

# 20160106 SLAC SEECS and UNIMAS Meeting

## Time & date

Wednesday Jan 6th 8:00pm Pacific Standard Time, Thursday Jan 7th, 2016 9:00am Pakistan time, Thursday Jan 7th 2016 12:00noon Malaysian time, Thursday Jan 7th 2016 02:00am Rio Standard Time.

## Coordinates of team members:

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

## Attendees

### Invitees:

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

Note returned to original time on the hour since no longer need to accomodate Johari.

+ Confirmed attendance

- Responded but Unable to attend:

? Individual emails sent

### Actual attendees:

Hassaan, Saqib, Aqsa, Johari, Adnan, Anjum, Fatima, Les, Bebo

## Administration

- Membership of pinger-my in <https://groups.google.com>
  - Aqsa Hameed has been added as of Dec 20, 2015. Johari has added her to <http://pinger.unimas.my/pinger/contact.php>.
- This will most likely be Anjum's last meeting. He has already resigned from the University of Malaya. He will be leaving Malaysia on 31st January or 1 February. He will stay for 15 days in Pakistan then travel to Canada. We will miss him, and wish him all the best in his new startup.
- Huffington Post article on PingER and African Internet performance, see [http://www.huffingtonpost.com/david-tereshchuk/a-giant-leap-in-2016-africa\\_b\\_8901556.html](http://www.huffingtonpost.com/david-tereshchuk/a-giant-leap-in-2016-africa_b_8901556.html)
  - Anjum said it would be good for Hassaan to attach this to his proposal.
  - Les will contact SLAC communications to get it in SLAC Today.
- NETAPPS2015 1-3 December 2015 was very successful, five papers were presented in the PingER session
  - Currently, NetApps organizers are in the process of sending NETAPPS2015 proceedings to be indexed in the major indexing services.
- 9 students from Amity University near New Delhi in India interested in coming to SLAC. Bebo sent reminder recently, response below:
  - Sorry for the delay. Students are giving end semester exams. They are busy till dec 20th. They will get back to work as soon as the exams are done. Once the exam is over we will be setting up "PingER measurement agent at Amity"
  - Three students have each written a short paper for the [6th International Conference on Cloud System and Big Data Engineering at Amity Jan 14-15, 2016](#). I have reviewed and made some corrections
    - 443: Analysis and clustering PingER network data
    - 463: Application for the emulation of pinger on Android devices
    - 432: Correlation analysis on Real Time Tab Delimited Network monitoring data
  - Btech computer science students would be graduating in May 2016. Their visas can be arranged after May 2016.
- Joao a Brazilian student identified by Renan is interested in a 3 month internship at SLAC similar to those completed by Renan and Thiago.
  - Have letter of Support for Academic Training from his organization. Les filled out form for Joao.

### Geolocation Anjum,

- This is a proposed project for one of the Indian students.

**Bebo has set up a Github codebase as a new project. It contains the PingER MA (pinger2.pl and the traceroute/ping server). Anyone needs to sign up for a Github account (if you don't already have one), so you can be added as a project member.**

- slac-pinger/pinger created by topherwhite. PingER project <https://github.com/slac-pinger/pinger>
- **Now we have it we can share with Amity to check it works.**
  - **Bebo reports that someone from Amity requested to be a member of the Android group.**

