MonALISA

MONitoring Agents using a Large Integrated Services Architecture

An Agent Based, Dynamic Service System to Monitor, Control and Optimize Distributed Systems

ICFA Workshop
Cracow, October 2006

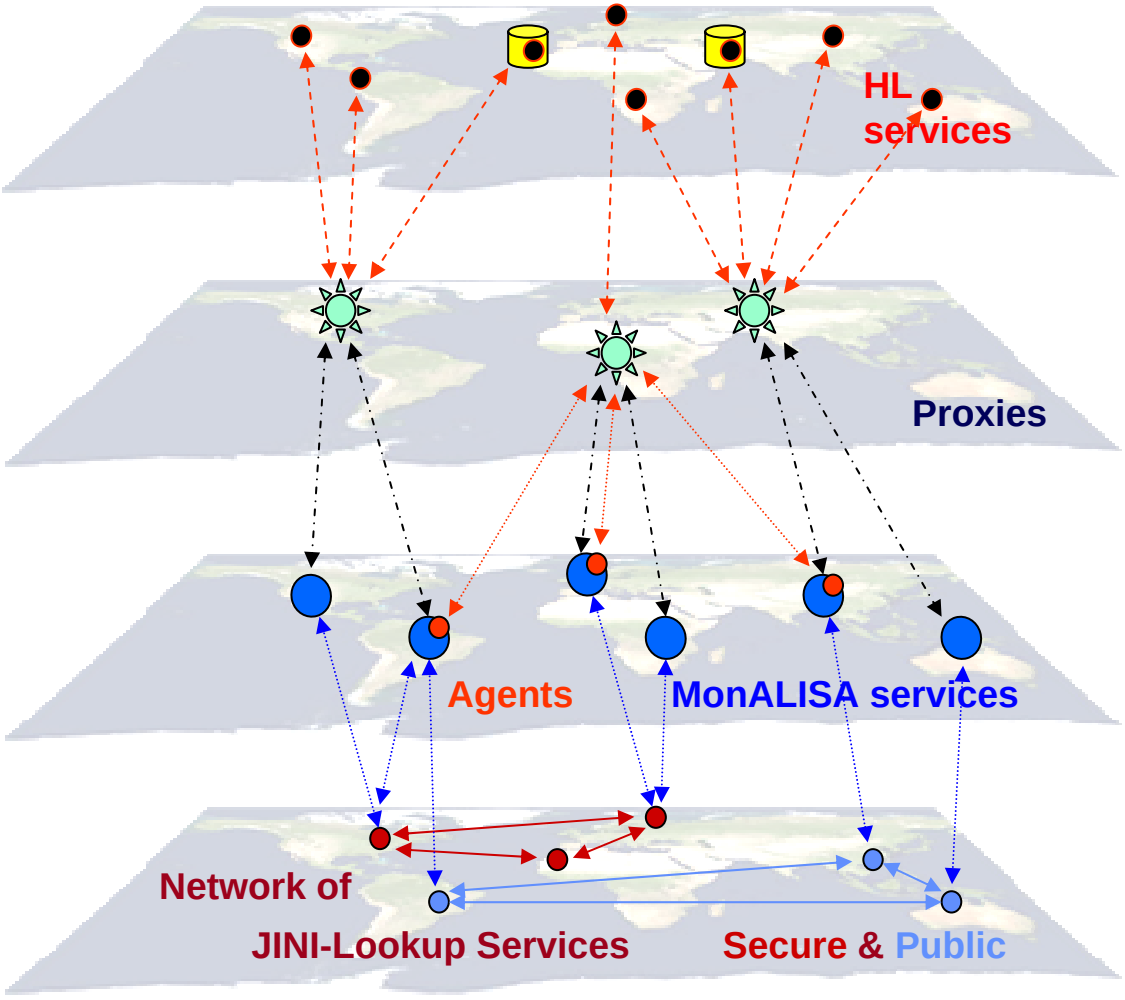
Iosif Legrand
California Institute of Technology

The MonALISA Framework

- **MonALISA is a Dynamic, Distributed Service System capable to collect any type of information from different systems, to analyze it in near real time and to provide support for automated control decisions and global optimization of workflows in complex grid systems.**

- **The MonALISA system is designed as an ensemble of autonomous multi-threaded, self-describing agent-based subsystems which are registered as dynamic services, and are able to collaborate and cooperate in performing a wide range of monitoring tasks. These agents can analyze and process the information, in a distributed way, and to provide optimization decisions in large scale distributed applications.**

The MonALISA Architecture



Regional or Global High Level Services, Repositories & Clients

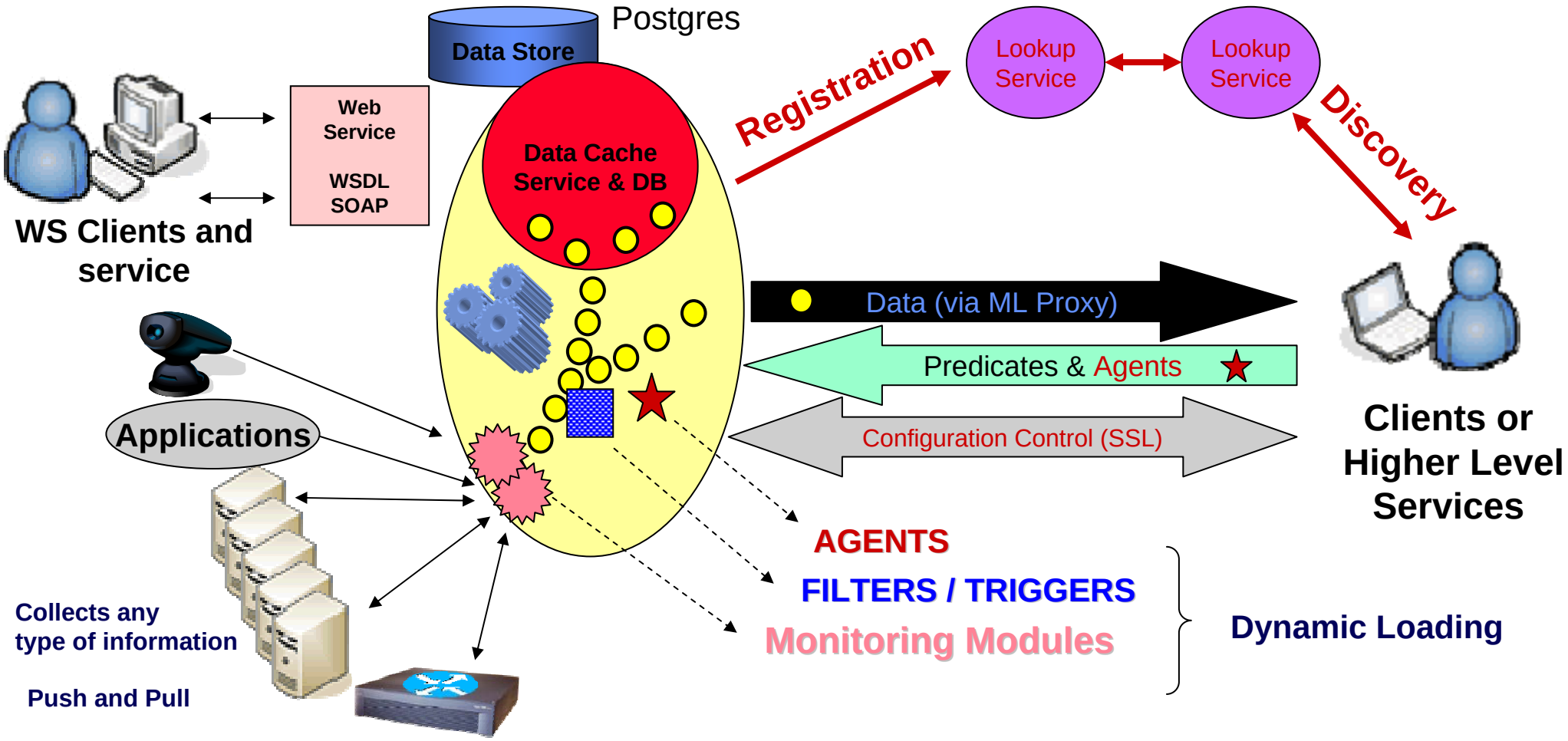
Secure and reliable communication
Dynamic load balancing
Scalability & Replication
AAA for Clients

Distributed System for gathering and analyzing information based on mobile agents:
Customized aggregation, Triggers, Actions

Distributed Dynamic Registration and Discovery-based on a lease mechanism and remote events

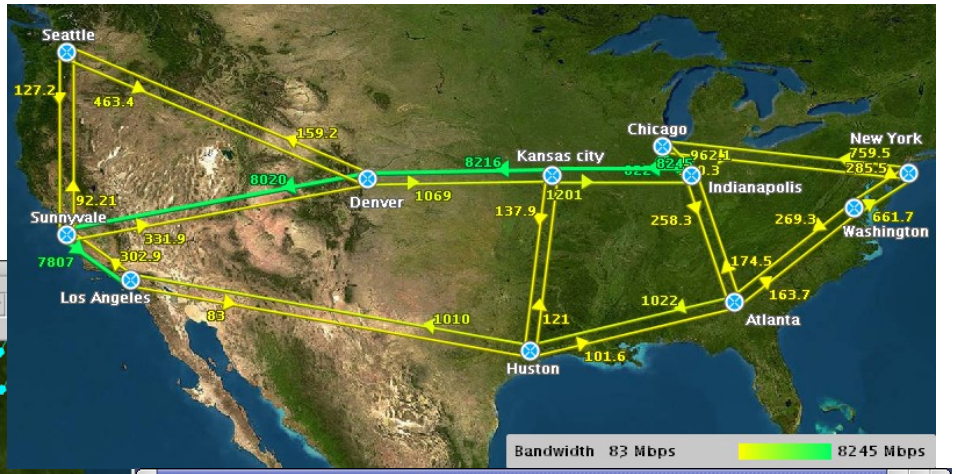
Fully Distributed System with no Single Point of Failure

