SDN) is a concept to decouple network traffic control decisions from the actual forwarding of packets within the network infrastructure. SDN enables customization of network paths for specific classes or types of traffic, up to a per-application level of customization. OpenFlow has emerged as a network flow-based open standard to implement SDN, and is gaining wide traction among router and switch vendors. It is just a short jump to envision end-to-end network paths customized for extreme data applications that would run across very high performance (400GE) network infrastructure. Functionally, such end-to-end customized network paths act as application-driven private virtual networks.

LHC data movement represents a very real use case for such application-driven private networks. The LHC experiments have developed a highly complex distributing computing model that moves extreme amounts of data across a global scale. To support this data movement, the experiments have been at the forefront of a trend within the R&E community to separate large-scale science data movement from "normal" Internet traffic. The objective is to avoid a classic "elephants and mice" scenario of bulk traffic interfering with user interactive types of traffic. Somewhat ironically, even the elephant (science) traffic can suffer major performance degradation at extreme data rates with just minor packet loss. To facilitate this data movement isolation, as well as enable customized network paths to satisfy application-specific requirements, the experiments are already utilizing physically private networks, such as the LHC Optical Private Network (LHCOPN), and making use of point-to-point virtual circuit technologies, such as OSCARs. It is a short step to project future LHC-class data movement based entirely on point-to-point and multi-point-to-multipoint virtual networks embedded within high capacity (400GE+) network infrastructure, using OpenFlow-like technologies.

PerfSONAR (PS) has emerged in recent years as the "open" standard for monitoring network infrastructure within the R&E community. PerfSONAR provides two essential functions for the current R&E network environment. First, it provides a standard framework for general access to network monitoring & measurement data across network paths that typically traverse multiple domains and mixed vendor equipment. Second, it conducts active network performance measurements (latency/loss and throughput) across the network infrastructure that supports those paths. However, it is the limitations in this second function that will likely prevent PerfSONAR from being useful within very high performance networks (100GE+) that will support virtualized "private" network paths. Specifically:

- PS servers are typically very low cost, commodity "white box" systems deployed at strategic points in the network. The mismatch between the network connections of these systems (1GE, slowly migrating to 10GE) and the high performance network links (100GE+) distorts the performance measurements. The disparity between network link performance capability and commodity system performance capability will almost certainly widen over time. Upgrading the PS server capabilities to be closer to leading edge network technology capabilities would be cost prohibitive and not realistically scalable.
- PS active measurements were developed for general routed networks. It will be extremely difficult to adapt these server-based measurement capabilities to cover private end-to-end virtual network paths. As these end-to-end paths

become more dynamic in nature, that challenge will become even more daunting.

II. Our Solution: Integrate active Performance measurement within the network infrastructure

Our proposed solution is to implement active performance measurement capabilities within the network infrastructure itself. We envision developing active measurement agents that will run directly on high performance routers and switches. These agents could be called by the PerfSONAR framework infrastructure for path specific active measurements, either on a persistent or demand basis. However, we don't target our work to be PerfSONAR-specific. Our objective is to develop a standard active measurement capability native to the network infrastructure. If widely deployed, our solution would provide the following benefits:

- There would no longer be a need to deploy external servers for active measurement purposes, and have to deal with the scaling issues of striving for ubiquitous coverage.
- Active performance measurements would be matched to the bandwidth characteristics of the network path, eliminating 1GE and 10GE-driven tests over 100GE+ infrastructure.
- A customizable active performance measurement capability would exist for virtualized private networks, not merely just from an end-to-end perspective, but also across individual segments along the private network path.

Our belief is that applications driving extreme data movement over high performance networks will eventually not only be capable of establishing customized private network paths, they will also be capable of monitoring those paths for expected performance levels. A native, integral performance measurement capability within the network infrastructure could eventually lead to active adjustment or correction of the virtual path by the application to satisfy its requirements.

Ultimately, we believe that active performance measurement should be integrated into the network infrastructure simply because that's where it belongs. SDN and OpenFlow seek to decouple routing decisions from packet forwarding, in order to create a more flexible and agile development environment for network paths. Similarly, we seek to decouple active network performance measurements from the general PerfSONAR framework collection, archiving, and presentation function in order to create similar flexibility and agility in development and use.

III. Challenges

We recognize our proposed work presents a significant number of technical challenges:

- Establishing co-operative efforts with router and switch vendors to adapt their platforms to support the capability we propose. Most routers/switches run proprietary O/S. We require their active support to be successful.
- Coupling of active measurement services with OpenFlow topology and discovery services.
- Router/switch resource contention issues. Our measurement processes must not impede device forwarding functions.
- The current active throughput testing (BWCTL) performed by PerfSONAR may be too intrusive at extremely high bandwidth. An alternative of using queue length and elasticity to derive throughput results in place of actual active measurement will be investigated.

IV. Project Road Map

Our proposal is intended to be a two-year design study on the feasibility of conducting active performance measurements natively from network infrastructure. Our project deliverables will be a design study document and the accompanying prototypes developed as part of those studies.

- Year 1: Work with a collaboration partner (vendor) to develop a prototype active performance monitoring capability that includes one-way delay & loss, as well as throughput tests. Then objective will be to support both persistent and on-demand tests. We will use the PerfSONAR services package to schedule and collect the measurements from our agents on the network platform, then archive the results. However, our focus will be on measurement agents themselves and on how well they work as part of the network device.
- Year 2: Pursue collaborative efforts with other major router/switch vendors to develop a similar capability on their platforms. We will also investigate passive throughput measurement through monitoring of queue lengths and elasticity.

Pre-Proposal Conflict-of-Interest (COI) Page

Project Title: Integrated Active and Passive Performance Monitoring within Advanced High-Bandwidth Networks

Applicant/Institution: Philip DeMar
Fermi National Accelerator Laboratory
Wilson & Kirk Roads
Batavia, Illinois 60510-5011

Postal Address: P.O. Box 500; M.S. 368
Batavia, Illinois 60510-5011

Philip DeMar – Project PI:

1. Past/Current Collaborators:

Collaborator	Institution	Project/Paper	Date(s)	Status
Dr. Dantong Yu	Brookhaven NL	ESCPS	2009-2012	complete
Dr Martin Swany	Indiana University	ESCPS	2009-2012	complete
Brian Tierney	Lawrence Berkeley NL	E-Center	2009-2012	complete

2. Past/Current C-Editors: None

3. Past/Current Graduate Advisors & Advisees: None

Dr Wenji Wu – Key Senior Personnel:

1. Past/Current Collaborators:

Collaborator	Institution	Project/Paper	Date(s)	Status
Dr. Dantong Yu	Brookhaven NL	MDTM	2013-2016	ongoing

2. Past/Current C-Editors: None

3. Past/Current Graduate Advisors & Advisees: None