

PingER data for an Economic Analysis

Another metric you could try from the PingER is TCP throughput which is a mixture of average RTT and loss, i.e. $TCP_throughput \sim 1460 \text{ Bytes} / (RTT \cdot \sqrt{\text{loss}})$ and so combines two PingER measurements in one. A problem is that since RTT depends on the distances between monitoring host and monitored host. Thus if measuring from say SLAC then a monitored host in the US will look much better than a monitored host in say Japan.

Loss on the other hand is usually at the edges and so more likely to be distance independent. Thus another PingER metric to try correlation with would be loss.

Jitter (or inter packet pair deviation - IPDV) is another metric that is pretty much distance independent.

One other possibly interesting thing might be to see if there is a correlation with hosts/countries that have international access via Geo Stationary satellite versus those that terrestrial access. One can tell it is a GEOS satellite connection if the minimum RTT is > say 400ms.

Finally there is another metric that is a combination of other direct measurements. This is the Mean Opinion Score (MOS) see http://en.wikipedia.org/wiki/Mean_opinion_score. It is calculated from the average RTT, jitter and loss. This basically gives the quality to be expected from VoIP, but is also relevant to other real time applications. PingER does not automatically provide the MOS. However I have put together an attached spreadsheet, in case it shows something.

Please keep me posted.

-----Original Message-----

From: John Horton john.joseph.horton@gmail.com

Sent: Saturday, October 30, 2010 6:52 PM

To: Cottrell, Les

Subject: Re: using PingER data for an economic analysis

This is really helpful. There is some worker-to-employer communication via email, but a lot of it happens across HTTP, on the intermediaries site. I'm going to try up time fractions you suggested w/ the data I have and see if there are any correlations. I'll keep you posted if you're interested.

Thanks again,
John

On Sat, Oct 30, 2010 at 2:14 AM, Cottrell, Les <cottrell@slac.stanford.edu> wrote:

Sounds interesting. Thinking aloud...

If someone is applying for a job I would imagine they would use email. Most emails are fairly small (<< 1 MB) and very resilient to poor network connections since the Mail Transfer Agents (MTAs) keep trying. Thus the problem is not one of large file transfers, or good VoIP connectivity (which needs low jitter, < 250ms round trip time, and low loss). I suspect the biggest problem would be the ability of someone to find and afford a reasonably reliable, convenient Internet connection. There are ITU statistics (e.g. see <http://www.internetworldstats.com/list4.htm>) that provide information on Internet penetration into countries. A second metric might be the reliability of Internet connections in a country. For example for what fraction of the time are nodes up and responding. If hosts do not respond then probably people cannot use them to send email. PingER does provide this info. See for example <http://www-iepm.slac.stanford.edu/pinger/intensity-maps/pinger-metrics-intensity-map.html>.

Not sure if this helps. Let me know if you have follow up questions or you think I can help.

-----Original Message-----

From: John Horton john.joseph.horton@gmail.com

Sent: Thursday, October 28, 2010 7:13 AM

To: Cottrell, Les

Subject: using PingER data for an economic analysis

Dear Dr. Cottrell,

We haven't met, but I was hoping I could ask you a question about using PingER data for a research project.

I'm an economist studying an online labor market (www.odesk.com) where large numbers of the workers are from developing countries. One thing I am interested in is how the number of applications a vacancy receives affects the probability that that job will be filled.

As you can probably imagine, this isn't straightforward---jobs that get lots of applicants might be systematically different than jobs that get few applicants. To get at this question causally, you need something that randomly "assigns" some jobs to get more or fewer numbers of applicants. In the statistics literature, this is called a instrumental variable <http://en.wikipedia.org/wiki/Instrumental_variable>.

My idea is that from an individual user's perspective, the quality of their internet connection on any particular day is independent from other factors and would affect how many jobs they could apply for (if any). It seems to my inexperienced eye that the PingER data might be perfect for this, but it's a little overwhelming and I'm not sure what measure would best map (latency, packet loss, unresponsiveness) to what I care about (end user ability to interact w/ the odesk.com servers), and whether it makes sense to use a node in a particular country as a proxy for nearby nodes (which the workers would actually be routing their data through).

Any thoughts you might have would be greatly appreciated,
John

—

John Horton
PhD Candidate in Public Policy
Harvard Kennedy School
Resident Tutor, Pforzheimer House
(617) 595-2437
<http://sites.google.com/site/johnjosephhorton/>