



# SAM/SAM-Grid

Lee Lueking

Oklahoma Workshop

July 10, 2002



# Overview



- SAM News
- Progress and plans
- SAM Station and remote node network stats
- Future SAM-Grid architecture and plans

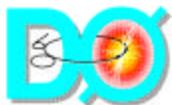




## SAM News



- We have lifted the station moratorium imposed in February. New stations requested for:
  - ◆ Cal State Fresno:csuf\_hep2-station
  - ◆ Tata Institute:Tata-mcfarm
  - ◆ Please send your requests to sam-admin
- CDF has committed to using SAM. They are suggesting features that may be useful to us too. Also contributing some manpower.
- Chris Jozwiak is leaving, so some things will be slower (like getting new stations going) for a while. Please be patient.



[d0db.fnal.gov/sam](http://d0db.fnal.gov/sam)



# DØ data handling News



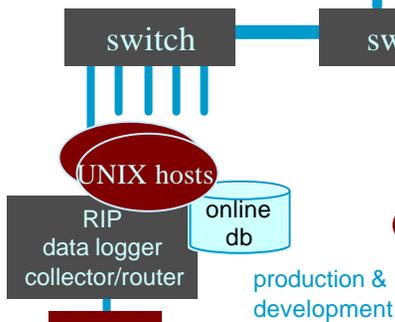
## LTO Test

- Only Raw to STK since mid-May
- MC, Secondary, User data to LTO
- Good results so far

STK  
powdermill ADIC AMEZ



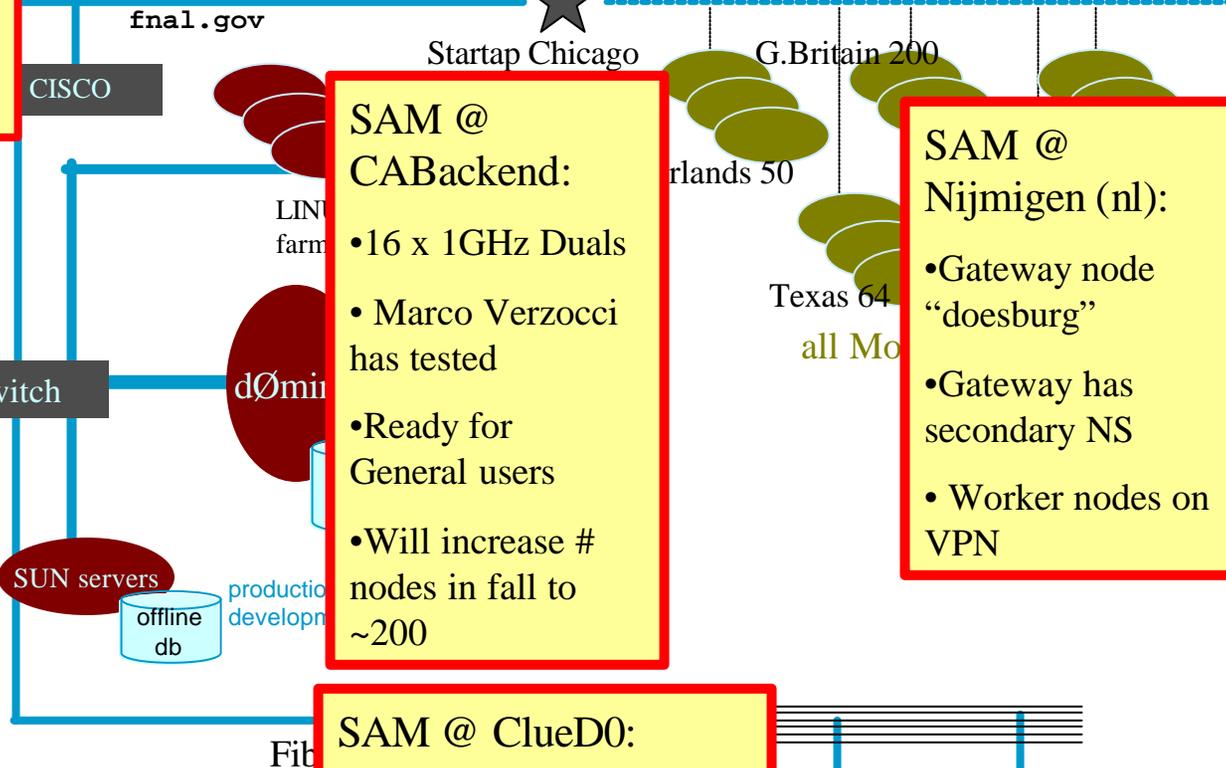
ENSTORE movers



L3 nodes



[d0db.fnal.gov/sam](http://d0db.fnal.gov/sam)



## SAM @ CABackend:

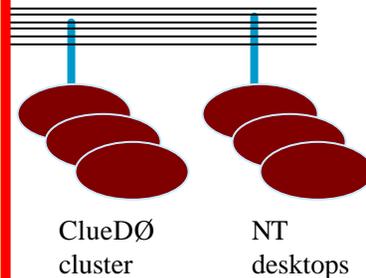
- 16 x 1GHz Duals
- Marco Verzocci has tested
- Ready for General users
- Will increase # nodes in fall to ~200

## SAM @ Nijmegen (nl):

- Gateway node "doesburg"
- Gateway has secondary NS
- Worker nodes on VPN

## SAM @ ClueD0:

- Use FCP
- Dedicated FS node lafite-clued0
- Test Harness Mods
- Closing in on success





June 2002  
CERN Courier  
•Data was transferred between IC, Lancaser and FNAL  
•Used DoE Science Grid Certificate



d0db.fnal.gov/sam

NEWS

NETWORKS

## D0 physicists 'shake hands' across the Atlantic in key grid test

In an important test of datagrid technology, members of the D0 collaboration at Fermilab have successfully communicated across the Atlantic with colleagues in the UK. The aim of the grid is not only to make it possible to access data remotely on different machines, but also to enable data processing to take place on remote machines.

