



Disk to Network Streaming at 42 Gbit/s

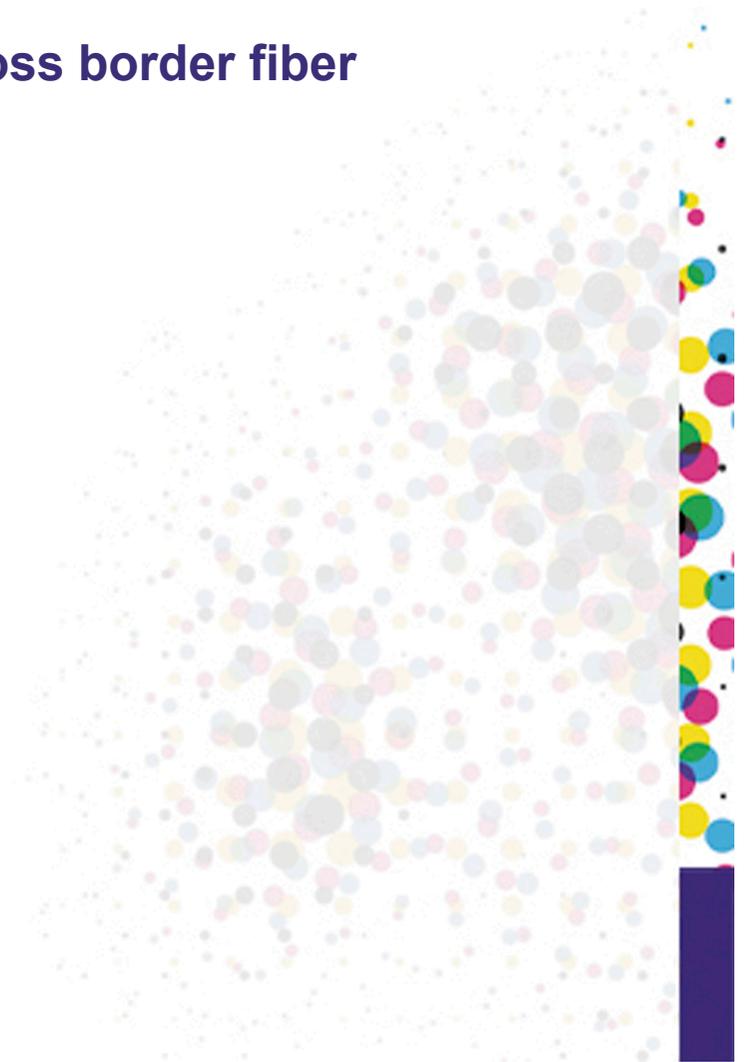
Ronald van der Pol <rvdp@sara.nl>

Paul Wielinga <wielinga@sara.nl>



Outline

- ▶ **Goal of the project**
- ▶ **Streaming server**
- ▶ **GLIF demo over 40G 1650 km cross border fiber**
- ▶ **SC10 demo over local 2x 40GE**
- ▶ **Conclusions**



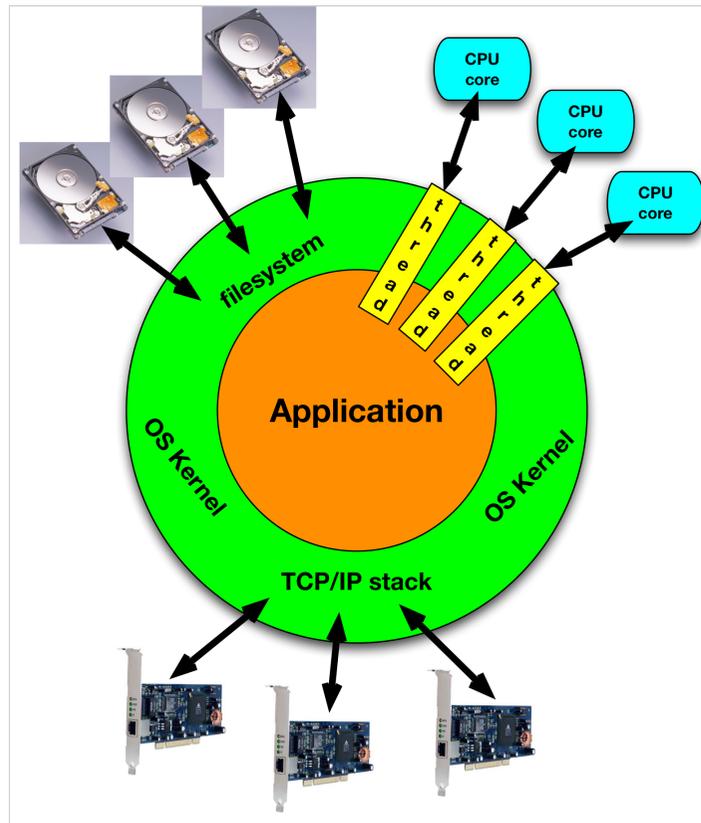


Goal of the project

- ▶ **Exploring the limits of a modern off the shelf server**
 - ▶ What network I/O can be reached from a single server?
 - ▶ What disk I/O can be reached from a single server?
- ▶ **Give demo at GLIF using SURFnet's 40G cross border fiber**
- ▶ **Explore availability of 40GE equipment**
- ▶ **Getting hand-on experience with 40GE equipment**

