

20150506 SLAC SEECS and UNIMAS Meeting

Time & date

Wednesday May 6th 2015 9:00pm Pacific Daylight Time, Thursday May 7th 2015 9:00am Pakistan time, Thursday May 7th 2015 noon Malaysian time, Thursday May 7th, 2015 02:00am Rio Standard Time.

Attendees

Invitees:

Hassaan Khaliq-, Kashif, Raja, Samad Riaz+ (SEECS); Johari+, Nara, Adnan Khan (UNIMAS); Abdullah, Badrul?, Anjum+, Ridzuan?, Ibrahim? (UM); Hanan, Saqib+ (UTM); Adib?, Fatima (UUM)+; Fizi Jalil- (MYREN); Les+, Bebo+ (SLAC)

+ Confirmed attendance

- Responded but Unable to attend:

? Individual emails sent

Actual attendees:

Anjum, Samad, Johari, Ibrahim, Saqib, Fatima, Bebo, Les

Administration

- Membership of pinger-my in <https://groups.google.com>
- How is the measurements, analysis and paper on GeoLocation coming along?
 - He has been looking at the alpha (directivity) behavior, there was an exponential behavior, but it was unclear how to take advantage of it. Anjum cut N. America into regions to facilitate improving the accuracy of the alpha prediction. He believes this will improve things.
 - **Anjum has a SLAC account. Privs have been set. he has copied many of Raja's scripts into his account. However many of them point to Raja's directories.**
- Johari has got the OK from the conference organizing committee to hold a colocated PingER/BigData workshop on August 3rd the day before the CITA 2015 (see <http://www.cita.my/> an International Conference 4th - 6th August 2015, on transforming Big Data into Knowledge. Johari will provide relevant information to Bebo. Bebo will be able to make a presentation. Les has sent Bebo some relevant slide decks. **Johari has an abstract from Bebo. There are PingER related papers submitted from UM**

UFRJ

Maria Luiza reported 4/13/2015:

'We are finally solving the problem with the network connection in our lab, and I guess the machines will be ready for some tests by the end of this week.'

Thiago Barbosa, a student from the Rural Federal University of Rio de Janeiro, who is in New York with a Science without Border grant has been invited to SLAC from June 1st to August 2015. He will be working on big data analysis for PingER.

Cristiane has studied the PinGER data and how to cast it into Linked Open Data form. The size of the [PingER hourly data for 1998-Sep 2014 archived via FTP in text form](#) amounts to ~ 5.12GB and this corresponds to 15.66×10^9 (billion) triples. Then using 5 triples for each measurement and using Turtle without compression gives us 685 Gbytes or an inflation factor of ~ 200.

When Christiane made the estimation of PingER triples, she wrote two documents that explain the process but they were in Portuguese. She has written the new versions in English.

- Counting PingER Measurements: https://www.dropbox.com/s/35itp7v6yasy3rb/Counting%20PingER%20Measurements%20_EnglishVersion.docx?dl=0
- PingER LOD Triples: https://www.dropbox.com/s/4oj5jqupwbujja5/PingERLOD%20Triples%20_EnglishVersion.docx?dl=0

Christiane's report is at: [Size Inflation of PingER Data for use in PingER LOD](#)

UUM

pinger.uum.edu.my is down Jan 31st, Feb 1st. **Adib reports 5/1/2015:**

"No update from UUM. I did contact and follow up last week.

Suppose configuration should be ready but still not OK. From my side, everything is ready.

Will try again on Tuesday,"

After the last meeting Les sent Fatima a link to the Cristiane's report and also introduced Cristiane and Fatima to one another.

Fatima is working with Hadoop/MapReduce not sparq. She is installing Hadoop on the first machine already. Hopefully by Thursday, she should start work on the second and third machine.

She is having problems loading <ftp://ftp.slac.stanford.edu/users/cottrell>. The link keeps loading for a very long time and wouldn't open. Anjum and Johari do not have problems. It could be a browser pop up is blocked or it maybe that port 21 is blocked at her site. Fatima will send a screen shot of the result of telnet ftp.slac.stanford.edu 21. Also what is the IP address of her computer is it a VM (it is not a VM), is it on a private network. Ibrahim reports that the link used to not work, but works now. Can Fatima create an account for Anjum on her computer? If we can't make this work, we may need to ship a disk. One can get a 64GB USB stick for \$25 these days.

