Quantifying the Digital Divide: A Scientific Overview of the Connectivity of South Asian and African Countries

Aziz A. Rehmatullah[1], R. Les. Cottrell[1], Jerrod D. Williams[1], Arshad Ali[2]

[1] Stanford Linear Accelerator Center [2] NUST Institute of Information Technology

Abstract

The future of computing for HEP applications depends increasingly on how well the global community is connected. With South Asia and Africa accounting for about 36% of the world's population, the issues of internet/network facilities are a major concern for these regions if they are to successfully partake in scientific endeavors. However, not only is the international bandwidth for these regions low, but also the internal network infrastructure is poor, rendering these regions hard to access for the global HEP community. In turn this makes collaborative research difficult and high performance grid activities essentially impractical. In this paper, we aim to classify the connectivity for academic and research institutions of these regions as a function of time, as seen from within, without and between the regions, and draw comparisons with more developed regions. The performance measurements are carried out using the PingER methodology; a lightweight approach using ICMP Ping packets. PingER has measurements to sites in over 110 countries that contain over 99% of the world's Internet connected population and so is well-positioned to characterize the world's connectivity. These measurements have been successfully used for quantifying, planning, setting expectations for connectivity and for identification of problems. The beneficiaries of this data range from international funding agencies and executive-level planners to network administrators.

Introduction:

The last decade has seen tremendous improvements in the Internet infrastructure with users experiencing, lower packet loss, Round Trip Times (RTT) and increased throughputs [1]. The PingER measurements [2] have been used for over a decade for monitoring Internet connectivity worldwide and more recently, the focus has shifted to the developing and under-developed regions, especially Africa [3] for the purpose of quantifying the Digital Divide.

In 2005, PingER monitoring hosts [4] were added at three major locations, in India at ERNet [5], in Pakistan at NIIT [6] and at NTC [7] and in South Africa at TENET [8]. This helped in measuring the connectivity of these developing regions from within. This paper focuses on reporting the initial results from these measurements. Unfortunately, since there are no remote or monitoring hosts in some South Asian and African countries, we are unable to include them in our analysis. However, we are striving to set up suitable monitoring infrastructures in Bangladesh, Sri Lanka and also working to enhance our coverage in Africa.

Viewed from the World

To compare the performance of these regions from the rest of the world, we first show in Figure 1 the packet losses from North America to selected regions of the world.

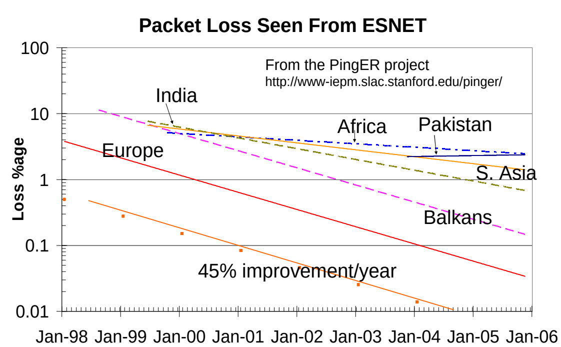

Figure 1: Packet Loss seen from ESNET to the World Regions

In general, the packet losses have declined by almost 45% each year. However, the progress for Africa and South Asia has been much slower. Since 2003 when PingER started monitoring the remote hosts in these regions, Pakistan, has shown little improvement, in fact the overall packet loss may be on the rise. For India, the progress has been slower compared to the rest of the world, but in comparison to Africa and Pakistan, the rate has been much better. Based on PingER's definition of connection quality [9], India, Pakistan and Africa all have acceptable quality levels (< 2.5% loss), with India expected to become good (< 1% loss) by 2007.

The minimum RTT seen in Figure 2 is dominated by the distance between the monitoring host and the remote (monitored) host and is usually independent of queuing in the routers.

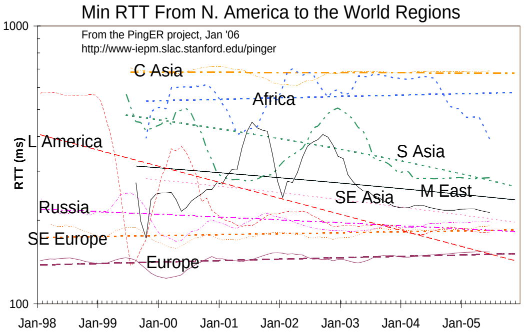

Figure 2: Minimum RTTs from North America to the World Regions. The straight lines are exponential fits, and the wiggly lines are moving averages over the last 6 months.