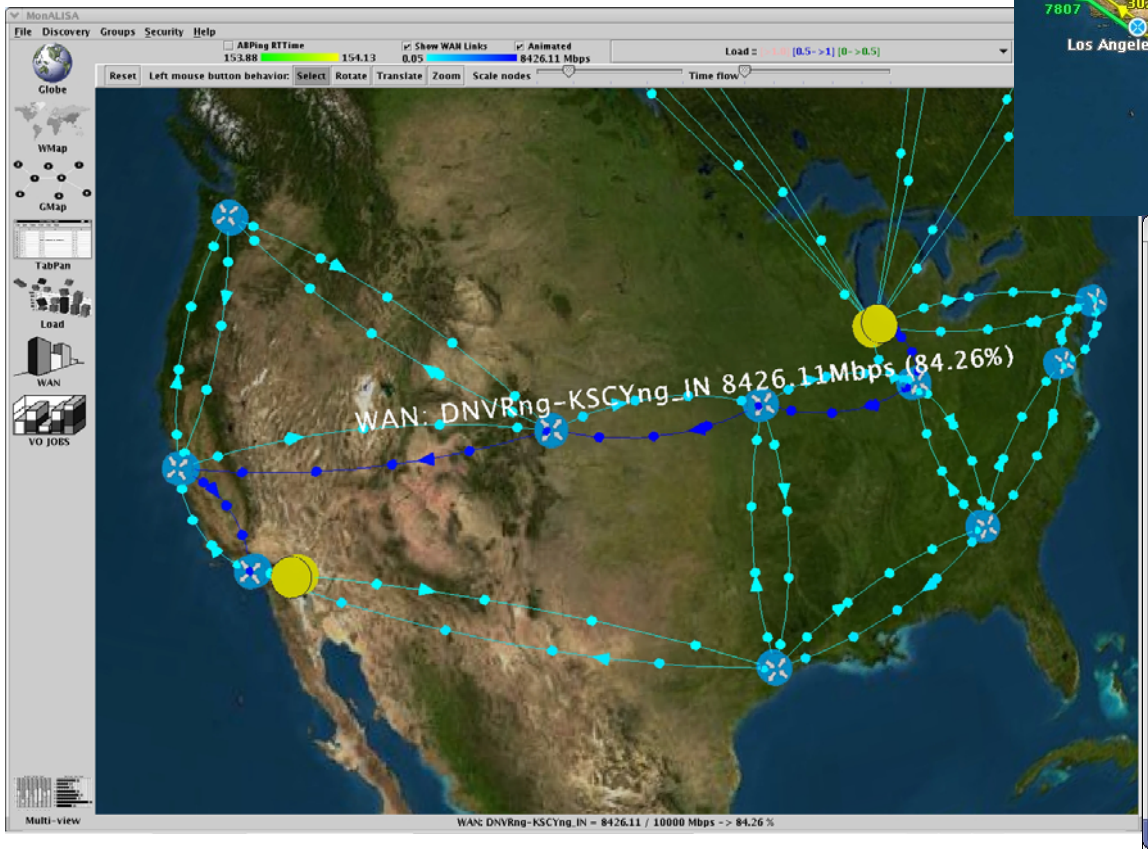
MonALISA service & Data Handling

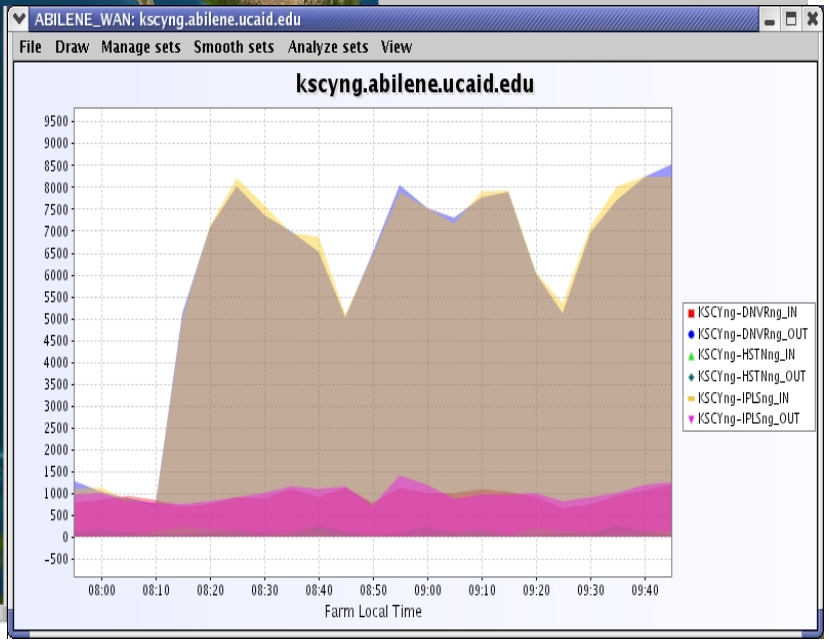


Monitoring Internet2 backbone Network

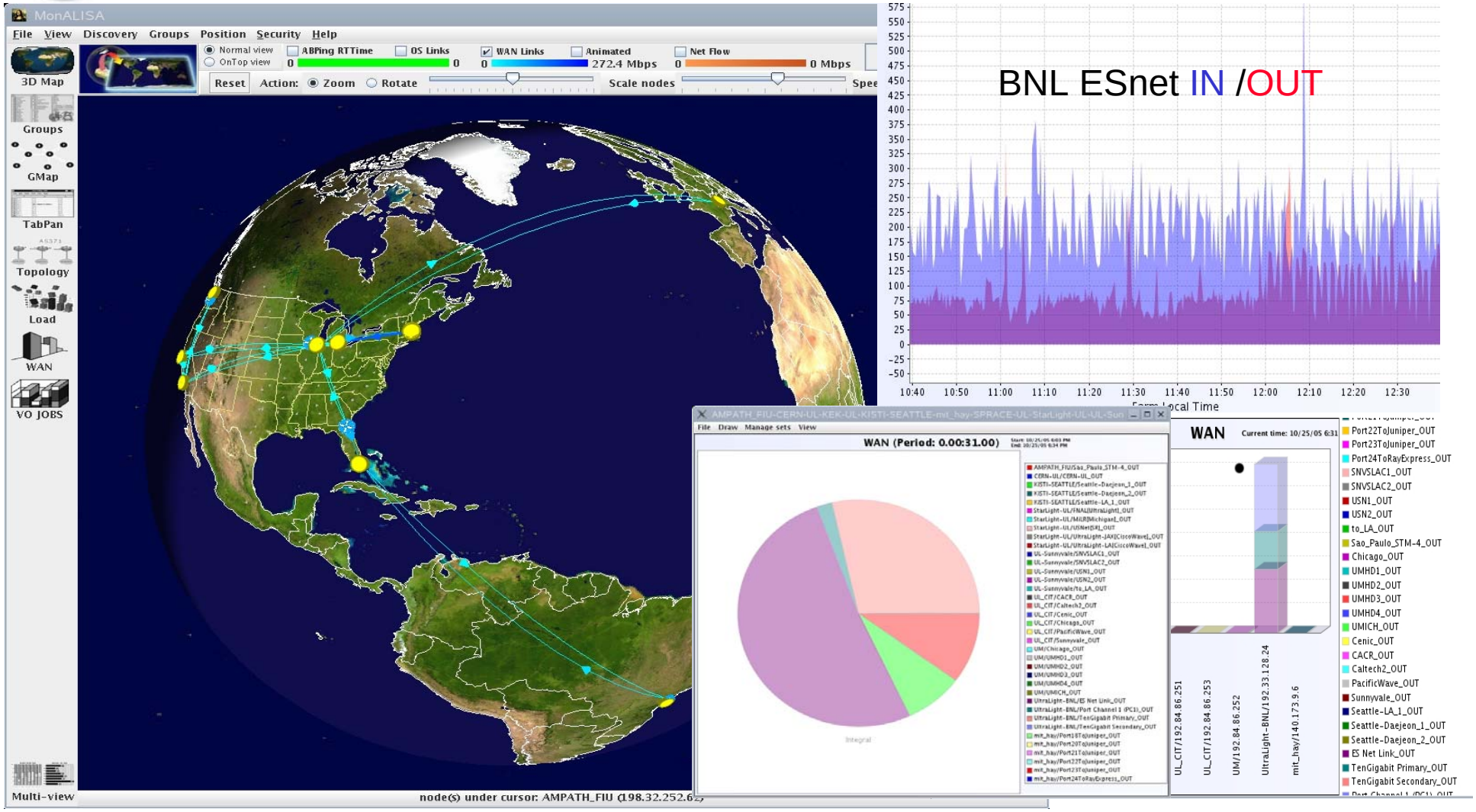
- Test for a Land Speed Record
- ~ 7 Gb/s in a single TCP stream from Geneva to Caltech



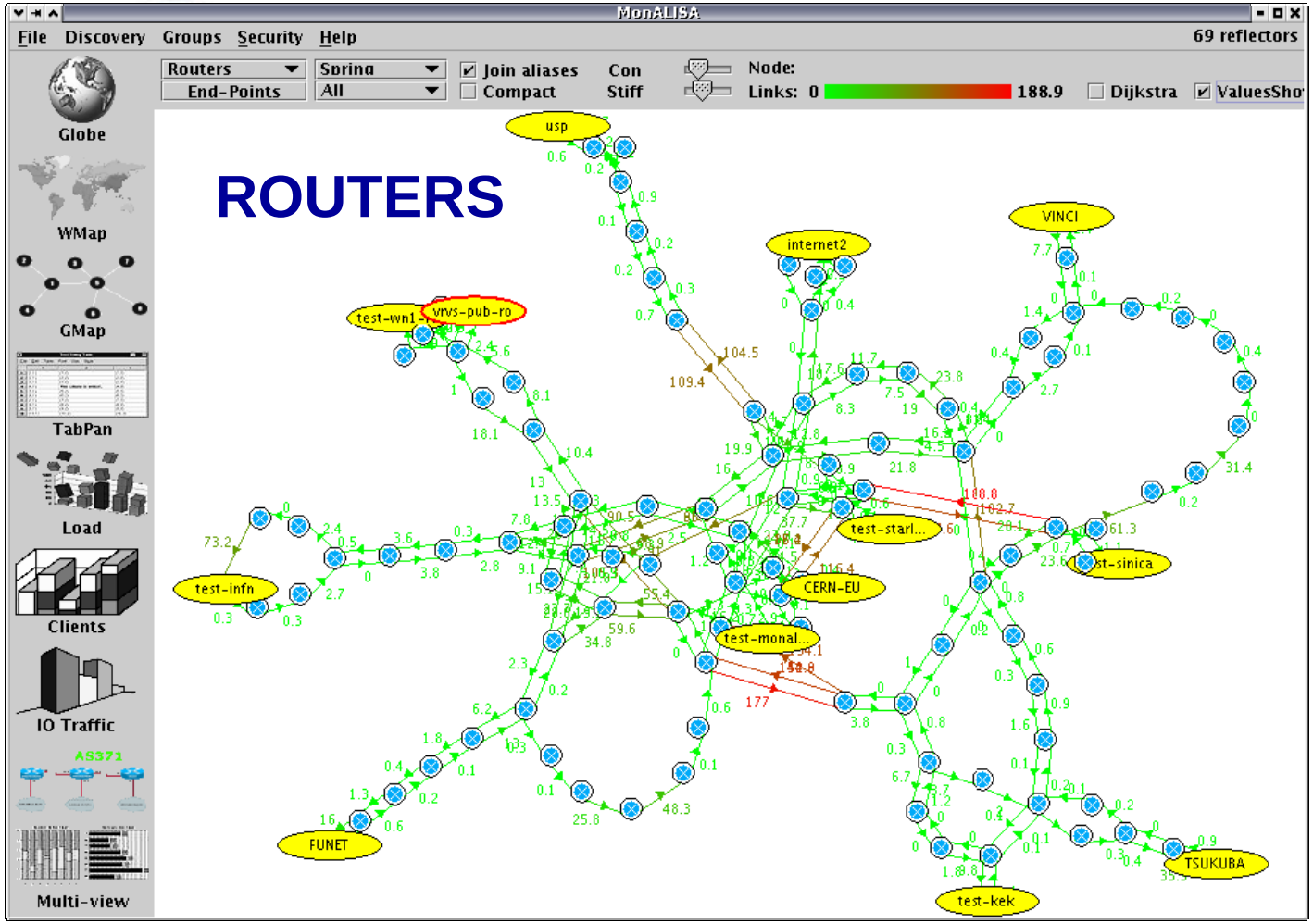




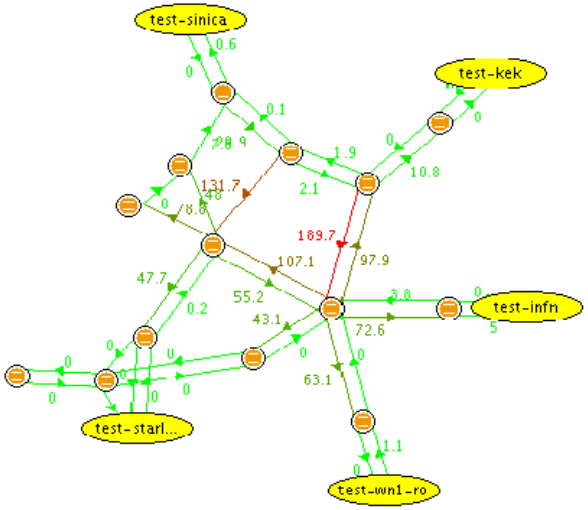
The UltraLight Network



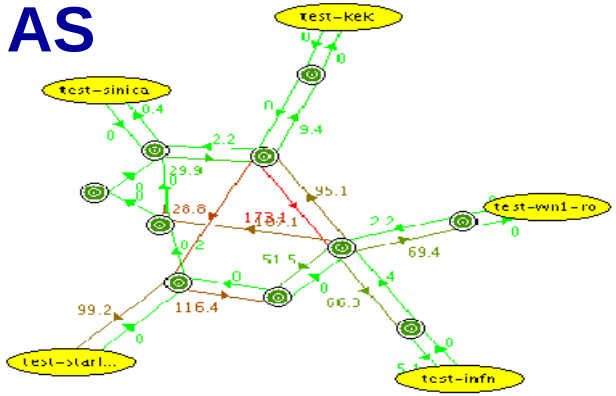
Monitoring Network Topology
Latency, Routers



