

Network Monitoring Tutorial

Tutorial on Internet Monitoring & PingER at SLAC

Translation into: [Russian](#), [Belorussian](#) (provided by [Websitehostingrating](#)), [Indonesian](#) (provided by [ChameleonJohn](#)), [Irish](#) (provided by [Travel Ticker](#)), [Hindi](#) (provided by [Dealsdaddy](#)), [Italian](#) (provided by [Headphones Encyclopedia](#)).
[Les Cottrell](#), Warren Matthews and Connie Logg, Created January 1996; last Update: December 1st, 2014.

[IEPM home page](#) | [PingER](#) | [PingER Detail Reports](#) | [PingER Site Map](#) | **Tutorial**

Work partially funded by a DOE/MICS Field Work Grant for [Internet End-to-end Performance Monitoring \(IEPM\)](#).

Web Page Contents

- [Introduction](#)
- [Mechanism](#) ICMP; Ping; [PingER](#); Network impact; Extent of measurements
- [Data Gathering](#)
- [Measurement Method](#)
- [Gotchas](#) - Calibration; High statistics ping measurements; Host requirements; ICMP Rate Limiting, Pathologies , SYN/ACK.
- [Validation](#) - HTTP GET vs Min ping response; Dynamic web server selection; Firehunter case study; [Internet quality of service assessment](#) ; [PingER and NetPerf](#); TCP throughput vs. Loss.
- [What do we Measure](#) - Loss; RTT; Visual Ping, distances, Jitter; Baselines & alerts.
- [Non Ping Based Tools](#) - Surveyor; [Correlations with ping](#); Modelling TCP throughput; [PingER and NetPerf](#); Throughput vs. RTT & Loss
- [Calculating Mean Opinion Scores \(MOS\)](#)
- [Calculating Network Contribution to Transaction Times \(NCTT\)](#)
- [Deriving TCP Throughput From Ping Measurements](#), Normalizing the Derived Throughput
- [Directness of connection](#)
- [Accessing the PingER data](#) - Raw & Summary data.

Pinging of Collaborators' Nodes

- [Data Gathering](#) - Big vs. small packets, timing of ping measurements.
- [Data Analysis & Presentation](#)
 - [Daily Plots](#)
 - [3D Plots of Host vs. Response vs. Time of Day](#)
 - [Last 180 Days](#)
 - [Monthly Ping Response & Loss Averages going back for years](#) - Long term trends; Impact of adding bandwidth & vacations.
 - [Quiescent Network Frequency](#) - Definition; Example.
 - [Jitter](#) - ITU proposal for measuring jitter; [PingER](#) jitter measurement methods; Non-Gaussian behavior; Comparison of 2 [PingER](#) jitter methods.
 - [Service Predictability](#) - Definition; Results.
 - [Unpredictability](#) - Definition.
 - [Reachability](#) - Definition; [Error-free seconds](#); MTBF; MTTR; [Outages](#); [Examples](#).
 - [Out of Order packets](#)
 - [Duplicate Packets](#)
 - [Combining the metrics](#) - Example and interpretation of a chart with most metrics on it.
 - [Quality](#)
 - [Delay](#) - Importance of delay.
 - [Loss](#) - Quality thresholds; ITU definitions; Relation to VoIP; MCI & Internet Weather definitions; Loss distributions.
 - [Jitter](#) - ITU definitions; Definition of ping jitter; Comparison of one & two way jitter.
 - [Utilization](#) Utilization and loss, network health and RTT variation.
 - [Throughput](#)
 - [Reachability](#) SLAs and availability
 - [Directivity](#)
 - [Grouping](#) Possible pair groups; Group distributions; Beacon sites.
- [One Way Measurements](#) Projects: Surveyor, Comparison of [PingER](#) & Surveyor, RIPE, NIMI, Waikato and sting; Relation to two way. Relevant IETF RFCs.
- [Traceroute](#)

[Internet NREN performance from U.S. to the world at the turn of Century.](#)
[Internet NREN performance from U.S. to world at the end of 2000.](#)

[Comparison of some Internet Active End-to-end Performance Measurement projects](#)

- [Comparison of Ping & PingER vs. Surveyor Results](#)
- [Comparison of Surveyor & RIPE Results](#)

[Comparison of Passive and Active Monitoring](#)

[Comparison of Performance between ESnet, .edu and XIWT sites](#)

[Impact of Routing on End-to-end Internet Performance](#)

[Glossary](#)

[Further Information](#)

[Index of Links on this Page](#)

Introduction

To provide a better expectation of the network performance between sites that SLAC collaborates with, in May 1996 the [PingER project](#) (initiated in January 1995) monitored about [100 hosts](#) worldwide from SLAC. Since 2000, the emphasis is more on measuring the Digital Divide. Nowadays (April 2007) there are over 35 [Monitoring Sites](#), over 600 [remote sites](#) being monitored in over 150 countries (containing over 99% of the worlds Internet connected population) and over 8000 monitor-site remote-site pairs included. More details on the deployment of PingER can be found in [PingER Deployment](#) and there is a [map](#) of the sites.

Mechanism

Measurement method

The set of remote hosts to ping is provided by a file called `pinger.xml` (for more on this see the [pinger2.pl documentation](#)). This file consists of two parts: **Beacon hosts** which are automatically pulled daily from SLAC and are monitored by all MPs; other hosts which are of particular interest to the administrator of the MP. The Beacon hosts (and the particular hosts monitored by the SLAC MP) are kept in an Oracle database containing their name, IP address, site, nickname, location, contact etc. The Beacon list (and list of particular hosts for SLAC) and a copy of the database, in a format to simplify Perl access for analysis scripts, are automatically generated from the database on a daily basis.

- The remote monitoring sites. These simply provide a passive remote-host with the appropriate [requirements](#).
- The monitoring site. The [PingER monitoring tools](#) needs to be installed and configured on a host at each of these sites. Also the ping data collected needs to be made available to the archive hosts via the HyperText Transport Protocol (HTTP) (i.e. there must be a Web server to provide the data on demand via the Web). There are also PingER tools to enable a monitoring site to be able to provide short term analysis and reports on the data it has in its local cache.
- The archive and analysis sites. There must be at least one each of these for each PingER project. The archive and analysis sites maybe located at a single site, or even a single host or they may be separated. The PingER project has two such sites, one at [NUST](#) in Islamabad, Pakistan, the other at [SLAC](#). Bot sites are archive and analysis sites. They complement each other by providing redundancy. The XIWT project had its archive /monitor site at the [CNRI](#).

The PingER architecture is illustrated below:



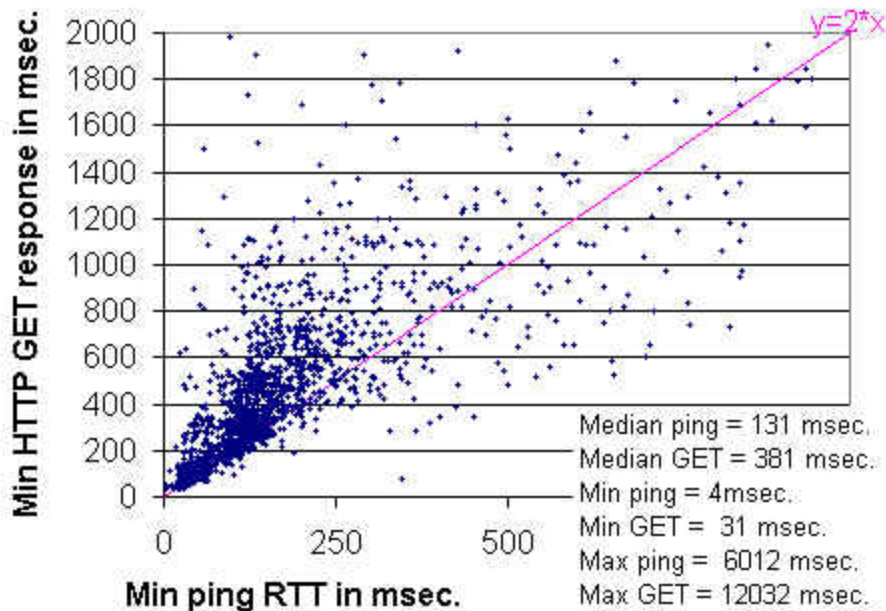
We have also observed various pathologies with various remote sites when using ping. These are documented in [PingER Measurement Pathologies](#).

The calibration and context in which the round trip metrics are measured are documented in [PingER Calibration and Context](#), and some examples of how ping results look when taken with high statistics, and how they relate to routing, can be found in [High statistics ping results](#).