For Pakistan and Africa, the minimum RTTs indicate an upward trend. For Pakistan, this can be attributed to increasing coverage with some newly added sites having higher RTT values. In the case of Africa, we have increased the coverage to less well connected countries and many sites with satellite links have been added. For example, in Dec-03, there were a total of 23 remote sites in Africa as compared to 56 in Dec-05. On the other hand, minimum RTTs to India have gradually improved at roughly 9% per year, approximately 40% in 6 years since 1999. Comparing with the rest of the world, Latin America has shown marked improvement, mainly due to a move from satellite links to terrestrial from 2000 through 2003. The minimum RTTs to most other regions have been more or less stable.

In the last two years, the minimum RTT trend has shown a sharp decline for Africa. This is dominated by factors such as Botswana and Namibia shifting from satellite to terrestrial links in Sep-04 and May-04 respectively, causing the minimum RTTs to drop from the 600ms+ range to around close to 300ms. The gradual decline in minimum RTT to India indicates a gradual improvement in its internal infrastructure and improved routing.

To get an overall idea of the throughput expectation we use the Mathis [10] formula to deriver the TCP throughput from the loss and RTT measurements.

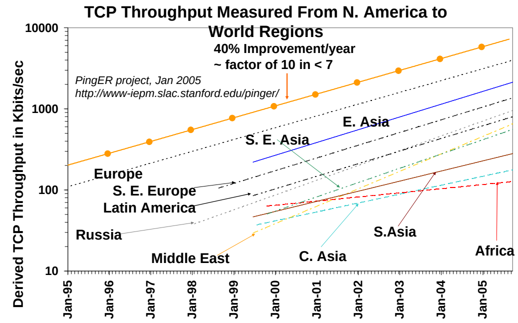

Figure 3: Throughput seen from North America to the World Regions

As seen in Figure 3, the throughputs for many world regions are increasing at an average of about 40% per year. Some developing regions like South Asia and Africa are behind Europe respectively by almost 8 and 11 years. On the basis of the existing trends, S. Asia and Africa will be 11 and 16 years behind Europe respectively by 2015.

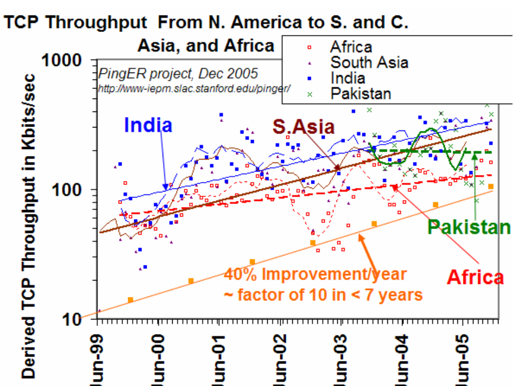

Figure 4: TCP Throughput seen from North America to South and Central Asia and Africa with exponential fits and bi-annual moving average trends.

More detail of the throughputs are seen in Figure 4. India is seen to be gradually progressing but not at the same pace as the rest of the world, for instance Russia over the last few years has improved remarkably. Africa is progressing, but the pace of progress is even slower as compared to India. Pakistan has barely managed to maintain its levels for the last 2 years and appears to be falling behind. There is a dire need in these regions, in particular Africa and Pakistan, to increase the internet bandwidth to the rest of the world. The current trends suggest that within a few years, there will be a Digital Divide within the Digital Divide.

India from within India

When we look at measurements taken from India to other remote hosts in India in Figure 5, we see a large variation in a government site being monitored in Hyderabad. Whereas the minimum RTT values and packet loss percentages are good, there is an average gap of almost 150ms between the minimum and average values, indicating that the path is heavily congested. The other two Indian sites seen from Bangalore have much smoother values with the links occasionally experiencing low congestion and the difference between the minimum and average RTT oscillating between 0 and 10ms most of the time.

