Pre-Proposal Cover Page

Project Title: Network Traffic Monitoring and Analysis with GPUs

Applicant/Institution: Dr. Wenji Wu
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: Dr. Wenji Wu
Email: wenji@fnal.gov
Tele.: 630-840-4541

Co-Principle Investigator: Phil DeMar
Email: demar@fnal.gov
Tele.: 630-840-3678

Administrative Point of Contact: Dr. Wenji Wu
Email: wenji@fnal.gov
Tele.: 630-840-4541

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): Terabit network technologies, high-performance networking

Network Traffic Monitoring and Analysis with GPUs

I. The Problem

Network traffic monitoring and analysis is the process of capturing network traffic and inspecting it closely to determine what is happening in the network. Since the inception of computer networks, traffic monitoring and analysis has been an indispensable tool in network operations and management, capacity planning, and performance troubleshooting. The DOE is working towards deploying terabit networks in support of distributed extreme-scale data movement. Existing backbone networks are now being upgraded with ultra-scalable 100-Gigabit technologies. At such high data rates, network traffic monitoring and analyzing applications, particularly those involved in traffic scrutiny on a per-packet basis, require enormous raw compute power and high I/O throughputs. These applications face extreme performance and scalability challenges.

II. Our Solution: Network Traffic Monitoring and Analysis with GPUs

To date, two major computing platforms have been used for packet-based network traffic monitoring and analysis. The first one is dedicated hardware based on NPU and/or ASIC technologies. The major drawback of this approach is that NPUs and ASICs have poor programmability and poor scalability, and high development costs. The other platform is the general-purpose CPU that is capable of running many types of applications. With recent advances in multicore technologies, a single CPU can provide multiple cores to process network traffic in parallel. However, we argue that CPU is not the appropriate computing platform for network traffic monitoring and analysis applications in future high-speed networks (100GE+) because the CPU architecture cannot adequately handle their high packet rates.

Recently, GPUs have been widely applied to accelerate general purpose scientific and engineering computing. The massive array of GPU cores offers an order of magnitude higher raw computation power than CPUs. GPU's data-parallel execution model and ample memory bandwidth can effectively hide memory access latency and effectively boost I/O intensive applications with inherent data parallelism. In addition, the CUDA and OpenCL programming frameworks provide GPU with ease of programmability.

We propose to research on *network traffic monitoring and analysis with GPUs*. In this work, we plan to achieve the following research goals:

(a) Investigate, design, and implement a high-performance packet I/O engine to support network traffic monitoring and analysis with GPUs in high-speed networks.

(b) Explore GPU-based network monitoring and analysis algorithms and data structures to enable fast packet processing and analysis. We will design and implement a GPU-accelerated library for network traffic monitoring and analysis. The library consists of various CUDA kernels, which can be combined in various ways to perform intended monitoring and analysis operations.

(c) Integrate the solutions developed in (a) and (b) to provide a unified solution to enable network traffic monitoring and analysis in high-speed networks.

III. A GPU-Accelerated Network Traffic Monitoring and Analysis System

a. Architecture

Our GPU-based network monitoring and analysis application runs in user mode, to take advantage of the friendly GPU programming framework (e.g., CUDA). As shown in Figure 1, it consists of four logical entities: Traffic Capture, Preprocessing, Monitoring & Analysis, and Output Display. (1) Traffic Capture. It captures network traffic and moves them from wire to the CPU domain. Traffic capture aims to capture packets without loss, even at high packet rates. (2) Preprocessing. It processes the captured network traffic and copies the packets from the CPU domain to the GPU domain. (3) Monitoring & Analysis. It performs network monitoring & analysis with GPUs. (4) Output Display. Network monitoring and analysis results are displayed or stored.

A logical entity runs on a worker thread. For each type of logical entity, one or multiple worker threads are spawned. On a multicore system, each worker thread is tied to a specific core to maximize overall performance. When the system is in operation, these worker threads run cooperatively to perform an intended task. For the same batch of

network traffic, these worker threads run in a pipeline mode with the sequence of "Traffic Capture → Preprocessing → Monitoring & Analysis → Output Display." For different batches of network traffic, these worker threads run in parallel to maximize the overall performance.