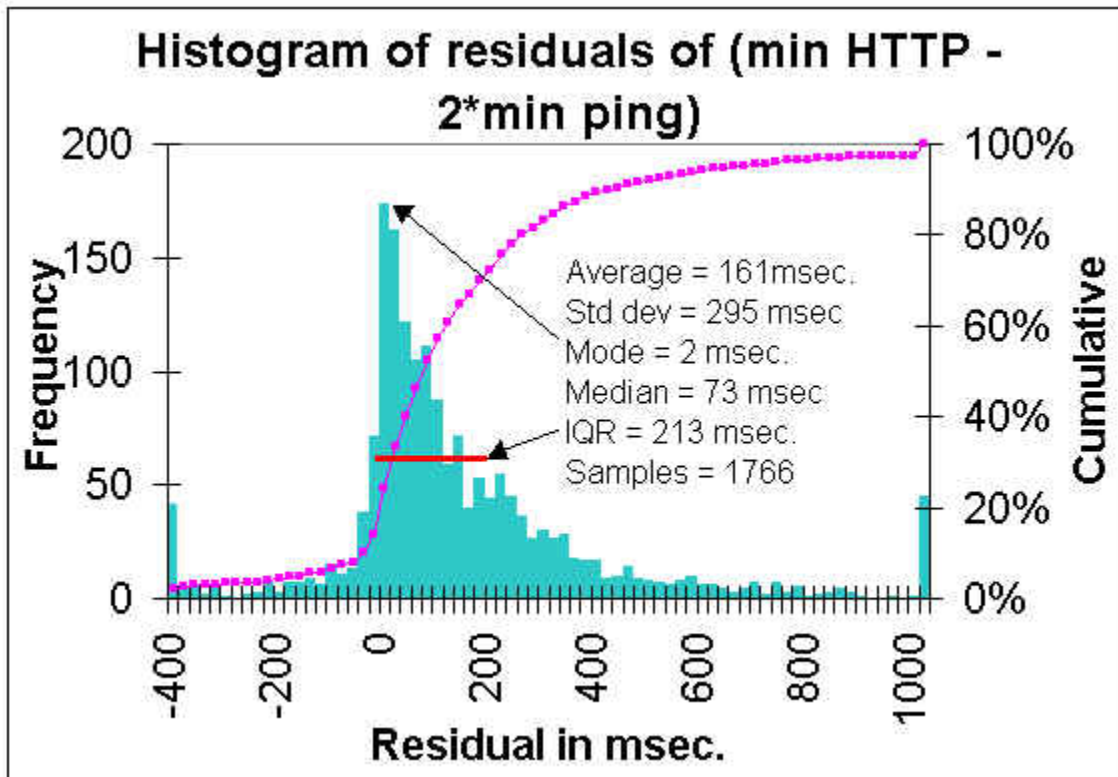
Validation

We have validated the use of ping by demonstrating that measurements made with it correlate with application response. The correlation between the lower bounds of Web and ping responses is seen in the figure below. The measurements were made on December 18th, 1996, from SLAC to about 1760 sites identified in the [NLANR caches](#). For more details see [Effects of Internet Performance on Web Response Times](#), by Les Cottrell and John Halperin, unpublished, January 1997.

The remarkably clear lower boundary seen around $y = 2x$ is not surprising since: a slope of 2 corresponds to HTTP GETs that take twice the ping time; the minimum ping time is approximately the round trip time; and a minimal TCP transaction involves two round trips, one round trip to exchange the second to send the request and receive the response. The connection termination is done asynchronously and so does not show up in the timing.



The lower boundary can also be visualized by displaying the distribution of residuals between the measurements and the line $y = 2x$ (where y = HTTP GET response time and x = Minimum ping response time). Such a distribution is shown below. The steep increase in the frequency of measurements as one approaches zero residual value ($y=2x$) is apparent. The Inter Quartile Range (IQR), the residual range between where 25% and 75% of the measurements fall, is about 220 msec, and is indicated on the plot by the red line.



An alternate way of demonstrating that ping is related to Web performance is to show that ping can be used to predict which of a set of replicated Web servers to get a Web page from. For more on this see [Dynamic Server Selection in the Internet](#), by Mark E. Crovella and Robert L. Carter.

The Firehunter Case Study of the Whitehouse Web Server showed that though ping response does not track abnormal Web performance well, in this case ping packet loss did a much better job.

[The Internet quality of service assessment](#) by Christian Huitema, provides measurements of the various components that contribute to web response. These components include: RTT, transmission speed, DNS delay, connection delay, server delay, transmission delay. It shows that the delays between sending the GET URL command and the reception of the first byte of the response is an estimate of server delay ("in many servers, though not necessarily all, this delay corresponds to the time needed to schedule the page requests, prepare the page in memory, and start sending data") and represents between 30 and 40% of the duration of the average transaction. To ease that, you probably need more powerful servers. Getting faster connections would obviously help the other 60% of the delay.

Also see the section below on [Non Ping based tools](#) for some correlations of throughput with round trip time and packet loss.

What do we Measure

We use ping to measure the response time (round trip time in milli-seconds (ms)), the packet loss percentages, the variability of the response time both short term (time scale of seconds) and longer, and the lack of reachability, i.e. no response for a succession of pings. For a discussion and definition of reachability and availability see [Internet Performance: Data Analysis and Visualization](#) a White Paper by the [XIWT](#). We also record information on out of order packets and duplicate packets.

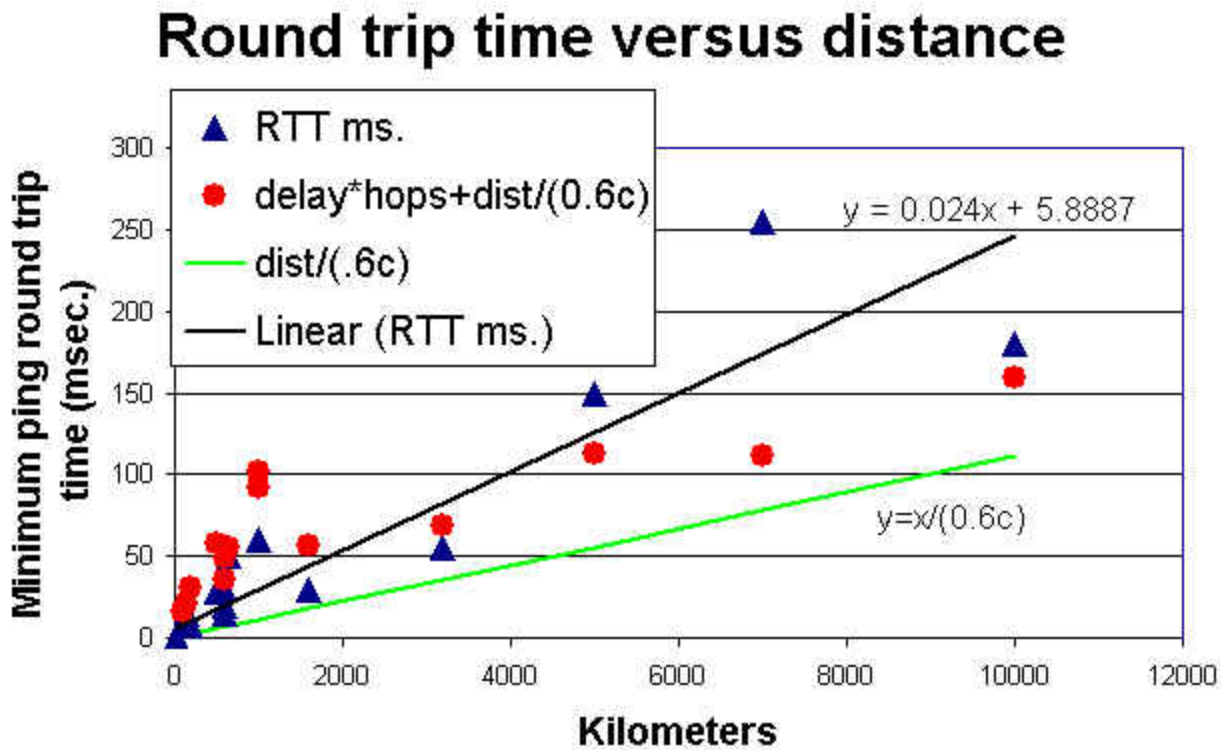
With measured data we are able to create long term baselines for expectations on means/medians and variability for [response time](#), [throughput](#), and [packet loss](#). With these baselines in place we can set expectation, provide planning information, make extrapolations and look for exceptions (e.g. today's response time is more than 3 standard deviations greater than the average for the last 50 working days) and raise alerts.

