Active Measurement Data Analysis Techniques

Todd Hansen[1]
Jose Otero[1]
Tony McGregor[2]
Hans-Werner Braun[1]

[1]National Laboratory for Applied Network Research (NLANR)
Measurement and Network Analysis Group
San Diego Supercomputer Center (SDSC)

[2]National Laboratory for Applied Network Research (NLANR)
Measurement and Network Analysis Group
University of Waikato, New Zealand

Contact author: Todd Hansen
National Laboratory for Applied Network Research (NLANR)
Measurement and Network Analysis Group
San Diego Supercomputer Center (SDSC)
University of California, San Diego
9500 Gilman Drive
La Jolla, Ca 92093-0505
tshansen@nlanr.net
(858) 822-0929

This work is sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. ANI-9807479. The government has certain rights in this material. We thank Maureen C. Curran for her editorial assistance.

Abstract

Active measurements, examine the capabilities of networks and the internet. Through the network analysis infrastructure (NAI), these measurements can provide useful data for network analysis. The National Laboratory for Applied Network Research (NLANR) team, has developed data sets which allow identifcation of a variety of problems which range from High Performance Connectivity and Commodity Network issues. to network hardware and routing issues. We present the steps necessary to secure and analyze such data.

Introduction

The National Laboratory for Applied Network Research (NLANR)[5] Measurement and Analysis Team[6] conducts both active and passive measurements in order to better understand the Internet and its capabilities. Active measurements are network probes developed to measure the capabilities of the internet. Active and passive measurements can be compared to car maintenance. When you are trying to determine what is wrong with your car, you can either check your car's oil level or give your car a test drive. A test drive would be an active measurement as it changes the state of the vehicle in question, while an oil check would be a passive measurement, which generally has no effect on the state of the car. With active measurements, one can generally retrieve additional information about a network's capabilities, at the cost of adding interference. In this paper, we discuss data analysis methods for NLANR's Active Measurement Project (AMP[7]). Our goal is to provide a basis for the understanding and analysis of active measurement data so that one can use this data to better understand their network connectivity. We detail methods, which will help one determine the overall performance of their network. This paper is designed to be an introduction to active measurements and analysis. As such, feedback[1] and suggestions on areas which are unclear, would be greatly appreciated.

The AMP System

The NLANR Active Measurement Project (AMP) is a distributed network of approximately 100 active monitors which systematically perform scheduled measurements between each other (Figure 1). These systems send data to a central data collector where it is made available for download (raw data) and public viewing (by Web browser[7]). Currently, we are working on additional methods for data presentation which may be more useful for general network analysis.

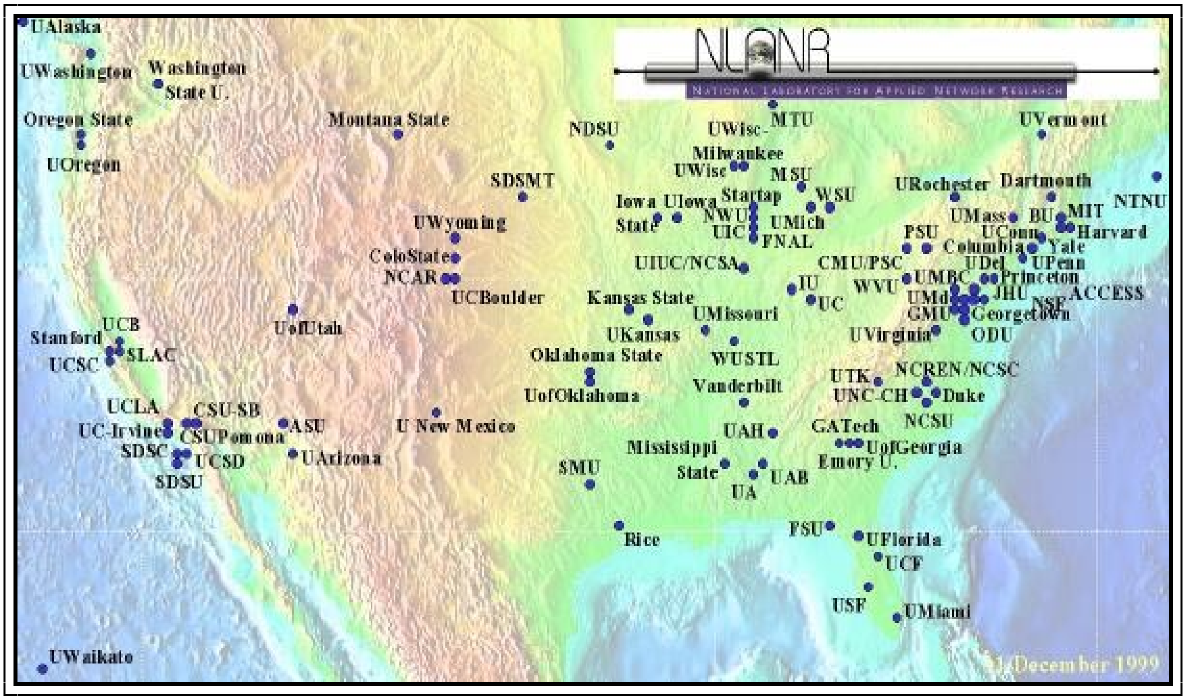


Figure 1: Map of AMP Probe constellation (Dec 1999)

Measurement Methods: what we measure

The AMP constellation measures round trip times (RTT) between each of its systems once a minute. Every ten minutes, `traceroutes` are preformed to determine routes between each of the measurement boxes. We choose to measure RTT instead of one-way delay because this measurement is easier to perform and does not rely on external devices to synchronize the time between each of the monitors. Other measurement projects use Global Position System (GPS) receivers to synchronize the time between hosts. We have found that these systems are too expensive and difficult to install, for the limited amount of additional information gained.

In greater detail, our RTT measurements are made once a minute (randomized by 15 seconds) using the `fping` program. This program sends an Internet Control Message Protocol (ICMP) echo packet to each host and waits for an ICMP reply packet. It then records the measured delay for each site. The benefit to `fping` is that it allows us to simultaneously ping multiple sites. (If we had to ping each site serially, it would take much longer than a minute.) A loss is determined when four ICMP echo request packets are sent and no replies are received before each timeout. Our route data is collected every 10 minutes (randomized by 15 seconds) using the `traceroute` program. Our Web interface allows one to view this data by host pairs. For example, after selecting a primary site (Figure 2), the Web interface shows the previous day's average RTT and loss, for each site within its mesh (Figure 3). The interface also allows one to select a secondary site, and view weekly graphs, which show the RTT between the primary and secondary sites over a weeks time (Figure 4). From this data, one can choose to look at daily RTT or loss graphs. One can also view a listing of route changes that may have occurred on a specified day (Figure 6).

Meshes and Peer Networks

A full mesh is a complete interlinking of objects, in which every object is directly connected to every other object. Our AMP constellation does a full mesh of measurements. This is a data and resource intensive process, and as such, is not scalable beyond 150 machines. Therefore, it is necessary to create peer networks, which have their own network of hosts doing measurements within the full mesh. Unfortunately, this does not allow for studying networks and routes which traverse multiple peer networks. For this reason, we configure mutual peering points, a couple of hosts in each peer's network, that perform measurements within both meshes. By strategically placing these peers we should, in theory, be able to gather useful information about the connectivity between the peer networks. We are currently in the process of experimenting with peer networks; the results of this work should be available sometime in the near future.

