



MonALISA

*MONitoring Agents using a Large
Integrated Services Architecture*

**An Agent Based, Dynamic Service System to Monitor,
Control and Optimize Grid based Applications**

CHEP 2004



Iosif Legrand
California Institute of Technology



Monitoring Services



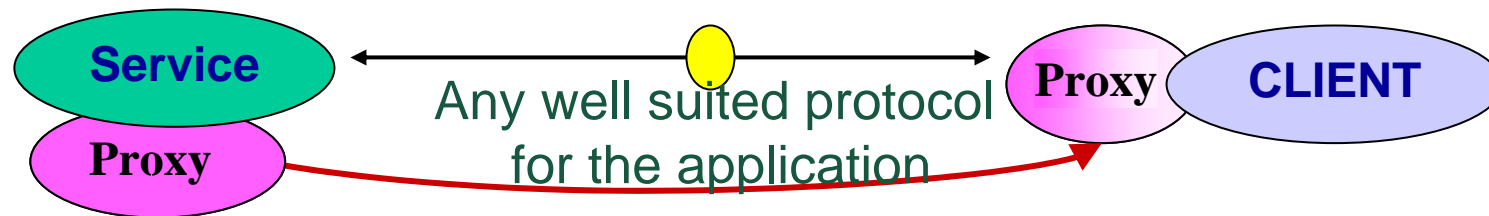
- **An essential part of managing a global Data Grid is a monitoring system that is able to monitor and track the many site facilities, networks, and the many tasks in progress, in real time.**
 - **System information for nodes and clusters**
 - **Network information Wan and LAN**
 - **Application monitoring**
- **The monitoring information gathered also is essential for developing the required higher level services, and components of the Grid system that provide decision support, and eventually some degree of automated decisions, to help maintain and optimize workflow through the Grid.**
- **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 and decisions in large scale distributed applications.**



MonALISA is A Dynamic, Distributed Service Architecture



- The MonALISA system is designed as an ensemble of autonomous multi-threaded, self-describing agent-based systems



Code Mobility Paradigm Dynamic Loading of Service

- Remote Services Proxy == RMI Stub
- Mobile Agents Proxy == Entire Service
- “Smart Proxies” Proxy adjusts to the client

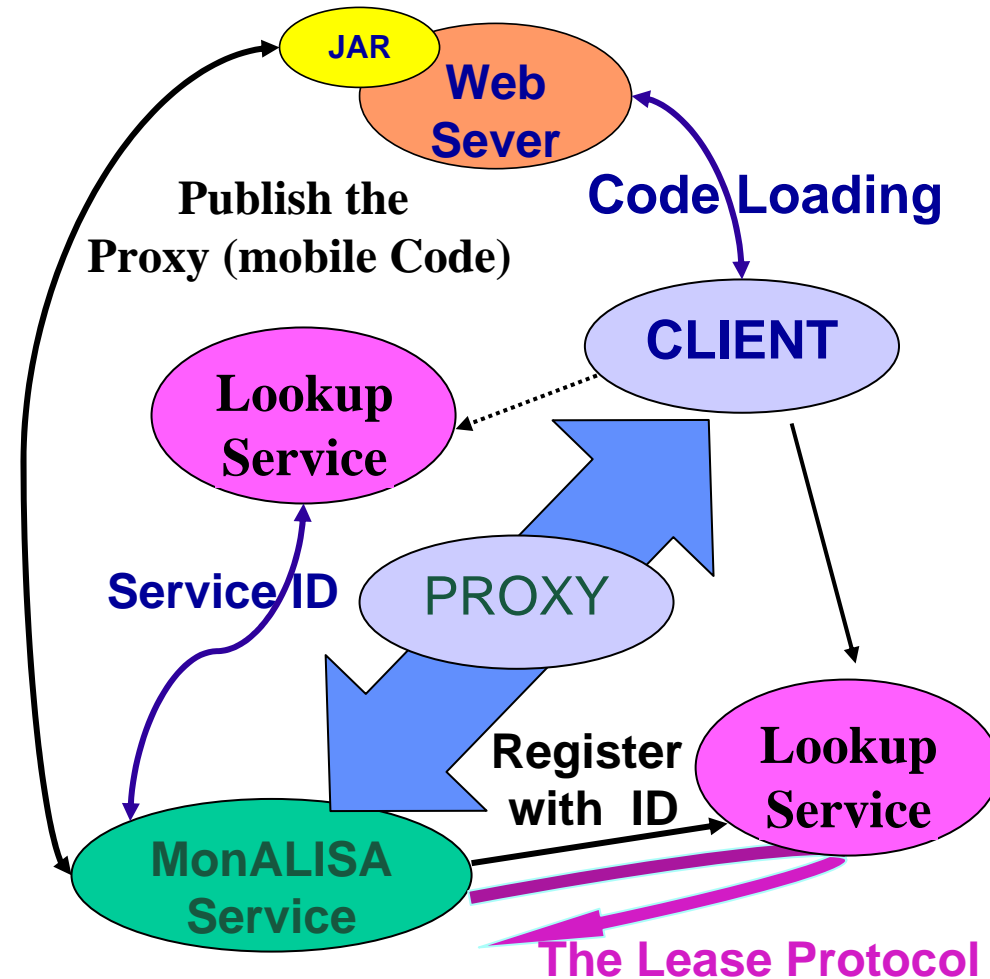
An agent-based architecture provides the ability to invest the system with increasing degrees of intelligence; to reduce complexity and make global systems (Grids and networks) manageable in real time



The Key MonALISA Features for a Reliable and Scalable Monitoring and Management System



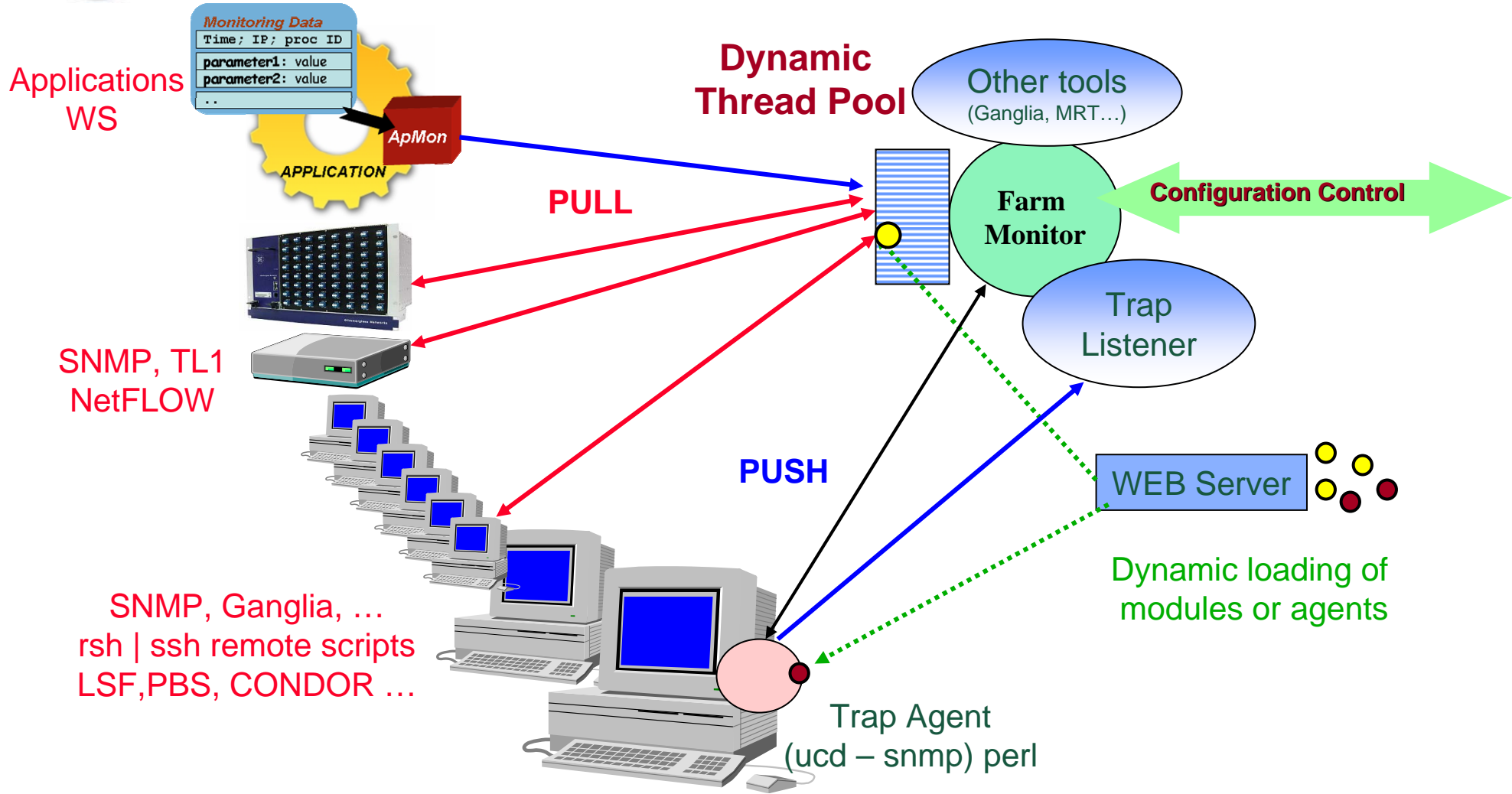
- ◆ MonALISA is able to dynamically **register** and **discover** all the other Services
- ◆ It is based on a **multi-threaded** services engine for global scalability
- ◆ The services are **self describing** and provide loadable proxies
- ◆ Automatic & secure code update
- ◆ Dynamic configuration for services. Secure Admin interface.
- ◆ **Active filter agents** to process the data and provided dedicated / customized information to other services or clients.
- ◆ **Mobile Agents** for decision support and global optimization.
- ◆ **Dynamic proxies** and WSDL & WAP pages for services.