Loss

The loss is a good measure of the *quality* of the link (in terms of its packet loss rates) for many TCP based applications. Loss is typically caused by congestion which in turn causes queues (e.g. in routers) to fill and packets to be dropped. Loss may also be caused by the network delivering an imperfect copy of the packet. This is usually caused by bit errors in the links or in network devices. Paxson (see [End-to-end Packet Dynamics](#)) from measurements made in 1994 and 1995 concluded that most corruption errors came from T1 links, and the typical rate was 1 in 5000 packets. This corresponds to a bit error rate for a 300byte packet average of about 1 in 12 million bits. IP has a 16 bit checksum, so the probability of not detecting an error in a corrupted packet is 1 in 65536, or 1 in about 300 million packets. A more recent study on [When the CRC and TCP checksum disagree](#) published in August 2000, indicates that traces of Internet packets over the last two years show that 1 in 30,000 packets fails the TCP checksum, even on links where link-level CRCs should catch all but 1 in 4 billion errors. These TCP checksum errors are higher level (e.g. they can be caused by bus errors in network devices or computers, or by TCP stack errors) than the link level errors that should be caught by the CRC checks.

RTT

The response time or Round Trip Time (RTT) when plotted against packet size can give an idea of *ping data rate* (kilo Bytes/sec (KB/s)) This becomes increasingly difficult as one goes to high performance links, since the packet range is relatively small (typically < 1500bytes), and the timing resolution is limited. The RTT is related to the distance between the sites plus the delay at each hop along the path between the sites. The distance effect can be roughly characterized by the speed of light in fiber, and is roughly given by $distance / (0.6 * c)$ where c is the velocity of light (the ITU in document G.144, table A.1 recommends a multiplier of 0.005 msec/km, or $0.66c$). Putting this together with the hop delays, the RTT R is roughly given by: $R = 2 * (distance / (0.6 * c) + hops * delay)$ where the factor of 2 is since we are measuring the out and back times for the round-trip. This is illustrated in the figure below, that shows the measured ping response as a function of distance for between 16 pairs of sites located in the U.S., Europe and Japan (Bologna-Florence, Geneva-Lyon, Chicago-U of Notre Dame, Tokyo-Osaka, Hamburg-Dresden, Bologna-Lyon, Geneva-Mainz, Pittsburg-Cincinnati, Geneva-Copenhagen, Chicago-Austin, Geneva-Lund, Chicago-San Francisco, Chicago-Hamburg, San Francisco-Tokyo, San Francisco-Geneva & Geneva-Osaka). The blue triangles are for the measured RTT (in milliseconds), the black line is a fit to the data, the green line is for $y = x(distance) / (0.6 * c)$, and the red dots fold in the hop delays with a delay/hop of about 2.25ms for each direction (i.e. the red dots are the theoretical RTT fit). We used the [How far is it?](#) web page to obtain the "as the crow flies" distances between major points along each route. A more recent measurement made by Mark Spiller in March 2001, to about 10 universities from UC Berkeley measured router delays of 500-700 usec with a few spikes in the 800-900 usec range.



Route length (R_{km}) may be used in the place of *distance* for some Frame Transfer Delay (FTD) performance objectives. If D_{km} is the air-route distance between the boundaries, then the route length is calculated as follows (this is the same calculation as found in the ITU document G.826).

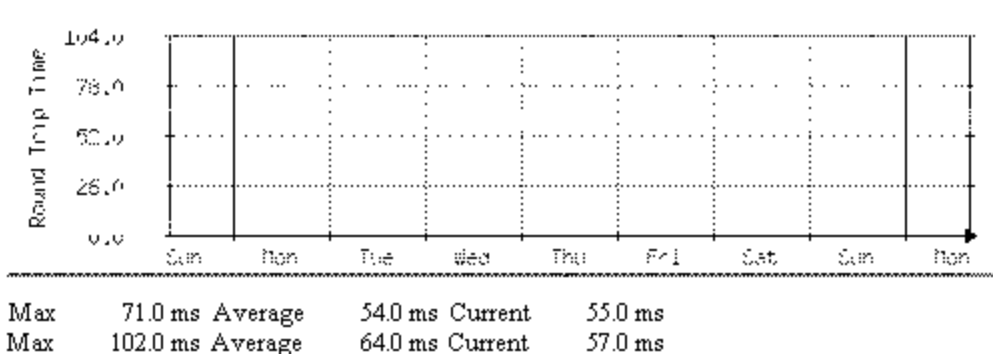
- if $D_{km} < 1000$ km, then $R_{km} = 1.5 * D_{km}$
- if $1000 \text{ km} \leq D_{km} \leq 1200$ km, then $R_{km} = 1500$ km
- if $D_{km} > 1200$ km, then $R_{km} = 1.25 * D_{km}$

This rule does not apply if there is a satellite in the route. If a satellite is present in any portion of the route, that portion is allocated a fixed FTD of 320 msec. The value of 320 msec takes into account factors such as low earth station viewing angles, and forward error correcting encoding. Most portions that contain a satellite are not expected to exceed 290 msec of delay. If it is a geostationary satellite then the lower limit on geosynchronous orbit is between 22,000 and 23,000 miles, the speed of light is ~ 186,000 miles, figure up and back is 45,000, and round-trip is 90,000 miles, so we get 500ms right there.

This elongated minimum RTT caused by geo-stationary satellites provides a useful signature to identify that the route between monitor and target includes a geo-stationary satellite. An example can be seen in Figure 5 of [2011 - 2012 Report of the ICFA-SCIC Monitoring Working Group](#).

The delays at each hop are a function of 3 major components: the speed of the router, the interface clocking rates and the queuing in the router. The former two are constant over short (few days) periods of time. Thus the minimum RTTs give a measure of the $\text{distance} / (0.6 * c) + \text{hops} * ((\text{interface speed} / \text{packet size}) + \text{minimum router forwarding time})$. This number should be a linear function of packet size. The router queuing effects, on the other hand, are dependent on more random queuing processes and cross-traffic and so are more variable. This is illustrated in the MRTG plot below that shows the very stable minimum RTT (green area) and the more random maximum RTTs (blue lines) measured from SLAC to University of Wisconsin from Sunday 25th February 2001, to Monday April 5th, 2001. The little blip in the RTT around mid-day Tuesday is probably caused by a route change.

'Weekly' Graph (30 Minute Average)



Non Ping Based Tools

SLAC was also a Surveyor site. Surveyor made one way delay measurements (not using ICMP), using Global Positioning System (GPS) devices to synchronize time, and dedicated monitoring/remote hosts. We [compared the PingER and Surveyor data](#) to compare and contrast the two methods and verify the validity of ICMP echo. One concern raised with ICMP echo is the possibility of Internet Service Providers (ISPs) [rate limiting ICMP echo](#) and thus giving rise to invalid packet loss measurements, for more on this see the section on [Gotchas](#) above.

We also use more complex tools such as [FTP](#) (to measure bulk transfer rates) and [traceroute](#) (to measure paths and number of hops). However, besides being more difficult to set up and automate, FTP is more intrusive on the network and more dependent on end node loading. Thus we use FTP mainly in a manual mode and to get an idea of how well the ping tests work (e.g. [Correlations between FTP and Ping](#) and [Correlations between FTP throughput, Hops & Packet Loss](#)). We have also compared [PingER predictions of throughput with NetPerf measurements](#). Another way to correlate throughput measurements with packet loss is by [Modeling TCP Throughput](#).

Calculating the Mean Opinion Score (MOS)

The telecommunications industry uses the Mean Opinion Score (MOS) as a voice quality metric. The values of the MOS are: 1=bad; 2=poor; 3=fair; 4=good; 5=excellent. A typical range for Voice over IP is 3.5 to 4.2 (see [VoIPtroubleshooter.com](#)). In reality, even a perfect connection is impacted by the compression algorithms of the codec, so the highest score most codecs can achieve is in the 4.2 to 4.4 range. For G.711 the best is 4.4 (or an R Factor (see ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning.") of 94.3) and for G.729 which performs significant compression it is 4.1 (or an R factor of 84.3).

