A decorative graphic consisting of a thin yellow circle and a thick yellow bracket on the right side, framing the title text. The title text is centered within a light olive-green rectangular background.

Deploying distributed network monitoring mesh for LHC Tier-1 and Tier-2 sites

Phil DeMar, Maxim Grigoriev **Fermilab**

Joe Metzger, Brian Tierney **ESnet**

Martin Swany **University of Delaware**

Jeff Boote, Eric Boyd, Aaron Brown, Matt Zekauskas, Jason Zurawski **Internet2**

Presented at **CHEP2009**
Prague, Czech Republic

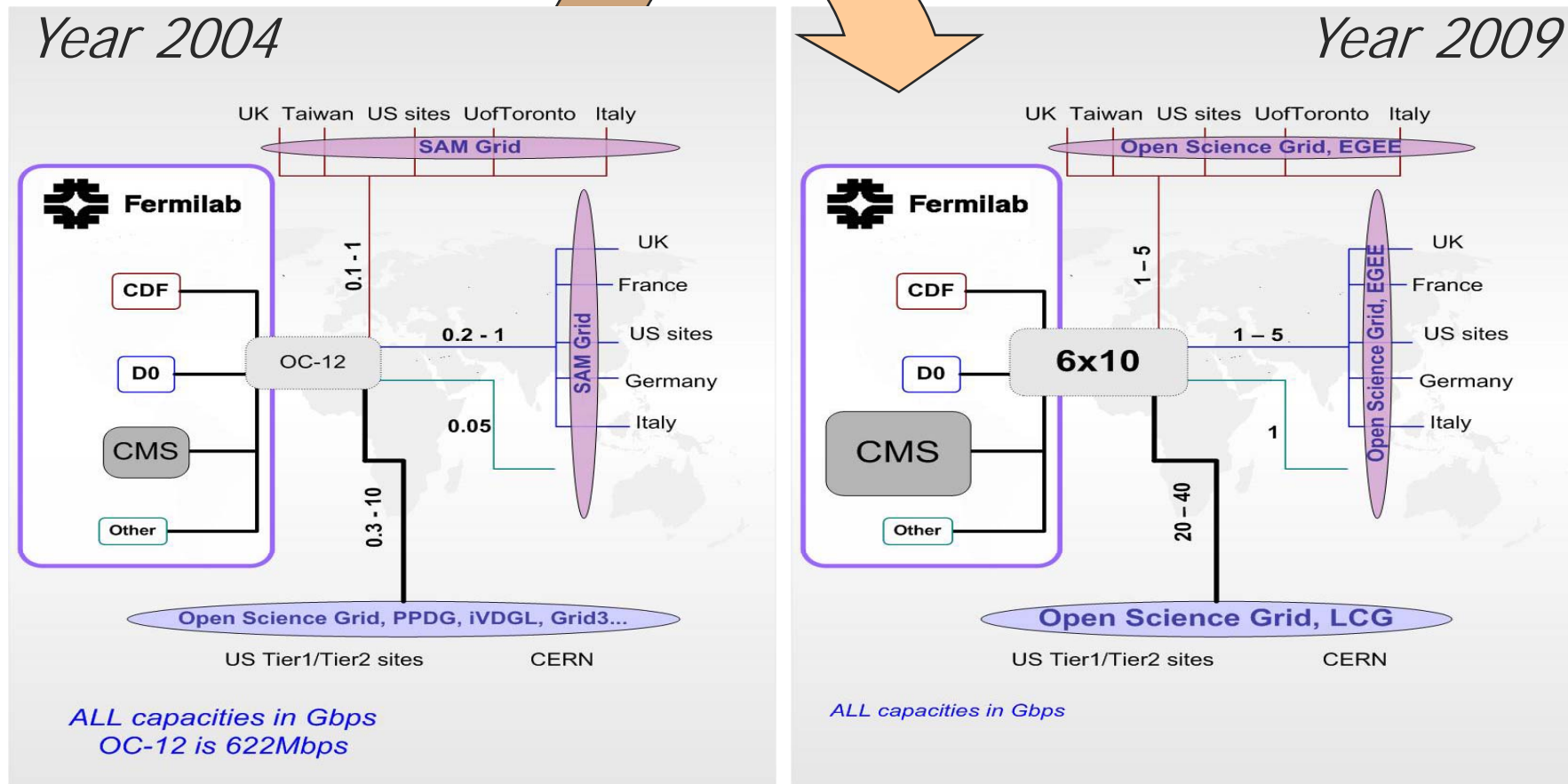
Outline

- Challenges of Wide Area Networking
- From centralized network monitoring model to distributed mesh of monitoring services
- perfSONAR-PS collection of webservicees
- Deployment at LHC Tier-1 and Tier-2 centers

