MonALISA

MONitoring Agents using a Large Integrated Services Architecture

Framework to monitor, control and optimize distributed systems

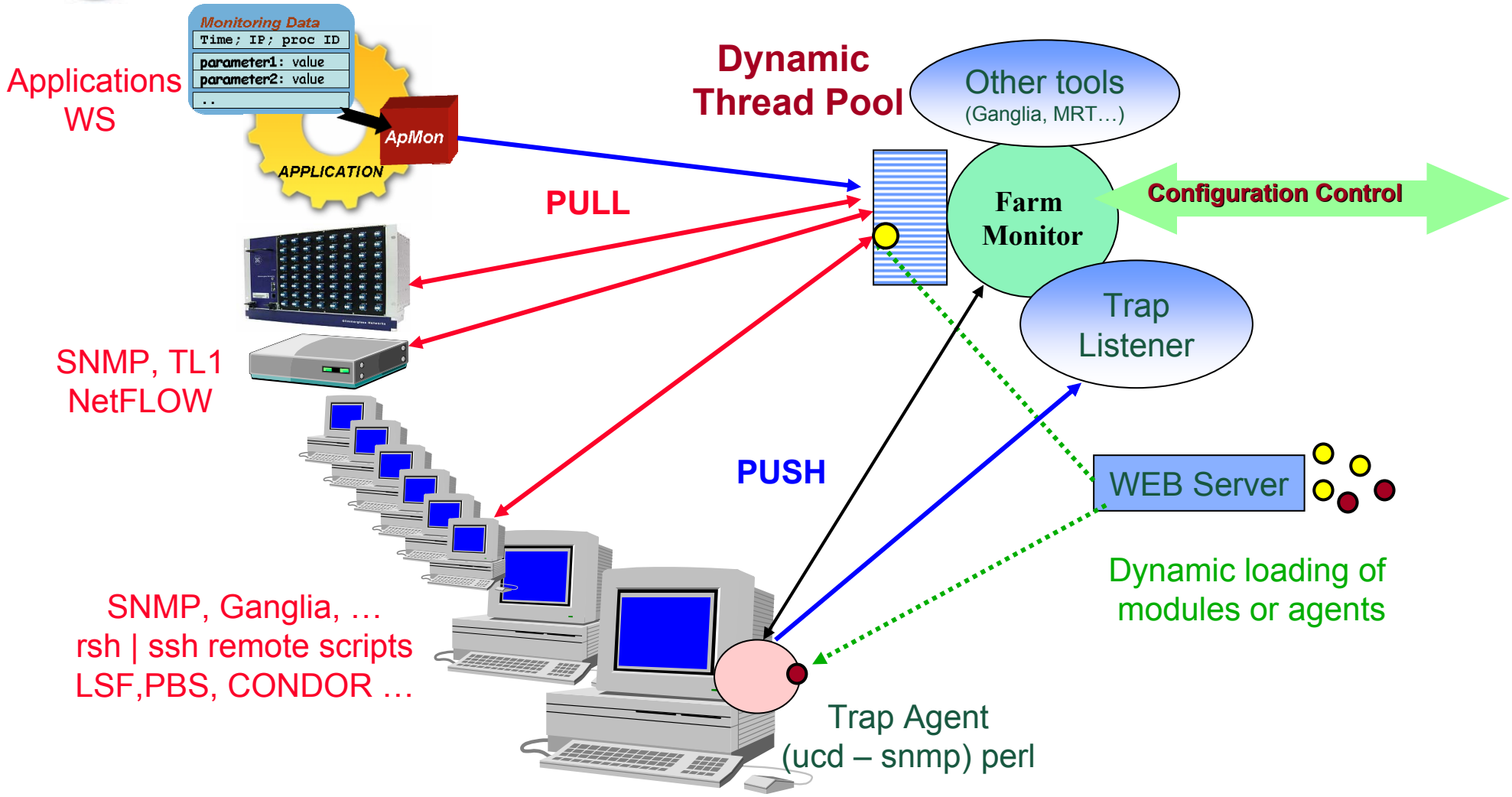
March 2005

Catalin Cirstoiu
UPB/CERN

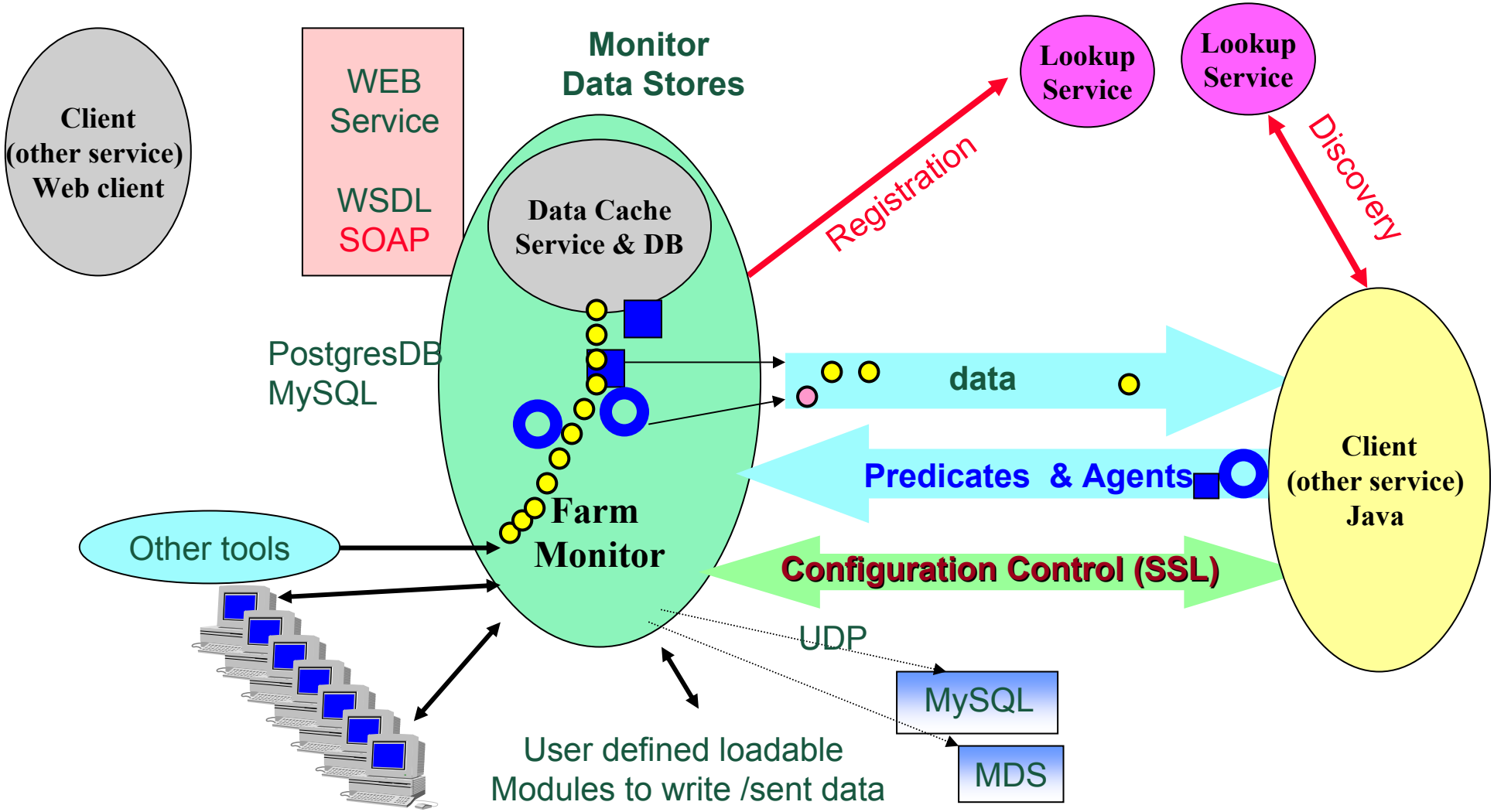
Monitoring Services

- 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.