There are three factors that significantly impact call quality: latency, packet loss, jitter. Other factors include the codec type, the phone (analog vs. digital), the PBX etc.) We show how we calculate jitter [later](#) in this tutorial. Most tool-based solutions calculate what is called an "R" value and then apply a formula to convert that to an MOS score. We do the same. This R to MOS calculation is relatively standard (see for example [ITU - Telecommunication Standardization Sector Temporary Document XX-E WP 2/12](#) for a new method). The R value score is from 0 to 100, where a higher number is better. Typical R to MOS values are: R=90-100 => MOS=4.3-5.0 (very satisfied), R=80-90=>MOS=4.0-4.3 (satisfied), R=70-80=>MOS=3.6-4.0 (some dissatisfaction), R=60-70=>MOS=3.1-3.6 (more dissatisfaction), R=50-60=>MOS=2.6-3.1 (Most dissatisfaction), R=0-50=>MOS=1.0-2.6 (not recommended). To convert latency, loss, jitter to MOS we follow [Nessoft's method](#). They use (in pseudo code):

```

#Take the average round trip latency (in milliseconds), add
#round trip jitter, but double the impact to latency
#then add 10 for protocol latencies (in milliseconds).
EffectiveLatency = ( AverageLatency + Jitter * 2 + 10 )
#Implement a basic curve - deduct 4 for the R value at 160ms of latency
#(round trip). Anything over that gets a much more aggressive deduction.

if EffectiveLatency < 160 then
    R = 93.2 - (EffectiveLatency / 40)
else
    R = 93.2 - (EffectiveLatency - 120) / 10

#Now, let us deduct 2.5 R values per percentage of packet loss (i.e. a
#loss of 5% will be entered as 5).
R = R - (PacketLoss * 2.5)

#Convert the R into an MOS value.(this is a known formula)
if R < 0 then
    MOS = 1
else
    MOS = 1 + (0.035) * R + (.000007) * R * (R-60) * (100-R)

```

Also see the following for some measurement tools and/or explanations:

- [phonepower](#),
- [FitSmallBusiness](#),
- [Packetizer VoIP Bandwidth Calculator](#),

Calculating Network Contribution to Transaction Times

The ITU have come up with a method to calculate the network contribution to transaction time in [ITU-T Rec.G1040 "Network contribution to transaction time"](#). The contribution depends on the *RTT*, loss probability (*p*), the Retransmission Time Out (*RTO*) and the number of round trips involved (*n*) in a transaction. The Network Contribution to Transaction Time (*NCTT*) is given as:

$$\text{Average}(NCTT) = (n * RTT) + (p * n * RTO)$$

Typical values for *n* are 8, for *RTO* we take 2.5 seconds, we take the *RTT* and loss probability (*p*) from the PingER measurements.

Deriving TCP throughput from ping measurements

[The macroscopic behavior of the TCP congestion avoidance algorithm](#) by Mathis, Semke, Mahdavi & Ott in Computer Communication Review, 27(3), July 1997, provides a short and useful formula for the upper bound on the transfer rate: $\text{Rate} < (MSS/RTT) * (1 / \sqrt{p})$ where:

Rate: is the TCP transfer rate

MSS: is the maximum segment size (fixed for each Internet path, typically 1460 bytes)

RTT: is the round trip time (as measured by TCP)

p: is the packet loss rate.

Strictly speaking the losses are the TCP losses which are not necessarily identical to the ping losses (e.g. standard TCP provokes losses as part of its congestion estimation). Also the ping RTTs are different from the way TCP estimates the RTT (see for example [Improving Round-Trip Time Estimates in Reliable Transport Protocol](#).) However, especially for lower performance links it is a reasonable estimator.

One may use the [Bit Error Rate \(BER\)](#) to estimate the losses. Typical values are BER = 10⁻⁹ (i.e. packet loss probability of ~0.001% for 100byte packets) for an electrical link, and 10⁻¹² for an optical link (see [CS244a: An Introduction to Computer Networks - Stanford University](#), and [10 Gigabit Ethernet and the XAUI interface](#)).

If you can't empirically measure *p*, start with the expectation bit error rate (BER) for 1000BaseT, 10⁻¹⁰ for the probability of loss. See [Error performance objective for 400GbE](#).

An improved form of the above equation can be found in: [Modeling TCP throughput: A simple model and its empirical validation](#) by J. Padhye, V. Firoiu, D. Townsley and J. Kurose, in *Proc. SIGCOMM Symp. Communications Architectures and Protocols* Aug. 1998, pp. 304-314.

$$\text{Rate} = \min(W_{\max}/RTT, 1/((RTT/\sqrt{2*b*p/3}) + \min(1, 3*\sqrt{3*b*p/8})*(1+32*p)))$$

where:

W_{max}: is the maximum congestion window size.

b: is the number of packets acknowledged by a delayed ACK. Many TCP receiver implementations send one cumulative ACK for two consecutive packets received (see W. Stevens. *TCP/IP Illustrated*, Vol. 1 The Protocols. Addison-Wesley, 1994), so *b* is typically 2.

The behavior of throughput as a function of loss and *RTT* can be seen by looking at [Throughput versus RTT and loss](#). We have used the formula Mathis to compare [PingER](#) and [NetPerf](#) measurements of throughput.

Normalizing the Derived Throughput

To reduce the effect of the 1/RTT in the Mathis formula for derived throughput, we normalize the throughputs by using

$$\text{norm_throughput} = \text{throughput} * \min_RTT(\text{remote region}) / \min_rtt(\text{monitoring_region})$$

Directness of connection

This is a metric to identify the directness of the connection between 2 nodes at known locations. Directness values close to one mean the path between the hosts follows a roughly great circle route. Values much smaller than 1 mean the path is very indirect. The derivation of directness coefficient *Directivity* based on the minimu RTT is given [here](#).

RTD=Round Trip Distance,

$RTD[km] = Directivity * \min_RTT[msec] * 200 [km/msec]$

Directivity allows for delays in network equipment and indirectness of the actual route.

$D = 1$ way distance

$Directivity = D(km) / (\min_RTT[msec] * 100 [km/msec])$

- $\text{Max}(Directivity) = 1$ = direct (great circle) route and no network delays
- Directivity obtained from minimum RTT is typically ~ 0.45
- Low values typically mean very indirect route, or satellite or slow connection (e.g. wireless)
- $Directivity > 1$ probably identifies bad coordinates for hosts.

In the case of the [VTrace visual traceroute](#), the difference between hop distance and end-to-end distance can provide an estimate of the *Directivity*. The end-to-end distance is the great circle path distance between source and destination, where as the total hop distance is the sum of great circle distances between all consecutive hops. In this case we estimate the *Directivity* as the $(\text{end_to_end_distance} / \text{total_hop_distance})$.

Access to data

The raw ping data is publicly accessible, see [Accessing the PingER data](#) for how to get the data and the format. The summarized data is also available from the web in Excel tab-separated-value (.tsv) format from the [PingER detail reports](#).