Fully Distributed System with no Single Point of Failure

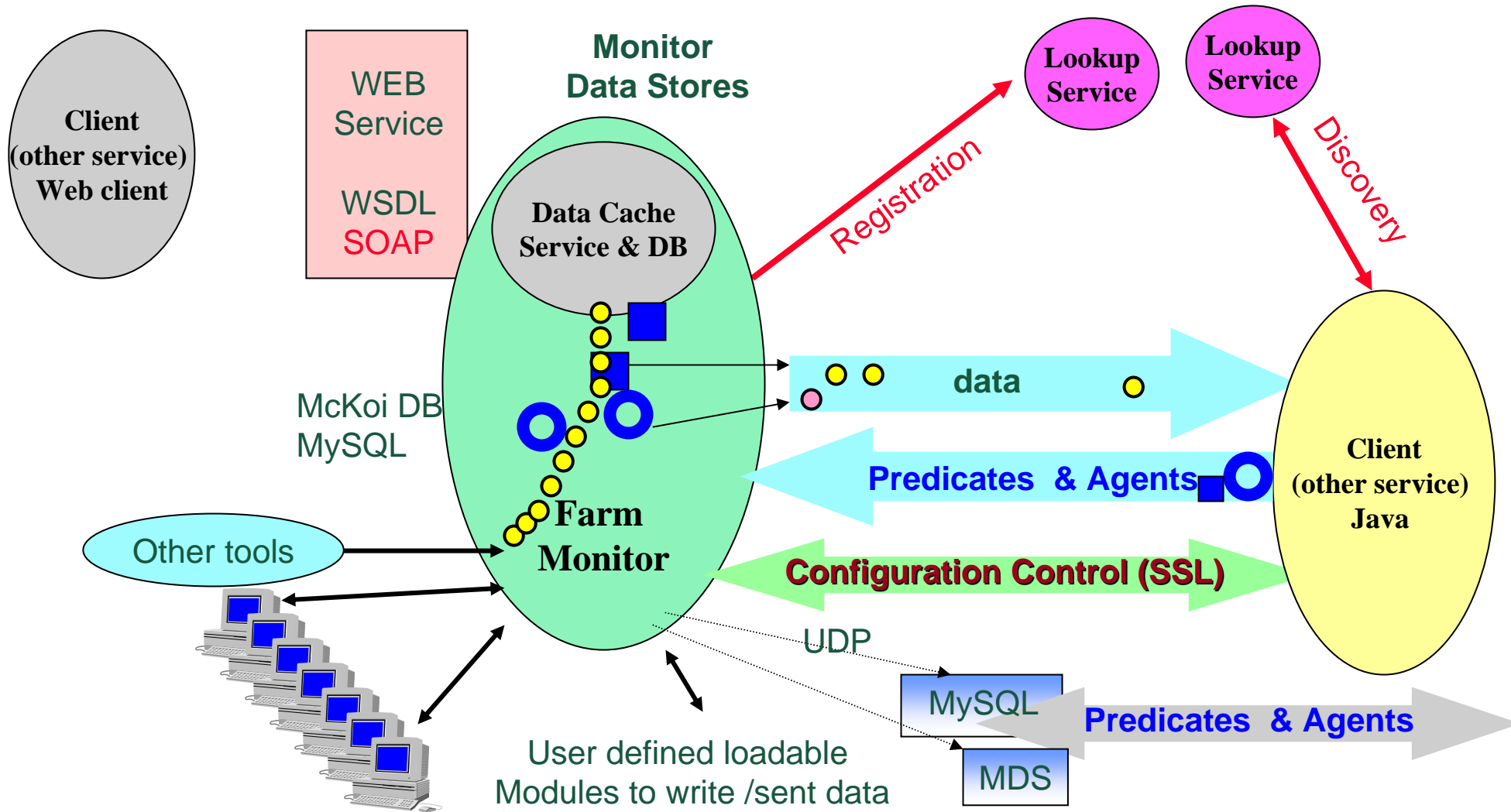


Monitoring: Data Collection



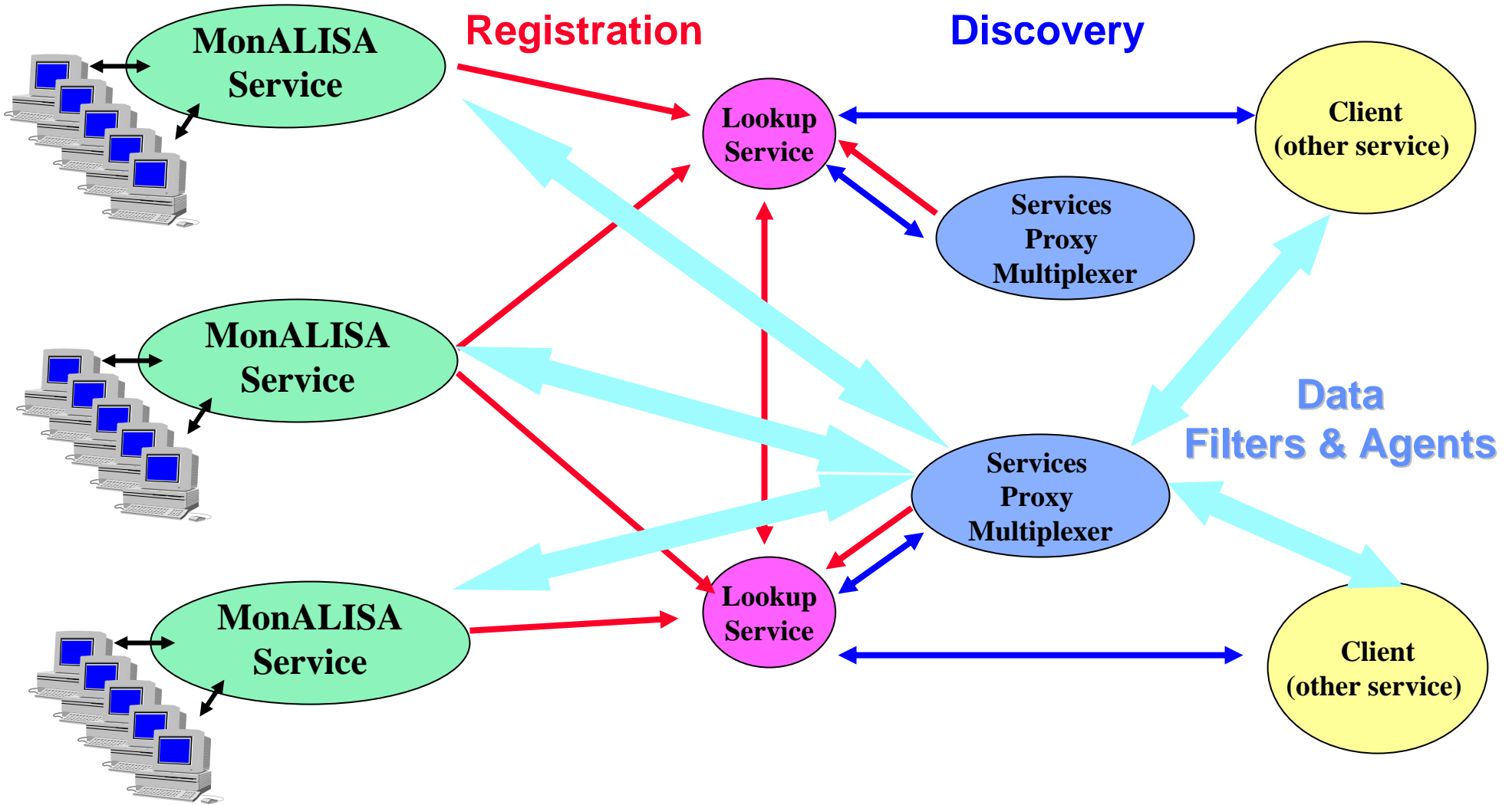


Service Monitor UNIT & Data Handling



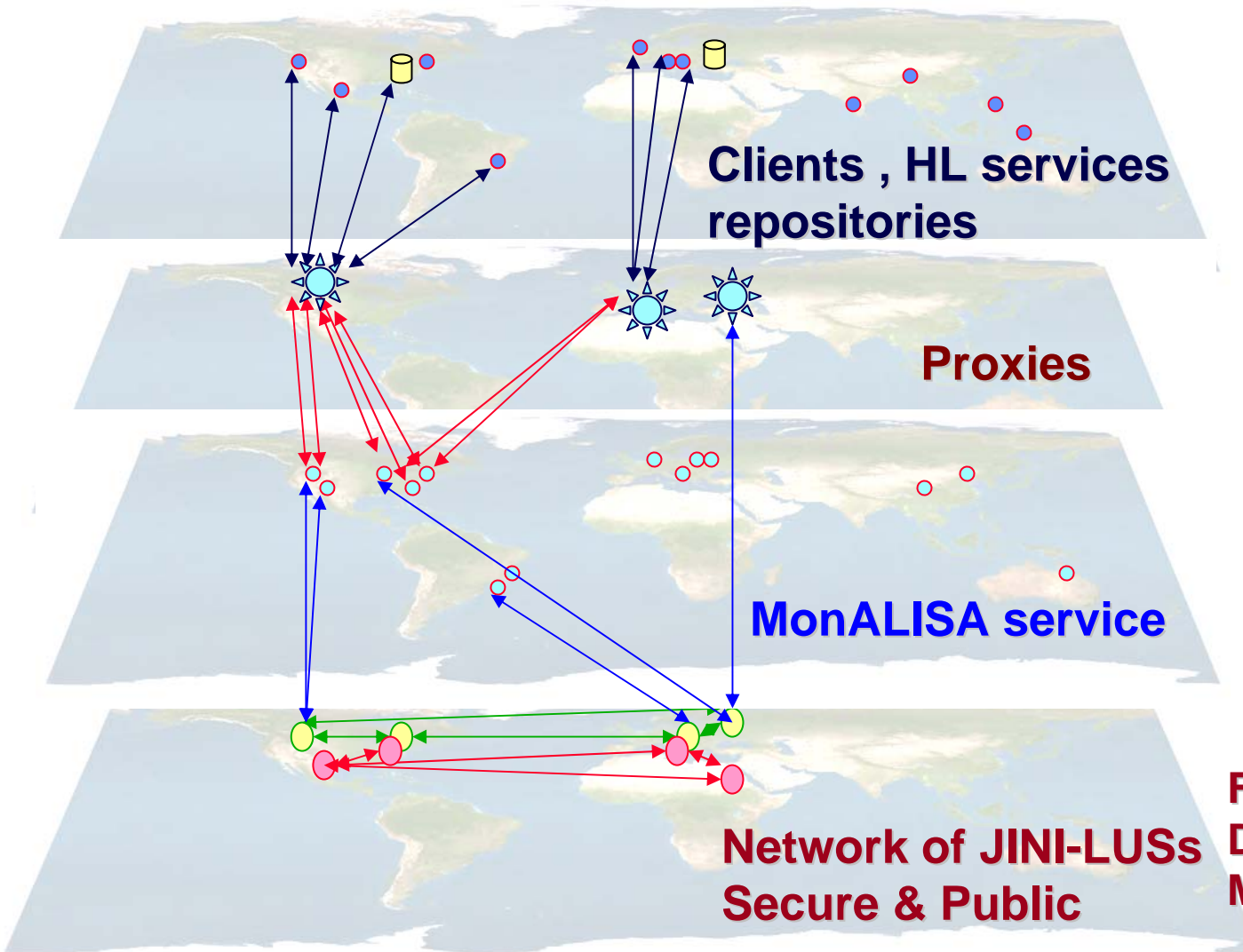


Registration / Discovery / Remote Notification





MonALISA Discovery System & Services



Global Services or Clients

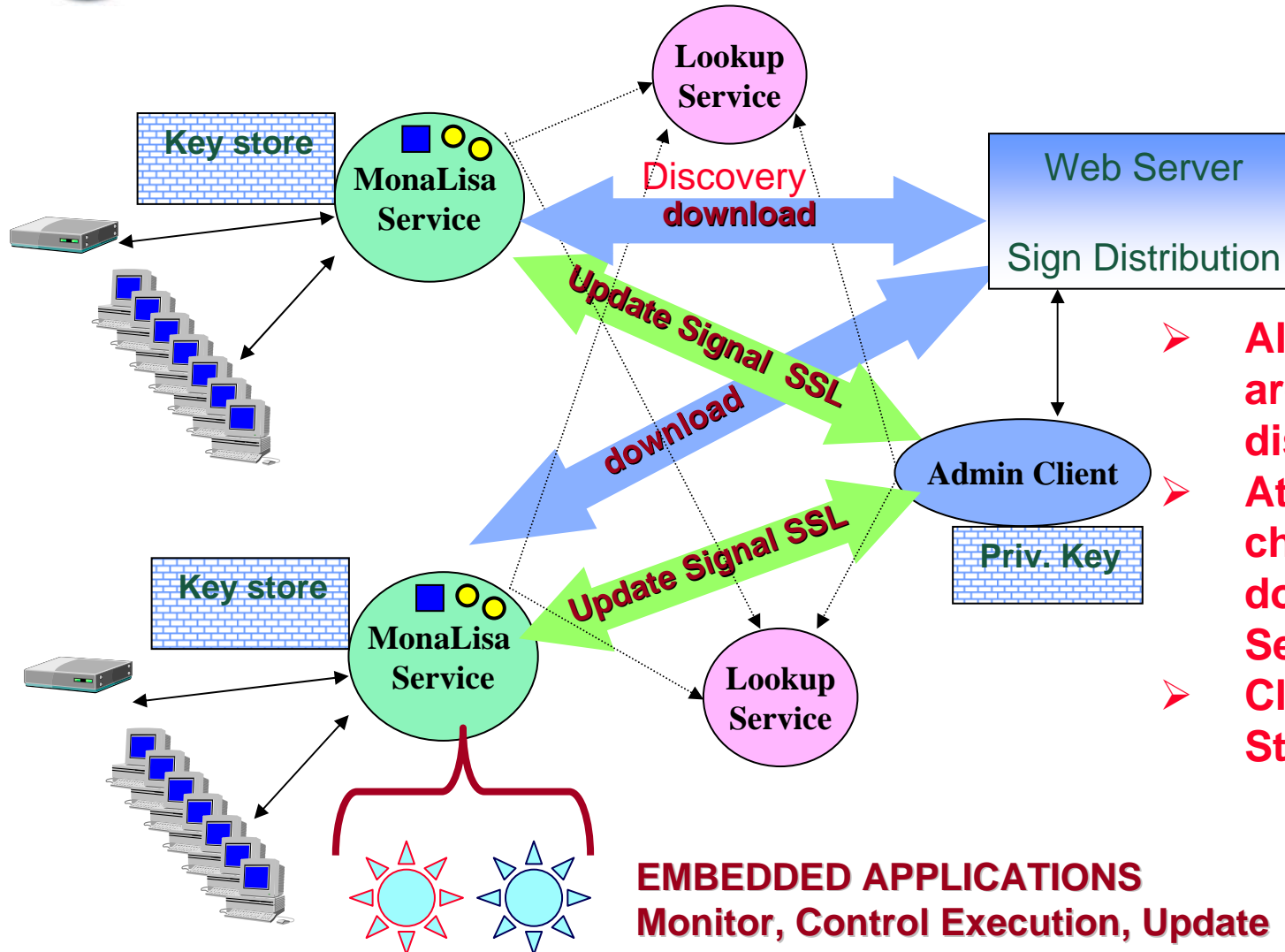
**Dynamic load balancing
Scalability & Replication
Security**

Distributed Information System.

**Fully Distributed Discovery
Dynamic - based on a lease
Mechanism and REN**



Secure – Automatic Update Mechanism for Services, Clients & Embedded Applications

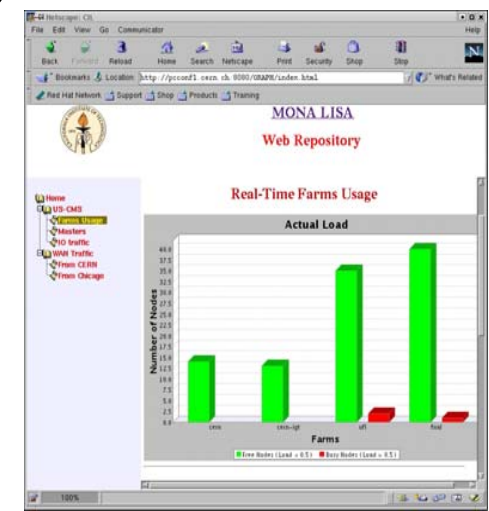
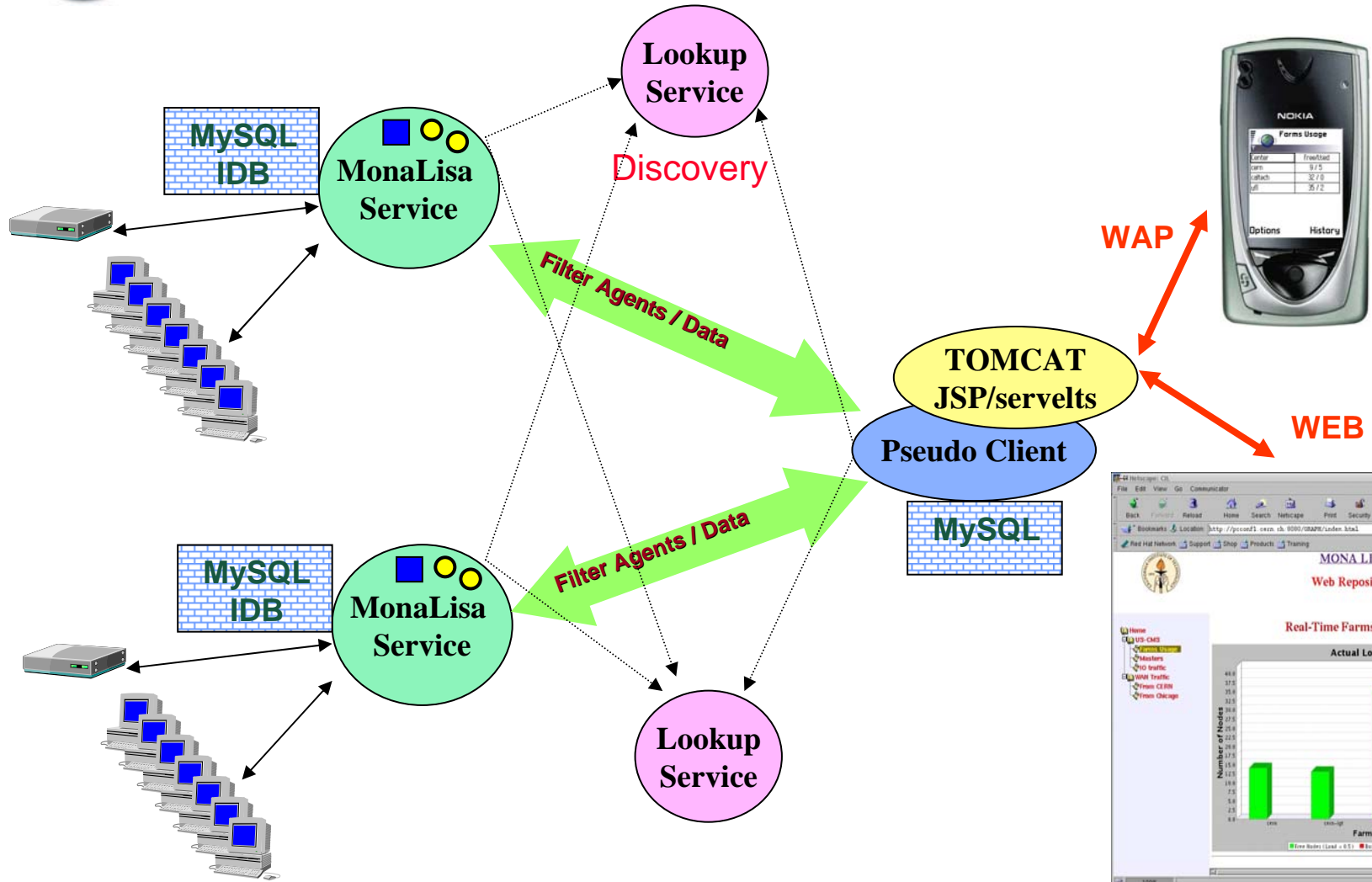


