10Gbit between GridKa and openlab (and their obstacles)

Forschungszentrum Karlsruhe GmbH
Institute for Scientific Computing
P.O. Box 3640
D-76021 Karlsruhe, Germany
http://www.gridka.de

Bruno Hoeft

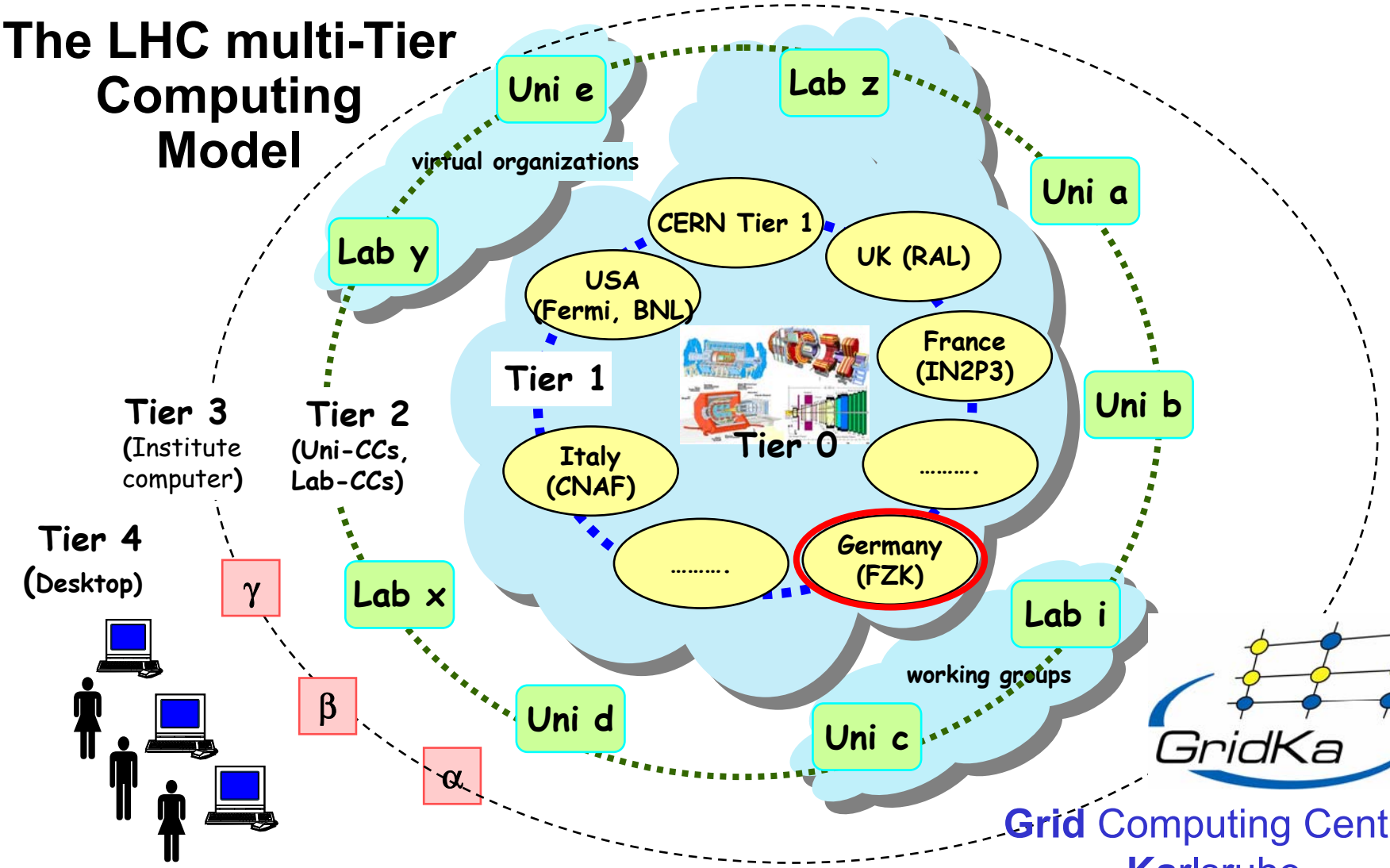
Outline

- LAN of GridKa
  - **Strucure**
  - **Prevision of installation in 2008**
- WAN
  - **History (1G/2003)**
  - **Current 10G**
  - **Testbed**
  - **Challenges crossing multi NREN (National Research and Education Network)**
  - **Quality and quantity evaluation of GridKa and openlap network connection**
  - **File transfer**
  - **Caching effect**

LHC Computing Grid Project - LCG

Enabling Grids for E-science in Europe

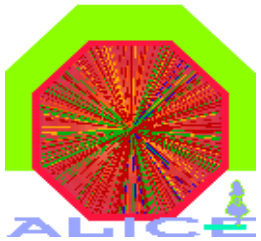
The LHC multi-Tier Computing Model



GridKa

Grid Computing Centre **Karlsruhe**

Projects at GridKa

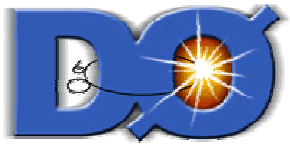


calculating jobs with "real" data

BABAR (SLAC, USA)

(FermiLab ,USA)

1,5 Mio. Jobs and 4,2 Mio. hours calculation in 2004



(FermiLab ,USA)

(CERN)

LHC Experimente

non-LHC Experimente

gradual extention of GridKa resources

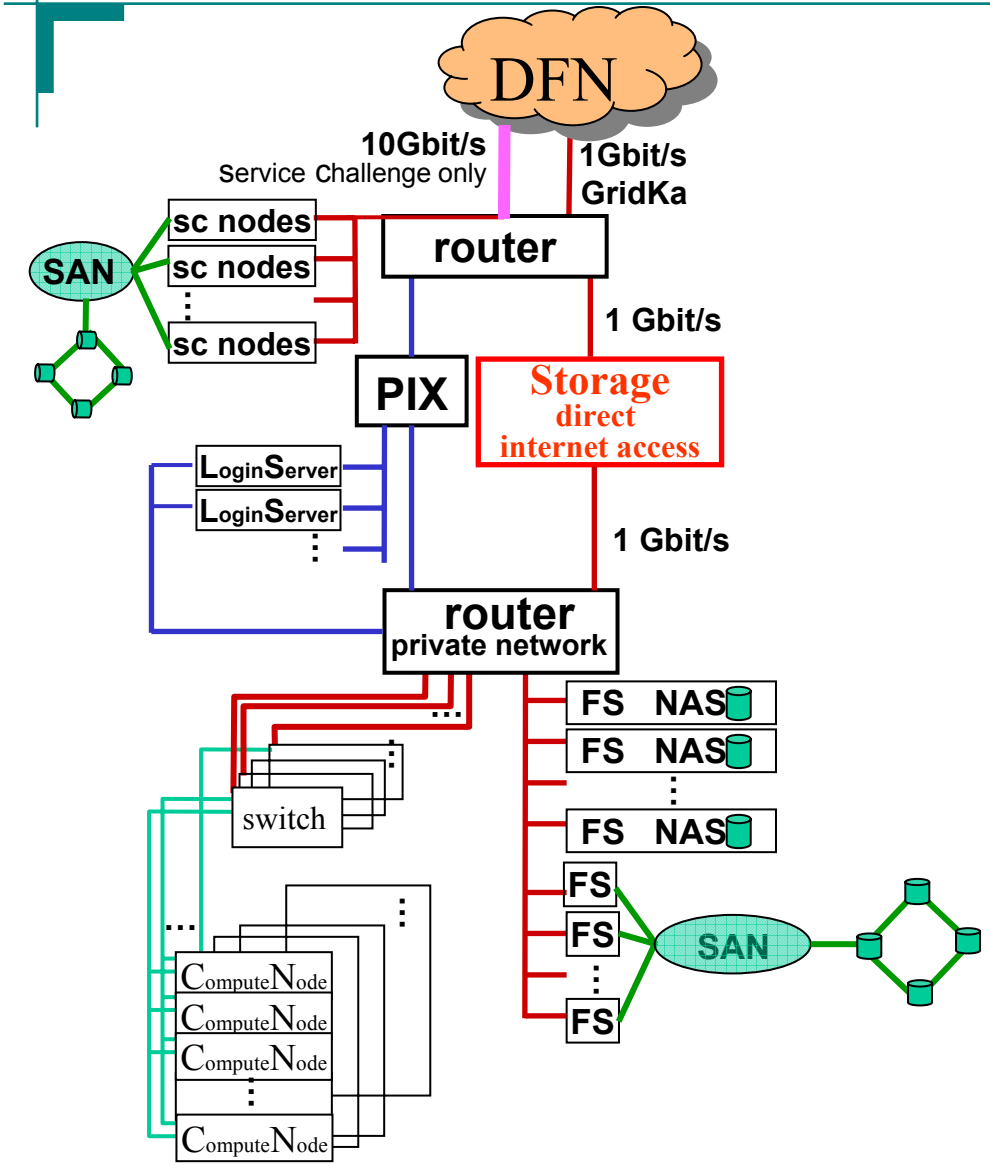
	Apr 2004	Okt 2004	Apr 2005	% of 2008
Processors	680	1070	1.280	30 %
Computing power / kSI2k	580	920	1290	12 %
Disk [TB]	160	220	270	18 %
Tape [TB]	280	375	475	12 %
Internet [Gb/s]	2	10*	10	50 %

* Internet connetion for sc

April 2005:

- biggest Linux-Cluster within the German Science Society
- largest Online-Storage at a single installation in Germany
- strongest Internet connection in Germany
- available at the Grid with over 100 installations in Europe

