Contribution ID: **245**   Type: **Oral presentation to parallel session**

Application Performance Evaluation and Recommendations for the DYNES

Tuesday 15 October 2013 17:45 (20 minutes)

Computing and networking infrastructures across the world continue to grow to meet the increasing needs of data intensive science, notably those of the LHC and other large high energy physics collaborations. The LHC's large data volumes challenge the technology used to interconnect widely-separated sites (and their available resources) and lead to complications in the overall process of end-to-end data distribution, analysis and management. A delicate balance is required to serve both long-lived, high capacity network flows, as well as more traditional end-user activities using general purpose infrastructure.

R&E networks have experimented with Virtual Circuits (VC) for a number of years as a mechanism that affords greater control over network capacity and traffic management [Oscars, ION, SDN, Autobahn]. This connection-oriented concept emulates a physical point-to-point connection, using the underlying technology of common packet-switched networks. In contrast to a physical circuit, VC technology allows for variable duration, guaranteed bandwidth channels, and fosters efficient use of common network infrastructures.

The DYNES instrument, an NSF funded cyberinfrastructure project designed to facilitate end-to-end dynamic circuit services, is built using this VC technology, as well as other common open source software packages for network monitoring and data movement [DYNES, FDT, perfSONAR-PS]. Dynamic circuits have been used in production for the last 6 years among a limited number of major laboratory and university sites. DYNES is extending this capability to many campuses with the goal of increasing the number of sites able to easily participate as end-points of virtual circuits.

A key observation during installation and testing of DYNES was related to the performance of standard data movement tools over virtual circuits: the observed performance did not match expectations related to the bandwidth reserved in circuits. In many cases the data movement reality was an order of magnitude lower than the initial bandwidth request; investigation as to a possible cause centered on the QoS mechanisms of the underlying network. In most cases bandwidth reservations are significantly below the "wire-speed" of the host's network interface card. This was deemed a likely source of at least part of the low performance typically observed.

Our study focused on factors commonly responsible for degrading network performance: buffer overruns due to lack of available memory in relation to application burst behavior and queuing on network devices that introduce out of order behavior in TCP streams. After experimenting with these behaviors we explored various techniques, some of which do not require modification of legacy applications, which can be used to mitigate these concerns at the end hosts.
We will present the results of our testing and list the benefits and shortcomings of the various options we explored. We will discuss our experiences with kernel network stack tuning, application pacing, tc, RoCE and TCP variants. When implemented correctly these mechanisms will improve (sometimes significantly) the end-to-end flow of traffic across VC resources.

Co-authors: BROWN, Aaron; Prof. TACKETT, Alan (VANDERBILT UNIVERSITY); MELO, Andrew Malone (Vanderbilt University (US)); BARCZYK, Artur Jerzy (California Institute of Technology (US)); MUGHAL,

Azher (California Institute of Technology (US)); Mr MEEKHOF, Ben (University of Michigan); FINKELSON, Dale (Internet2); BOYD, Eric (Internet2); NEWMAN, Harvey (California Institute of Technology (US)); ZURAWSKI, Jason (Internet2); BINKLEY, Mathew (V); SHELDON, Paul (Vanderbilt University (US)); VOICU, Ramiro (California Institute of Technology (US)); BALL, Robert (University of Michigan (US)); BROWN, Robert (Vanderbilt University); ROZSA, Sandor (California Institute of Technology (US)); WOLFF, Stephen (Internet2)

Session Classification: Facilities, Infrastructures, Networking and Collaborative Tools

Track Classification: Facilities, Production Infrastructures, Networking and Collaborative Tools