- All running services are update using the discovery mechanism
- At startup each service check if it an update is done at a set of Web Servers
- Clients use the Web Start mechanism

EMBEDDED APPLICATIONS
Monitor, Control Execution, Update



Pseudo – Clients & Dedicated Repositories





Data Collection and Interfacing with Other Tools



MonALISA is interfaced with many monitoring tools and is capable to collect information from different applications:

Computing Nodes / Farms (system information , network traffic...)

→ **SNMP, Ganglia, dedicated scripts**

Routers , Switches

→ **SNMP, MRTG, WS, NetFlow**

End to End Network performance

→ **IPERF, Pipes, Abing, ABping ...**

Batch Queuing Systems

→ **LSF, PBS, Condor, NQS, Grid Job Manager**

Applications

→ **VRVS, ORCA, GridFTP, TMDB, Clarens, Apache, RRD ...**



Communities using MonALISA



❖ Grid3

~20 sites in US and 1 Korea

❖ CMS-US sites

❖ CMS – DC04

We collected ~ 50 million monitoring records from the 6 T1

❖ CDF

❖ DO SAR

❖ ABILENE backbone

❖ GLORIAD

❖ STAR

❖ ALICE

❖ VRVS System

❖ RoEduNET backbone

❖ INTERNET2 PIPES

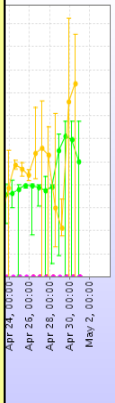


It has been used for Demonstrations at:

➤ SC2003

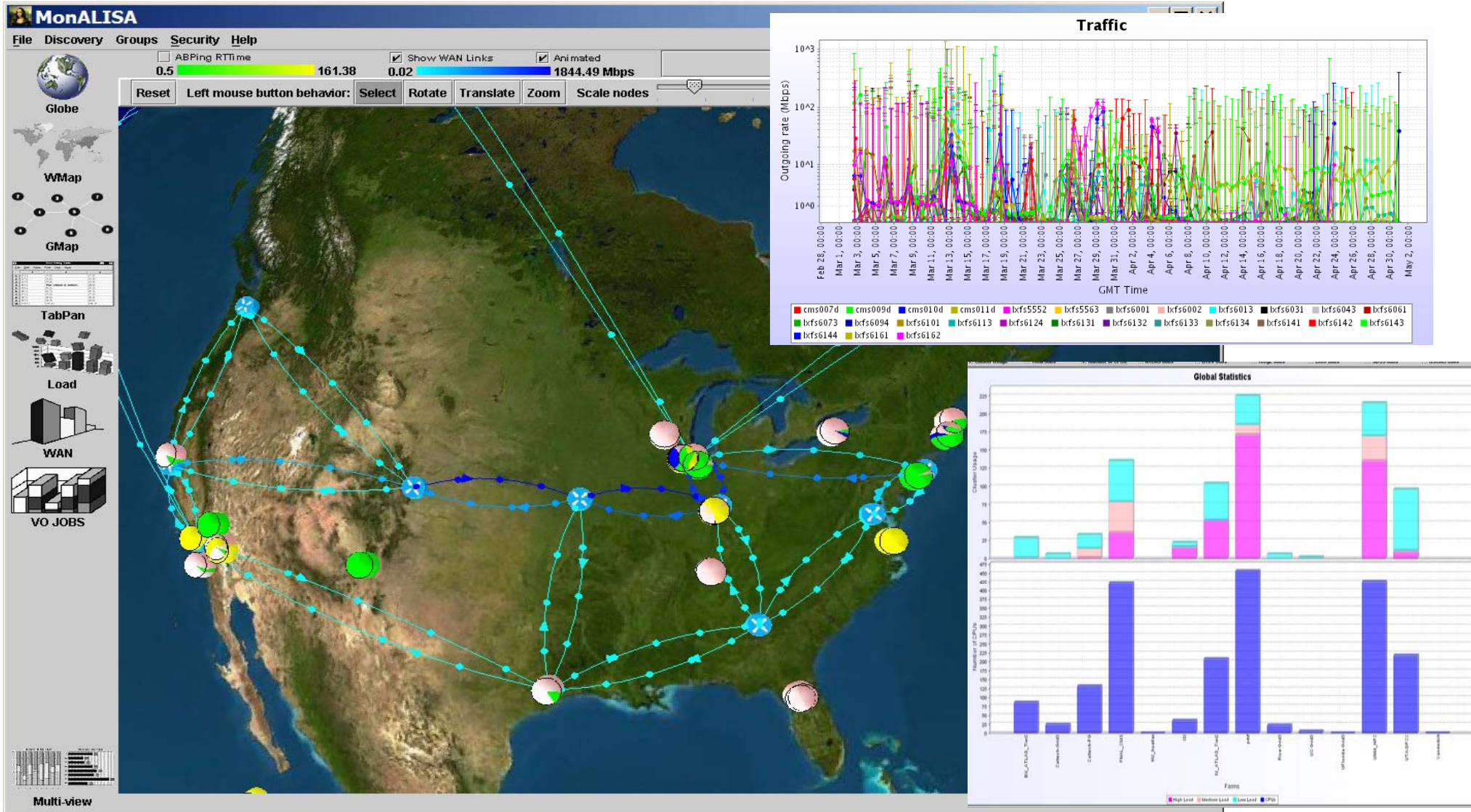
➤ Telecom 2003

➤ WSIS 2003



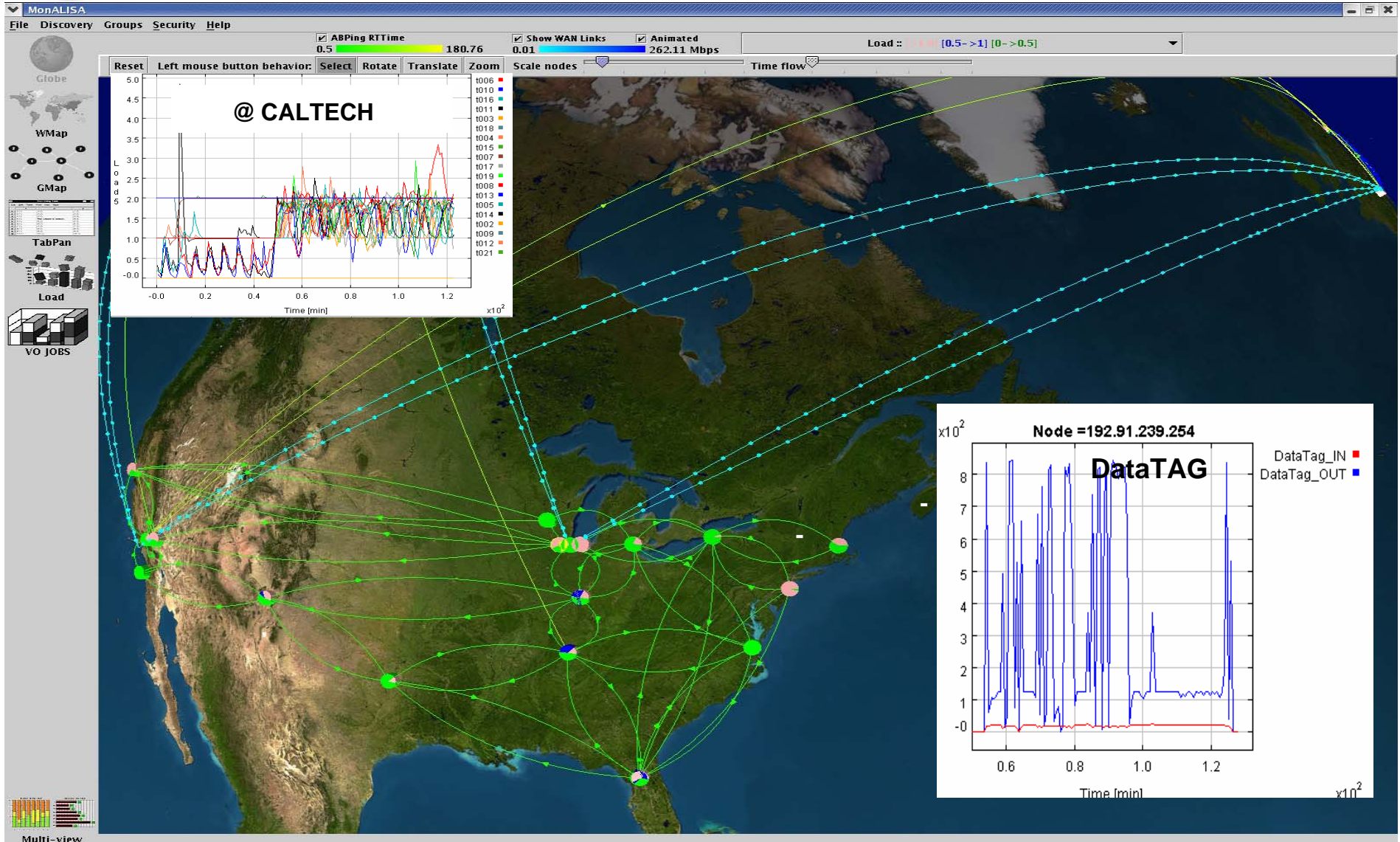


Monitoring I2 Network Traffic, Grid03 Farms and Jobs





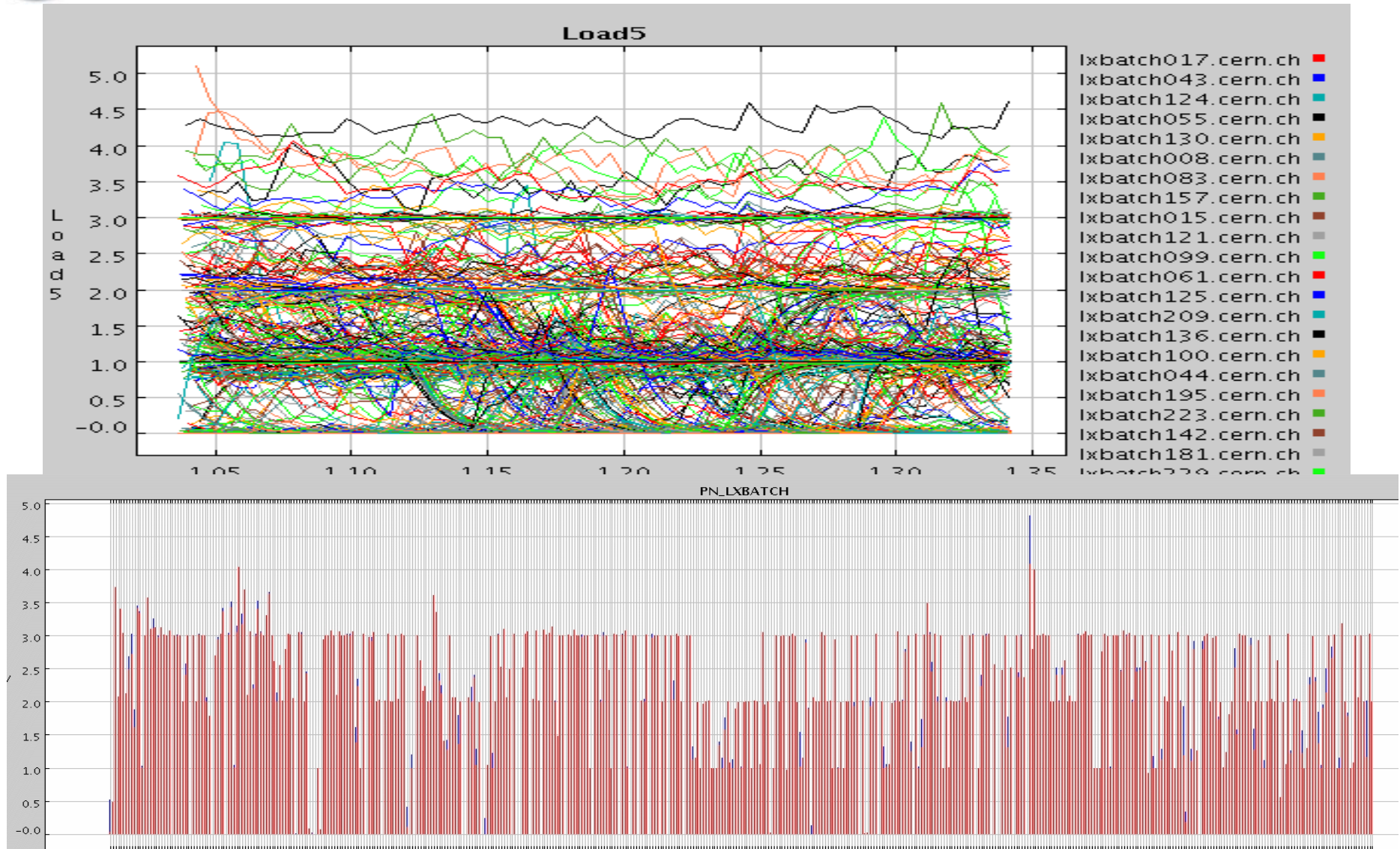
Global Client for HEP Grid Sites Network Traffic and Connectivity