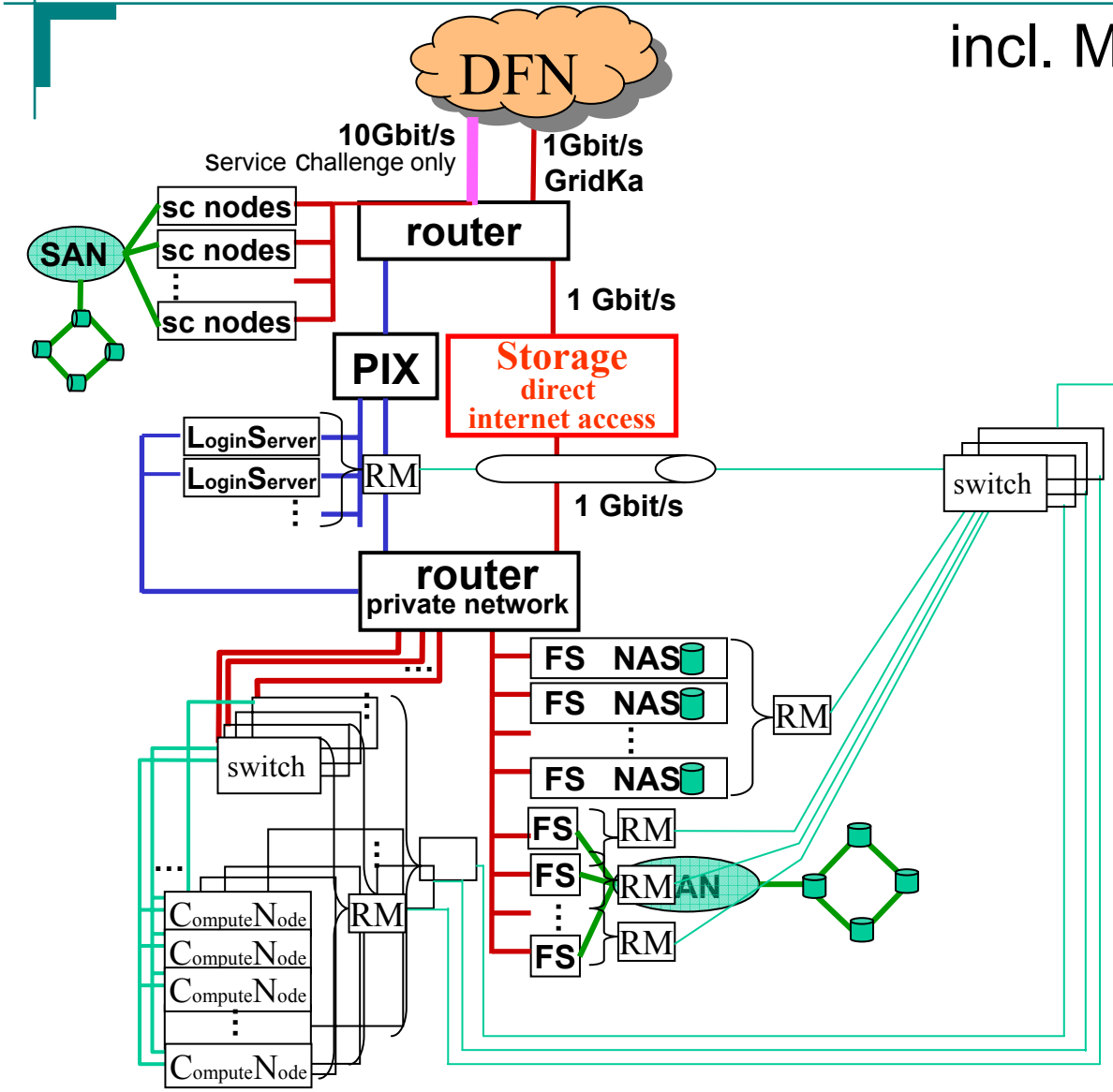
Network installation



Ethernet 320 Mbit
Ethernet 100 Mbit
Ethernet 1 Gbit
FiberChannel 2Gbit
Ethernet 10 Gbit