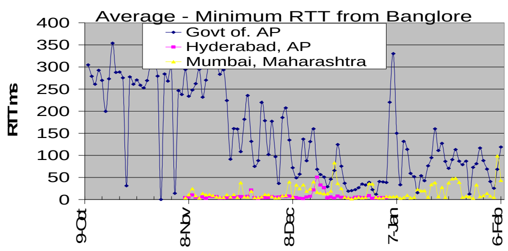

Figure 5: Minimum RTT from PingER Monitoring host in Bangalore to three remote nodes in India

Also, in a random sample of weekdays during the last few weeks, throughputs from SLAC and CERN to Bangalore and Gujarat dropped by a factor of almost 2 between 7:00 pm to 5:00 am India time indicating a marked decrease in congestion during those hours.

Africa from within Africa

Since the PingER monitoring host was deployed in South Africa as recently as August '05, the amount of data is insufficient to suggest any major indications. However, based on the minimum RTT values, traceroutes and ping triangulation [11], we deduced the routing to the monitored hosts in the African countries.

Except for two immediate neighbors of South Africa and one out of the 3 sites in Tunisia, all traffic to all the African countries was routed via USA or Europe (see Figure 6). Not only does this increase the RTT by 150-250 ms (for terrestrial links), but also consumes expensive international bandwidth. For the satellite links, the choice of routing limited. Theoretically, the minimum RTT has an approximate lower bound of 85ms[1] to anywhere in Africa from South Africa.

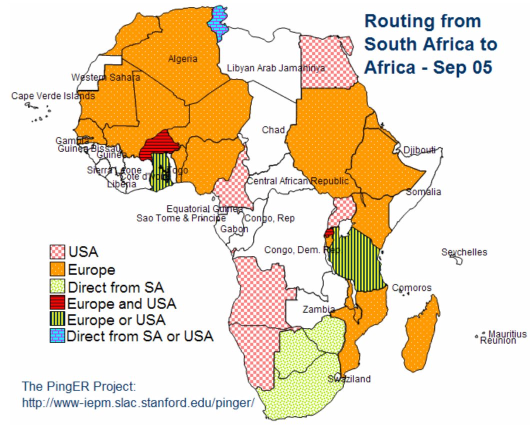

Figure 6: Routing of traffic from South Africa to Africa.

Due to the lack of exchange points in Africa, the minimum RTT values from South Africa to other parts of Africa are unusually high. In Figure 7, only the countries in red (darkest) are satellite links. For terrestrial links, minimum RTT values greater than 100ms are indicative of the countries having indirect routing from South Africa.

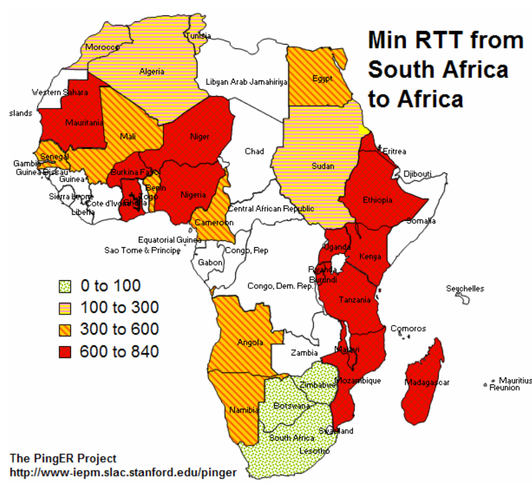

Figure 7: Minimum RTT from South Africa to Africa during the month of September 05

Pakistan from within Pakistan

In January and December of 2005, two PingER monitoring hosts were deployed at the NUST Institute of Information Technology (NIIT) Pakistan, one (NIIT-2) uses the National Telecommunication Corporation's [12] link and the other (NIIT-4) uses Micronet Broadband's [13] link. Another (NTC-1) monitoring host was deployed at NTC in November, which is very important as NTC is providing services to the Pakistan Education and Research Network (PERN) [14].

Measuring the minimum RTT to the same set of 7 nodes (all part of PERN and serviced by NTC)

[1] Assuming the speed of light in fiber optics, or 100km/msec

in Pakistan, produced fairly variable results with NIIT-2, NIIT-4 and NTC-1 having median minimum RTTs to Pakistan's other academic institutions of 10.37ms, 57.8ms and 5.8ms respectively. The large value for NIIT-4 is presumably due to the slower backbone and/or different routes used by Micronet, which add almost 50ms of time to the minimum RTTs.