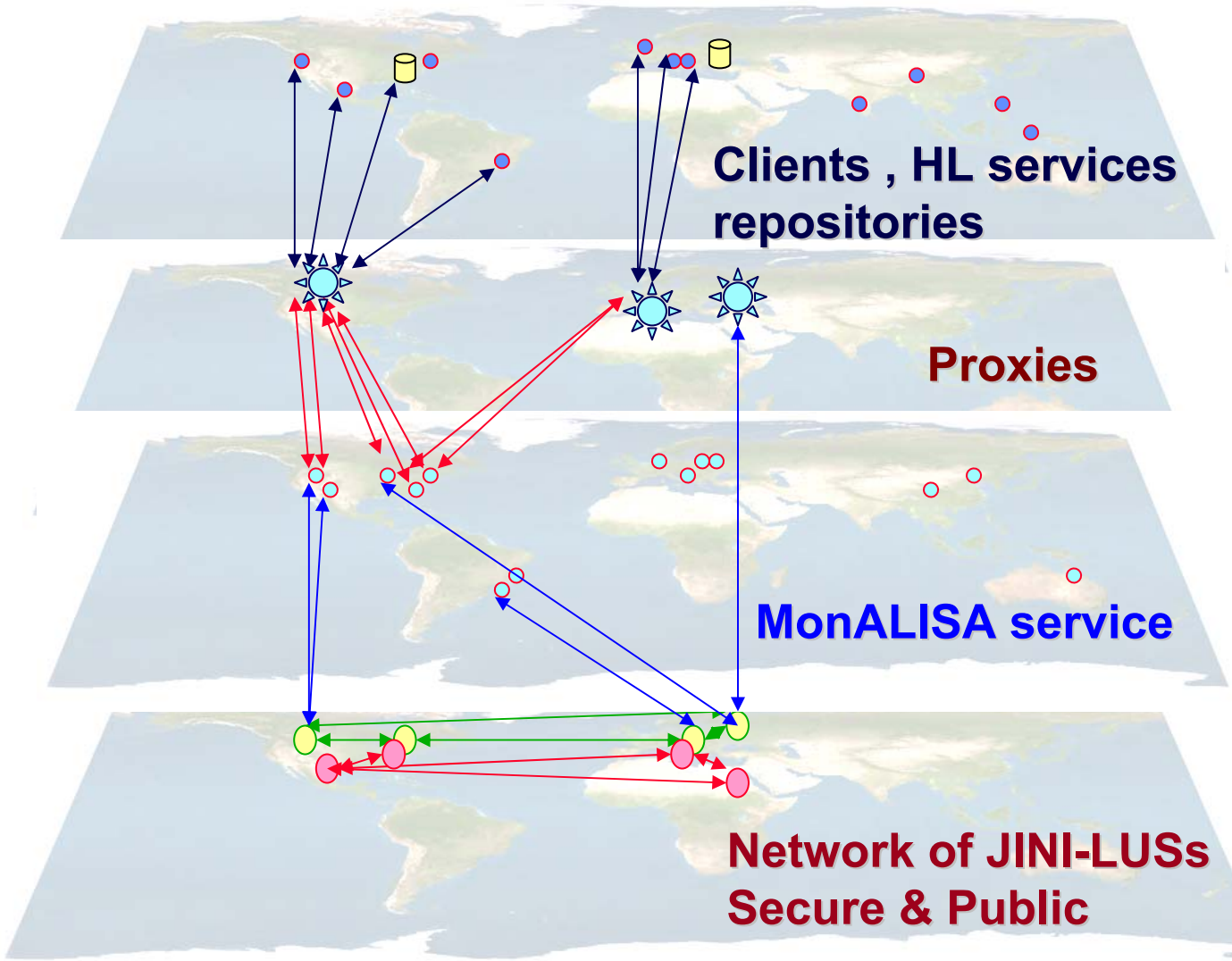
Monitoring: Data Collection



Service Monitor UNIT & Data Handling



MonALISA Discovery System & Services



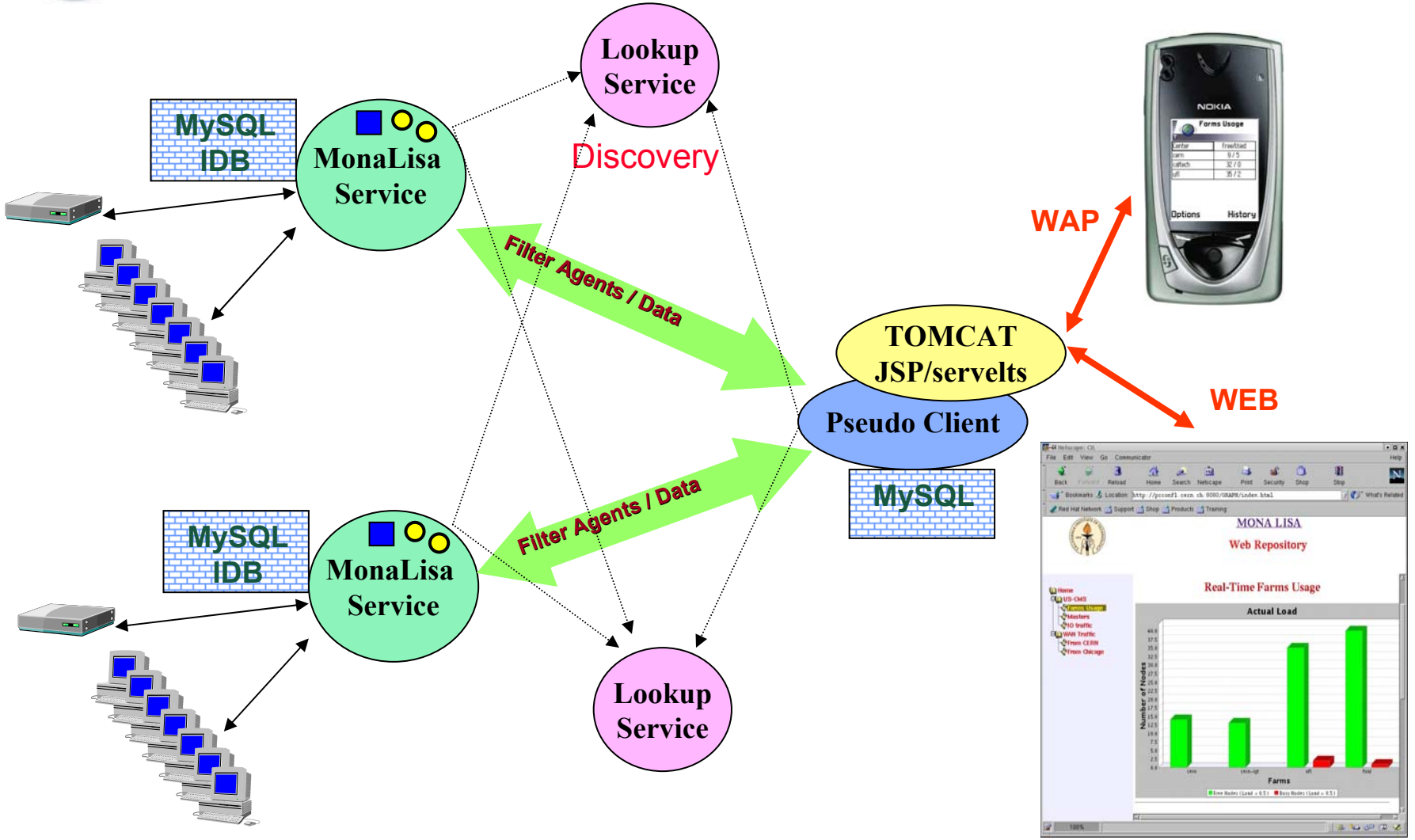
Global Services or Clients

Dynamic load balancing
Scalability & Replication
Security

Distributed Information System.

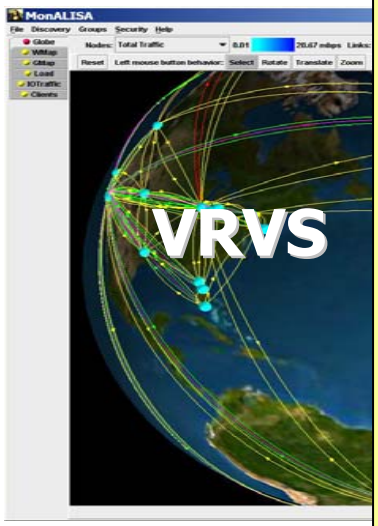
Fully Distributed Discovery
Dynamic - based on a lease Mechanism and REN

