CC-IN2P3 — IHEP connectivity issues

Progress report

Fabio Hernandez

fabio@in2p3.fr

July 2nd, 2012

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DES AFFAIRES ÉTRANGÈRES ET EUROPÉENNES

中國科學院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

CCIN2P3

Context

- Summary of the previous meeting (2012-03-12):

  link aggregation algorithm used by file transfer machines at CC-IN2P3 suspected to be the cause of the low throughput

  need to make detailed file transfer tests between IHEP and CC-IN2P3

  measured network throughput from GEANT routers in London to IHEP considered good (799 Mbps)

  details: https://indico.in2p3.fr/conferenceDisplay.py?confId=6652

- Today, we summarize the progress collectively made since then

Information sharing

- Reference document at IHEP's Twiki

  http://twiki.ihep.ac.cn/twiki/bin/view/InternationalConnectivity/IHEP-CCIN2P3

- All the historical information about this issue can be found there

- Timely updated

First test campaign

- Extensive tests performed by Yvan Calas in April

both file transfer and memory-to-memory tests

comparing throughput using dual-NIC and single-NIC machines at CC-IN2P3

detailed results: http://tinyurl.com/ccin2p3-ihep-net

First test campaign (cont.)

- Memory-to-memory tests CC-IN2P3 → IHEP

  low throughput

  results extremely variable in time: 10 Mbps to 409 Mbps (Linux sender) and 19 Mbps to 187 Mbps (Solaris sender)

- Disk-to-disk file transfer CC-IN2P3 → IHEP

  very low throughput: 64 Mbps to 560 Mbps (using 20 streams), depending on the size of the file (119 MB to 4 GB)

  sending machine runs Linux and used a single 1Gbps network card

- Disk-to-disk file transfer IHEP → CC-IN2P3

  good results, consistent with the observed memory-to-memory throughput measured by Perfsonar

  not detailed here to keep focused on the real problem

Link aggregation algorithm

- On May 22nd, the link aggregation algorithm was modified for the file transfer machines at CC-IN2P3

from 'round-robin' to 'IEEE 802.3ad dynamic link aggregation'

relevant only for machines with dual network interface card

memory-to-memory tests between 2 Linux machines showed improved results: from 16Mbps to 220Mbps

network throughput between the 2 Perfsonar machines remains low: 30Mbps

Second test campaign

- On June 13th, Yvan Calas made a second round of file transfer tests

  sending machine: Linux with 2 x 1Gbps network card using 'IEEE 802.3ad dynamic link aggregation'

  receiving machine: dedicated Linux machine, single 1Gbps network card

- Results

  CC-IN2P3 → IHEP: poor throughput, ***32Mbps to 40Mbps***

  IHEP → CC-IN2P3: ***71Mbps to 380Mbps***

  similar test conditions in both cases: 20 simultaneous streams, file sizes ranging from 119MB to 4GB

- Conclusion

  low throughput in the direction CC-IN2P3 → IHEP remains after changing the link aggregation algorithm

ATLAS file transfers

- The network issue between CC-IN2P3 and IHEP was discussed (among other things) during the ATLAS FR-Cloud Regional Centers meeting

  https://indico.cern.ch/conferenceDisplay.py?confId=181944

- A significant increase in the average file transfer rates for ATLAS from CC-IN2P3 to IHEP was presented