Multiple Disks, Cores and NICs

- Storage I/O speedup with multiple disks (RAID-1/RAID-Z)
- Compute speedup with multi-core systems
- Network I/O speedup with multiple NICs

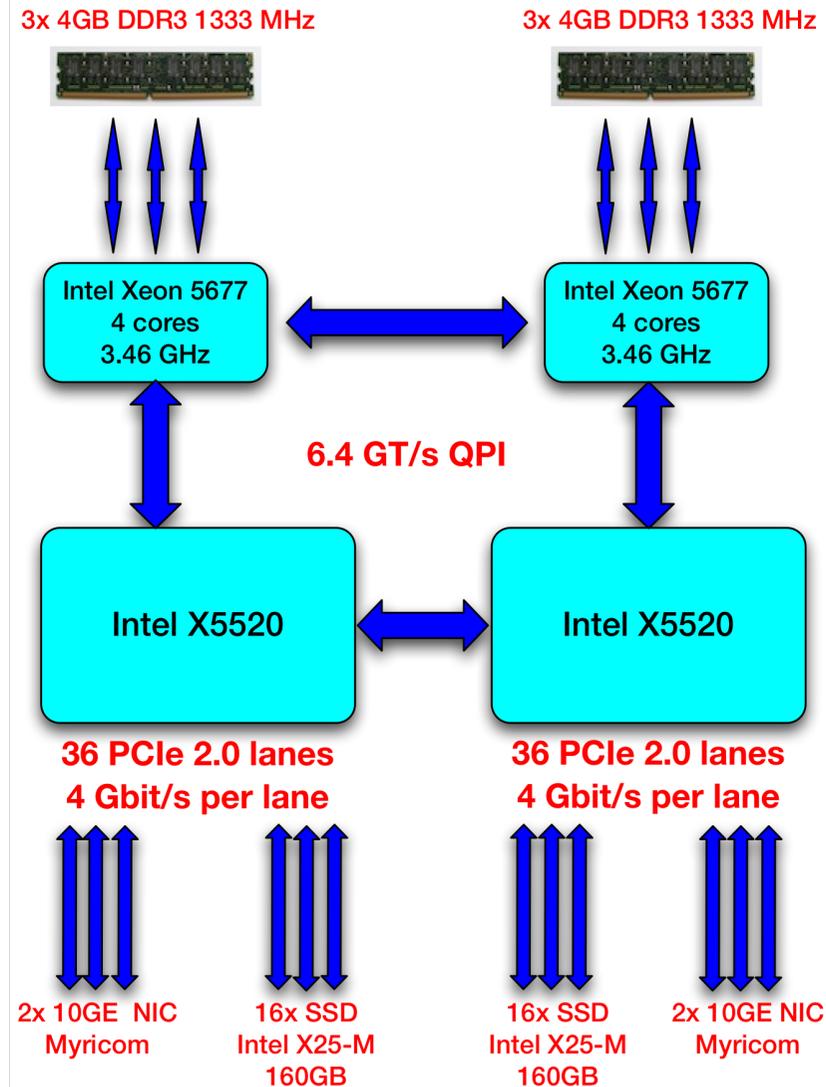




Streaming Server Architecture

Supermicro X8DAH+-F motherboard
2x PCI-E 2.0 x16
4x PCI-E 2.0 x8
1x PCI-E 2.0 x4
8 cores @ 3.46GHz (Intel Xeon 5677)
24GB DDR3 @ 1333 MHz
4x 10GE Myricom (2 dual port NICs)
32x SSD Intel X25-M 160GB

Eur 22,000 (June 2010)
SSDs take up 50% of price





SARA's 5x3 Tiled Panel



- 15x LCD 2560x1600
- 12,800x4,800 pixels
- 8x Dell T5500 servers
 - 10GE NIC
 - Nvidia GeForce GTX460
 - 2 LCDs per server



CosmoGrid

- ▶ Dutch supercomputing challenge project
 - ▶ prof. Portegies Zwart (University of Leiden)
- ▶ Simulation of 256^3 and 2048^3 bodies of dark matter
- ▶ Simulation shows formation of clusters after big bang
- ▶ Distributed application using several European and Japanese supercomputers, including Huygens @ SARA