Pseudo – Clients & Dedicated Repositories



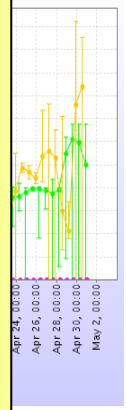
Communities using MonALISA

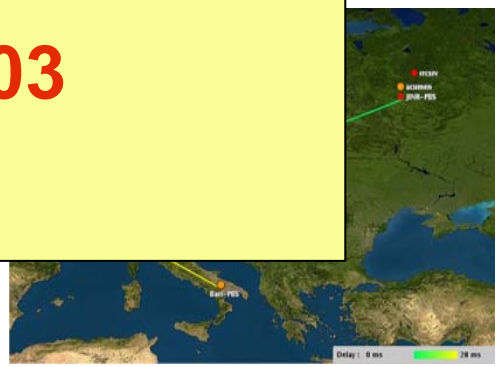
- Grid 3
- CMS sites
- CMD – DC04
- CDF
- D0SAR
- ABILENE backbone
- GLORIAD
- STAR
- ALICE
- VRVS System
- RoEduNET backbone
- Internet2 PIPES
- OSG



It has been used for Demonstrations at:

- SC2003
- Telecom 2003
- WSIS 2003
- SC 2004

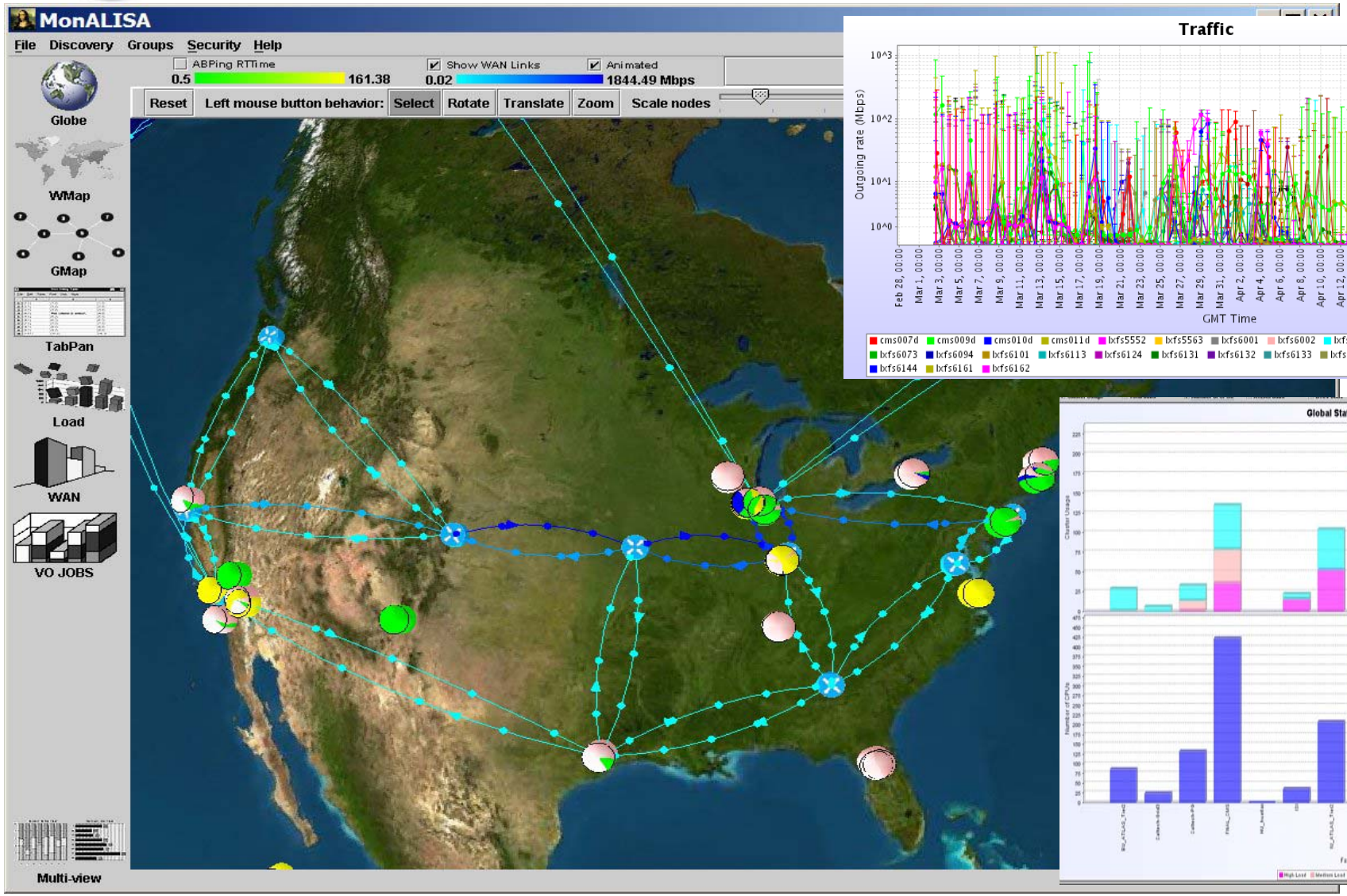