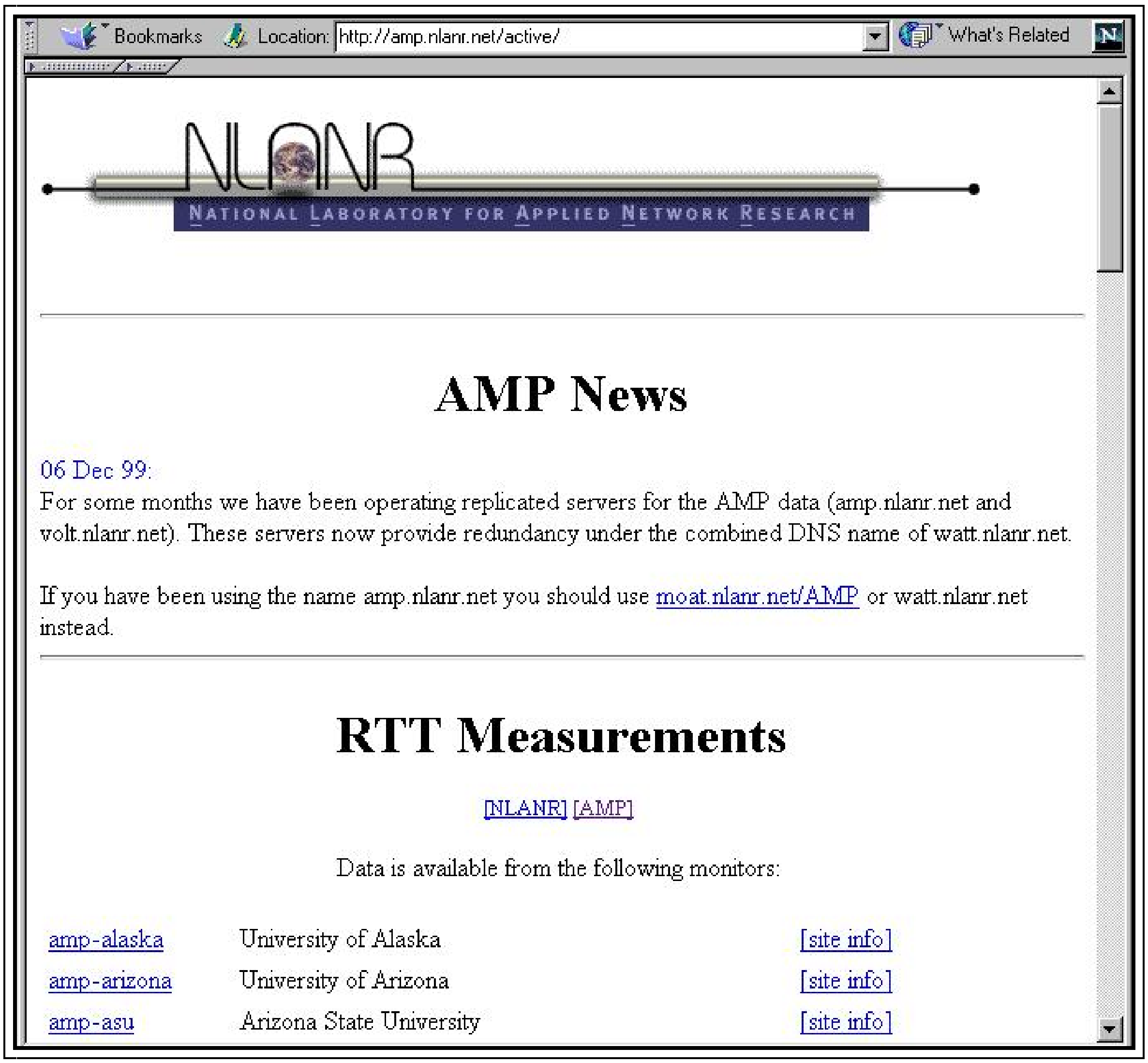
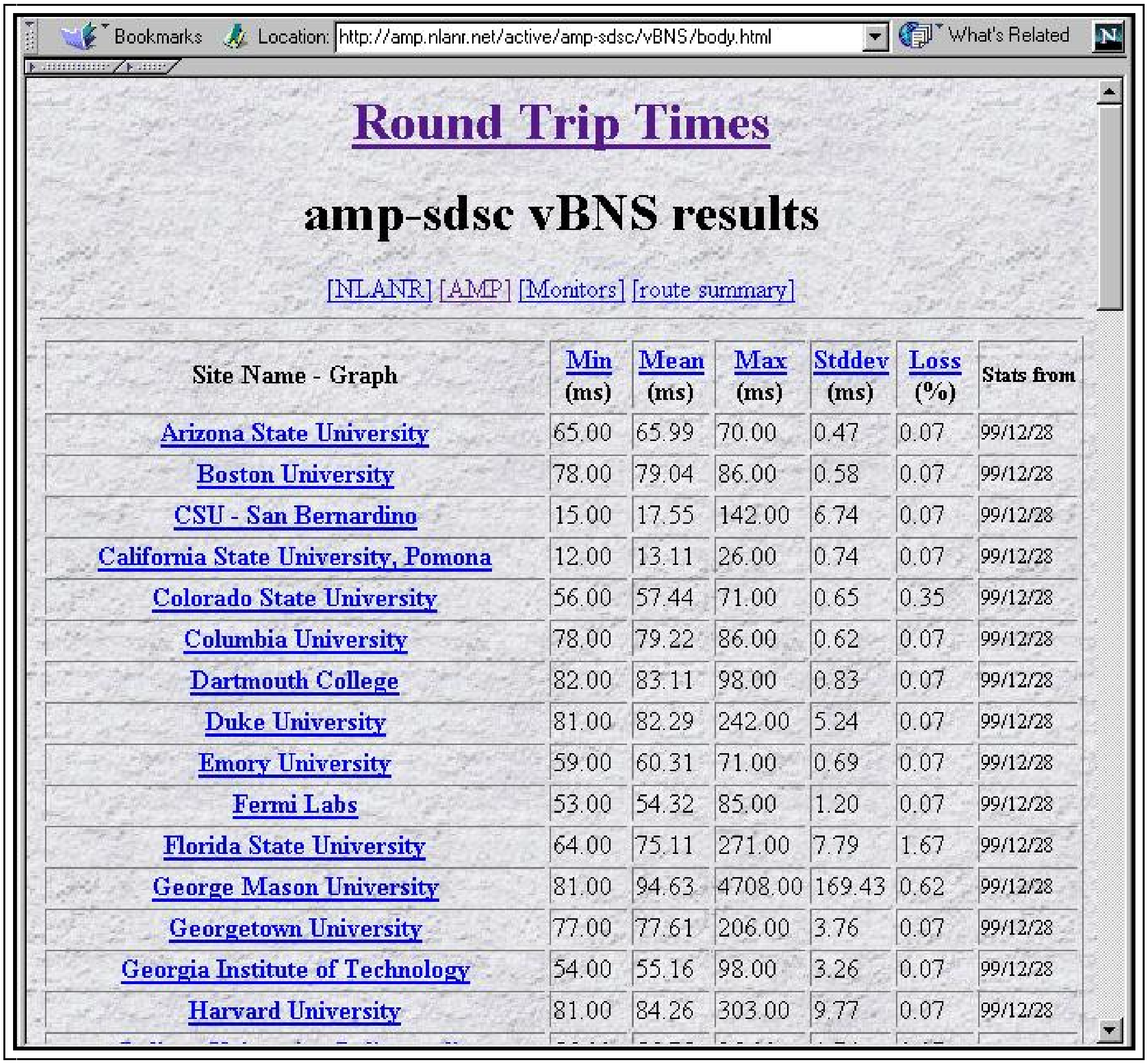


Figure 2: Main AMP Web interface page, select primary monitor to view.



Bookmarks Location: http://amp.nlanr.net/active/amp-sdsc/vBNS/body.html What's Related

Round Trip Times

amp-sdsc vBNS results

[NLANR] [AMP] [Monitors] [route summary]

Site Name - Graph	Min (ms)	Mean (ms)	Max (ms)	Stddev (ms)	Loss (%)	Stats from
Arizona State University	65.00	65.99	70.00	0.47	0.07	99/12/28
Boston University	78.00	79.04	86.00	0.58	0.07	99/12/28
CSU - San Bernardino	15.00	17.55	142.00	6.74	0.07	99/12/28
California State University, Pomona	12.00	13.11	26.00	0.74	0.07	99/12/28
Colorado State University	56.00	57.44	71.00	0.65	0.35	99/12/28
Columbia University	78.00	79.22	86.00	0.62	0.07	99/12/28
Dartmouth College	82.00	83.11	98.00	0.83	0.07	99/12/28
Duke University	81.00	82.29	242.00	5.24	0.07	99/12/28
Emory University	59.00	60.31	71.00	0.69	0.07	99/12/28
Fermi Labs	53.00	54.32	85.00	1.20	0.07	99/12/28
Florida State University	64.00	75.11	271.00	7.79	1.67	99/12/28
George Mason University	81.00	94.63	4708.00	169.43	0.62	99/12/28
Georgetown University	77.00	77.61	206.00	3.76	0.07	99/12/28
Georgia Institute of Technology	54.00	55.16	98.00	3.26	0.07	99/12/28
Harvard University	81.00	84.26	303.00	9.77	0.07	99/12/28

Figure 3: Site information specific to the primary site chosen, showing averages for previous 24 hours. From here, the user can select a secondary site and view further measurements.

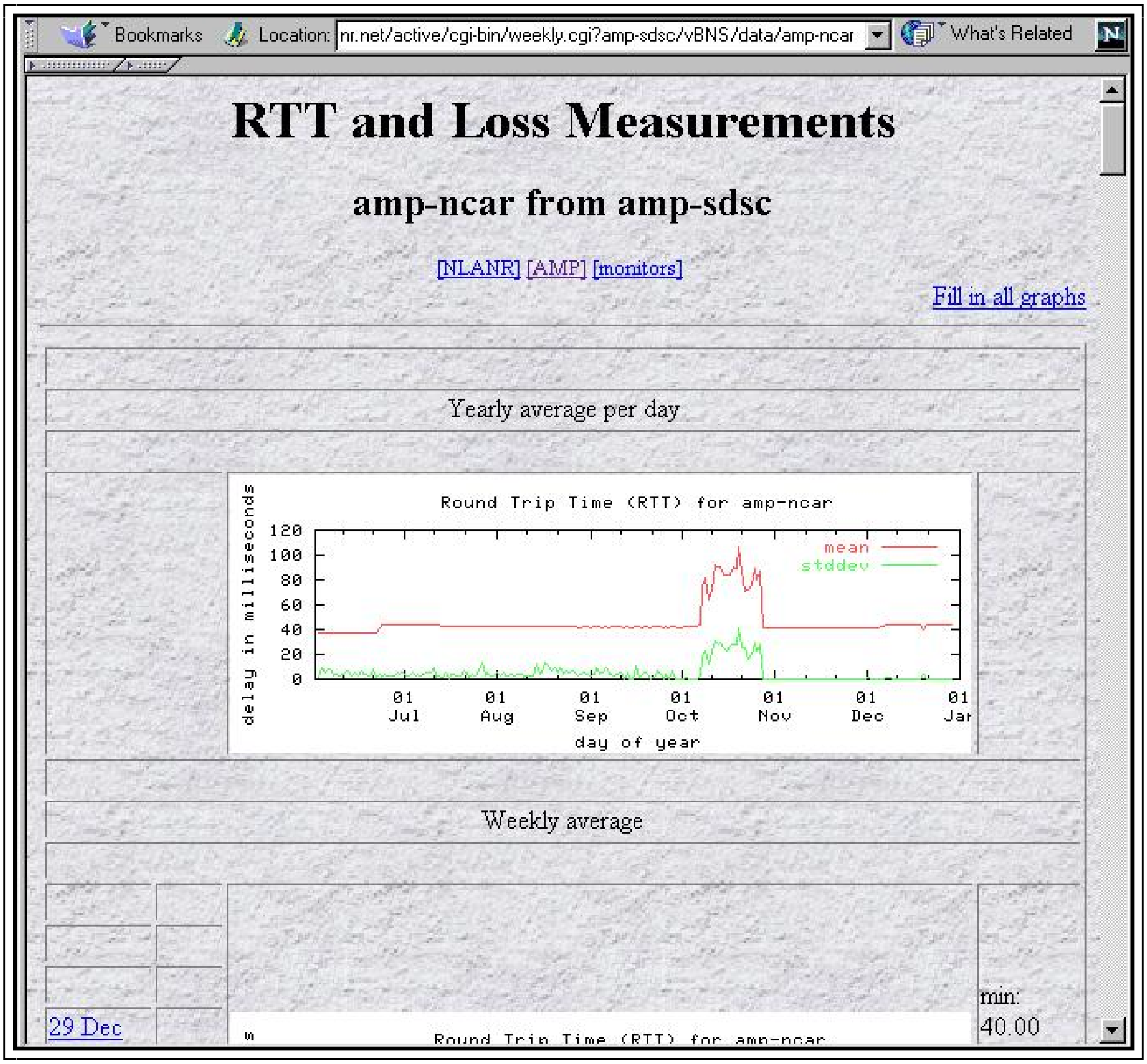


Figure 4: Primary (amp-sdsc) to secondary (amp-ncar) site data is displayed as round trip time (RTT) graphs. Yearly and weekly graphs are also presented. From this page you can choose to view a graph of a single day's data (Figure 5) or you can choose to look at the routes for a specific day (Figure 6).