Network installation

incl. Management Network

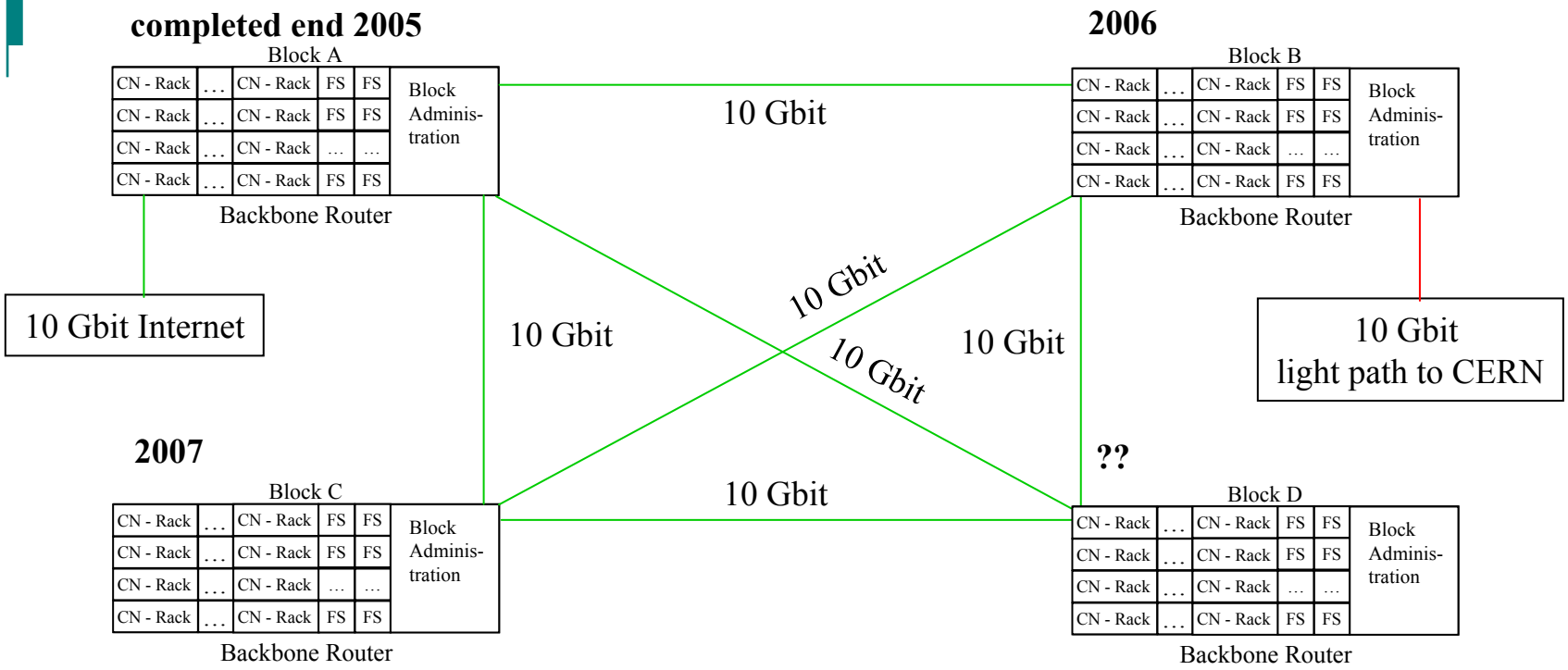


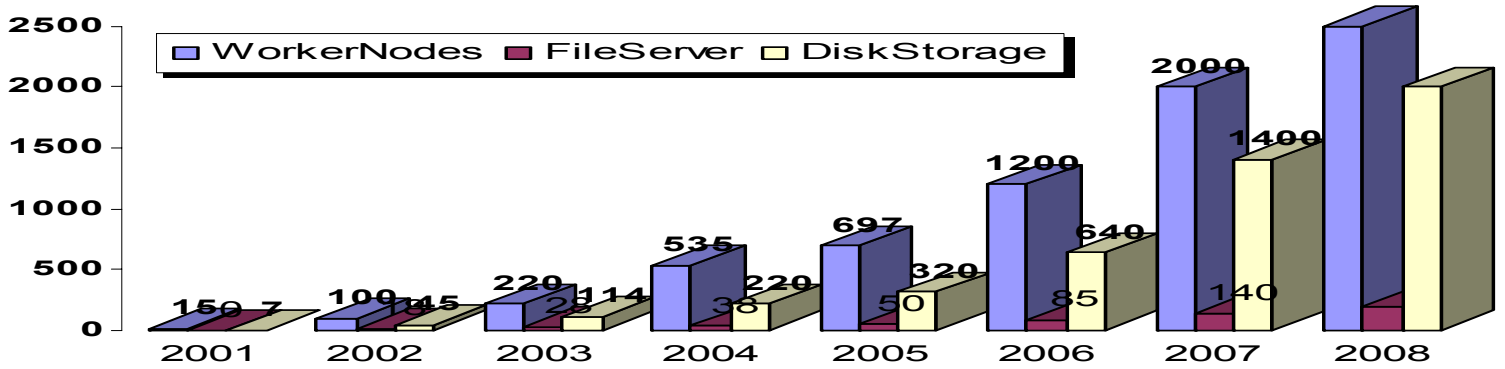
Master Controller

-Ganglia
-Nagios
-Cacti

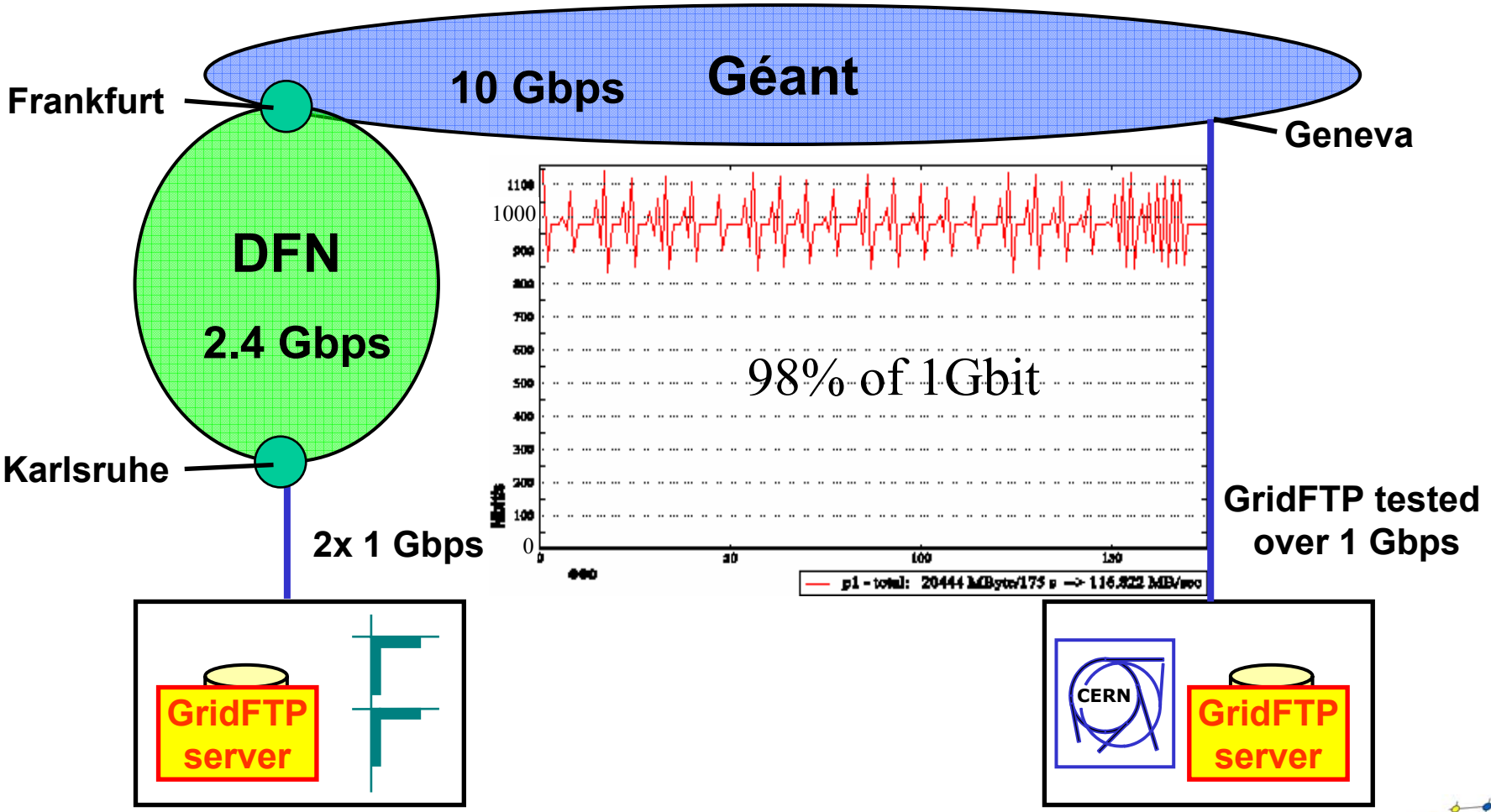
Ethernet 320 Mbit
Ethernet 100 Mbit
Ethernet 1 Gbit
FiberChannel 2Gbit
Ethernet 10 Gbit

Projection of installation in 2008

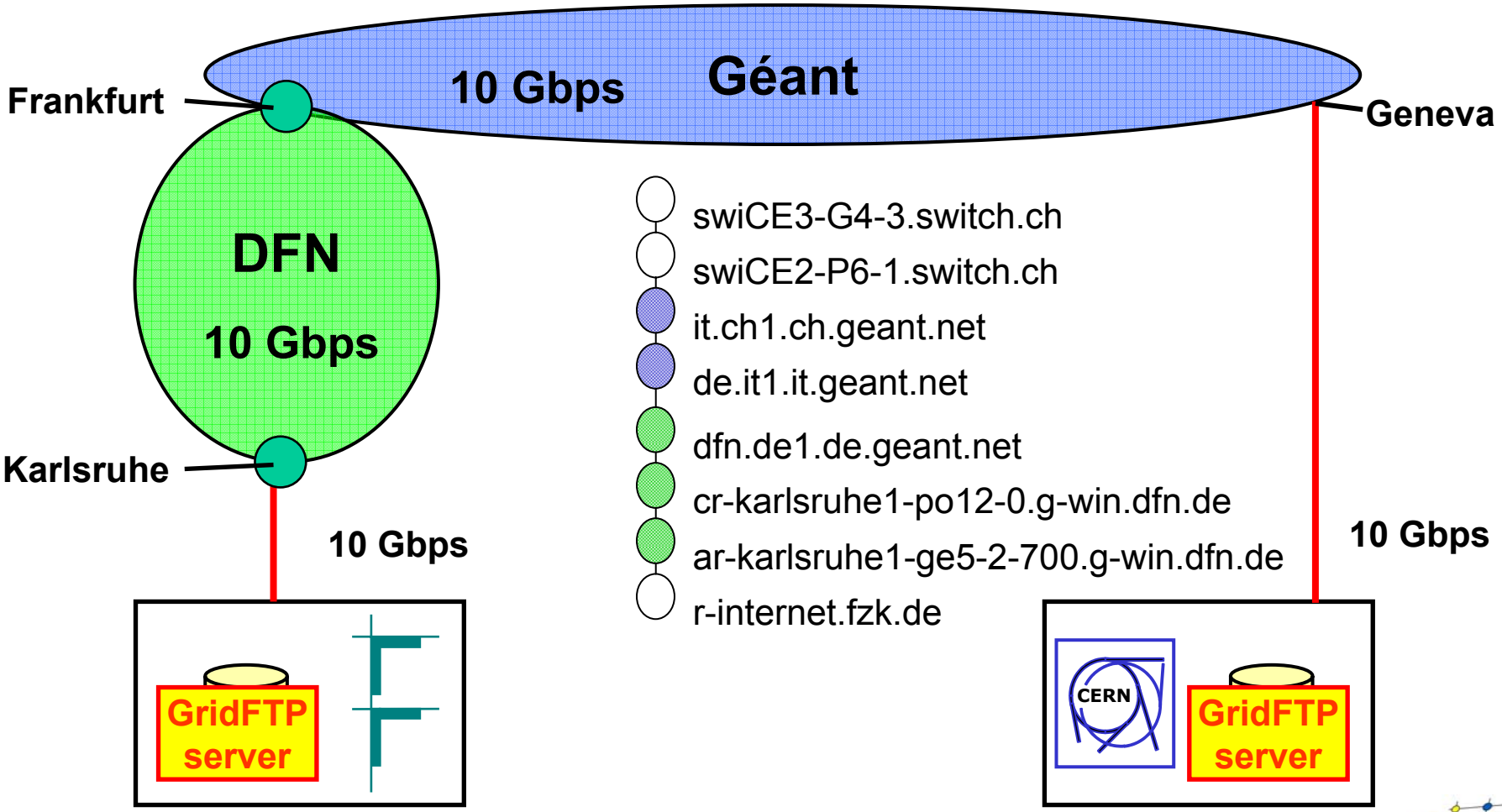




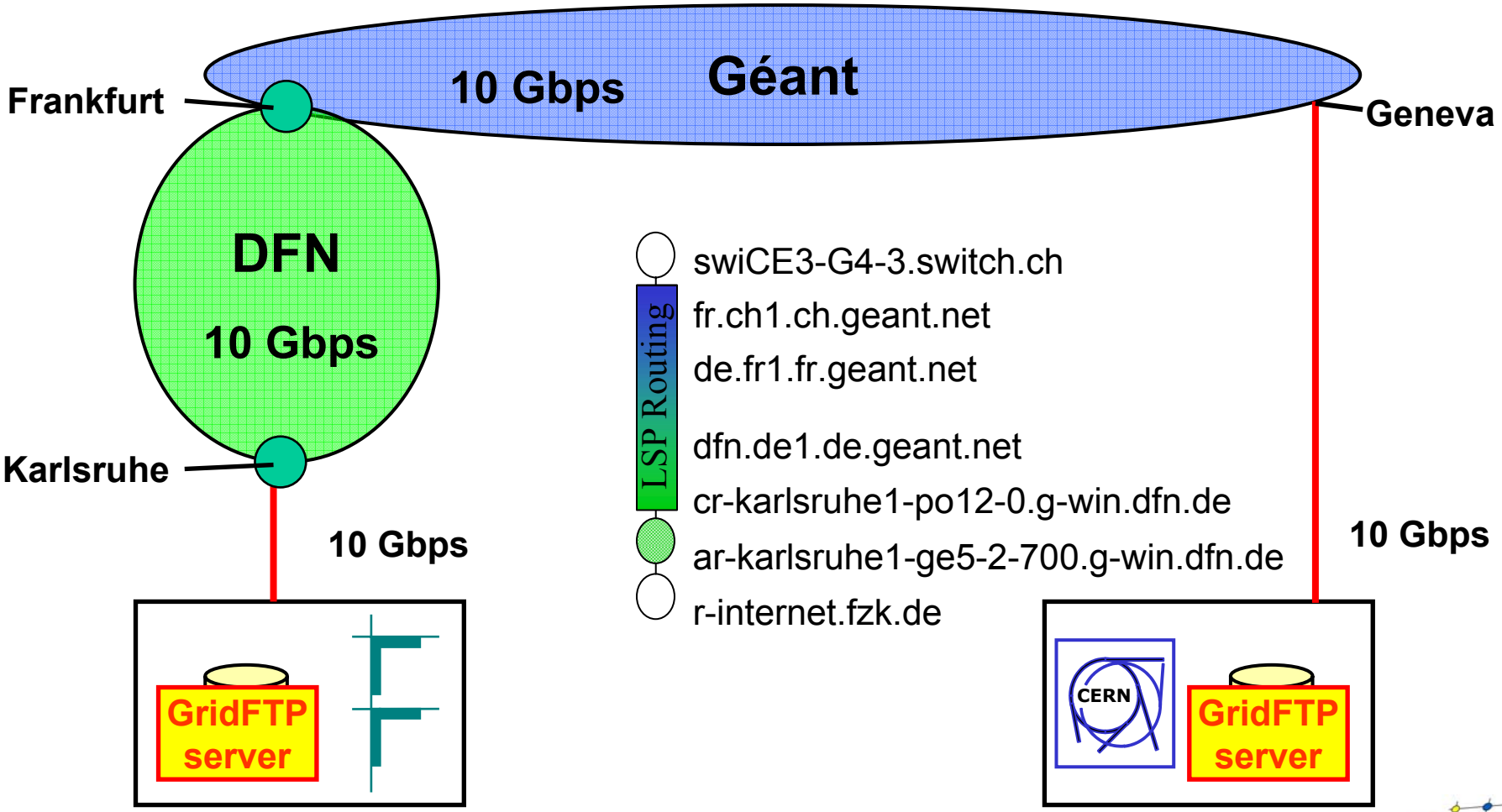
WAN 2003/4 -- Gigabit GridKa – CERN (DataTag)



