SLAC-PUB-10186
October 2003

Measuring the Digital Divide with PingER

R. Les Cottrell*, Warren Matthews+
Stanford Linear Accelerator Center, 2575 Sand Hill Road, Menlo Park, CA 94025, USA,
Office of Information Technology (OIT), Georgia Institute of
Technology, Rich Building, 258 Fourth Street, Atlanta, GA 30332, USA.
cottrell@slac.stanford.edu, Warren.Matthews@oit.gatech.edu

Abstract

We introduce the PingER project/toolkit and show its relevance to monitoring sites in developing countries. We then show results from PingER that illustrate the extent of the Digital Divide in terms of Internet performance between developed and developing regions, which developing regions are catching up, keeping up, or falling behind and the magnitude of the differences in performance between developed regions and developing regions.

Introduction

Modern science increasingly requires collaborations of scientists from diverse parts of the globe. This may range from large High Energy Physics experiments with 10-20% of the collaborators coming from countries in the developing world to individual scientists in say Africa needing access to journals via the Internet [1]. To set expectations for the quality of Internet connectivity, understand where upgrades are needed and in some cases suggest how they may be achievable, provide trouble-shooting information, and provide quantitative information to planners and policy makers, it is critical to monitor the performance of the Internet to all major regions of the world.

One of the goals of the PingER project [2] is to make end-to-end active monitoring information publicly available to meet the above needs. PingER has very low network impact (< 100bits/s per monitoring-remote host by default and can be set to < 10bits/s for hosts with especially poor connectivity). Further since it uses the ubiquitous ping facility no software has to be installed or configured on the remote hosts, and no special hardware is needed, the remote host can be any host connected to the Internet that runs most of the time (e.g. a web or name or mail server). These two features (low network impact and no special hardware or software at the remote site) mean PingER is particularly well suited to gathering and providing the information for countries with poor connectivity. PingER also has historical data going back to January 1995, so there is a wealth of trend information available.

The methodology employed by PingER is described in a companion paper [3].

Results

To understand the effects of the Digital Divide at an upper level it is necessary to aggregate the measurements by regions. From this one can drill down to understand details shown by the various countries and sites in a region. We try to use regions that are well understood and accepted and also include countries with similar Internet performance and challenges. Hopefully by choosing countries/sites within a region to have similar challenges, in some cases at least, common solutions may be possible. The regions we use in this paper are outlined in Table 1.

Table 1: Regions together with countries monitored in the region, number of sites monitored in each region and the average losses and RTT for August 2003. The regions are ordered by decreasing average loss

Region	Countries	Sites	Loss Avg	RTT Avg ms
Africa	Ghana, Namibia, Nigeria, S. Africa, Uganda	6	16%	1100

Work supported in part by the Department of Energy contract DE-AC03-76SF00515.

Invited talk presented at the 2003 Round Table on Developing Countries Access to Scientific Knowledge The Abdus Salem ICTP, Trieste, Italy, 10/23/2003 - 10/24/2003

C Asia	Kazakhstan, Kyrghzstan, Mongolia, Tajikistan, Turkmenistan, Uzbekistan	9	12%	650
S Asia	Bangladesh, India, Indonesia, Malaysia, Pakistan, Thailand, (Vietnam)	16	5.3%	600
M East	Egypt, Iran, Israel, Jordan, Saudi Arabia, Turkey	10	4.7%	500
C America	Costa Rica, Guatemala, Mexico	4	2.4%	220
S America	Argentina, Brazil, Columbia, Chile, Peru, Uruguay, Venezuela	13	2.3%	360
Russia	Russia	5	2.2%	385
China	China including Hong Kong	5	2%	290
Caucasus	Armenia, Azerbaijan, Georgia	5	1.5%	630
SE Europe	(Albania), Bulgaria, Croatia, Macedonia, Moldova, Romania, Serbia/Montenegro, Slovenia	13	1.5%	280

The Internet performance RTT, loss and derived throughput (*~MSS/(RTT*sqrt(loss))* [4] where MSS~1460Bytes), seen from SLAC to each of these regions sorted by derived throughput, is seen in Fig. 1.

Figure 1: RTT, loss and derived throughput statistics for developing regions.