However, a more interesting observation (see Figure 8) is the difference in the queuing as shown by the *average(RTT) - min(RTT)* from the three monitoring hosts. The links from NIIT-4 and NTC-1 experience little or no queuing. NIIT-2 experiences huge amounts of queuing, that may be attributed to the high usage of the link that in turn causes high congestion.

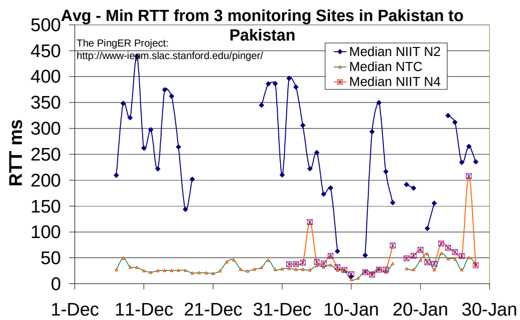

Figure 8: Median of *Average(RTT) – Min(RTT)* from three monitoring hosts in Pakistan to a set of 7 remotes nodes (NIIT to NIIT was removed to eliminate the effect of LAN)

In Figure 9, the high average RTT observed to NIIT's NTC link from the NTC and NIIT/ Micronet monitoring hosts shows that the NIIT NTC link itself is heavily congested. Moreover, when monitored from NIIT via the NTC link, the average RTTs to the Pakistani sites oscillate between 200ms and 400ms indicating heavy congestion/utilization at a common point. We believe this is due to the inadequate capacity of the last mile connectivity between NIIT and NTC. NIIT is situated far enough away from the telephone exchange that there is little choice but to migrate to fiber from copper in the last mile for NIIT to improve connectivity.

Figure 9 also highlights the fragility of the network, with the blue circles (light) showing outages at NIIT and red (dark) showing outages at remote hosts. These outages occur for a variety of reasons including power and system failures. In 2005, Pakistan has also been affected by several country-wide Internet outages, with a major outage in Jun-July [15], lasting almost a fortnight, due to a fault in the sole undersea fiber optics cable to the country. The backup satellite link was inadequate (it could only carry 25% of the traffic) to support the country's growing Internet traffic, almost doubled the RTTs, and its use was prioritized for call center traffic, resulting in dreadful performance for those universities that could use the satellite link.

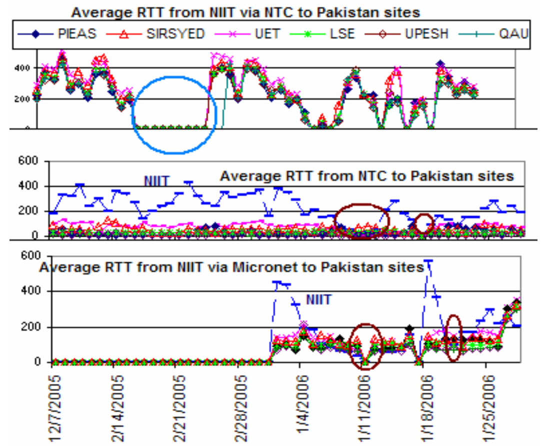

Figure 9: Average RTTs from NIIT via NTC and Micronet, and NTC to hosts in Pakistan. The blue dashed lines are for connections to NIIT via NTC. The circles show various outages.

Queuing as Seen from S. Asia and Africa to Various Countries

One of the major issues experienced in some developing countries is the amount of queuing in their networks. Although the sites in India and South Africa are less congested compared to Pakistan, their results, in Figure 10, show a dominant effect of congestion in the connectivity of the NIIT/NTC monitoring host.

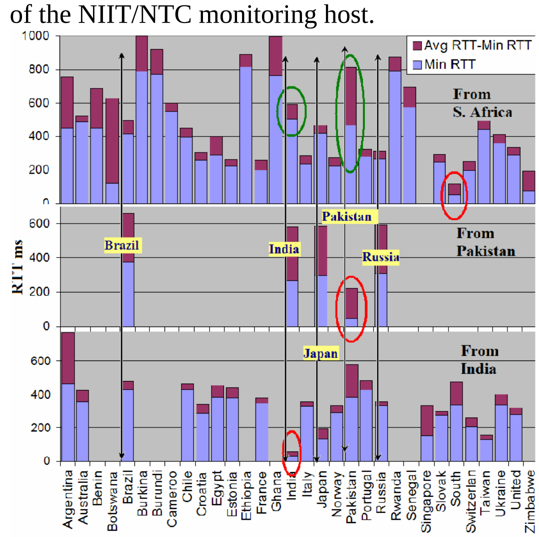

Figure 10: Minimum and Average RTTs from India, Pakistan and South Africa to different Countries. The red ellipses indicate within region measurements. The green ellipses are to highlight the differences between Pakistan and India from South Africa.

