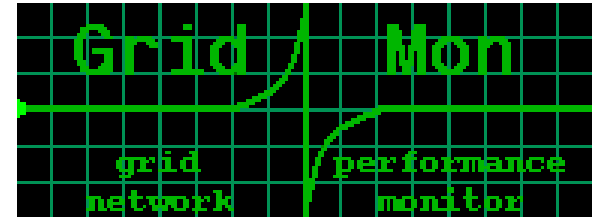




CCLR
Daresbury Laboratory



Network Performance Monitoring for the Grid

UK e-Science All Hands Meeting
Nottingham, 22nd September 2005

Mark Leese

Money by...



GridPP
UK Computing for Particle Physics

JISC

The Joint Information Systems Committee



Extra brainpower by...



eGEE
Enabling Grids for
E-science in Europe



Mark Leese
Daresbury Laboratory

- Why do this? It's the morning after the conference dinner. I'm tired. The red went down a bit too easily. Do I really need to I pay attention?
- What happened in the past?
- Where are you now?
 - UK
 - GGF work
 - EGEE
- Outreach work
- Conclusion
- Questions

There's lots happening, so this is just a taster!

For more info, read the conference paper or get in touch.

Network performance monitoring has always been important to the operation of networks of any significant size:

- fault detection
- determining expected performance

Q: So it's well understood then?

A: Yes, but the Grid is a special case:

- Grid middleware and applications could use network data to optimise their performance, adapting to changing network conditions
- Data intensive applications (LHC, VLBI, RealityGrid...) need networks debugged for efficiency
- Measurable SLAs

Why monitor network performance (2)?

Q: Okay, so why don't we just throw some more bandwidth at the problem? Upgrade the links. That seems to work with my broadband connection at home.

A: For want of a more interesting term to make sure you're awake, this is what I call the Heroin Effect ("Well I used to do a little, but a little was doing, so a little got more and more. I just keep trying to get a little better, I said a little better than before" Mr. Brownstone, Guns 'n Roses).

- You keep buying more and more (bandwidth) to take you to new highs but it's never quite as good as you thought it would be
- Simple over-provisioning is not sufficient
- Doesn't address the key **end-to-end** performance
- Network backbone in most cases is genuinely not the source of the problem
- Last mile (campus network → end-user system → your app) often cause of the problem: firewall, wiring, hard disc, **application** and many more potential culprits

Why monitor network performance (3)?

Q: Okay, so why don't we use dedicated optical fibre everywhere?

A: Costs are still prohibitive...perhaps if you've been a lot of selling heroin :)