Pinging of Collaborators' Nodes

Data Analysis & Presentation Daily Plots The ping response times are plotted for each half hour for each node, for example see the Smokeping plot if RTT and loss. This is mainly used for trouble shooting (e.g. see if it got dramatically worse in the last few hours). 3D Plots of Node vs Response vs Time of Day By plotting a 3D plot of node versus time versus response we can look for correlations of several nodes having poor performance or being unreachable at the same time (maybe due to a common cause), or a given node having poor response or being unreachable for an extended time. To the left is an example showing several hosts (in black) all being unreachable around 12 noon. Last 180 Days plots: Long term graphs showing the response time, packet loss and unreachability for the last 180 days can also indicate whether a service is getting worse (or better). Monthly Ping Response & Loss Averages going back for years: Tables of monthly medians of the prime time (7am- 7pm weekday) 1000 byte ping response time and 100 byte ping packet loss allow us to view the data going back for longer periods. This tabular data can be exported to Excel and charts made of the long term ping packet loss performance. Quiescent Network Frequency When we get a zero packet loss sample (a sample refers to a set of n pings), we refer to the network as being quiescent (or non-busy). We can then measure the percentage frequency of how often the network was found to be quiescent. A high percentage is an indication of a good (quiescent or non-heavily loaded) network. For example a network that is busy 8 work hours per week day, and quiescent at other times would have a quiescent percentage of about $75\% \sim (\text{total_hours/week} - 5 \text{ weekdays/week} * 8 \text{ hours/day}) / (\text{total_hours/week})$. This way of representing the loss is similar in intent to the phone metric of error-free seconds. The Quiescent Network Frequency table shows the percentage (frequency) of the samples (where a sample is a set of 10 100 byte pings) that measured zero packet loss. The samples included in each percentage reported are all the samples for each site for each month (i.e. of the order of 30 days * 48 (30 min periods) or about 1440 samples) per site/month. Jitter, see also Jitter, The short term variability or "jitter" of the response time is very important for real-time applications such as telephony. Web browsing and mail are fairly resistant to jitter, but any kind of streaming media (voice, video, music) is quite susceptible to jitter. Jitter is a symptom that there is congestion, or not enough bandwidth to handle the traffic. The jitter specifies the length of the VoIP codec de-jitter buffer to prevent over- or under-flow. An objective could be to specify that say 95% of packet delay variations should be within the interval [-30msec, +30msec]. One method requires injecting packets at regular intervals into the network and measuring the variability in the arrival time. The IETF has IP Packet Delay Variation Metric for IP Performance Metrics (IPPM) (see also RTP: A Transport Protocol for Real-Time Applications, RFC 2679 and RFC 5481). We measure the instantaneous variability or "jitter" in two ways. Let the i-th measurement of the round trip time (RTT) be R_i , then we take the "jitter" as being the Inter Quartile Range (IQR) of the frequency distribution of R . See the SLAC \Leftrightarrow CERN round trip delay for an example of such a distribution. In the second method we extend the IETF draft on Instantaneous Packet Delay Variation Metric for IPPM, which is a one-way metric, to two-way pings. We take the IQR of the frequency distribution of dR , where $dR_i = R_i - R_{i-1}$. Note that when calculating dR the packets do not have to be adjacent. See the SLAC \Leftrightarrow CERN two-way instantaneous packet delay variation for an example of such a distribution. Both of the above distributions can be seen to be non-Gaussian which is why we use the IQR instead of the standard deviation as the measure of "jitter". Also see RFC's 1889/3550. By viewing the Ping "jitter" between SLAC and CERN, DESY & FNAL it can be seen that the two methods of calculating jitter track one another well (the first method is labelled IQR and the second labelled IPD in the figure). They vary by two orders of magnitude over the day. The jitter between SLAC & FNAL is much lower than between SLAC and DESY or CERN. It is also noteworthy that CERN has greater jitter during the European daytime while DESY has greater jitter during the U.S. daytime. We have also obtained a measure of the jitter by taking the absolute value dR , i.e. $|dR|$. This is sometimes referred to as the "moving range method" (see Statistical Design and Analysis of Experiments, Robert L. Mason, Richard F. Guest and James L. Hess. John Wiley & Sons, 1989). It is also used in RFC 2598 as the definition of jitter (RFC 1889 has another definition of jitter for real time use and calculation). See the Histogram of the moving range for an example. In this figure, the magenta line is the cumulative total, the blue line is an exponential fit to the data, and the green line is a power series fit to the data. Note that all 3 of the charts in this section on jitter are representations of identical data. In order to more closely understand the requirements for VoIP and in particular the impacts of applying Quality of Service (QoS) measures, we have set up a VoIP testbed between SLAC and LBNL. A rough schematic is shown to the right. Only the SLAC half circuit is shown in the schematic, the LBNL end is similar. A user can lift the phone connected to the PBX at the SLAC end and call another user on a phone at LBNL via the VoIP Cisco router gateway. The gateway encodes, compresses etc. the voice stream into IP packets (using the G.729 standard) creating roughly 24kbps of traffic. The VoIP stream includes both TCP (for signalling) and UDP packets. The connection from the ESnet router to the ATM cloud is a 3.5 Mbps ATM permanent virtual circuit (PVC). With no competing traffic on the link, the call connects and the conversation proceeds normally with good quality. Then we inject 4 Mbps of traffic onto the shared 10 Mbps Ethernet that the VoIP router is connected to. At this stage, the VoIP connection is broken and no further connections can be made. We then used the Edge router's Committed Access Rate (CAR) feature to label the VoIP packets' by setting Per Hop Behavior (PHB) bits. The ESnet router is then set to use the Weighted Fair Queuing (WFQ) feature to expedite the VoIP packets. In this setup voice connections can again be made and the conversation is again of good quality. Service Predictability A measure of variability of service (or ping predictability) may be obtained by means of a scatter plot of the dimensionless variables daily average ping data rate / maximum ping data rate versus the daily average ping success / maximum ping success (where % success = (total packets - packets lost) / Total Packets). Here ping data rate is defined as $(2 * \text{bytes in ping packet}) / \text{response time}$. The 2 is since the packet has to go out and come back. Another way of looking at the ratios is that numbers near 1 indicate that average performance is close to the best performance. Numbers not close to 1 are typically caused by large variations in ping time between work hours and non-work hours, see for example the UCD ping response for Oct 3, 1996 for an example of the diurnal variations. Some examples of ping predictability scatter plots for various parts of the Internet as measured from SLAC for July 1995 and March 1996 can be seen below. Date All Hosts ESnet N. America International Jul '95 Mar '96 One can reduce this scatterplot information further by plotting the monthly average ping packet success / maximum ping packet success versus the monthly average ping throughput / maximum ping throughput for different months to see the changes. Such a plot for a few N. American nodes for July 1995 and March 1996 shows big changes, in all cases being for the worse (more recent points are more to the lower left of the plot). Unpredictability One can also calculate the distance of each predictability point from the coordinate (1,1). We normalize to a maximum value of 1 by dividing the distance by $\sqrt{2}$. I refer to this as the ping unpredictability, since it gives a