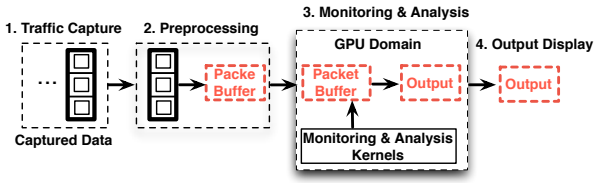

Figure 1 System Architecutre

b. A high-performance packet I/O engine

We must capture and move data packets from the wire to the GPU domain without packet loss, even at high packet rates. Clearly, loss during packet capturing will significantly degrade the accuracy and integrity of the monitoring and analysis results. We propose to use commodity NICs to capture network traffic into the CPU domain and then copy the packets into the GPU domain. We will use commodity 40GE NICs that are available in the marketplace, and evaluate 100GE NICs when prototypes become available. To date, various packet I/O engines have been proposed to boost the performance of commodity NICs in packet capture. NETMAP [1] and PSIOE [2] are the most recent of such developments. However, neither NETMAP nor PSIOE are designed to avoid packet drops. To address the deficiencies, we propose a new and novel packet capture technology. Our packet I/O engine works as follows. It divides a NIC receive ring into *descriptor segments*. Each descriptor segment consists of multiple receive packet descriptors. The packet I/O engine also pre-allocates a number of empty large *packet buffer chunks*. A packet buffer chunk consists of multiple fixed-size cells, with each cell corresponding to a ring buffer. To capture packets, each descriptor segment in the receive ring will be attached with a pre-allocated packet buffer chunk; each cell in the attached packet buffer chunk is sequentially tied to the corresponding packet descriptor in the descriptor segment. Our packet I/O engine provides operations to allow a user space application to capture packets. The application can access the operations through the *ioctl* interface. Once an attached packet buffer chunk is filled with network packets, it will be "captured" to a user space application. To reduce the capture cost, all packet buffer chunks are mapped into the application's process space. Therefore, an attached packet buffer chunk can be moved to user space via pointer passing; no copying is required. When an attached packet buffer chunk is moved to user space, the corresponding descriptor segment will be attached with a new "free" packet buffer chunk. In the user space, the data in a *captured* packet buffer chunk is finally processed. Subsequently, the chunk will be recycled for future use.

c. GPU-based network monitoring and analysis algorithms and data structures

We must design and implement high-performance GPU-based network monitoring and analysis algorithms. We will implement a GPU-accelerated library for network traffic capturing, monitoring, and analysis. The library will consist of various CUDA kernels, which can be combined in various ways to perform monitoring and analysis tasks.

We must design and implement network monitoring and analysis data structures, which can be efficiently handled and manipulated by both CPU and GPU. Captured packets will be processed in the GPU domain to compute various statistics on network traffic and status. However, these statistics need to be transported back into the CPU domain for storage and display purposes. Well-designed data structures that are suitable for both GPU and CPU will accelerate overall performance.

IV. Preliminary Results

We have run studies on our proposed work. Preliminary results were presented at GTC'13 [3]. Our work has recently been reported on in *Network World* [4], *PC World* [5], and other computing publications.

Reference:

[1] Luigi Rizzo, Netmap: a novel framework for fast packet I/O, In Proc. USENIX ATC'12.

[2] S. Han et al. PacketShader: a GPU-accelerated software router, In Proc. ACM SIGCOMM'10.

[3] http://on-demand.gputechconf.com/gtc/2013/presentations/S3146-Network-Traffic-Monitoring-Analysis-GPUs.pdf

[4] http://www.networkworld.com/news/2013/112113-sc13-gpus-would-make-terrific-276246.html

[5] http://www.pcworld.com/article/2066120/sc13-gpus-would-make-terrific-network-monitors.html

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

Project Title: Network Traffic Monitoring and Analysis with GPUs

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

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

Dr. Wenji Wu – Project PI:

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

Phil DeMar – Project Co-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