[Overview]

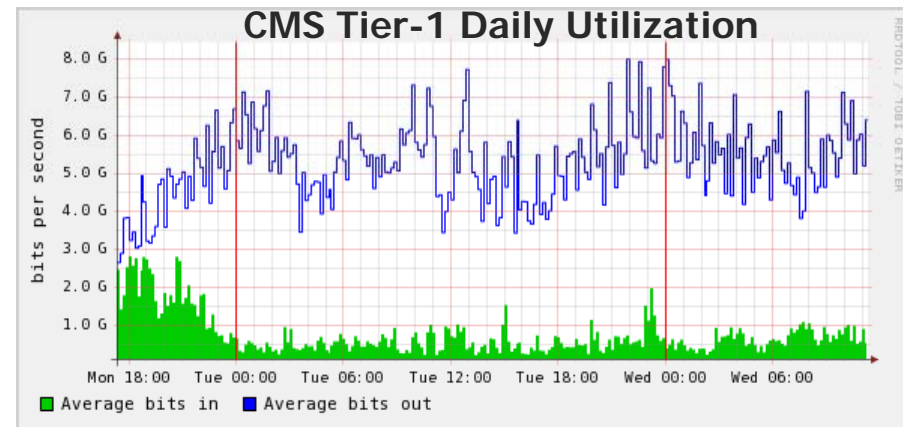
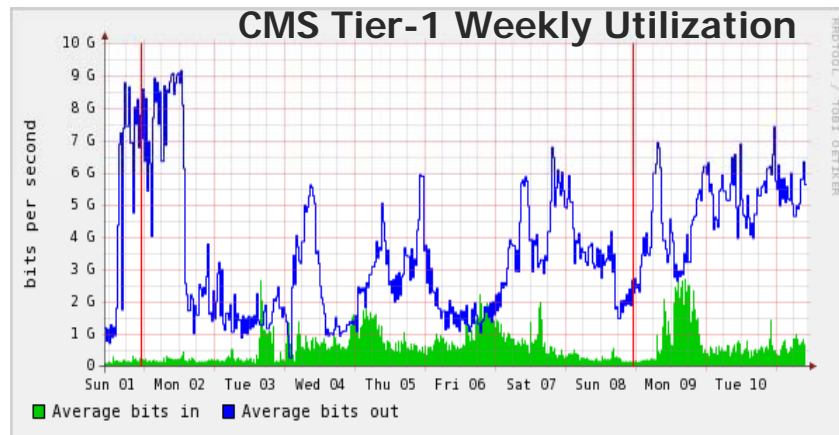
- Everyone know how to “**ping**” but how many know how to share results of it ?
- Centralized monitoring models failed to deliver scalable robust network monitoring solutions
- Everything is a service, I mean **everything**
 - Network
 - Computational facility
 - Storage ...
- Let's think about network monitoring as Service Oriented Architecture

Fermilab's WAN connectivity



Just Numbers

- 4x10Gbps ESnet Science Data Network channels with dynamic circuit reservation system
- 2x10Gbps routed channels
- It's very easy to saturate 10Gbps (March 2009)



[perfSONAR]

- Collection of interoperable webservices
- New set of XML schema and protocols
- Every network monitoring tool as a service
- Mesh of deployed monitoring services as
 - *Network Monitoring Service*
- **perfSONAR-PS is perfSONAR services implemented in perl**

[perfSONAR-PS services]

- **PingER** – based on ping, very lightweight
- **SNMP** – used for interface utilization/errors, possible to extend for any MIBs
- **perfSONAR-BUOY** – active measurements
 - **BWCTL** – iperf on demand, scheduling, AA
 - **OWAMP** – one way delay, scheduling, on demand
- **Information Service** - services discovery, two-tiered
 - Lookup Service
 - Topology Service

Current state of perfSONAR-PS

- about 100 services are running
- **ESnet** – US Energy Science network is covered
- **Internet2** – largest R&D network in US is covered
- Tier-1 sites in US – BNL and FNAL are running LHCOPN Layer2 monitoring, LHC monitoring nodes
- plan to deploy ~ 200 services on 30 networks by the end of Year 2009

[NPToolkit]