percentage indicator of the unpredictability of the ping performance. ReachabilityBy looking at the ping data to identify 30 minute periods when no ping responses were received from a given host, one can identify when the host was down. Using this information one can calculate ping unreachability= (# periods with Node down / total number of periods), # Down periods, Mean Time Between Failure (MTBF or Mean Time To Failure MTTF) and Mean Time To Repair (MTTR). Note that MTBF = sample_time/ping_unreachability where for PingER sample time is 30 minutes. The reachability is very dependent on the remote host, for example if the remote host is renamed or removed, the host will appear unreachable yet there may be nothing wrong with the network. Thus before using this data to provide long term network trends the data should be carefully scrubbed for non-network effects. Examples of ping reachability and Down reports are available. One can also measure the frequency of outage lengths using active probes and noting the time duration for which sequential probes do not get through. Another metric that is sometimes used to indicate the availability of a phone circuit is Error-free seconds. Some measurements on this can be found in Error free seconds between SLAC, FNAL, CMU and CERN. There is also an IETF RFC on Measuring Connectivity and a document on A Modern Taxonomy of High Availability which may be useful. Out of Order packets PingER uses a very simple algorithm for identifying and reporting out of order packets. For each sample of 10 packets, it looks to see if the sequence numbers of the responses are received in the same order as the requests were sent. If not then that sample is marked as having one or more out of order responses. For a given interval (say a month) the value reported for out of order is the fraction of samples that were marked with out of order ping responses. Since the ping packets are sent at one second intervals it is expected that the fraction of out of order samples will be very small, and may be worth investigating whenever it is not. Duplicate Packets Duplicate ping responses can be caused by: More than one host has the same IP address, so all these hosts will respond to the ICMP echo request. The IP address pinged may be a broadcast address. The host has multiple TCP stacks bound to an Ethernet adaptor (see http://www.doxpara.com/read.php/tcp_chorusing.html). A router believes it has two routes by which it can reach the end host and (presumably mistakenly) forwards the ICMP echo requests by both routes, thus the end host sees two echo requests and responds twice. There may be two or more (non-routed) paths to the end host and each request is forwarded by more than one path. A misbehaving NAT box. Some tests that may help include: Pinging the routers along the route to see if any of them respond with duplicates. Capture the ping packets and look to see if all the packets are returned from the same Ethernet address. An idea of the prevalence of duplicate ping packets comes from PingER measurements on March 31st 2012 to 703 hosts in over 600 countries. Of these hosts 15 responded with duplicate pings. For 13 of the 15 hosts it occurred on both 100 and 1000 Byte pings. Out of 10 pings sent 6 hosts had 1 ping duplicated, 5 had 2 pings duplicated, 2 had 4 pings duplicated, 1 had 3 pings duplicated and 1 returned 12 pings for each ping sent. The sites of the hosts range from national labs (CERN, IHEP SU), developed countries (Israel), developing countries (Burkina Faso, Malawi, Mauritius, Sierra Leone, Swaziland, Zambia), and educational sites (SDSC). PingER simply reports whether there were duplicates or not. A useful metric is to report the number of pings received/number pings sent. The number received may depend on the ping command options. One option will send a given number of pings until it receives that many back or times out. Another option will send 10 pings and wait (or time out) until they are received. So the metric value may also depend on the ping command. Combination of all ping measures One can put together a plot of all the above ping measures (loss, response, unreachability and unpredictability) to try and show the combination of measurements for a set of hosts for a given time period. The plot below for March 1-11, 1997, groups the hosts into logical groups (ESnet, N. America West, ...) and within the groups ranks the hosts by % 100 byte ping packet loss for SLAC prime time (7am - 7pm weekdays), also shown by a blue line is the prime time ping response time, and the negative of the ping % unreachability and unpredictability. In the above plot, the loss and response time are measured during SLAC prime time (7am - 7pm, weekdays), the other measures are for all the time. The loss rates are plotted as a bar graph above the y=0 axis and are for 100 byte payload ping packets. Horizontal lines are indicated at packet losses of 1%, 5% and 12% at the boundaries of the connection qualities defined above. The response time is plotted as a blue line on a log axis, labelled to the right, and is the round trip time for 1000 byte ping payload packets. The host unreachability is plotted as a bar graph negatively extending from the y=0 axis. A host is deemed unreachable at a 30 minute interval if it did not respond to any of the 21 pings made at that 30 minute interval. The host unpredictability is plotted in green here as a negative value, can range from 0 (totally unpredictable) to 1 (highly predictable) and is a measure of the variability of the ping response time and loss during each 24 hour day. It is defined in more detail in Ping Unpredictability. The following observations are also relevant: ESnet hosts in general have good packet loss (median 0.79%). The average packet losses for the other groups varies from about 4.5% (N. America East) to 7.7% (International). Typically 25%-35% of the hosts in the non-ESnet groups are in the poor to bad range. The response time for ESnet hosts averages at about 50ms, for N. America West it is about 80ms, for N. America East about 150ms and for International hosts around 200ms. Most of the unreachable problems are limited to a few hosts mainly in the International group (Dresden, Novosibirsk, Florence). The unpredictability is most marked for a few International hosts and roughly tracks the packet loss. Quality In order to be able to summarize the data so the significance can be quickly grasped, we have tried to characterize the quality of performance of the links. Some interesting reports are below: ESnet reports on bandwidth requirements. QoS Requirements of Network Applications on the Internet. Network QoS Needs of Advanced Internet Applications: A Survey Below are some other measures organized by metric. Delay The scarcest and most valuable commodity is time. Studies in late 1970s and early 1980s by Walt Doherty of IBM and others showed the economic value of Rapid Response Time: 0-0.4s High productivity interactive response 0.4-2s Fully interactive regime 2-12s Sporadically interactive regime 12s-600s Break in contact regime 600s Batch regime For more on the impact of response times see The Psychology of Human-Computer Interaction, Stuart K. Card, Thomas P. Moran and Allen Newell, ISBN 0-89859-243-7, published by Lawrence Erlbaum Associates (1983). There is a threshold around 4-5s where complaints increase rapidly. For some newer Internet applications there are other thresholds, for example for voice a threshold for one way delay appears at about 150ms (see ITU Recommendation G.114 One-way transmission time, Feb 1996) - below this one can have toll quality calls, and above that point, the delay causes difficulty for people trying to have a conversation and frustration grows. For keeping time in music, Stanford researchers found that the optimum amount of latency is 11 milliseconds. Below that delay and people tended to speed up. Above that delay and they tend to slow down. After about 50 milliseconds (or 70), performances tended to completely fall apart. The human ear perceives sounds as simultaneous only if they are heard within 20 ms of each other, see http://www.mercurynews.com/News/ci_27039996/Music-at-the-speed-of-light-is-researchers-goal For real-time multimedia (H.323) Performance Measurement and Analysis of H.323 Traffic gives one way delay (roughly a factor two to get RTT), of: 0-150ms = Good, 150-300ms = Acceptable, and > 300ms = poor. The SLA for one-way network latency target for Cisco TelePresence is below 150 msec. This does not include latency induced by encoding and decoding at the CTS endpoints. All packets that comprise a frame of video must be delivered to the TelePresence end point before the replay buffer is depleted. Otherwise degradation of the video quality can occur. The peak-to-peak jitter target for Cisco TelePresence is under 10 msec. The paper on The Internet at the Speed of Light gives several examples of the importance of reducing the RTT. Examples include search engines such as Google and Bing, Amazon sales and the stock exchange For real time haptic control and feedback for medical operations, Stanford researchers (see Shah, A., Harris, D., & Gutierrez, D. (2002). "Performance of Remote Anatomy and Surgical Training Applications Under Varied Network Conditions." World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002(1), 662-667) found that a one way delay of <=80msec. was needed. The Internet Weather map identifies as bad any links with delays over 300ms. Loss For the quality characterization we have focussed mainly on the packet losses. Our observations have been that above 4-6% packet loss video conferencing becomes irritating, and non native language speakers become unable to communicate. The occurrence of long delays of 4 seconds or more at a frequency of 4-5% or more is also irritating for interactive activities such as telnet and X windows. Above 10-12% packet loss there is an unacceptable level of back to back loss of packets and extremely long timeouts, connections start to get broken, and video conferencing is unusable (also see The issue of useless packet transmission for multimedia over the internet, where they say at page 10 "we conclude that for this video stream, the video quality is unintelligible when packet loss rates exceeds 12%". On the other hand MSF (Multi Service Forum) officials said as a result of tests up on next-generation networks for IPTV "testing showed that even one half of 1% of packet loss in a video stream can make the video quality unacceptable to end users" (see Computerworld, October 29, 2008). Originally the quality levels for packet loss were set at 0-1% = good, 1-5% = acceptable, 5-12% = poor, and greater than 12% = bad. More recently, we have refined the levels to 0-0.1% excellent, 0.1-1% = good, 1-2.5% = acceptable, 2.5-5% = poor, 5%-12% = very poor, and greater than 12% = bad. Changing the thresholds reflects changes in our emphasis, i.e. in 1995 we were primarily concerned with email and ftp. This quote from Vern Paxson sums up the main concern at the time: Conventional wisdom among TCP researchers holds that a loss rate of 5% has a significant adverse effect on TCP performance, because it will greatly limit the size of the congestion window and hence the transfer rate, while 3% is often substantially less serious. In other words, the complex behaviour of the Internet results in a significant change when packet loss climbs above 3%. In 2000 we were also concerned with X-window applications, web performance, and packet video conferencing. By 2005 we were interested in the real-time requirements of VoIP and are starting to look at voice over IP. As a rule, packet loss in VoIP (and VoFi) should never exceed 1 percent, which essentially means one voice skip every

