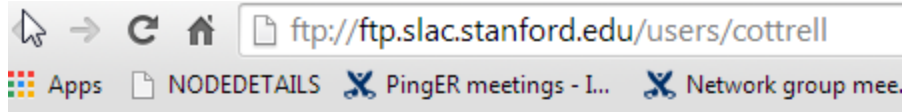


# Retrieving Archived PingER hourly data from Anonymous FTP

## Using a Web browser

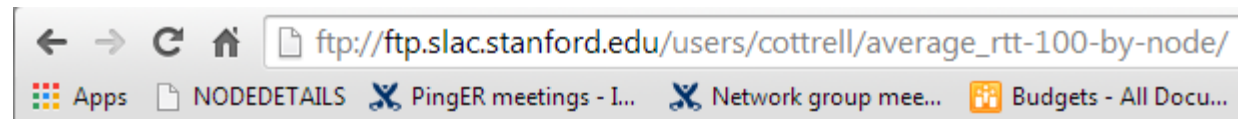
Opening the ftp site using Chrome gives a result similar to below:



## Index of /users/cottrell

Name	Size	Date Modified
[parent directory]		
MOS-100-by-node/		9/11/14, 11:27:00 PM
MOS-100-by-node.tar	489 MB	9/11/14, 11:33:00 PM
MOS-1000-by-node/		9/16/14, 3:47:00 AM
MOS-1000-by-node.tar	234 MB	9/16/14, 3:49:00 AM
alpha-100-by-node/		9/11/14, 11:50:00 PM
alpha-100-by-node.tar	720 MB	9/12/14, 12:00:00 AM
alpha-1000-by-node/		9/16/14, 3:55:00 AM
alpha-1000-by-node.tar	339 MB	9/16/14, 3:59:00 AM
average_rtt-100-by-node/		9/11/14, 5:33:00 PM
average_rtt-100-by-node.tar	1.5 GB	9/11/14, 8:58:00 PM
average_rtt-1000-by-node/		9/15/14, 5:36:00 PM
average_rtt-1000-by-node.tar	1.4 GB	9/16/14, 2:12:00 AM
conditional_loss_probability-100-by-node/		9/12/14, 1:55:00 AM
conditional_loss_probability-100-by-node.tar	280 MB	9/12/14, 1:59:00 AM
conditional_loss_probability-1000-by-node/		9/16/14, 4:53:00 AM
conditional_loss_probability-1000-by-node.tar	280 MB	9/16/14, 4:57:00 AM

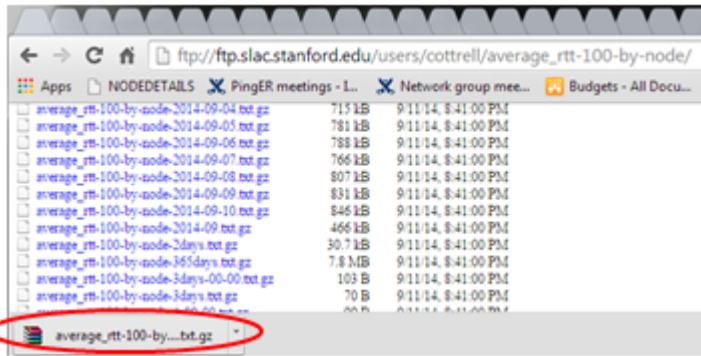
Clicking on say average\_rtt-100-by-node/ results in:



## Index of /users/cottrell/average\_rtt-100-by-node/

Name	Size	Date Modified
[parent directory]		
average_rtt-100-by-node-120days-00-00.txt.gz	105 B	9/11/14, 8:21:00 PM
average_rtt-100-by-node-120days.txt.gz	3.1 MB	9/11/14, 8:21:00 PM
average_rtt-100-by-node-1900-03.txt.gz	384 kB	9/11/14, 8:21:00 PM
average_rtt-100-by-node-1900-04.txt.gz	67.6 kB	9/11/14, 8:21:00 PM
average_rtt-100-by-node-1970-01-01.txt	63 B	9/11/14, 8:21:00 PM

Scroll down to some recent data, e.g. average\_rtt-100-by-node-2014-09-10.txt.gz. Click on it and in Chrome the file appears at the bottom left



Drag this file to a more permanent place (e.g. your desktop). Then you have to unzip it. How you do that depends on your OS. I would do it in Unix. So I copy it to a Unix host and use gunzip. On windows I copy it to the desktop and then use WinRAR (you may need to download) to unzip and look at.

## From a Unix command line

Use [wget](#), e.g to download a file into the working directory:

```
$wget 'ftp://ftp.slac.stanford.edu/users/cottrell/average_rtt-100-by-node-2014-09-10.txt.gz'
```

You will probably want to write a script to get all the data.