## Saqib

- Saqib has a 5 MS students from the Database team
  - Sara Massoud came up with ideas to improve the info put together by Amity
  - Sabah Massoumil working on Linked Open Data (please excuse spelling)
  - Aqsa Hameed working on big data/analysis of PingER data. The initial idea was to set up a big data /big analysis
  - Aqsa Hameed has been working with Anjum to look at a project to create a Hadoop/Cloudera PingER warehouse to enable easier, more powerful access to PingER historical data.
- Thinking a bit on Aqsa her work appears to be related to the work done by Thiago on PingER warehouse using a cluster/Cloudera/HDFS/Impala earlier this year. In particular see the presentation at NETAPPS2015 (see [https://confluence.slac.stanford.edu/download/attachments/123309267/NETAPPS\\_PRESENTATIONrevLuiza.pptx](https://confluence.slac.stanford.edu/download/attachments/123309267/NETAPPS_PRESENTATIONrevLuiza.pptx)). Once the paper is published we can also provide her with that (**Adib will let us know when this is OK**).
- **Les emailed the relevant people to put them in contact with one another. It appears there is a lot of overlap between what Aqsa proposed as a Masters project and what Thiago has already done. However Thiago's system was mainly a proof of concept and not in production. We need to look at the next steps: Internet access, auto-updating of information in near realtime, production service, maintainable etc. With this in place one could really mine the data looking for all kinds of interesting correlations, clustering, event impacts, comparisons etc.**
  - **Given that such a warehouse is available, then the next step would be to automatically create queries that would produce in near real-time the plots we produce manually for the PingER annual reports.**
  - **Following this publish the data in RDF to tie in with the Southampton RDF web observatory repository (similar to what Renan did as a proof of concept).**
  - **Since Thiago's warehouse is only available at SLAC, it may assist to get Aqsa an account at SLAC, alternatively Aqsa will need to set up a repository at here home base.**
  - **There is some documentation written by Les on the usage of the warehouse at SLAC. It is at: [PingER Data Warehouse using Big Data with Cloudera on Nebula](#).**
  - **Les will contact Thiago to see if there is other documentation or where to find the programs etc.**
  - **We need to put together all we know:**
    - **Paper from Thiago on PingER warehouse presented in Malaysia at NETAPP2015 Thiago M. Da S. Barbosa, Renan F. Souza, Sergio M. S. Da Cruz, Maria L. Campos and R. Les Cottrell. [Applying Data Warehousing and Big Data Techniques to Analyze Internet Performance](#)**
    - **Document on Warehouse**
      - **[PingER Data Warehouse using Big Data with Cloudera on Nebula](#)**
    - **Documents on PingERLOD**
      - **[Possible follow on to PingERLOD project, PingER#PingERLOD](#)**

- [Linked Open Data Publication Strategies: Application in Networking PPerformance Measurement Data](#), Renan Souza, Les Cottrell, Bebo White, Maria Campos, Marta Mattoso, poster presented at the [BigDataScience - Stanford](#) conference, CA, USA May 27-31, 2014.
  - **Reviews/proposals/ from Aqsa/Fatima**
    - [Survey on Big Data Indexing strategies](#), Fatima Bintu Adama, Adib Habbal, Suhaidi Hassan, R. Les Cottrell. Bebo White, Ibrahim Abdullahi.
    - Saqib will provide Aqsa's
  - **Tehseen is working on missing PingER data.**
- Saqib has submitted a project with title "A Fundamental Active Internet Performance Monitoring Framework for Pakistan Education & Research Network (PERN) in University of Agriculture, Faisalabad"
  - Project is accepted.
  - **working on a project to develop a node using Raspberry Pi 2 and IoT to measure the air and soil quality.**
  - **Have 2 RPis and setting up air quality with sensors from market, idea is to distribute on campus.**
  - **Currently, students are analyzing the project to develop a problem statement for their research project.**
- For the GeoLocation Saqib should contact Anjum
- **Project on Internet performance has been accepted, but no funding yet.**

## UUM

Fatima presented her work at NetApps. Fatima has passed her master oral examination (Viva) with minor correction and she is submitting her final dissertation coming Sunday to UUM school of computing. Indeed, we are very grateful to PingER team members support and help. Currently, Fatima is planning to pursue her PhD and possibly on a topic related to her master dissertation. We need PingER team suggestions and advice before we can decide on the topic. **Adib will attend next meeting to further discuss on this.**

## UM

Ibrahim has extracted the PingER Zip manually. He is reconstructing 11Gb of data, has 15GBytes of data there. He is trying to use SPAR to classify the data. He also was looking at RDF. the next step is to use MapReduce to organize and reduce the output of data so can visualize it. He will be using the same techniques he used for looking at 1996-2006 weather data. However at the moment he cannot access his VMs from myren for the last two weeks, even for the myren site, He has emailed them but till now they have not fixed the issue. He has updated all his data in their cloud. **No update 1/6/2016**

## UNIMAS

Custom ISO still in progress, unable to release it due to bug that causes kernel panic error when booting from customized iso. Update 1/6/2016.  
 Random seem OK on Intel but not on AMD. Still debugging  
 Will revisit traceroute server.  
 Update PingER Malaysia site.  
 Get back to RPi2 at Datacenter  
 Latest RPi0 for \$5. Spec 2\*good of RPi2

## UTM

Johari has had no response from UTM so far. Not sure whether they still want to be in the project. Last straw, need to contact someone from the network group so that they understand and appreciate the Pinger project. Update 1/6/2016.