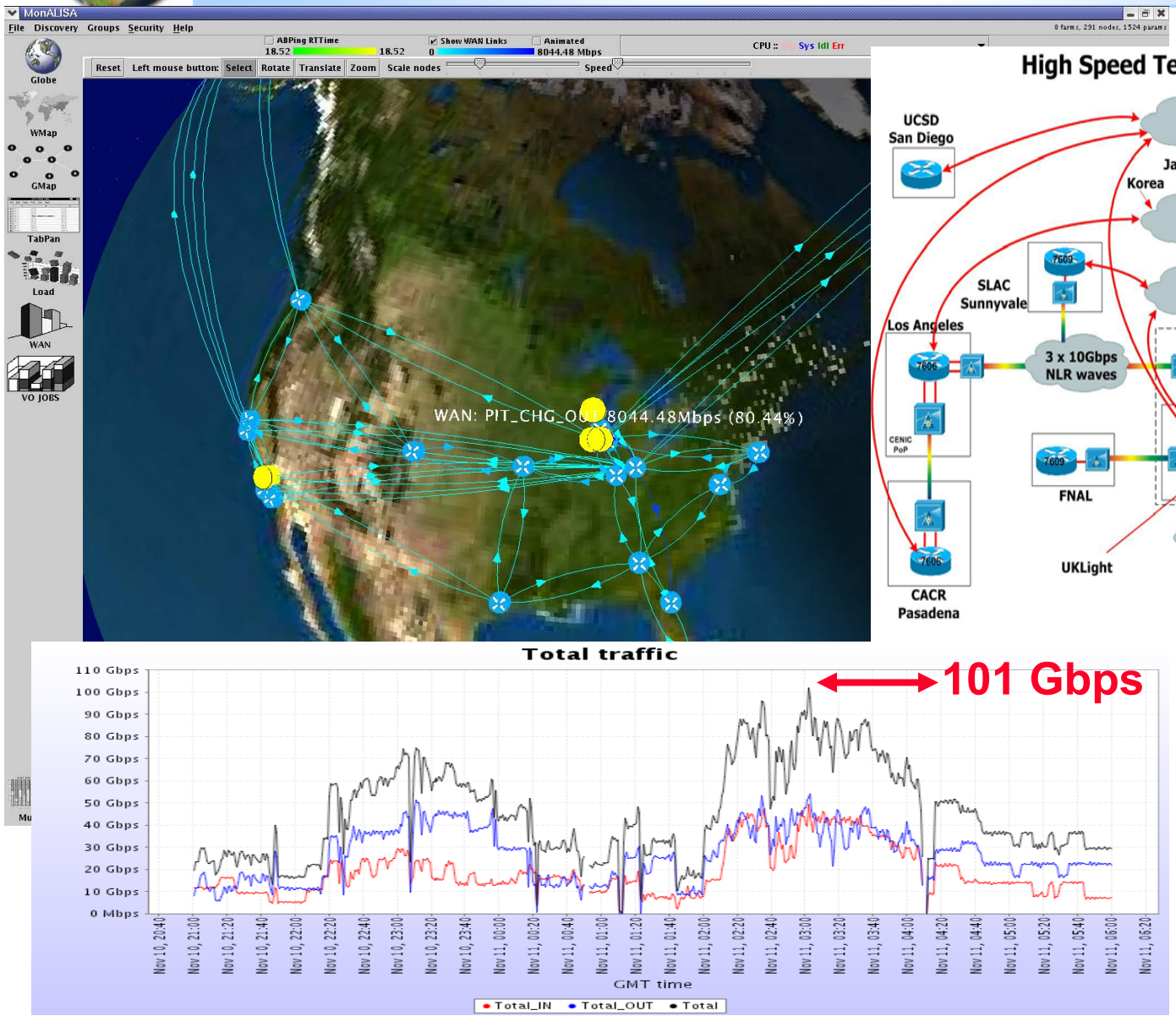
~180 Sites running MonALISA

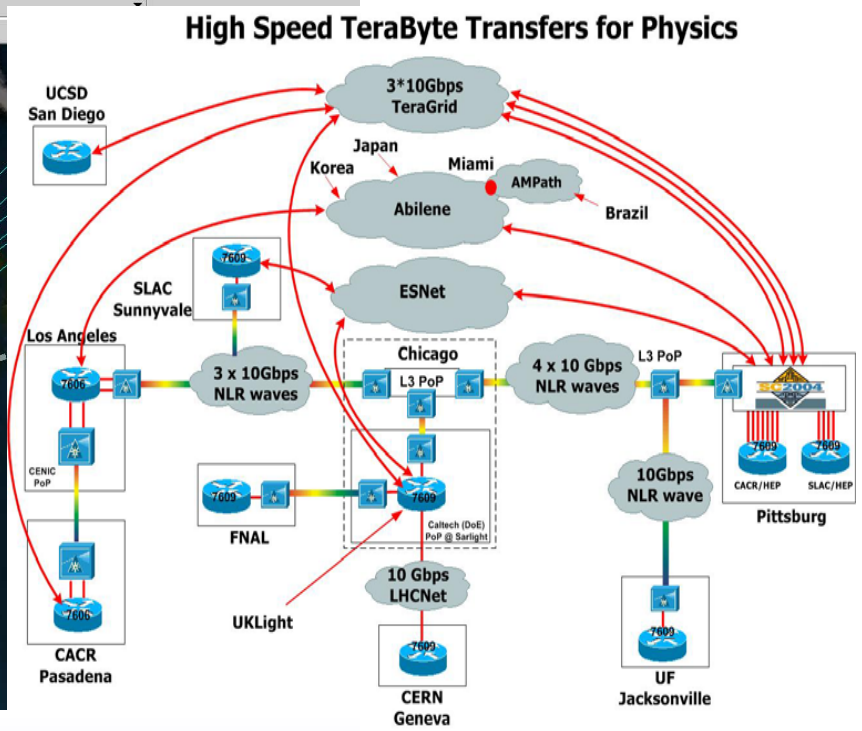
10 000 nodes / 150 000 parameters

Monitoring I2 Network Traffic, Grid03 Farms (~ 2800 CPUs) and Jobs



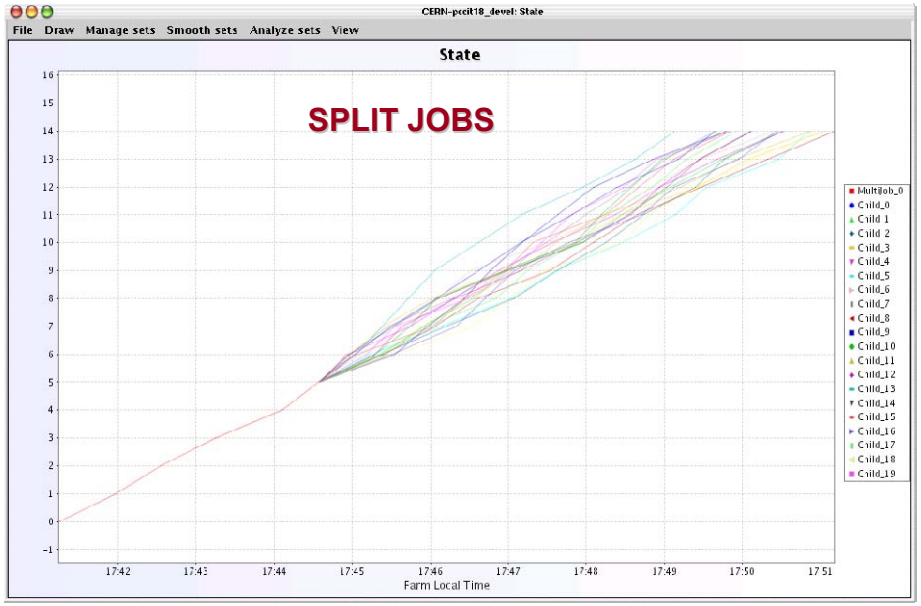