Then use gunzip to unzip the data

## Metrics available

average_rtt/	out_of_order_packets/
conditional_loss_probability/	packet_loss/
duplicate_packets/	throughput/
ipdv/	unpredictability/
iqr/	unreachability/
minimum_packet_loss/	zero_packet_loss_frequency/
minimum_rtt/	maximum_rtt/
MOS/	alpha/

See the [Tutorial](#) for more on the metrics.

## Updating the information

It takes a long time to download all the information (dating back to 1998) from the PingER archive. Thus this is typically done once. Thus hourly data for the last 6 months is therefore made available in directories of the form /users/cottrell/<metric-size-by>/. This data is updated daily. See below for an example:

←

→

↻

🏠

🔍

ftp://ftp.slac.stanford.edu/users/cottrell

📁

Apps

🔗

Network group meeting

🔗

PingER meetings - Int

🔗

One-on-One Meeting

new










ServiceNov

Index of /users/cottrell

	Name	Size	Date Modified
📁	[parent directory]		
📁	MOS-100-by-node/		12/21/16, 2:27:00 PM
📁	MOS-100-by-node.tar	595 MB	12/21/16, 2:31:00 PM

These directories contains the daily files that contain the hourly data for the last 6 months:

## Index of /users/cottrell/average\_rtt-100-by-node/

Name	Size	Date Modified
 [parent directory]		
 <a href="#">average_rtt-100-by-node-2016-06-14.txt.gz</a>	490 kB	8/2/16, 11:15:00 PM
 <a href="#">average_rtt-100-by-node-2016-07-13.txt.gz</a>	460 kB	8/2/16, 11:15:00 PM
 <a href="#">average_rtt-100-by-node-2016-07-23.txt.gz</a>	430 kB	8/2/16, 11:15:00 PM
 <a href="#">average_rtt-100-by-node-2016-08-07.txt.gz</a>	424 kB	8/8/16, 11:24:00 AM
 <a href="#">average_rtt-100-by-node-2016-08-11.txt.gz</a>	418 kB	8/12/16, 11:24:00 AM
 <a href="#">average_rtt-100-by-node-2016-09-01.txt.gz</a>	371 kB	9/2/16, 11:24:00 AM
 <a href="#">average_rtt-100-by-node-2016-09-02.txt.gz</a>	396 kB	9/3/16, 11:24:00 AM
 <a href="#">average_rtt-100-by-node-2016-09-18.txt.gz</a>	365 kB	9/19/16, 11:24:00 AM

The idea is that the analysis site uploads all the available data from the tar files and then one time can update its files for the last few days using this data from the metric directory, rather than transferring the full tar file (that contains many Gbytes compared to the hundreds of GBytes in the daily files).

### Format of the file

There is a copy of the file [here](#).

Looking at the first few lines of the file in [https://confluence.slac.stanford.edu/download/attachments/187016936/average\\_rtt-100-by-node-2014-09-10.txt](https://confluence.slac.stanford.edu/download/attachments/187016936/average_rtt-100-by-node-2014-09-10.txt)

It appears as :

```

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
pingerkhi.pern.edu.pk pingerpwr.pern.edu.pk . . . . . pingerkhi.pern.edu.
pk pingerpwr.pern.edu.pk
pinger.sprace.org.br www.kist.ac.rw . . . . . pinger.sprace.org.br www.
kist.ac.rw
airuniversity.seecs.edu.pk mail.gnet.tn 141.522 141.515 141.500 141.720 144.745 141.918 144.594 142.411 142.123
142.355 142.495 141.863 141.760 142.403 140.645 139.498 139.637 139.677 142.402 140.002 139.565 139.670 139.752
139.398 airuniversity.seecs.edu.pk mail.gnet.tn

```

As described in <http://www-iepm.slac.stanford.edu/pinger/tools/retrievedata.html#rawdataformat>:

The first line is special and labels the columns (with the hour of the day) that start after the source (monitor) and destination/target/monitored/remote host names (e.g. in line 2 after pingerkhi.pern.edu.pk pingerpwr.pern.edu.pk). Following the two names are the values of the metric for the hours in the first line. A dot (.) means there was no data. At the end of the line the source and destination names are given again. The name of the file gives the metric (in the above case average\_rtt). Thus in the line starting:

```
airuniversity.seecs.edu.pk mail.gnet.tn 141.522 141.515 141.500
```

airuniversity.seecs.edu.pk is the source (monitoring host), mail.gnet.tn is the name of the host being monitored, 141.522 is the RTT seen at 0 hours, 141.515 is the RTT at 1am etc.

### More information