Saqib please contact Dr Md Asri Ngadi. Will provide update after talk.

Johari will try and contact Hafizi

Saqib reports:

"Faculty management in the Faculty of Computing-UTM is changed. They ask me to re-locate the PingER server from the faculty server room as students are not allowed for physical access of the server room. Only faculty members are allowed to physically manage the servers in the server room. Therefore, I asked my supervisor (Dr Md Asri Ngadi/ Prof Hanan) from UTM to take the responsibility of the server. He is agreed, however, I have to manage the server remotely. Further, he needs an official invitation from your side to fulfill the official requirements. Further, he wants to make this project as a source of an international collaboration and linkages between UTM and SLAC-Stanford."

Les sent back:

The UTM PingER server is part of a worldwide collaboration led by SLAC that includes about 70 sites in over 20 countries. These sites are monitoring on-going Internet end-to-end performance, both currently and in the past with a history of 18 years of collected data.

SLAC is happy to host postgraduate students from UTM for a period of up to a year to contribute to this project as visiting scientists.

The students participate in collaborative research and development on the Internet End-to-end Performance Measurement (IEPM) project to monitor, analyze and present information on the end-to-end performance of Internet links. This project includes researching and developing techniques to optimize the measurements for the purposes of characterizing the "Digital Divide". Dr. Les Cottrell of SLAC's Office of the Chief Information Officer is the host. All support for the visit is provided by sources other than SLAC.

The students are expected to comply with all policies and procedures, including safety guidelines. Your mentor will provide them with any specific policies, procedures or safety guidelines for their assigned area. As a program participant, the student does not enter an employee/employer relationship with DOE, SLAC or any other federal agency.

## MYREN

No update 8/12/2015, no Update 9/2/2015.

[perfonar-unimas.myren.net.my](http://perfonar-unimas.myren.net.my) is back up and running

We have been unable to gather data from [pingersonar-utm.myren.net.my](http://pingersonar-utm.myren.net.my) since September 11th, 2015 [pingersonar-utm.myren.net.my](http://pingersonar-utm.myren.net.my). It is pingable however there is no response from <http://pingersonar-utm.myren.net.my/cgi-bin/traceroute.pl?function=ping&target=www-wanmon.slac.stanford.edu&options=-i%200.2>

## NUST

Hassaan reports: "There are total of 48 active nodes for Pakistan out of which 18 nodes are working properly whereas 4 nodes have configuration issues. Lahore college for Woman has changed its data center and they want to know what do have to do with the new one. Faisalabad PERN had some critical issues and burnt their power supply so they also wanted to know what can be done to overcome that damage. FAST University Karachi and Air University, Islamabad are not in a mood to help as they say that they will put the node up, but haven't done that in past 2 months [Hassaan would like to discuss this with Dr. Anjum for excluding them from the list]. Last but not the least Quetta HEC regional center wanted to change their static IP to 121.52.157.148."

Les has Disabled the following Pakistani monitors since they have no DNS name and are not responding by IP address. [pinger.numl.edu.pk](http://pinger.numl.edu.pk), [pinger.uet.edu.pk](http://pinger.uet.edu.pk), [pinger.uob.edu.pk](http://pinger.uob.edu.pk), [pinger.usindh.edu.pk](http://pinger.usindh.edu.pk). We are still monitoring hosts at the site.

## Pinger at SLAC

- Working on the following hosts to be able to gather data
- Pinger is now on a VM.

Host	State	last seen	Status
<a href="http://web.hepgrid.uerj.edu">web.hepgrid.uerj.edu</a>	emails 5/1/2015	Dec, 2014	does not ping electrical problems
<a href="http://pinger.sesame.jo">pinger.sesame.jo</a>	email 1/6/2016		Does not ping, server works
<a href="http://Pinger.stanford.edu">Pinger.stanford.edu</a>	email 1/6/2016	May 2015	does not ping

## Next Meeting

Next meeting: Wednesday Feb 3 8:00pm Pacific Standard Time, Thursday Feb 4, 2016 9:00am Pakistan time, Thursday Feb 4 2016 12:00noon Malaysian time, Thursday Feb 4 2016 02:00am Rio Standard Time.

