

20150603 SLAC SEECS and UNIMAS Meeting

Time & date

Wednesday June 3rd 2015 9:00pm Pacific Daylight 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.

Coordinates of team members:

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

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); Thiago+, Les+, Bebo+ (SLAC)

+ Confirmed attendance

- Responded but Unable to attend:

? Individual emails sent

Actual attendees:

Anjum (in Canada), Johari, Ibrahim, Thiago, Les, Bebo. Saqib was online but we were unsuccessful in contacting him by Skype.

Administration

- [Membership of pinger-my](https://groups.google.com) in <https://groups.google.com>
- Thiago Barbosa, a student from the Rural Federal University of Rio de Janeiro, who was in New York with a Science without Border grant has been invited to SLAC from June 1st to August 2015. He is now at SLAC. He will probably be working on big data analysis for PingER. I have added him to the pinger-my email list. We need to add him to the contacts (Johari). His email address is thiago.marcos.13@gmail.com, his Skype ID is Thiego Barbosa.
- Arshad has left SEECS/NUST to become the Rector of the National Textile University in Faisalabad. We have contacted the Rector at NUST to discuss continued support for PingER at SEECS. He has been very supportive and responded as follows:

We will surely continue with PingER project. Dr Zaidi who has succeeded Arshad has spent as many years in NIIT and later SEECS as Arshad. He will take care of the project.

I have copied this email to Director Research and Director Academics too. They will fully support Zaidi. Hope your worries are removed. If you find any issues, please give an email to me and I will ensure that we achieve all the targets of this project. Thanks and regards. I have emailed Dr Zaidi.

Acting principal (Engineer Habeel) asked for details about project that Anjum provided. They have also contacted Hassaan and asked him to be faculty in-charge of the project. Dr. Zaidi and Habeel Ahmed are not clear of the status and direction of the project so Dr. Arshad told them he can discuss it during his next visit of SEECS. That is what I last know. I think Dr. Arshad visits SEECS every 15 days or so and is helping with smooth transition.

Dr Zaidi writes:

Thank you for your email and for posing confidence in NUST-SEECS for the collaboration that has been on-going for a number of years and has only grown stronger with time. It is indeed a pleasure for us here at NUST-SEECS to be working in collaboration with your team at SLAC. We certainly value this collaborative effort and look forward to take it to newer heights.

I have recently conducted a special review meeting to ensure that things are back on track. Through this email, I would like to reiterate our commitment to this collaboration. We assure you of our full support for taking this a step further. As for your requirement of a System Administrator and a Faculty member in-charge of this project, you will be happy to know that Dr Hassan Khaliq -- a faculty member at SEECS who is already involved in the project, has been made responsible to manage the affairs from our end. In addition, we have deployed an MS research student to assist him in this effort. We have also raised the requirement of a full-time resource for this project and we are hoping to have this resource with us soon

Please do not hesitate to contact me directly should you face any problem from this day forward.

- Anjum How are 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 not started on this yet, he hopes to get to it when he returns from Canada**
- 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.'

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 wrote 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

Adib reports 5/23/2015:

"I did follow up this issue personally as I promise. I spent the whole day on Thursday (07 May 2015) with the network engineer from the computer center to troubleshoot this connectivity issue. I was hoping to report good news last meeting, unfortunately, without any tangible output. In fact, there is no issue at my side with UUM Pinger server, except the connectivity. Even UPS, I am waiting to get one soon to avoid the electricity problem.

As a result of this, we find out that there is a problematic switch connects UUM Pinger server with the center. I can not do much here cos it may involve buying a new switch/approval from their boss. I am waiting for update and follow up on this."

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 had problems loading <ftp://ftp.slac.stanford.edu/users/cottrell>. She has decided to get the data on an external hard drive.**

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-NEKleLII79ZFNmUnhiVGJ0Nmc&usp=sharing_eid. It is incomplete and needs updating. Anjum will find a new PhD student to look at this.

UM

Ibrahim had downloaded PingER in Zip files format, however, when he stored them in the Hadoop distributed file system (HDFS) and try to process them, the file got corrupted, so he had to extract the file, but one file zip has more than 10000 zip files with small size. **So he is trying to create a mapreduce job which can accept zip format, that will save lot of his time. Currently mapreduce can only read from files like .txt, and any doc file format or database. He will have meeting with Dr. Anjum on 11 of june asking for advice and seeking of how we can work on this together.**

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 5:30-7pm in the evening (Malaysia time). He suggested looking to see if PingER could spot the effect. **Les provided the UM host ping responses for the relevant period. Anjum did the very basic analysis of the RTT times and found that there RTT as well as packet loss was higher during 11th march until 13th march at the times of attack. PTM at UM believed that attack was being launched for a period of roughly 2 hours at the times 12noon and 6pm. The data showed other times, almost 6 hours periodic repetition. Data shows the problem remained for 3 days. There were packet drops at other random times before and after the attack, which cannot be related to the attack. However, such drops are unusual for a well established site like UM. Anjum asked Ibrahim to look further into data but has not heard further from him.**

Anjum pointed out he can probably assist with getting pinger.fsktm.um.my back running. It has been down since 27th May, 2015.