Real-time Data for Large Systems

“lxshare” cluster at cern ~ 600 nodes



September 2004

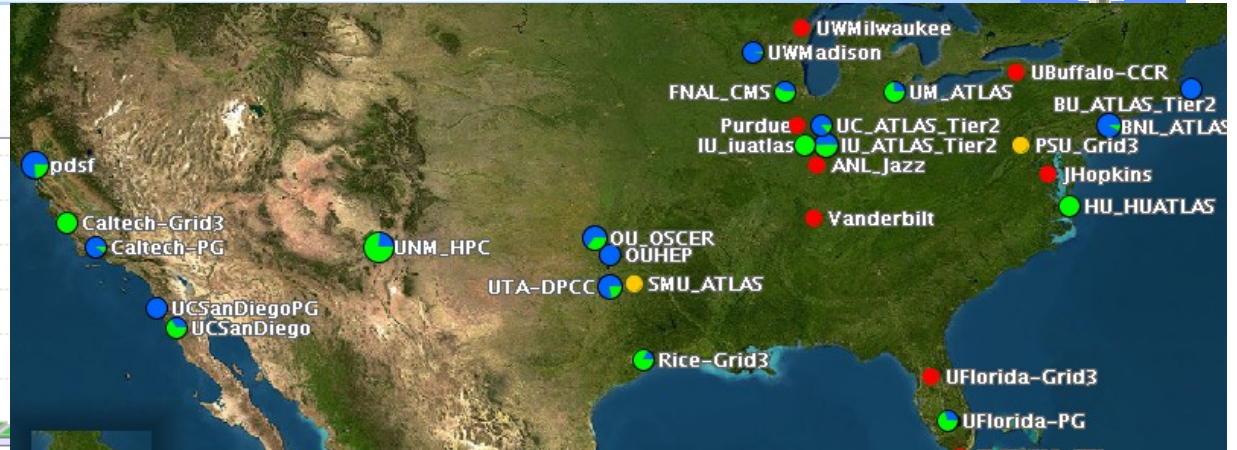
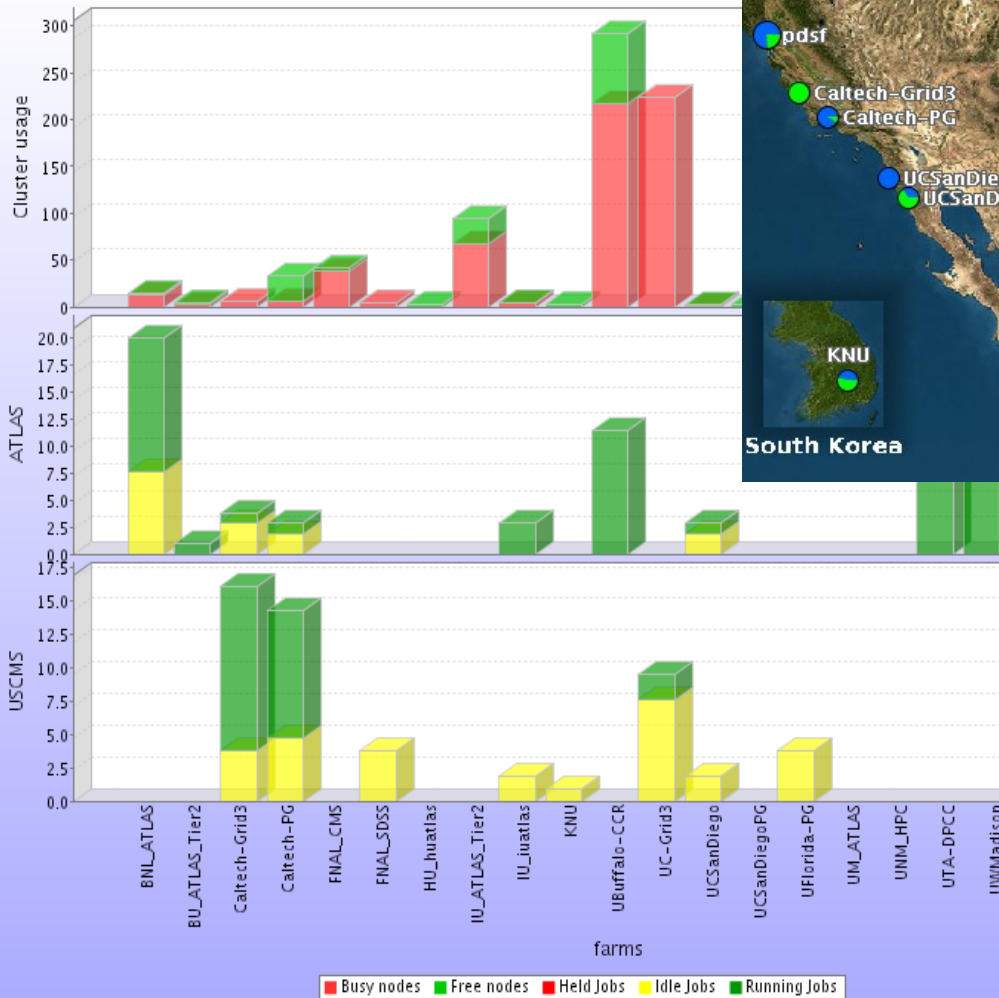
Iosif Legrand



MonALISA repositories Grid03



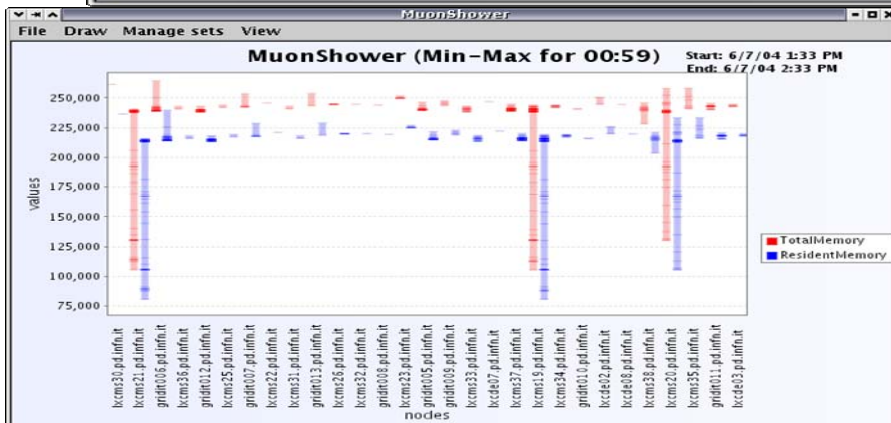
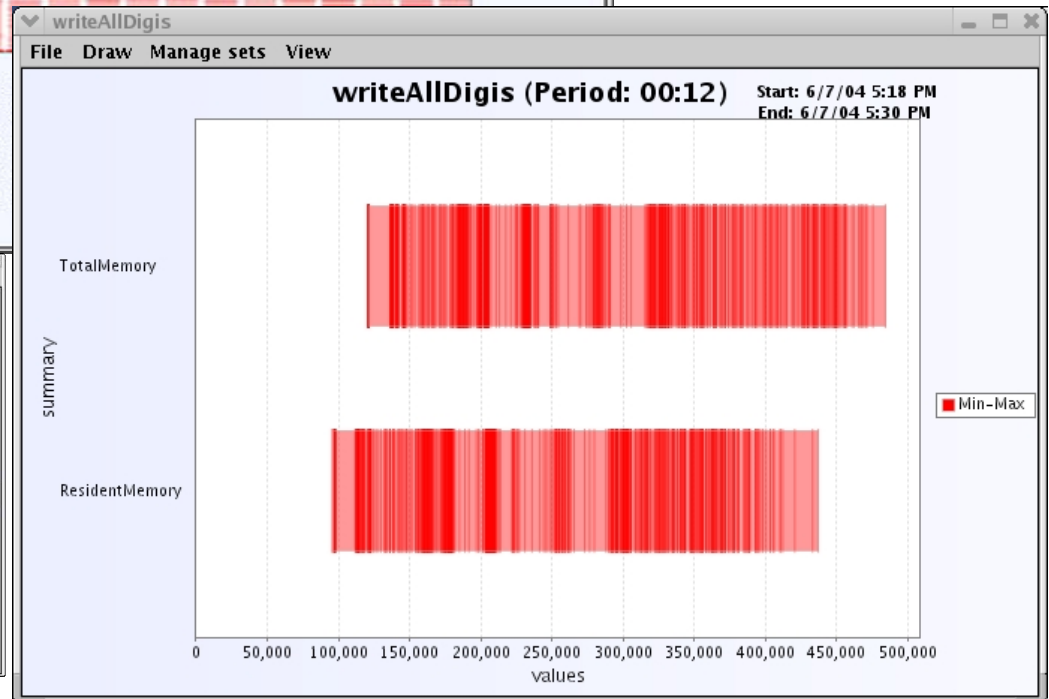
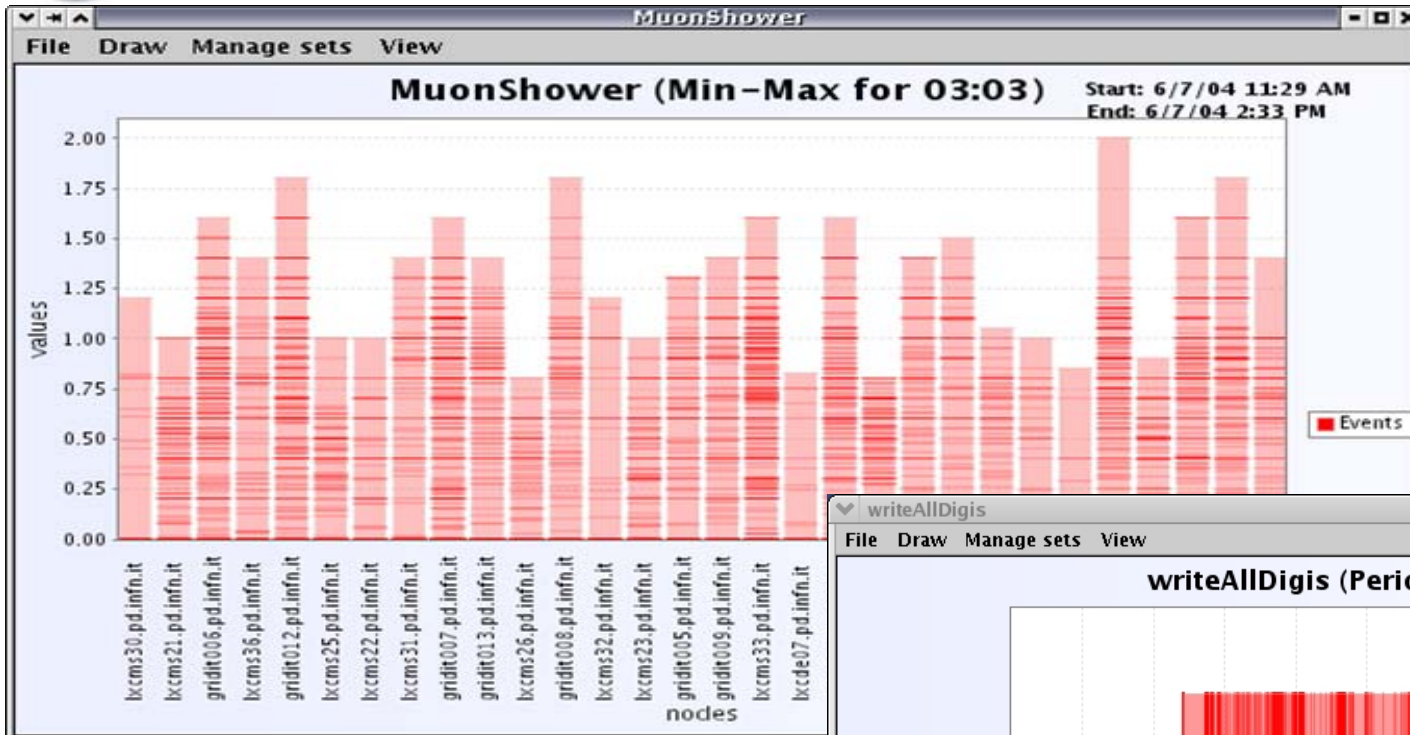
Global statistics