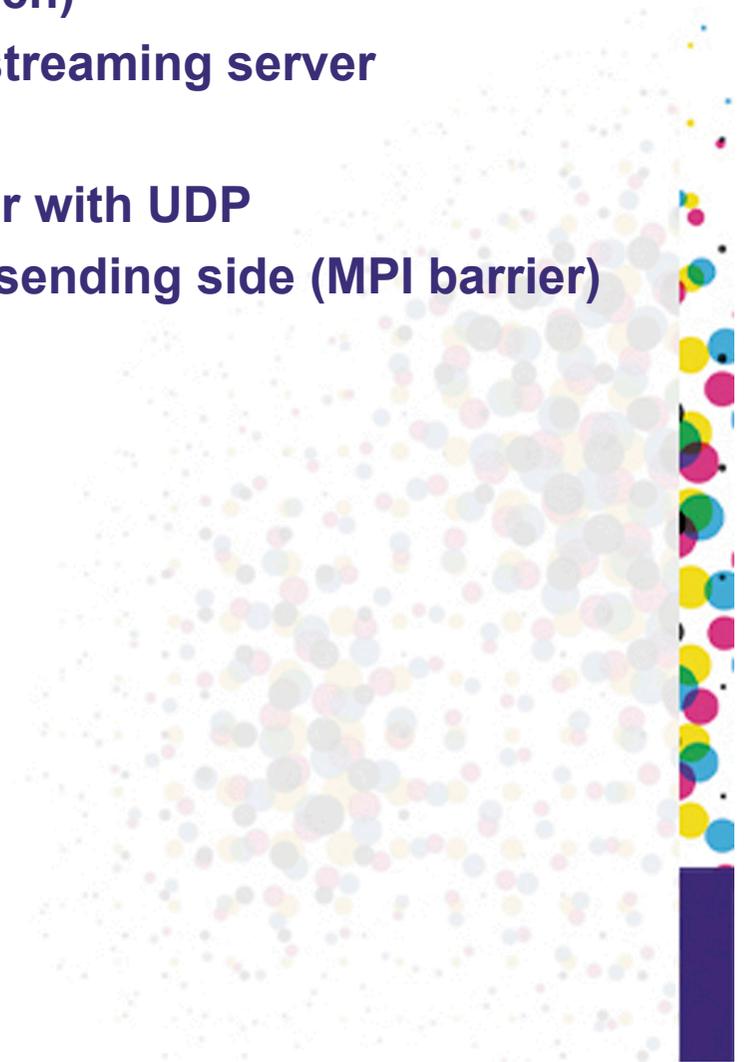


CineGrid Workshop, 12-15 dec, La Jolla, USA



CosmoGrid Visualisation

- ▶ **Pre-rendered images**
- ▶ **15 movies (1 movie for each screen)**
- ▶ **Movies stored on SSD disks on streaming server**
- ▶ **24 bit RGB uncompressed**
- ▶ **Each movie streamed from server with UDP**
- ▶ **15 MPI threads, synchronized at sending side (MPI barrier)**