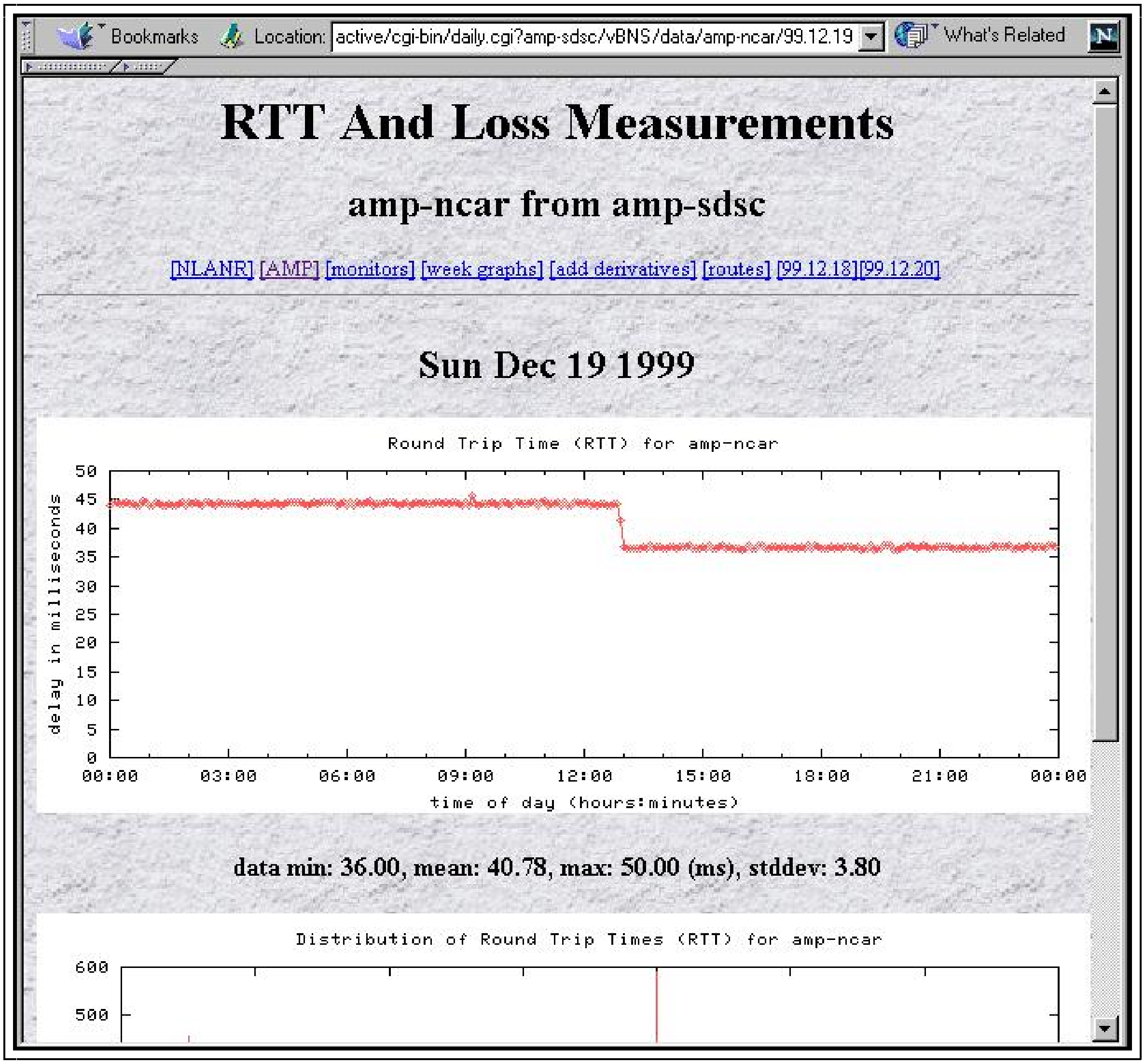
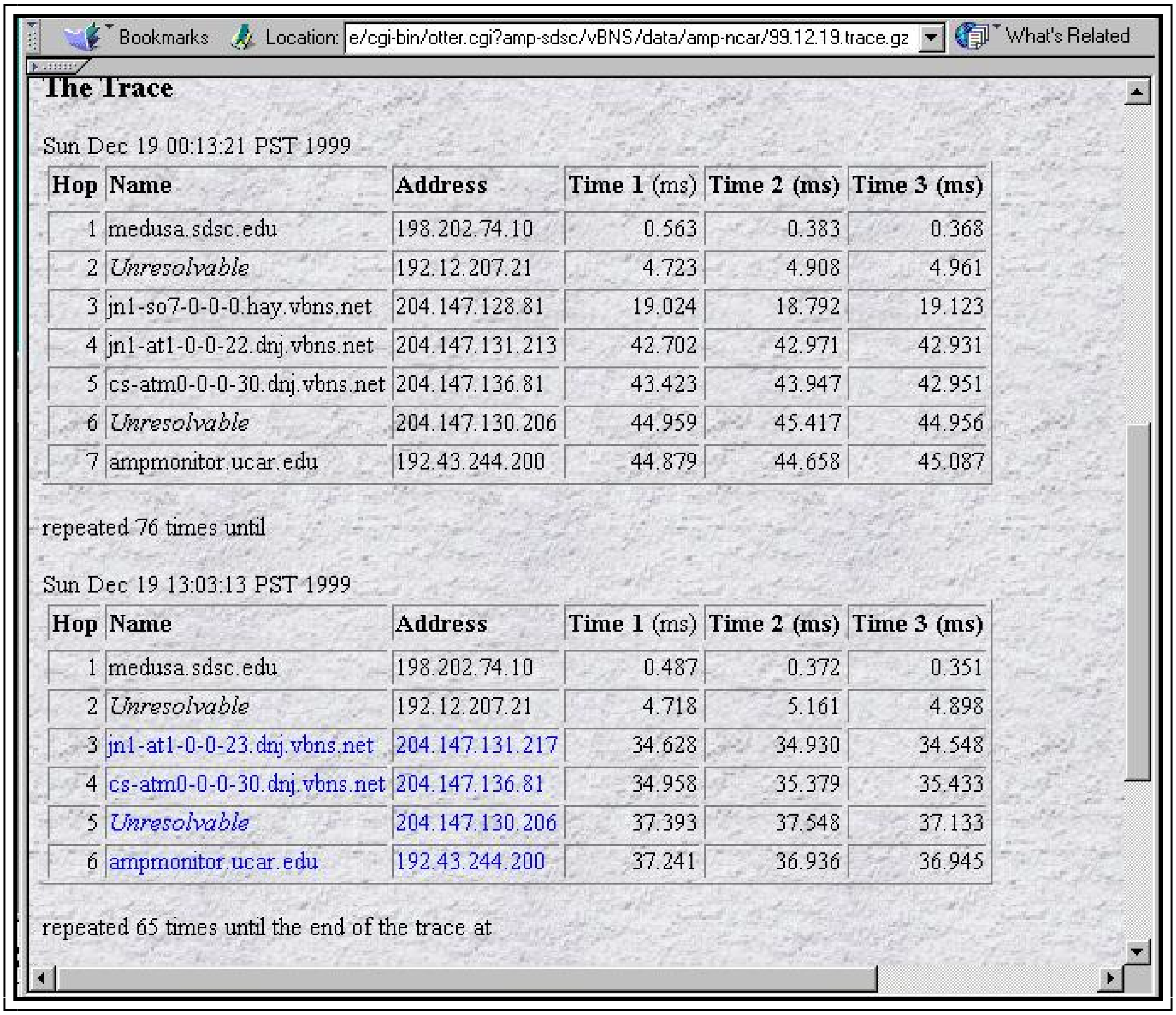


Figure 5: Graph of a single day's measurements between the primary and secondary sites. This can be used to closely examine interesting anomalies within data or to more closely examine events which may not show clearly in the weekly summaries. Packet loss and round trip times (RTT) for a particular day are also presented in graphical form, on these pages. The graph above exhibits a step down in RTT as the result of a router switch.



Bookmarks Location: e/cgi-bin/otter.cgi?amp-sdsc/vBNS/data/amp-ncar/99.12.19.trace.gz What's Related

The Trace

Sun Dec 19 00:13:21 PST 1999

Hop	Name	Address	Time 1 (ms)	Time 2 (ms)	Time 3 (ms)
1	medusa.sdsc.edu	198.202.74.10	0.563	0.383	0.368
2	*Unresolvable*	192.12.207.21	4.723	4.908	4.961
3	jn1-so7-0-0-0.hay.vbns.net	204.147.128.81	19.024	18.792	19.123
4	jn1-at1-0-0-22.dnj.vbns.net	204.147.131.213	42.702	42.971	42.931
5	cs-atm0-0-0-30.dnj.vbns.net	204.147.136.81	43.423	43.947	42.951
6	*Unresolvable*	204.147.130.206	44.959	45.417	44.956
7	ampmonitor.ucar.edu	192.43.244.200	44.879	44.658	45.087

repeated 76 times until

Sun Dec 19 13:03:13 PST 1999

Hop	Name	Address	Time 1 (ms)	Time 2 (ms)	Time 3 (ms)
1	medusa.sdsc.edu	198.202.74.10	0.487	0.372	0.351
2	*Unresolvable*	192.12.207.21	4.718	5.161	4.898
3	jn1-at1-0-0-23.dnj.vbns.net	204.147.131.217	34.628	34.930	34.548
4	cs-atm0-0-0-30.dnj.vbns.net	204.147.136.81	34.958	35.379	35.433
5	*Unresolvable*	204.147.130.206	37.393	37.548	37.133
6	ampmonitor.ucar.edu	192.43.244.200	37.241	36.936	36.945

repeated 65 times until the end of the trace at

Figure 6: This image shows routing information between the primary and secondary sites for the day selected. As you can see the page highlights route changes in blue. Using this page one can view, in graphical form, all routes for a particular day, using the Otter utility.