SC2004

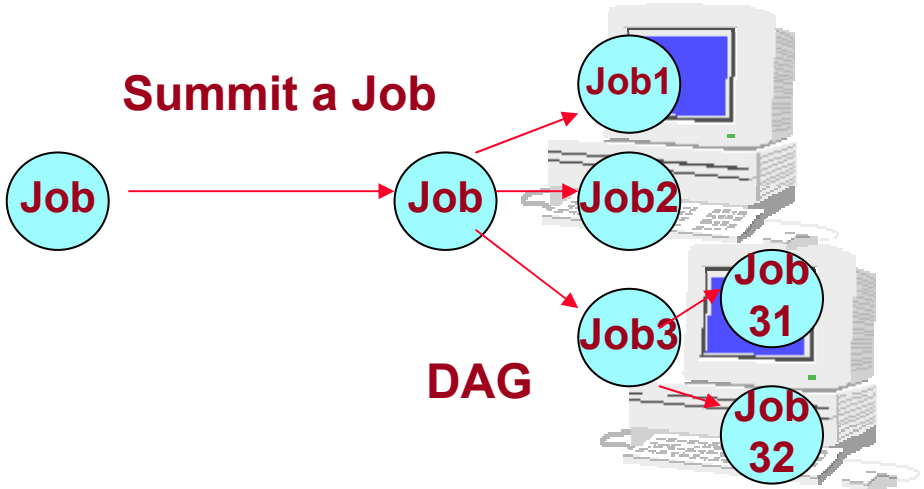




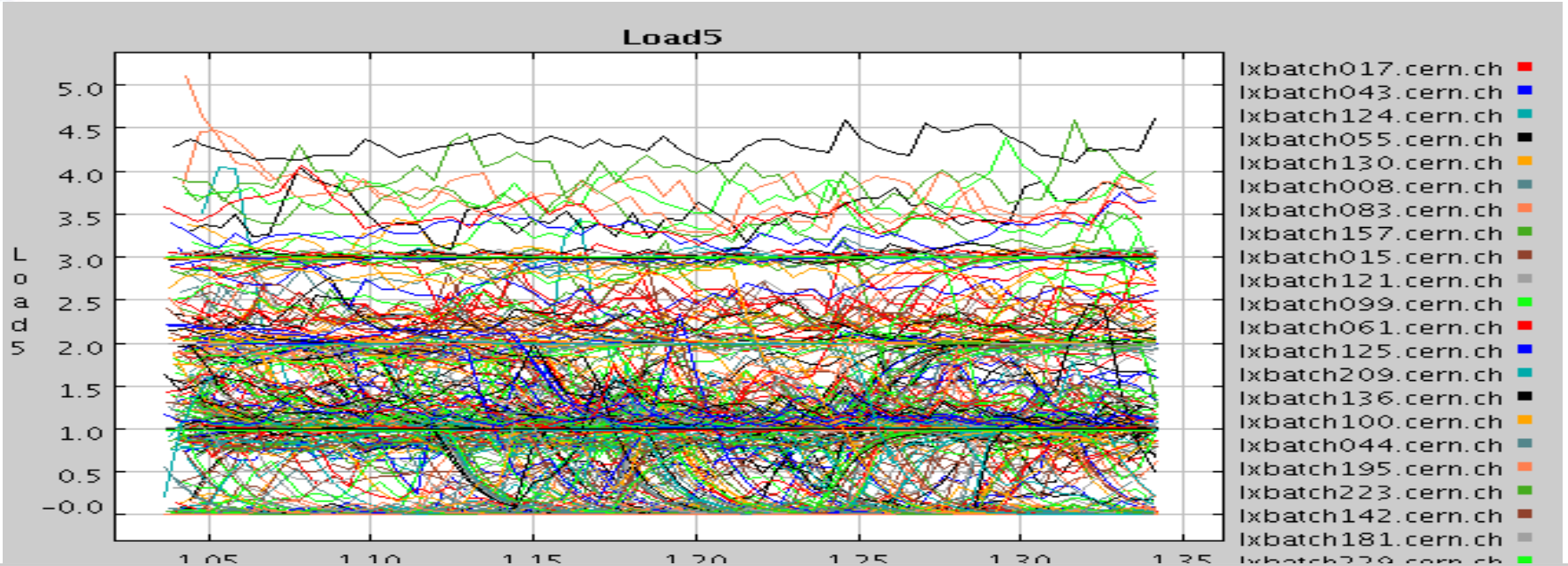
Catalin Cirstoiu

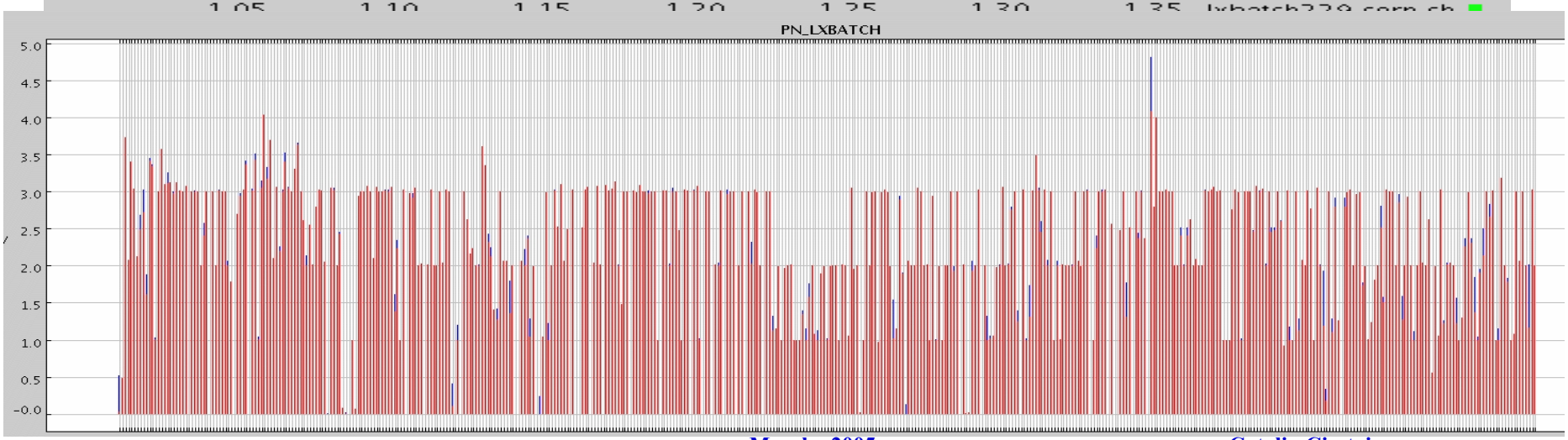
Monitoring the Execution of Jobs and the Time Evolution