However, the results from Africa as well as India to Pakistan also show that the difference in the average and minimum RTT is quite high indicating congestion. Nodes in Argentina had much higher difference between Average and Minimum RTTs, both from S. Africa and India.

Conclusion

The growth of Internet users from the year 2000 to 2005 has been 182% [16] and almost 400% in Africa. Although Africa is still far behind the rest of the world, the PingER results do show a trend of improvement over the last few years. One of the major improvements seen in the African countries is that many countries now have terrestrial links, which reduces the RTTs, thus making the infrastructure more viable for real-time applications. More countries, especially in East Africa will also get terrestrial links under the EASSy Project [17]. Establishment of African exchange points and increase in bandwidth are needed in Africa. The networks in India are indicated to be better than those in Pakistan and Africa, probably reflecting the ever growing influence of India as a technology hub. We need more monitoring and remote sites in India to make stronger statements. The end-to end performance of the networks in Pakistan has not shown major improvements over the last two years. As a result they have become more and more congested as the utilization increased by almost 5000% [18] since the year 2000. This calls for greater increase in bandwidth, improvement of the internal network infrastructure in the country and in particular improvements in last mile performance. We also observe major variations within the Digital Divide itself with regions such as S. Asia and Africa being worse off than other developing regions, and massive variations within Africa itself. We also observe that in some cases the problems are in the last mile.

Acknowledgements

We thank Harvey Newman and ICFA for their continued enthusiasm and support. For the site in Capetown, South Africa we are grateful to Duncan Martin and others at TENET. Anil Srivastava has been phenomenal in helping us find new sites in India. We are also grateful to the NTC chairman and his team for the support that we received setting up a monitoring site at their headquarters. We also thank the entire PERN team for their help and support. ICTP Italy has also been a great help in spreading the coverage of PingER and FNAL provides support for the PingER management and infrastructure. We also thank Michael Foley of the World Bank for his efforts in facilitating contacts in Bangladesh.

References

[1] January 2006 Report of the ICFA-SCIC Monitoring Working Group: http://www.slac.stanford.edu/xorg/icfa/icfa-net-paper-jan06/

[2] The PingER Project: http://www-iepm.slac.stanford.edu/pinger/

[3] The PingER Brochure: http://www-iepm.slac.stanford.edu/monitoring/bulk/sc2005/pinger.pdf

[4] PingER Monitoring and Remote Sites: http://202.83.166.181/worldkit/index.html

[5] Education and Research Network: http://www.ncb.ernet.in

[6] NUST Institute of Information Technology http://www.niit.edu.pk

[7] National Telecommunication Corporation, Pakistan http://www.net.net.pk

[8] Tertiary Education Network, South Africa http://www.tenet.ac.za

[9] Tutorial on Internet Monitoring and PingER at SLAC http://www.slac.stanford.edu/comp/net/wan-mon/tutorial.html

[10] *The macroscopic behavior of the TCP congestion avoidance algorithm* by Mathis, Semke, Mahdavi & Ott in Computer Communication Review, 27(3), July 1997

[11] Constraint Based Geolocation of Internet Hosts, IMC'04, October 25–27, 2004, Taormina, Sicily, Italy Bamba Gueye et al

[12] National Telecommunication Corporation (NTC) http://www.ntc.net/.

[13] Micronet Broadband (pvt) Ltd. http://www.dsl.net.pk

[14] Pakistan Education and Research Network (PERN) http://www.pern.edu.pk

[15] Fiber outage to Pakistan June 27th to July 8, 2005, http://www.slac.stanford.edu/grp/scs/net/case/pakjul05/jun-july.htm

[16] Internet Usage Statistics - The Big Picture: http://www.internetworldstats.com/stats.htm

[17] The East African Submarine System http://eassy.org/

[18] Internet Usage and Population in Asia: http://www.internetworldstats.com/stats3.htm