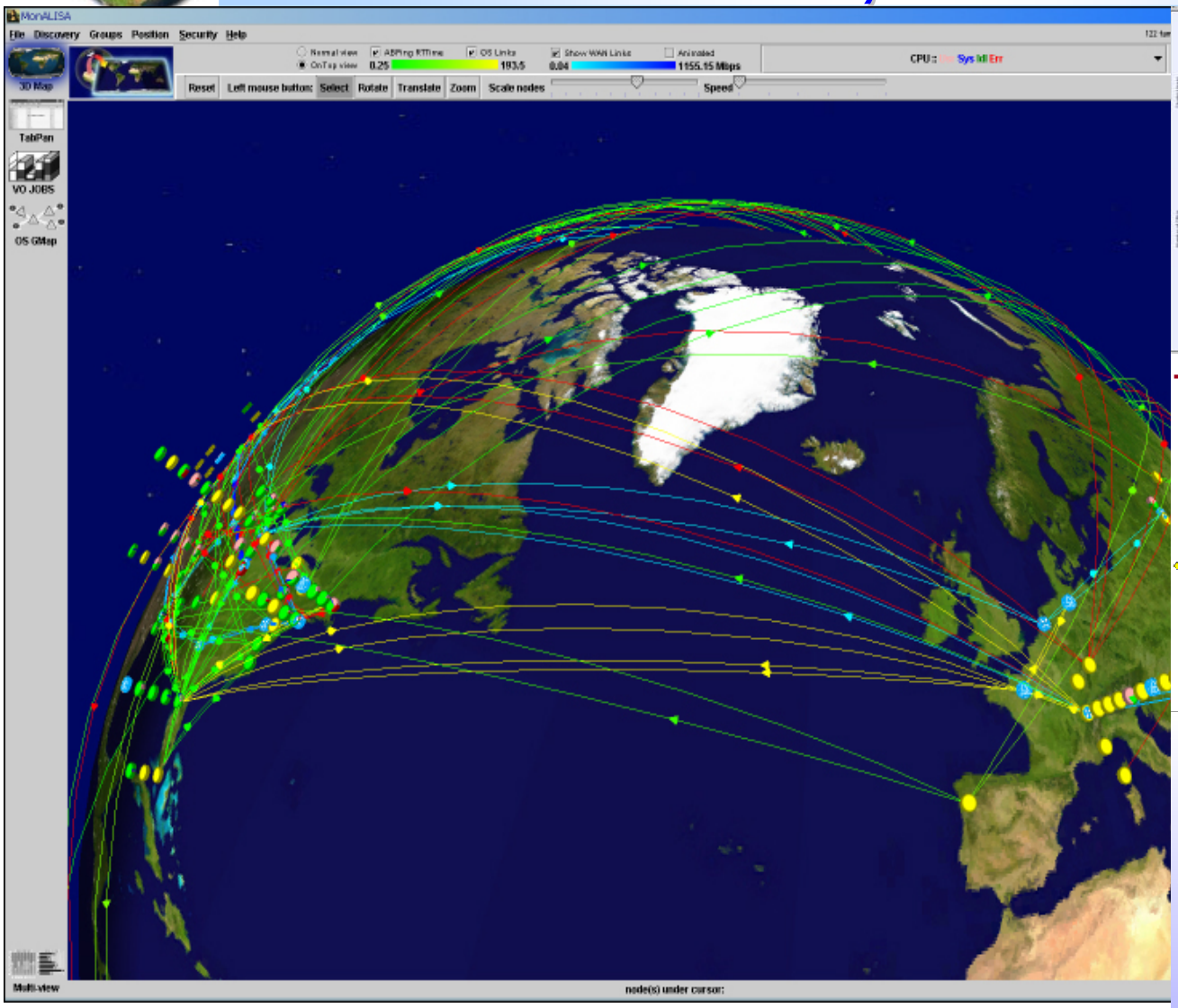
NETWORKS

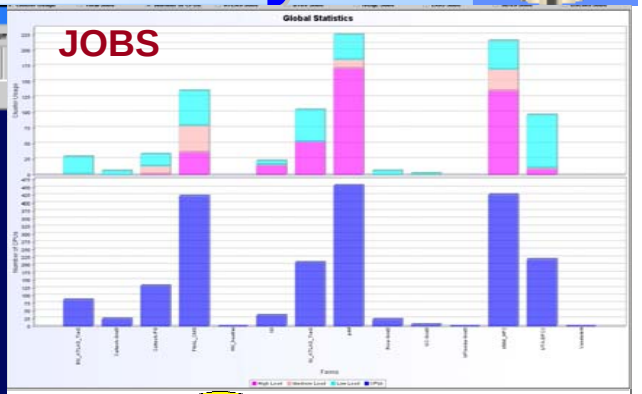


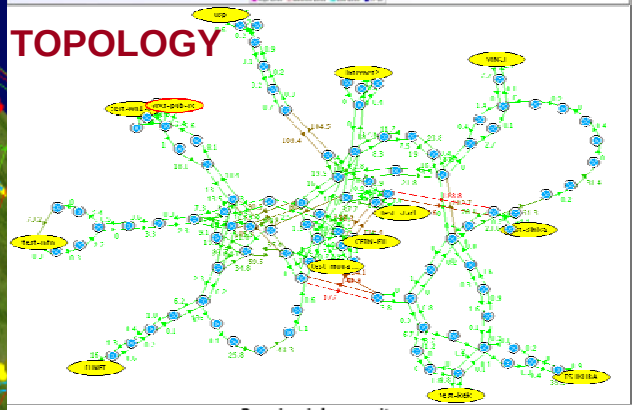
AS

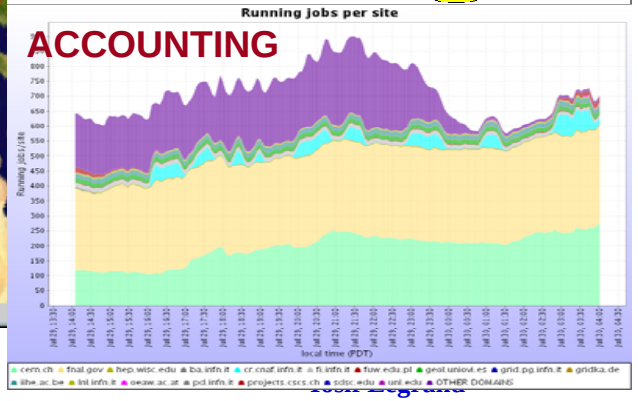


Monitoring Grid sites, Running Jobs, Network Traffic, and Connectivity

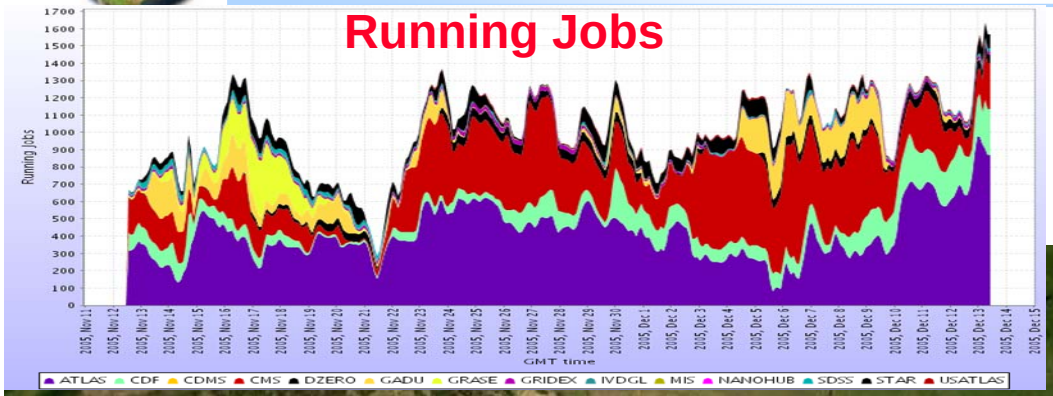


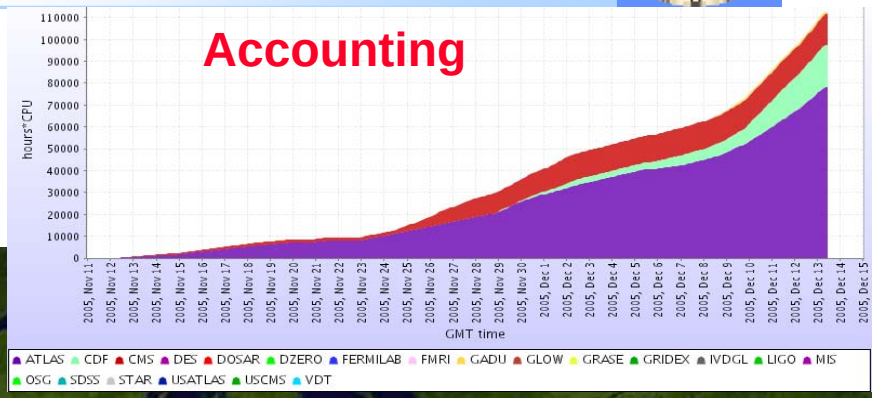




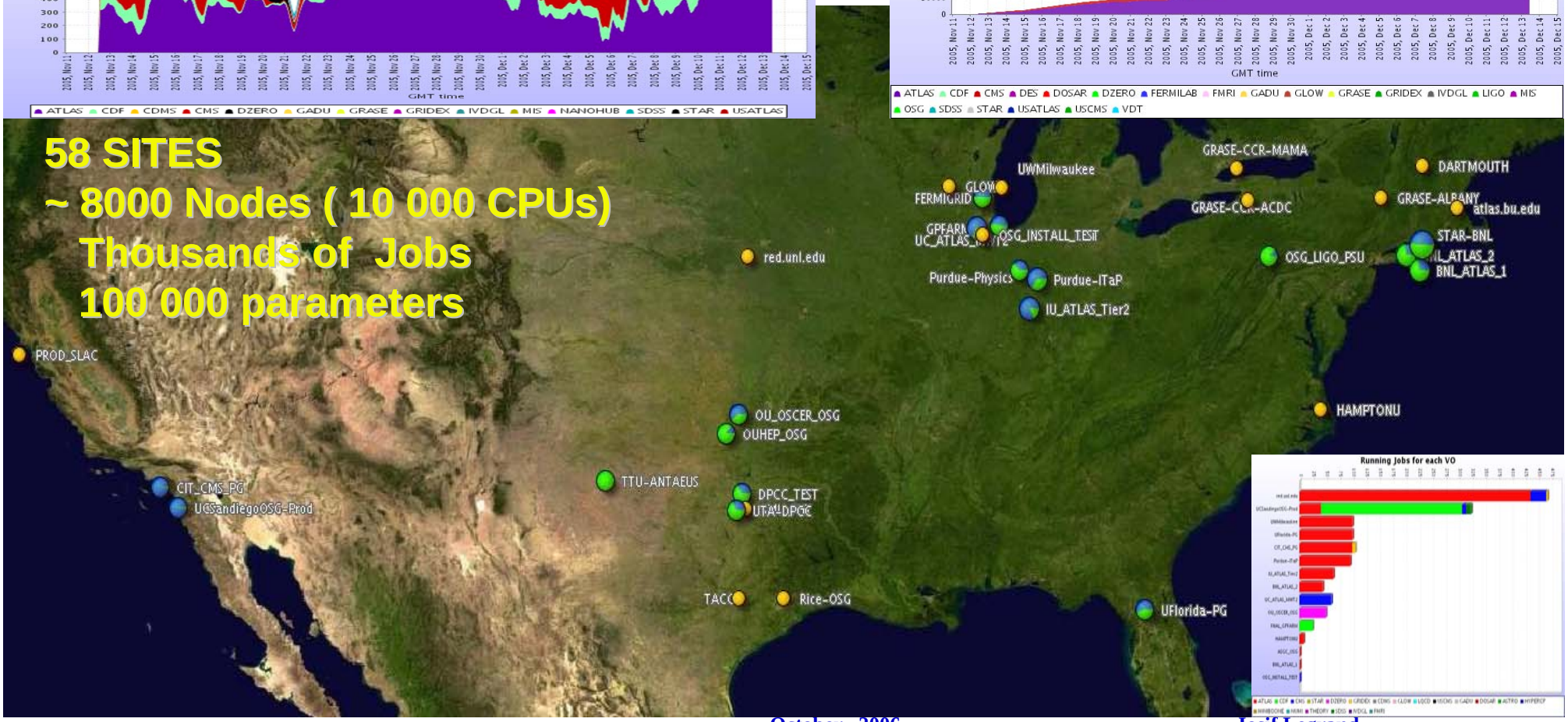


Monitoring OSG: Resources, Jobs & Accounting





58 SITES
~ 8000 Nodes (10 000 CPUs)
Thousands of Jobs
100 000 parameters



CMS Aggregate Job Monitoring

CMS Repository

- Global view
  - Map
  - Running jobs
  - Jobs per Site
  - Processing Events
  - Total Events
- Real time views
  - Running jobs
  - Site activity
- Sites
  - Masters Load5
- CMS File Servers
  - Load
  - Memory usage
  - Disk space
  - Disk rates
  - eth traffic
- WAN
  - Peering in Starligh
  - History
- Site info

Sites : (check all | uncheck all)

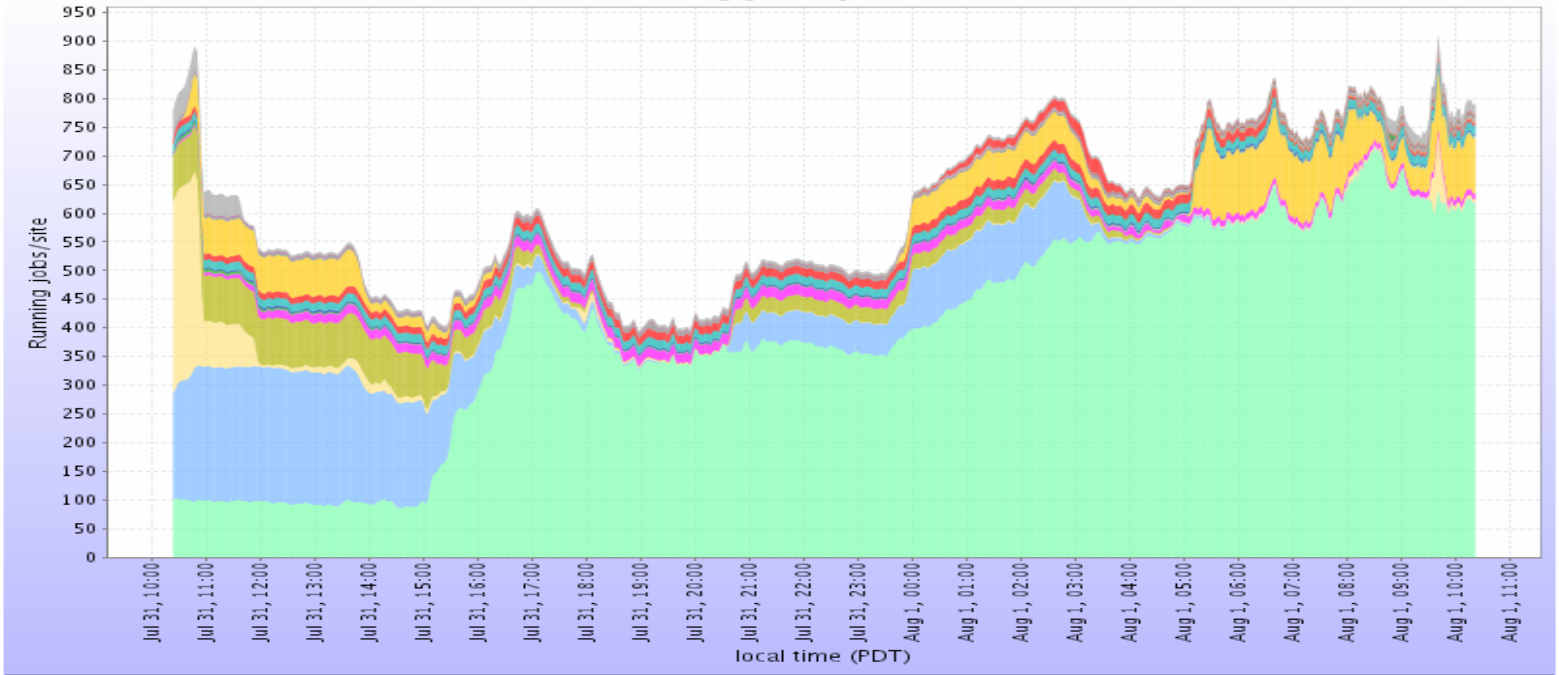
ba.infn.it ciemat.es cmsfarm2.bo.infn.it cr.cnaf.infn.it ehep.tifr.res.in ekpplus.cluster fi.infn.it fuw.edu.pl
fynu.ucl.ac.be geol.uniovi.es grid.pg.infn.it gridka.de gridpp.rl.ac.uk hep.ph.ic.ac.uk ifca.org.es
ihep.ac.cn iihe.ac.be in2p3.fr inrne.bas.bg kfki.hu lip.pt lmcg.wisc.edu lnl.infn.it oeaw.ac.at