Adib is involved in NETAPPS 2015. It is a forum for scientists, researchers, students, and practitioners from all over the world to present their latest research results, ideas, and developments in the area of Future Internet and discuss advancement of next generation networks. More details about NETAPPS2015 can be found in the following link: <http://internetworks.my/netapps2015/v2/index.php>. Included in the advisory committee are: Bebo, Anjum, Les. the conference is in December in KL. There was a discussion of how to engage PingER. There is a track on Internet protocols and service. This was agreed to be a good place to present some PingER papers on Internet measurements and a Malaysian Case Study. Saqib started a case study. It is at: https://drive.google.com/folderview?id=0B-NEKleLI79ZFNmUnhiVGJ0Nmc&usp=sharing_eid. It is incomplete and needs updating. Anjum will find a new PhD student to look at this.

UM

Ibrahim is conducting experiments on using Big data with PingER data and writing a conference paper. He has a cloud of 4 computers at UM running Hadoop and 15 at MYREN. The experiment is on going, have have not been unable to complete yet. For the conference he just submitted an abstract about join processing in MapReduce. He is expecting result in coming days. 5/3/2015.

Anjum reported that UM had experienced a TCP syn DOS attack prior to Mar 12th (when an IDS was put in place). It occurred mainly for several days before between the hours on noon- 2pm and 7-7 in the evening (Malaysia time). He suggested looking to see if PingER could spit the effect. Ibrahim, Les and Anjum will look at.

UNIMAS

Johari has a Raspberry Pi 2. It has double the RAM and a better processor. The previous one had a problem that it randomly rebooted and then it was not possible to ssh into it. the new one will go to the Data Center around May 11, 2015.

Johari still has to uncover the problem of the traceroute from UNIMAS. UDP has been unblocked. The MYREN host works fine and share most of the hops. Thus the problem must be in the first few hops.

Johari has a student starting on the custom iso (5/9/2015): Johari could get as far as the boot screen, but was unable to get to the desktop. This will be a few month project for the student.

They are also looking at anomaly detection: <http://slac.stanford.edu/pubs/slacpubs/13250/slac-pub-13399.pdf> or <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.363.1087> for comparisons of some techniques and http://people.cs.missouri.edu/~calyamp/publications/ontimedetect_mascots10.pdf. Next they will look at performance among correlated routes. There are quite a lot of papers in this area so a literature search is highly recommended.

UTM

Saqib will work on the Malaysia Case study and change it into a conference paper. in June he hopes to present it at an Advanced Network Workshop at UM in August.

The current case study is available in Google drive as a "Shared-PingER" document for review at https://drive.google.com/folderview?id=0B-NEKleLI79ZFNmUnhiVGJ0Nmc&usp=sharing_eid

Saqib points out a number of Malaysian routes are IPv6 which could have problems for traceroute. Saqib is checking

MYREN

Fizi reports that the MYREN PingER host at UTM was reported down on 22 March, it came back up again the next day.

He has been reading up about the PingER project.

- A goal would be to compare and contrast the benefits of perfSONAR and PingER.

NUST

Samad reports there are 13 hosts were not working, 5 are fine now. The IP address of one host changed. Samad will send information.

The following hosts are sometimes unable to be found via DNS when using ping, for example:

Host	Time PDT
pinger.uob.edu.pk	Apr 8, 11:01am, Apr 11, 11:00am
pinger.usindh.edu.pk	Apr 6, 1:04am, Apr 5, 1:05am Apr 4, 1:02am
pinger.isra.edu.pk	Apr 25, 12:58am Apr 6, 1:04am Apr 4, 1:02am
pinger.kohat.edu.pk	Apr 25, 12:58am Apr 24, 1:02am Apr 4, 1:02am

So far this month (30 April 2015) we have been unable to gather any data from the following Pakistani hosts:

- ns3.pieas.edu.pk pingable
- nuisb.seecs.edu.pk
- pinger-ncp.ncp.edu.pk
- pinger.cemb.edu.pk pingable
- pinger.lcwu.edu.pk
- pinger.nca.edu.pk
- pinger.pern.edu.pk
- pinger.numl.edu.pk pingable
- pinger.uum.edu.my
- pingerisl-fjwu.pern.edu.pk
- pingerlhr.pern.edu.pk
- pingerqta.pern.edu.pk
- sau.seecs.edu.pk
- www.upesh.edu.pk pingable

We have been unable to gather data from the following hosts for 50% or more of the days of April 2015, now below refers to 8:40am 4/30/2015 PDT

- airuniversity.seecs.edu.pk not pingable now
- comsatsswl.seecs.edu.pk not pingable now
- pinger.usindh.edu.pk not pingable now
- pingerisl-air.pern.edu.pk able to collect data now for entire month
- pingerisl-qau.pern.edu.pk not pingable now
- pingerlhr-pu.pern.edu.pk able to collect data now for entire month