A vital first step in achieving this goal is to allow individuals wishing to access the grid to identify themselves and show that they are authorized users. A two-way trust must be established between the individual and the machine that is being used.

In the tests, carried out in February, Fermilab exchanged files with Lancaster University, UK and Imperial College, London, after the transfers had been authenticated using certificates issued by the Department of Energy ScienceGrid and the UK High Energy Physics Certificate Authority.

The "firewalls" installed in many computer systems are making it increasingly difficult to access computers remotely, which is the antithesis of the philosophy behind the grid. The authentication system is intended to provide a means of allowing secure access so that the grid can operate effectively. In February's tests the certificates were used to establish trust between users and machines at Fermilab, Imperial College and Lancaster.

Although in this case the users were members of the same collaboration, the transfers took place as though the users were completely unknown to one another. This approach was used to test the Globus Toolkit - the software tool that was used to build the authentication system.

The software is currently being developed by the US-based Globus Project to bring about the higher level of computer access that will be essential if the grid is to fulfil its promise in a wider context.



D0's international mixture of experimenters may sometimes get the chance to meet at Fermilab, but their data analysis relies on the power of processors at collaborating institutes across the world, which in future will be connected by the datagrid. (Fermilab Visual Media Services.)

8

CERN Courier June 2002





“With great power comes great responsibility”

We need to be sure the product, and project, are on track to meet the needs of all users for Run II.



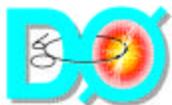


# What's New in SAM?

(Latest version v4\_2\_0\_3p1)



- Many improvements in distributed cluster environment (CAB, Farms, ClueD0)
- Much easier installation and configuration
- New improved installation instructions
- Ability to use distributed Naming Service, and run station with workers on VPN (Frank's presentation)
- Ability to set "route" and "constrained-delivery" (at the same time!)
- Removed some cache accounting problems and other inconsistencies
- Improved name service and sam-at-a-glance monitoring





# What is in the Works

(just some highlights for next 4 months)



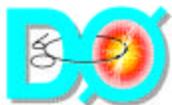
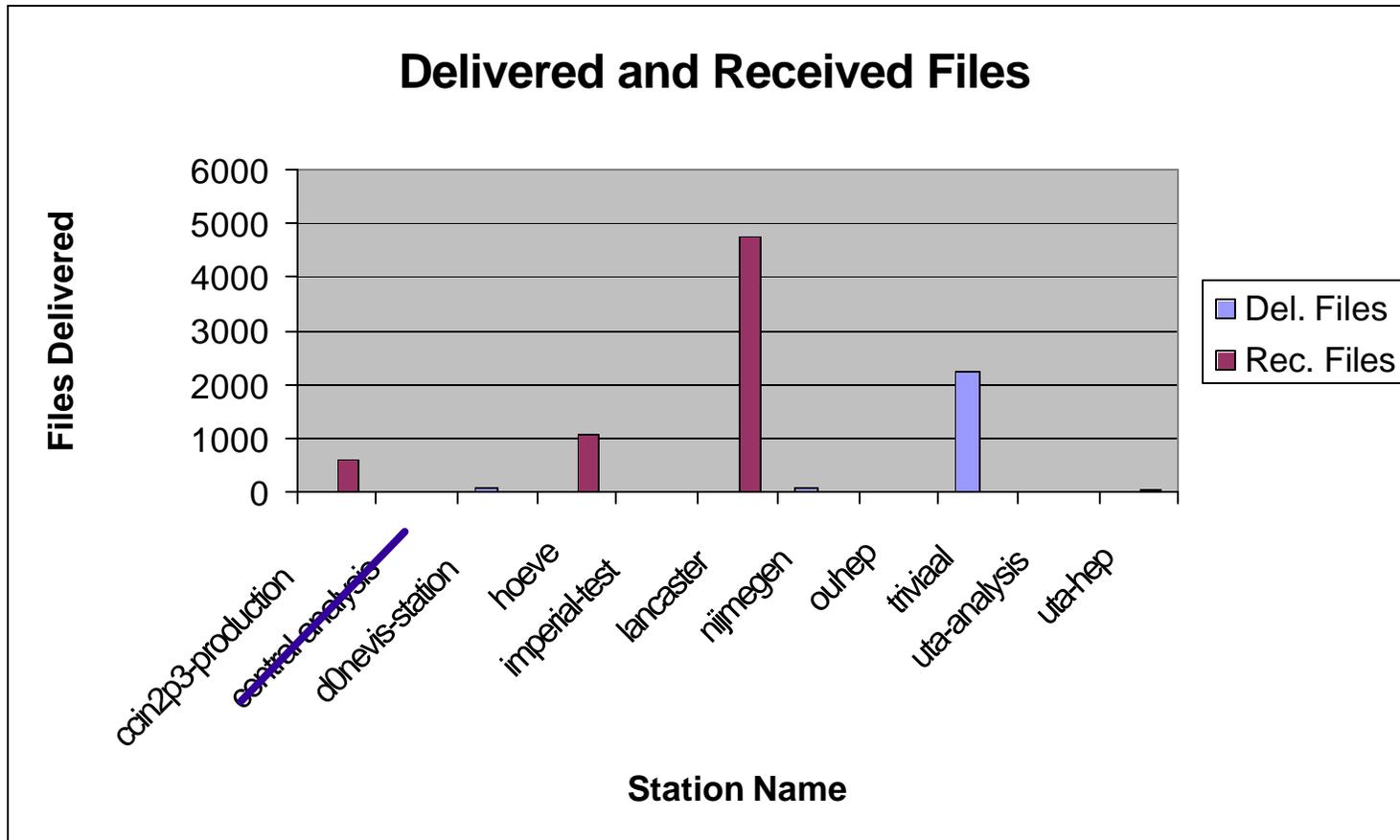
- Clean up additional caching issues
- Add CRC checking on file transfers
- Move to GridFTP (from bbftp) for extra-domain file transfer
- More monitoring, including file transfer and station tracking
- Ability to share NFS mounted disks in cluster.
- Looking into use of dCache, and ability to use network files served by TdCache, using Tnet in ROOT to read them.