Farm	Load5		Current CPU usage (%)				Average CPU usage (%)				Farm info				
	Last	Average	user	sys	nice	idle	user	sys	nice	idle	No. CPUs	Busy (last)	Busy (avg)	Idle (last)	Idle (avg)
ASCC_Grid3_Testbed	0.03	0.029	1	1	0	98	2	1	0	98	4	0	0	2	2
BNL_ATLAS	3.88	2.671	18	18	0	64	13	6	0	81	172	79	79	7	7
BNL_ATLAS_BAK	0.23	0.762	16	4	0	80	15	4	1	80	172	79	79	7	7
BU_AGT_Tier2	0.72	0.915	1	3	0	96	1	3	0	96	62	1	1	15	15
BU_ATLAS_Tier2	2.74	3.019	8	9	0	83	7	6	0	87	114	29	29	0	0
Caltech-Grid3	0.12	0.143	2	1	0	97	2	1	0	98	24	0	0	6	6
Caltech-PG	0.53	0.228	2	1	0	97	2	1	0	97	132	30	30	3	3
FNAL_CMS	3.31	4.616	10	11	0	79	12	12	0	76	98	31	33	43	41
FNAL_CMS2	0.13	0.316	1	1	0	98	3	2	0	95	112	43	43	15	15
HU_HUATLAS	0.22	0.163	4	1	0	95	4	1	0	95	2	0	0	2	2
IU_ATLAS_Tier2	8.34	9.434	22	31	2	45	18	32	2	49	208	53	53	51	51
IU_uatlas	0.09	0.688	4	1	0	95	13	3	0	84	1	0	1	1	0
KNU	-	-	49	51	0	0	38	61	0	0	2	8	8	10	10
OUHEP	2.23	3.214	17	7	76	0	12	8	77	4	19	14	14	0	0
OU_OSCER	1.8	1.908	0	1	1	98	1	3	3	93	272	89	114	47	22
PSU_Grid3	6.4	9.263	35	12	0	33	44	19	0	37	-	-	-	-	-
Purdue-Physics	5.65	5.79	-	-	-	-	-	-	-	-	43	13	13	10	10
Rice-Grid3	4.21	3.472	13	12	0	75	7	14	0	80	22	1	1	5	5
SLACdev	0.02	0.198	3	1	0	96	3	2	0	96	2	0	0	1	1
SMU_ATLAS	3.99	4.568	13	7	80	0	9	3	38	31	-	-	-	-	-
UCSanDiego	0.04	0.169	4	1	0	95	7	1	0	92	3	1	1	2	2
UCSanDiegoPG	0.72	0.776	3	9	0	87	4	8	0	88	84	21	21	0	0
UC_ATLAS_Tier2	1.01	3.472	6	5	0	89	7	7	0	86	80	32	32	5	5
UFlorida-PG	42.05	41.82	31	5	0	64	27	4	0	69	82	12	12	29	29
UM_ATLAS	4.03	2.602	20	35	0	45	11	29	0	60	33	3	4	9	8



Monitoring CMS applications : Summary Plots for Probability Density for Several Parameters

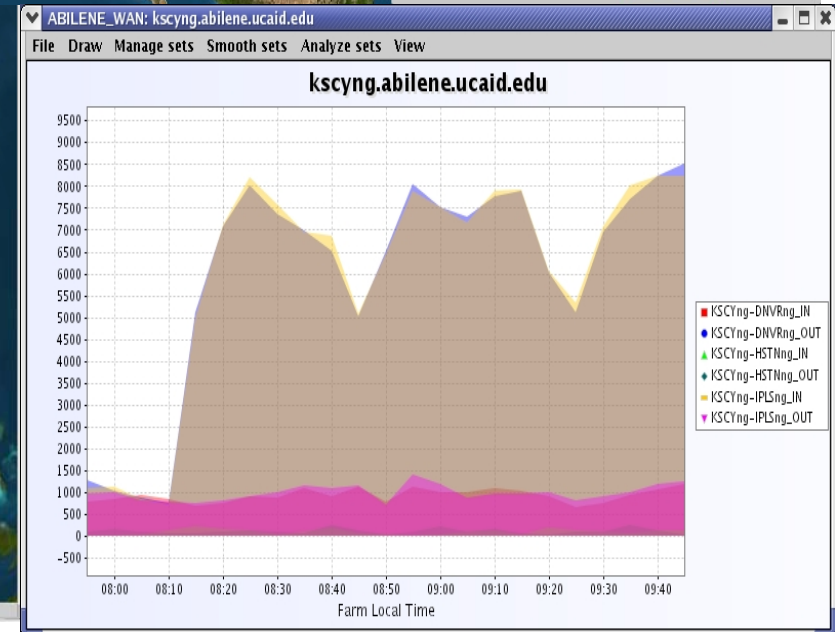
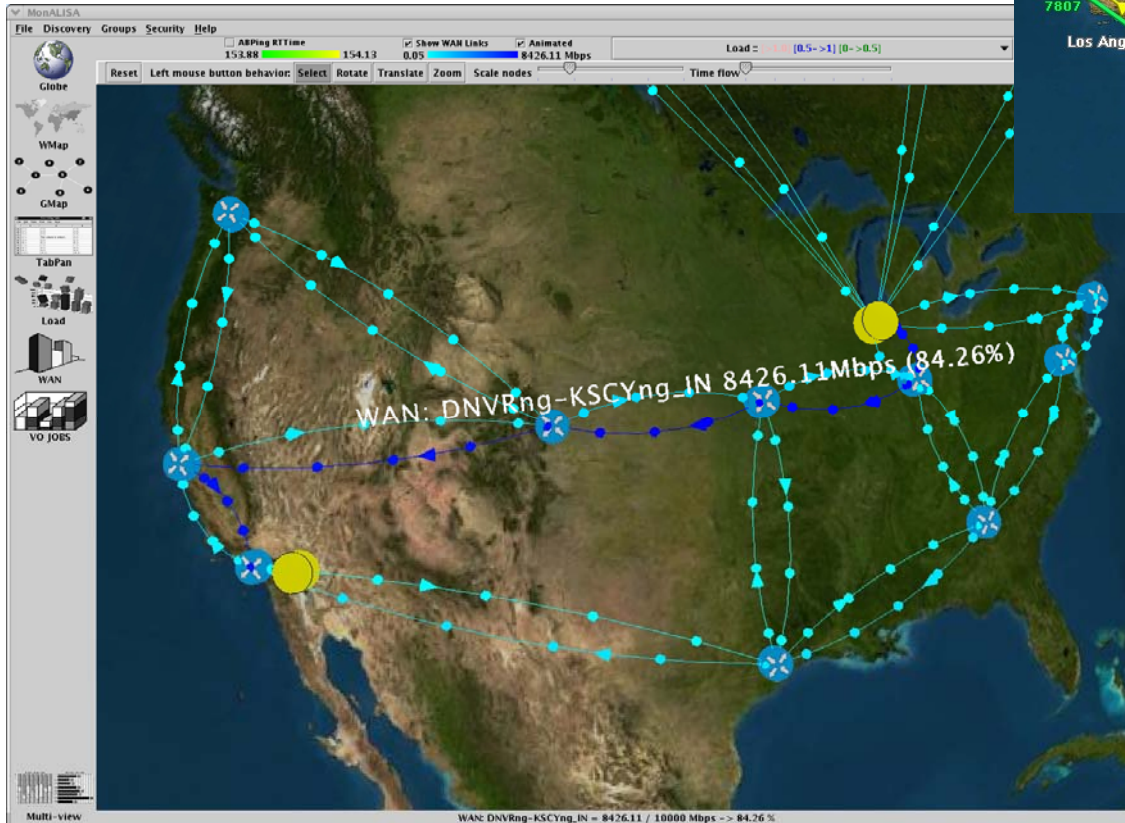
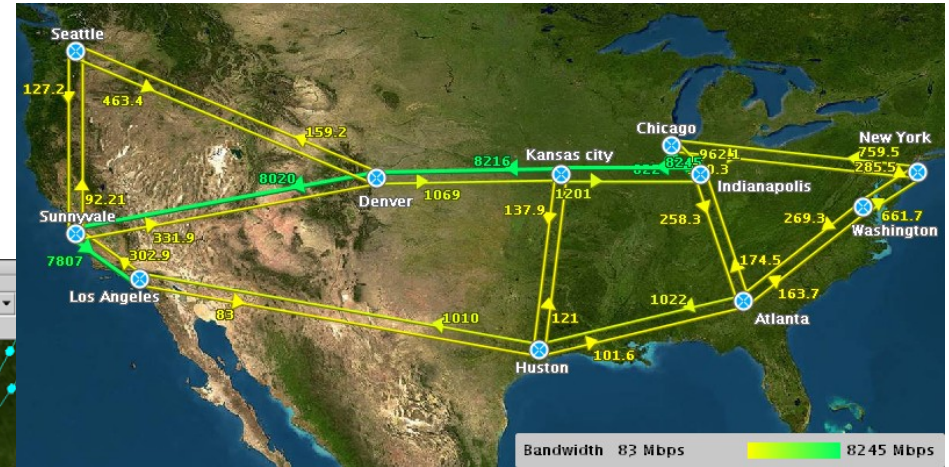




Monitoring ABILENE backbone Network

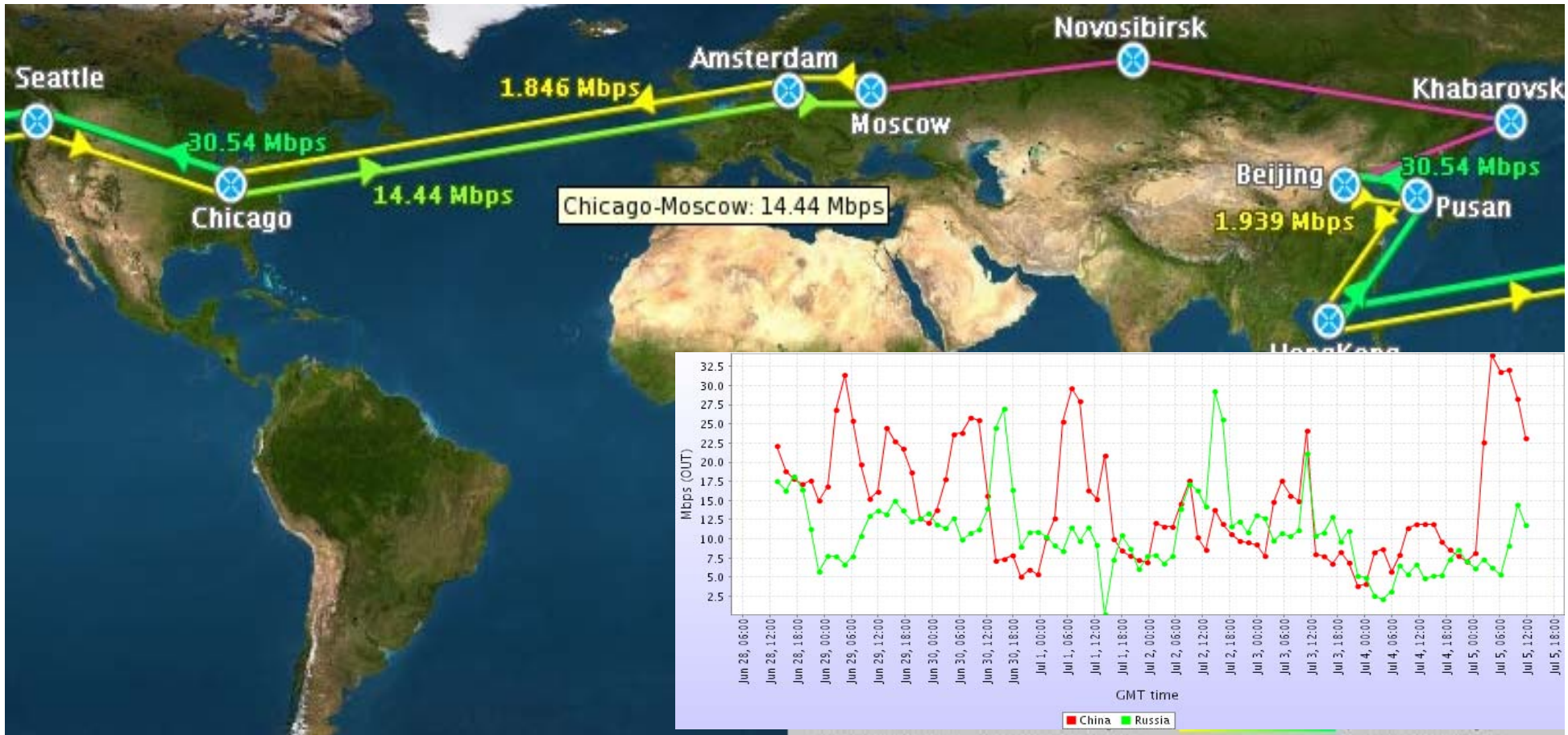


- ◆ Test for a Land Speed Record
- ◆ ~ 7 Gb/s in a single TCP stream from Geneva to Caltech





