

Archiving PingER data by tar for retrieval by anonymous ftp

Table of Contents

- [Table of Contents](#)
- [Format of data etc](#)
- [Saving Analyzed data](#)
 - [For 1 metric \(average_rtt\)](#)
 - [For all metrics with 100Byte pings by node](#)
- [Updating Analyzed data in the archive](#)
 - [Tar file](#)
 - [Directory](#)
- [Retrieving](#)
- [Saving Raw Data](#)

Format of data etc

See [PingER data flow at SLAC](#)

Saving Analyzed data

The most likely data to be of use to others is the analyzed/aggregated data. This is kept in

`/nfs/slac/g/net/pinger/pingerreports/hep/`

There is a lot of data so you will need a lot of space. I suggest you use the `/tmp` directory as intermediate storage, do one metric at a time (e.g. `average_rtt`), one size (e.g. 100) and one by (e.g. by-node). If this does not work send an email to `unix-admin` temporary requesting space in `/afs/slac/public/users/cottrell` (this is accessible from anonymous FTP). Note you will need space for the copied directory and for the tar's and zipped file. I requested 100Gbytes.

For 1 metric (average_rtt)

```
$mkdir /afs/slac/public/users/cottrell/average_rtt-100-by-node
```

```
$cp -v /nfs/slac/g/net/pinger/pingerreports/hep/average_rtt/average_rtt-100-by-node* /afs/slac/public/users/cottrell/average_rtt-100-by-node
```

#There are about 6500 files per metric. Copy takes about 20 mins per metric.

```
$tar -cvzf /afs/slac/public/users/cottrell/archive-average_rtt-100-by-node.tar /afs/slac/public/users/cottrell/average_rtt-100-by-node
```

#A metric takes about 6 minutes to tar and compress and each tar file occupies ~ 1.5GBytes.

For all metrics with 100Byte pings by node

```
$mkdir /afs/slac/public/users/cottrell/metrics-100-by-node
```

```
$cp -v /nfs/slac/g/net/pinger/pingerreports/hep/*/*-100-by-node* /afs/slac/public/users/cottrell/metrics-100-by-node
```

```
$tar -cvzf /afs/slac/public/users/cottrell/archive-average_rtt-100-by-node.tar /afs/slac/public/users/cottrell/average_rtt-100-by-node
```

However:

```
$ls /nfs/slac/g/net/pinger/pingerreports/hep/*/*-100-by-node*
```

/bin/ls: Argument list too long.
Exit 1

To provide maximum flexibility we decided to write a script (`/afs/slac/package/pinger/pinger-tar.pl`) to copy and tar the data.