# Remote SAM Station Stats

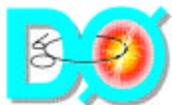
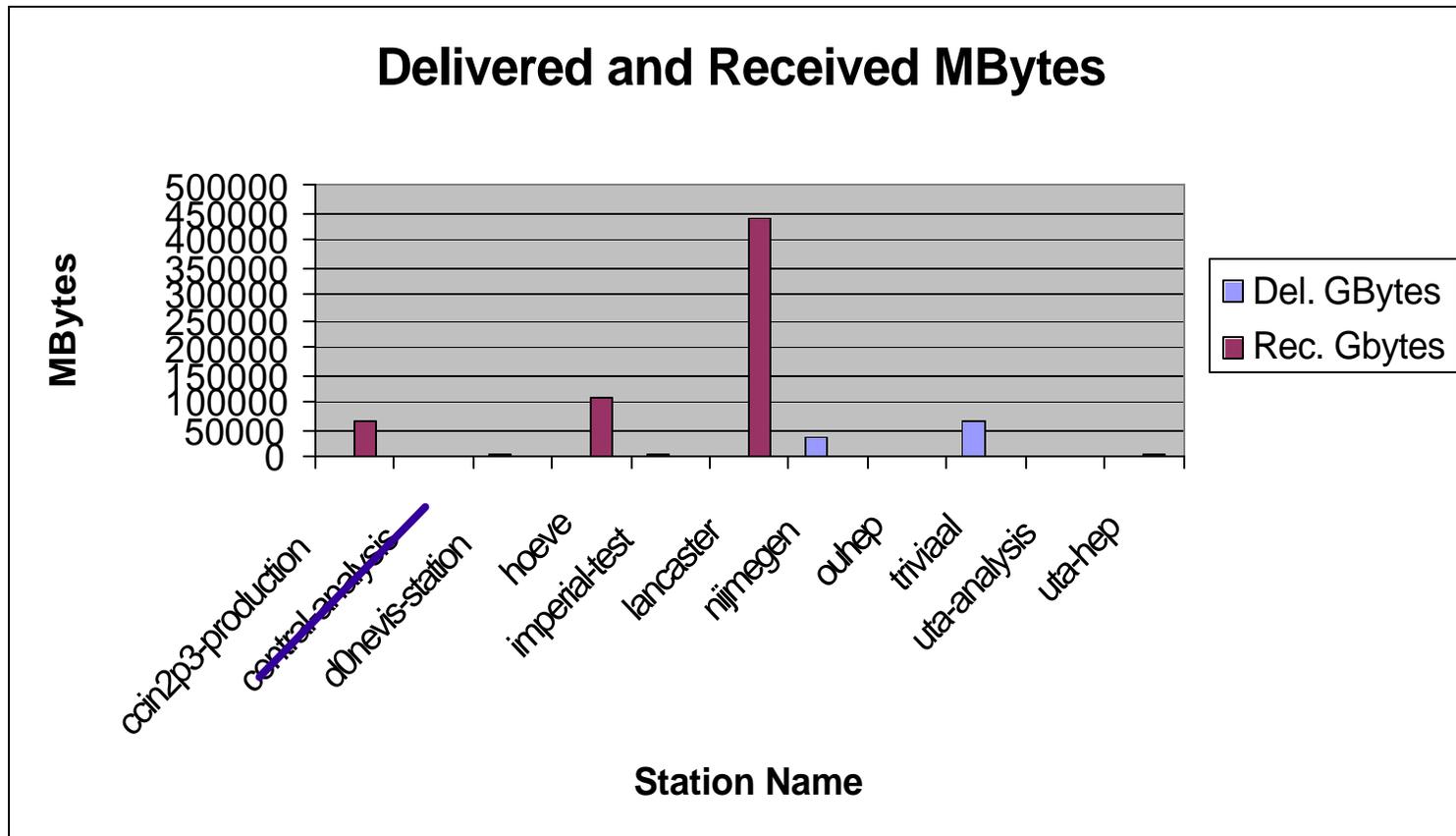
(For the Last 21 Days)





