Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Bridging the gap between peak and average loads on science networks

Sam Nickolay[a], Eun-Sung Jung[b,*], Rajkumar Kettimuthu[c], Ian Foster[a,c]

[a] Department of Computer Science, University of Chicago, Chicago, IL, USA
[b] Hongik University, South Korea
[c] Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA

HIGHLIGHTS

- Researching reducing the gap between peak and average loads in research networks.
- Categorized transfers into on-demand and best-effort, giving priority to on-demand.
- Simulated high data transfer loads and studied performance of the data transfers.
- Current network capacity can handle up to 2× the current load.
- Impact on on-demand transfers is negligible, and best-effort transfers is minimal.

ARTICLE INFO

Article history:
Received 21 September 2016
Received in revised form
7 April 2017
Accepted 10 May 2017
Available online xxxx

Keywords:
Network utilization
Data transfer scheduling
Network planning

ABSTRACT

Backbone networks are typically overprovisioned in order to support peak loads. Research and education networks (RENs), for example, are often designed to operate at 20–30% of capacity. Thus, Internet2 upgrades its backbone interconnects when the weekly 95th-percentile load is reliably above 30% of link capacity, and analysis of ESnet traffic between major laboratories shows a substantial gap between peak and average utilization. As science data volumes increase exponentially, it is unclear whether this overprovisioning trend can continue into the future. Even if overprovisioning is possible, it may not be the most cost-effective (and desirable) approach going forward. Under the current mode of free access to RENs, traffic at peak load may include both flows that need to be transferred in near-real time – for example, for computation and instrument monitoring and steering – and flows that are less time-critical, for example, archival and storage replication operations. Thus, peak load does not necessarily indicate the capacity that is absolutely required at that moment. We thus examine how data transfers are impacted when the average network load is increased while the network capacity is kept at the current levels. We also classify data transfers into on-demand (time-critical) and best-effort (less time-critical) and study the impact on both classes for different proportions of both the number of on-demand transfers and amount of bandwidth allocated for on-demand transfers. For our study, we use real transfer logs from production GridFTP servers to do simulation-based experiments as well as real experiments on a testbed. We find that when the transfer load is doubled and the network capacity is fixed at the current level, the gap between peak and average throughput decreases by an average of 18% in the simulation experiments and 16% in the testbed experiments, and the average slowdown experienced by the data transfers is under 1.5×. Furthermore, when transfers are classified as on-demand or best-effort, on-demand transfers experience almost no slowdown and the mean slowdown experienced by best-effort transfers is under 2× in the simulation experiments and under 1.2× in the testbed experiments.

1. Introduction

Scientific applications in various domains such as high-energy physics, cosmology, genomics, etc., generate large datasets that need to be transported over the network for a variety of reasons. In order to support these applications, federal agencies in different countries fund organizations to build and operate high-speed

* Corresponding author.
E-mail addresses: samnickolay@uchicago.edu (S. Nickolay), ejung@hongik.ac.kr (E.-S. Jung), kettimut@anl.gov (R. Kettimuthu), foster@anl.gov (I. Foster).

http://dx.doi.org/10.1016/j.future.2017.05.012
0167-739X/

Please cite this article in press as: S. Nickolay, et al., Bridging the gap between peak and average loads on science networks, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.05.012

research and education networks. These networks are typically overprovisioned so that all science users continue to have adequate network resources even at times of peak load. Thus, they are underutilized most of the time. Research and education networks (RENs), such as Internet2, have a policy of operating their networks at light loads (25%–30%) to allow the networks to absorb surges in traffic [1]. Other RENs such as ESnet and GEANT are also engineered to operate at similar average load levels.

Several recent studies project that, given predictions of science traffic's exponential growth, it may not be feasible to continue such overprovisioning [2–4]. Recent reports on science network requirements note that different transfers have different needs [5–7]. Some transfers, such as those in which remote analysis result of one experiment is used to guide the next, are time-critical and have tight constraints. In contrast, some transfers, such as certain data replication, backup, and archiving use cases, have more flexibility and may only need to be completed within a window several times longer than the transfer time under average load. Under the current mode of free access to RENs, the traffic at peak load may include a combination of different types of transfers including some that are less time-critical. We argue that measures to spread the load and keep the peak under control are important, and use simulations based on real traffic traces to quantify the benefits that may be gained from such measures.