- Based on Knoppix Live Linux CD disk
- **Web100** kernel
- **perfSONAR-PS** services + **NPAD** and **NDT**
- Packaged Apache webserver, MySQL DB, Oracle XML DB
- Cacti, RRDtools, Cricket
- Zero Configuration, Out of Box Service

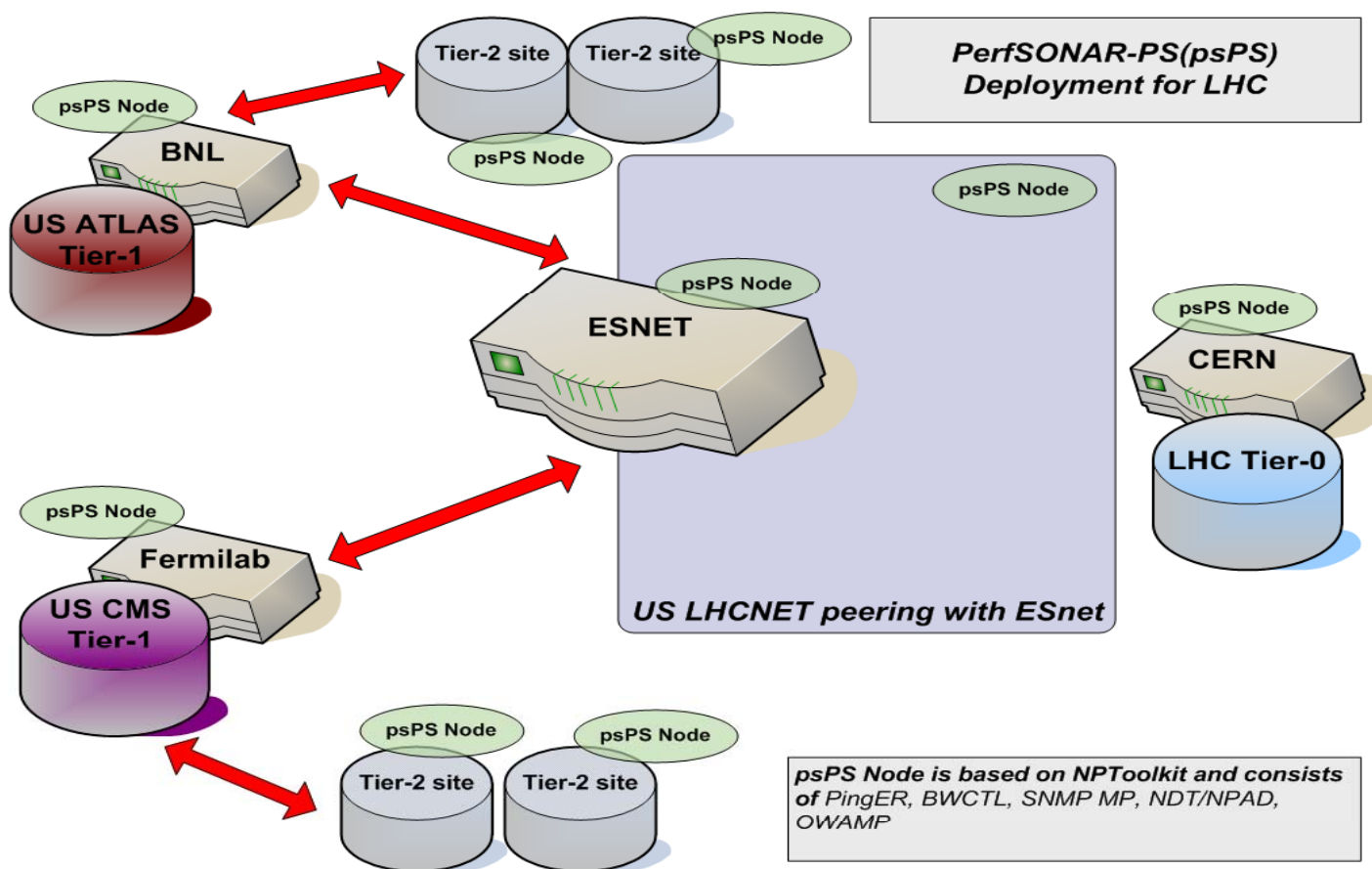
[LHC network monitoring node]

- Network Monitoring appliance
- Based on *NPToolkit*
- Modest hardware configuration ~ *600USD* a box
- Easy updates – just insert CD with updated package