# Remote SAM Station Stats

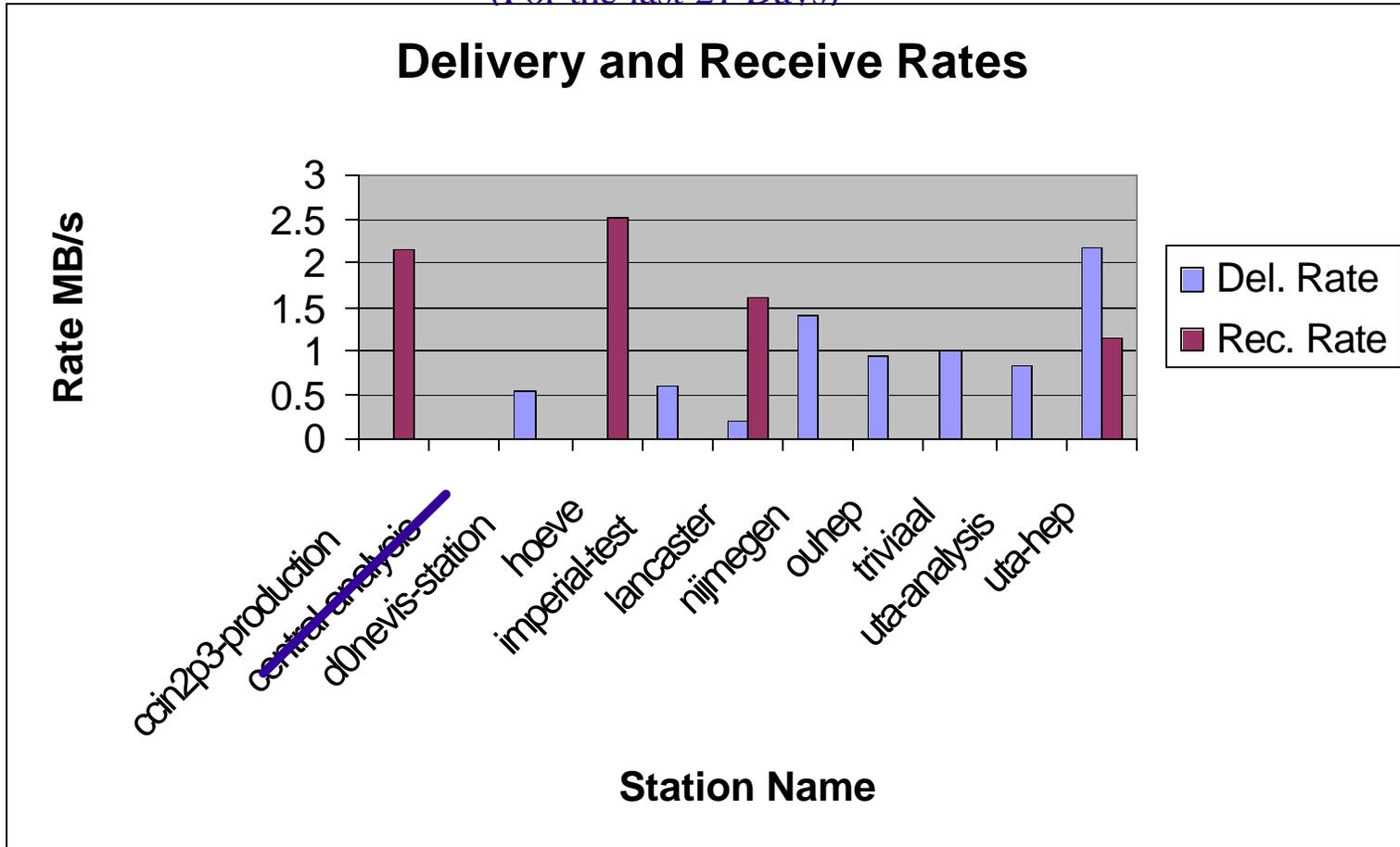
(For the last 21 Days)





# Remote SAM Station Stats

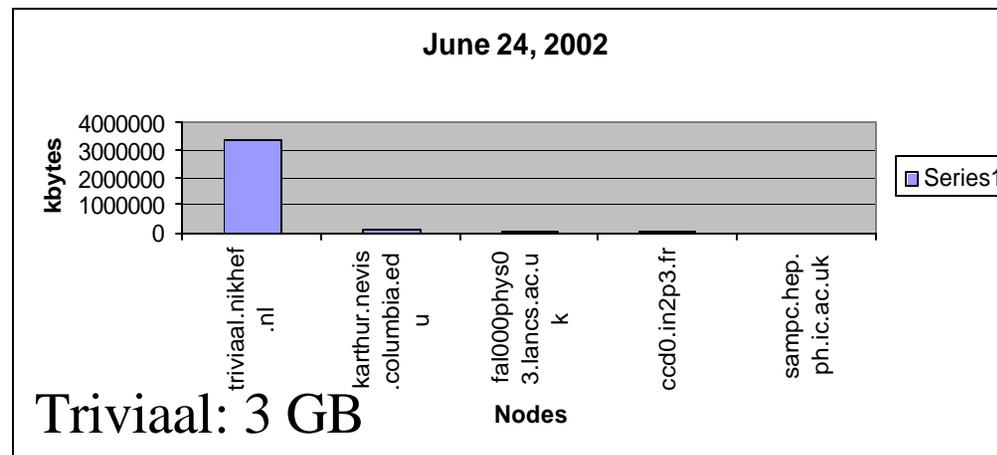
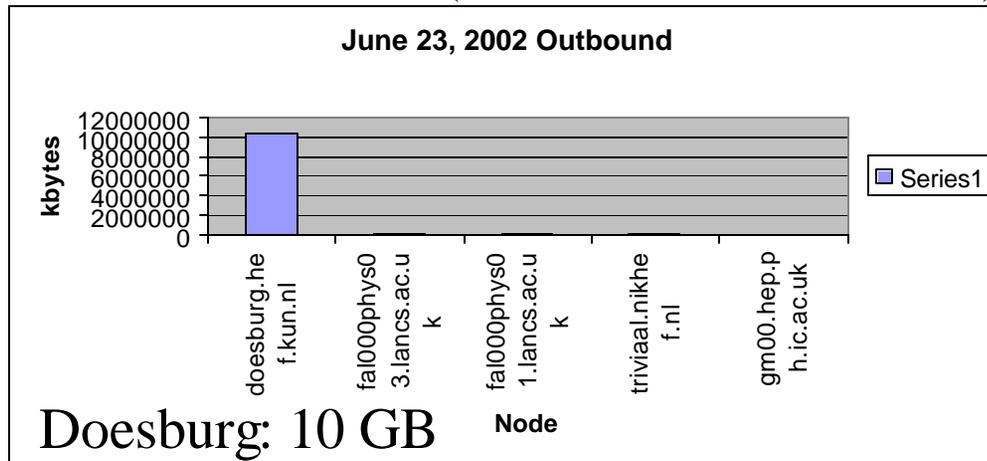
(For the last 21 Days)