three minutes. DSP algorithms may compensate for up to 30 ms of missing data; any more than this, and missing audio will be noticeable to listeners. The Automotive Network eXchange (ANX) sets the threshold for packet loss rate (see ANX / Auto Linx Metrics) to be less than 0.1%. The ITU TIPPHON working group (see General aspects of Quality of Service (QoS), DTR/TIPHON-05001 V1.2.5 (1998-09) technical Report) has also defined < 3% packet loss as being good, > 15% for medium degradation, and 25% for poor degradation, for Internet telephony. It is very hard to give a single value below which packet loss gives satisfactory/acceptable/good quality interactive voice. There are many other variables involved including: delay, jitter, Packet Loss Concealment (PLC), whether the losses are random or bursty, the compression algorithm (heavier compression uses less bandwidth but there is more sensitivity to packet loss since more data is contained/lost in a single packet). See for example Report of 1st ETSI VoIP Speech Quality Test Event, March 21-18, 2001, or Speech Processing, Transmission and Quality Aspects (STQ); Anonymous Test Report from 2nd Speech Quality Test Event 2002 ETSI TR 102 251 v1.1.1 (2003-10) or ETSI 3rd Speech Quality Test Event Summary report, Conversational Speech Quality for VoIP Gateways and IP Telephony. Jonathan Rosenberg of Lucent Technology and Columbia University in G.729 Error Recovery for Internet Telephony presented at the V.O.N. Conference 9/1997 gave the following table showing the relation between the Mean Opinion Score (MOS) and consecutive packets lost. Consecutive packet losses degrade voice quality. Consecutive frames lost 12345 MOS 3.4 2.3 2.2 4.2 1.1 7 where: Mean Opinion Score Rating Speech Quality Level of distortion 5 Excellent Imperceptible 4 Good Just perceptible, not annoying 3 Fair Perceptible, slightly annoying 2 Poor Annoying but not objectionable 1 Unsatisfactory Very annoying, objectionable If the VoIP packets are spaced apart by 20 msec then a 10% loss (assuming a random distribution of loss) is equivalent to seeing 2 consecutive frames lost about every 2 seconds, while a 2.5% loss is equivalent to 2 consecutive frames being lost about every 30 seconds. So we set "Acceptable" packet loss at < 2.5%. The paper Performance Measurement and Analysis of H.323 traffic gives the following for VoIP (H.323): Loss = 0%-0.5% Good, = 0.5%-1.5% Acceptable and > 1.5% = Poor. The above thresholds assumes a flat random packet loss distribution. However, often the losses come in bursts. In order to quantify consecutive packet loss we have used, among other things, the Conditional Loss Probability (CLP) defined in Characterizing End-to-end Packet Delay and Loss in the Internet by J. Bolot in the Journal of High-Speed Networks, vol 2, no. 3 pp 305-323 December 1993 (it is also available on the web). Basically the CLP is the probability that if one packet is lost the following packet is also lost. More formally $Conditional_loss_probability = Probability(loss(packet\ n+1)=true \mid loss(packet\ n) = true)$. The causes of such bursts include the convergence time required after a routing change (10s to 100s of seconds), loss and recovery of sync in DSL network (10-20 seconds), and bridge spanning tree re-configurations (~30 seconds). More on the impact of bursty packet loss can be found in Speech Quality Impact of Random vs. Bursty Packet losses by C. Dvorak, internal ITU-T document. This paper shows that whereas for random loss the drop off in MOS is linear with % packet loss, for bursty losses the fall off is much faster. Also see Packet Loss Burstiness. The drop off in MOS is from 5 to 3.25 for a change in packet loss from 0 to 1% and then it is linear falling off to an MOS of about 2.5 by a loss of 5%. Other monitoring efforts may choose different thresholds possibly because they are concerned with different applications. MCI's traffic page labelled links as green if they have packet loss < 5%, red if > 10% and orange in between. The Internet Weather Report we colored < 6% loss as green and > 12% as red, and orange otherwise. So they are both more forgiving than we or at least have less granularity. Gary Norton in Network World Dec. 2000 (p 40), says "If more than 98% of the packets are delivered, the users should experience only slightly degraded response time, and sessions should not time out". The figure below shows the frequency distributions for the average monthly packet loss for about 70 sites seen from SLAC between January 1995 and November 1997. Due to the high amount of compression and motion-compensated prediction utilized by TelePresence video codecs, even a small amount of packet loss can result in visible degradation of the video quality. The SLA for packet loss target for Cisco TelePresence should be below 0.05 percent on the network. For real time haptic control and feedback for medical operations, Stanford researchers found that loss was not a critical factor and losses up to 10% could be tolerated. However for high performance data throughput over long distances (high RTTs), as can be seen in ESnet's article on Packet Loss, losses of as little as 0.0046% (1 packet loss in 22,000) on 10Gbps links with the MTU set at 9000 Bytes (the impact is greater with default MTU's of 1500 Bytes) result in factors of 10 reduction in throughput for RTTs > 10 msec. Jitter The ITU TIPPHON working group (see General aspects of Quality of Service (QoS) DTR/TIPHON-05001 V1.2.5 (1998-09) technical report) defines four categories of network degradation based on one-way jitter. These are: Levels of network degradation based on jitter Degradation category Peak jitter Perfect 0 msec. Good 75 msec. Medium 125 msec. Poor 225 msec. We are investigating how to relate the one-way jitter thresholds to the ping (round-trip or two-way) jitter measurements. We used the Surveyor one-way delay measurements (see below) and measured the IQRs of the one-way delay ($J_a \Rightarrow b$ and $J_b \Rightarrow a$, where the subscript $a \Rightarrow b$ indicates the monitoring node is at a and is monitoring a remote node at b) and inter packet delay difference ($J_a \Rightarrow b$ and $J_b \Rightarrow a$). Then we add the two one way delays for equivalent time stamps together and derive the IQRs for the round trip delay ($J_a \Leftrightarrow b$) and inter packet delay difference ($J_a \Leftrightarrow b$). Viewing a Comparison of one and two way jitter one can see that the distributions are not Gaussianly distributed (being sharper and yet with wider tails), the jitter measured in one direction can be very different from that measured in the other direction and that for this case the above approximation for the round trip IQR works quite well (within two percent agreement). Web browsing and mail are fairly resistant to jitter, but any kind of streaming media (voice, video, music) is quite susceptible to jitter. Jitter is a symptom that there is congestion, or not enough bandwidth to handle the traffic. The jitter specifies the length of the VoIP codec playback buffers to prevent over- or under-flow. An objective could be to specify that say 95% of packet delay variations should be within the interval [-30msec, +30msec]. For real-time multimedia (H.323) Performance Measurement and Analysis of H.323 Traffic gives for one way: jitter = 0-20ms = Good, jitter = 20-50ms = acceptable, > 50ms = poor. We measure round-trip jitter which is roughly two times the one way jitter. For real time haptic control and feedback for medical operations, Stanford researchers found that jitter was critical and jitters of < 1msec were needed. Throughput Performance requirements (from AT&T) 768k - 1.5Mbps: sharing photos, downloading music, emailing, web surfing. 3.0Mbps - 6.0Mbps - streaming video, online gaming, home networking. > 6Mbps - hosting websites, watching TV online, downloading movies. Here are a couple more guidelines: The following is from Design Considerations for Cisco Telepresence over a PIN Architecture. The bandwidth utilized per Cisco TelePresence endpoint varies based on factors that include the model deployed, the desired video resolution, interoperability with legacy video conferencing systems, and whether high or low speed auxiliary video input is deployed for the document camera or slide presentation. For example, when deploying 1080p best video resolution with high speed auxiliary video input and interoperability, the bandwidth requirement can be as high as 20.4 Mbps for a CTS-3200 and CTS-3000, or 10.8 Mbps for a CTS-1000 and CTS-500. FCC Broadband Guide Utilization Link utilization can be read out from routers via SNMP MIBs (assuming one has the authorization to read out such information). "At around 90% utilization a typical network will discard 2% of the packets, but this varies. Low-bandwidth links have less breadth to handle bursts, frequently discarding packets at only 80% utilization... A complete network health check should measure link capacity weekly. Here's a suggested color code: RED: Packet discard > 2 %, deploy no new application. AMBER: Utilization > 60%. Consider a network upgrade. GREEN: Utilization < 60%. Acceptable for new application deployment. "High-speed slowdown, Gary Norton, Network Magazine, Dec. 2000. The above does not indicate over what period the utilization is measured. Elsewhere in Norton's article he says "Network capacity ... is calculated as an average of the busiest hours over 5 business days". "Queueing theory suggests that the variation in round trip time, σ , varies proportional to $1/(1-L)$ where L is the current network load, $0 \leq L \leq 1$. If an Internet is running at 50% capacity, we expect the round trip delay to vary by a factor of +20, or 4. When the load reaches 80%, we expect a variation of 10." Internetworking with TCP/IP, Principle, Protocols and Architecture, Douglas Comer, Prentice Hall. This suggests one may be able to get a measure of the utilization by looking at the variability in the RTT. We have not validated this suggestion at this time. Reachability The Bellcore Generic Requirement 929 (GR-929-CORE Reliability and Quality Measurements for Telecommunications Systems (RQMS) (Wireline), is actively used by suppliers and service providers as the basis of supplier reporting of quarterly performance measured against objectives. Each year, following publication of the most recent issue of GR-929-CORE, such revised performance objectives are implemented) indicates that the the core of the phone network aims for a 99.999% availability, That translates to less than 5.3 minutes of downtime per year. As written the measurement does not include outages of less than 30 seconds. This is aimed at the current PSTN digital switches (such as Electronic Switching System 5 (ESS) and the Nortel DMS250), using today's voice-over-ATM technology. A public switching system is required to limit the total outage time during a 40-year period to less than two hours, or less than three minutes per year, a number equivalent to an availability of 99.99943%. With the convergence of data and voice, this means that data networks that will carry multiple services including voice must start from similar or better availability or end-users will be annoyed and frustrated. Levels of availability are often cast into a Service Level Agreements. The table below (based on Cahners In-Stat survey of a sample of Application Service Providers (ASPs)) shows the levels of availability offered by ASPs and the levels chosen by customers. Levels offered Chosen by customer Less than 99% 26% 19% 99% availability 39% 24% 99.9% availability 24% 15% 99.99% availability 15% 5% 99.999% availability 18% 5% More than 99.999% availability 13% 15% Don't know 13% 18% Weighted average of availability offered 99.5% 99.4% For more information on availability etc. see: the Cisco White Paper on Always-On Availability for Multiservice Carrier Networks for how Cisco is striving for high

availability on data networks; A Modern Taxonomy of High Availability; and the IETF document RFC 2498: IPPM Metrics for Measuring Connectivity.