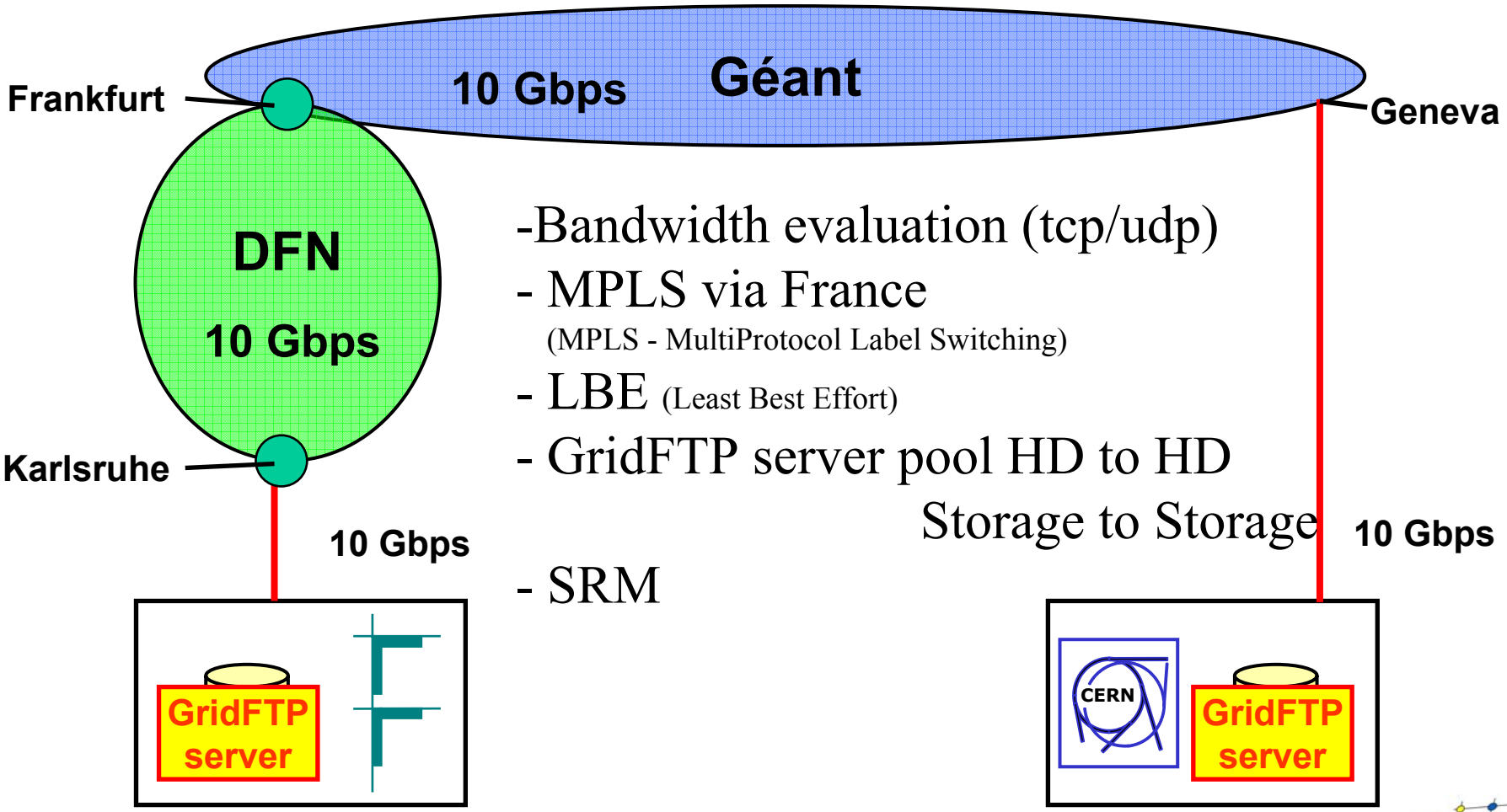
10Gigabit WAN
SC GridKa – CERN(openlab)



10Gigabit WAN
SC GridKa – CERN(openlab)



10Gigabit WAN
SC GridKa – CERN(openlab)



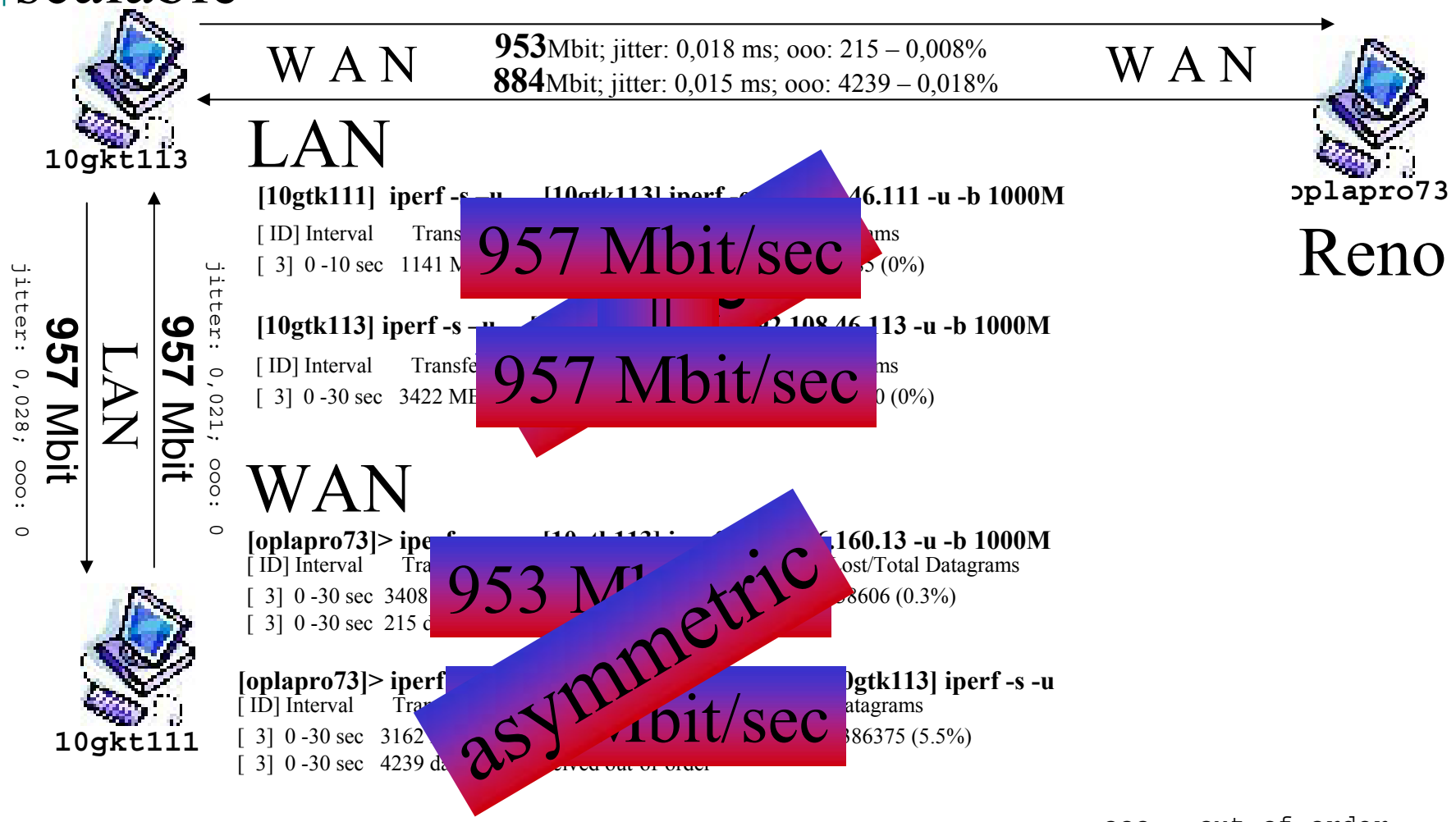
Hardware

- Various Xeon dual 2.8 and 3.0 GHz IBM x-series (Intel and Broadcom NIC)
- Recently added 3.0 GHz EM64T (800 FSB)
- Cisco 6509 with 4 10 Gb ports and lots of 1 Gb
- Storage:Datadirect 2A8500 with 16 TB
- Linux RH ES 3.0 (U2 and U3), GPFS
- 10 GE Link to GEANT via DFN (least best effort ☹)
  TCP/IP stack
  - 4 MB buffer
  - 2 MB window size

Quality Evaluation (UDP stream)

Enabling Grids for E-science in Europe

scalable



ooo - out of order

Bandwidth evaluation (TCP stream)