pakgrid.org.pk pd.infn.it phy.ncu.edu.tw physics.ucdavis.edu physics.uoi.gr physik.uni-karlsruhe.de
projects.cscs.ch rice.edu sdsc.edu sinp.msu.ru stat.wisc.edu tier2.hep.man.ac.uk ucr.edu unl.edu
wn.iihe.ac.be OTHER DOMAINS

Most active: cern.ch cs.wisc.edu fnal.gov hep.wisc.edu physik.rwth-aachen.de pic.es

Quick interval: last day | Interval selection : Jul 31, 10:00 - Aug 1, 10:00

Sum series: disabled | Point shape: Small circle | Area view: Enabled | Stacked areas: Enabled | Plot

Running jobs per site



cern.ch cs.wisc.edu fnal.gov hep.wisc.edu physik.rwth-aachen.de pic.es ba.infn.it cmsfarm2.bo.infn.it
ehep.tifr.res.in fi.infn.it fuw.edu.pl geol.uniovi.es grid.pg.infn.it gridka.de ihep.ac.cn iihe.ac.be lmcg.wisc.edu
lnl.infn.it oeaw.ac.at pd.infn.it sdsc.edu sinp.msu.ru ucr.edu unl.edu wn.iihe.ac.be OTHER DOMAINS

In-depth or abstracted high-level information, as needed

ALICE : Jobs & resource usage monitoring

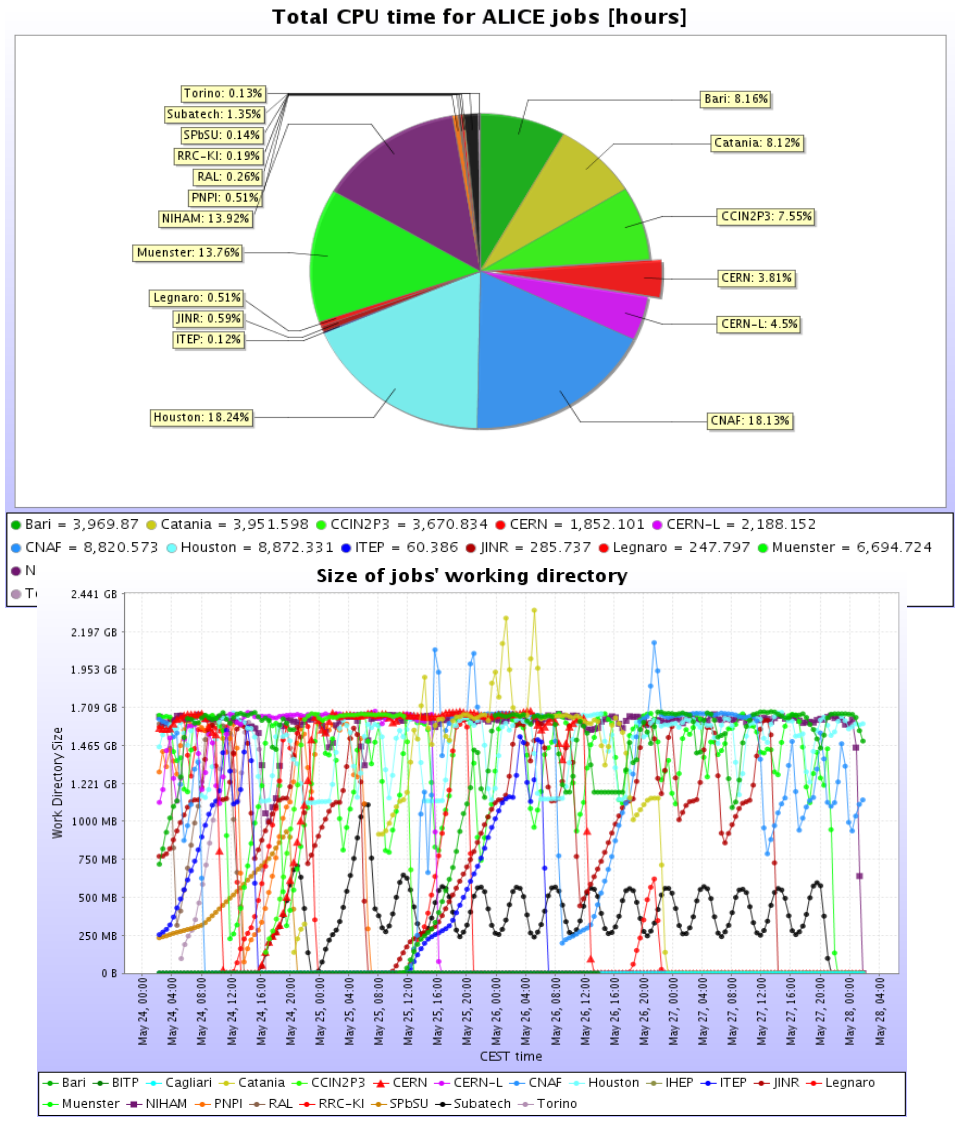
Cumulative parameters

- ➔ CPU Time
- ➔ Wall time
- ➔ Input & output traffic (xrootd)
- ➔ Read & written files

Running parameters

- ➔ Resident memory
- ➔ Virtual memory
- ➔ Open files
- ➔ Workdir size
- ➔ Disk usage
- ➔ CPU usage

Aggregated per site



Available Bandwidth Measurements

Embedded Pathload module.