DirectivityThe theoretical limits on Directivity are that it must be 0 and 1. A value of 1 indicates the route is a great circle route and the only delay is due to the speed of light in fibre or electrons in copper. Values > 1 usually indicate the source or destination or both have incorrect locations and so make Directivity a useful diagnostic for host locations. Typical values of Directivity between research and educational sites in the US, Canada, Europe, East Asia, and Australia/New Zealand vary from 0.15 - 0.75 with a median of about 0.4. This corresponds to about 4 times slower than the speed of light in vacuum. Low values of Directivity typically mean very indirect route, or satellite or slow connection (e.g. wireless) GroupingAs the number of host pairs being monitored increased it becomes increasingly necessary to aggregate the data into groups representing areas of interest. We have found the following grouping categories to be useful:by area (e.g. N. America, W. Europe, Japan, Asia, country, top level domain);by host pair separation (e.g. trans-oceanic links, intercontinental links, Internet eXchange Points);Network Service Provider backbone that the remote site is connected to (e.g. ESnet, Internet 2, DANTE ...);common interest affiliation (e.g. XIWT, HENP, experiment collaboration such as BaBar, the European or DOE National Laboratories, ESnet program interests, perfSONAR)by monitoring site;one remote site seen from many monitoring sites. We need to be able to select the groupings by monitoring sites and by remote sites. Also we need the capability to include all members of a group, to join groups and to exclude members of a group. Some examples of how many ~1100 PingER monitor-host remote-site pairs were in the global area groups and the affinity groups can be found in PingER pair grouping distributions. At the same time it is critical to choose the remote sites and host-pairs carefully so that they are representative of the information one is hoping to find out. We have therefore selected a set of about 50 "Beacon Sites" which are monitored by all monitoring sites and which are representative of the various affinity groups we are interested in. An example of a graph showing ping response times for groups of sites is seen below: The percentages shown to the right of the legend of the packet loss chart are the improvements (reduction in packet loss) per month for exponential trend line fits to the packet loss data. Note that a 5%/month improvement is equivalent to 44%/year improvement (e.g. a 10% loss would drop to 5.6% in a year).One Way MeasurementsSLAC is also collaborating in the perfSONAR project to make one way delay and loss measurements between perfSONAR sites. Each perfSONAR site has a measurement point consisting of an Internet connected computer with a GPS receiver. This allows accurate synchronized time stamping of packets which enables the one way delay measurements. The delay estimates generated are more detailed than PingER's and clarify the asymmetry on Internet paths in the two directions. For more on comparing the two methods see the Comparison of PingER and Surveyor. RIPE also has a Test Traffic project to make independent measurements of connectivity parameters, such as delays and routing-vectors in the Internet. A RIPE host is installed at SLAC. The NLANR Active Measurement Program (AMP) for HPC awardees is intended to improve the understanding of how high performance networks perform as seen by participating sites and users, and to help in problem diagnosis for both the network's users and its providers. They install a rack mountable FreeBSD machine at sites and make full mesh active ping measurements between their machines, with the pings being launched at about 1 minute intervals. An AMP machine is installed at SLAC. More detailed comparisons of Surveyor, RIPE, PingER and AMP can be found at Comparison of some Internet Active End-to-end Performance Measurement projects. SLAC is also a NIMI (National Internet Measurement Infrastructure) site. This project may be regarded as complementary to the Surveyor project, in that it (NIMI) is more focussed on providing the infrastructure to support many measurement methodologies such as one way pings, TReno, traceroute, PingER etc. Waikato University in New Zealand is also deploying Linux hosts each with a GPS receiver and making one way delay measurements. For more on this see Waikato's Delay Findings page. Unlike the AMP, RIPE and Surveyor projects, the Waikato project makes passive measurements, of the normal traffic between existing pairs, using CRC based packet signatures to identify packets recorded at the 2 ends. The sting TCP-based network measurement tool is able to actively measure the packet loss in both the forward and reverse paths between pairs of hosts. It has the advantage of not requiring a GPS, and not being subject to ICMP rate limiting or blocking (according to an ISI study ~61% of the hosts in the Internet do not respond to pings), however it does require a small kernel modification. If the one way delay (D) is known for both directions of an Internet pair of nodes (a,b), then the round trip delay R can be calculated as follows: $R = D_{a \Rightarrow b} + D_{b \Rightarrow a}$ where $D_{a \Rightarrow b}$ is the one way delay measured from node a to node b and vice versa. The two way packet loss P can be derived from the one way losses (p) as follows: $P = p_{a \Rightarrow b} + p_{b \Rightarrow a} - p_{a \Rightarrow b} * p_{b \Rightarrow a}$ where $p_{a \Rightarrow b}$ is the one way packet loss from node a to b and vice versa. There are some IETF RFCs related to measuring one way delay and loss as well as a round trip delay metric.

Traceroute

Another very powerful tool for diagnosing network problems is traceroute. This enables one to find the number of hops to a remote site and how well the route is working.

John MacAllister at Oxford developed *Traceping Route Monitoring Statistics* based on the standard traceroute and ping utilities. Statistics were gathered at regular intervals for 24-hour periods and provided information on routing configuration, route quality and route stability.

TRIUMF also has a very nice [Traceroute Map](#) tool that shows a map of the routes from TRIUMF to many other sites. We are looking at providing a simplification of such maps to use the [Autonomous Systems](#) (AS) passed through rather than the routers.

One can also plot the [FTP throughput versus traceroute hop count](#) as well as ping response and packet loss to look for correlations.

Many sites are appearing that run [traceroute servers](#) (the [source code](#) (in Perl) is available) which help in debugging and in understanding the topology of the Internet.

Some sites provide access to network utilities such as [nslookup](#) to allow one to find out more about a particular node. A couple of examples are [SLAC](#) and [TRIUMF](#).

Impact of Routing on Internet End-to-end Performance

Glossary

- **DSCP** Differentiated Services CodePoint. The Differentiated Services CodePoint is 6 bits in the IP header field that are used to select the per-hop-behaviour of a packet. The 6 bits for DSCP and 2 unused bits are intended to supersede the existing definitions of the IPv4 TOS octet, see [RFC 2474](#) for more details.
- **MTU** Maximum Transfer Unit. The maximum transfer unit is the largest size of IP datagram which may be transferred using a specific data link connection.
- **MSS** Maximum Segment Size. MTU-40.
- **QBSS** QBone Scavenger Services. QBone Scavenger Services is an additional class of best-effort service. A small amount of network capacity is allocated (in a non-rigid way) for this service; when the default best-effort capacity is underutilized, QBSS can expand to consume unused capacity.
- **Receive Window (Rwin)** The size of the TCP buffer, the number of packets your machine will send out without receiving an ACK.

Further Information

- [SLAC Talks and Papers on Network Monitoring](#).

- [Network Monitoring Tools](#)
- [ICFA-NTF Monitoring Working Group](#) provides a lot of useful information on the results of end-to-end Internet monitoring.
- [Why Latency matters to Mobile Backhaul](#).
- [Research Network User's Guide to Performance: Knowledge Base](#) from the Geant2 Performance Enhancement Response Team.
- [Network Management Protocols](#) from Realtime Publishers.
- [CAIDA Metrics Working Group](#).
- [Network management Tutorial](#).
- [References on Network Management](#).
- [Requirements of Network Monitoring for the Grid](#) from the EuroGrid WP7 provides an interesting overview of methods for monitoring etc.

cron A cronjob calls the appropriate scripts. At SLAC this cronjob is in pinger@pinger.slac.stanford.edu/.trs/crontab. For PingER1 the script is called [timeping.pl](#), for PingER2 it is called [pinger2.pl](#). At SLAC the PingER scripts are all in [Perl](#) and unless otherwise indicated are found in the path /afs/slac/package/netmon/pinger/

gather The data is gathered at SLAC by the [getdata.pl](#) script that uses [Lynx](#) to get the data from the [ping_data.pl](#) CGI script at each monitoring site. [Getdata.pl](#) is called daily to gather the data to the SLAC archiving site. It is called from the same cronjob as calls [timeping.pl](#) (see above).

analysis Information on the analysis chain of scripts can be found in: "[Restoring Historical Data](#)" which explains the whole analysis chain.

pingtable The main reporting mechanism is via the [pingtable.pl](#) script that at SLAC is stored in the PingER CGI path /afs/slac/g/www/cgi-wrap-bin/net/offsite_mon/pinghistory/.

134 Pageviews
Apr. 04th - May. 04th

[[Feedback](#) | [Reporting Problems](#)]

[Les Cottrell](#)