Reno

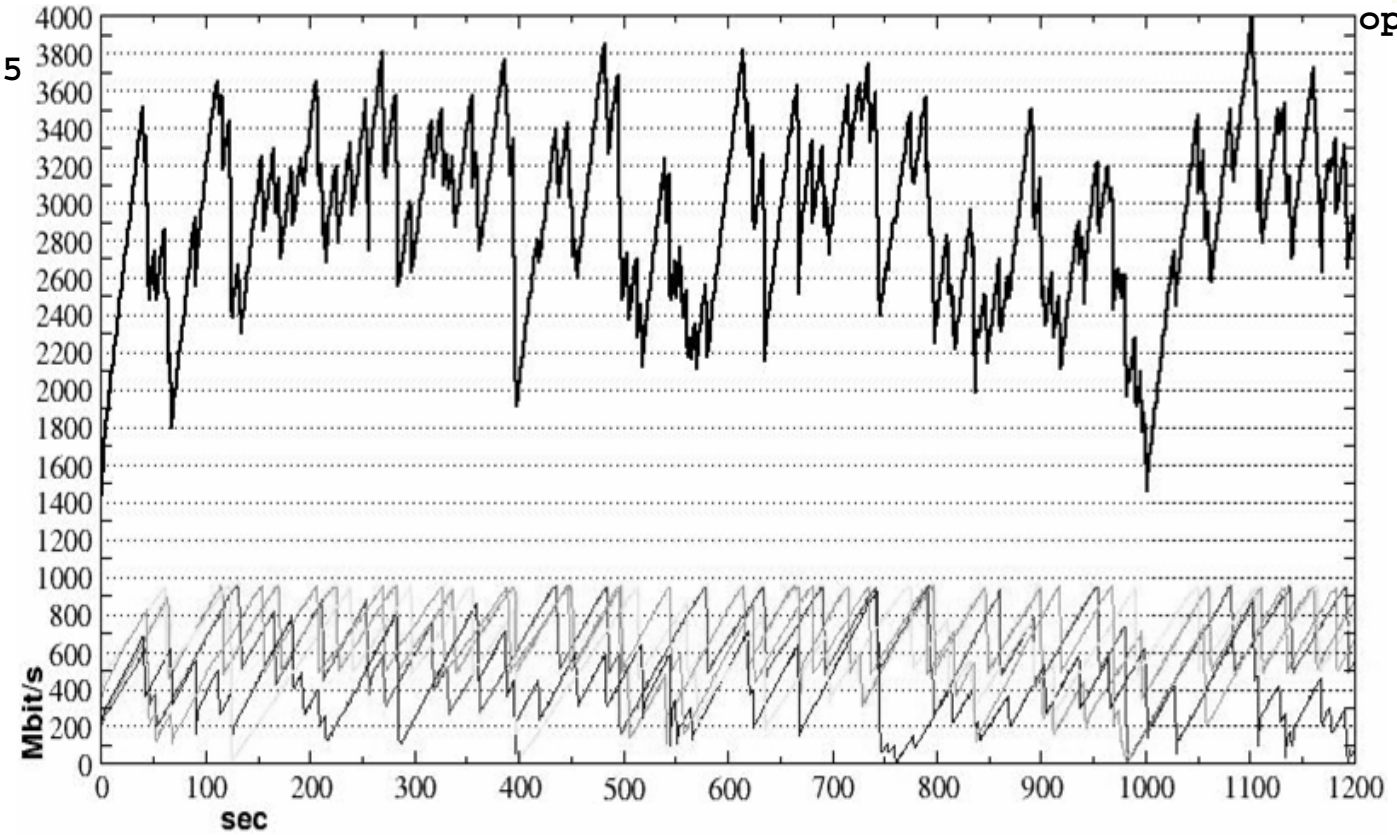
W A N — 348Mbit/sec – single stream — W A N

iperf

10gkt101 – 10gtk105

oplapro73

Reno



Bandwidth evaluation

Reno

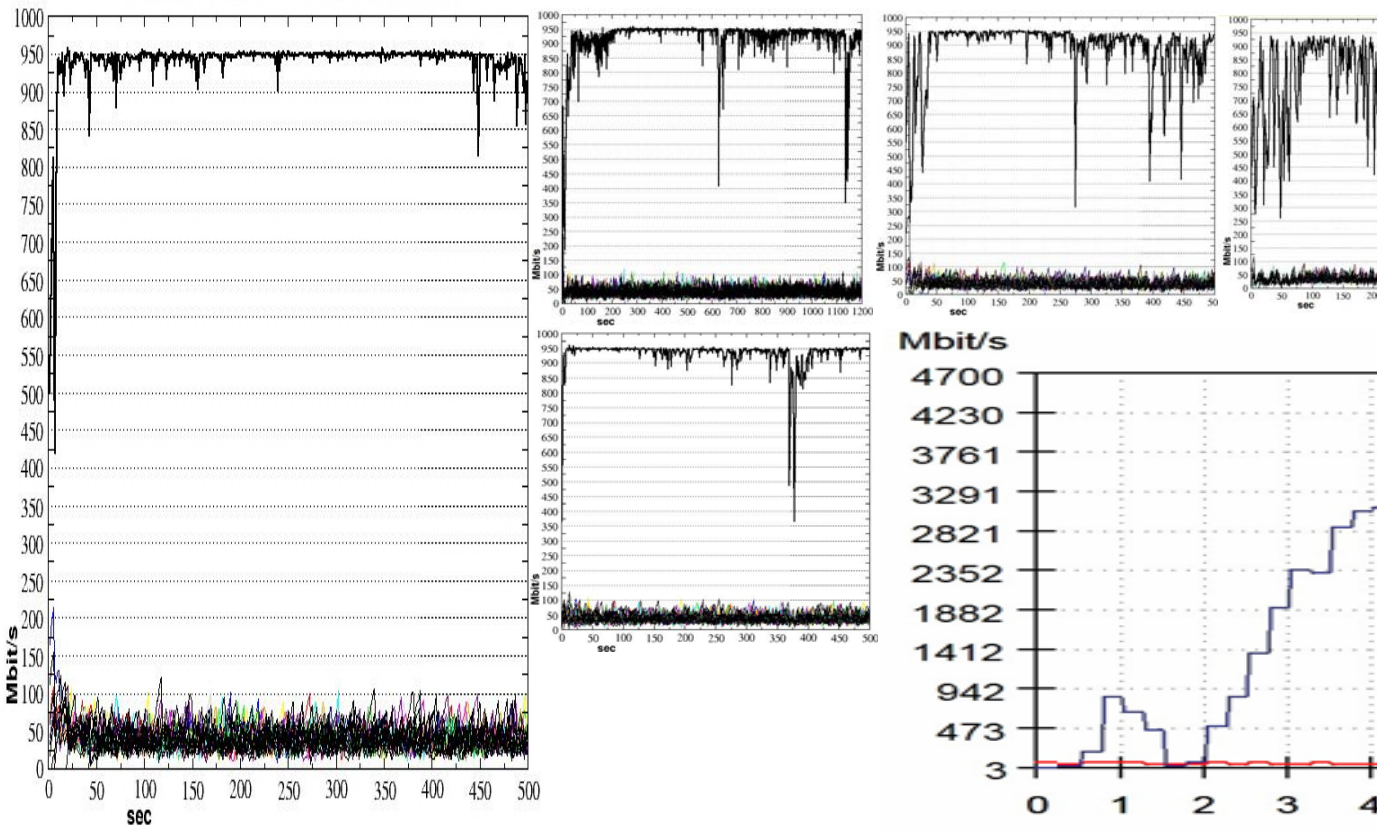
W A N — 5 nodes a 112MByte/sec – 24 parrallel streams — W A N

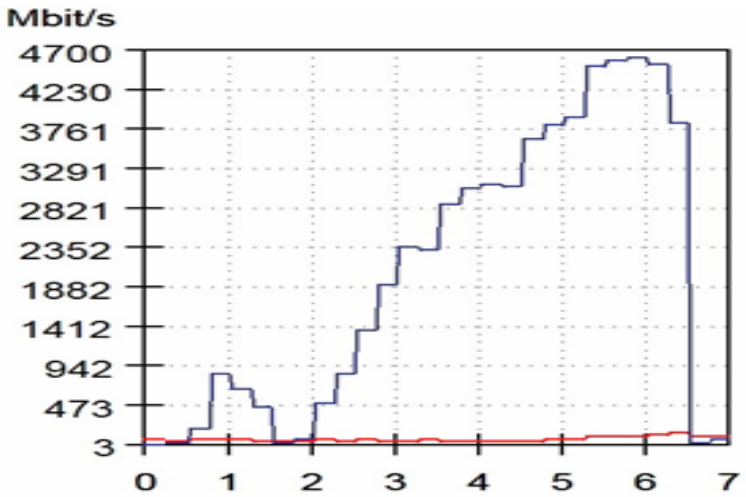
iperf

10gkt10[1-5]

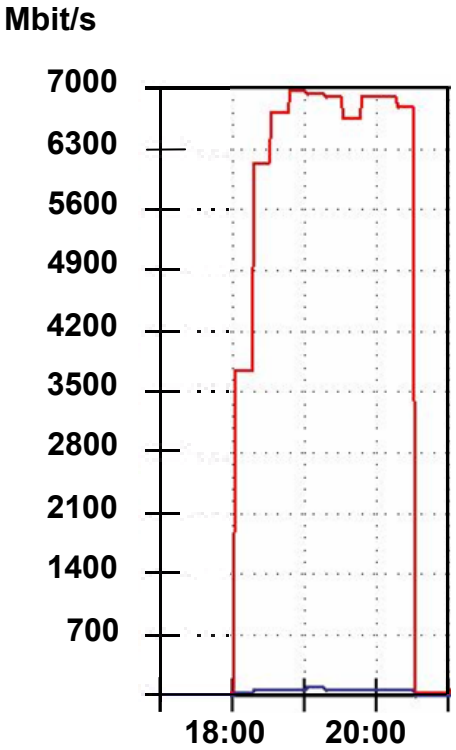
Oplapro7[1-5]