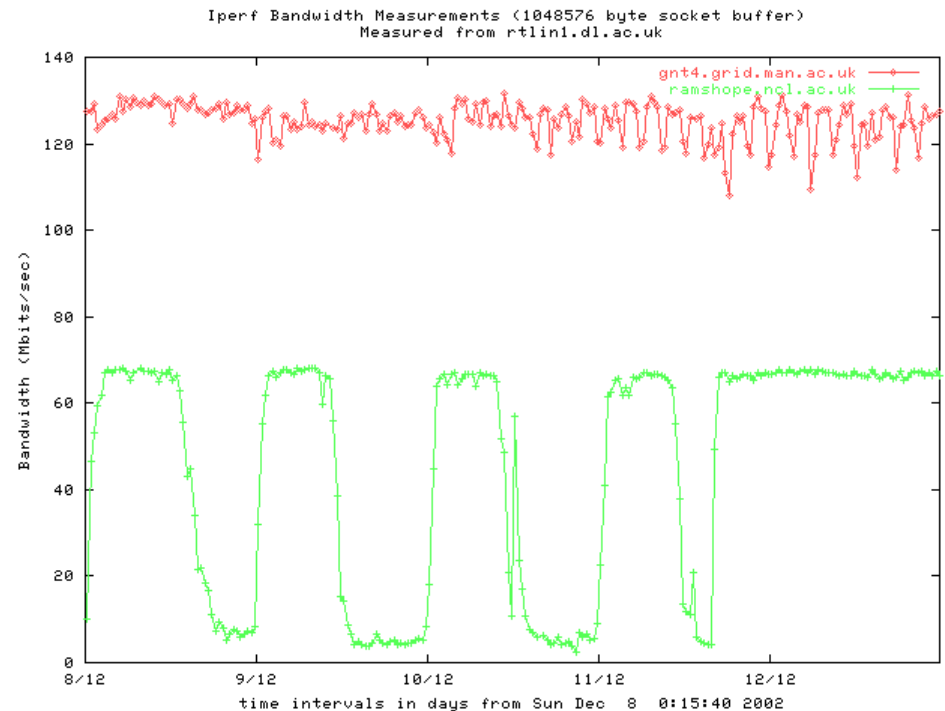
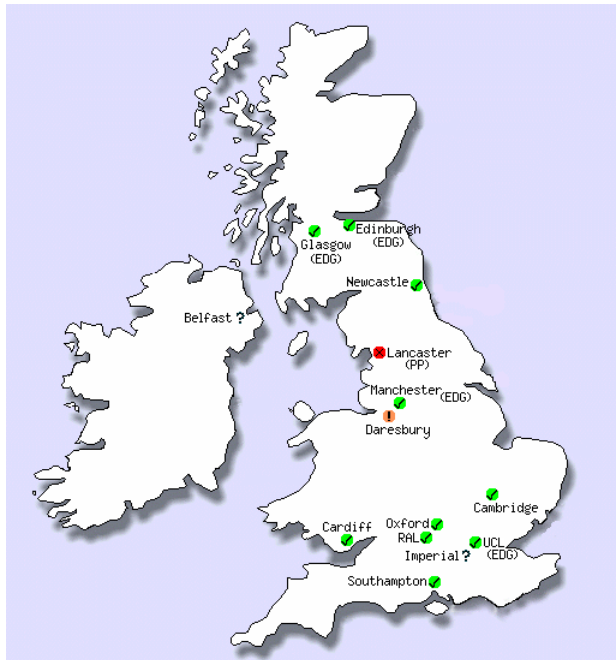
Q: Okay then smarty pants, what if we share existing fibre, and used circuit-switched lightpaths? That's dedicated bandwidth, but without the cost of dedicated fibre.

A: Good idea in theory, and we can see the benefits in UKLight and ESLEA*, but it still doesn't address the end-to-end issue. Take a **real-life** ESLEA example (thanks to ESLEA for the figures)...

- The UK wants to transfer data from FermiLab (Chicago) to UCL for analysis by physicists, before sending results back.
- Datasets = **currently** 1-50TB
- 50TB would take > 6 mths on production net, or one week at 700Mbps
- So a 1Gbps circuit-switched light path is provisioned
- Result = disc-to-disc transfers @ 250Mbps (1/4 of theoretical max)
- Tests show it's a problem at an end site

Previous Work

- “...design and deploy an infrastructure for network performance monitoring within the UK e-Science community” – June 2002
- MPs (Monitoring Points) at each e-Science Centre
- Full mesh of tests
- Human access (www interface) to monitor performance, find faults
- Work with GGF NM-WG to develop unified interface for requesting and publishing performance data (Web Services) to Grid m/ware and apps