## Old Items

- Jan 5, 2015 Hassaan reports "I have received revisions on my proposal and these days I am revising my proposal. In the meanwhile, I have also added another student (Anas Abrar) on this project. He is in learning phase and will follow the nodes which are not working. I shall give you an update very soon."
- Hassaan is very hopeful that if the proposal is accepted then we can easily have a full time RA for the project.

Hassaan has re-submitted the proposal after revisions. He would like to get Anas Abrar more trained on monitoring operation and then will inform us to add him to the mailing list at <http://pinger.unimas.my/pinger/contact.php>.

- Oct 2015. Following the last meeting, Anjum, Hassaan and Les met to discuss the way forward.

- "Adnan currently is unable to find resources for handling the project. Similarly, there is no progress on hiring of a full time RA by NUST HQ.
- However, I (Hassaan) checked from HEC about the proposal that I submitted last year. They have informed me that 2 reviewers have asked for revisions while they are waiting for the third review. I am very hopeful about it. If the proposal is accepted then we can easily have a full time RA for the project. I have plans to talk to Dr. Zaidi about hiring an RA on assuming that our proposal will be accepted by HEC. We can then get his salary deducted later from the HEC project. I shall update you very soon in this regard."
- Hassaan is waiting to hear from HEC about the comments on the proposal.
- Moreover, he has asked a student to work on the project for the time being. His name is Mian Anas however he will need few weeks to understand the project.
- Thiago completed setting up the PingER data SQL Impala warehouse running on a Nebula/Cloudera cluster using the Hadoop File System (HDFS). Unfortunately it is not currently accessible from outside SLAC. There have been several attempts to provide outside access, but no success yet, we need to engage the subject matter experts. Thiago is now a SLAC associate so he still has an account at SLAC. There was a cyber security alert on the version of java installed with Cloudera. Les has replaced the cloudera version of java which should fix the vulnerability. However the new version has not been tested.

## Geolocation

Anjum believes the TULIP Geolocation application can be improved significantly. At least there are few ideas that we can try. For this, either a group of undergraduate students or an active masters student is required. The resultant work can easily be the thesis of masters level. Who is interested?

- Saqib at Faisalabad has an MS student interested to work on Geolocation project. He requests an initial paper on the project. Les has responded to Saqib. He also has some other students. Anjum will contact him. Potential projects/asks include: take over management of PingER monitoring in Pakistan (say 5 monitors/student; case study of how Pakistan's network performance/connectivity has improved over the years especially as function of funding etc; geolocation with variable alpha; indoor geolocation
- Johari will contact Anjum to learn more of the requirements. Update Johari/Adnan
- See <http://www.slac.stanford.edu/comp/net/tulip/>. Basically TULIP uses pings to a target from landmarks at known locations and converts the minimum RTTs to estimate the distances. Then uses the distances with multilateration to estimate the location of the target
- To improve TULIP one needs the right selection of landmarks, i.e. good (working landmarks) at the right locations (not too far from the target), straddling the target, and with a reasonable estimate of the indirectness (directivity or alpha) of the path from the landmark to the target (so we can reasonably accurately estimate the distance). One also needs a reasonable density of landmarks (e.g. number of targets/100,000sq km)
- The landmarks come from PingER and perfSONAR sites. We have a reasonable density in the US, Pakistan and Europe. Currently Anjum is getting better than 20km accuracy for Pakistani targets
- As the number of landmarks goes up so does the accuracy, but so does the time to make the measurements (pings).
- One needs to find the optimal density
- Anjum proposes to speed up the measurements using a cluster for parallelization and also proposes to improve the adaptation of alpha based region. He regards the adaptive geolocation and parallelization as MS projects.
- He is also interested in geolocation in small proximity (e.g. indoors), e.g. using cell tower signals. This is a new area of research. It is possible that the port of PingER to an Android could be related to this. This is a PhD project
- Anjum reports he can supervise the students on Geolocation. He will need to know when the students are ready. We can start with a joint meeting involving Les and the students. Later on, Anjum can have the meeting with students every week while Les can join if he has time.

## NUST/SEECS Pakistani PingER nodes status

Pink Background indicates host was bad last month, strike through says it is fixed, yellow is an new bad host.

