Developing network-aware operating systems

The Net100 project is a collaboration among Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, National Center for Atmospheric Research, and the Pittsburgh Supercomputing Center (PSC). The Department of Energy's Office of Science funds the three-year project.

Objectives and Impact:

Despite increasing network speeds, many distributed applications manage to utilize only a small percentage of the available bandwidth. Often a "network wizard" can tune the application and/or operating system to dramatically improve network performance. The objective of the Net100 research project is to improve the network performance of scientific applications without wizard intervention by dynamically tuning the network components of the application and operating system. We have already seen improvements in bulk transfer applications, and we expect improvements in grid-based applications as well.

Project Components:

The project is based on three major components:

1. An instrumented and tunable network transport layer in the operating system. This component is based on the Web100 Linux kernel funded by the National Science Foundation (NSF). The Web100 kernel provides passive monitoring of the TCP transport protocol for each active flow. Net100 extensions to the Web100 kernel include additional TCP transport options and mechanisms for tuning TCP for each individual flow.
2. Development and analysis of network transport protocols using simulation, emulation, and deployment on high-speed wide-area networks (e.g., ESnet and Internet2). For TCP, the objectives are to eliminate memory buffer bottlenecks, avoid packet loss, and, in the event of packet loss, speed recovery. The Net100 protocol optimizations, as well as other TCP optimizations provided by the greater network research community, are implemented in the Web100 kernel and can be selectively enabled and tuned on a per-flow basis.
3. A Network Test and Analysis Facility (NTAF) provides a collection of network probes and sensors located across the Internet. These sensors measure network characteristics (bandwidth, jitter, congestion) of various network links. The data are archived for historical analysis and provided to network tuning daemons on the Web100 hosts for tuning the transport layer.

Novel Approaches:

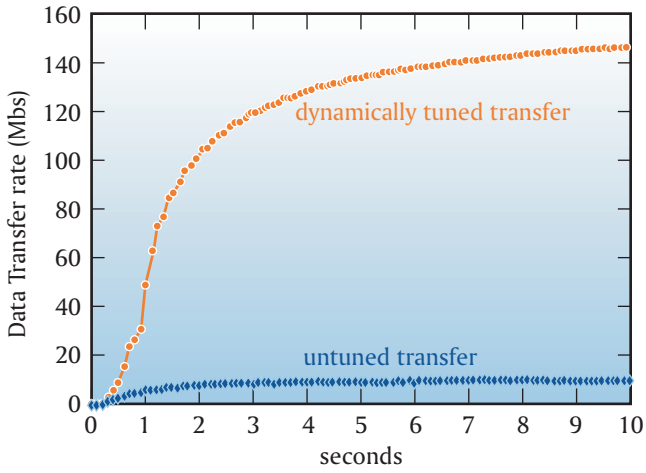
The project employs several novel approaches:

- The NTAF sensors and probes use the Web100 kernel to gather information on the behavior of TCP during the probe test. Besides just bits-per-second, the NTAF collects and reports on packet losses, duplication, round-trip times, and congestion events.
- A tuning daemon, or work-around daemon (WAD), can tune even legacy (network-challenged) applications, and no changes need to be made to the application.
- Tuning can be done selectively based on source and destination, so transport tuning can be optimized for designated paths using the metrics provided by the NTAF.

Results (year 1):

The Net100 research team achieved all of its first-year milestones, including these:

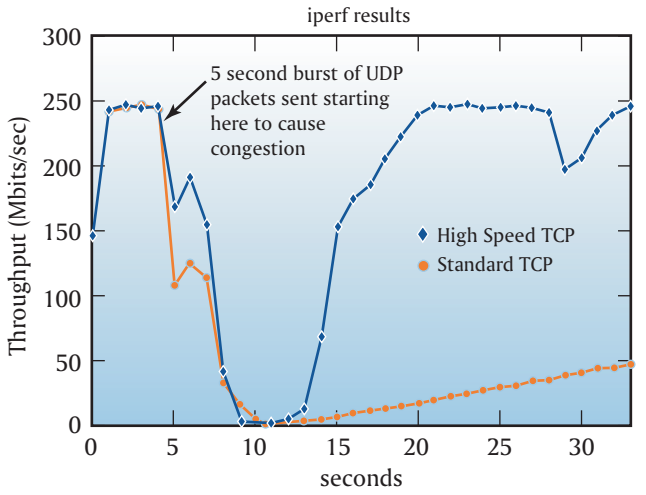
- Developed and deployed network probes and sensors across the Internet.
- Added TCP tuning options to the kernel, including buffer sizes, burst management, AIMD-tuning, virtual MSS, and Sally Floyd's modified slow-start and high-speed TCP AIMD
- Developed and tested event-based tuning daemon (WAD). Used the tuning daemon to tune both single and parallel stream transfers, including gridFTP transfers, between DOE and PSC .



Interactions:

The Net100 team has worked closely with the NSF-funded Web100 development team in getting additional instrumentation in the kernel and in adding TCP tuning optimizations. The Net100 research team has been interacting with other DOE- and NSF-funded projects that are looking at ways to measure path bandwidth and available capacity. In turn, several of the network measurement research projects have experimented with using a Web100 kernel to enhance their measurement tools. We have had good interactions with the various data grid projects and as a result have Net100 nodes in Western Europe. We have also had interactions with the SciDAC Super Nova project, seeing how Net100 might be used to achieve its network requirements.

The Net100 team has given numerous presentations and published a paper on preliminary results of TCP tuning.



Future Work:

In the coming year, we will continue to evaluate protocol optimizations, as well as non-TCP protocols and parallel TCP streams. We will make the tuning daemon more user-friendly so that it can be easily deployed and utilized. We have already had some contacts with computer vendors to pursue adding the Web100/Net100 extensions to operating systems other than Linux. Longer-term research includes exploring the use of the NTAF metrics for selecting alternate or multiple paths for data transport using application-based routing daemons.