Reno





Evaluation of max throughput



9 nodes each site

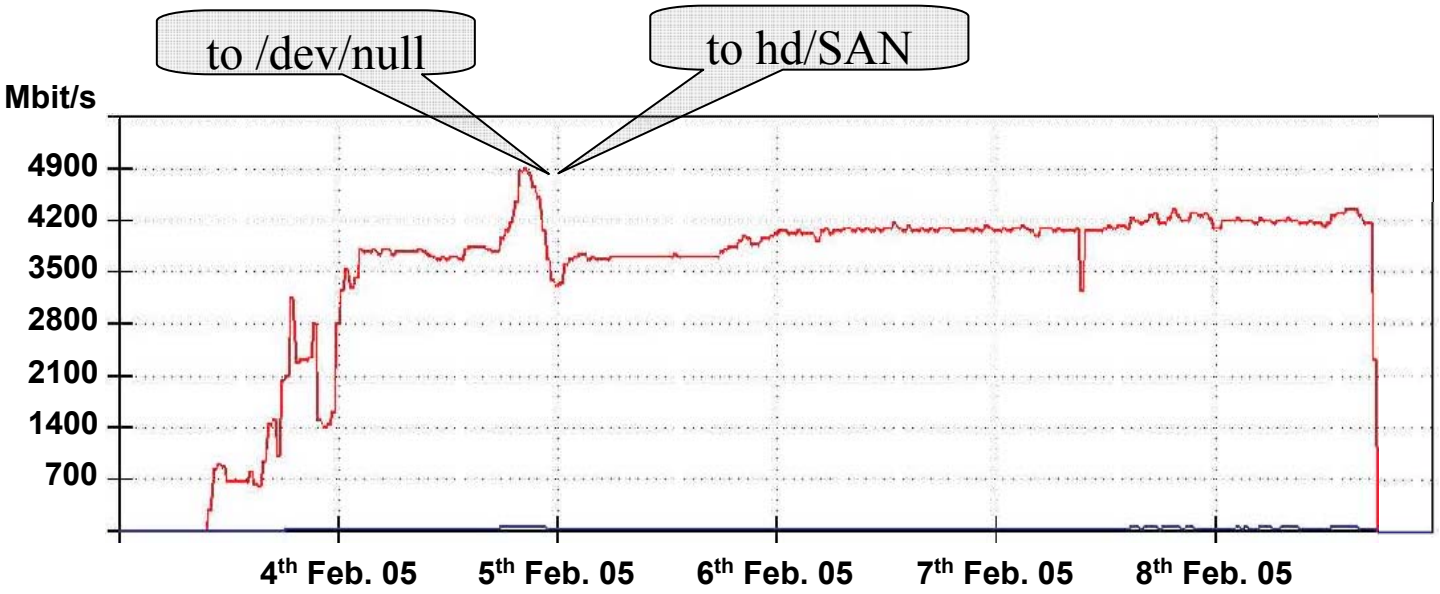
-8 * 845 Mbit

-1 * 540 Mbit

higher speed at one stream is resulting in a packet loss

Gridftp sc1 throughput

Sc1 – 500MByte sustained

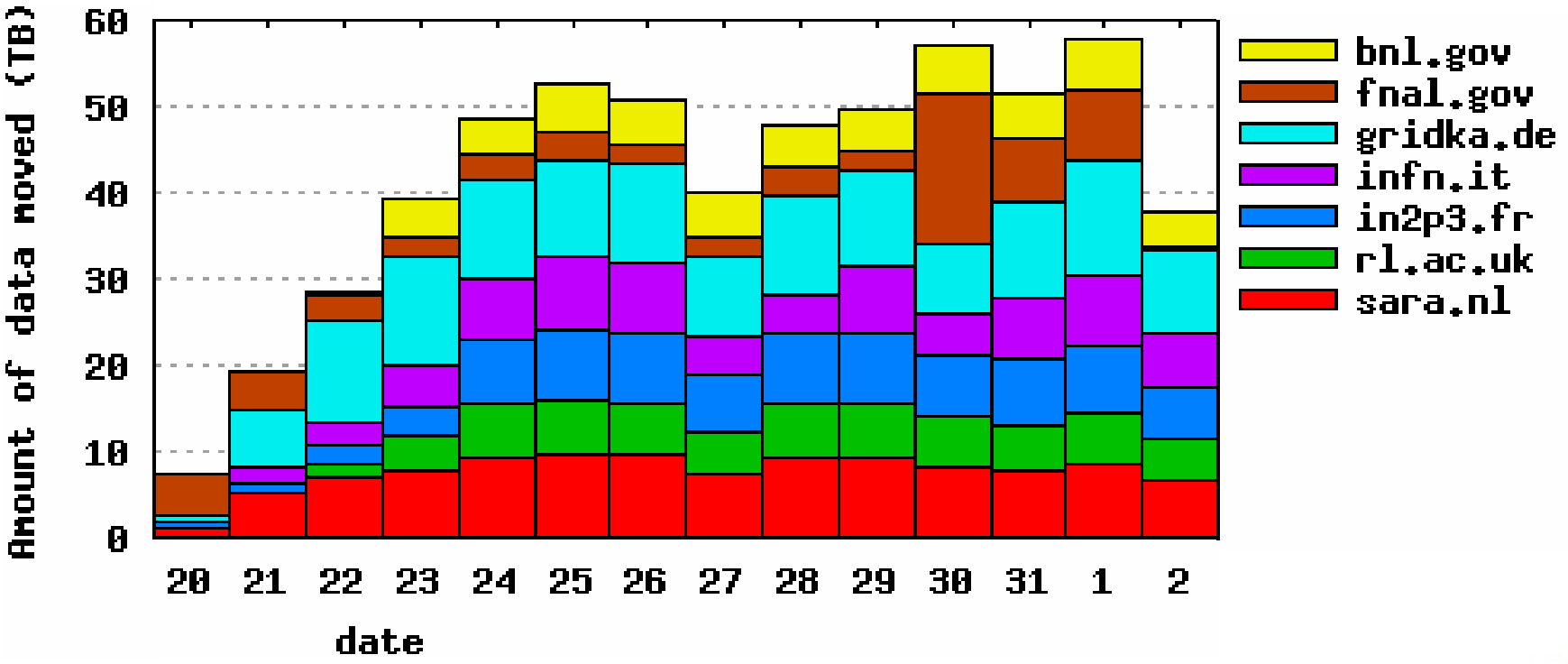


19 Nodes

- 15 WorkerNodes * 20MByte IDE/ATA HD
- 1 FileServer * 50MByte SCSI HD
- 3 FileServer * 50MByte SAN