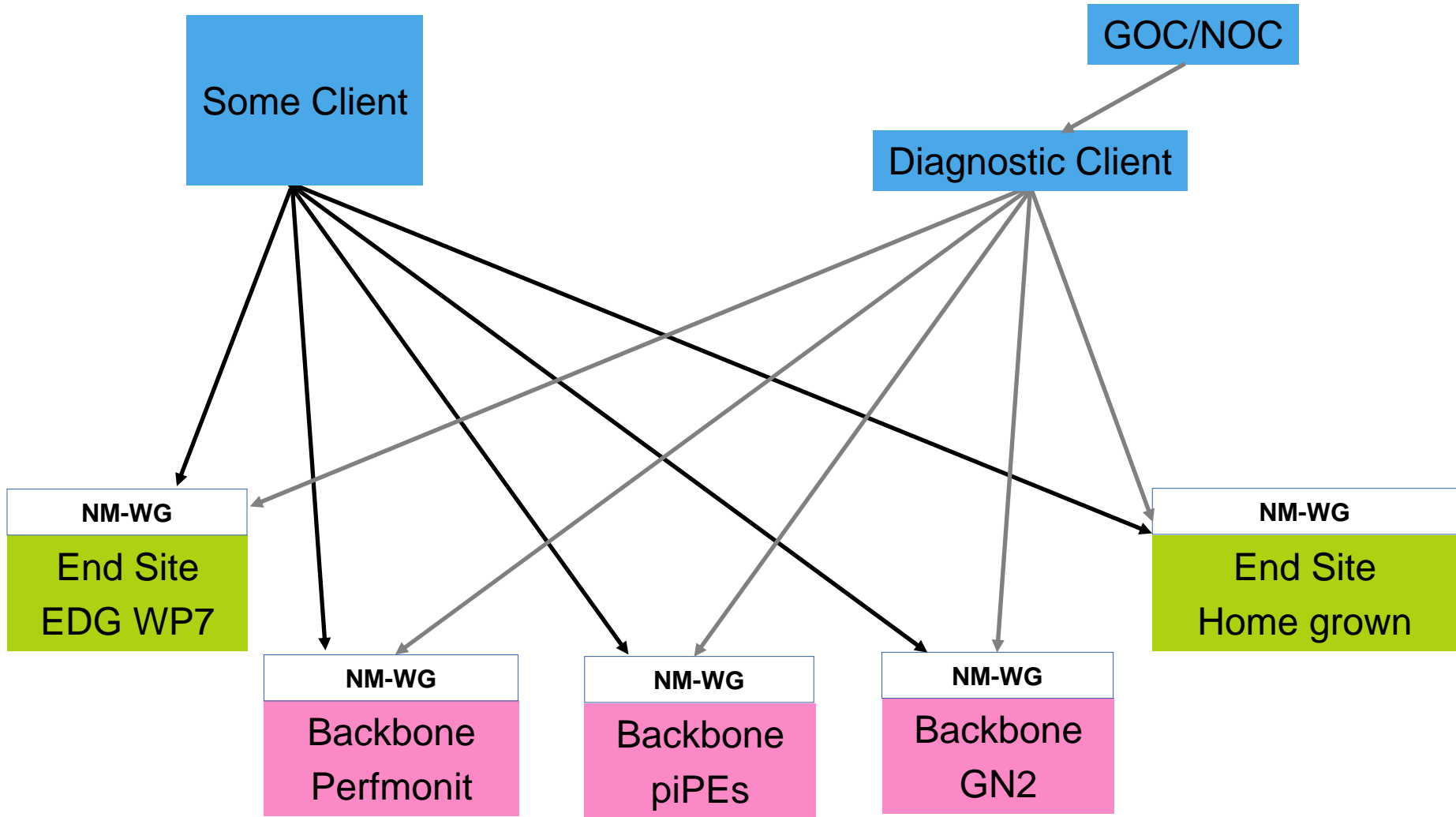


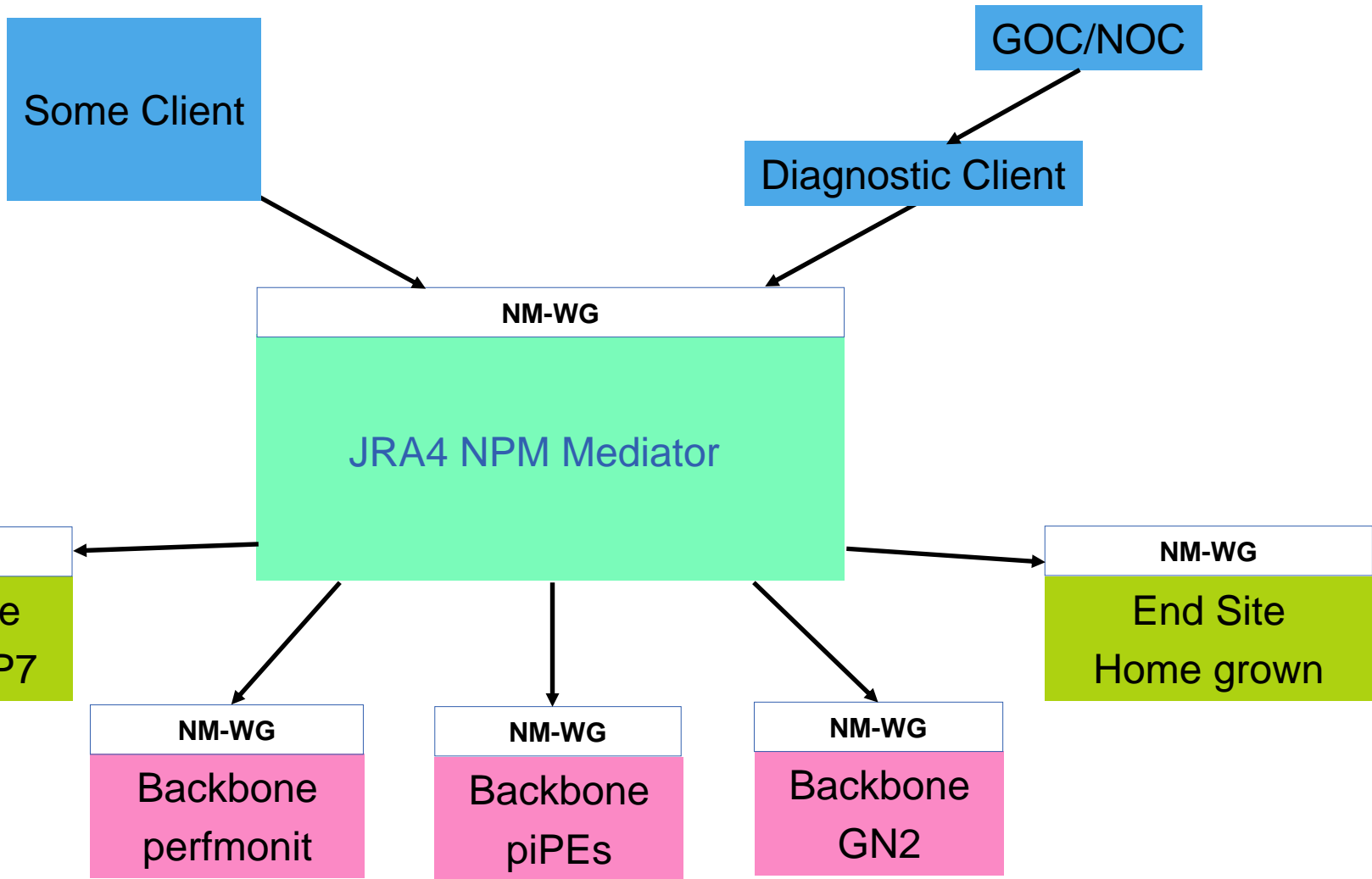
- Well received and grew interest (e.g. UK HEP/PP community), but...
- Version 1 infrastructure proved to be unsustainable
 - most institutions were helpful, but...
 - varying spec of machines, flavours of Linux, security rules etc.
- V1 MP:
 - Ran tests
 - Stored data locally
 - Served data to human users using web server running on the MP
 - Would have provided WS i/f using Tomcat running local
 - **Grew interest and a useful learning exercise**
- V2 MP will:
 - Run tests
 - Write data back to central DB at DL and one other
 - Revised web i/f and WS i/f will be provided by machines co-located with DBs
 - MP is thus much simpler, and brains of the operation are centralised at two, more accessible, sites