The directory looks as follows:

```
[cottrell@pinger ~]$ ls -l /afs/slac.stanford.edu/public/users/cottrell/*.tar
-rw-rw-r-- 1 cottrell sf 512297529 Sep 11 16:33 /afs/slac.stanford.edu/public/users/cottrell/MOS-100-by-node.tar
-rw-rw-r-- 1 cottrell sf 755387730 Sep 11 17:00 /afs/slac.stanford.edu/public/users/cottrell/alpha-100-by-node.tar
-rw-rw-r-- 1 cottrell sf 1640854797 Sep 11 13:58 /afs/slac.stanford.edu/public/users/cottrell/average_rtt-100-by-node.
tar
-rw-rw-r-- 1 cottrell sf 293583749 Sep 11 18:59 /afs/slac.stanford.edu/public/users/cottrell
/conditional_loss_probability-100-by-node.tar
-rw-rw-r-- 1 cottrell sf 259801856 Sep 11 19:40 /afs/slac.stanford.edu/public/users/cottrell/duplicate_packets-100-
by-node.tar
-rw-rw-r-- 1 cottrell sf 1018523005 Sep 11 14:27 /afs/slac.stanford.edu/public/users/cottrell/ipdv-100-by-node.tar
-rw-rw-r-- 1 cottrell sf 751314261 Sep 11 16:16 /afs/slac.stanford.edu/public/users/cottrell/iqr-100-by-node.tar
-rw-rw-r-- 1 cottrell sf 1896982390 Sep 11 18:03 /afs/slac.stanford.edu/public/users/cottrell/maximum_rtt-100-by-
node.tar
-rw-rw-r-- 1 cottrell sf 297380470 Sep 11 19:53 /afs/slac.stanford.edu/public/users/cottrell/minimum_packet_loss-
100-by-node.tar
-rw-rw-r-- 1 cottrell sf 1464424282 Sep 11 13:16 /afs/slac.stanford.edu/public/users/cottrell/minimum_rtt-100-by-
node.tar
-rw-rw-r-- 1 cottrell sf 240217457 Sep 11 19:24 /afs/slac.stanford.edu/public/users/cottrell/out_of_order_packets-
100-by-node.tar
-rw-rw-r-- 1 cottrell sf 425757104 Sep 11 15:54 /afs/slac.stanford.edu/public/users/cottrell/packet_loss-100-by-node.
tar
-rw-rw-r-- 1 cottrell sf 1995419214 Sep 11 15:33 /afs/slac.stanford.edu/public/users/cottrell/throughput-100-by-node.
tar
-rw-rw-r-- 1 cottrell sf 587342674 Sep 11 18:39 /afs/slac.stanford.edu/public/users/cottrell/unpredictability-100-by-
node.tar
-rw-rw-r-- 1 cottrell sf 273680921 Sep 11 18:16 /afs/slac.stanford.edu/public/users/cottrell/unreachability-100-by-
node.tar
-rw-rw-r-- 1 cottrell sf 269081359 Sep 11 19:15 /afs/slac.stanford.edu/public/users/cottrell
/zero_packet_loss_frequency-100-by-node.tar
```

The compression ratio is about 3.2:1.

To increase the number of files we could change the ping size from 100 to 1000 (in average_rtt-100-by-node* for example) and the by-node to by-site, i.e. a factor of 4.

As of 9/13/2014 there are about 20GBytes of compressed data and 100,000 files. By 6/24/2019 there are about 33GBytes (do an `ls -l /afs/slac.stanford.edu/public/users/cottrell/*.tar`, import into Excel and use space to divide into columns and sum up the Bytes column).

Multiply by say 3.2 to get uncompressed data and it is about 100GBytes.

Nb the following files do not gunzip:

```
[root@sc2u0n0 afs]# find . -name "*.gz"
```

```
./slac/public/users/cottrell/conditional_loss_probability-100-by-node/conditional_loss_probability-100-by-node-1998-02-20.txt.gz
```

```
./slac/public/users/cottrell/minimum_packet_loss-100-by-node/minimum_packet_loss-100-by-node-2004-07-18.txt.gz
```

I get:

```
268cottrell@pinger:/afs/slac/public/users/cottrell/conditional_loss_probability-100-by-node$gunzip /tmp/conditional_loss_probability-100-by-node-1998-02-20.txt.gz
```




```
gzip: /tmp/conditional_loss_probability-100-by-node-1998-02-20.txt.gz: invalid compressed data--crc error
```

```
gzip: /tmp/conditional_loss_probability-100-by-node-1998-02-20.txt.gz: invalid compressed data--length error
```

Updating Analyzed data in the archive

The FTP directory appears as:

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

I.e. there are a directory and a tar file for each metric. There is also a trscron job:

```
lnxcron:480 00 4 * * * /opt/TWWfsw/bin/perl /afs/slac/package/pinger/ftp-update_2.pl #Joao that
makes the updates at 4:00am each morning.
```

This updates the above files.

Tar file

The tar file contains files for each day with the hourly data for all the days in the PingER archive