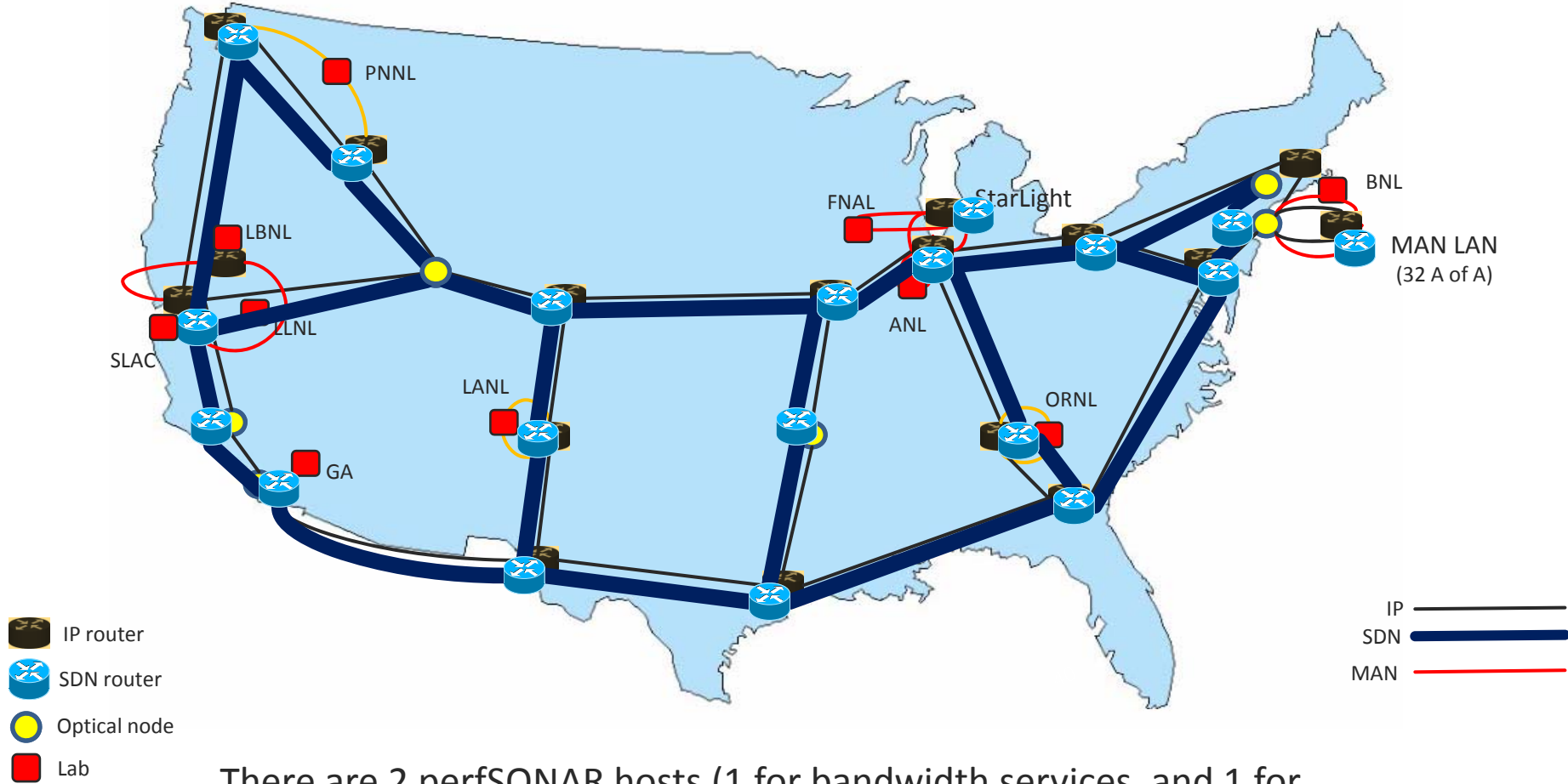
- Two boxes required - one for *latency* tests, another for *throughput* tests

- Each box is dual homed - one NIC for production network, another for high impact circuit(s)

Deployment for LHC



[ESnet PerfSONAR Locations]



There are 2 perfSONAR hosts (1 for bandwidth services, and 1 for latency services) at each SDN router location, and at most DOE labs