Data Analysis Methods: What you can determine

Our data is collected and organized in a hierarchical structure on the Web.[7] After entering the Web interface, select a primary site (Figure 2). Upon selection of the primary site, measurements are performed and data is sent to central data collector. You will then be presented a summary page (Figure 3), which shows the day's measurements from the primary site, to each of the secondary sites. (The secondary sites are the sites to which the measurements are made). These sites are the same ones, as seen in the primary site list. The presented summary is useful for a quick view of status and for identifying possible problem sites. From this page, choose a secondary site. After selecting the secondary site, you will be able to view measurement data collected between the primary and secondary sites (Figure 4). The top graph is a yearly summary of the measurements, followed by weekly summary graphs giving detailed information. The weekly graphs are very useful in determining the time of a network event. From the weekly graphs you can select to view an even more detailed graph of measurements for a particular day (Figure 5). This is done by clicking on the graph or by clicking on the date to the left of the graph. You can also choose to view a listing of route changes for the day (Figure 6). In theory, using this hierarchical structure, one can view all the data that is recorded by our systems. Unfortunately, due to the high volume of data (70 Gb) this is not very releasitic.

To analyze the network connectivity for a specific site: select that site as the primary site; then go through each of the secondary sites listed, looking at each of the weekly graphs. For example, if a site shows 100% loss, it is most likely because of a filter (a router that blocks certain packets). This can be verified by looking at the problem log for the site, which is available from the primary site selection page (Figure 2). It is also useful to look at the recent weekly graphs for any network events, problems or trends.

In the case studies detailed below, we attempt to give you many different examples of network events. For example, you might notice that your site's connection to UC San Diego exhibits commodity link congestion curves (Case Study #2) and upon further investigation you find that this is due to excessive use of the network or simply the fact that your route to UCSD is via your commodity connection. In general, a little bit of investigation with the information we provide can lead to useful insights. However, we can never provide all of the answers. In general, if you go through all of the secondary sites for your primary site, you should get a good idea of your connectivity and where any problems may lie. Keep in mind that it may be useful to investigate a few reverse path measurements from other sites to your own. This may give you a better idea on how the world connects with you.

If you want to get an idea of the overall status of the network, we recommend going through a number of different primary and secondary site pairs, almost at random. Perhaps over time you can come up with a useful list of whic sites are important or significant to you. Unfortunately, with 100 sites it would take too long to look at each of the 10,000 possible pairs. Therefore, you must decide on some way to limit your search in order to view only relevant information. Another method worth mentioning is to look at two different secondary sites for each primary site. This way you can get a good overall picture.

The point of general network analysis is to get an idea of the stability of the network and the performance of the links. We are currently working on several tools to aid in this. One of these is the AMP data reports page. This allows you to get a quick summary of all the routes between each of the different sites; this can be very useful in finding problem sites. Another useful tool is the Cichlid 3-D visualization system, which can graph all of the sites RTT/Loss over the previous 24 hours. We are also working on ideas to develop event triggers and route decomposition to pinpoint when and where network problems occur. The intent of these tools is to give you an idea of the stability of the networks, as well as to alert you when something interesting occurs.

One of the best ways to find trouble or to diagnose problems with a site is to look for bumps or steps in the round trip time (RTT). These bumps tend to show a routing change, or some configuration change that alters the RTT between two sites. See Figure 7 for an example.

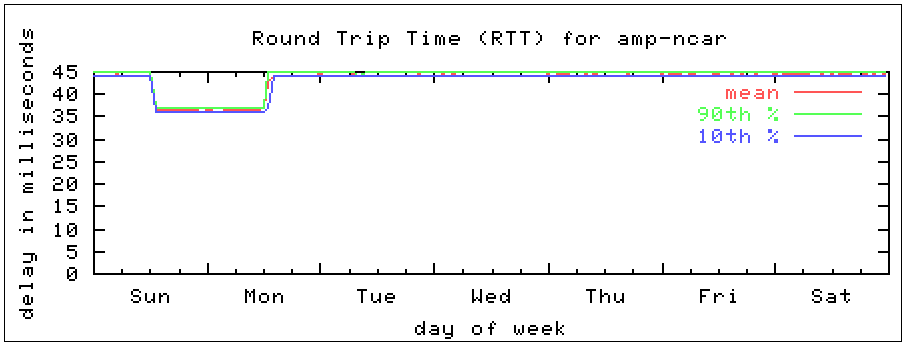


Figure 7: Weekly graph (Dec 19 – Dec 25) from SDSC to NCAR. The temporary step down is caused by the removal of a router by vBNS engineering. The router was replaced on Monday.

Another thing to look for is high loss. If a site has high loss it may have faulty hardware. This analysis is very basic. A more interesting understanding of your site's performance can be acquired by comparing its connectivity with that of other sites. For example, is there the same congestion on your commodity links that other sites exhibit on theirs? What about the round trip times of your vBNS[9] or Abilene[10] connections? Are they relatively good or is there something in your local network that is making the RTT unstable. Another thing to look at, is how the congestion curves (of the RTT graph) are shaped: do they have a smooth shape? Are they bigger than those for other sites? If the graph shows congestion curves, then how does it affect your traffic?

In general, our feeling is that a high performance network should have low RTT and no significant congestion curve. One factor to examine is what TCP window sizes within the TCP stacks, would yield maximum performance. This can show a great deal about the feasibility of high speed communication over these networks. Also, how much different is a commodity connection at noon on Friday compared to a vBNS link? Hopefully the answers to these questions will help give an accurate picture of the connectivity and performance of your network connection.

Cases Studies

We feel that the following case studies illustrate some of the important aspects of network performance analysis. The first two case studies show the characteristics of high performance networks and commodity networks. We invite your comments and suggestions on ways to improve the tractability of these issues.

Index for Case Studies:

Please note: Case Studies #1 and #2 can be used as base line references for normal vBNS and commodity connections behaviour.

Case Study #1: A normal vBNS connection.

Case Study #2: A normal commodity connection.

Case Study #3: Long-term network outage

Case Study #4: A typical HPC network failure

Case Study #5: An unusual situation where RTT decreases after a switch from HPC to commodity network connection

Case Study #6: Route Change

Case Study #7: Asynchronous Routing

Case Study #8: Reverse Path Router Change

Case Study #9: Increased Router Hops Lower RTT

Case Study #10: End Network Failure

Case Study #1: A normal vBNS connection

week of:	14 Nov 1999
primary site:	San Diego Supercomputer Center (SDSC)
secondary site:	National Center for Supercomputer Applications (NCSA)

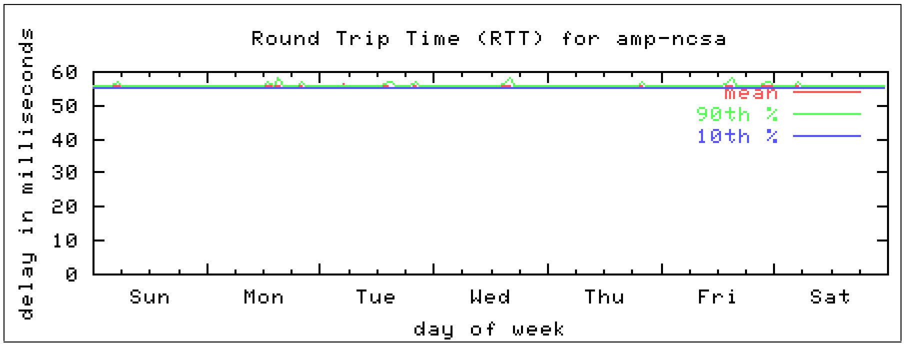


Figure 1-1: RTT delay for SDSC to NCSA for week of 11-14-99

Explanation: This is a classic example of a high performance network connection. The most important characteristic of this type of network is the lack of round trip time (RTT) variance. Figure 1-1 shows that the network does not suffer from congestion in the same way as most commodity networks. (Case Study #2, will discuss a commodity link connection.) A second feature of this Case Study is the relatively low RTT. It averages 55ms while standard commodity links are well over 150ms. The combination of these features (lack of congestion and low RTT) differentiate high performance networks different from commodity networks.

In this vBNS example, we do see minor variance in the round trip time. While it is only averages 1-3 milliseconds, this is unusual for a vBNS connection. This type of behavior, may indicate an overly used link. A benefit to studying vBNS, connections is that when a major network event occurs, it becomes very pronounced when graphed. On a commodity connection, this is not the case, as there is much more noise which obscures the event. One reason we are seeing such a low RTT is due to the fact that the link travels only over the vBNS, as opposed to multiple networks (three to five) which are common for a commodity links. Traveling over the vBNS, which attempts to achieve only two hops between universities, reduces the total number of hops in each route.

Supporting Documentation: N/A

Case Study #2: A normal commodity connection

week of:	10 Oct 1999
primary site:	University of Alaska
secondary site:	California State University, Pomona

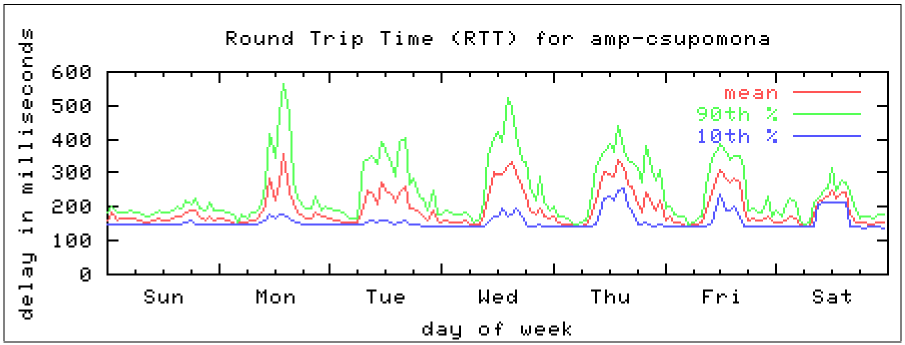


Figure 2-1: RTT delay for U. of Alaska to CSU Pomona for week of 10-10-99

Explanation: Figure 2-1 shows a weekly graph of a commodity network connection. It clearly shows the daily congestion curves. These curves are characteristic of commodity networks, which tend to become saturated during the week days. As a result, this causes RTT to drastically increase, and packets to experience queuing delays and higher latency.

It would be interesting to study the correlation between loss data statistics and congestion phenomena. However, on our systems, a loss is defined when four consecutive request packets do not receive a reply. This tends to only occur during network failures, and does not show the severity of network congestion. Current work is being done, at this time, to correct the problem.