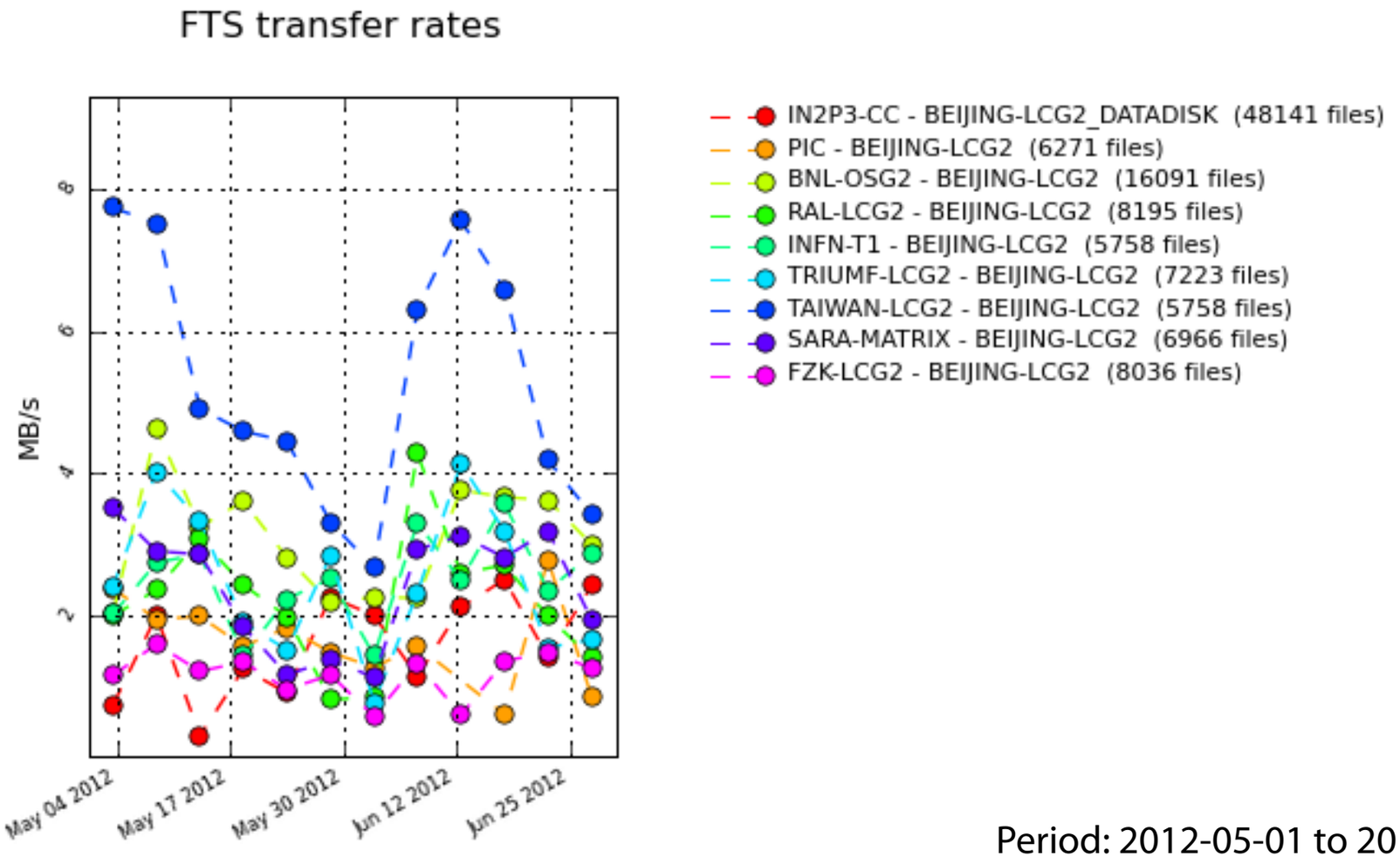
  from 1.5MB/sec to 3MB/sec on average

  still very low compared to what the network link is able to deliver and below the requirements for ATLAS Direct tier-2s

ATLAS file transfers (cont.)

Transfer rates of ATLAS files

all tier-1s → IHEP



Period: 2012-05-01 to 2012-07-01
Number of files: 112.439
File size: any size (75% are bigger than 100MB)
Source: http://bourricot.cern.ch/dq2/ftsmon