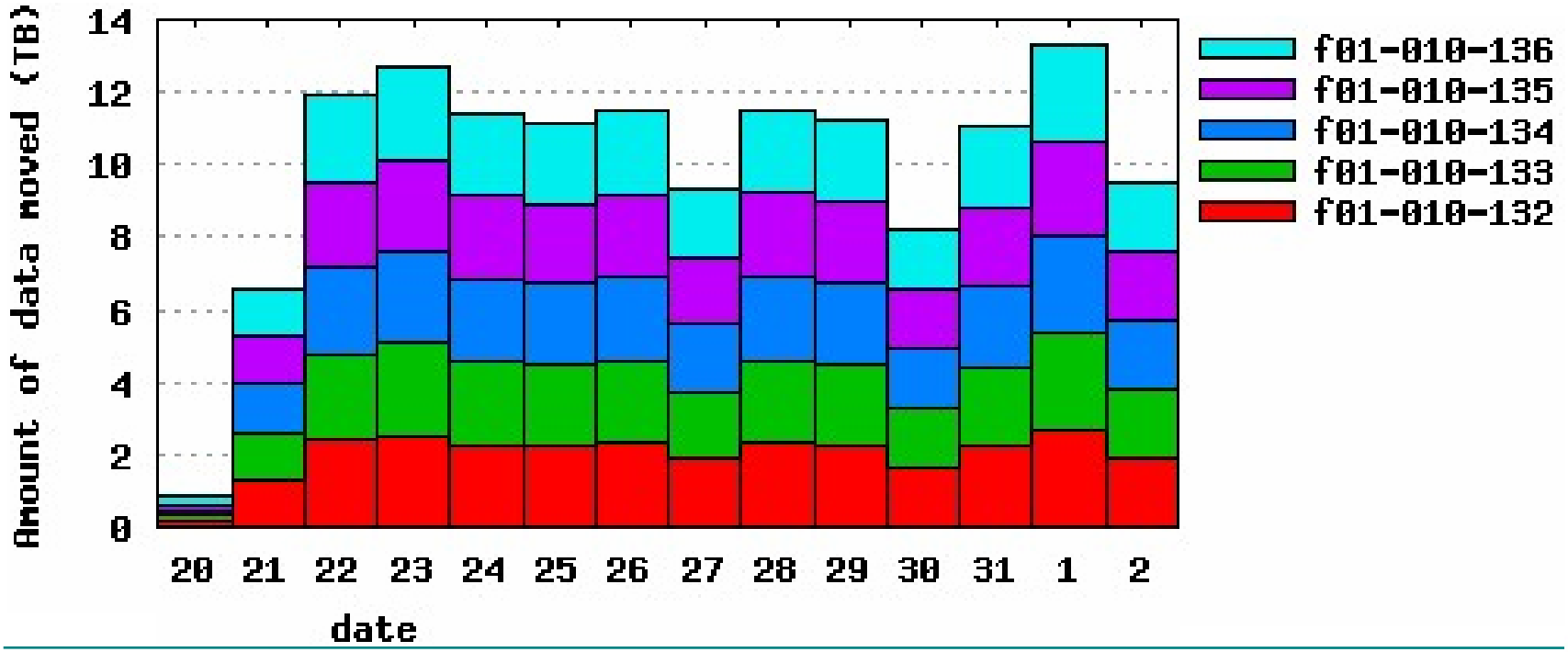
SC2

- approx 1/5 of the load



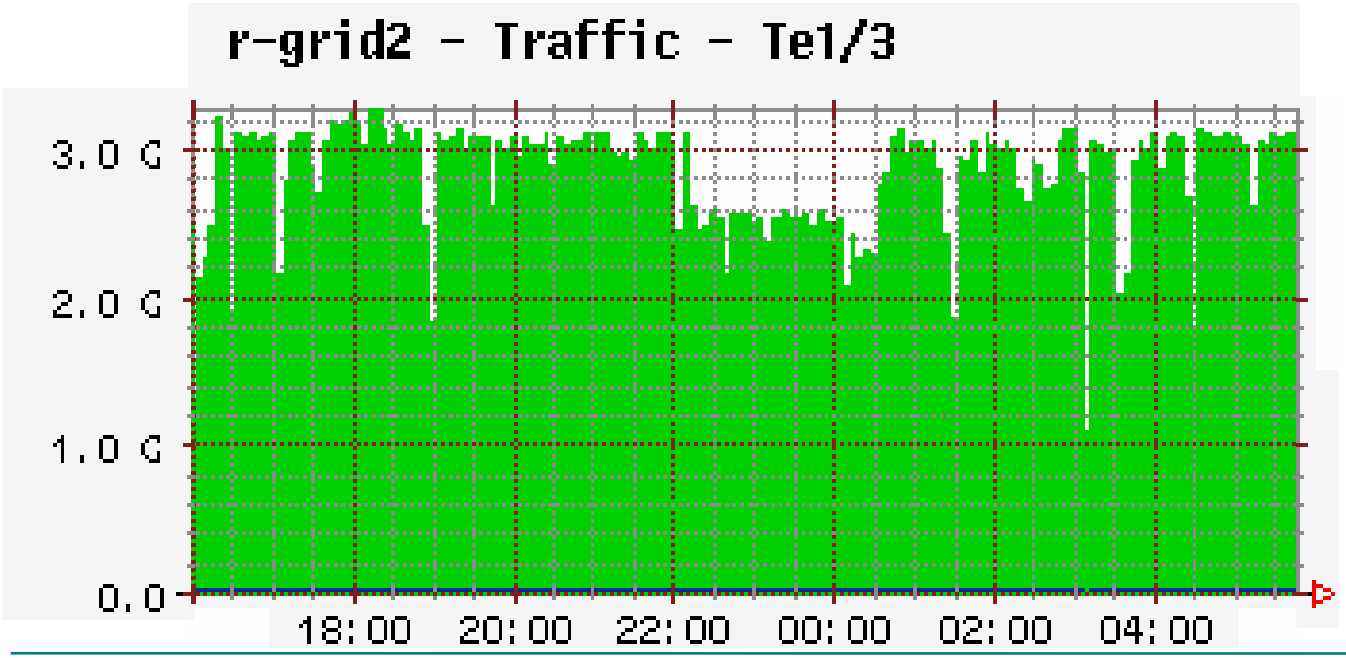
SC2

- five nodes at GridKa
- gridftp to gpfs



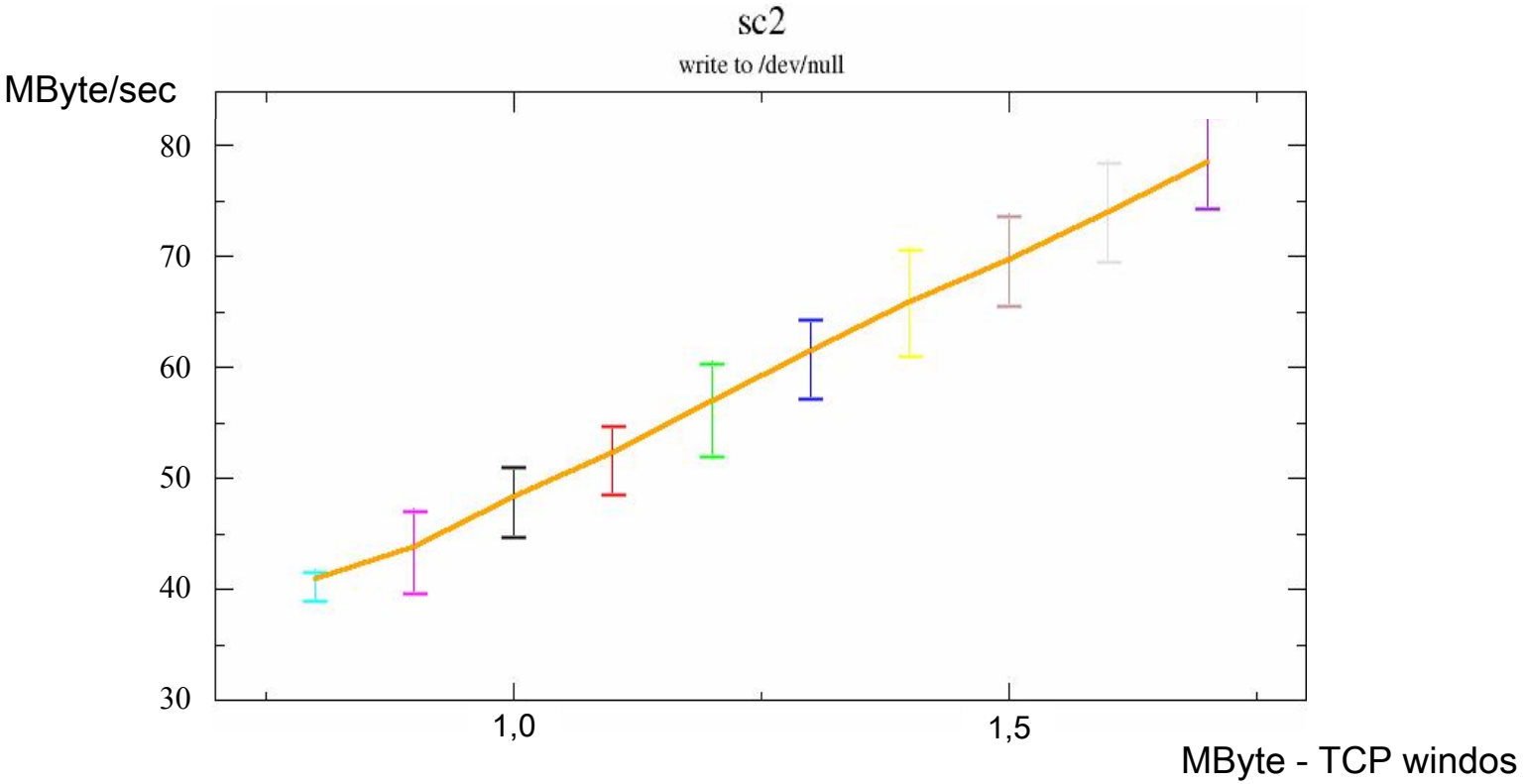
SC2-part 2

- Trouble shooting with radiant
- Shaping diffrent host performances (load bolancing)
- Parallel threads did not perfom better
- Best performance
  - •20 parallel file copies
  - •Equal Nodes (no performance diffrenz)



