BWCTL: Making e-VLBI Happen

A Use Study for E2Epi

www.internet2.edu

What is Iperf? Iperf is a tool to measure TCP bandwidth, delay jitter, datagram loss; more information can be found at: http://dast.nlanr.net/Projects/Iperf/.

What is BWCTL? It is the Bandwidth Control tool, a bandwidth test scheduler developed by Internet2's End-to-End Performance Initiative (E2Epi); more information can be found at: http://e2epi.internet2.edu/bwctl/.

What is VLBI? Very-Long-Baseline Interferometry (VLBI) is a high-resolution imaging technique used in radio astronomy. VLBI techniques involve using multiple radio telescopes simultaneously in an array to record data, which is then stored on magnetic tape and shipped to a central processing site for analysis. Using high-bandwidth networks, electronic transmission of VLBI data (known as "e-VLBI") from this worldwide array of telescopes is now becoming a reality. The Internet2 VLBI Birds of a Feather (BOF) group focuses on the needs of the global VLBI community in the new era of high speed networks.

Do you have to create a tuned path for a real-time point-to-point data transfer? Like many, you may want to use **Iperf** to establish the expected performance on your existing path. Iperf can help determine the achievable (or available) bandwidth between a pair of hosts.

It is often useful to test to multiple points along a network path to determine the network characteristics along that path. Typically, users who want to conduct path decomposition must directly contact the network or system administrators who control the hosts along the path. Iperf tests can consume a great deal of bandwidth and, as a result, many network administrators are wary about to whom, when, and for how long they will allow such tests on their networks. Each administrator must either run half of the test for the user or provide a user account on the host, which is untenable for all but well-known testers.

Most network paths involve multiple administrators. This is made even more complicated when the test locations are in different countries (even different continents, as in the case of radio astronomy) and backbone networks. These hurdles make path decomposition difficult in practice.

BWCTL can help with this problem; it was designed to enable non-specific Iperf tests to a host without having to give full user accounts on the given system. BWCTL is a command line client application and a scheduling and policy daemon that executes the Iperf command line program. System administrators can configure a given host as an Iperf endpoint, which can be shared by multiple users without concern that those users will interfere with each other.

Case In Point

On March 25, 2004, the first successful real-time international transmission and processing of **VLBI** data was conducted between MIT Haystack Observatory's Westford antenna in Massachusetts and the Onsala Space Observatory antenna in Sweden. Data from the Onsala observatory were transmitted over the GEANT network to New York and then over Internet2's Abilene Network to the **Haystack Observatory**. Although the data streamed at relatively low bandwidth (32 Mbps), it was still a milestone demonstration of real-time simultaneous correlation of data from two antennas. Using the BWCTL component of Internet2's E2E piPEs and intermediate BWCTL beacons in Washington and London, researchers at Haystack could run tests between them to assist in diagnosing transcontinental network problems during the experiment.

Dr. David Lapsley is a Research Engineer at MIT Haystack Observatory, in Massachusetts; he has been working to enable radio astronomers practicing VLBI to shift vast amounts of data in short periods of time across research networks spanning the globe, using high-speed networks. This "e-VLBI" will allow astronomers at various locations to tune their radio telescopes in real-time, to catch interesting events as they are identified by other telescope locations.

Lapsley works with VLBI researchers in the United States and around the world. The data collected is transported on a wide variety of networks, including Abilene, GEANT, APAN, SURFnet, and NorduNet, among others.

In trying to optimize the paths between Haystack and Onsala telescope locations, Lapsley discovered that he was not getting the performance he needed to make e-VLBI more efficient than the existing system of downloading data onto tapes and shipping them to the nearest correlator. Lapsley heard

What is Haystack? The Haystack Observatory at MIT is one of the pioneers in the field of VLBI in radio astronomy. VLBI is a high-resolution imaging technique used in radio astronomy. VLBI techniques involve using multiple radio telescopes simultaneously in an array to record data, which is then stored on magnetic tape and disks and shipped to a central processing site for analysis. Using high-bandwidth networks, electronic transmission of VLBI data (known as "e-VLBI") from this worldwide array of telescopes is now becoming a reality.

What is E2Epi? E2Epi is Internet2's End-to-End Performance Initiative. For more information about us, see our website at: http://e2epi.internet2.edu/. There you'll find links to projects and presentations, as well as other information related to end-to-end performance issues, measurement tools, upcoming events, and related activities.

about BWCTL at an Internet2 Member Meeting and joined the early adopters group.

For several of the telescopes to which he is attempting to setup tests, Lapsley does not have root privileges or administrative access needed to run standard Iperf tests; also, many of the machines to which he'd like to test are heavily scheduled. By having the remote locations download BWCTL and install it on a server at their location, Lapsley can schedule tests whenever machines are available.

Using BWCTL, he was able to pinpoint a packet loss problem to local congestion in the Haystack Observatory area; as a result, this bottleneck link was updated in May 2004. BWCTL, which he calls "an excellent diagnostic tool," allows Lapsley to perform partial path analysis, quickly eliminating portions of the path from consideration.

When setting up similar tests with Kashima Space Research Center in May 2004, he discovered that the end-to-end bandwidth was ~ 1 Mbps (which was not surprising, given that they were using 100 Mbps "commodity" access at each end). They decided to cache data in Washington, where they have a server, connected at 1 Gbps to ISI-E, MAX, and Abilene, that is able to get high bandwidth to Tokyo across Transpac. They were very surprised to find that the bandwidth was less than 4 Mbps using eight parallel TCP streams. When faced with the usual situation of trying to isolate the problem using pairwise Iperf testing between nodes and then trying to construct an overall picture of the network status, a tedious, time consuming and often error prone situation, they decided instead to "instrument" their test network with BWCTL:

http://web.haystack.mit.edu/staff/dlapsley/tsev7.html

Although the website is still basic, it is the first time they have had a good overall view of what is going on in the network and how it changes with time. Lapsley reports that it took only one afternoon to setup the BWCTL partial mesh using some simple scripts that he had developed for the Onsala-Haystack tests.

Lapsley reports that, via BWCTL, they were able to isolate the cause of the performance problem to a particular network segment. Some additional detective work by his colleagues isolated the problem to a faulty NIC in a server on this segment. By isolating the problem, the NIC was replaced just in time for the e-VLBI experiment. In collaboration with their Japanese colleagues, Lapsley's team was able to reduce the time for a particular type of e-VLBI experiment by a factor of > 4.

BWCTL significantly reduces the problem isolation loop; instead of exchanges of email, Lapsley feels he can just point network engineers to this page and have them draw their own conclusions about the location of the problem.

In the future, Lapsley expects to use this tool to optimize the time of day for non-real-time bulk data transfers and identify the most aggressive application usable at any given time of day for real-time transfers. Lapsley is currently working to have BWCTL installed at all of his test sites; currently, it is available at Haystack, Onsala, and Kashima.

Who's Involved?

Active involvement in the "e-VLBI" project includes Lapsley, Dr. Alan Whitney, Associate Director of the MIT Haystack Observatory, and Jeff Boote, Internet2 Network Engineer, developer of the BWCTL tool. MIT Haystack Observatory, in collaboration with the MIT Laboratory for Computer Science, has recently been awarded a grant by NSF to develop new IP protocols specifically tailored to applications such as e-VLBI.

MIT Haystack gratefully acknowledges the support, assistance, and suggestions of Masaki Hirabaru and Yasuhiro Koyama (National Institute of Information and Communications Technology), Rüdiger Haas (Chalmers Onsala Space Observatory), Tom Lehman (University of Southern California, Information Sciences Institute) and staff of the APAN Tokyo XP.