Perfsonar at IHEP

- Fresh installation since June 28th

  more stable platform now

- Two dedicated machines

  bandwidth tests: http://perfsonar.ihep.ac.cn

  latency tests: http://perfsonar2.ihep.ac.cn

- Reconfiguration

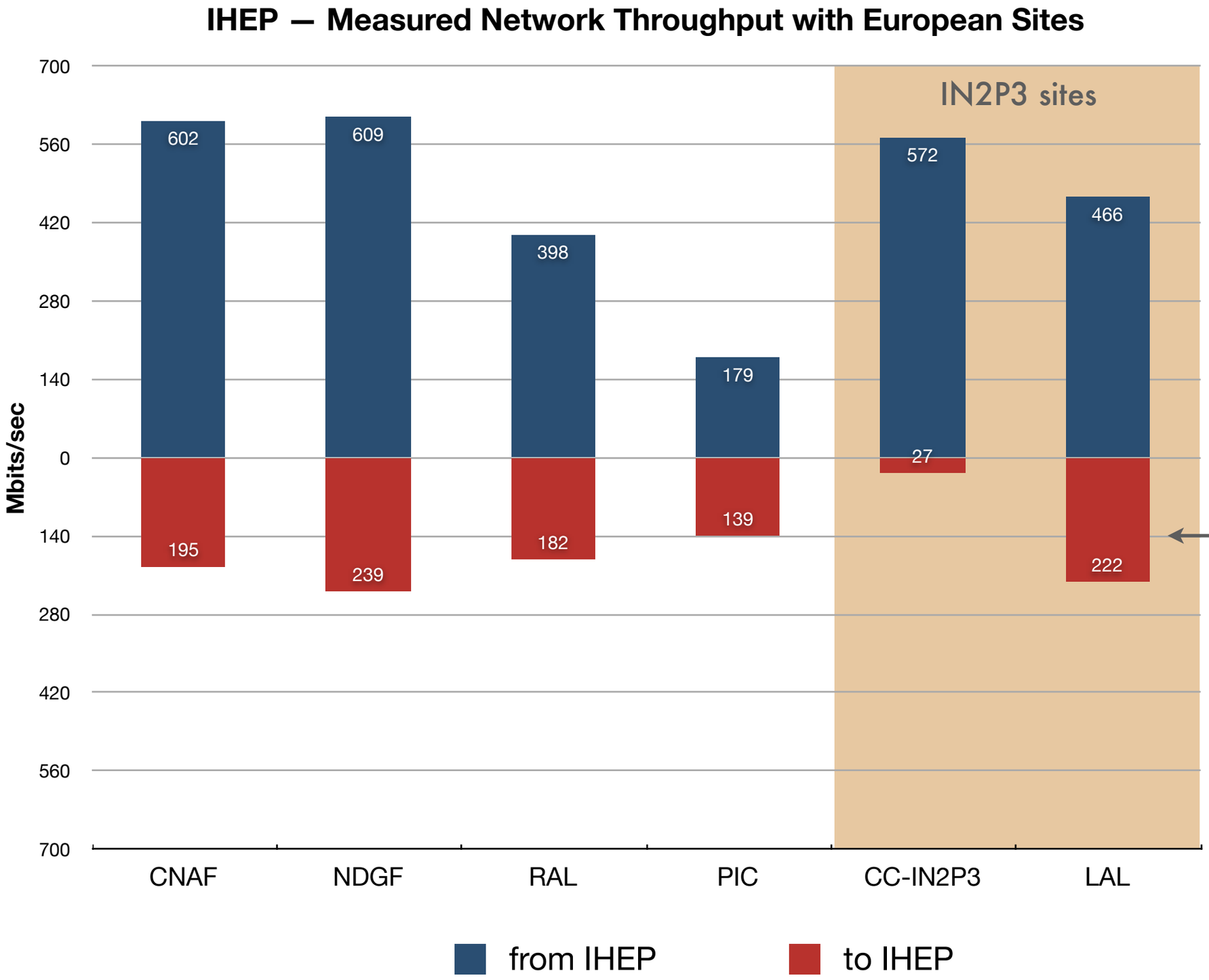
  added IHEP to the LHC-FR Cloud Perfsonar dashboard:
  https://perfsonar.usatlas.bnl.gov:8443/exda/?page=25&cloudName=LHC-FR

  added regular throughput tests with KIT (DE)

  added regular throughput and latency tests with GRIF-LAL (FR)