MonALISA Repository
UltraLight

MonALISA Client
Click on the button below to start the Monalisa Client.
Client

- MonALISA Repository
- Global Views
- Node Info
- Available Bandwidt
  - Spider View
  - Matrix View
  - History
- Sites Status
- Site Info

close all

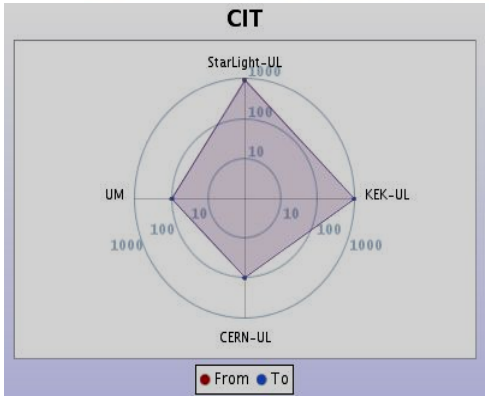
ABPing Configuration
Site Administration

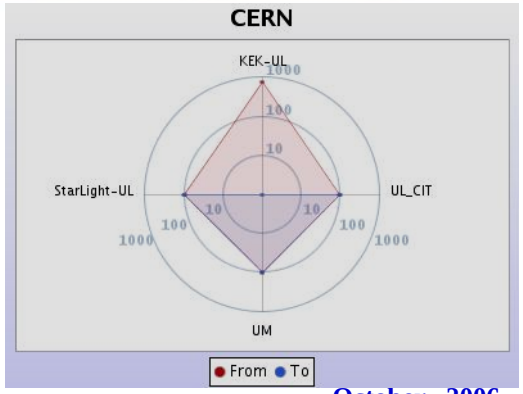
MonALISA
MONitoring Agents using a Large Integrated Services Architecture

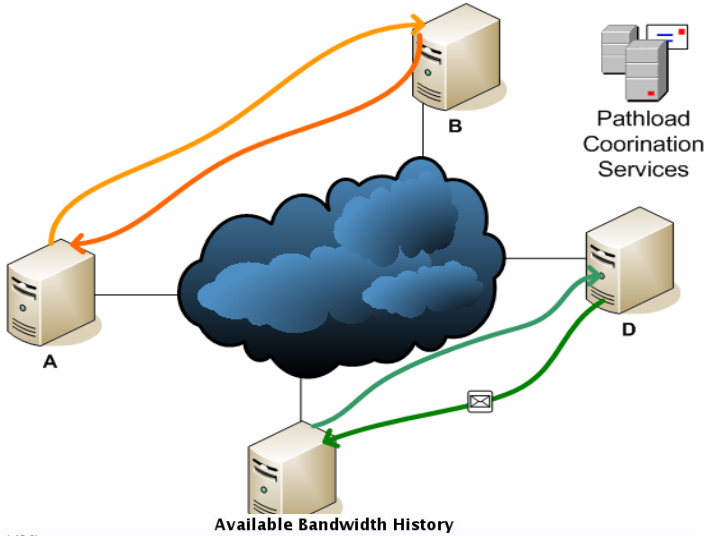
Available bandwidth measurements using pathload

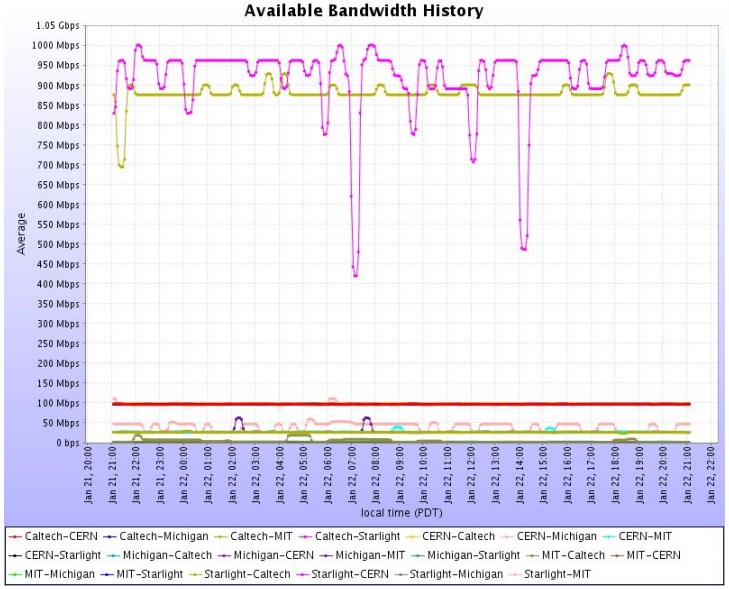
Available bandwidth between UL sites (average)

Site (from->to)	CERN-UL	KEK-UL	SPRACE-UL	UL_CIT	UM
CERN-UL	-	794.4 Mbps	97.37 Mbps	97.73 Mbps	97.85 Mbps
KEK-UL	750 Mbps	-	96.32 Mbps	993.2 Mbps	96.34 Mbps
StarLight-UL	97.4 Mbps	97.53 Mbps	-	875 Mbps	97.5 Mbps
UL_CIT	97.5 Mbps	993.2 Mbps	876 Mbps	-	96.63 Mbps
UM	97.48 Mbps	96.84 Mbps	96.55 Mbps	96.85 Mbps	-
Min	**97.4 Mbps**	**97.53 Mbps**	**96.32 Mbps**	**96.85 Mbps**	**96.34 Mbps**
Max	**750 Mbps**	**993.2 Mbps**	**876 Mbps**	**993.2 Mbps**	**97.85 Mbps**