In addition to the daily cycle of the congestion curve, you can also see a weekly cycle, where the weekends exhibit much lower congestion. You can also see how congestion falls slightly as we approach the end of the work week. Comparing this Case Study with Case Study #1, a vBNS connection, we can see a definite advantage to being on a high performance network connection. Not only does the vBNS have high bandwidth and lower latency, but it also lacks the congestion of typical commodity networks. Another interesting feature of commodity networks is that traffic tends to flow over many different networks before it reaches the end host, whereas with the vBNS and Abilene, high performance networks, the traffic flows over only 1-2 networks.

Supporting Documentation: None

Case Study #3: Long-term network outage

week of: 29 Sep 1999
primary site: University of Alaska
secondary site: California State University, Pomona

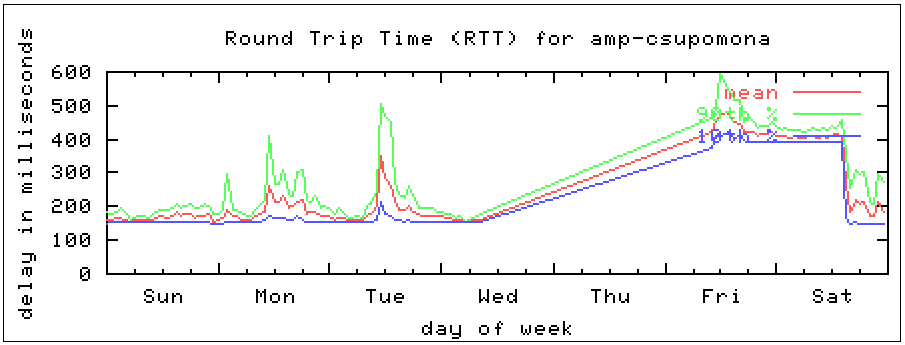


Figure 3-1: RTT delay for U. of Alaska to CSU Pomona for week of 9-29-99

Explanation: In this example, the University of Alaska lost its commodity connection to the internet for a couple of days. Because CSU Pomona was not connected to either the vBNS or Abilene, there was no alternate route for the data flow. Therefore, during the commodity connection outage, this route suffered 100% loss. On Friday the problem was fixed and the commodity connection was reestablished. From Wednesday through early Friday the slope of the graph is smooth, because during this time, in which there was 100% packet loss, no data was available for the plotting engine. As a result, the plotting program connected the last measurement on Wednesday with the first measurement on Friday. (Creating a straight line.) In order to help us analyze the situation depicted in Figure 3-1, we used the daily graphs (Figure 3-2) which clearly show 100% packet loss during the network outage. These data sets show a lack of measurements from Wednesday until Friday morning. We also see from the route data sets (Figure 3-3) for Wednesday, that the campus connection, originally traveling over commodity networks, ends at the campus edge during network failure. This example, is a good argument for redundant networking. In the even of another network outage, traffic would be able to flow over alternate paths, eliminating network down time.

Supporting Documentation: Examples of trace route data before and during network outage.

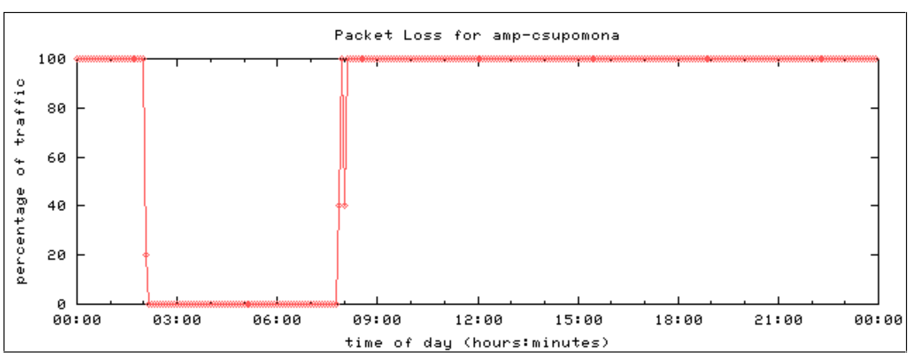


Figure 3-2: Packet loss for Wednesday, notice how 100% were lost for a majority of the day. This 100% loss continued until the problem was solved on Friday.

Routes Before Outage (Sep 27 07:30:00 PST)	**Routes during outage** (Sep 27 08:00:00 PST)
Unresolvable	Unresolvable
Unresolvable	Unresolvable
Maewestbr-aip.att-disc.net	
sl-w1-mae-0-0-0-100M.sprintlink.net	
sl-bb2-stk-4-0-45M.sprintlink.net	
sl-bb10-stk-2-2.sprintlink.net	
sl-gw11-stk-0-0-0.sprintlink.net	
sl-csuhay-1-0-0-T3.sprintlink.net	
WESTED-HAY-ATM.CSU.net	
POM-WESTED-ATM.CSU.net	
AMP-CSU	

Figure 3-3: Routing information for Wednesday before and during the network outage. This data helps pinpoint the cause for network failure.

Case Study #4: A typical HPC network failure

week of:	26 Dec 1999
primary site:	University of Alaska
secondary site:	South Dakota School of Mines (SDSMT)

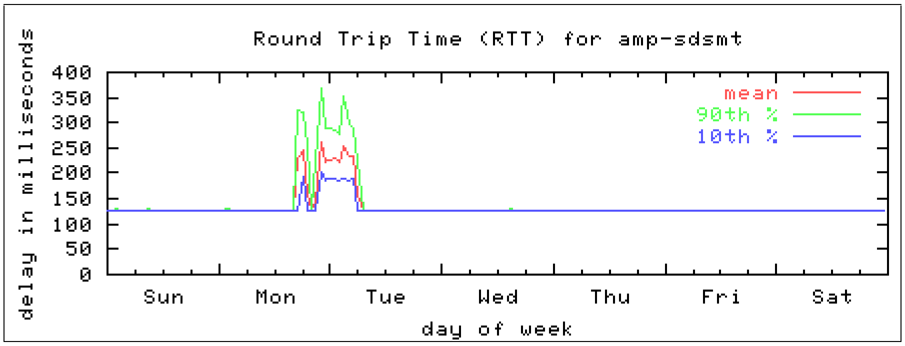


Figure 4-1: RTT delay for U. of Alaska to SDSMT for week of 12-26-99

Explanation: Figure 4-1 shows a normal High Performance Connection (HPC), with the exception of a RTT jump on Monday until early Tuesday. It is clear that there is little to no variance for most of the week. During late Monday and early Tuesday, however, hardware problems with the Abilene router caused the RTT to increase in excess of 225 ms from its average 125 ms. Inspection of the routing data (Figure 4-2), reveals that the hardware failure caused the path to change from the normal high performance connection to the slower and more congested commodity network. During HPC network outages, it is common for paths to re-route over commodity links. Once the failure was resolved, the path and RTT returned to normal.

Supporting Documentation: See next page.

Supporting Documentation: Examples of trace route data previous to, during, and after hardware failure.

Before hardware failure (Mon Dec 27 00:08:16 PST 1999)	Δ	**During hardware failure** (Mon Dec 27 17:18:57 PST 1999)
Unresolvable		Unresolvable
M40	*	Unresolvable
Uacore1-ge-0-0-0-0.pnw-gigapop.net	*	Ua-gw.alaska.edu
Westincore1-so-0-1-0-0.pnw-gigapop.net	*	d1-1-3-1.a06.sttlwa01.us.ra.verio.net
Abilene-gsr-GE4-0.pnw-gigapop.net	*	Ge-6-0.r02.sttlwa01.us.bb.verio.net
Sttl-pnwgp.abilene.ucaid.edu	*	Sea2.ord0.verio.net
Dnvr-sttl.abilene.ucaid.edu	*	Ord0.ord2.verio.net
Kscy-denv.abilene.ucaid.edu	*	Ord2.stl1.verio.net
ks-2-p00.r.greatplains.net	*	Stl1.vcp1.verio.net
ks-2-sdsmt.r.greatplains.net	*	Stl0-s600.cp.verio.net
Amp-SDSMT	*	Kcm0-s400.cp.verio.net
	*	Kcm1-fa100.cp.verio.net
	*	Oma0-s100.cp.verio.net
	*	Sxf0-h20.cp.verio.net
	*	Sxf0-bit-s0.cp.verio.net
	*	Sd-sdsmt-2-rc-s3.sd.net
	*	Hardrock-out.sdsmt.edu
	*	Cisco081.sdsmt.edu
	*	Amp-SDSMT

Figure 4-2a: Routing information before and during the hardware failure.

After hardware failure (Tue Dec 28 23:58:21 PST 1999)
Unresolvable
m40
uacore1-ge-0-0-0-0.pnw-gigapop.net
Westincore1-so-0-1-0-0.pnw-gigapop.net
abilene-gsr-GE3-0.pnw-gigapop.net
sttl-pnwgp.abilene.ucaid.edu
dnvr-sttl.abilene.ucaid.edu
kscy-denv.abilene.ucaid.edu
ks-2-p00.r.greatplains.net
ks-2-sdsmt.r.greatplains.net
Amp-SDSMT

Figure 4-2b: Routing information after the hardware failure.

Case Study #5: An unusual situation where RTT decreases after a switch from HPC to commodity network connection

week of: 25 Jul 1999
primary site: Southern Methodist University (SMU)
secondary site: San Diego Supercomputer Center (SDSC)

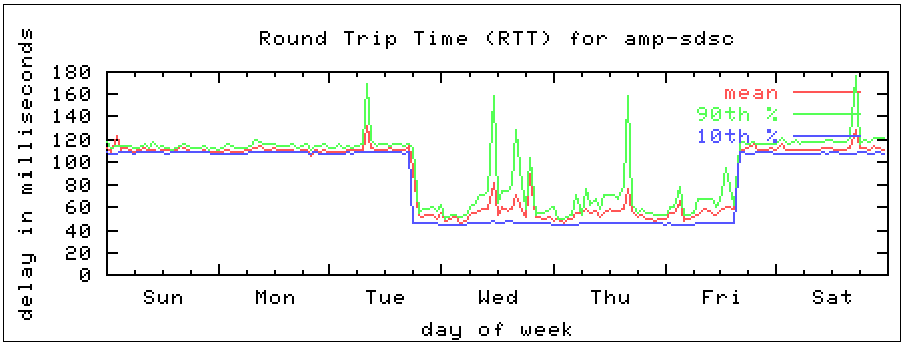


Figure 5-1: RTT delay for SMU to SDSC for week of 7-25-99

Explanation: This is an interesting example, in which the same thing happens as in Case Study #4 (network failure), but with different results. In this case, when the High Performance Connection failed, traffic was switched to the commodity link as expected (Figure 5-2a). However, in this case the commodity route was *faster* than the HPC connection. On Friday, the problem was resolved and the HPC route was re-established (Figure 5-2b).