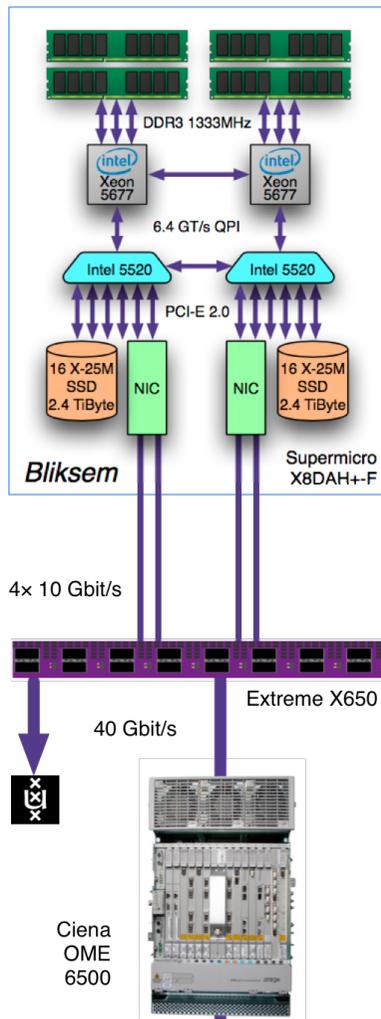


GLIF Streaming Demo

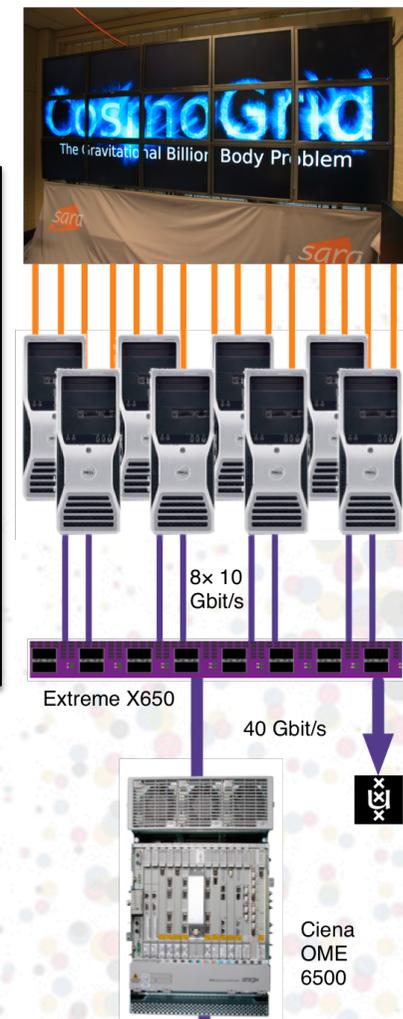
- ▶ **Annual Global LambdaGrid Workshop**
 - ▶ 13-14 October @ CERN in Geneva
- ▶ **Streaming demo over SURFnet's 40G cross border fiber between Amsterdam and Geneva**
- ▶ **Streaming from a single server in Amsterdam to the TPD in Geneva**
- ▶ **4x 10GE in streaming server**
- ▶ **15 UDP streams evenly balanced over 4x 10GE NICs of streaming server**

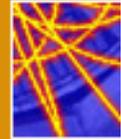


Topology GLIF Demo



Amsterdam to Geneva
1650 km 40G link
4x 10GE client side
Extreme x650
Ciena 65000
33 Gbit/s disk to network
38 Gbit/s disk to network (no synch)





[Home](#) > [iSGTW - 3 November 2010](#) > Feature - Ultra-fast networks: the final frontier

Feature - Ultra-fast networks: The Final Frontier

Researchers from Holland have demonstrated a network infrastructure that could potentially help scientists save time and even transform the movie business. This could be done without the need for large computer clusters or grids, just off-the-shelf hardware components combined with human ingenuity and one of the world's fastest research networks. The team were from [SARA](#), a Dutch supercomputing and e-science support center.

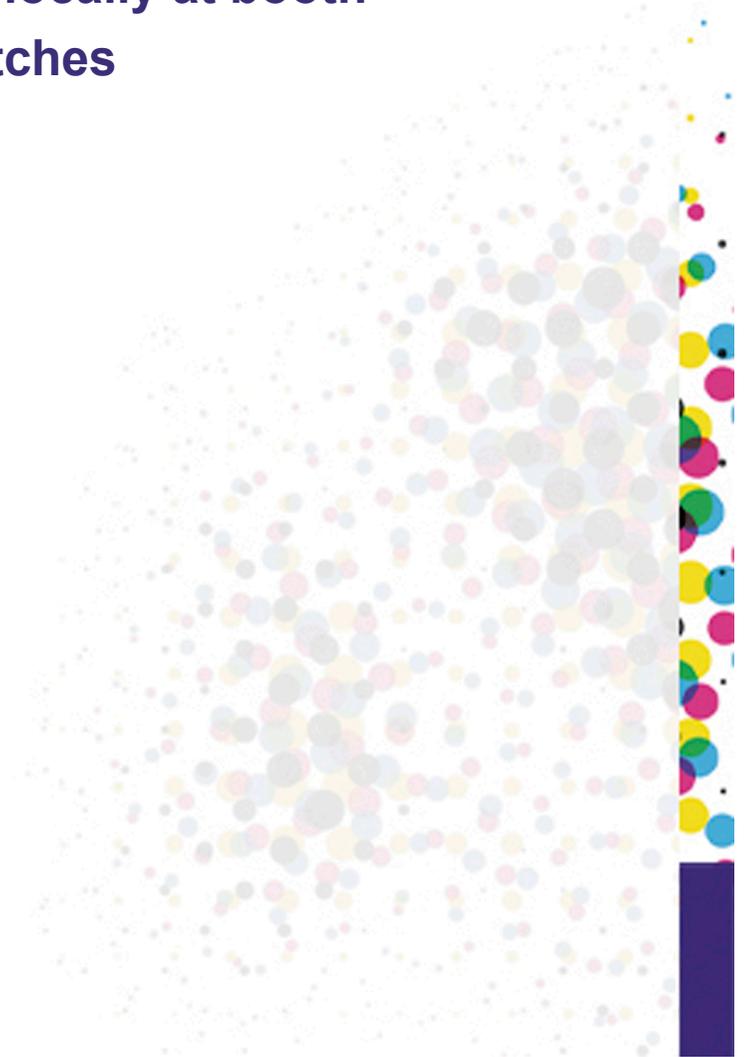


A network researcher in awe of the billions of dark matter particles simulated on 15 ultra-high definition monitors. *Image courtesy Freek Dijkstra*