Requirements for setting up LHC Network Monitoring Node

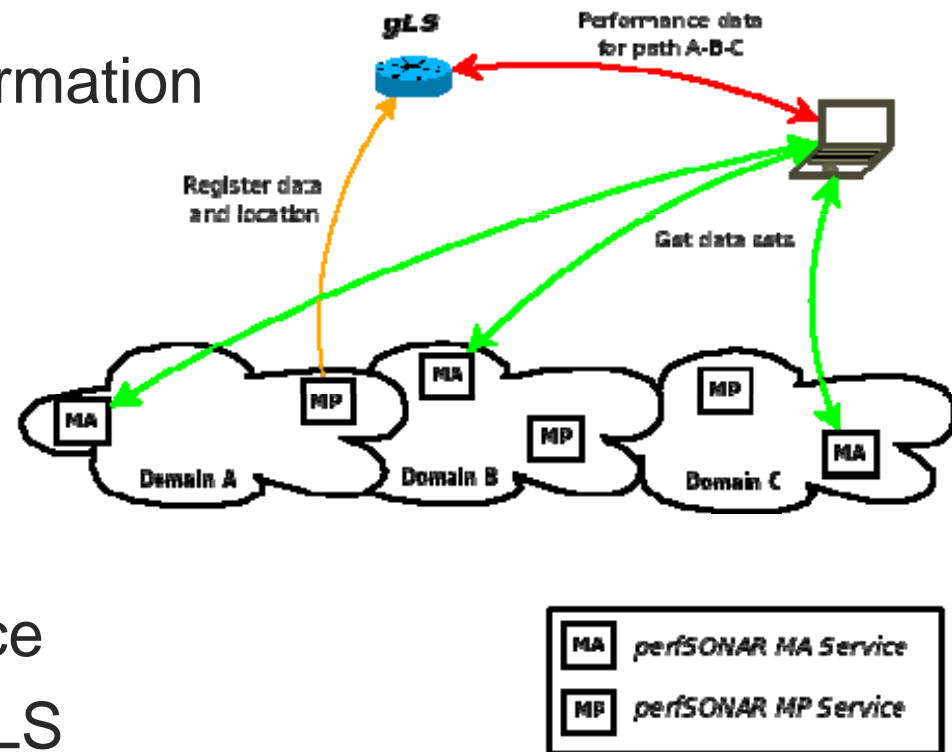
- LHC Tier-1/2/3 center
- 1 Gbps connectivity
- Thats it !

Why do you need it ?

- Network issues troubleshooting
 - Applying Network performance troubleshooting methodology
 - Isolation of the network segments
 - End-system vs networking problem
- Setting up expectations
 - Network capacity planning
 - Networking resources allocation
 - Dynamic circuits reservation

Information Service (IS)

- Global Lookup (gLS) + Topology Service (TS)
- Network Topology Information
- Services discovery
- Services registration
- End-to-end performance troubleshooting with gLS



PingER data UI

PingER GUI

URL of the remote PingER MA

Enter URL(s)

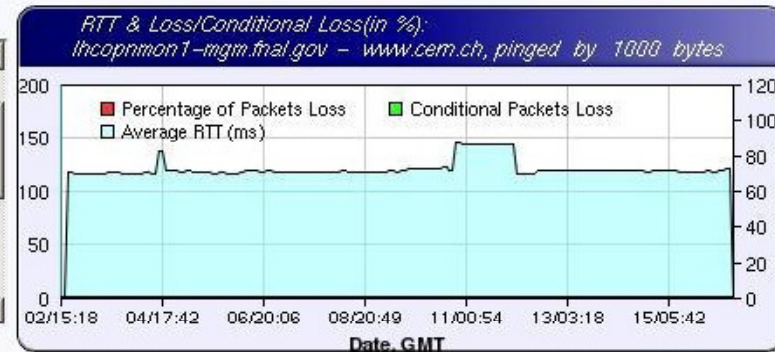
`http://lhcopnmon1-mgm.fnal.gov:8075/perfSONAR_PS/services/pinger/ma`

Use wildcard to get hostname or packetsize of interest. List of links will be updated automatically.