For the latest see: <http://www-iepm.slac.stanford.edu/monitoring/checkdata/>

At the last meeting Samad was working on:

- www.upesh.edu.pk
- pingerlhr.pern.edu.pk
- pinger.pern.edu.pk
- ns3.pieas.edu.pk

- pinger.cemb.edu.pk
- pinger.nca.edu.pk

PingER has been installed at the naml node in Bolochistan. It is gathering data, it needs a tweak to its configuration

Pinger at SLAC

Working on the following hosts to be able to gather data:

Host	State	last seen	Status
hunnas.learn.ac.lk	emailed 2/26/2015, 5/2/2015	Nov 13, 2014	traceroute.pl works
multivae.sdsc.edu	emailed 2/26/2015, 3/6/2015. Fixed 3/9/2015	Jan 20, 2015	Fixed 3/9/2015
web.hepgrid.uerj.edu	emails 12/2/2014, 12/8/2014, 2/26/2015, 4/30/2015	Oct 23, 2014	traceroute.pl works but no response from ping_data.pl
www.umss.edu.bo	emails 8/30/2014, 9/12/2014, 11/27/2014, 2/27/2015, 4/30/2015. Disabled in NODEDETAILS as Monitor	Jul 6, 2014	No response from traceroute.pl
pinger.sesame.jo	email 3/14/2015. Fixed 3/16/2015, Fails again 3/17/2015, email 4/30/2015, Fixed 5/4/2015	Mar 4, 2015	BeaconList was missing, Does not ping
pinger.stanford.edu	email 3/14/2015	Feb 18, 2015	Works
pinger.fnal.gov	email 3/21/2015, 4/30/2015	Mar 18, 2015	Talked with Phil Demar of FNAL 4/3/2015. Pings, no web server
pinger@sonar-utm-myron.net.my	email 3/21/2015, Fixed 3/23/2015	Mar 9, 2015	Fixed 3/23/2015

Bebo arranged a meeting with the Colombia RENATA NREN folks and the minister of IT to discuss the use of PingER in Colombia. There is a web page at: [Colombia](#). Les has sent an email asking them to install pinger2.pl at at least one site in Columbia. Sent a reminder email 2/27/2015. Bebo will send a gentle reminder to the RENATA people of Columbia to see whether they continue to be interested and need a meeting.

Next meeting

Next meeting: Wednesday June 3rd 2015 9:00pm Pacific Standard Time, Thursday June 4th 2015 9:00am Pakistan time, Thursday June 4th 2015 noon Malaysian time, Thursday June 4th, 2015 02:00am Rio Standard Time.

Old Items

Traceroute at UTM 5/9/2015

The traceroute problem regarding maximum reachable hops (i.e. 11 hops) may be since the Unix/Linux/OSX traceroute uses UDP to send the requests. The first request is sent to a particular port (33434), with a ttl to tell it how many hops to go to. The ttl starts at 1 is incremented as it tries the next hop, also the port is incremented (up to 33465). It looks like the first few UDP ports are enabled and then they are blocked. The Windows traceroute uses ICMP to send the probes so does not see the problem.

Raspberry Pi 5/9/2015

The two major issues with the Raspberry Pi would be:

- are the results statistically the same as for the other monitor at UNIMAS (e.g. use the Kolmogorov-Smirnov test); There is Advanced Project (Master by coursework student) working on the statistics of the data from the raspberry Pi and the production PingER monitor at UNIMAS to see how much they differ.
- is it reliable/robust is it clear what to do to debug problems remotely (e.g. if it is at Bario). Looking at the monitoring data I have been unable to collect any from it (it is pingable, and port 80 responds, however the remote traceroute and ping_data.pl are not working) since Oct 20th which does not sound promising. Will need to evaluate the robustness of the unit by doing simulated scenario of various events such as power failure, hard and cold reboot, etc. Johari will need access to computer center to verify it comes up correctly after reboot etc.

If/when it works it would be instructive to look at the data from pinger and raspberry pi to Malaysia since the distances are shorter and the differences may show up better. For Sep-Oct 2014 when there was data measured from both Oct-Nov the averages for 20 paths was 52+-21ms (from pinger.unimas.my to 20 other Malaysian hosts) and 56+-21ms for raspberry pi to 20 other Malaysian hosts.

Linked Open Data

Feb 2015

The plan is still the one seen before (see [project proposal](#)), experimenting those alternatives. Right now, they managed to triplify the data according to a new ontology that takes advantage of a combination of a current standard for multidimensional data (called data cube vocabulary) and a revised version of Renan's Moment ontology adaptation. With this we expect to have a better data organization than the previous solution.

