HPC Graphics and Visualization, not Games, Drive Advances in Networked Graphics Technology

John Shalf and Wes Bethel
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

1.0 Introduction

A mantra we've heard over the years is that the consumer gaming market drives graphics technology. Our thesis is that the same does not hold true for networking technology. Instead, it is applications from high performance graphics and visualization efforts, not games, that push networks past their breaking point, which in turn stimulates work that results in better networks. This case study shows how a high performance computational science and visualization application has set new levels of networking performance by winning the SC2001 Application Bandwidth Challenge [1] by novel use of existing network protocols.

2.0 The Scientific Problem

One of the most challenging problems in science is the computational simulation of Einstein's General Theory of Relativity. These equations are among the most complex in the world of physics, containing millions of terms if fully expanded. The General Relativity Group at the Albert Einstein Institute has developed a code (the Cactus Code [2]) for solving these equations on supercomputers in order to simulate the most extreme of astrophysical phenomena, such as the collision of two black holes and the gravitational waves that radiate from that event. The Cactus simulation codes runs on some of the largest supercomputers in the world, including NERSC's SP2, the largest unclassified supercomputer in the world. The simulations are so large that it is impossible to use traditional visualization tools to see and understand the results of these simulations. Typically, computation is performed at NERSC, with the results visualized and made available to a remote viewer.

3.0 Seeing the Science from Remote Locations

The NERSC/LBNL Visualization group has developed the Visapult [3] tool to attack these sorts of "Grand Challenge" problems. Visapult is a distributed, parallel volume rendering application that allows us to use computers and high performance networking resources that are on the same order of scale as the supercomputers that these massive simulation codes consume. The Visapult code was first demonstrated at Supercomputing 2000, where it won that year's Bandwidth Challenge award by rendering data stored on a distributed parallel network filesystem. At SC2000, Visapult obtained peak bandwidth of 1.5Gbps, and a sustained rate of about 660Mbps. These rates represent about 60% and 25% of the theoretical line rate of the OC-48 link used during the contest.

For SC2001, Visapult was modified to connect directly to a running simulation code in order to interactively visualize the simulation results during a live run. Visapult sustained 3.3 Gigabits/sec. over the WAN using OC-48 and OC-12 links over ESnet from Berkeley to SC2001, as well as an OC-12 link between NCSA to SC2001. This represents an 88% sustained utilization of theoretical line rate - a dramatic improvement over our earlier results.

4.0 The Dirty Tricks

In order to achieve these very high line utilization rates, we were required to take a fundamentally different approach in our use of the network. In SC2000, all WAN communication occurred using the connection-oriented TCP protocol over "striped" connections. TCP guarantees reliable delivery of data across networks, but at a heavy cost in terms of network performance. The machinations and tuning necessary to get TCP to perform well on a wide area network far exceed the initial benefits that make it attractive as a lossless transport mechanism. Indeed, we encountered such egregious difficulties with TCP performance, that we dropped it entirely to use unreliable protocols that focus on interactivity.

For SC2001, we switched to a connectionless UDP-based transport mechanism. This simple change resulted in dramatic improvements in line utilization. Since UDP is a "lossy" protocol, packets aren't guaranteed to be delivered. However, by understanding those circumstances that cause packet loss, they can be effectively avoided. In addition, use of UDP required changes to Visapult in order to encode and decode subsets of scientific data into small network packets. After a brief period of performance tuning to match packet load with line rate, we experienced virtually no packet loss with UDP.

We submit that the trade-offs involved in using an unreliable protocol are absolutely necessary to achieve adequate performance on the WAN, and that the gains in interactivity and high fidelity are far more important than the occasional visual artifact that results from such an approach.

5.0 Acknowledgement and More Information

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Dept. of Energy under Contract No. DE-AC03-76SF00098. The SC2001 Application Bandwidth Challenge required the successful coordination of many people. We gratefully acknowledge Ed Siedel and his team (AGI), Eli Dart (NERSC), John Christman (LBL), George "Chip" Smith (NERSC), and Nick Cardo (NERSC).

6.0 References

[1] The SC Annual Conference Series, at http://www.supercomp.org/.
[2] Cactus Code, at http://www.cactuscode.org/.
[3] Visapult, at http://vis.lbl.gov/.