The GLORAI Network





VRVS traffic and connectivity

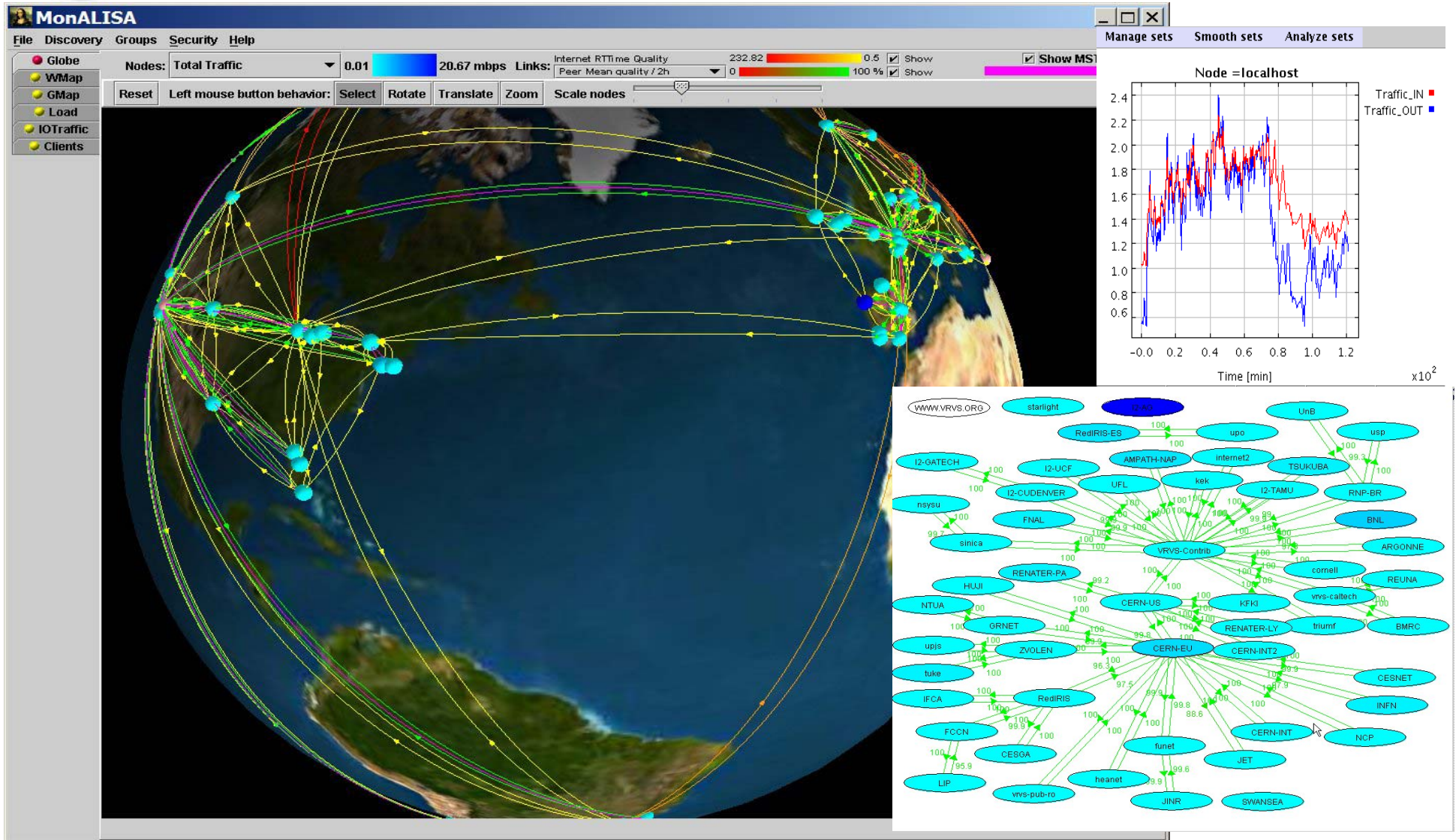
Real-time and historical data



The screenshot displays the MonALISA software interface. The main window, titled "MonALISA", features a menu bar with "File", "Discovery", "Groups", "Security", and "Help". On the left, a sidebar contains icons for "Globe", "WMap", "GMap", "Load", "IOTraffic", and "Clients". The central area shows a 3D globe with several nodes (represented by colored spheres) and connecting lines, illustrating network connectivity. The top status bar displays "Nodes: No. of video clients" with a value of 0 and a color gradient from blue to yellow. It also shows "Links: Internet RTTime Quality" with a value of 160.33 and a color gradient from red to yellow, and "Peer Mean quality / 2h" with a value of 0 and a color gradient from red to green. A "Scale nodes" slider is visible below the status bar. A secondary window titled "vrvs-us@vrvs-us.cern.ch:9002" is open, showing a file browser with a tree structure: "RC" containing "vrvs-us" containing "Reflector" containing "vrvs-us.cern.ch". A third window titled "Node = vrvs-us.cern.ch" displays a line graph of CPU usage over time. The graph has three data series: "Cpu_usr" (red), "Cpu_nice" (blue), and "Cpu_sys" (cyan). The x-axis is labeled "Time [min]" and ranges from -0.0 to 1.2, with a multiplier of $\times 10^2$. The y-axis ranges from -0.0 to 2.5. The "Cpu_usr" series shows significant fluctuations between 1.0 and 2.0. The "Cpu_nice" series is near 0.0. The "Cpu_sys" series fluctuates between 1.0 and 1.5. The interface also includes a "Parameters" panel on the right with a list of system metrics and a "Modules" panel with a list of active modules.



Monitoring VRVS Reflectors Communication Topology & Dynamic MST



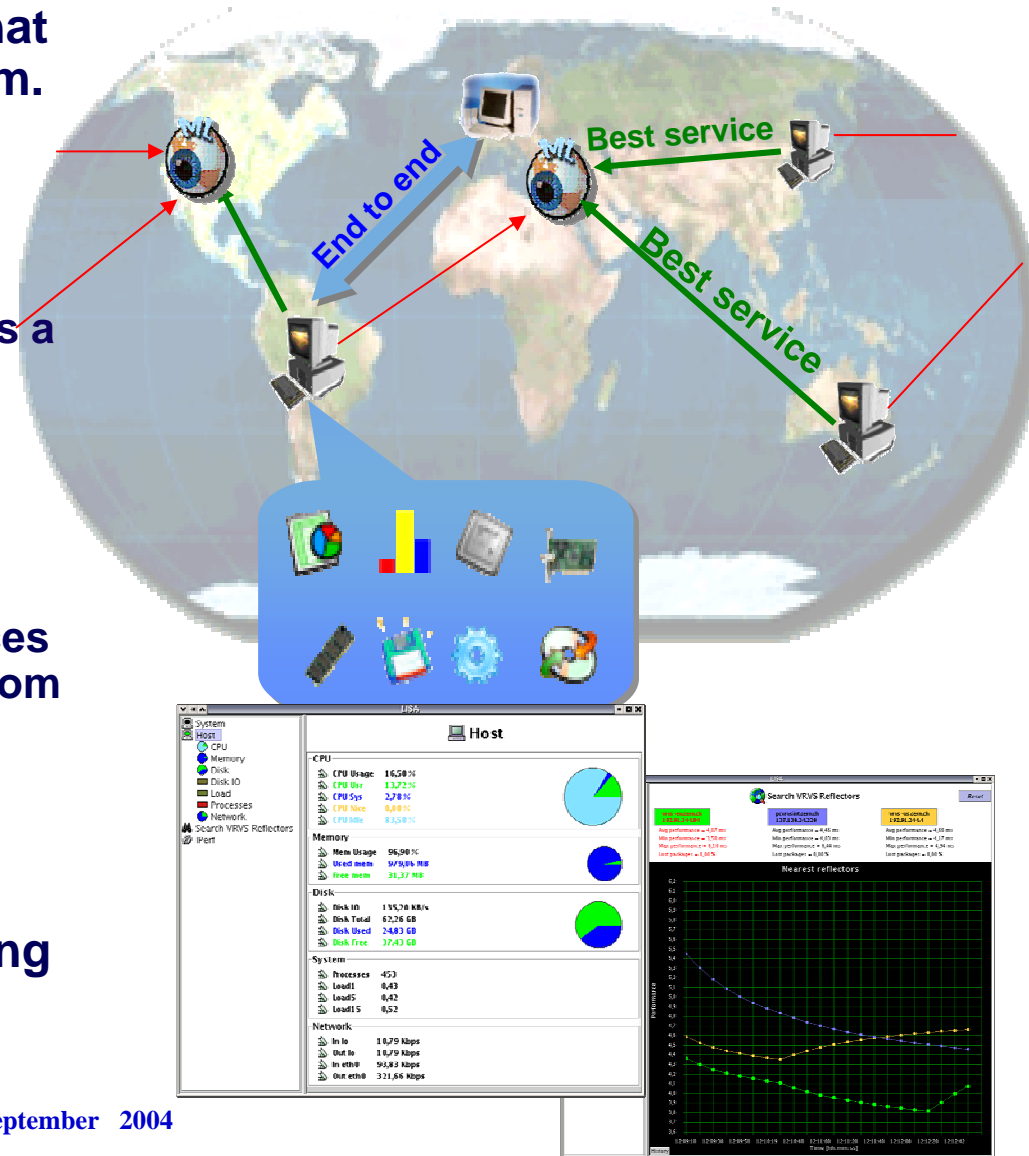


LISA- Localhost Information Service Agent

End To End monitoring & Efficient Integration

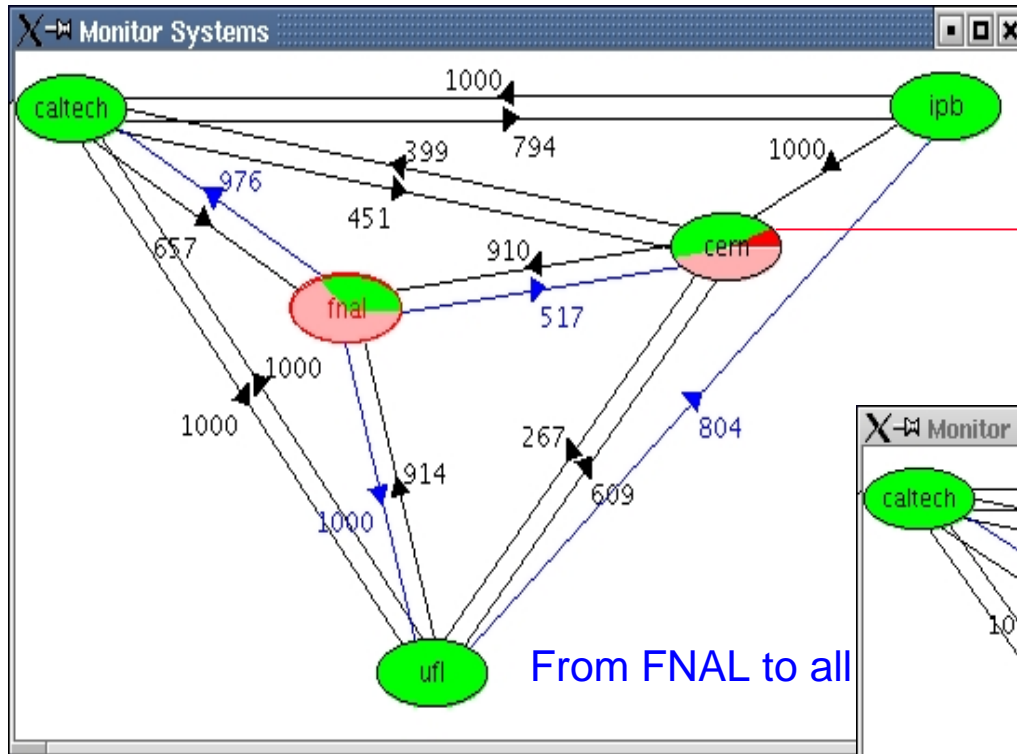


- ◆ A lightweight Java Web Start application that provides a complete monitoring of the system.
- ◆ It can be easily deployed on any system.
- ◆ Discovers MonALISA services
 - Based on information such as Network, AS number, continent, etc. it determines a list with the best possible services.
 - Registers as a listener for other service attributes (eg. number of connected clients).
 - Continuously monitors the network connection with several selected services and provides the best one to be used from the client's perspective.
 - Detects network quality faults, informs upper layer services and takes appropriate decisions
- ◆ A user friendly GUI to present the monitoring information.