They are now preparing a test plan (like a small benchmark) to be used on all alternatives so that we can compare the results accordingly.

Aug 2014

Renan finished the new pingerlod web site. The new thing is that it should be much easier now to modify the info texts. What Renan did was to put the texts into a separate [file](#). The new version has been loaded on the server and some text added to describe how to use the map. However there is a bug that prevents it from executing the map. Renan reports that the bugs should be easy to fix. He has talked to his professor who suggested trying RDF OwlLink, it should have faster responses to queries. Renan will research this. It will probably mean reloading the PingER data so is a lot of work, hopefully this will improve performance. Before the rebuild he will make the fixes and provide a new WAR for us to load on pingerlod.slac.stanford.edu. He is also working on documentation (he has finished the ontology and has a nice interactive tool for visualizing it, since the ontology is the core of the data model of our semantic solution, this will be very helpful for anyone who uses our system, both a developer of the system and a possible user) and his thesis. Bebo pointed out that to get publicity and for people to know about the data, we will need to add pingerlod to lod.org.

Things he will soon do regarding documentation:

1. A task/process flow writing all java classes involved on all those batch jobs;
2. A Javadoc <<http://www.oracle.com/technetwork/java/javase/documentation/index-jsp-135444.html>> which will explain all classes and how they are used.

For the Linked Open Data / RDF which is in pre-alpha days, you can go to <http://pingerlod.slac.stanford.edu>. As can be seen this page is not ready for prime time. However the demos work as long as one carefully elects what to look at:

- Click on Visualizations, there are two choices:
 - Multiple Network Metrics: Click on the image: gives a form, choose from Node pinger.slac.stanford.edu pinging to www.ihep.ac.cn, time parameters yearly, 2006 2012, metrics throughput, Average RTT Packet loss and display format Plot graph, then click on submit. In a few seconds time series graph should come up. Mouse over to see details of values at each x value (year).
 - A mashup of network metrics x university metrics Click on image: gives another form, pinging from pinger.slac.stanford.edu, School metric number of students, time metric years 2006 2012, display format plot graph, click on submit. Longer wait, after about 35 seconds a google map should show up. Click on "Click for help." Area of dots = number of students, darkness of dots = throughput (lighter is better), inscribing circle color gives university type (public, private etc.) Click on circle for information on university etc.
- Renan will be working on providing documentation on the programs, in particular the install guide for the repository and web site etc. This will assist the person who takes this over.

Renan is using OWLIM as RDF Repository. He is using an evaluation version right now. Renan looked into the price for OWLIM (that excellent RDF Database Management System he told us about). It would cost 1200EUR minimum (~ 1620 USD, according to Google's rate for today) for a one time eternal license. It seems too expensive. No wonder it is so good. Anyhow, he heard about a different free alternative. Just not sure how good it would be for our PingER data. He will try it out and evaluate. He will also get a new evaluation of the free OWLIM lite.

He has also made some modifications on the ontology of the project (under supervision of his professor in Rio) hence he will have to modify the code to load the data accordingly.

Maria and Renan are advancing in some approaches to deal with PingER data, making it easier to be analyzed and integrated. In particular they have been busy studying and evaluating alternatives, analyzing results from the latest benchmarks on NoSQL (including RDF and graph based storages) database management, distributed processing and mediated solutions over relational databases, and also other experiments with multidimensional analyses on Linked Data. The new students involved are now understanding better the scenario and they have been interacting with Renan regularly.

UM

Moved here 3/4/2015:

Ibrahim has setup distributed hadoop clusters. He has 2TB of disk space. Les has provided information on getting a subset of PingER data by anonymous ftp via <ftp://ftp.slac.stanford.edu/users/cottrell>. It was put there last September. Information on how the data was put together is at <https://confluence.slac.stanford.edu/display/IEPM/Archiving+PingER+data+by+tar+for+retrieval+by+anonymous+ftp>. There is information on formatting etc at <http://www-iepm.slac.stanford.edu/pinger/tools/retrievedata.html> and some on the dataflows at <https://confluence.slac.stanford.edu/display/IEPM/PingER+data+flow+at+SLAC>. Renan at UFRJ has successfully used this data, he has also characterized the data in terms of bytes/metric per year etc.

Ibrahim has started downloading all zip files in the local machines. 6 weeks ago he downloaded 2 GB of Weather data to test his nodes cluster, he wrote a simple Java program (Map, Reduce) to find the Average and it was working fine.

NUST

The following is from Samad 2/24/2015.