SC10 Streaming Demo

- ▶ **Supercomputing 2010 in New Orleans**
- ▶ **Streaming server and tiled panel locally at booth**
- ▶ **Two Extreme Networks x650 switches**
 - ▶ **2x 40GE between switches**
- ▶ **Two topologies tried**
 - ▶ **6x 10GE in streaming server**
 - ▶ **2x 40GE in streaming server**





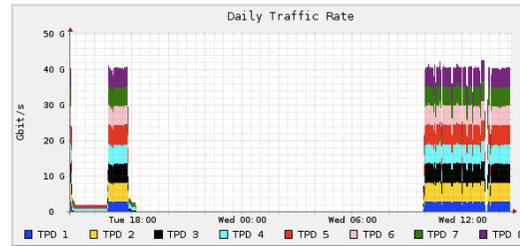
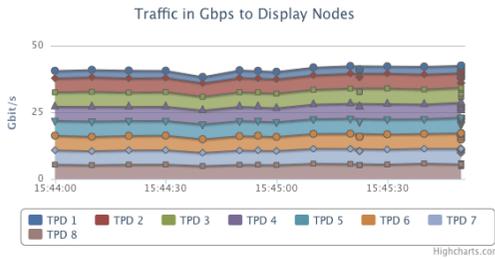
SC10 Demo



Single Server 2x 40GE Disk to Network Streaming

Live Monitoring

Network Throughput



Sending Server

6x 10GE



Extreme X650

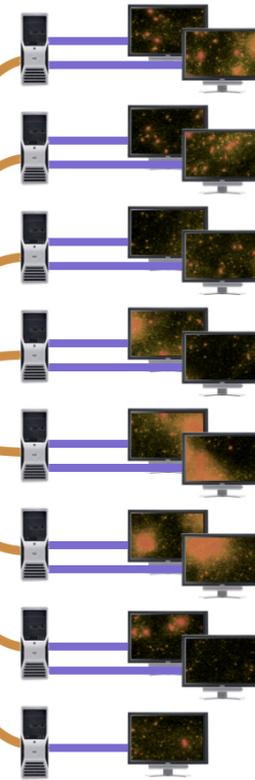
Current traffic rate: 40.7 Gbit/s

2x 40GE

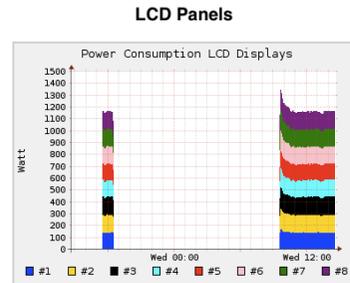
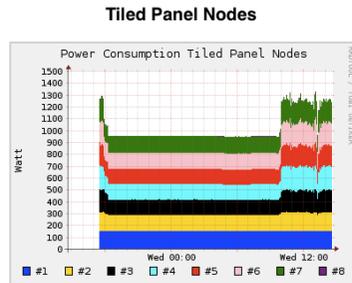
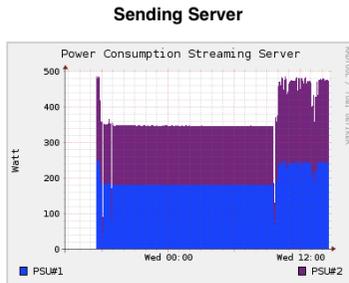


Extreme X650

8x 10GE



Power Consumption



Sponsored by:



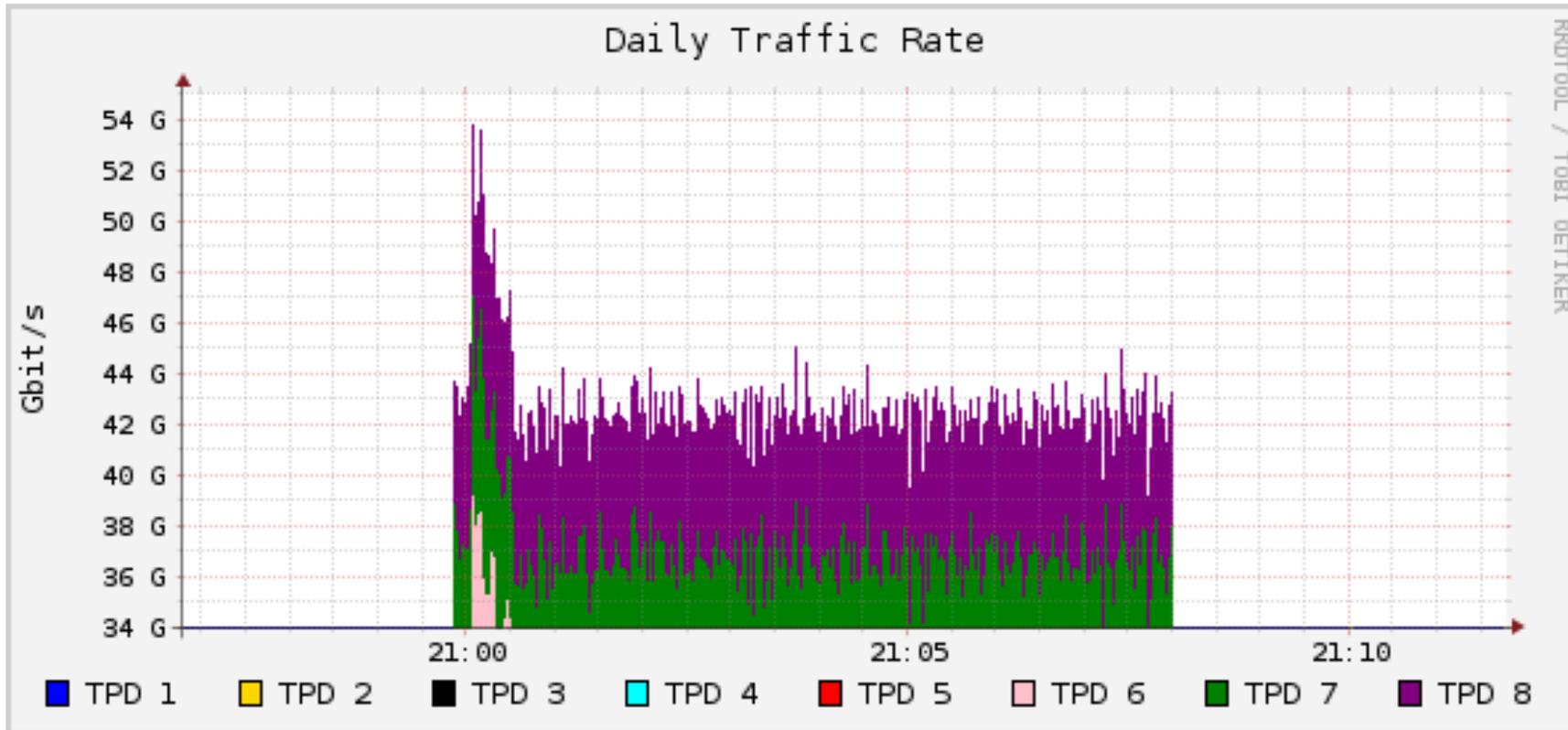
In Collaboration with:



CineGrid Workshop, 12-15 dec, La Jolla, USA



Results





Server & Extreme X650s



CineGrid Workshop, 12-15 dec, La Jolla, USA



2x 40GE Disk to Network

sara 2×40 Single Server 6×10 GE Disk to Network Streaming

Single Server 2x 40GE Disk to Network Streaming

SARA has built a server with off-the-shelf components that is capable to transmit over 30 Gbit/s from disk to the network. Previously, clusters and grid setups were needed to reach these application speeds. This opens new possibilities for demanding applications.

A systematic approach was used to tune the end-to-end performance, with studies on the systems, I/O system design, processor and interrupt affinity, link bonding and 4x Gbit/s interworking. The result is a scientific visualization demonstration that shows how applications can benefit from an integrated approach towards end-to-end performance optimization.

Server Architecture
The total capacity in a high-end server is sufficient to achieve high bandwidth between disk, memory, CPU and network. The speed of PCI Express slots was a practical limitation.

Disk Performance
Solid state disks (SSD) promise low energy consumption, high reliability (due to the lack of moving parts), and near-zero access times. Solid State Disks achieve a read speed of 500 MB/s (four times as fast as hard disk drives). We were able to reach a sustained read speed of 500 MB/s (5x Gbit/s) with 4 solid state disks. 2x (mSATA), 2x (SATA) and 2x (SATA) could all achieve these speeds, provided that the checksumming used hardware acceleration.

Streaming Server

Sending Server

Motherboard	Supermicro X8DML-F	Monitor	19" Dell 3000 WFP 30"
CPU	2x Intel Xeon X5375	GPU	1x NVIDIA 9800 GTX
Memory	24 GB DDR3	Frame	3x3 members
Disk	2x 4x 500 GB SATA	Switch	1x Cisco WS-C4506R
PCI System Controller	2x 4x SATA 9750-BI	PC	1x Dell Precision T3200
Network	2x Mellanox ConnectX-2 4x GbE	Video card	1x NVIDIA GeForce GTX 480
OS	Linux 2.6.31 (64 bit)		

Tiled Panel Display

Visualization Cluster

Remote Visualization

A tiled panel display (TPD) consists of any number of monitors, stacked in a grid pattern. It allows a researcher to view high resolution data, both in real detail and in a bird's eye view at the same time.