The question here is, why was the commodity route faster? One reason is because when traffic flowed over the commodity link it had to transverse one backbone network, whereas when it was flowed over the high performance connection it transversed two networks (Abilene and vBNS) and an exchange point (Figure 5-2a). Although high performance networks may be faster, data is often sent out of the way in transversing these networks, causing it to pass through more network hops thereby increasing RTT. In this case using the commodity network connection resulted in lower latency and faster RTT. Making a permanent switch to the commodity network, however, would result in congestion patterns like those of Case Study #2, or those seen for Wednesday and Thursday in figure 5-1. Aside from these issues, this case study shows successful redundant networking. When the HPC connection failed, redundant networking (commodity routes) allowed data to continue flowing between SMU and SDSC.

Supporting Documentation: Examples of trace data before, during and after RTT drop

Route before drop in normal RTT (Tue Jul 27 00:41:36 PDT 1999)	Δ	**Route during drop in normal RTT** (Tue Jul 27 20:20:43 PDT 1999)
Isem.smu.edu		Isem.smu.edu
Cisco.smu.edu		Cisco.smu.edu
Abilfar.smu.edu	*	S2-0-0-6.dfw-bb1.cerf.net
Atla-hous.Abilene.ucaid.edu	*	Unresolvable
Nycm-atla.abilene.ucaid.edu	*	Pos8-3-155M.lax-bb4.cerf.net
Vbns-abilene.abilene.ucaid.edu	*	Atm1-0-2-622M.san-bb6.cerf.net
Cs-atm0-0-20.sdsc.vbns.net	*	Pos10-0-0-155M.san-bb1.cerf.net
Unresolvable	*	Sdsc-gw.san-bb1.cerf.net
AMP-SDSC	*	Medusa-atm.sdsc.edu
		AMP-SDSC

Figure 5-2a: Routing information before and during switch from HPC to commodity networks.

Route after return to normal RTT (Fri Jul 30 16:00:38 PDT 1999)
Isem.smu.edu
Cisco.smu.edu
Abilfar.smu.edu
Atla-hous.Abilene.ucaid.edu
Nycm-atla.abilene.ucaid.edu
Vbns-abilene.abilene.ucaid.edu
Cs-atm0-0-20.sdsc.vbns.net
Unresolvable
AMP-SDSC

Figure 5-2b: Routing information after switching back to HPC networks.

Case Study #6: Route Change

week of: 10 October 1999
primary site: University of Alaska
secondary site: University of California, Irvine (UCI)

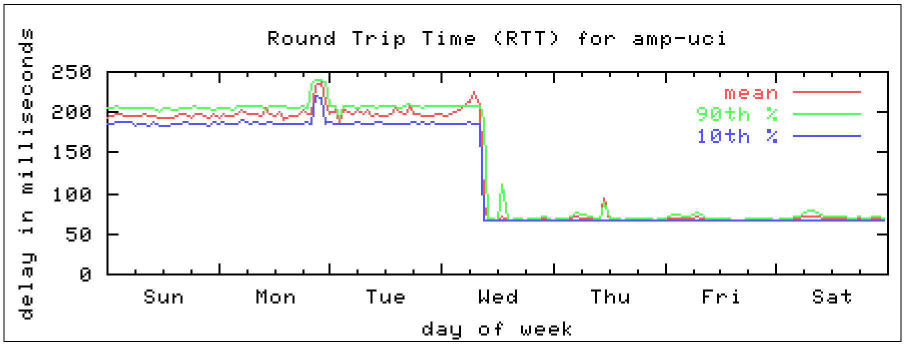


Figure 6-1: RTT delay for U. of Alaska to UCI for week of 10-10-99

Explanation: This example shows a beneficial route change for UCI. From Figure 6-1, we notice a large decrease (step down) in RTT. Looking at the daily data for October 10th (Figure 6-2) we can again see a very clear step down in RTT values. Upon inspection of the trace route data (Figure 6-4) the following conclusions can be made. Prior to October 13th, 1999 traffic from Alaska to UCI went over the HPC Abilene connection, through an HPC exchange, over vBNS and to calren2 (a regional network connection). This route had an average RTT of 200 ms. After October 13th, a direct connection from Abilene to calren2 eliminated the vBNS exchange point and decreased RTT. RTT variance also lowered. The long-term effects of this beneficial change can best be seen in Figure 6-3, where the data is graphed by month.

Supporting Documentation: Examples of weekly and yearly RTT data; trace route data

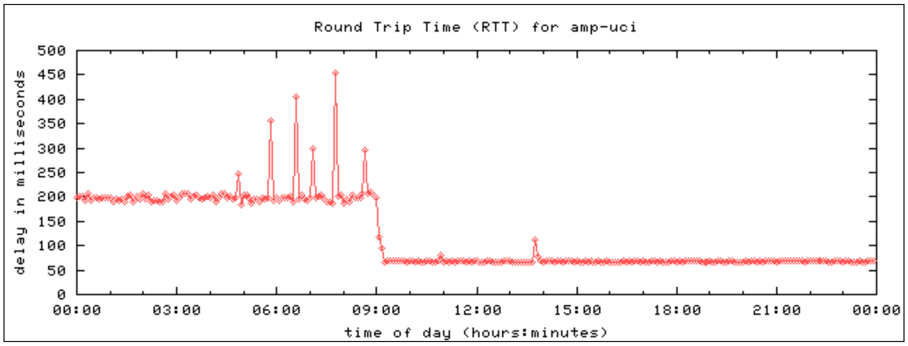


Figure 6-2: **RTT data for October 10, 1999**

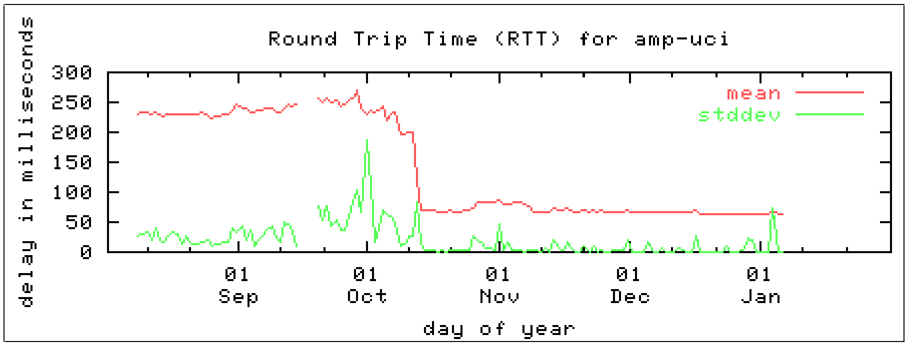


Figure 6-3: Yearly RTT data for U. of Alaska to UCI

Before Route Change (Wed Oct 13 00:13:41 PDT 1999)	Δ	**After Route Change** (Thu Oct 14 11:47:15 PDT 1999)
Unresolvable		Unresolvable
Fai-fw-e1.sons.alaska.edu		Fai-fw-e1.sons.alaska.edu
M40		M40
Uacore1-ge-1-0-0-0.pnw-gigapop.net		Uacore1-ge-0-0-0-0.pnw-gigapop.net
Westincore1-so-0-1-0-0.pnw-gigapop.net		Westincore1-so-0-1-0-0.pnw-gigapop.net
Abilene-gsr-GE4-0.pnw-gigapop.net		Abilene-gsr-GE4-0.pnw-gigapop.net
Sttl-pnwgp.abilene.ucaid.edu		Sttl-pnwgp.abilene.ucaid.edu
Scrm-sttl.abilene.ucaid.edu		Scrm-sttl.abilene.ucaid.edu
Denv-scrm.abilene.ucaid.edu	*	Losa-scrm.abilene.ucaid.edu
Kscy-denv.abilene.ucaid.edu	*	USC—abilene.ATM.calren2.net
Ipls-kscy.abilene.ucaid.edu	*	UCI—USC.POS.calren2.net
Unresolvable		Durin.gw.nts.uci.edu
Cs-atm0-0-0-24.rto.vbns.net		Unresolvable
UCI-vBNS.Calren2.net	*	Cpl-cdl-rsm1-vl401.nts.uci.edu
Durin.gw.nts.uci.edu		AMP-UCI
Unresolvable		
Unresolvable		
AMP-UCI		

Figure 6-4: Routing information before and route change.

Case Study #7: Asychronous Routing

week of:	12 December 1999
primary site:	San Diego Supercomputer Center (SDSC)
secondary site:	Washington University, St. Louis (WUSTL)

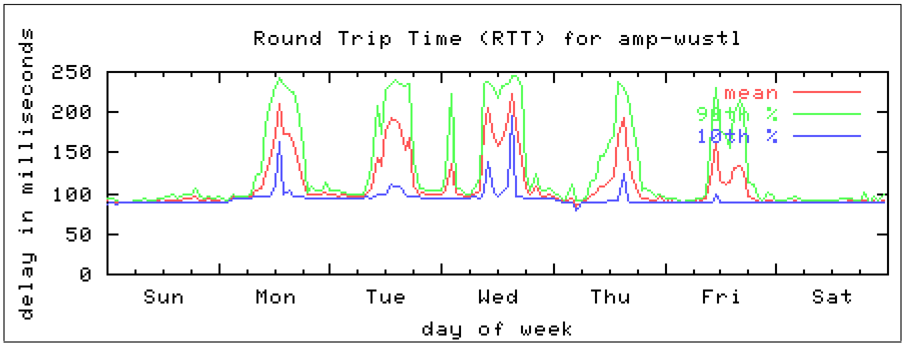


Figure 7-1: RTT delay for SDSC to WUSTL for week of 12-12-99

Explanation: Figure 7-1 shows the typical characteristics of a commodity route. Notice how the daily RTT values increase and decrease throughout the work week. This signature is commonly associated with increased and decreased network congestion throughout the day. Upon inspection of the trace route data in figure 7-2, however, we see that traffic from SDSC to WUSTL goes over the vBNS, meaning that both sites are connected to high performance networks, resulting in a seeming paradox. Instead of seeing traffic patterns like those of Case Study #1 (a high performance connection), we see signatures much like those in Case Study #2 (a standard commodity connection). The reason for this 'commodity-like' congestion needs explanation. With further investigation, we find that data from SDSC to WUSTL is routed over the vBNS network, while traffic from WUSTL to SDSC transverses the commodity network (Figure 7-2). Possible reasons for this curiosity can be: 1) a bad default route on the amplet, or, 2) a router that does not know to send data over the vBNS. We have found, due to routing policy and configuration decisions at WUSTL, that point (2) is the reason. Routing policy at WUSTL sends all amp subnet traffic over the commodity network, regardless of its destination. WUSTL's reason for this policy is that their vBNS network endpoints use ATM adapters while our AMP boxes use 10/100 baseT.