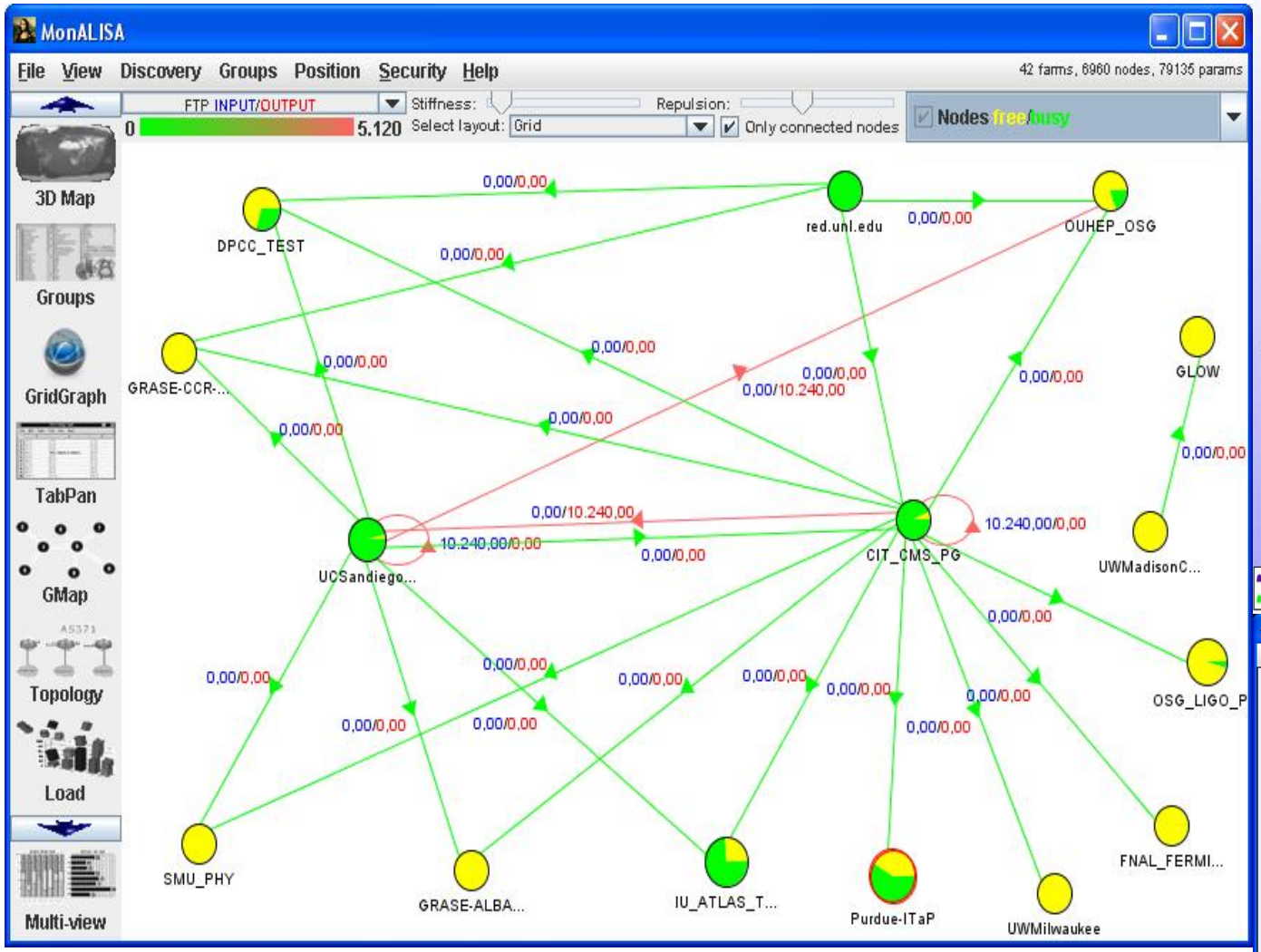


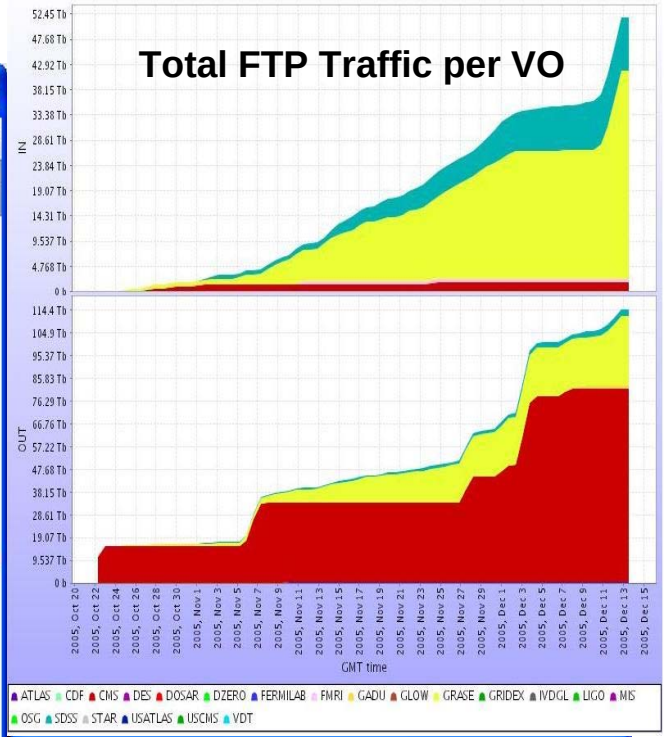


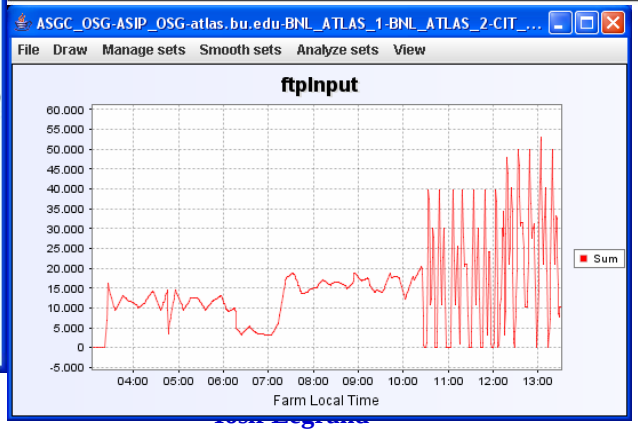




FTP Data Transfer between GRID sites







ApMon – Application Monitoring

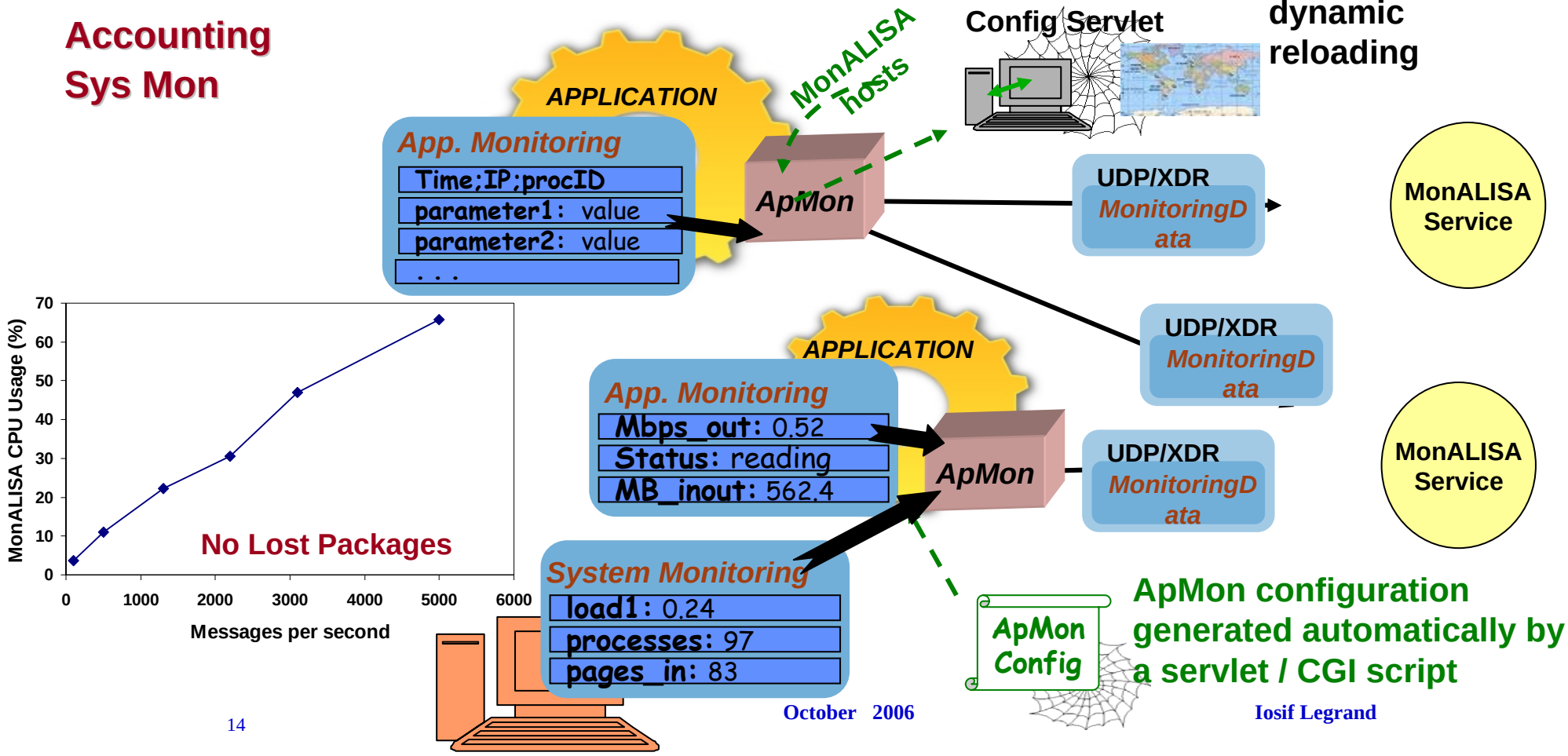
Lightweight library of APIs (C, C++, Java, Perl, Python) that can be used to send any information to MonALISA Services

High comm. performance

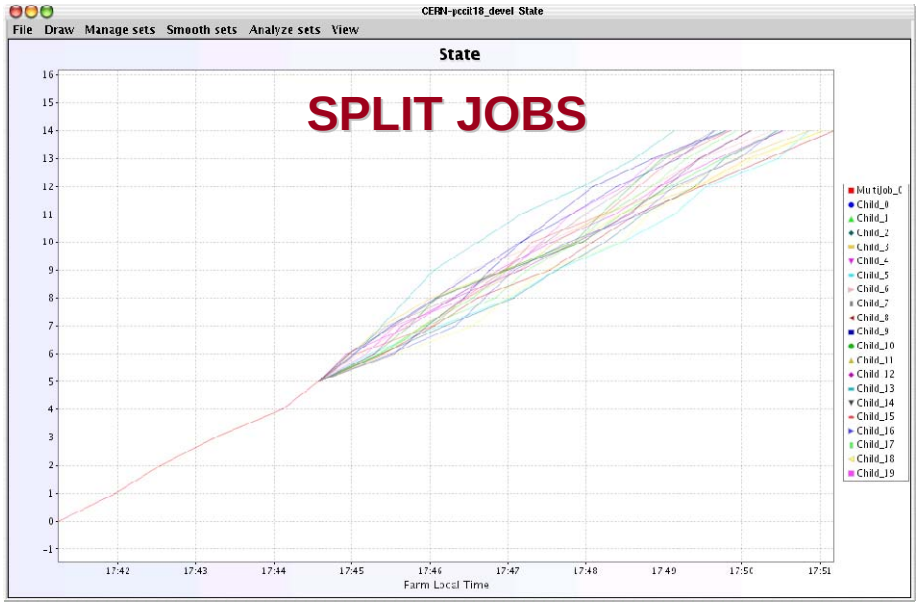
Flexible

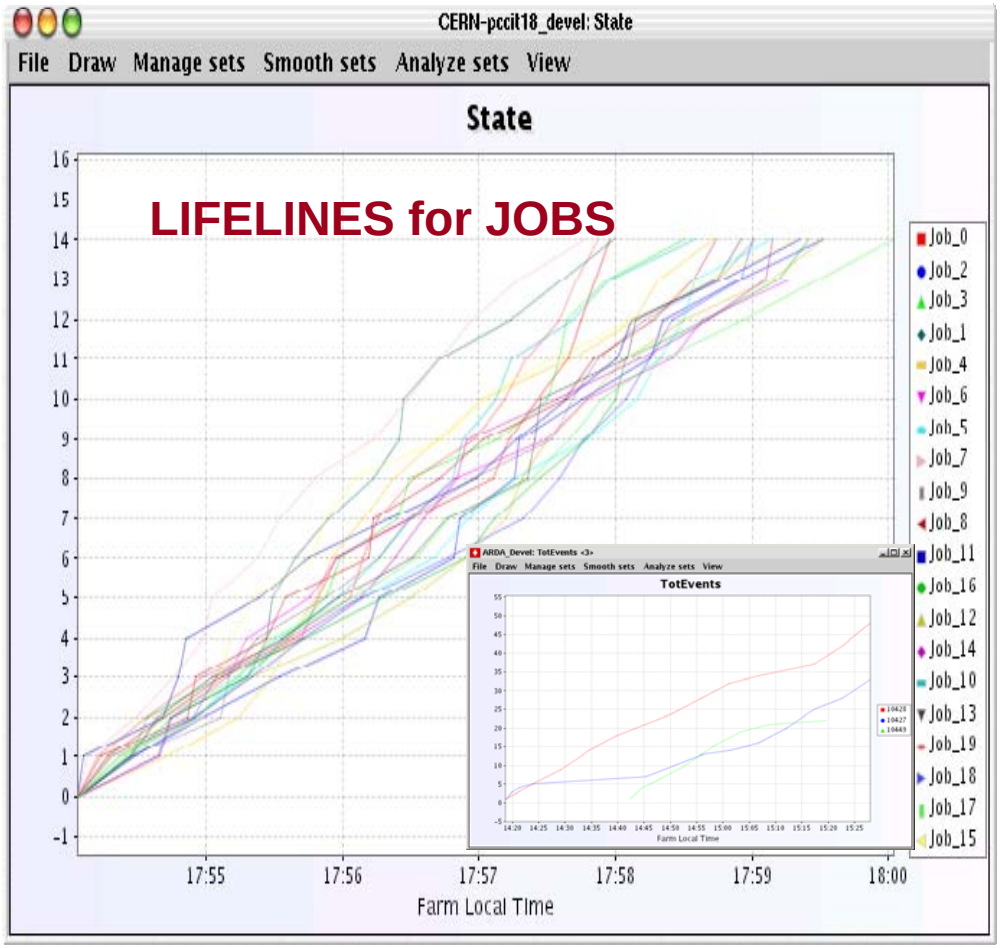
Accounting

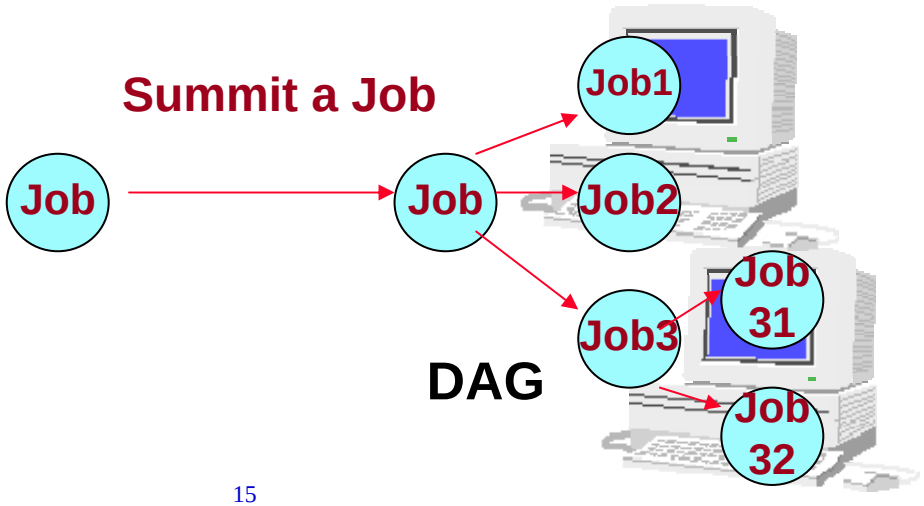
Sys Mon



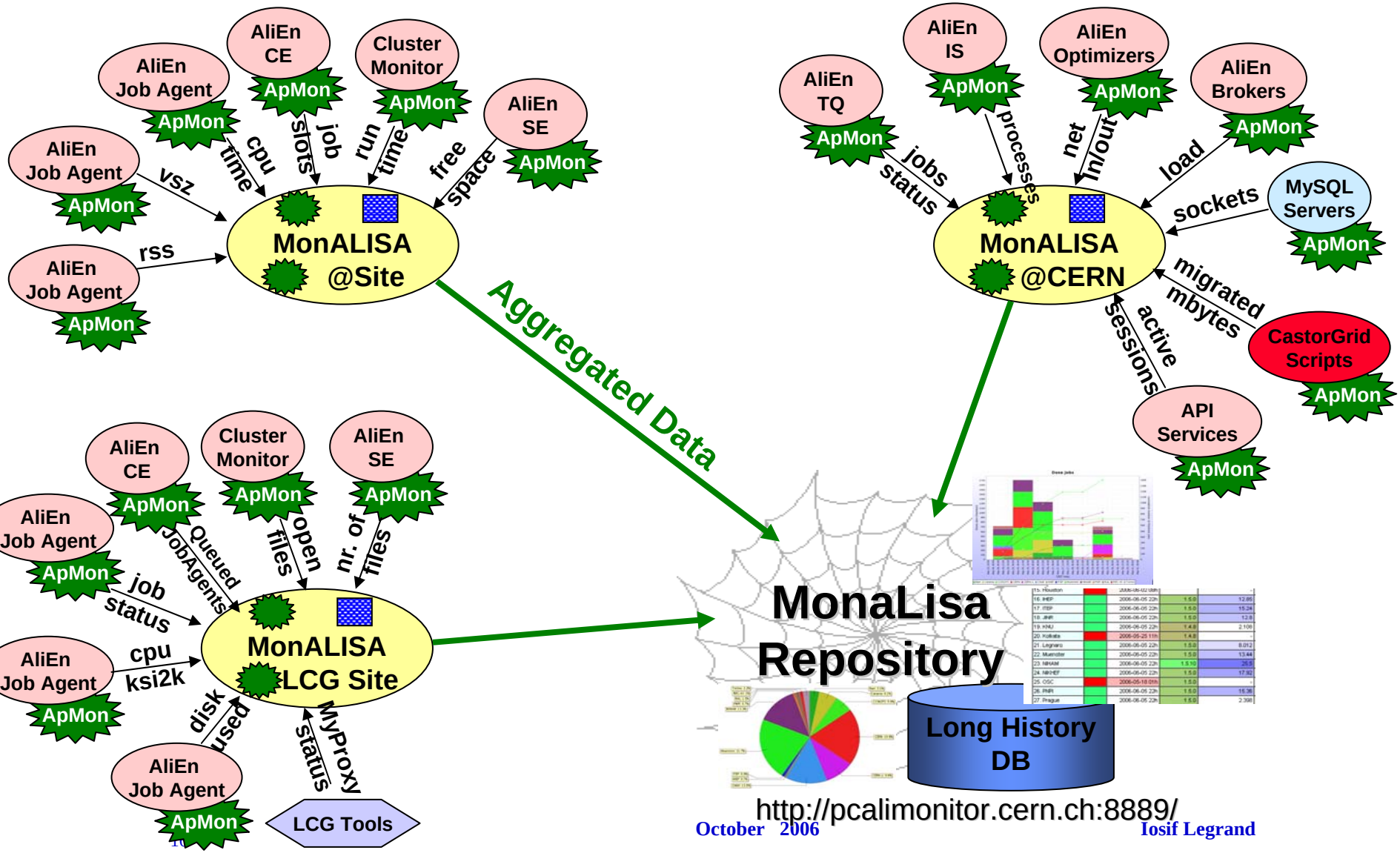
Monitoring the Execution of Jobs and the Time Evolution