Source Hostname: Destination Hostname: Packetsize:

End-to-End Links

- lhcopnmon1-mgm.fnal.gov -> pinger.fnal.gov (1000)
- lhcopnmon1-mgm.fnal.gov -> www.cmsaf.mit.edu (1000)
- lhcopnmon1.fnal.gov -> phedex.rcac.purdue.edu (1000)
- lhcopnmon1.fnal.gov -> t2.unl.edu (1000)
- lhcopnmon1-mgm.fnal.gov -> www.cern.ch (1000)**
- lhcopnmon1-mgm.fnal.gov -> tier2.ihepa.ufl.edu (1000)
- lhcopnmon1.fnal.gov -> www.cern.ch (1000)
- lhcopnmon1-mgm.fnal.gov -> newmon.bnl.gov (1000)
- lhcopnmon1-mgm.fnal.gov -> cithep130.ultralight.org (1000)
- lhcopnmon1-mgm.fnal.gov -> t2.unl.edu (1000)



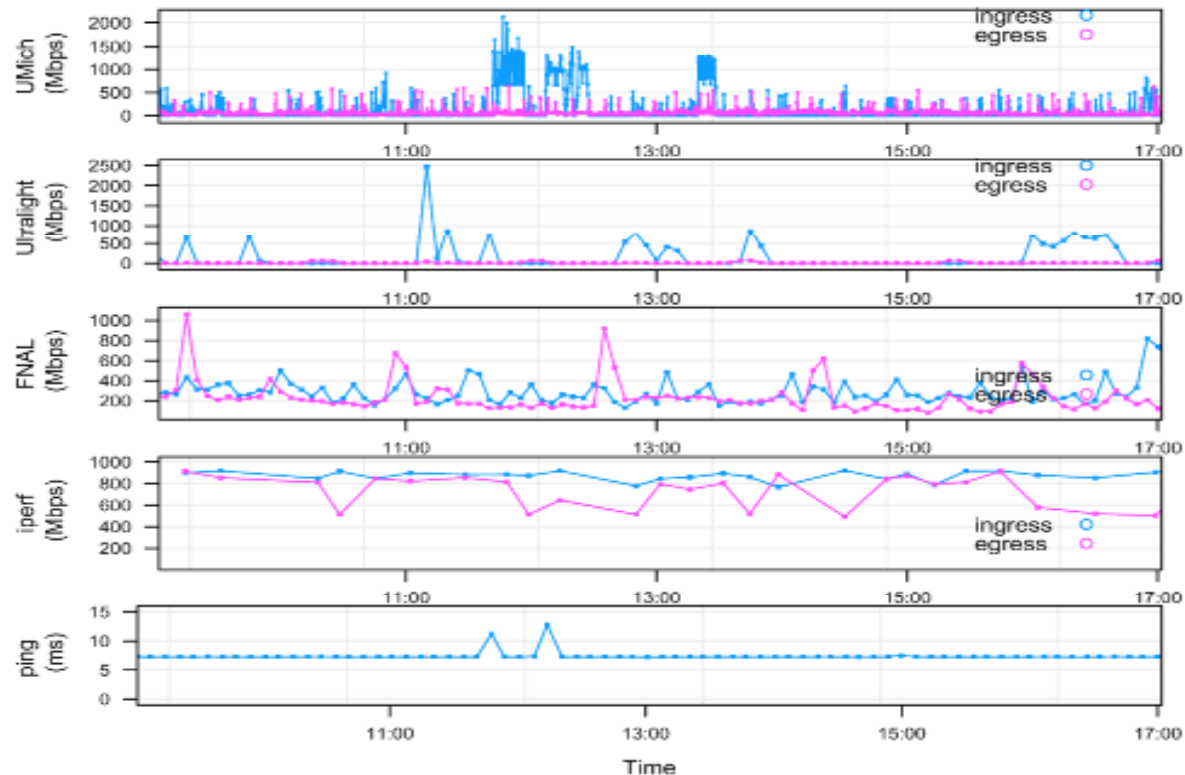
Graph parameters

2009-03-02T10:18:13 Start Time 2009-03-16T10:18:17 End Time GMT offset: GMT

Type of metric: RTT/Loss/Conditional Loss Upper RTT(or IPD) Limit: Auto Graph type: LINES

Sample Test results

- This plot shows both **ping** and **iperf** results for an **8 hour** window on the network path from **FNAL** to **UMich**.
- Note the latency spikes around **11:30** that are clearly related to the traffic spike on the **UMich** router during that same time.



[Future Deployment plans]

- Every Tier-2 in US, full interoperability with European perfSONAR MDM deployments
- All federated networks involved with LHC computing
- Orchestration level for the monitoring services, higher level data fusion and analysis
- Advance visualization layer
- Network issues tracking service

[Useful links]

- perfSONAR-PS project - <http://code.google.com/p/perfsonar-ps/>
- NPToolkit – <http://code.google.com/p/perfsonar-ps/wiki/NPToolkit>
- perfSONAR - <http://www.perfsonar.net>
- Fermilab Wide Area Networking Group - <https://plone3.fnal.gov/P0/WAN/>

[Questions]

?