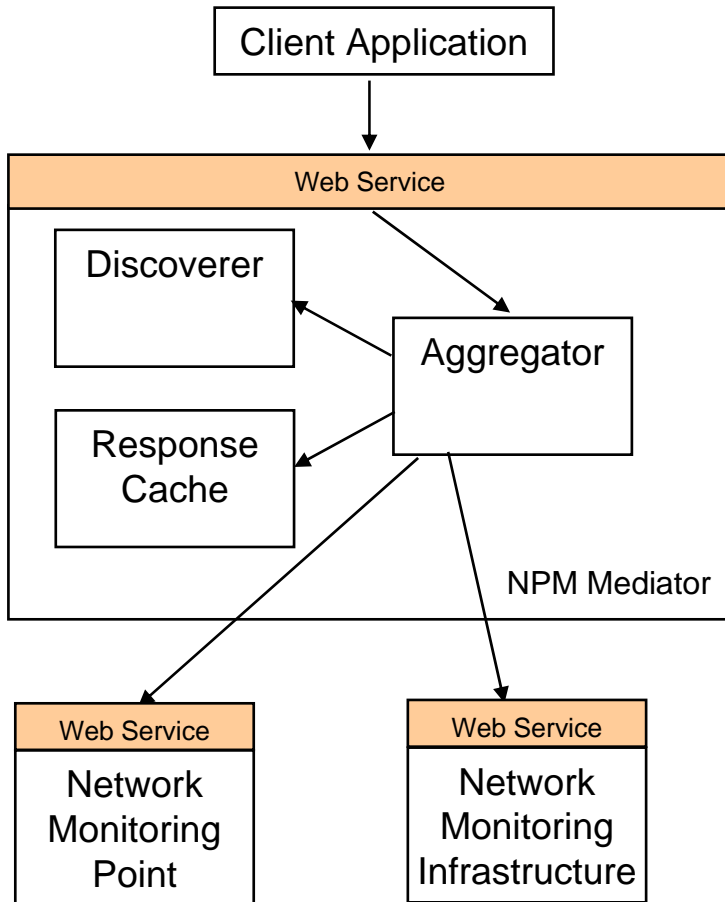
- Days, weeks and months have gone into this work. Fortunately it can be summarised by two slides.
- Unified schemas to request and subsequently publish network performance data (e.g. results of bandwidth tests)
- All you really need to concentrate on is that they provides a unified interface for network operators, and Grid m/ware and apps to share data
- Last year we reported on V1 schemas
 - Monolithic, “one size fits all” solution
 - Just two schemas: less to maintain, less to worry about parsing etc.
 - Test deployments identified some gaps in group thinking
 - So here comes V2, but as with V1 Gridmon, a useful learning exercise, and **crucially** proved usefulness of idea to people, e.g. EGEE JRA4 Mediator demo at GGF13 (March 2005) - obtaining end-to-end and backbone (multiple administrative domain) data using the same method

- NM-WG now define certain base elements from which schemas for specific tools and/or specific characteristics are developed.
- As a result, schemas only contain elements relevant to that particular tool or characteristic.
- Separation of data and metadata
 - Data = measurement value + associated time
 - Metadata = the who, what and how of the measurement
 - Undoubtedly neater
 - Less data sent on wire: metadata only needs to be sent once, can then be referred to by a metadata ID
 - Beneficial implications for “chaining” of operations
- V1 = direct request-response model. V2 lends itself to supporting more e.g. subscribing to event notifications and reporting those events
- Need gatekeeper for “approving” new schemas

- **There are various monitoring tools and frameworks available:**
 - End-to-end: e.g. EDG::WP7, Gridmon
 - Backbone: e.g. Internet2 and DANTE's perfSONAR collaboration
- **We (JRA4 - Development of Network Services) in our NPM (Network Performance Monitoring) work are not building another one! The work is about standardising access to NPM data across multiple domains and using it.**
 - NM-WG schemas are the selected basis for standardisation
- **So what is being produced:**
 - Mediator (standardise access to NPM data)
 - Diagnostic tool (use the data)
 - Publisher (provide NPM data to Grid middleware)
- **1st Mediator prototype (deliverable DJRA4.2) was produced in PM9 (Dec '04):**
 - Proves we can harness (multi-domain) backbone and end-site tools together
 - Low level framework only
 - Improved in subsequent iterations
 - For more info (and prototype design doc): <https://edms.cern.ch/document/533215/>