# Remote Node Network Stats

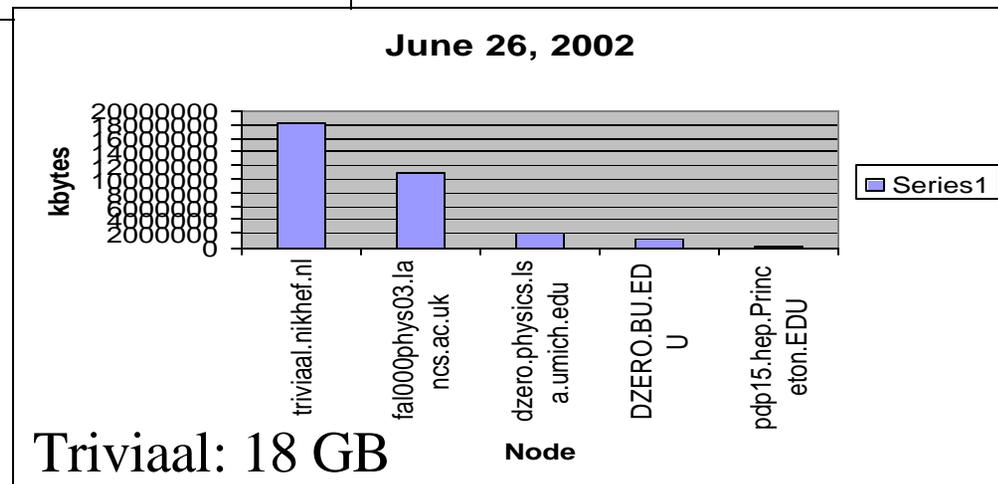
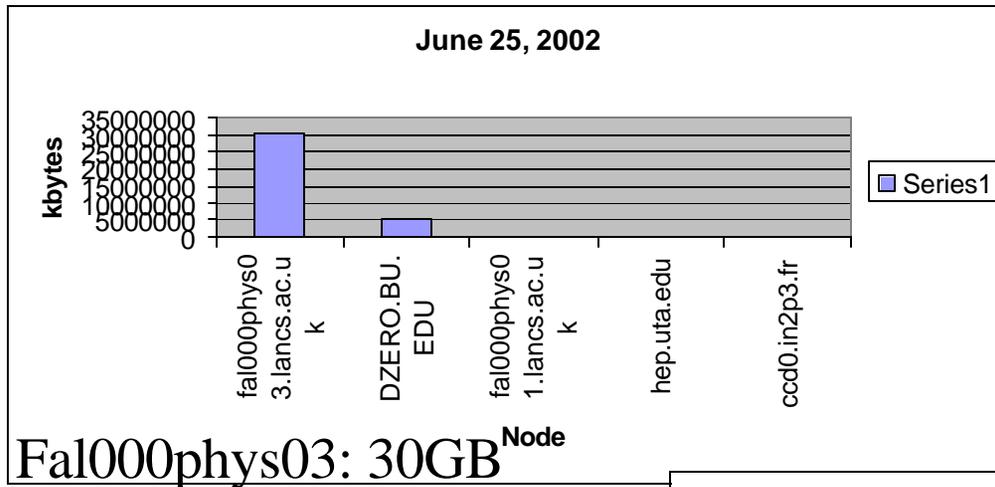
(Outbound traffic from d0mino)





# Remote Node Network Stats

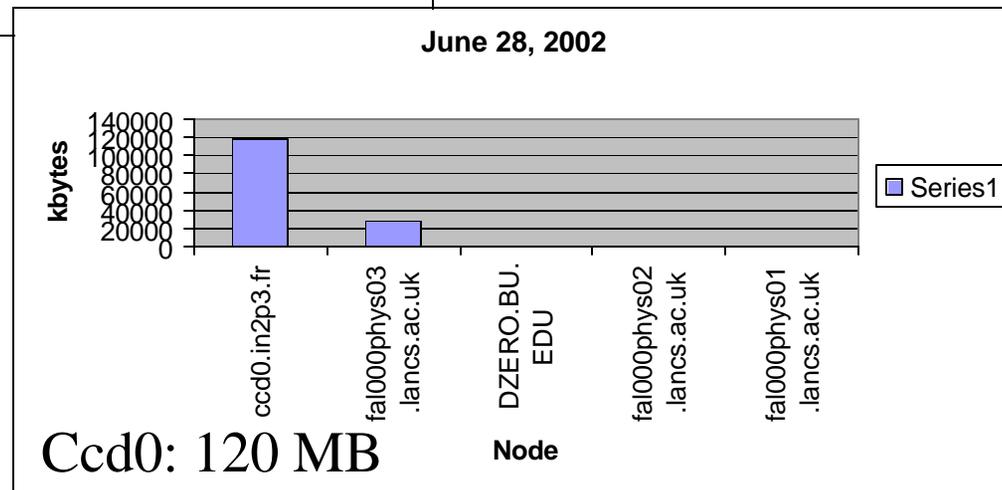
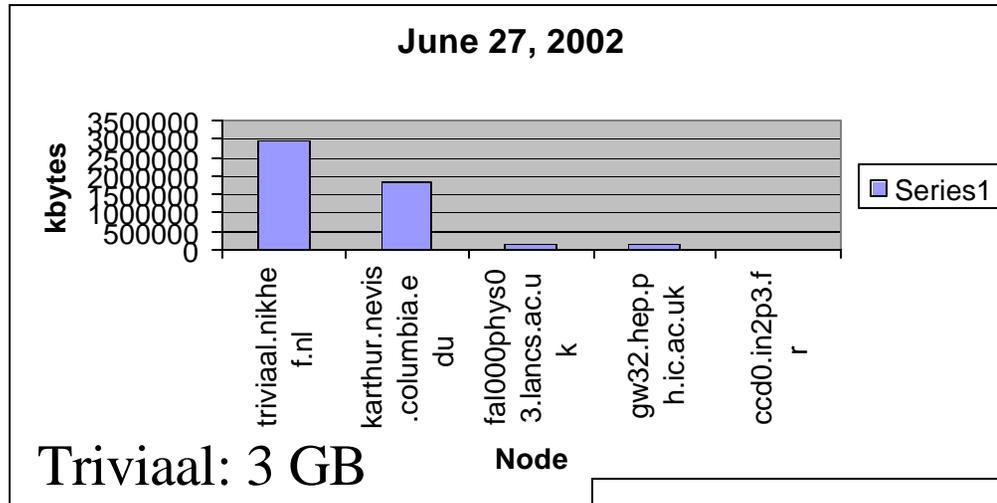
(Outbound traffic from d0mino)