Current status of Pakistani Hosts 7/1/2015

1. <a href="http://airuniversity.seecs.edu.pk">airuniversity.seecs.edu.pk</a>	Down	Called (Person Not Responding).
<del>2. <a href="http://comsatscw1.seecs.edu.pk">comsatscw1.seecs.edu.pk</a></del>	<del>Down</del>	<del>Called (Link Issue)</del>
<a href="http://ns3.pieas.edu.pk">ns3.pieas.edu.pk</a>	Pingable	
3. <a href="http://nuisb.seecs.edu.pk">nuisb.seecs.edu.pk</a>	Down	Called (Not Responding)
4. <a href="http://nukhimain.seecs.edu.pk">nukhimain.seecs.edu.pk</a>	Down	Called (Will be up within two days)
5. <a href="http://pinger.cemb.edu.pk">pinger.cemb.edu.pk</a>	Pingable	Called (Need Access)
6. <a href="http://pinger.kohat.edu.pk">pinger.kohat.edu.pk</a>	Down	Email sent to the concern Person (DNS Entry issue)
7. <a href="http://pinger.lhr.nu.edu.pk">pinger.lhr.nu.edu.pk</a>	Down	Called (Person Not Responding)
8. <a href="http://pinger.lcwu.edu.pk">pinger.lcwu.edu.pk</a>	Down	Working now?
9. <a href="http://pinger.nca.edu.pk">pinger.nca.edu.pk</a>	Down	Called (Will be up within two days)
10. <a href="http://pinger.numl.edu.pk">pinger.numl.edu.pk</a>	Pingable	Need Visit
11. <a href="http://pinger.pern.edu.pk">pinger.pern.edu.pk</a>	Down	Need Visit
12. <a href="http://pinger.usindh.edu.pk">pinger.usindh.edu.pk</a>	Down	Called (Person Not Responding)
13. <a href="http://pingerisl-fjwu.pern.edu.pk">pingerisl-fjwu.pern.edu.pk</a>	Down	Need Visit
<a href="http://pingerisl-qau.pern.edu.pk">pingerisl-qau.pern.edu.pk</a>	Down	
<a href="http://pingerkhi.pern.edu.pk">pingerkhi.pern.edu.pk</a>	Down	

pingerlhr.pern.edu.pk	Down	
14. <a href="http://pingerqta.pern.edu.pk">pingerqta.pern.edu.pk</a>	Pingable	Email sent to the concern person (DNS Entry Issue)
15. <a href="http://www.upesh.edu.pk">www.upesh.edu.pk</a>	Pingable	Called (Person not cooperating)
sau.seecs.edu.pk	Down	

Is it time to start paring down the list of PingER monitor hosts in Pakistan, starting with those that have been down for a while and despite your efforts they are not cooperating. One might also look at the coverage by region in Pakistan and try and keep good coverage for all regions.

#### 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.

#### Linked Open Data

Cristiane reports (7/1/2015): "I am trying to automatize the triplification of PingER data on Kettle. For now, part of the transformation is made on Kettle and another is made by a Java code. Although this solution works for a data sample, is important to have the entire process on Kettle because it facilitates to understand, modify and control the triplification process."

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.



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  $15.66 \times 10^9$  (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](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](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](#)

## 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.

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. Les analyzed the data and sent it to Anjum

## 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.uettaxila.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.

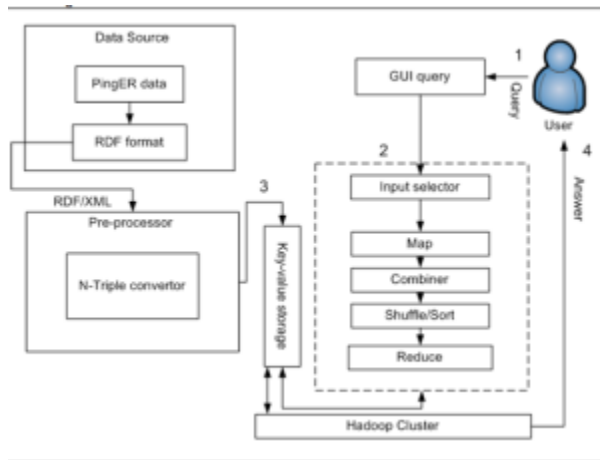
1. Need to identify a few key areas of research related to PingER Malaysia Initiative and this can be shared/publicized through the website. These might include using the infrastructure and data for: anomaly detection; correlation of performance across multiple routes; and for GeoLocation. Future projects 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](#)