UNIMAS

Johari has a Raspberry Pi 2. It has double the RAM and a better processor. **It will go to the Data Center in June.**

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 for 6 months. He will working on the custom iso (6/3/2015):

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

Johari will contact Hanan to request someone to support PingER at UTM, now Saqib has left. (6/3/2015),

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

No update 6/3/2015

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

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

List of Hosts Status Comments

Current status of Pakistani Hosts 6/3/2015.		
1. airuniversity.seecs.edu.pk	Down	Called (Person Not Responding)
2. comsatsswl.seecs.edu.pk	Down	Called (Link Issue)
3. nuisb.seecs.edu.pk	Down	Called (Not Responding)
4. nukhimain.seecs.edu.pk	Down	Called (Will be up within two days)
5. pinger.cemb.edu.pk	Pingable	Called (Need Access)
6. pinger.kohat.edu.pk	Pingable	Email sent to the concern Person (DNS Entry issue)
7. pinger.lhr.nu.edu.pk	Down	Called (Person Not Responding)
8. pinger.lcwu.edu.pk	Pingable	Working now
9. pinger.nca.edu.pk	Down	Called (Will be up within two days)
10. pinger.numl.edu.pk	Pingable	Need Visit
11. pinger.pern.edu.pk	Down	Need Visit
12. pinger.usindh.edu.pk	Down	Called (Person Not Responding)
13. pingerfjwu.pern.edu.pk	Down	Need Visit
14. pingerqta.pern.edu.pk	Pingable	Email sent to the concern person (DNS Entry Issue)
15. www.upesh.edu.pk	Pingable	Called (Person not cooperating)

In addition pinger.isra.edu.pk does not respond by name.

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.

PingER at SLAC

Bebo, Thiago & Les met to go over possible projects for Thiago.

1. Extend the work done by Renan to provide open analysis using Linked Open Data of PingER data. Part of this would be to publish and provide querying of PingER data through linkdata.org. A major requirement for this is a Hadoop cluster. Les will investigate. This may be a block.
 - a. Thiago is looking at Impala, Les will see if we have access to a Hadoop cluster.
 - b. Thiago and Renan have scheduled a Skype meeting to discuss ways forward.
2. Configure a Raspberry Pi as a PingER monitor. Look at reliability and compare with a regular PingER monitor. Bebo Provided a Raspberry, Les has provided a SD drive.

We have put together a case study on the impact of the Nepalese earthquakes on the Internet connectivity, see [Nepal Earthquake 2015](#). It is interesting that the connectivity continued for an hour or so after the shock wave. Caltech points out that this could be important for projects measuring the impact of the initial shockwave.

Working on the following hosts to be able to gather data (Bold face = change since last meeting, green background fixed, red background = disabled, strike through completed at a previous meeting):

Host	State	last seen	Status
hunnas.learn.ac.lk	emailed 2/26/2015, 5/2/2015. Disabled as Monitor still a Beacon 5/10/2015.	Nov-13-2014	traceroute.pl works
multivac.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, 6/1/2015	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 NOEDDETAILS 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. Moved to VM with new name pinger-host.fnal.gov, enabled ICMP request/ response and port 80, works 5/19/2015.	Mar 18, 2015	Talked with Phil Demar of FNAL 4/3/2015. Pinge, no web server
pinger.sonar-utm-myron.net.my	email 3/21/2015, Fixed 3/23/2015	Mar 9, 2015	Fixed 3/23/2015
pinger.unesp.br	email 11/28/2014, 5/22/2015, 6/1/2015.	Nov 3, 2014	Host is pingable from SLAC.
pinger.daffodilvarsity.edu.bd	email May 30th, Fixed June 1st, 2015	May 25th	Host not pingable
pinger.fsktm.um.edu.my	email May 29th, June 6th, host in need of maintenance, Anjum may be able to assist.	May 24th	Host not pingable
ping.riken.jp	email May 31, 2015, replaced DIMMS, Fixed 6/2/2015, but may still have a problem	May 22nd	Host not pingable

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 July 1st 2015 9:00pm Pacific Standard Time, Thursday July 2nd 2015 9:00am Pakistan time, Thursday July 2nd 2015 noon Malaysian time, Thursday July 2nd 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.

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.

- buittms.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.

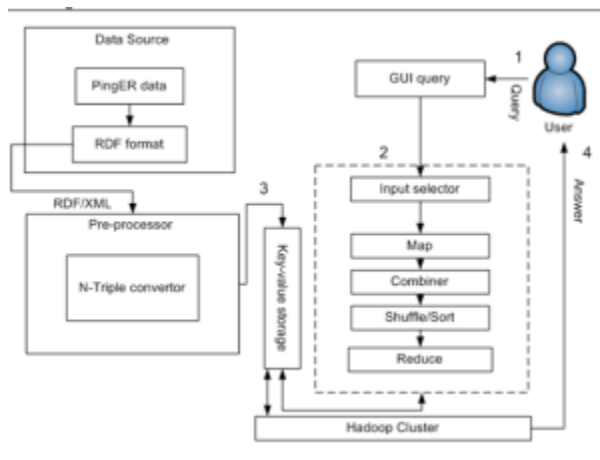
1. Need to identify a few key areas of research related to PingER Malaysia Initiative and this can be shared/publicized through the website. These might include using the infrastructure and data for: anomaly detection; correlation of performance across multiple routes; and for GeoLocation. Future projects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