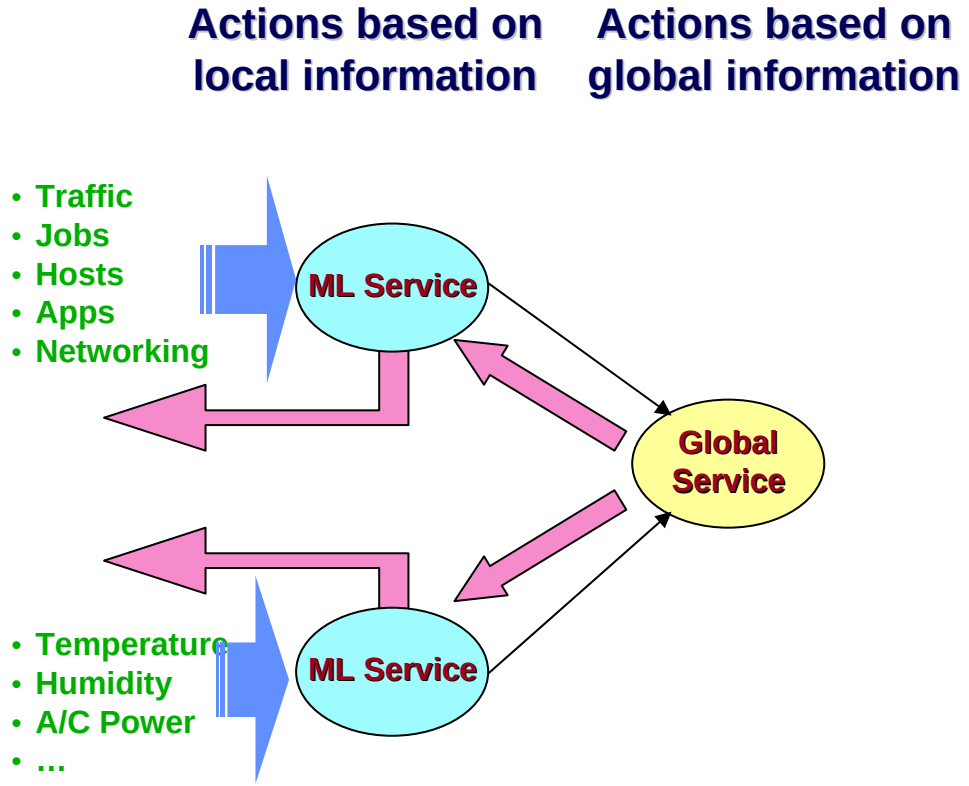


The Alien Monitoring Architecture

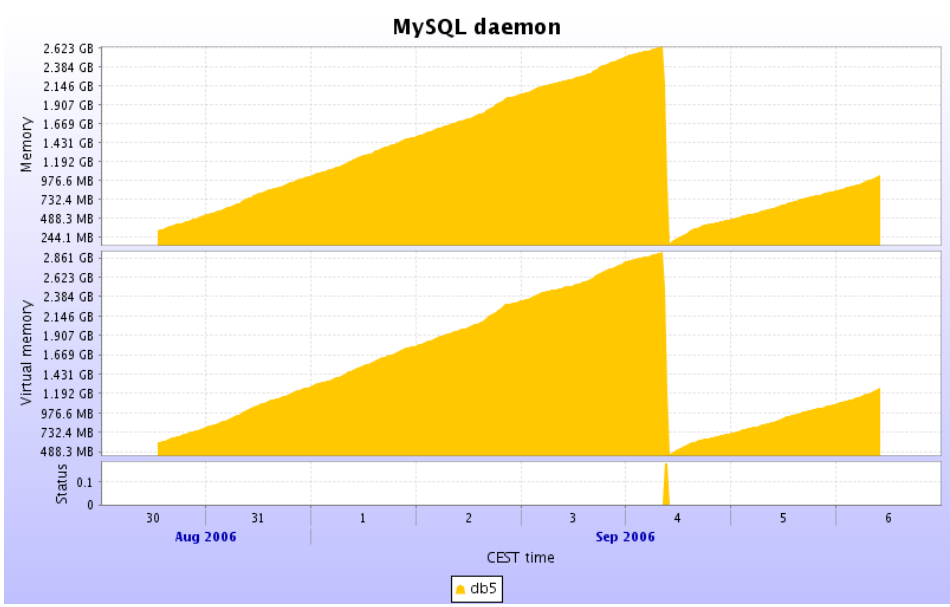


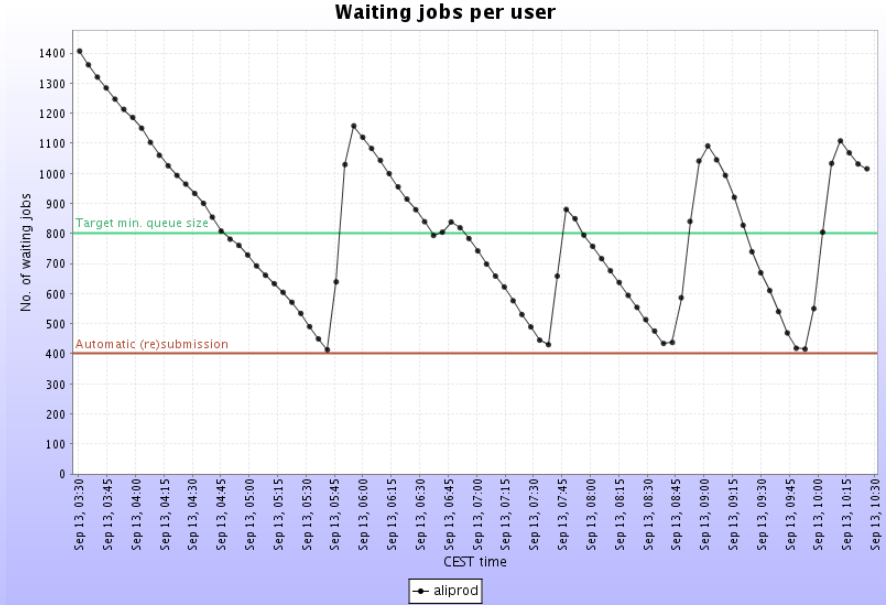
Operational Decisions and Actions

- Based on monitoring information, actions can be taken in
  - ML Service
  - ML Repository
- Actions can be triggered by
  - Values above/below given thresholds
  - Absence/presence of values
  - Correlation between multiple values
- Operational actions
  - Alerts
    - e-mail
    - Instant messaging
  - Supervision for Services
  - External commands
  - Event logging



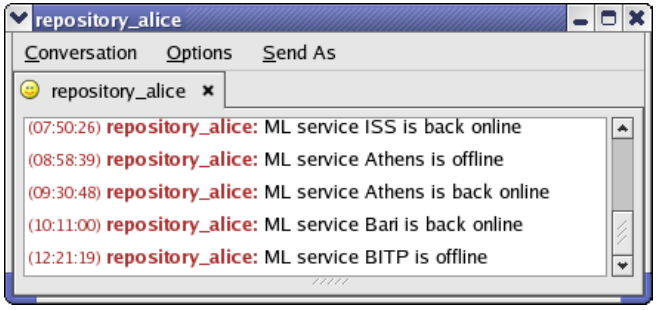
ALICE Examples





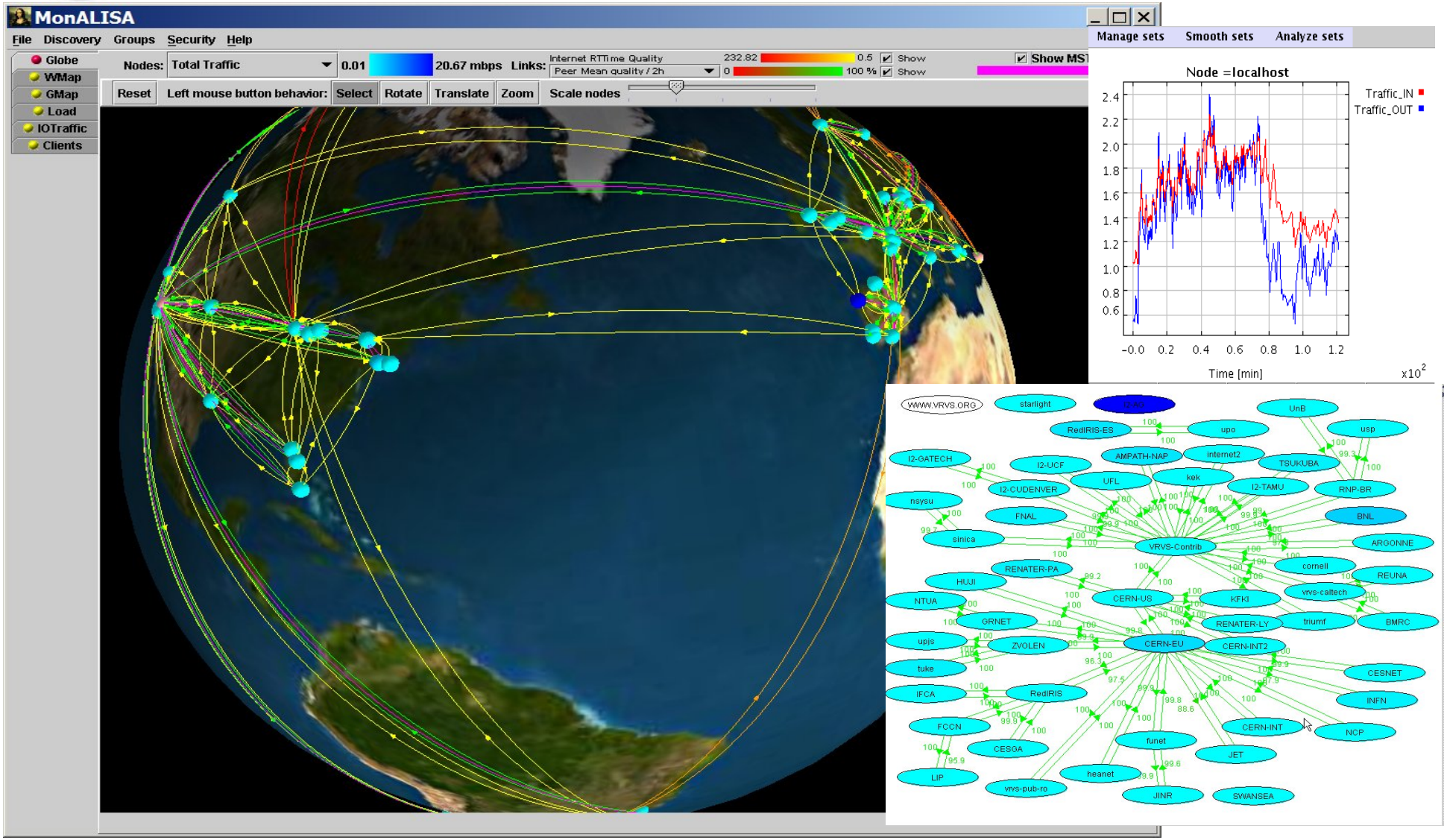
MySQL daemon is automatically restarted when it runs out of memory
Trigger: threshold on VSZ memory usage

