International Networking in support of Extremely Large Astronomical Data-centric Operations

Jeronimo Bezerra,[1] Vinicius Arcanjo,[2] Julio Ibarra,[1] Jeff Kantor,[3] Ron Lambert,[3] Matt Kollross,[4] Albert Astudillo,[5] Shahram Sobhani,[3] Sandra Jaque,[5] Donald Petravick,[4] Heidi Morgan[6] and Luiz Lopez[7]
[1]*Florida International University, Miami, FL, USA*

[2]*Rede Nacional de Ensino e Pesquisa, Campinas, SP, Brazil*

[3]*LSST Project Office, Tucson, AZ, USA*

[4]*National Center for Supercomputing Applications, Champaign, IL, USA*

[5]*Red Universitaria Nacional, Santiago, Chile*

[6]*University of Southern California, Los Angeles, CA, USA*

[7]*Academic Network of Sao Paulo, Sao Paulo, SP, Brazil*

Abstract. New international academic collaborations are generating daily data sets in the order of terabytes in size. Often these data sets need to be moved in real-time to a central location to be processed, archived and then shared. In the field of astronomy, data sets are usually created in remote locations, often away from the processing facilities. To send data sets to processing facilities, high bandwidth network infrastructures need to be provided to transport these data sets through very long distances, relying on multiple network operators. Creating an end-to-end path involving multiple network operators, technologies and interconnections often adds conditions that make the real-time movement of big data sets challenging. The Large Synoptic Survey Telescope (LSST) is an example of an astronomical application imposing new challenges on multi-domain network provisioning activities. This paper introduces the ongoing effort to guarantee an efficient network configuration capable of handling LSST data transfers across multiple network operators from South America to North America.

1. Introduction

Each night, terabytes of raw data are generated at the telescopes in Chile. These data sets must be transported to remote archives for processing and distribution to the astronomers waiting to pursue their scientific analyses. For some projects, scientific work flows require rapid processing for time-sensitive results. These projects drive even higher bandwidth needs to transport and manage the data in near real time. Leveraging the Research and Academic Networks (REN) is the most efficient and effective approach to achieve a reliable and scalable configuration. However, building international end-to-end paths across multiple RENs to support these new astronomical applications is a complex activity, requiring intensive coordination.

To illustrate such complexity, the Large Synoptic Survey Telescope (LSST) (Ivezic et al. 2008) network will be used as a use case. LSST is transforming how networks and cyberinfrastructure are used across geographically distributed sites in the U.S. and Chile. During operation, the LSST will have to transfer its 12.8 GB images generated approximately every 30 seconds from Chile to the U.S. in less than 5 seconds each time the camera shutter closes. Feature catalogs and telescope commands and controls also need to be transported in parallel to the LSST images.

This paper addresses some of the challenges to achieve such a reliable international end-to-end path, with multiple network operators, high-delay links and complex network requirements and bandwidth guarantees.

2. Network Challenges to Support Astronomical Data-centric Operations

This section has a non-extensive list of challenges network operators must manage when deploying data-centric international end-to-end paths that cross multiple network domains.

2.1. Inter-domain Network Provisioning and Coordination

To achieve effective high-bandwidth connections across multiple network domains, network operators must coordinate and understand the differences in their networks. Network operators in a commercial environment exercise control and privacy on the use of their resources because they operate in a competitive environment. On the other hand, Research and Education Networks (REN) are accustomed to collaborating and sharing information. Nevertheless, a multi-domain network scenario for research and education network operators remains a challenge, because decisions on the use of network resources are constrained not only on technology and protocol differences, but also on administrative policies in place between the network operators.

Effective provisioning of network services in a multi-domain network environment refers to the ability of multiple network operators successfully exchanging information along the following classifications: routing, Quality of Service (QoS) provisioning, path protection and restoration, and inter-domain control plane interactions. For LSST, two of these classifications are salient: QoS and path protection and restoration. To achieve end-to-end QoS provisioning, neighboring domains must agree to exchange information on multiple levels of the network architecture which is non-trivial, because of the absence of full network topology information across domains.

2.2. Network Performance and Measurement

Most astronomy projects in Chile face the same challenge during the data transfer phase: the end-to-end path has a high propagation delay, which affects traditional transport protocols performance. Currently, round trip times (RTT) from Chile to Illinois, in the U.S., vary from 144 to 280ms[1].