“Mediator” for GOCs, NOCs, end users

- Human & machine users interact via client application, “speaking” NM-WG
- Discoverer locates MP(s) or infrastructures that can answer the client’s query
- Aggregator
 - obtains query results from MP(s)
 - aggregates results (if necessary)
- To improve performance and reduce loading, results of recent requests will be cached

- **Web interface access to any network data accessible via the Mediator, i.e. any data that the Mediator can access via the unified NM-WG data – potentially a lot!**
- **Aimed at NOCs and GOCs**
- **JSPs linked to the Mediator**
- **Must have a valid certificate to access it**
- **DT can access lots of data (EGEE, DANTE etc.) but must do so through a Web Services interface – not very efficient for graphing.**
- **The Gridmon web interface will be Gridmon only, but can access data more natively using a simple TCP connection, or a DB interface such as PerlDBI**
- **This is a prototype – usual disclaimers apply!**
- **<https://egee.epcc.ed.ac.uk:28443/npm-dtD/>**



- 1.5 day workshop showing end-users how to obtain better network performance for their applications
 - NFNN1 = July '04, UCL
 - NFNN2 = June '05, NeSC
- NFNN2 was “new and improved”:
 - 97% of feedback respondents rated the workshop as good or better
- Workshop materials available online
 - videos also, if sufficient demand

<http://gridmon.dl.ac.uk/nfnn/>

Register interest to: nfnn@dl.ac.uk



Networking Primer

- Written by NFNN attendees, and available from the NFNN website: <http://gridmon.dl.ac.uk/nfnn/resources.html>
- This is an AVO (Astrophysical Virtual Observatory) project document, and a new draft is being produced with input from DL, but this is more than good enough to start with
- Many thanks to Clive Davenhall (NeSC), Eckhard Sutorius and Bob Mann (ROE)

Web Services

- WS is not the core business of NM-WG, may be others in similar position
- Organised a half day Web Services pseudo-tutorial for GGF15 (Boston, 3-6th October 2005)
- Aimed at anyone seeking to reinforce or expand their Web Services knowledge, or make use of the NM-WG schemata
- http://www.ggf.org/GGF15/ggf_events_schedule_NMApps.htm
- Many thanks to Marlon Pierce and Shrideep Pallickara (Indiana) for agreeing to speak

The Wider Picture

- Important as it is, the network is part of a wider Grid performance effort
- International Grid Performance Workshop, NeSC, June '05:
<http://www-unix.mcs.anl.gov/~schopf/GPW2005/>
- Just one finding published in the report:
 - Mismatch between what application scientists need and what performance tool developers provide. Apps people generally want simple but reliable tools that solve basic problems, while tool builder's kudos and funding demands demonstration of novel and usually complex techniques.
 - e.g. b/w available = yes/no vrs detailed plots
- Follow up BoF at SC|05:
http://sc05.supercomputing.org/schedule/event_detail.php?evid=5267

Forget the talk. Just take this away with you...

- Network performance monitoring is crucial to the Grid
 - Adaptive behaviour, predicting performance, making the network efficient, fault detection, monitoring SLAs
- We're interested in the **end-to-end** network performance
- Network backbones generally perform well, indeed the network is **frequently** not the cause of performance problems:
 - Your hard disc, your application (GridFTP vrs HTTP)...
 - If it is the network, look in the last mile (campus) e.g. the firewall
- GridMon is the infrastructure for UK academic network performance monitoring
- There's lots happening. The UK has its fingers in many worthwhile pies, but consensus and development take time
- For more info
 - Networks for non-Networkers
 - Networking primer
 - Read the conference paper or get in touch