# Remote Node Network Stats

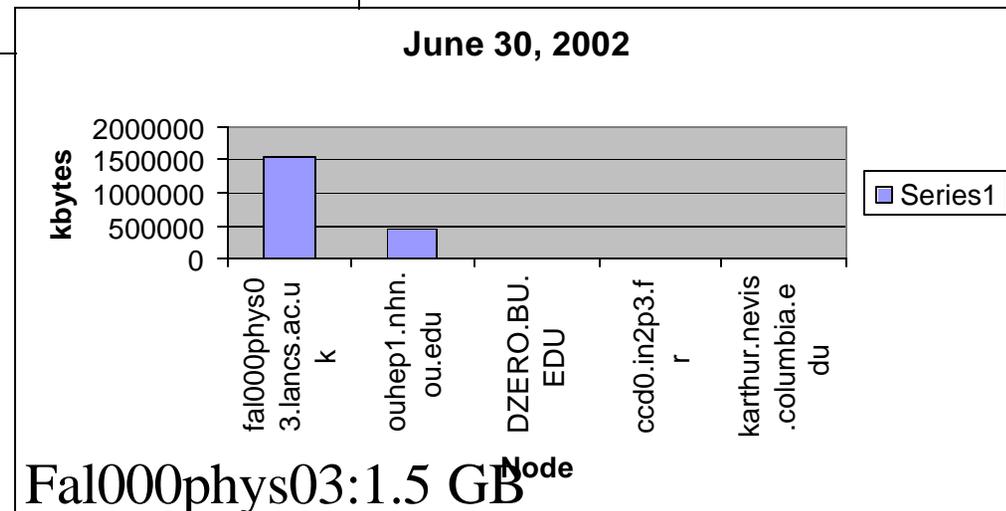
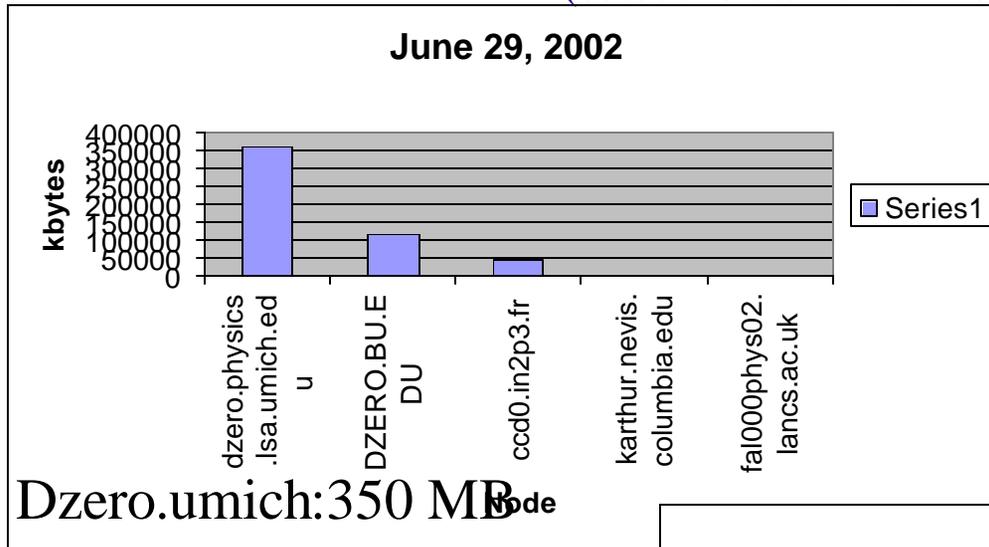
(outbound traffic from d0mino)





# Remote Node Network Stats

(outbound traffic from d0mino)

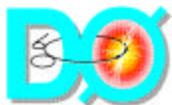
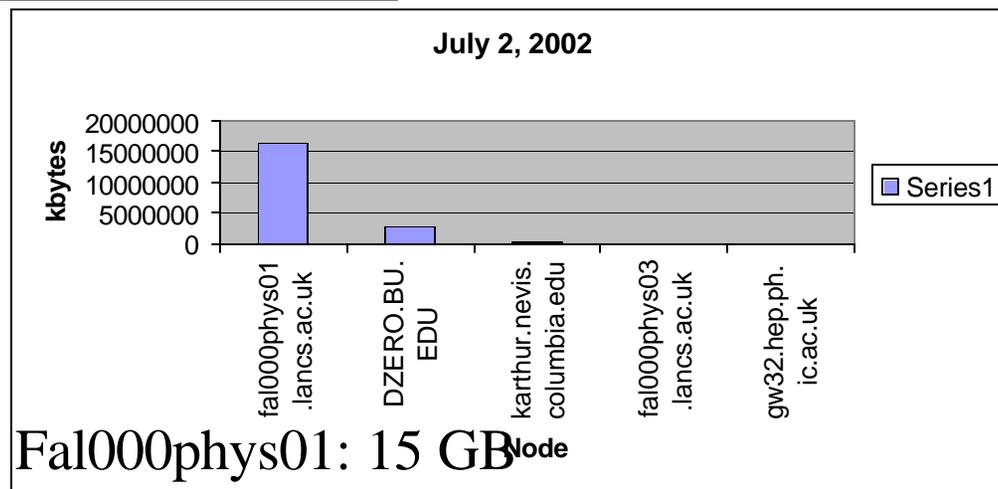
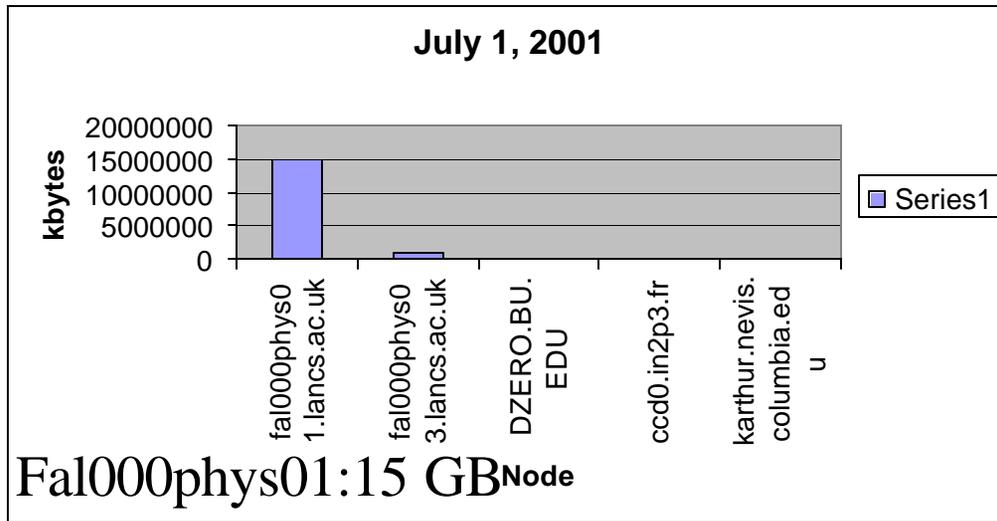