↶

average_rtt

average_rtt-1000-by-node.tar

📄

Type: Text Document

📄

Size: 0.13 KB → 0.13 KB

📄

average_rtt-1000-by-node-1998-02-23.txt

Type: Text Document

Date modified: 8/28/2015 1:38 PM

Size: 6.10 KB → 6.10 KB

📄

average_rtt-1000-by-node-1998-02-24.txt

Type: Text Document

Date modified: 8/28/2015 1:43 PM

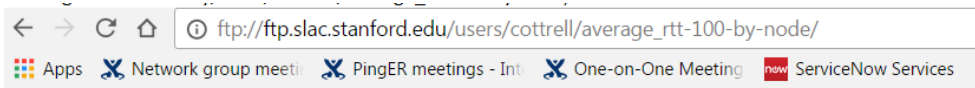
Size: 6.28 KB → 6.28 KB

The files appear 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
pinger-host.fnal.gov www-05.nexus.ao 312.466 290.641 284.717 293.712 272.983 286.128 282.721 280.938 305.152
309.595 302.656 293.713 310.015 282.700 299.544 297.096 288.092 286.728 285.082 285.459 297.283 300.526 290.449
295.673 pinger-host.fnal.gov www-05.nexus.ao
pingersonar-usm.myren.net.my www.afe.mr 631.613 588.780 611.781 629.670 580.149 631.359 690.517 526.923 584.238
635.444 749.802 786.702 869.685 981.443 837.803 956.452 858.124 652.750 723.169 704.219 1060.705 642.455
598.416 612.010 pingersonar-usm.myren.net.my www.afe.mr
```

Directory

The directory 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]		
average_rtt-100-by-node-2016-06-14.txt.gz	490 kB	8/2/16, 11:15:00 PM
average_rtt-100-by-node-2016-07-13.txt.gz	460 kB	8/2/16, 11:15:00 PM
average_rtt-100-by-node-2016-07-23.txt.gz	430 kB	8/2/16, 11:15:00 PM
average_rtt-100-by-node-2016-08-07.txt.gz	424 kB	8/8/16, 11:24:00 AM
average_rtt-100-by-node-2016-08-11.txt.gz	418 kB	8/12/16, 11:24:00 AM
average_rtt-100-by-node-2016-09-01.txt.gz	371 kB	9/2/16, 11:24:00 AM
average_rtt-100-by-node-2016-09-02.txt.gz	396 kB	9/3/16, 11:24:00 AM
average_rtt-100-by-node-2016-09-18.txt.gz	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). The format of the daily files is the same as for the untarred files above.

Retrieving

The data is available via anonymous ftp via <ftp://ftp.slac.stanford.edu/users/cottrell>, see [here for more on retrieving the data](http://www-iepm.slac.stanford.edu/pinger/tools/retrievedata.html). For more information see: <http://www-iepm.slac.stanford.edu/pinger/tools/retrievedata.html>.

[Volume of analyzed data](#) May, 2015. There is a script to assist with getting the data volumes at `~cottrell/bin/sumdir-regexp.pl`.

Saving Raw Data

The raw data is the data initially recorded every 30 minutes by the `pinger2.pl` measurement script as opposed to the analyzed and aggregated into hourly interval data above. For the record unzipped raw data records look like:

```
pinger.slac.stanford.edu 134.79.240.30 www.eldjazair.net.dz 193.194.64.80 100 1178841602 10 10 195.572 196.680
198.257 0 1 2 3 4 5 6 7 8 9 196 195 198 195 197 196 197 195 196 196
```

The hourly analyzed data has the advantage that it is cleaner since it has had filtering for some bad data, and the FTP archive for the analyzed data is updated daily etc. Gathered raw data for 1998-01 - 2007-07 is saved in

```
/nfs/slac/g/net/pinger/pingerdata/hep/data/<YYYY>/<host>/
e.g.
ls2007/brunsvigia.tenet.ac.za/
With files of the form:
-rw-rw-r-- 1 pinger ieipm 584920 Jan 2 2007 ping-2007-01-01.txt.gz
```

Gathered data from 2007-07 onwards is saved in:

```
/nfs/slac/g/net/pinger/pingerdata/hep/data/<host>/ping-<YYYY>--<MM>--<DD>.txt.gz
e.g.
/nfs/slac/g/net/pinger/pingerdata/hep/data/pinger.slac.stanford.edu/ping-2011-03-22.txt.gz
/nfs/slac/g/net/pinger/pingerdata/hep/data/pcgiga.cern.ch/ping-2006-09-28.txt.gz
```

There are about 270K compressed raw data files in `/nfs/slac/g/net/pinger/pingerdata/hep/data` (found using

ls -l /nfs/slac/g/net/pinger/pingerdata/hep/data/* >! junk; wc junk), occupying 300GBytes.

There is also data in the space below but I am unsure of its provenance