ALICE Production jobs queue is automatically kept full by the automatic resubmission
Trigger: threshold on the number of *aliprod* waiting jobs

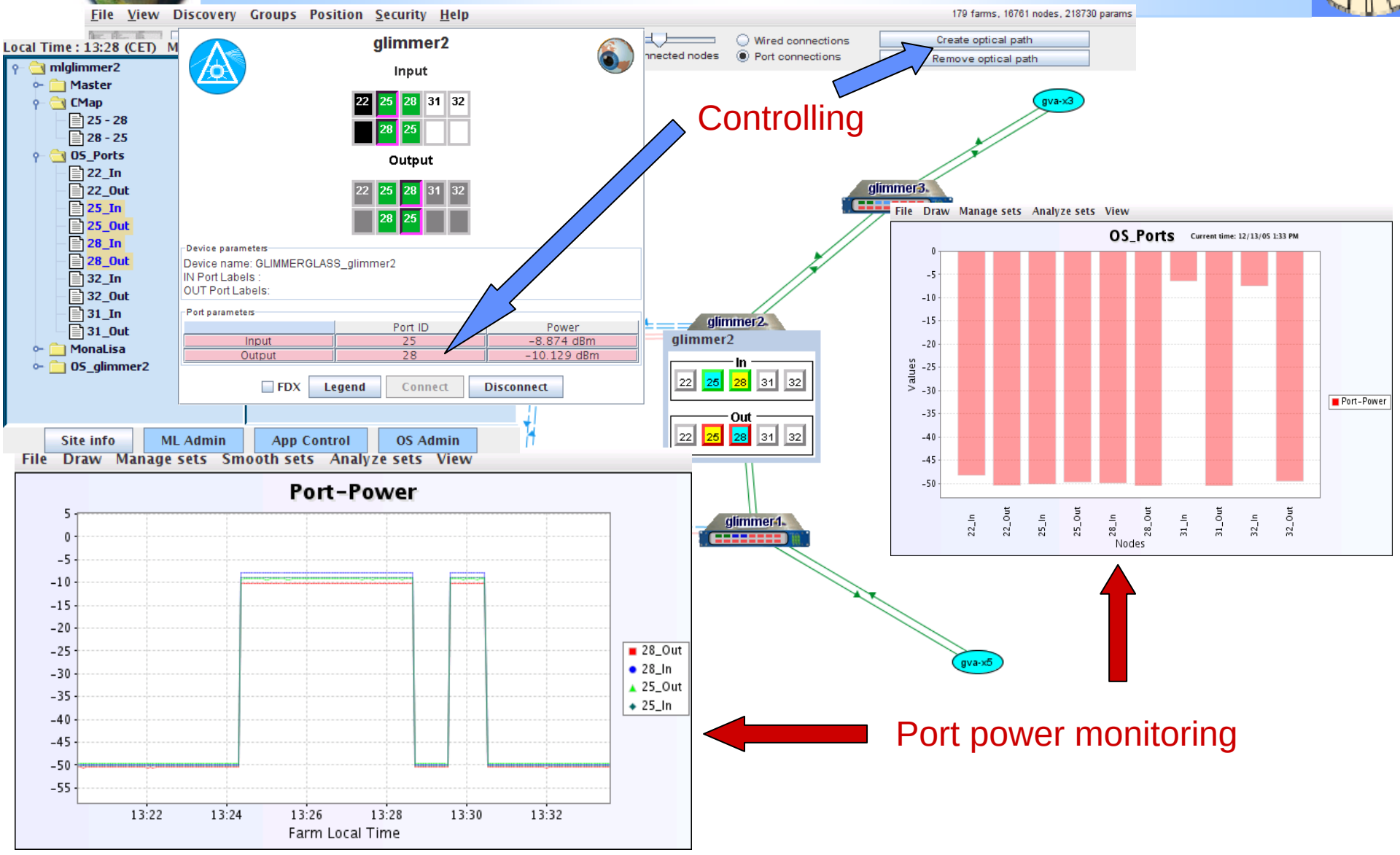


Administrators are kept up-to-date on the services' status
Trigger: presence/absence of monitored information

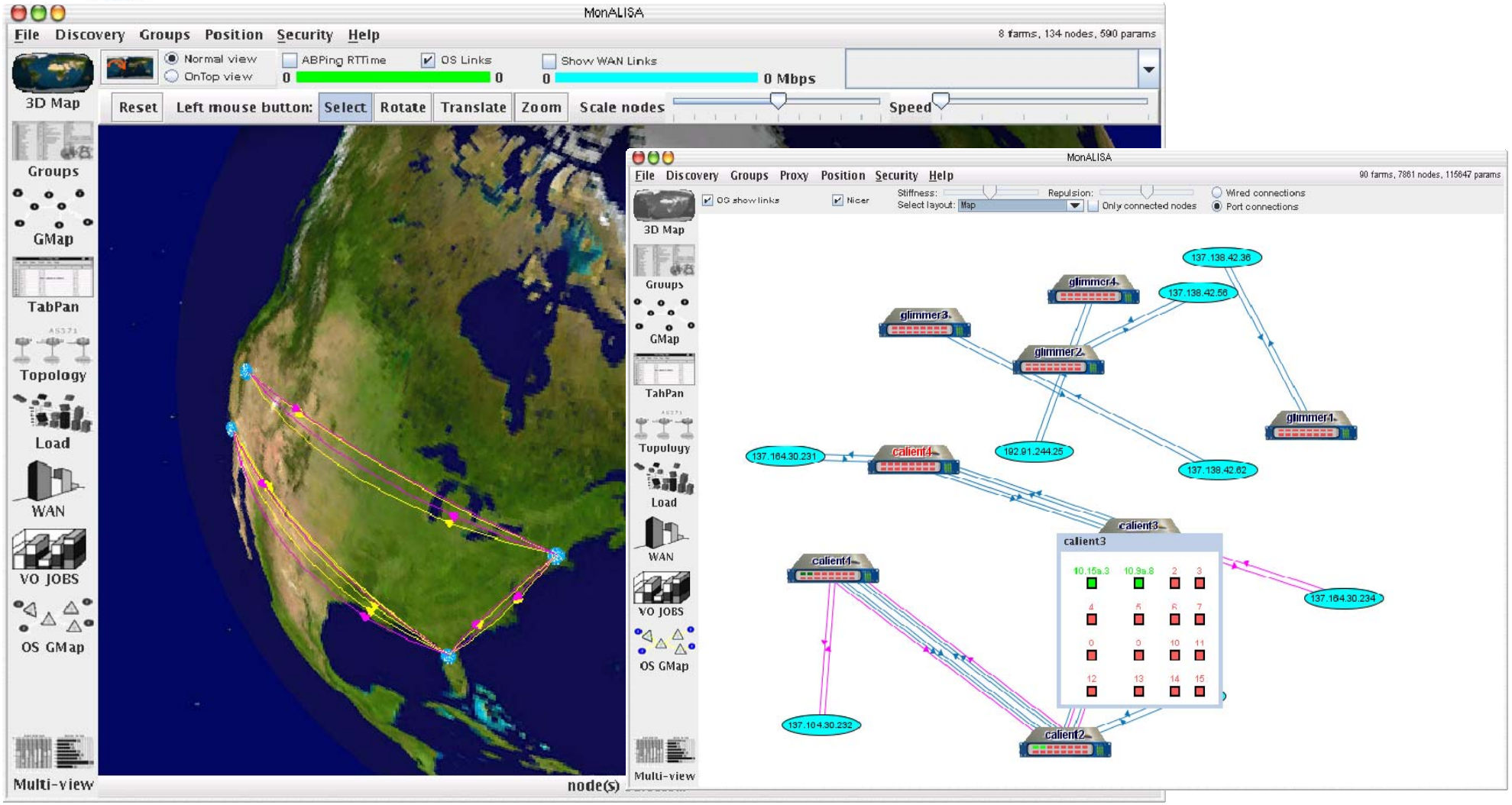
Monitoring Video Conference System: Reflectors and Communication Topology



Monitoring and Controlling Optical Planes

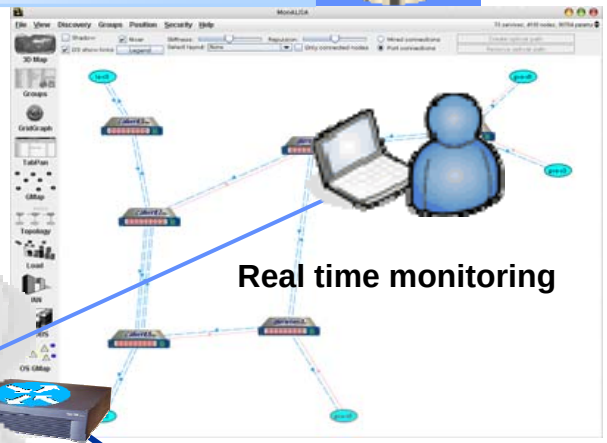


Monitoring Optical Switches
Agents to Create on Demand an Optical Path



“On-Demand”, Dynamic Path Allocation

CREATES AN END TO END PATH < 1s



Internet

MonALISA Distributed Service System

Regular IP path

APPLICATION

A

DATA

B

Control

Monitor

MonALISA Service

OS Agent

LISA Agent

TL1

Optical Switch

Active light path

Detects errors and automatically recreate the path in less than the TCP timeout

LISA AGENT

LISA sets up
- Network Interfaces
- TCP stack
- Kernel parameters
- Routes

LISA → APPLICATION
“use eth1.2, …”

Communities using MonALISA

Major Communities

- OSG
- CMS
- ALICE
- D0
- STAR
- VRVS
- LGC RUSSIA
- SE Europe GRID
- APAC Grid
- UNAM Grid (Mx)
- ITU

- ABILENE
- ULTRALIGHT
- GLORIAD
- LHC Net
- RoEduNET
- Enlightened

MonALISA Today

Running 24 X 7
at ~300 Sites

- Collecting > 600,000 parameters in near real-time
- Update rate of 20,000 parameter updates per second
- Monitoring
  - 12,000 computers
  - > 100 WAN Links
- Thousands of Grid jobs running concurrently

Demonstrated at:

- Telecom World
- WSIS 2003
- SC 2004
- Internet2 2005
- TERENA 2005
- IGrid 2005
- SC 2005
- CHEP 2006
- CENIC 2006
  Innovation Award for High-Performance Applications

The MonALISA Architecture Provides:

- Distributed Registration and Discovery for Services and Applications.
- Monitoring all aspects of complex systems :
  - System information for computer nodes and clusters
  - Network information : WAN and LAN
  - Monitoring the performance of Applications, Jobs or services
  - The End User Systems, its performance
  - Environment; Video streaming
- Can interac… ustomized information…
- Secure, re…
- Agents to … econfigure them, and to notify other services when certain conditions are detected.
- The MonALISA framework is used to develop higher level decision services, implemented as a distributed network of communicating agents, to perform global optimization tasks.
- Graphical User Interfaces to visualize complex information

http://monalisa.caltech.edu