The two main contributions of this study are as follows:

- We characterize how data transfers are impacted when the gap between peak and average throughput is reduced by increasing the average network load while keeping network capacity at the current levels.
- We determine how these impacts change when data transfers are classified as either on-demand (ones that are time-critical) or best-effort (ones that are less time-critical), with on-demand transfers getting a relatively larger share of the bandwidth.

These results suggest that the utilization of research networks can be increased significantly with modest or no impact on the performance of deadline-sensitive data transfers, simply by identifying deadline-sensitive transfers as such.

We use transfer logs from production GridFTP servers for our study. Our results indicate that when load is doubled with network capacity fixed at the current level, the gap between peak and average load decreases by an average of 18% and the average slowdown experienced by data transfers is still less than 1.5×. For the logs in which the peak load is 5× or more than the average load, average slowdown experienced by the data transfers is under 1.1×. Under the same scenario of doubled network load with the same network capacity, when the transfers are classified into on-demand and best-effort, on-demand transfers experience almost no slowdown and the slowdown experienced by best-effort transfers is under 2×, even when 50% of transfers were treated as on-demand and on-demand transfers are given a reasonably higher share (70%) of the bandwidth. For the logs in which the peak load is 5× or more than the average load, the average slowdown experienced by best-effort data transfers is under 1.2×.

The rest of the paper is organized as follows. Section 2 introduces the two characteristics of contemporary networks and science workloads that motivate our research. Section 3 presents the problem our research addresses and the metrics we use to evaluate our results. Section 4 defines the algorithm we developed for our research. Section 5 evaluates the results from our experiments. Section 6 describes the testbed experiments and their results. Section 7 discusses how this research can be generalized. Section 8 examines related work. Section 9 discusses the conclusions from our research.

2. Background and motivation

Our research is motivated by the following two characteristics of contemporary networks and science workloads.

2.1. Big gap between peak and average network load

Fig. 1, which shows the wide-area network traffic over a one-month period for two HPC facilities, illustrates the substantial gap between average and peak loads that is common on RENs.

We also obtained logs from the anonymized usage statistics that Globus GridFTP servers send to a usage collector. These logs include transfer size, start time, and end time. We collected the logs of the four servers that transferred the most bytes in a one month period. For each of those servers, we then picked the log for the day in that month in which it transferred the most bytes. Fig. 2 shows the aggregate throughput over that 24-h period for each server. Once again, we see substantial difference between mean and peak throughputs.

2.2. Some transfers can tolerate delays

While certain transfer requests are time-sensitive, others can tolerate delays. For example, science communities replicate large quantities for performance [8,9], fault tolerance [10], and/or preservation [11]. Such replication activities are often cited as a common relatively time-insensitive data transfer use case [5–7], as when a multi-TB dataset needs to be copied to a remote site overnight. Because subsequent processing involves manual steps, there is no advantage in completing the transfer earlier. Replicating 100 TB within a month is another use case cited. For these use cases, transfer times can vary by at least an order of magnitude without compromising science goals.

3. Problem statement and goals

Current RENs experience highly variable load patterns and offer no differentiation in the service provided to different applications. These networks are overprovisioned (see Definition 1 in Section 3.2) to absorb traffic surges and to ensure that no flow is delayed due to aggregate load exceeding capacity. Consequently, the average usage is limited to X%, where X is usually small (as low as 10% in some cases). Thus the gap between peak and average aggregate throughput on these networks is often high.

If the service provided by a REN could be differentiated so that traffic requiring instant service (on-demand) was treated differently from traffic that can be delayed by a certain amount (best-effort), then we could reduce the gap between peak and average throughput (e.g., by increasing the average usage to Y%) on these networks without any flows being delayed beyond their service level. The question then is under what circumstances is this true and how do X and Y relate to the service levels chosen for the best-effort and on-demand traffic classes. This is the problem we study in this work.

3.1. Background of the problem

Our target networks are wide-area networks composed of many components such as switches and data transfer nodes (DTNs). The networks may be either circuit-switched networks, in which paths are reserved ahead of data transfers, or packet-switched networks, in which packets from multiple data transfers can share network links. In reality, most networks used by data transfer users such as scientists are packet-switched networks, except for special networks such as ESnet [12] and Internet2 [13] which support circuits. However, we assume that both kinds of networks offer ways (e.g., tc [14] Linux) to perform traffic engineering on each flow.

Please cite this article in press as: S. Nickolay, et al., Bridging the gap between peak and average loads on science networks, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.05.012