Remote visualization allows geographic separation between the creation (rendering) and display of data. A display system can be relatively simple in terms of hardware and software. One possible display system is a tiled panel display which allows high resolution content to be displayed.

CosmoGrid

CosmoGrid simulates the structure of the universe using 8,579,939,100 bodies of dark matter, and was generated on multiple supercomputers, including SARA's Huggins. The project is led by prof. dr. Simon Portegies Zwart of the Leiden Observatory.

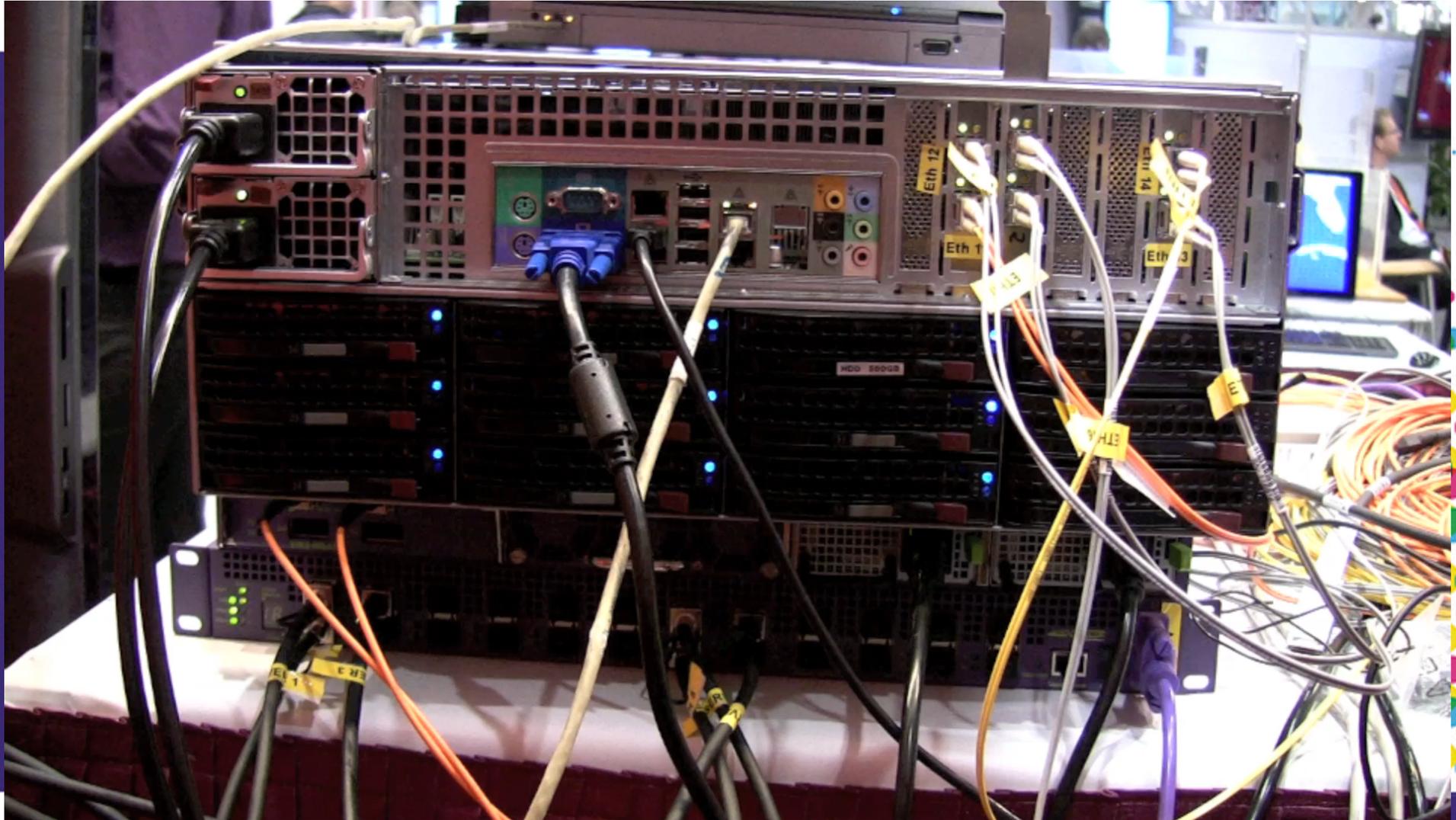
From two different simulation datasets visualizations were created by SARA. For this demonstration the visualizations were pre-rendered and stored on disk, while still allowing the user to interactively choose which visualization to show.