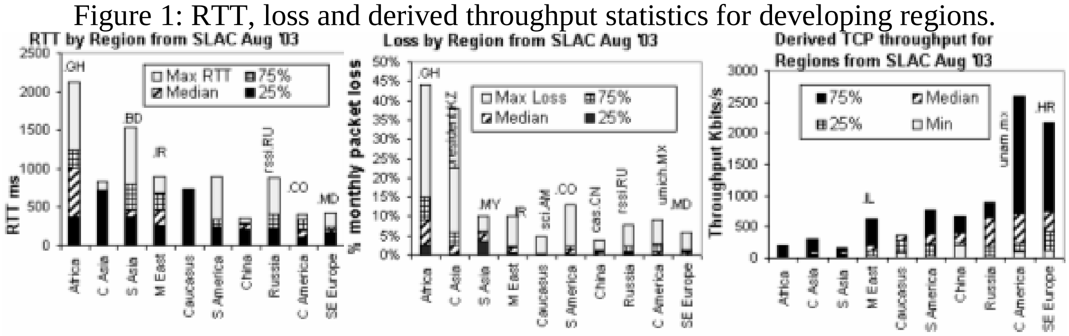


Fig. 1 also identifies countries or sites (by their Internet country code) that have large outliers compared to the rest of the region. In some of the RTT and loss cases such outlier country/sites may benefit by following the lead of others in that region (e.g. Columbia getting access to AMPATH [5] similar to Argentina, Brazil and Chile), and the large derived throughputs indicate what is achievable, e.g. Israel (.IL) and Croatia (.HR) excel in their respective regions, while in Mexico a single site unam.MX is much better than other Mexican sites.

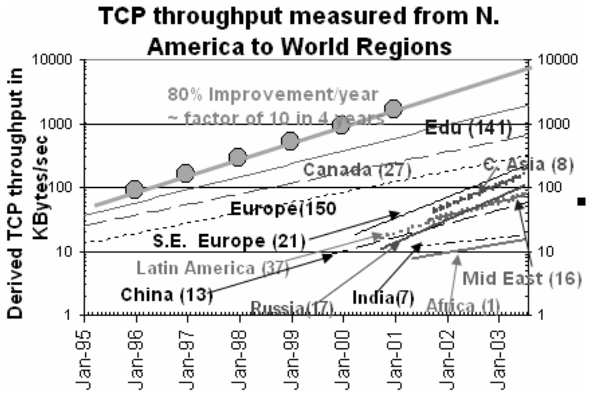
Table 2: Derived throughputs between regions for August 2003. Throughputs of < 200kbits/s are considered bad (this is below the typical DSL line), between 200 and 1000kbits/s is poor to acceptable, and above 1000kbits/s is good.

August-03	AU	CA	EDU	GOV	CH	DE	DK	HU	IT	UK	JP	NET	ORG	RU	SU	Avg
N America	234	716	9346	1993	322	344	305	329	337	405	197	3231	4121	278	148	1487
Australasia	7833	72	258	224				134	132	164		279	194		96	939
Balkans	149	233	276	365	1111	1544	939	4745	1093	1148	155	321	351	670	346	896
Europe	148	234	316	380	1662	1714	1049	1840	1596	2538	164	390	367	720	298	895
Baltics	126	183	265	331	644	1009	1693	851	896	804	149	298	313	1050	213	588
Russia	129	170	220	225	508	593	912	545	587	561	321	285	244	809	768	458
E Asia	142	196	219	209	143	163	157	152	147	181	3559	205	219	163	157	401
M East	125	187	168	234	227	505	386	453	645	572	125	243	283	331	185	311
L America	131	150	220	191	165	162	152	159	164	175	117	198	186	139	100	161
S Asia		73	179		112											121
Caucasus			87													87
Central Asia			64		68											66
Africa		40	56	27	44				32			31				38
Avg	1002	205	898	418	455	754	699	1023	563	727	598	548	698	520	257	496

Table 2 shows the derived throughput performance from monitoring hosts in various regions (columns) to remote hosts in various regions (rows) for August 2003. It is seen that when the monitoring host is close to a remote host, performance is generally better, see for example Canadian or EDU or GOV hosts monitoring hosts in N. America, Japanese hosts monitoring E. Asian hosts, or Russian hosts monitoring Russian or Baltic state hosts, and European hosts monitoring other European hosts. This is partially explained by the shorter physical distances resulting lower RTT, and also fewer hops and inter Internet

Service Provider exchange points resulting in lower losses. One can also see that hosts in Africa, the Caucasus, Central and South Asia all have bad performance.

Figure 2: Derived throughput performance from N. America to regions of the world.



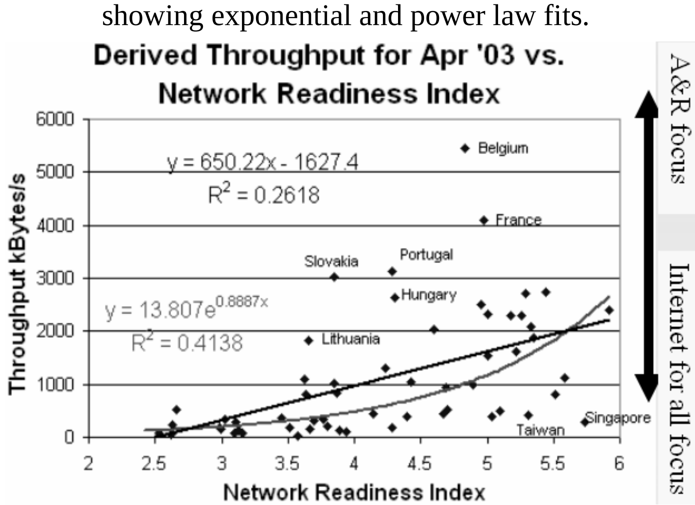
The derived throughput trends for the various major regions are shown in Fig.2. The numbers in parentheses are the number of monitor-remote hosts pairs included. RTTs greater than 500-600ms are usually caused by satellite links. In the case of Africa we have data to 3 countries, but only show throughput to Uganda since Ghana and Nigeria are a factor of 2 to 3 worse, have only been measured for 6 months, and aggregating them causes the performance for the region to apparently drop, even though most individual sites are improving. We also have data measured from CERN, but for many of the regions it only goes back to August 2001, so we do not show it here. It does, however, confirm the major conclusions seen in Fig.2, i.e.:

- In the long term, performance to all regions is improving;
- For the developed regions performance is improving by roughly a factor of 10 in 4-6 years;
- Performances to the developing regions are a factor of 5 to 50 times worse than to the developed regions;
- Performance to developing regions is typically on a par with what was seen 2-7 years ago to the developed regions;
- For the developing regions, S. E. Europe, C. Asia and Russia are catching up, Latin America, the Middle East and China are keeping up, and India and Africa are falling behind.

Comparisons with Economic/Development Indices

For the UN Gross Domestic Product per capita and Human Development Index (HDI) [6], R^2 (the square of the correlation *coefficient)*, or in this case the proportion of variation in PingER attributable to the HDI was positive and about about 24%, and for the GDP/capita about 28%. For the combined primary, secondary and tertiary education enrollment index and adult literacy rate the correlation was weaker (R^2 *~ 10%* in both cases). For European countries alone the correlation between HDI and PingER derived throughput was about R^2 *~ 55%.*

Figure 3: Derived throughputs vs. NRI for April 2003, showing exponential and power law fits.



Throughputs by country against the Network Readiness Index (NRI [7]) yielded $R^2 \sim 41\%$ (see Fig. 3). Looking at the outliers it appears that countries with large values of performance relative to NRI are for countries (e.g. Belgium, France, Slovakia) where there is a focus on providing excellent Academic and Research networking. Countries where there is more of a focus on Internet access for all (e.g. Singapore and Taiwan) typically have lower PingER performance relative to the NRI.

Challenges

Though the resources needed at the remote site are negligible (see [3]), it is important to make contact with people at remote sites in developing regions to ensure that even the light PingER load is acceptable, and to have someone to follow up in case of problems such as ping rate-limiting or blocking [3]. This initial contact often requires a careful explanation of the benefits, the host and network impacts, possible problems, and may be just part of a multi-step process as more network/computer technical people are involved. Ongoing contacts are needed to respond to pathologies such as hosts being inaccessible (e.g. due to moves or ping blocking) or high loss rates caused by rate limiting. Typically we contact about one remote site/week to follow up on pathologies.

Conclusions

PingER is a valuable, simple, light-weight tool for end to end active Internet performance monitoring and has been particularly effective in measuring performance to developing countries providing both historical and near-real time data. It has been used for trouble-shooting, setting expectations, identifying needed upgrades and their effects, choosing providers, and for providing information to policy makers and funding bodies.

Internet performance is improving worldwide. The performance between developed countries can be 5-50 times better than to developing countries. Some developing regions are catching up to the developed regions, others are keeping up and Africa and S. Asia are falling behind. As might be expected, there is a positive correlation between countries with better development indices and their Internet performance measured by PingER.

Acknowledgements

This work would not be possible without the help of hundreds of volunteer contacts at the remote sites, the maintainers of the monitoring sites, Jerrod Williams and Connie Logg of SLAC, and Maxim Grigoriev of FNAL. We would also like to thank the members of the ICFA-SCIC and the ICTP/eJDS people for providing contacts in many countries, as well as many suggestions on ways to analyze and report on the data.

References

[1] “Electronic Journals Delivery Service”, available at http://www.ejds.org/
[2] W. Matthews & R. L. Cottrell, “The PingER Project: Active Internet Performance Monitoring for the HENP Community”, IEEE Communications Magazine, May 2000.
[3] J. Williams & R. L. Cottrell, “PingER Methodology”, ICTP/eJDS Round Table: Developing Countries Accesss to Scientific Knowledge, October 2003
[4] M. Mathis, J. Semke, J. Mahdavi, T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", Computer Communication Review, volume 27, number 3, pp. 67-82, July 1997
[5] “AMPATH: The AmericasPATH Network”, available at http://www.ampath.fiu.edu/
[6] “Synack”, available at http://www-iepm.slac.stanford.edu/tools/synack/
[7] “UN Human Development Reports”, available at http://www.undp.org/hdr2003/
[8] “Network Readiness Index”, available at http://www.weforum.org/pdf/Global_Competitiveness_Reports/Reports/GITR_2002_2003/GITR_Rankings.pdf