High RTTs affects the TCP congestion control mechanisms as described in (Bhandarkar et al. 2005). In addition to RTT, TCP performance is also determined by packet loss. With packet loss, the Mathis equation (Mathis et al. 1997) can be used to estimate

[1]LSST Performance Dashboard - http://ps-dashboard.ampath.net

the theoretical maximum achievable throughput. Using Mathis equation and the RTTs observed (144 and 280ms), with a 0.001% packet loss, the total time required to move each of LSST data set would be from 5 to 10 seconds. Isolating a 0.001% packet loss across multiple network operators is challenging and time-consuming. For this reason, being able to measure the network performance is crucial and perfSONAR (Hanemann et al. 2005) is the *de facto* tool for such measurement.

In network environments created to support high-bandwidth data transfers, each domain must deploy perfSONAR nodes to measure the performance of each link, or set of links, and results might uploaded to a measurement archive. The measurement archive might used to define a performance *baseline*. This baseline will be used in future troubleshooting activities. Once network operators complete their local performance measurement installations, inter-domain tests must be established for each inter-domain connection and also for end-to-end, as close as possible from both source and destination nodes or networks.

2.3. Quality of Service

Traffic bursts can potentially happen at any time and could lead to congestion and packet loss. In a complex scenario such as LSST, controls need to be in place to handle bursts; otherwise, the LSST Data Management application will miss its 5 seconds data transfer window. Therefore, Quality of Services (QoS) techniques play a crucial role in making sure that, in the case of congestion, high profile network traffic types are prioritized.

Five traffic types were defined by the LSST Data Management team and the highest priority traffic type requires a minimum bandwidth of 35Gbps. QoS policies have to be deployed in all networking devices in the path. As far as enforcing and deploying QoS policies, there are several techniques and components involved in the final solution to achieve a consistent end-to-end result. Some underlay technologies and techniques are coupled with vendor implementation and lack flexibility, and new approaches, such as Software-defined Networking, could take advantage of certain QoS abstractions and policy frameworks to try to achieve a more ubiquitous QoS policy enforcement. Due to these differences, each network operator's underlay transport technologies have to be evaluated to guarantee the end-to-end performance.

Because each network operator might have a different underlay technology, and as QoS is technology-dependent, there is an overhead to be handled by the network operators supporting the LSST network. Topics to be coordinated are: how to classify traffic and hand it off across network operators; which scheduling and queue management approached to use; and how to evaluate if the requirements provided are being respected. Once these decisions are made, each network will need to validate the results. perfSONAR is a possible approach to instrument and validate QoS policies.

2.4. Complex Reliability

In addition to high bandwidth, traffic prioritization, and an efficient performance measurement system, network resilience is another key component of a data-centric network. In such network infrastructure, single points of failure need to be eliminated. In multi-domain networks, single points of failure or network failures may happen inside any of the network operators (intra-domain) and between network operators (inter-domain).

Intra-domain events are usually hidden from the outside users and handled directly by Network Operation Teams (NOC). For most protocols, network recovery can happen under a few dozens of milliseconds, being imperceptible to users and applications. Inter-domain recovery may take dozens of seconds to occur when backup links are present.

Fast recovery and maintaining the bandwidth guarantees is key for real-time data-centric applications such as LSST. The LSST's bandwidth requirements request for highest priority traffic type 35Gbps independently of network failures. This condition implies that all network operators must consider multiple paths with, at least, 40Gbps. Also, this condition implies that multiple connections must be established between network operators. With each network operator deploying its own technologies and protocols for intra-domain transport, for inter-domain, the solution has to be carefully planned by all operators.

3. Conclusion

Leveraging existing research and education networks to create an end-to-end path is possible and encouraged. However, inter-domain connectivity needs to be properly addressed in advance, as some requirements might require a complete network reconfiguration, for example, to support bandwidth reservation. The real challenge is to combine a set of protocols and tools that could be integrated to achieve a reliable network infrastructure to support data-centric network applications.

Acknowledgments. This material is based upon work supported in part by the National Science Foundation through Cooperative Agreement 1258333 managed by the Association of Universities for Research in Astronomy (AURA), and the Department of Energy under Contract No. DE-AC02-76SF00515 with the SLAC National Accelerator Laboratory. Additional LSST funding comes form private donations, grants to universities, and in-kind support from LSSTC Institutional Members.

References

Bhandarkar, S., Jain, S., & Reddy, A. L. N. 2005, in Proceedings of the Third PFLDNET Workshop. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.65.2194`

Hanemann, A., et al. 2005, in Service-Oriented Computing - ICSOC 2005: Third International Conference, Amsterdam, The Netherlands, December 12-15, 2005. Proceedings (Springer Berlin Heidelberg), 241

Ivezic, Z., et al. 2008, ArXiv e-prints. `arXiv:0805.2366`

Mathis, M., Semke, J., Mahdavi, J., & Ott, T. 1997, ACM SIGCOMM Computer Communication Review, 27, 67