Challenges in Designing a Network Service Model for Efficient Data Sharing

Lan Wang
lanwang@memphis.edu
Department of Computer Science, University of Memphis

As the capacity of computing and experiment equipment grows at an unprecedented rate, the large-scale experiments, simulations, and observations conducted by DOE scientists will soon produce exabytes of data easily [1]. Given that moving a petabyte of data takes 21 hours even at the rate of 100 Gb/s, it will be **extremely difficult for the scientists to access and share the huge amount of data** without profound changes to the DOE networks. In particular, scientific users would like to quickly visualize and analyze the data, ideally as the data is produced. However, the data is often produced and stored at a few powerful facilities that may be far away [4]. Moreover, they would like to retrieve only the portion of the data they need, but there is no easy way to identify the relevant data in big data files, so they typically have to transfer the entire data from an experiment, which leads to much longer delay and unnecessary bandwidth consumption. Similarly, when a site wants to replicate data for its users, it often needs to replicate the big data files instead of only the portion the users need.

The above inefficiencies stem from the following *fundamental issue* in today's practice: ***shipping large files using a host-based point-to-point communication model***. First, applications have *no easy way to identify the needed data at a granularity most efficient for the network to transfer*. Second, the network and transport layers are designed to *ship opaque data bytes from one address/process to another address/process*, rather than getting the data from anywhere that replicates the data. Finally, this model is inherently *incompatible with today's multiparty scientific collaborations*. A 2008 DOE workshop [2] pointed out that "**underlying network service models will have to extend beyond existing 'point-to-point' models, and support broader application-specific topologies that can interconnect multiple user sites**". A 2014 DOE workshop [3] also identified the need for research and developments to encompass multi-site, multi-domain connections.

The apparent incongruity leads us to pose this *what-if question*: **what would the research challenges be if the network service model changes from "delivering packets to given destination addresses" to "getting the desired data"**? Envision a network where users (applications) can specify exactly what data they need and the network will find the data from any repository, cache, or device that has a copy of the data. This can make data distribution and sharing among computing/experiment facilities and user sites much more efficient. In fact, a limited version of this is provided by today's web and content distribution networks (CDNs). The question is whether this service model can be implemented at the network layer to support all applications and users (including extreme-scale scientific applications). Below are a few **research challenges and opportunities**:

(a) How to *scale routing* so that the network can find the right data without maintaining an entry for each piece of data? This is related to how the data is named and how adaptive the forwarding plane is.

(b) How to *control congestion*? There is no longer a concept of flows (between two end points) in this model. Conventional TCP congestion control schemes that are based on flows may no longer apply. Each piece of data that a user needs is independently retrieved from anywhere that the copy of data exists. Along with the challenges, there are also new opportunities. For example, the network may play a more active role in relieving congestion as it has more information about data – it can cache more popular data and slow down data requests.

(c) How to *reserve resources* for data transfer? In the current DOE networks, users are able to reserve bandwidth on-demand for their data transfer using a separate signaling protocol (MPLS-TE). In the new model, the focus is on data, so the goal is to give the requested data more resources and higher priority. It would be ideal if such information can be encoded in data requests from users so that no special signaling protocol is necessary and all necessary resources (bandwidth, computation and storage) are reserved with the same data request.

References

[1] DOE Office of Science. Data crosscutting requirements review. `http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/ASCR_DataCrosscutting2_8_28_13.pdf`, March 2013.

[2] Donald Petravick *et al.* Workshop report on advanced networking for distributed petascale science: R&D challenges and opportunities. `http://www.orau.gov/ioninfrastructure2014/intelligent-network-infrastructure-workshop-report.pdf`, April 2008.

[3] Tom Lehman, Inder Monga, Nagi Rao, and John Wu. Software defined networking for extreme-scale science: Data, compute, and instrument facilities – report of DOE ASCR intelligent network infrastructure workshop. `http://www.orau.gov/ioninfrastructure2014/intelligent-network-infrastructure-workshop-report.pdf`, August 2014.

[4] Peter E. Nugent and J. Michael Simonson. Data and communications in basic energy sciences: Creating a pathway for scientific discovery – report of a workshop linking experimental user facility needs with advances in data analysis and communications. `http://science.energy.gov/~/media/ascr/pdf/research/scidac/ASCR_BES_Data_Report.pdf`, June 2012.