  investigating the reason why TEIN3/Orient is not used for traffic CERN → IHEP

Network throughput



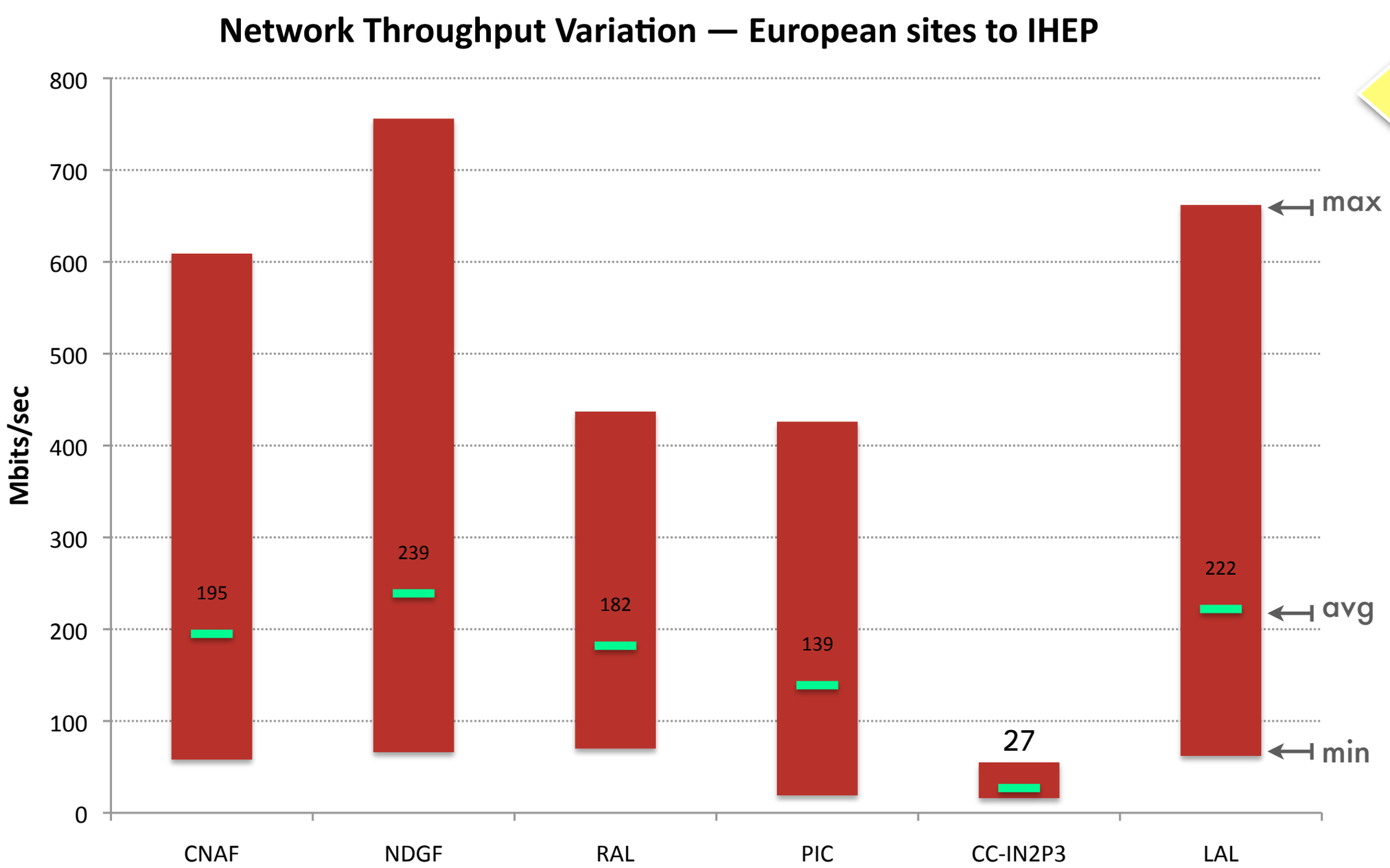
Average network throughput over 1 month up to 2012/06/26

Source: IHEP's Perfsonar
https://perfsonar.ihep.ac.cn

Throughput from GRIF-LAL→IHEP significantly higher than CC-IN2P3→IHEP.