Supporting Documentation: Examples of trace route data

Outgoing Path (Wed Dec 15 00:13:47 PST 1999)	**Return Path** (Wed Dec 15 00:53:19 PST 1999)
Medusa.sdsc.edu	Ncrc-onc.wustl.edu
Unresolvable	Wustl-fddi-starnet.wustl.edu
Gsr-atm0-0-24.dng.vbns.net	Fe0-0.starnet1.starnet.net
Unresolvable	Stl-core1-s010.cp.verio.net
Unresolvable	Vcp1.stl1.verio.net
AMP-WUSTL	Stl1.ord2.verio.net
	Ord2.nyc2.verio.net
	Nyc2-1.nyc1-1.verio.net
	Nyc1.phl02.verio.net
	Phl02-0.pne0-0.verio.net
	Pne0.cerfnet.verio.net
	Pos6-0-155M.phl-bb2.cerf.net
	Atm1-0-3.nyc-bb4.cerf.net
	Pos2-0-622M.nyc-bb8.cerf.net
	Pos5-0-622M.chi-bb4.cerf.net
	So4-0-0-622M.dfw-bb2.cerf.net
	Pos2-0-622M.lax-bb4.cerf.net
	Atm1-0-2-622M.san-bb6.cerf.net
	Pos10-0-0-155M.san-bb1.cerf.net
	Sdsc-gw.san-bb1.cerf.net
	Medusa-atm.sdsc.edu
	AMP-SDSC

Figure 7-2: Routing information for forward and reverse path connections between SDSC and WUSTL.

Case Study #8: Reverse Path Router Change

week of: 10 October 1999
primary site: San Diego Supercomputer Center (SDSC)
secondary site: Arizona State University (ASU)

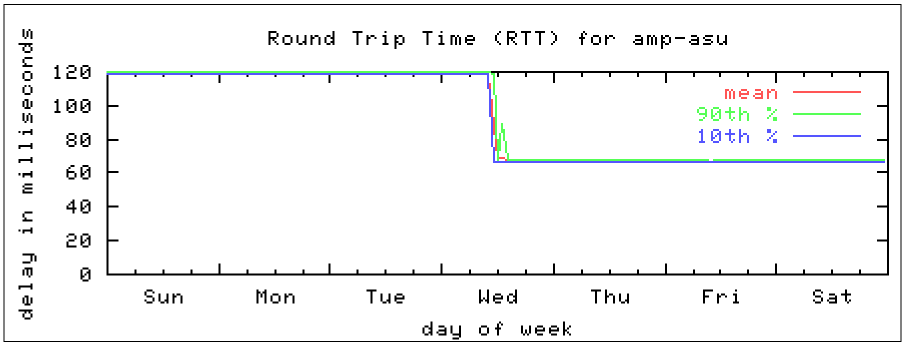


Figure 8-1: RTT delay for SDSC to ASU for week of 10-10-99.

Explanation: Figure 8-1 shows a sudden decrease in RTT from 120 ms to 75 ms sometime during mid-Wednesday. One could guess that this is due to a path change between SDSC and ASU. Upon inspection of Wednesday's RTT traceroute data (Figure 8-2a), however, we see no path change between SDSC and ASU. Upon inspection of the reverse path (ASU to SDSC), however, we see changes in the routing information (Figure 8-2b). What makes this case study unusual, is that the router change eliminated the vBNS and substituted it with a regional network which are traditionally somewhat slower. The elimination of extra hops and and exchange points (needed to get data from Abilene to the vBNS) became possible through this replacment, thus decreasing RTT.

Supporting Documentation: Examples of trace route data before and after the change

Forward Path Before Router Change (Wed Oct 13 00:21:26 PDT 1999)	Δ	**Forward Path After Router Change** (Thu Oct 14 00:23:43 PDT 1999)
Medusa.sdsc.edu		Medusa.sdsc.edu
Unresolvable		*Unresolvable*
Gsr-atm0-0-24.dng.vbns.net		Gsr-atm0-0-24.dng.vbns.net
Unresolvable		*Unresolvable*
Kscy-ipls.abilene.ucaid.edu		Kscy-ipls.abilene.ucaid.edu
Denv-kscy.abilene.ucaid.edu		Denv-kscy.abilene.ucaid.edu
Scrm-denv.abilene.ucaid.edu		Scrm-denv.abilene.ucaid.edu
Losa-scrm.abilene.ucaid.edu		Losa-scrm.abilene.ucaid.edu
Unresolvable		*Unresolvable*
Unresolvable		*Unresolvable*

Figure 8-2a: Routing information for forward paths before and after router change.

Reverse Path Before Router Change (Wed Oct 13 00:33:19 PDT 1999)	∆	**Reverse Path After Router Change** (Thu Oct 14 00:35:02 PDT 1999)
Main-gw.inre.asu.edu		Main-gw.inre.asu.edu
Unresolvable		*Unresolvable*
Scrm-losa.abilene.ucaid.edu	*	USC--abilene.ATM.calren2.net
Denv-scrm.abilene.ucaid.edu	*	UCSD--USC.POS.calren2.net
Kscy-denv.abilene.ucaid.edu	*	Sdsc1--UCSD.ATM.calren2.net
Ipls-kscy.abilene.ucaid.edu		AMP-SDSC
Unresolvable		
Cs-atm0-0-0-24.sdsc.vbns.net		
Unresolvable		
AMP-SDSC		

Figure 8-2b: Routing information for reverse paths before and after router change.

Case Study #9: Increased Router Hops Lower RTT

week of: 2 January 2000
primary site: West Virginia University (WVU)
secondary site: Arizona State University (ASU)

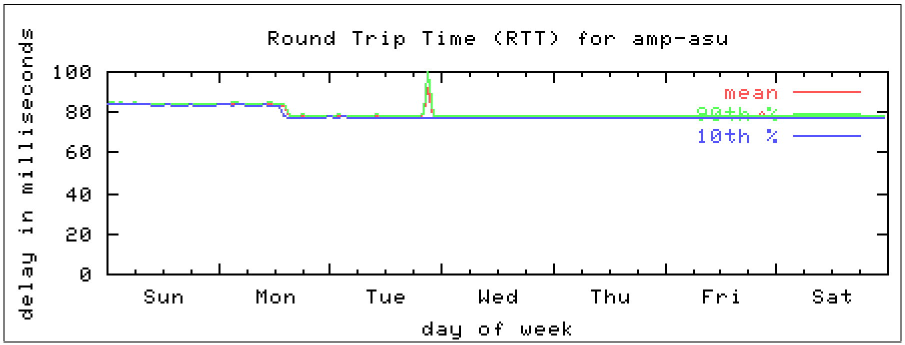


Figure 9-1: RTT delay for WVU to ASU for week of 01-02 -2000.

Explanation: This example shows a small RTT decrease on Monday from about 84 ms to 77 ms. While this in itself is not significant, what is interesting is that the number of hops in the path route data (Figure 9-2) increased as the RTT time decreased. This can be contrasted with Case Studies #6 and #8 in which RTT increased with an increase in the number of hops. The reason for this peculiarity stems from the removal of ATM based components and the replacement of POS based components. While more POS components show up in the traceroute data, there are actually fewer total hops. This case study was particularly challenging and it's solution was provided by vBNS experts.

Supporting Documentation: Examples of trace route data

Before Router Change (Mon Jan 3 00:31:53 PST 2000)	Δ	**After Router Change** (Mon Jan 3 14:56:09 PST 2000)
Unresolvable		*Unresolvable*
Unresolvable		*Unresolvable*
Vcisco-gp.psc.net		Vcisco-gp.psc.net
gsr-atm0-0-8.dng.vbns.net	*	jn1-at1-0-0-8.nor.vbns.net
Unresolvable	*	jn1-so5-0-0-0.dng.vbns.net
kscy-ipls.abilene.ucaid.edu	*	*Unresolvable*
denv-kscy.abilene.ucaid.edu	*	*Unresolvable*
scrm-denv.abilene.ucaid.edu		kscy-ipls.abilene.ucaid.edu
losa-scrm.abilene.ucaid.edu		denv-kscy.abilene.ucaid.edu
Unresolvable		scrm-denv.abilene.ucaid.edu
Unresolvable		losa-scrm.abilene.ucaid.edu
		Unresolvable
		Unresolvable

Figure 9-2: Routing information before and after router change.