SARA - LIVE 40GE monitoring

CineGrid Workshop, 12-15 dec, La Jolla, USA



SC10 6x 10GE Demo



CineGrid Workshop, 12-15 dec, La Jolla, USA

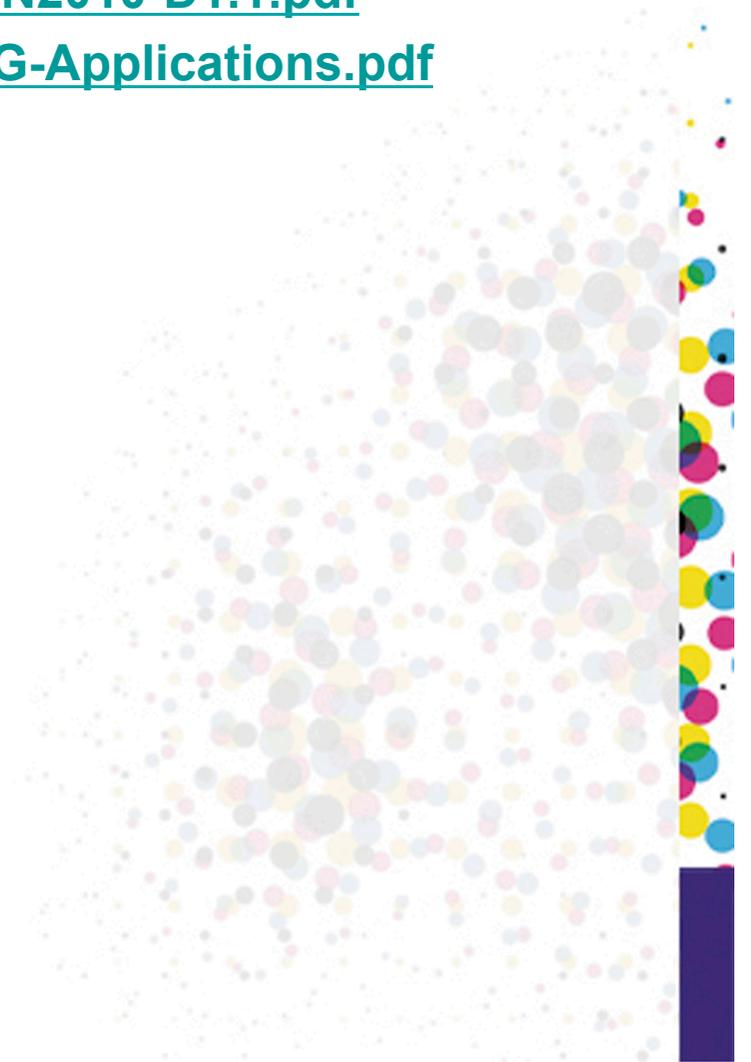
Conclusions

- ▶ **Modern off the shelf server can transmit 42 Gbit/s from disk to network**
- ▶ **54 Gbit/s memory to network**
- ▶ **40 Gigabit Ethernet not production quality yet**
 - ▶ **Most equipment still in beta or development**
 - ▶ **More interoperability testing needed**
- ▶ **Testing with real applications useful to find problems**



Additional Information

- ▶ <http://nrg.sara.nl/>
- ▶ <http://nrg.sara.nl/publications/RoN2010-D1.1.pdf>
- ▶ <http://nrg.sara.nl/publications/40G-Applications.pdf>
- ▶ Email: nrg@sara.nl





Acknowledgements

- ▶ **SARA:** *Pieter de Boer, Freek Dijkstra, Igor Idziejczak, Tijs de Kler, Paul Melis, Hanno Pet, Peter Tavenier, Paul Wielinga*
- ▶ **SURFnet:** *Gerben van Malenstein, Erik-Jan Bos*
- ▶ **Extreme Networks:** *Rene Huntelerslag, Gihad Ghaibeh, Ramon Semmekrot, Trung Tran, Renuke Mendis*
- ▶ **Mellanox:** *Bill Lee, Yossi Avni, Glenn Church*
- ▶ **Ciena:** *Martin Bluethner, Jan Willem Elion, Kevin Mckernan, Harry Peng, David Yeung*
- ▶ **CERN:** *Edoardo Martelli*
- ▶ **Leiden Observatory:** *Simon Portegies Zwart*

- ▶ **Partly funded by GigaPort3 and CineGrid.nl (EFRO, Pieken in de Delta, Provincie Noord-Holland, Gemeente Amsterdam)**



Thank you

Ronald van der Pol
rvdp@sara.nl

CineGrid Workshop, 12-15 dec, La Jolla, USA