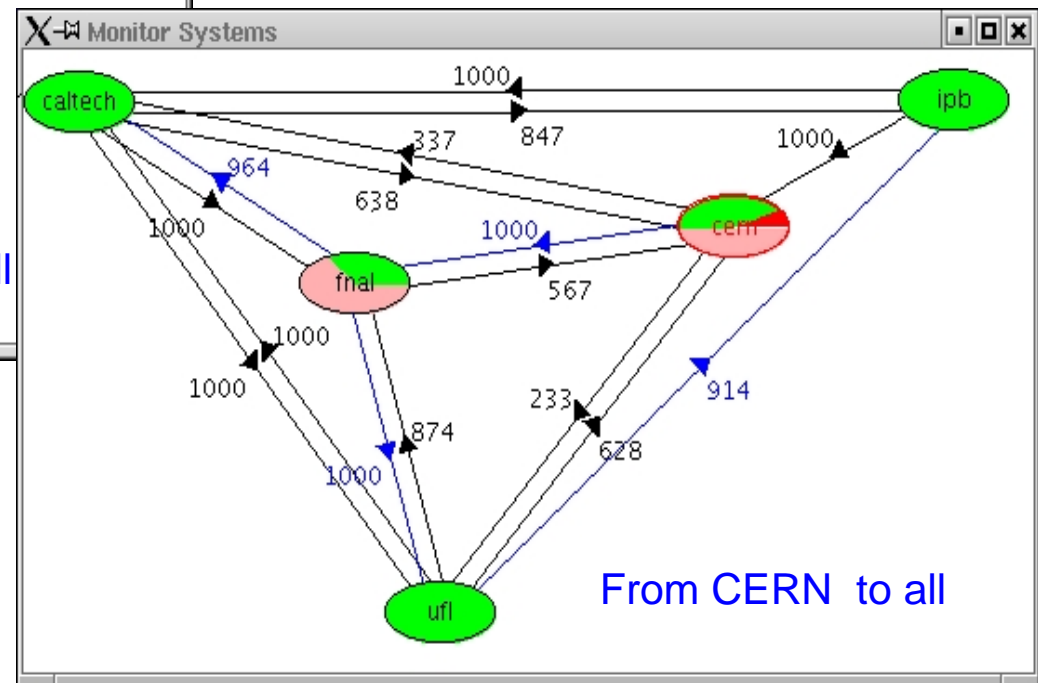




Mobile Agents for Global Optimization



Simple "Global Load" filter agent



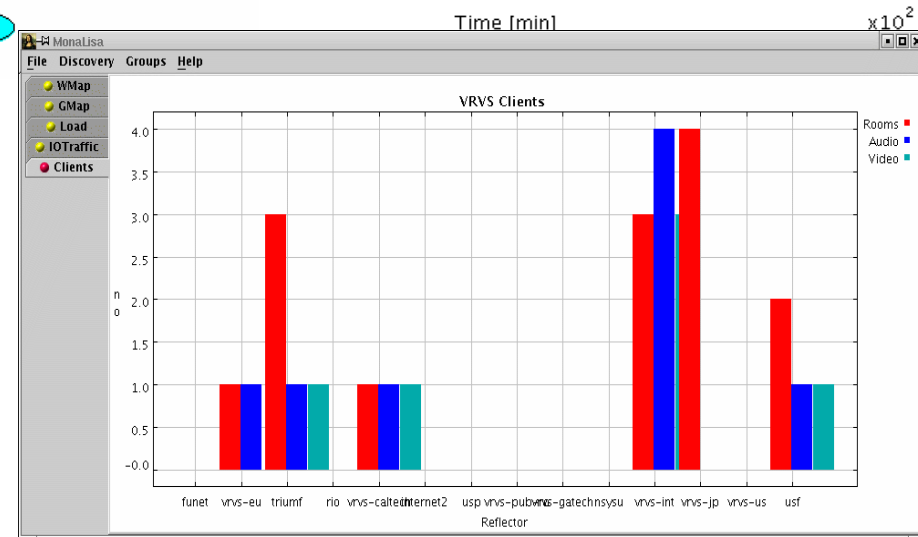
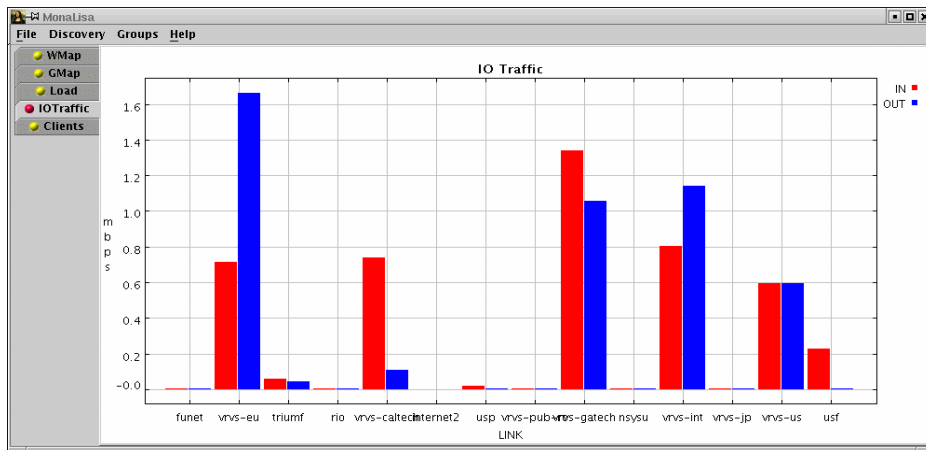
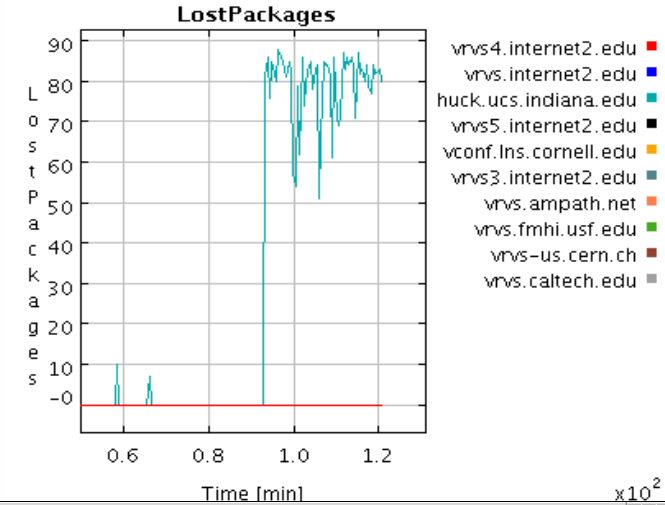
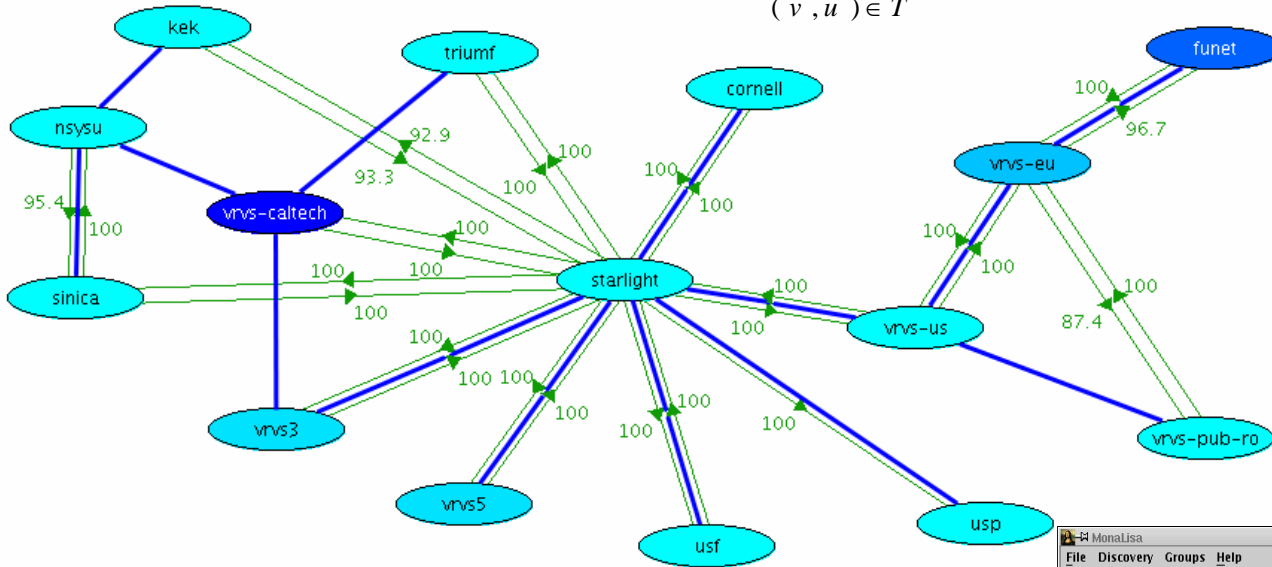
Maximum Flow Data Replication Path Agent Deployed to each RC and evaluates the best path for real-time data replication



Monitoring VRVS Reflectors ; Agents for Creating a Dynamic Minimum Spanning Tree

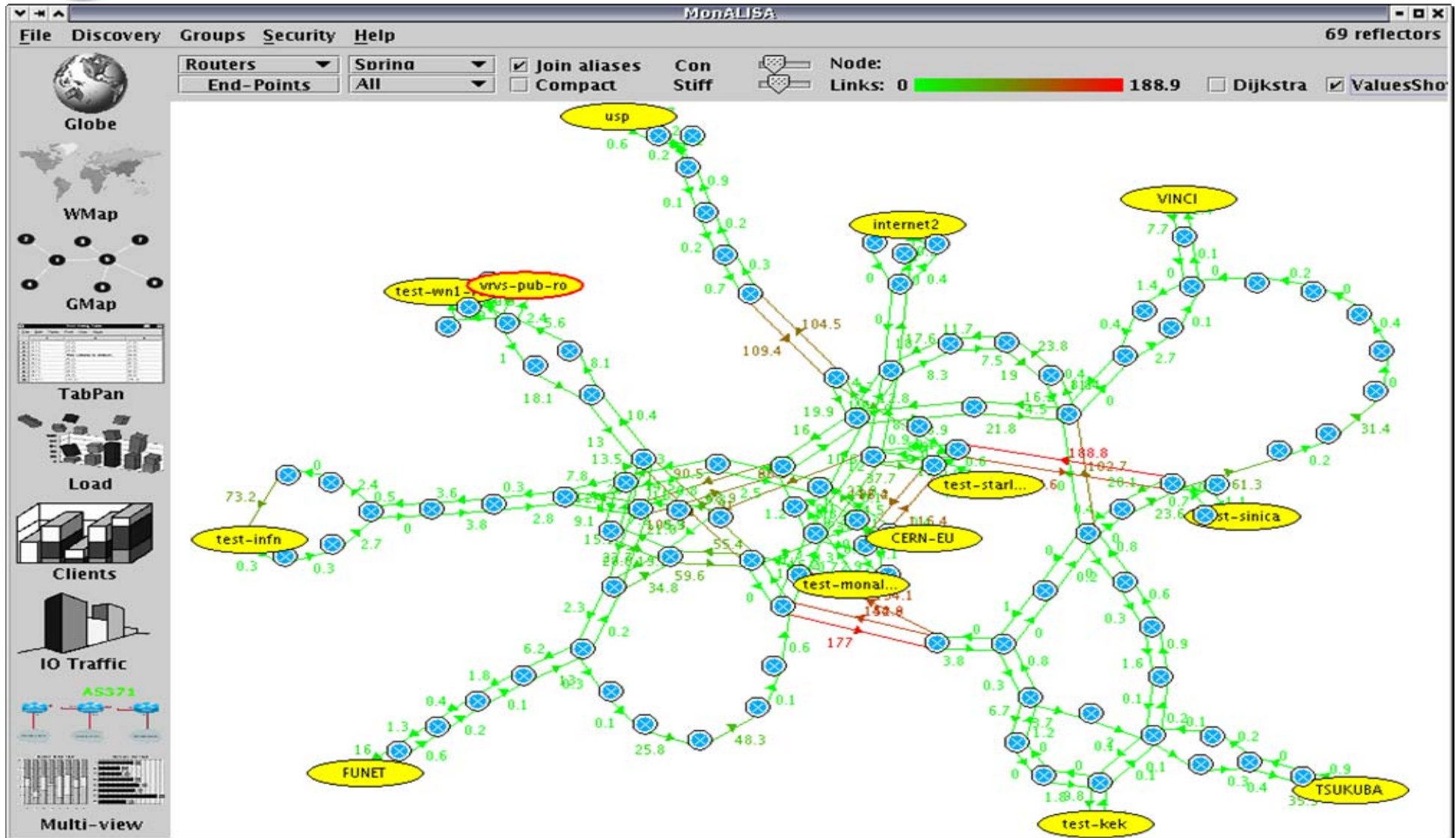


$$w(T) = \sum_{(v,u) \in T} w((v,u))$$





Monitoring Network Topology Latency, Routers

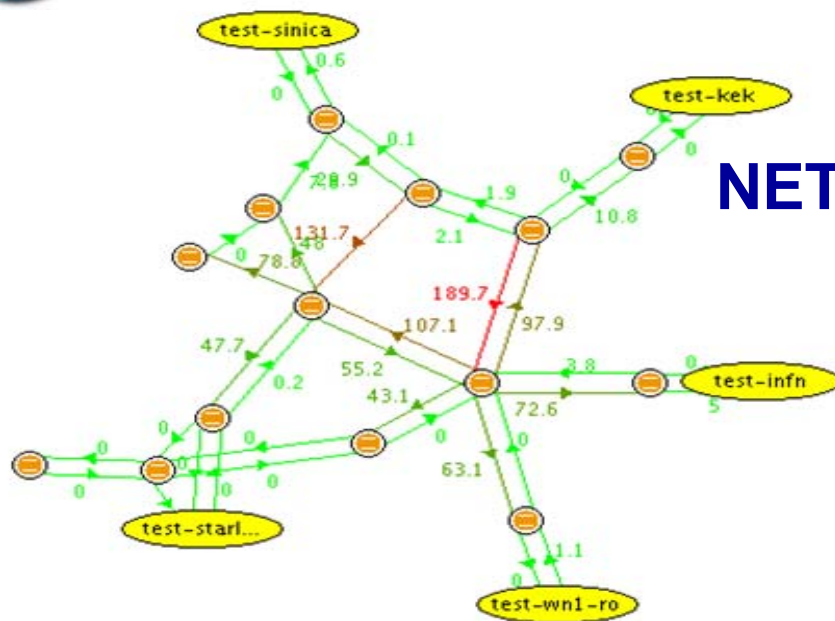




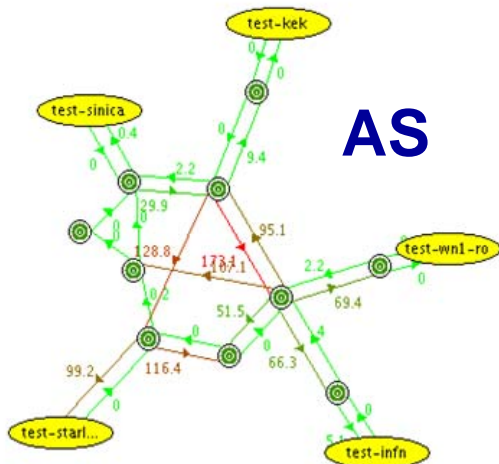
“Tomographic” reconstruction of the WAN Topology & Quality of the connectivity