fabio hernandez
fabio@in2p3.fr
CAS/IHEP Computing Centre
CNRS/IN2P3 Computing Centre

Network throughput variation



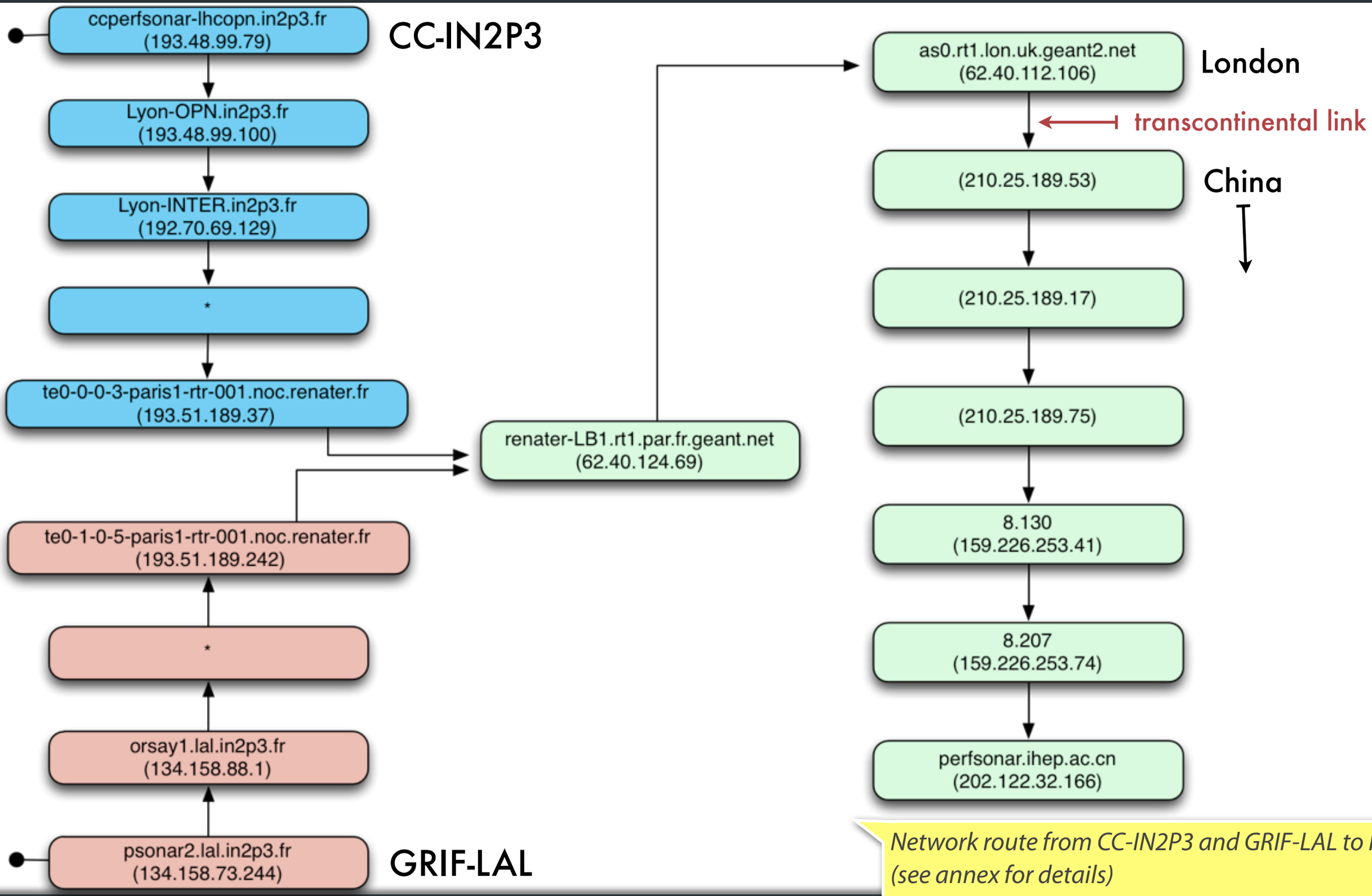
Network throughput variation from selected European sites to IHEP.

