Challenges with Petabyte-Scale Flows and Beyond

Violet R. Syrotiuk† and Ibrahim Matta‡

†School of Computing, Informatics, and Decision Systems Engineering, Arizona State University
‡Computer Science Department, Boston University
†syrotiuk@asu.edu ‡matta@bu.edu

Instruments of science and engineering are measuring phenomena at rates that result in data sets that challenge storage systems for both short-term access and long-term archival, the development of software systems for their analysis, and the design of protocols to share and access them across the network. In this white paper, we focus on network protocols, including in-network computing and storage and software-defined networking, for petabyte-scale sized data and beyond inspired by [1].

At present, our efforts focus on modeling with the aim of informing the requirements of the infrastructure, and also designing our network protocols before moving to implementation. We consider two scenarios:

1. Moving a huge data set to a cloud or other end system: Here, we model a source transmitting a petabyte-sized data set to a destination through an intermediate router.

2. Transporting an application program to a cloud or other end system for remote execution on a petabyte-size data set and returning the result of the computation.

Ultimately, we intend to implement our protocols for FABRIC because it is an everywhere programmable nationwide instrument made up of extensible network elements equipped with large amounts of compute and storage, interconnected by high speed, dedicated optical links [2]. Furthermore, FABRIC connects a number of testbeds (including clouds) and high-performance computing facilities appropriate for our experimental activities.

The first scenario involves moving a huge data set to a cloud or other end system. This problem may arise when the data set is collected by the scientific instrument for post-processing. Our model considers:

- The use of a window-based versus rate-based transport: In window-based transport, both the window and the segment size may be varied. For example, IPv6 permits jumbograms [3], i.e., payloads of up to one byte less than 4 GiB. In rate-based transport, the rate of the network is matched, using network feedback [4].

- The memory requirements on the end system: This includes both main memory and secondary storage requirements. Some work considers the use of parallel file systems; bypassing the file system stack is also considered [5].

- The computing requirements of routers: If a router is modeled as a store-and-forward device, then any intermediate processing must consider the complexity of algorithms run at the router. These may include algorithms to compress the data, error-checking, among other in-networking processing. It is also important to model interface buffering, and associated queuing, transmission, and propagation delays.

- The use of CRCs versus checksums, and including the checksum in the packet header versus the packet trailer: To select an appropriate CRC requires knowledge of the distribution of the data set and also the error characteristics of the underlying medium. Compressing data is known to improve the performance of checksums [6]; the practice of placing checksums in the header is to be avoided in favor of appending them as a trailer [6].

- The use of forward-error correction (FEC) schemes, such as fountain coding [7], versus backward-error correction schemes must be investigated.
• The effect of compression schemes versus hashing schemes, or other in-networking processing that may be performed on the data to preserve its completeness and integrity.

• The use of multi-level error control, such as that performed at the link versus the transport layer [8].

Once the data set is located where desired, the second scenario considers transporting an application program for remote execution on the data set and returning the result of the computation. Gaining an understanding of the size of the results is an important first step.

Our initial work proposes to use data sets gathered by Ed Llewellyn, a Professor of Volcanology in the Department of Earth Sciences at the University of Durham in the United Kingdom.

References


[2] FABRIC Core Team, Fabric, an adaptive programmable research infrastructure for computer science and science applications. URL https://www.whatisfabric.net/