Case Study #10: End Network Failure

week of: 26 December 1999
primary site: University of Tennessee Knoxville (UTK)
secondary site: The University of Oregon (U. of Oregon)

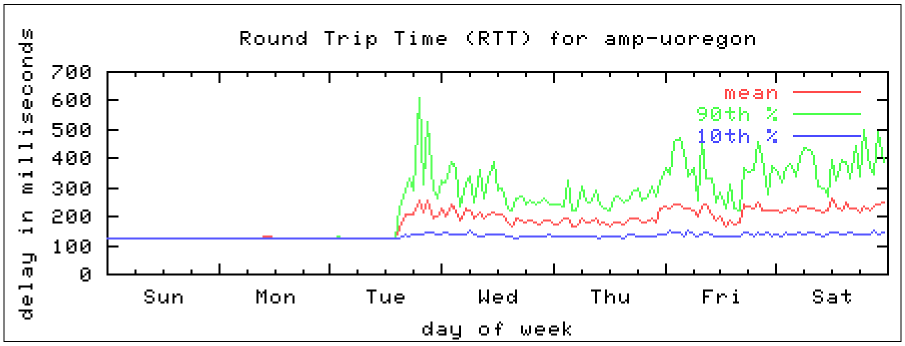


Figure 10-1a: RTT delay for UTK to U. of Oregon for weeks 12-12-99

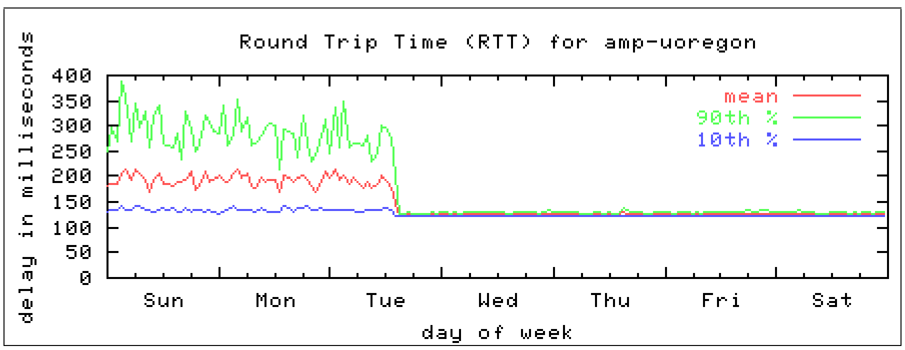


Figure 10-1b: RTT delay for UTK to U. of Oregon for weeks 12-26-99

Explanation: This is an interesting example in which RTT was representative of a normal HPC connection but became erratic for a period of two weeks. This erratic behavior began on Tuesday, December 14th (Figure 10-1a) and ended sometime on Tuesday, December 28th (Figure 10-1b). In order to understand the high variance in RTT we examined the trace route data from UTK to University of Oregon (Figure 10-4ab) in the hopes of finding a routing change or an other such hardware related event. To our surprise, there was no such change. We did notice, however, that the last hop took a disproportionately high amount of time, compared to other hops and varied radically from ping request to ping request. With this information, we proceeded to find other examples of this anomaly. The weekly graphs (Figure 10-2a and Figure 10-2b) from SDSC to University of Oregon gave us this data. Finding that both data sets [UTK and U of Oregon] and [U of Oregon and SDSC] reflected the same problem we were able to verify that the erratic RTT problem existed at University of Oregon. In other words, it was not related to the SDSC or UTK hosts. While the exact problem, which led to an increase in RTT, is not clearly identified we feel that it can be attributed to one of three pieces of hardware: either the 0car-0gw.oregon-gigapop.net router, the nlanr-amp.oregon-gigapop.net box or some level 1 or level 2 failure in between these pieces of hardware (such as a bad switch). The problem was fixed sometime on Tuesday, December 28th.

Supporting Documentation: Further examples of RTT graphs, examples of trace route data.

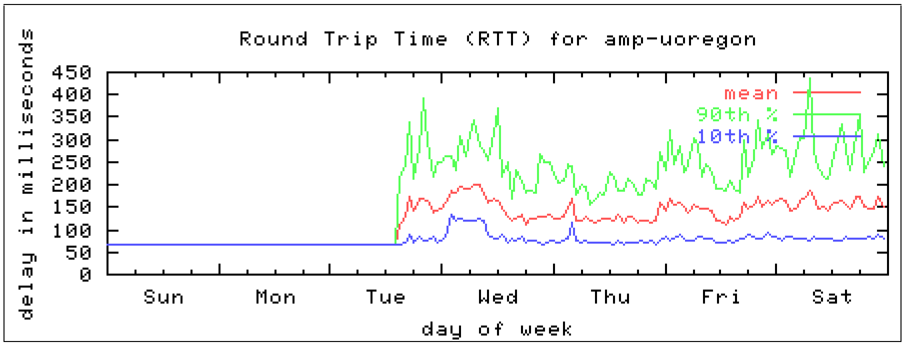


Figure 10-2a: RTT delay for SDSC to U of Oregon for the week of 12-12-99

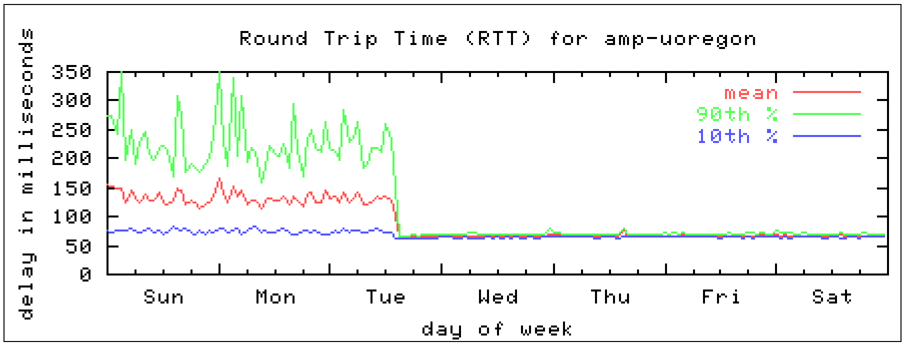


Figure 10-2b: RTT delay for SDSC to U. of Oregon for week of 12-26-99

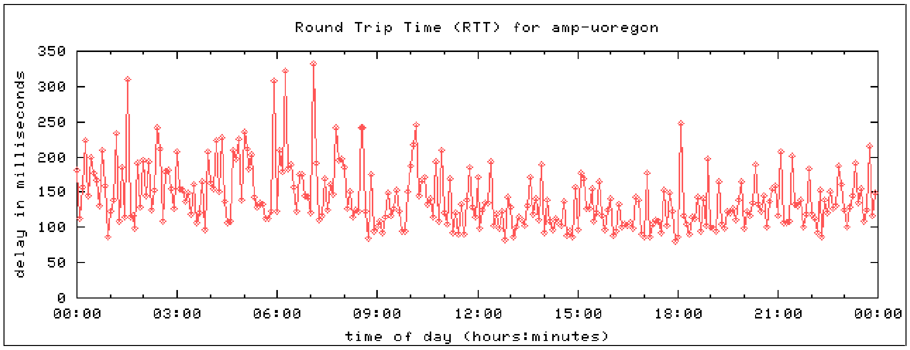


Figure 10-3: RTT delay for SDSC to U. of Oregon for 12-20-99

Trace Route Data: UTK to U. of Oregon (Tue Dec 28 00:09:40 PST 1999)	**Time 1 (ms)**	**Time 2 (ms)**	**Time 3 (ms)**
R5HM01V277.NS.UTK.EDU	0.755	0.644	0.631
Unresolvable	1.679	1.405	1.466
UTK-GATECH.NS.UTK.EDU	32.950	32.952	32.940
ast-dng.vbns.sox.net	34.450	33.972	34.462
jn1-at1-0-0-0.ast.vbns.net	35.012	35.416	34.975
jn1-so5-0-0-0.wae.vbns.net	49.491	49.460	49.330
Abilene-vbns.abilene.ucaid.edu	54.557	54.673	54.732
clev-nycm.abilene.ucaid.edu	65.940	66.938	66.923
ipls-clev.abilene.ucaid.edu	72.936	73.027	72.866
kscy-ipls.abilene.ucaid.edu	82.493	81.974	81.490
denv-kscy.abilene.ucaid.edu	92.968	92.444	92.406
ogig-den.oregon-gigapop.net	124.488	124.379	124.000
0car-0gw.oregon-gigapop.net	123.892	123.506	124.507
AMP-Oregon	669.401	1052.441	268.516

Figure 10-4a: Trace route data from UTK to University of Oregon.

Trace Route Data: SDSC to U of Oregon (Tue Dec 28 00:09:22 PST 1999)	**Time 1 (ms)**	**Time 2 (ms)**	**Time 3 (ms)**
medusa.sdsc.edu	0.529	0.395	0.322
Unresolvable	4.681	4.946	4.946
gsr-atm0-0-24.dng.vbns.net	52.173	51.482	51.950
Unresolvable	54.648	54.774	54.481
kscy-ipls.abilene.ucaid.edu	54.764	54.635	55.165
denv-kscy.abilene.ucaid.edu	55.292	55.467	55.493
ogig-den.oregon-gigapop.net	65.434	64.930	64.807
0car-0gw.oregon-gigapop.net	65.179	64.865	65.939
AMP-Oregon	112.823	820.868	109.104

Figure 10-4b: Trace route data from SDSC to University of Oregon.

Related Projects

There are a couple of related projects which have attempted to do the same sort of active analysis that we have done. Some of them, Surveyor, for example have attempted to measure one-way delays instead of round trip times. They feel that this will give them more useful information. However, all design decisions come at a cost. It would be wise to look at the platforms on which the tests are running and the size of their mesh. However, if the measurements are done properly, then their results should be as useful as ours. A listing of these related projects is included in the references section.[8]

Acknowledgements

This work was sponsored by the National Science Foundation under Cooperative Agreement No. ANI-9807479. We thank Maureen C. Curran for her editorial assistance.

References

1. Todd Hansen, tshansen@nlanr.net, San Diego Supercomputer Center, UCSD, 9500 Gilman Dr., La Jolla, CA, 92093-0505, USA
2. Jose Otero, jotero@nlanr.net, San Diego Supercomputer Center, UCSD, 9500 Gilman Dr., La Jolla, CA, 92093-0505, USA
3. Hans-Werner Braun, hwb@nlanr.net, San Diego Supercomputer Center, UCSD, 9500 Gilman Dr., La Jolla, CA, 92093-0505, USA
4. Tony McGregor, tonym@nlanr.net, The University of Waikato, Private Bag 3107, Hamilton, New Zealand
5. National Laboratory for Applied Network Research (NLANR), http://www.nlanr.net/
6. NLANR's Measurement and Network Analysis Team, http://moat.nlanr.net/
7. NLANR's Active Measurement Project (AMP) http://moat.nlanr.net/AMP/
8. A listing of related projects:
    - DREN AMP http://www.sd.wareonearth.com/amp (AMP peer network)
    - Internet End-to-End Performance Monitoring at SLAC http://www-iepm.slac.stanford.edu/
    - National Internet Measurement Infrastructure http://www.ncne.nlanr.net/nimi/
    - RIPE's Test Traffic Measurements http://www.ripe.net/test-traffic/index.html
    - Surveyor http://www.advanced.org/surveyor/
    - CAIDA's Skitter Project http://www.caida.org/Tools/Skitter/
9. MCI's vBNS Network http://www.vbns.net/
10. Internet2's Abilene Network http://www.internet2.edu/abilene/
11. AMP Introduction http://moat.nlanr.net/AMP/active/intro.html
12. A. J. McGregor, H-W Braun, J.A. Brown. The NLANR Network Analysis Infrastructure. Accepted for publication by *IEEE Communications Magazine*.