The graph shows the minimum, maximum and average iperf-based throughput over 1 month up to 2012/06/26.

Source: IHEP's Perfsonar
https://perfsonar.ihep.ac.cn

fabio hernandez
fabio@in2p3.fr
CAS/IHEP Computing Centre
CNRS/IN2P3 Computing Centre

Network route to IHEP



Network route from CC-IN2P3 and GRIF-LAL to IHEP (see annex for details)

Data collected on 2012/07/02 at 14h10 Beijing Time

What is next?

- Can we investigate further the reasons of the throughput discrepancies between GRIF-LAL and CC-IN2P3 to IHEP?
- Is it possible to perform throughput tests for each segment of the route from CC-IN2P3 at least up to the GEANT routers in Paris?
- Other ways for making progress?

Questions & Comments

fabio hernandez
fabio@in2p3.fr
CAS/IHEP Computing Centre
CNRS/IN2P3 Computing Centre

Annex

Traceroute: CC-IN2P3 → IHEP

```
Source machine: 193.48.99.79 (ccperfsonar-lhcopn.in2p3.fr)


Executing exec(traceroute, -m 30 -q 3, 202.122.32.166, 140)
traceroute to 202.122.32.166 (202.122.32.166), 30 hops max, 140 byte packets
 1  Lyon-OPN.in2p3.fr (193.48.99.100)  0.370 ms  0.375 ms  0.419 ms
 2  Lyon-INTER.in2p3.fr (192.70.69.129)  0.298 ms  0.329 ms  0.373 ms
 3  * * *
 4  te0-0-0-3-paris1-rtr-001.noc.renater.fr (193.51.189.37)  8.910 ms  8.902 ms  8.906 ms
 5  renater-LB1.rt1.par.fr.geant.net (62.40.124.69)  5.600 ms  5.635 ms  5.653 ms
 6  as0.rt1.lon.uk.geant2.net (62.40.112.106)  12.924 ms  12.899 ms  12.903 ms
 7  210.25.189.53 (210.25.189.53)  189.170 ms  189.209 ms  189.228 ms
 8  210.25.189.17 (210.25.189.17)  191.023 ms  190.401 ms  191.154 ms
 9  210.25.189.75 (210.25.189.75)  189.245 ms  189.212 ms  189.190 ms
10  8.130 (159.226.253.41)  189.391 ms  189.382 ms  189.337 ms
11  8.207 (159.226.253.74)  189.475 ms  189.559 ms  189.504 ms
12  perfsonar.ihep.ac.cn (202.122.32.166)  189.720 ms  189.683 ms  189.671 ms
```

Traceroute: GRIF-LAL → IHEP

```
Source machine: 134.158.73.244 (psonar2.lal.in2p3.fr)


Executing exec(traceroute, -m 30 -q 3, 202.122.32.166, 140)
traceroute to 202.122.32.166 (202.122.32.166), 30 hops max, 140 byte packets
 1  orsay1.lal.in2p3.fr (134.158.88.1)  0.481 ms  0.528 ms  0.616 ms
 2  * * *
 3  te0-1-0-5-paris1-rtr-001.noc.renater.fr (193.51.189.242)  5.970 ms  5.971 ms  5.963 ms
 4  renater-LB1.rt1.par.fr.geant.net (62.40.124.69)  1.058 ms  1.110 ms  1.160 ms
 5  as0.rt1.lon.uk.geant2.net (62.40.112.106)  8.436 ms  8.485 ms  8.535 ms
 6  210.25.189.53 (210.25.189.53)  184.540 ms  184.685 ms  184.617 ms
 7  210.25.189.17 (210.25.189.17)  185.779 ms  186.685 ms  187.369 ms
 8  210.25.189.75 (210.25.189.75)  205.459 ms  200.847 ms  200.823 ms
 9  8.131 (159.226.253.45)  184.914 ms  184.856 ms  184.899 ms
10  8.207 (159.226.253.78)  184.929 ms  184.915 ms  184.895 ms
11  perfsonar.ihep.ac.cn (202.122.32.166)  185.258 ms  185.139 ms  185.203 ms
```