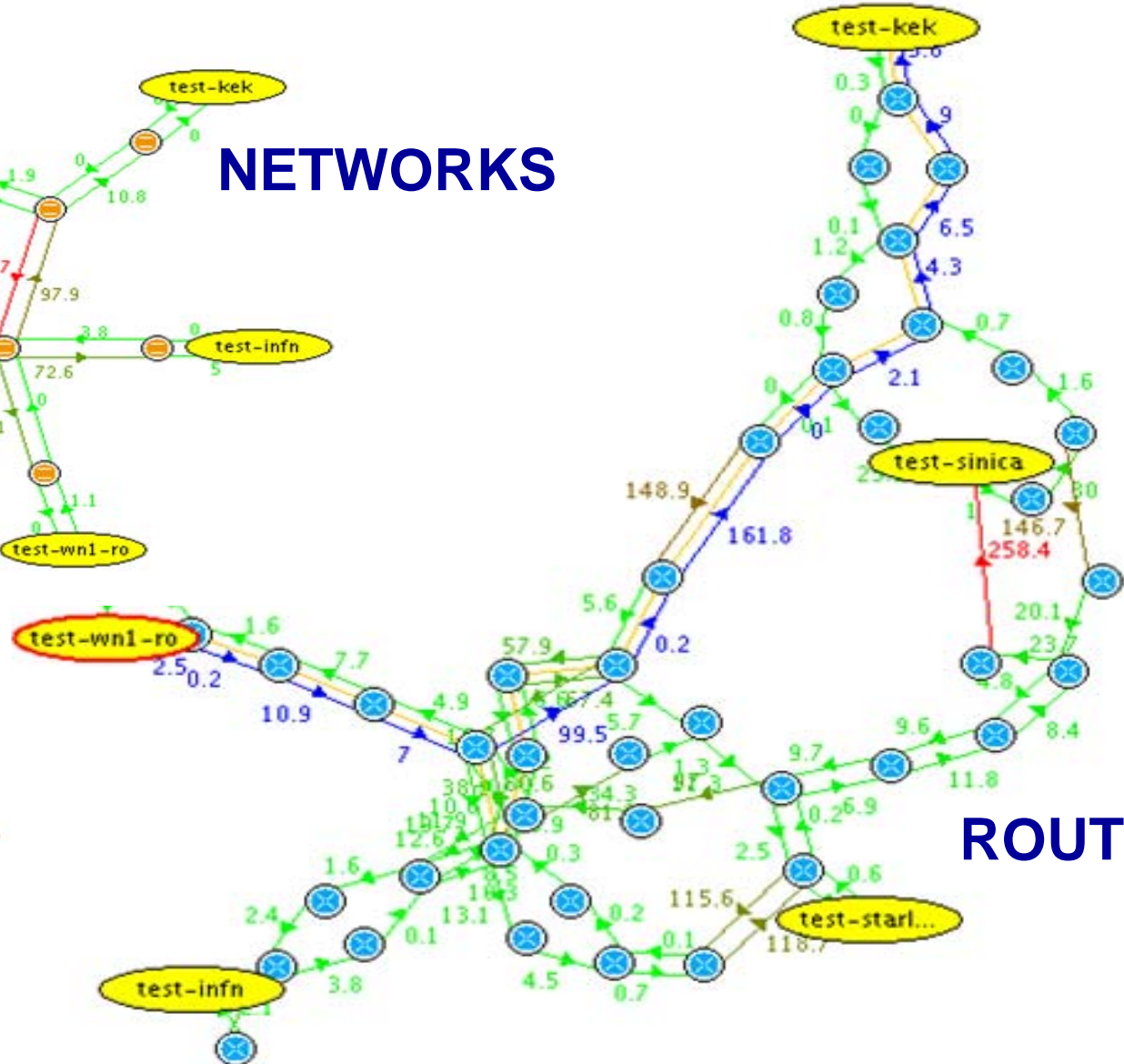
NETWORKS



AS

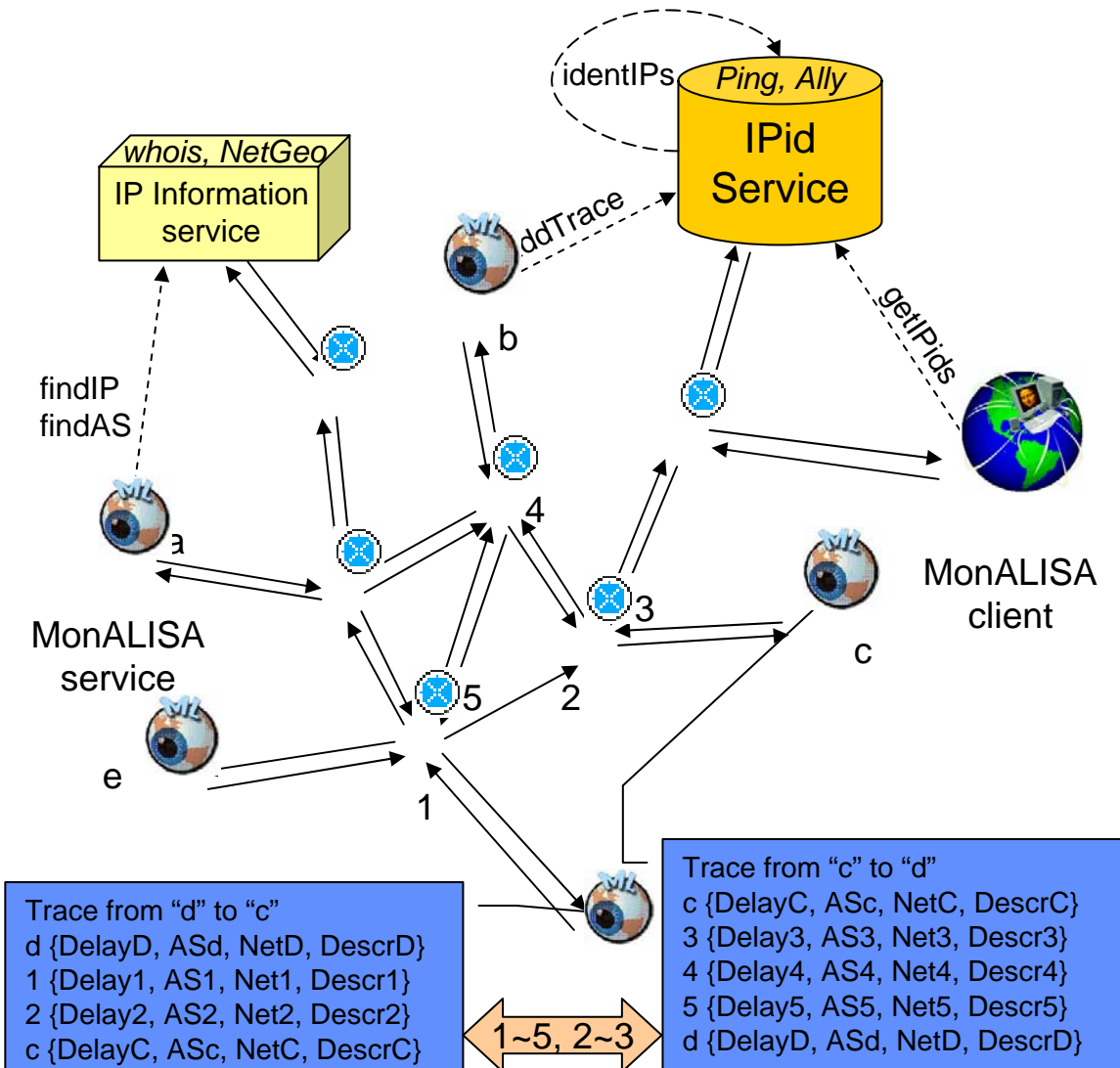


ROUTERS





Network Topology Service



MonALISA service

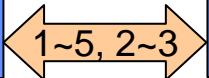
- Performs traces
- Gets relevant info for each hop
- Sends traces to the interested clients
- Sends new IPs to the IPid Service

MonALISA client

- Discover all services
- Get traces data
- Resolve IP aliases
- Display selected data
- Perform algorithms

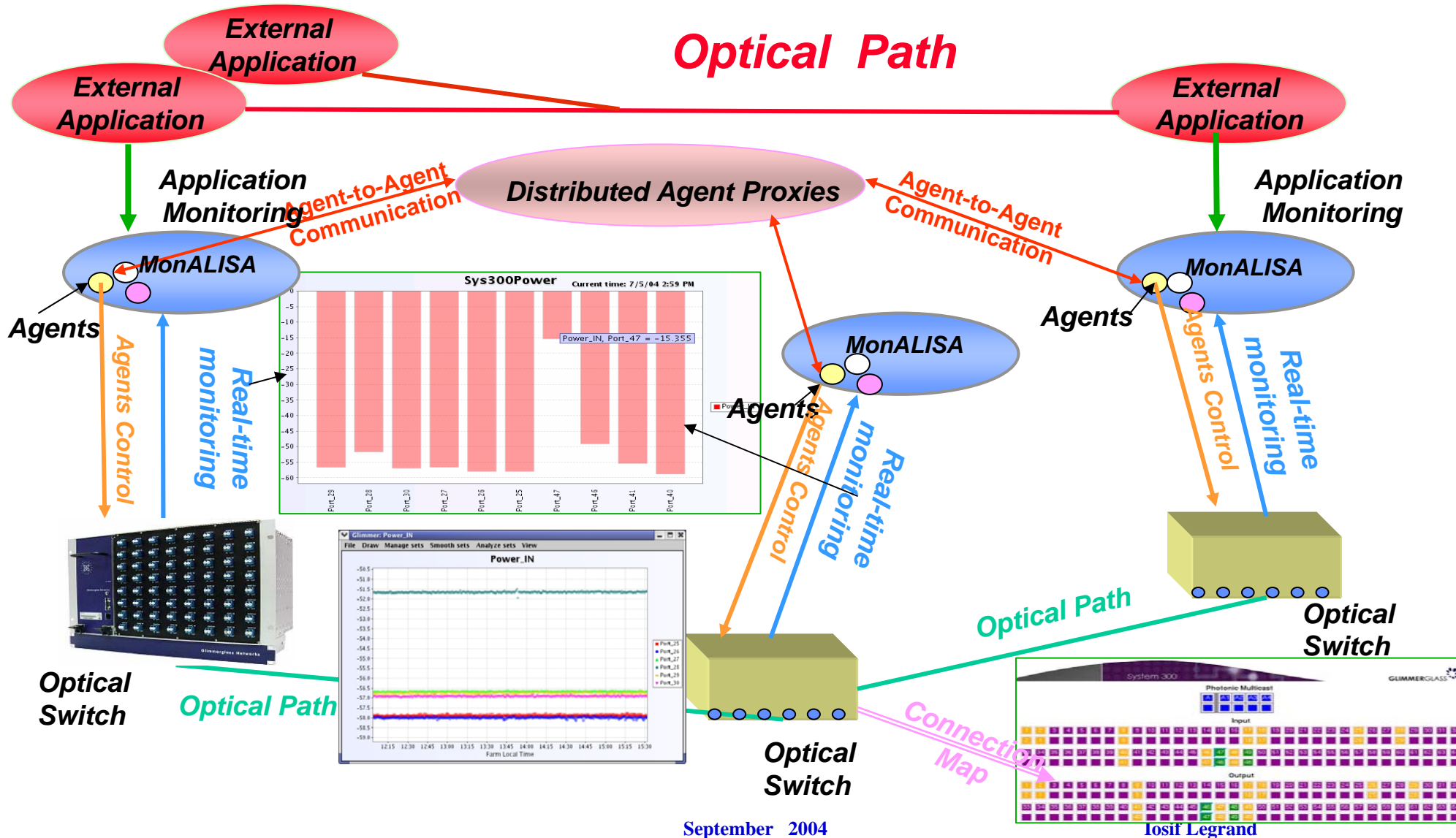
Trace from "d" to "c"
 d {DelayD, ASd, NetD, DescrD}
 1 {Delay1, AS1, Net1, Descr1}
 2 {Delay2, AS2, Net2, Descr2}
 c {DelayC, ASc, NetC, DescrC}

Trace from "c" to "d"
 c {DelayC, ASc, NetC, DescrC}
 3 {Delay3, AS3, Net3, Descr3}
 4 {Delay4, AS4, Net4, Descr4}
 5 {Delay5, AS5, Net5, Descr5}
 d {DelayD, ASd, NetD, DescrD}





Monitoring And Controlling Optical Switches

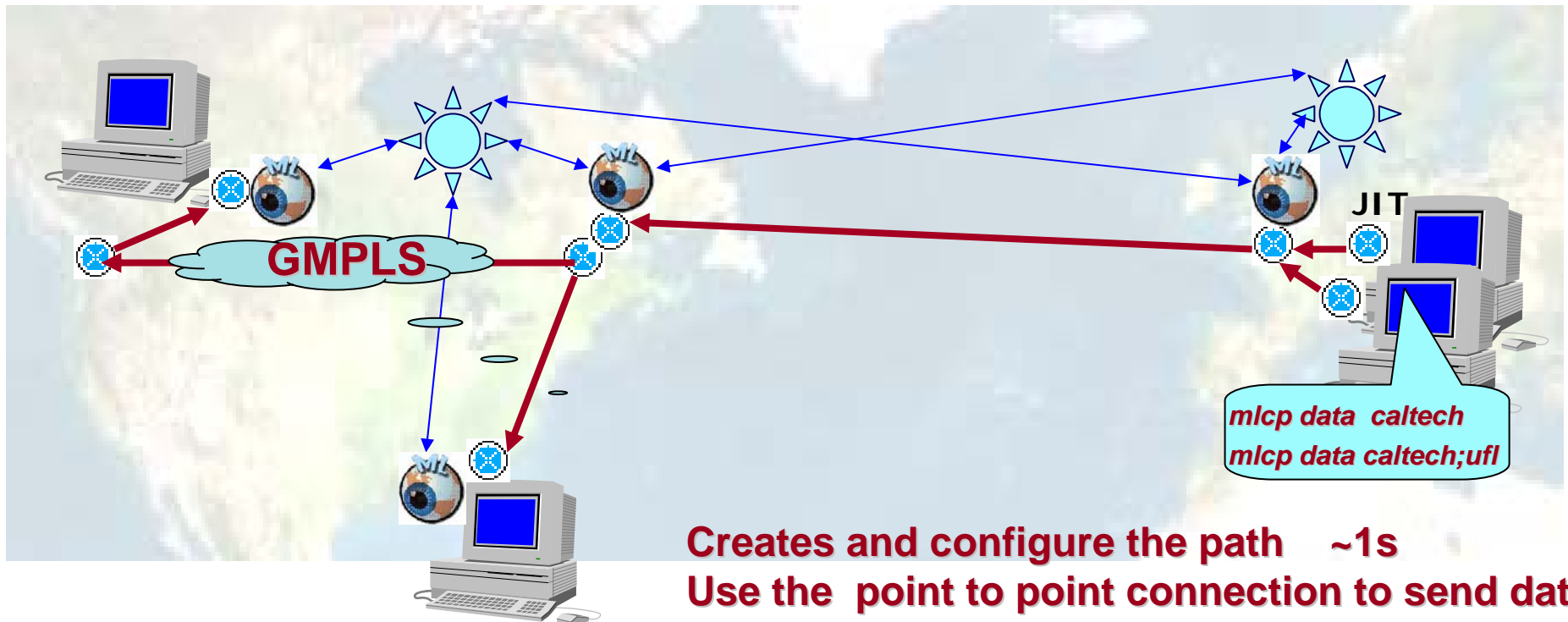




MonALISA agents to create on demand on an optical path or tree



- ◆ Control optical switches using TL1 /SNMP
- ◆ Interface with GMPLS
- ◆ Interface with JIT
- ◆ Monitor power / connectivity





Other Current Developments



- ◆ **GRID Scheduler for STAR (presented by Stratos)**
- ◆ **LISA (host monitoring, end to end network performance measurements and optimization for distributed applications – load balancing ; best connectivity)**
- ◆ **We evaluate using MonALISA for Resource Accounting in CDF**
- ◆ **Started to work with D0 in monitoring data services**
- ◆ **LAN topology discovery (using Level 2/3)**
- ◆ **Generate on demand network measurements for I2-PIPES**
- ◆ **NetFlow measurements ; Development of filters to detect attracts at the site and propagate this information to peer systems**
- ◆ **Lightweight Network Bandwidth measurement tools**



SUMMARY



- ◆ MonALISA is a **distributed service system** able to dynamically discover all the "Service Units" used by a community and provides support for handling (monitoring) information in a flexible way using self describing protocols.
- ◆ Automatic & secure code update (services, embedded applications and clients) .
- ◆ Dynamic configuration for services. Secure Admin interface.
- ◆ Access to aggregate or derived parameters and all the details for each node
- ◆ Selected real time / historical data for any subscribed listeners
- ◆ **Active filter agents** to process the data and provided dedicated / customized information to other services or clients.
- ◆ **Mobile Agents** for decision support and global optimization.
- ◆ Dynamic proxies and WSDL & WAP pages for services.
- ◆ Dedicated pseudo-clients for repository, WAP access or decision making units
- ◆ **It proved to be a stable and reliable distributed service system.**

It is currently running at ~ 160 sites

<http://monalisa.caltech.edu>