d0db.fnal.gov/sam



# Remote Node Network Stats

(outbound traffic from d0mino)

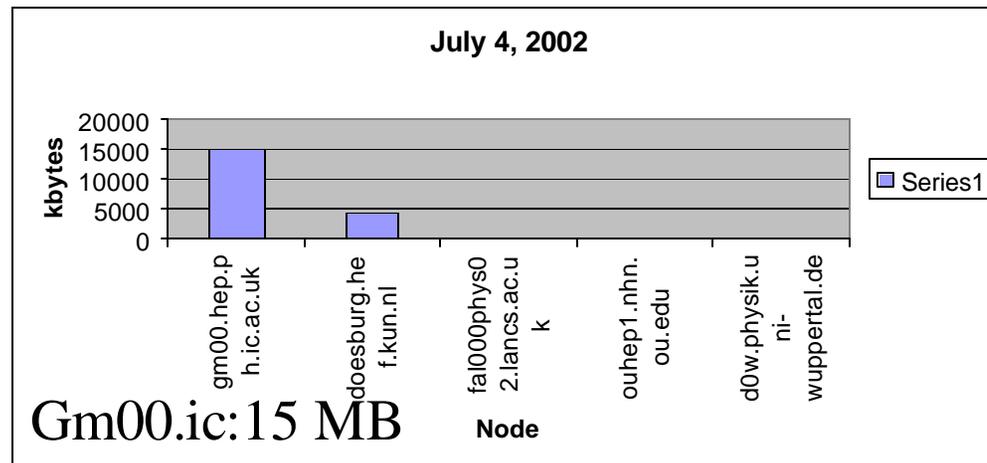
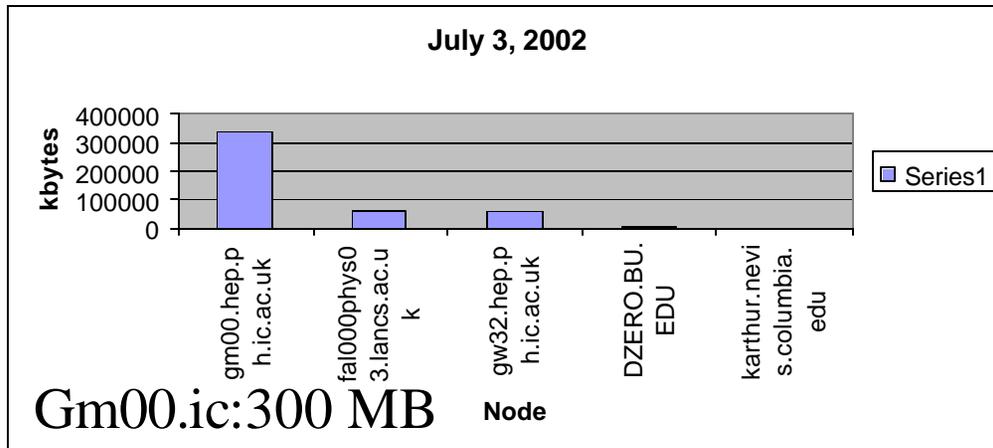


d0db.fnal.gov/sam



# Remote Node Network Stats

(outbound traffic from d0mino)





## SAM and the Grid

# SAM-Grid



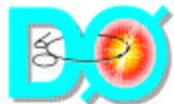
[d0db.fnal.gov/sam](https://d0db.fnal.gov/sam)



## What is SAM-Grid?



- Project to include Job and Information Management with the SAM Data Management System
- Project started in 2001 as part of the PPDG collaboration to handle D0's expanded needs.
- Recently included CDF
- Current SAM-Grid team includes:
  - ◆ Andrew Baranovski, Gabriele Garzoglio, Lee Lueking, Dane Skow, Igor Terekhov, Rod Walker (Imperial College), Jae Yu (UTA), Drew Meyer (UTA), Tomasz Wlodek
  - ◆ Collaboration with U. Wisconsin Condor team.



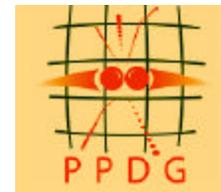
■ <http://www-d0.fnal.gov/computing/grid>



## The Goal



- Enable fully distributed computing for the DZero and CDF, by enhancing SAM and incorporating standard Grid tools and protocols. Developing new solutions for Grid computing in a secure and accountable environment.
- The SAM grid-ification is funded by PPDG and GridPP. The collaborators we are working with include the Condor Team (via PPDG) and Imperial College (via GridPP)
- We are communicating with other groups working on Grid technologies as well (EDG among them).
- Regular CDF/DZero joint grid meetings
- We promote interoperability and code reuse





# Major Components



- **Job Definition and Management:** The preliminary job management architecture is aggressively based on the Condor technology provided by through our collaboration with University of Wisconsin CS Group.
- **Monitoring and Information Services:** We assign a critical role to this part of the system and widen the boundaries of this component to include all services that provide, or receive, information relevant for job and data management.
- **Data Handling:** The existing SAM Data Handling system, when properly abstracted, plays a principal role in the overall architecture and has direct effects on the Job Management services.