- buitms.seecs.edu.pk #We have to disable gathering data from this host because the person still don't want to continue with us as i have tried once again to convince him but the answer is same. Les has disabled from SLAC.
- nukhimain.seecs.edu.pk # We were unable to gather data since 20th November, 2014 and now the Node is working fine and collecting data as well.
- pinger.uetaxila.edu.pk #The node is working fine from last two weeks.
- sau.seecs.edu.pk #This Node is working fine now.
- pingerjms.pern.edu.pk #This node is working now.
- pinger.uet.edu.pk # this was also not working from so many days. and now its working fine and collecting data as well.

- pinger.isra.edu.pk # This node is also working fine now.
- pingerlhr-pu.pern.edu.pk # This is also working fine now.
- pinger.kohat.edu.pk # Collecting data now.
- The IP of "pingerqta.pern.edu.pk" has been changed, Les has updated the databas at SLAC with the following

DNS: pingerqta.pern.edu.pk

Old IP: 121.52.157.157

New IP: 121.52.157.148

Follow up from workshop

- Hossein Javedani of UTM is interested in anomalous event detection with PingER data. Information on this is available at <https://confluence.slac.stanford.edu/display/IEPM/Event+Detection>. We have sent him a couple of papers and how to access the PingER data. Hossein and Badrul have been put in contact. Is there an update Badrul?

The Next step in funding is to go for bigger research funding, such as LRGS or eScience. Such proposals must lead to publications in high quality journals. They will need an infrastructure such as the one we are building. We can use the upcoming workshop (1 specific session) to brainstorm and come up with such proposal. We need to do some groundwork before that as well. Johari will take the lead in putting together 1/2 page descriptions of the potential research projects.

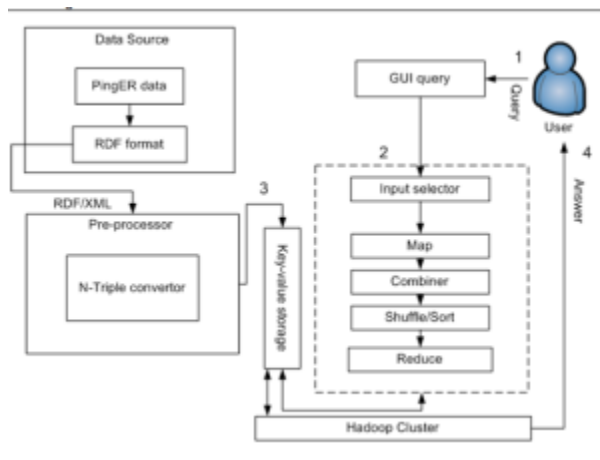
1. Need to identify a few key areas of research related to PingER Malaysia Initiative and this can be shared/publicized through the website. These might include using the infrastructure and data for: anomaly detection; correlation of performance across multiple routes; and for GeoLocation. Future projects as Les listed in Confluence here <https://confluence.slac.stanford.edu/display/IEPM/Future+Projects> can also be a good start and also Bebo's suggestion.
2. Need to synchronize and share research proposals so as not to duplicate research works. how to share? Maybe not through the website, or maybe can create a member only section of the website to share sensitive data such as research proposal?

Anjum suggested Saqib, Badrul and Johari put together a paper on user experiences with using the Internet in Malaysia as seen from Malaysian universities. In particular round trip time, losses, jitter, reliability, routing/peering, in particular anomalies, and the impact on VoIP, throughput etc. It would be good to engage someone from MYREN.

Ibrahim

Ibrahim Abaker is planning to work on a topic initially entitled "leveraging pingER big data with a modified pingtable for event-correlation and clustering". Ibrahim has a proposal, see <https://confluence.slac.stanford.edu/download/attachments/17162/leveraging+pingER+big+data+with+a+modified+pingtable+for+event-correlation+and+clustering.docx>. Ibrahim reports 7/15/2014 "I have spent the last few months trying to understand the concept of big data storage and its retrieval as well as the traditional approach of storing RDF data. I have integrated a single hadoop cluster in our cloud. but for this project we need multiple clusters, which I have already discussed with Dr. Badrul and he will provide me with big storage for the experiment." No Update 8/20/2014.

"I have come up with initial proposed solution model. This model consists of several parts. The upper parts of the Figure below shows the data source, in which PingER data will be convert into RDF format. Then the data pre-processor will take care of converting RDF/XML into N-triples serialization formats using N-triples convertor module. This N-triple file of an RDF graph will be as an input and stores the triples in storage as a key value pair using MapReduce jobs"



Potential projects

[See list of Projects](#)

Coordinates of team members:

See: <http://pinger.unimas.my/pinger/contact.php>