```
/nfs/slac/g/net/pinger/pingerdata/hep/data.unite/<host>/ping-<YYYY>-<MM>-<DD>.txt.gz
with files of the form:
-rw-rw-r-- 1 cottrell sf 52464462 Jul 20 18:11 ping-2007-04.txt
```

There is a script `pinger-tar.pl`. To save the raw data use the command `/afs/slac/package/pinger/pinger-tar.pl`. You can get help on this script by going to <http://www-iepm.slac.stanford.edu/pinger/scripttable.html> and choosing `pinger-tar.pl` from the pull down list. It lists the directory to find all hosts with data. For each of these hosts it copies all files with names starting `ping-*` into the anonymous ftp space, then it un-compresses it and tars and compresses it into the FTP space. Partial output from running the script is seen below.

```
Time so far=7688 secs cmd=rm /afs/slac/public/users/cottrell/vle.iuu.edu.pk/ping-*
Fri Nov 13 10:14:50 2015 Took 1 secs, for rm /afs/slac/public/users/cottrell/vle.iuu.edu.pk/ping-*
Copy data for host(10/199/210)=drwxrwsr-x 2 iepm iepm 31744 Jul 21 05:23 wanmoninst1.cern.ch debug=0,
production=true
Time so far=7689 secs cmd=cp -p /nfs/slac/g/net/pinger/pingerdata/hep/data/wanmoninst1.cern.ch/ping-* /afs/slac
/public/users/cottrell/wanmoninst1.cern.ch/
Fri Nov 13 10:15:28 2015 Took 38 secs, for cp -p /nfs/slac/g/net/pinger/pingerdata/hep/data/wanmoninst1.cern.ch
/ping-* /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/
Time so far=7727 secs cmd=gunzip -f /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/ping-*
Fri Nov 13 10:17:06 2015 Took 98 secs, for gunzip -f /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/ping-*
Time so far=7825 secs cmd=tar -Pczf /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/wanmoninst1.cern.ch.tar
/afs/slac/public/users/cottrell/wanmoninst1.cern.ch/ping-*.txt
Fri Nov 13 10:18:04 2015 Took 58 secs, for tar -Pczf /afs/slac/public/users/cottrell/wanmoninst1.cern.ch
/wanmoninst1.cern.ch.tar /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/ping-*.txt
Fri Nov 13 10:18:04 2015: 168840707 bytes in /afs/slac/public/users/cottrell/wanmoninst1.cern.ch/wanmoninst1.
cern.ch.tar(11/199/210), time so far=7883secs
```

This saves about 200 tar files in <ftp://ftp.slac.stanford.edu/users/cottrell>. They are for data from 2007-07 forwards. They appear as directories with the directory name being the name of the PingER monitoring agent (MA) host. For example

 bofur.rmki.kfki.hu/	11/9/15, 4:36:00 AM
 brunsvigia.tenet.ac.za/	11/9/15, 11:37:00 PM
 buitms.seecs.edu.pk/	11/9/15, 11:41:00 PM
 cache01.ansp.br/	11/9/15, 5:10:00 AM
 cae.seecs.edu.pk/	11/9/15, 11:48:00 PM
 caia.swin.edu.au/	11/9/15, 5:18:00 AM
 cc.if.ufjf.br/	11/9/15, 11:57:00 PM
 cithep130.ultraviolet.org/	11/10/14, 12:00:00 AM
 comsatsswl.seecs.edu.pk/	11/10/14, 12:01:00 AM

The files are tar files and compressed. Together they occupy about 70GBytes compressed or about 300GBytes uncompressed.

Information on each host MA can be found by going to <http://www-iepm.slac.stanford.edu/pinger/pingerworld/all-nodes.cf> which can be require'd in a perl script, or for currently working MAs <http://www-iepm.slac.stanford.edu/pinger/pingerworld/slaonly-nodes.cf> or in the XML file <http://www-iepm.slac.stanford.edu/pinger/pingerworld/rss.xml>.

A directory of the files (using `ls -l /afs/slac/public/users/cottrell/*/*tar`) can be found [here](#).