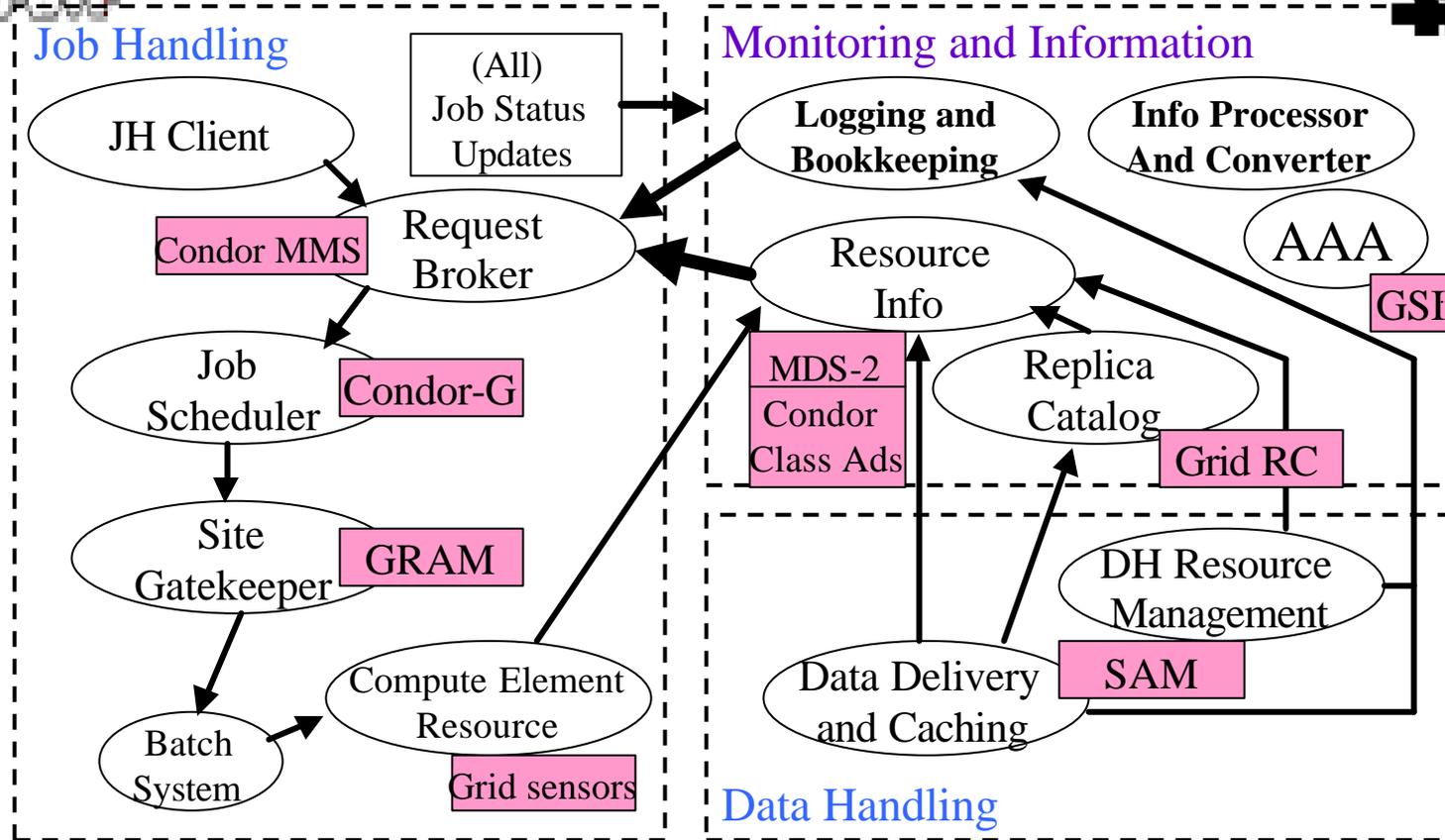


[d0db.fnal.gov/sam](http://d0db.fnal.gov/sam)





# SAM-Grid Architecture



Principal Component

Service

Implementation Or Library

Information



[d0db.fnal.gov/sam](http://d0db.fnal.gov/sam)



**Condor**  
High Throughput Computing



the globus project™



# Kerberized Certificate Authority

Dane Skow



- The KCA (Kerberized Certificate Authority) serves two immediate purposes for us and allows for use of PKI authenticated transactions that comply with Fermilab Policy.
  - ◆ First, use of the KCA means that all users with FNAL accounts (principals) the ability to get GRID proxies and need not register with any other CA or learn the details of the care of longterm identity credentials.
  - ◆ Second, use of a FNAL KCA proxy for authentication is sufficient proof of authorization to use FNAL resources that full authority of the Kerberos principal is allowed.





# KCA and CA acceptance Schedule

Dane Skow



- A prototype KCA is running now at FNAL and available for earlybird users.
- The detailed contents of the KCA proxy are likely to change slightly as discussions with the DOE Science Grid folks proceed toward making the FNAL (K)CA a subordinate CA of the DOE Science Grid CA.
- The service is running on a single dedicated server running in "Best Effort" support mode.
- The immediate plans are to bring a second KCA online for redundancy and resolve the KCA proxy format questions (by the end of August).
- Next would come resolving the subordinate CA issues with DOE Science Grid and implementing methods for non-user PKI credential generation (kcron jobs, host certificates, etc.) (perhaps the end of September ?)
- Then would come discussions about conditions and methods for acceptance of external PKI credentials (no schedule).







## Conclusion



- CDF is committed to use SAM. Many new features and issues will arise from this expanded use.
- Several new features and fixes have been added in the current SAM release, v4\_2\_0\_3p1.
- We are starting to monitor the SAM data transfers to remote stations and encourage more use of SAM at remote sites.
- Comparing network statistics to the remote sam usage gives some hints about how the network is being used.
- SAM-Grid is an architecture for a computing grid that is being developed and will enable easy and efficient use of compute resources around the world. This system is being built with “standard Grid Middleware” including Globus toolkit and Condor technology.
- Work is under way to understand and use various CA's.
- Encourage all to be involved