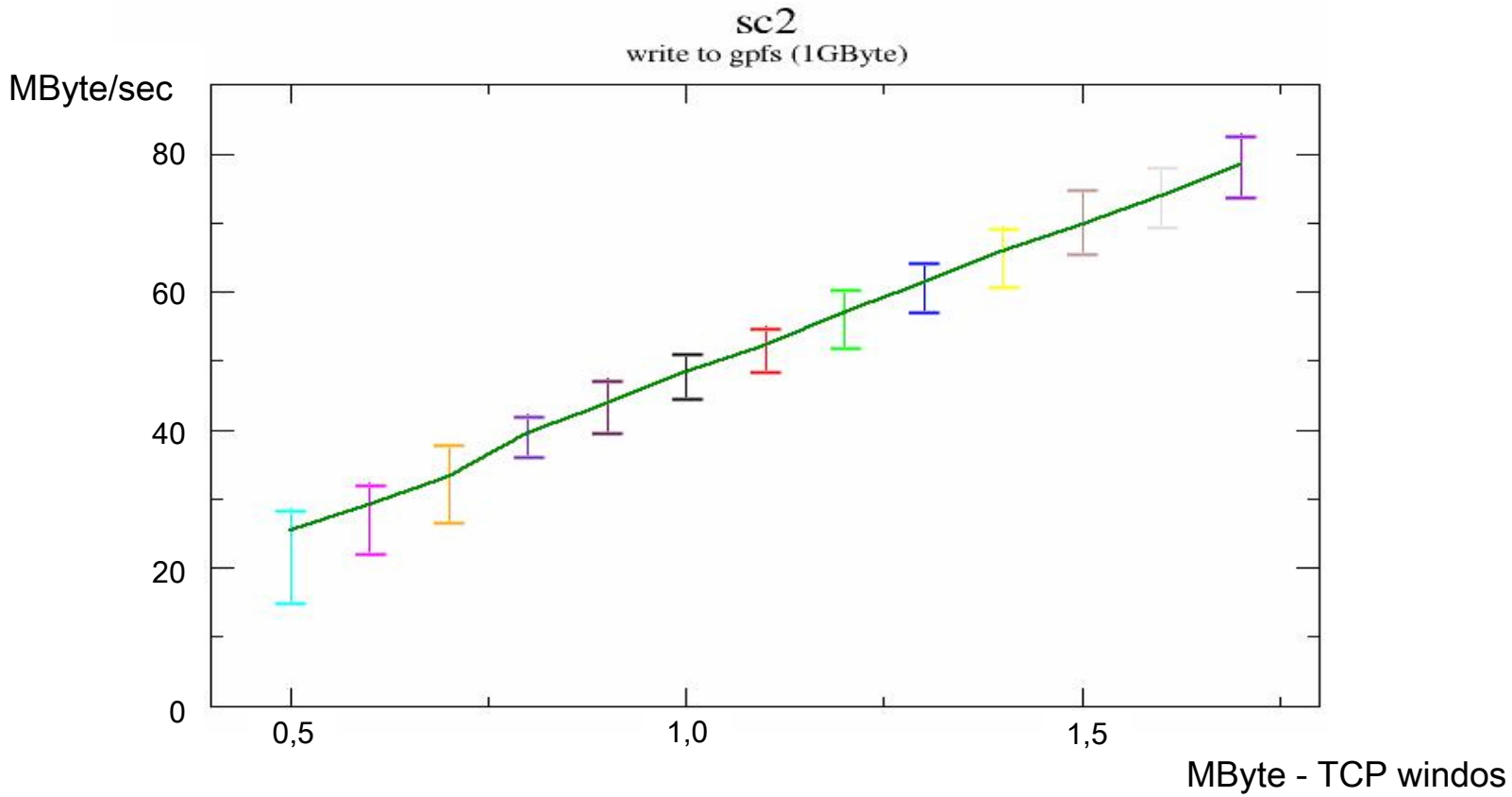
cache

Gridftp 2 /dev/null

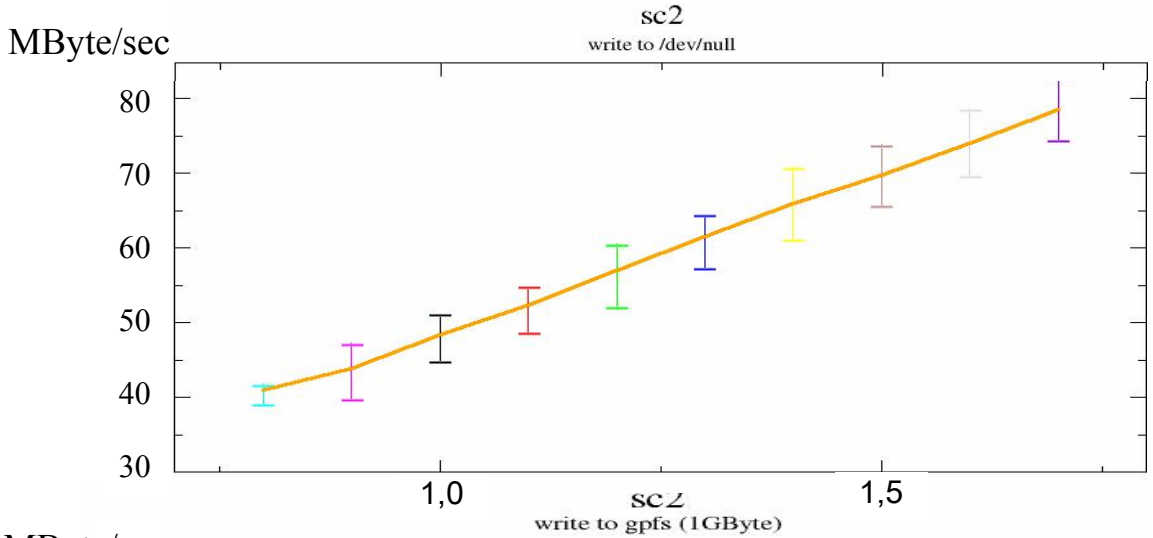


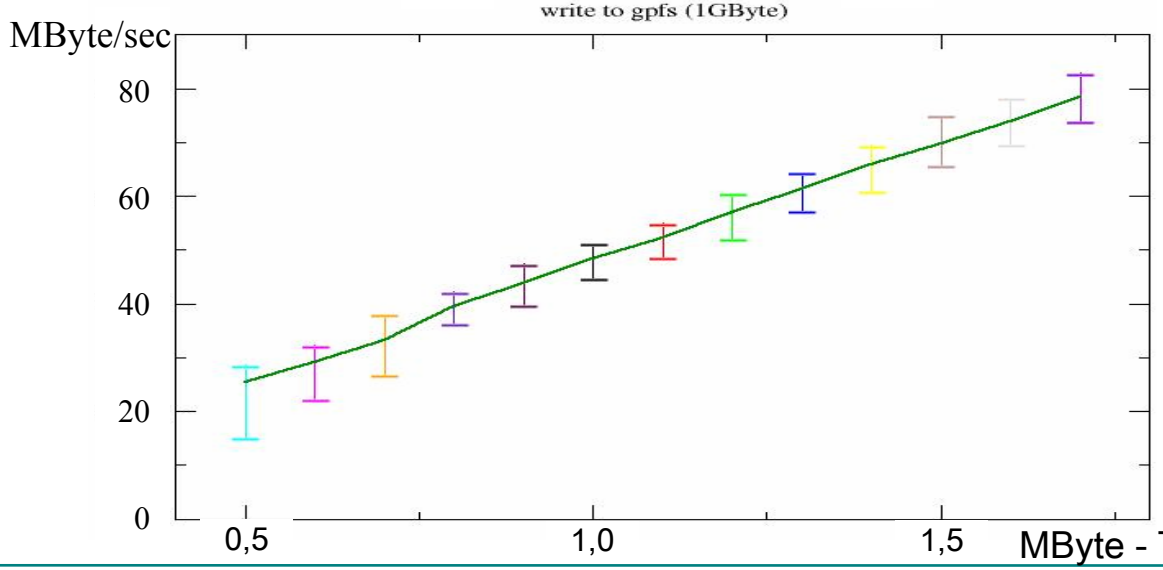
cache

Gridftp 2 gpfs



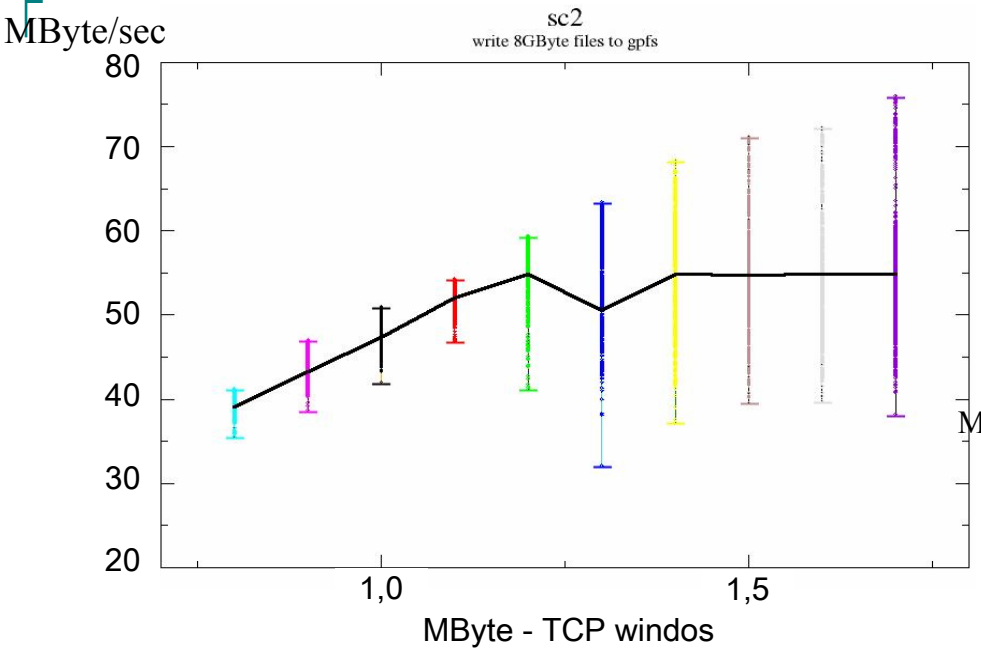
cache





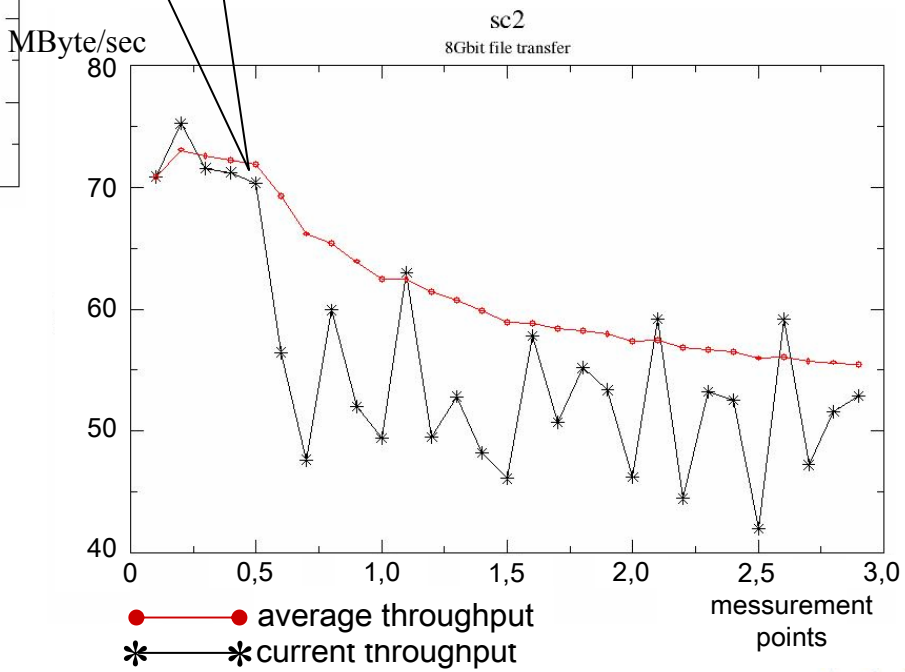
MByte - TCP windos

cache



Gridftp 8Gbyte File

2Gbyte
16% of time



Conclusion

- multi NREN 10Gbit link up to 7.0Mbit usable
- SC2 part 1 : approx. 1/5 of the CERN aggregated load
- SC2 part 2 : 250 MByte stable, peak over 400MByte
- TCP for WAN (un-)modified?

Future Work

- Gpfs for data dest. ✓ since sc2 end of March
  single stream up to 115MByte
  sustained multistream 60 to 70MByte
- Digging into HW details to discover bottlenecks (packet drop due to bad PCI timing) some solved, but still ongoing
- Stabilise the transport since sc2 is not aproaching the edge, the impreession is far more stable
- Installation of SRM
- Installation of dcache

for SC, migrate to production

- Planned 2006
  - Lightpath via X-Win (DFN) / Geant2 to CERN


thanks for your attention
?
Questions