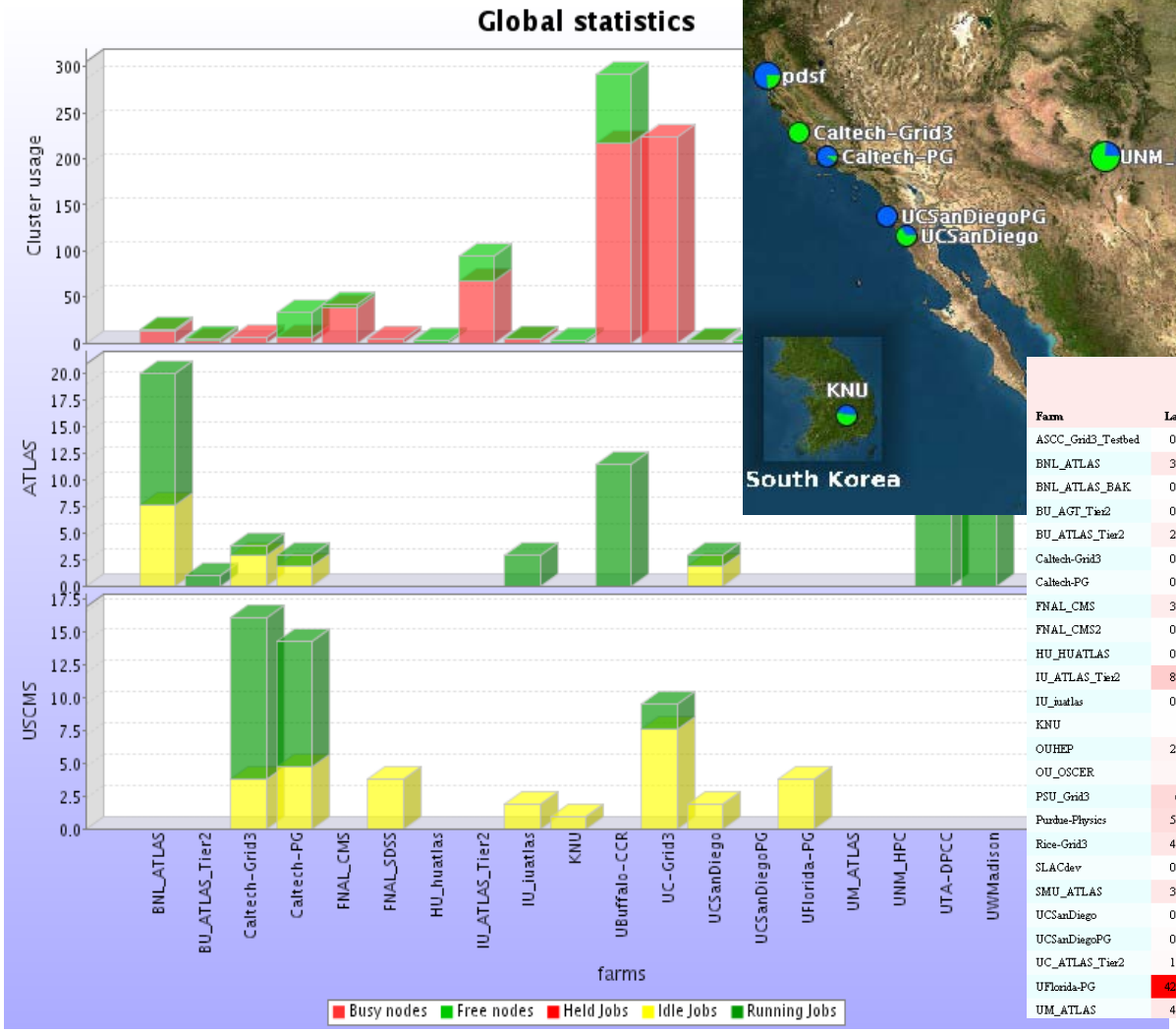


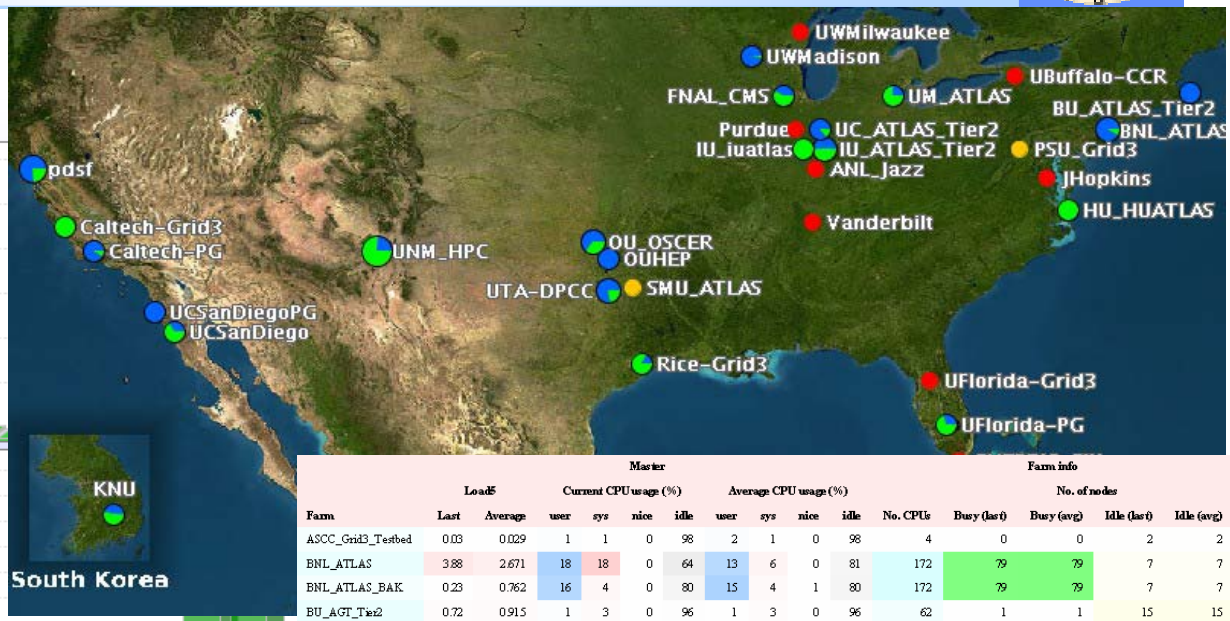
Real-time Data for Large Systems
"lxbatch" cluster at cern ~ 1200 nodes





MonALISA repositories
Grid03 : ~ 40 Sites in US + Korea





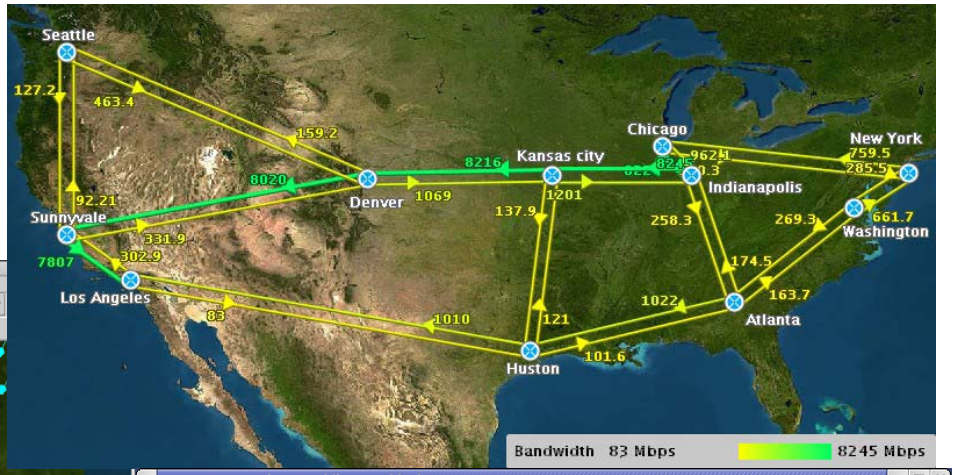
	Master										Farm info				
	Load5		Current CPU usage (%)				Average CPU usage (%)				No. of nodes				
Farm	Last	Average	user	sys	nice	idle	user	sys	nice	idle	No. CPUs	Busy (last)	Busy (avg)	Idle (last)	Idle (avg)
ASCC_Grid3_Testbed	0.03	0.029	1	1	0	98	2	1	0	98	4	0	0	2	2
BNL_ATLAS	3.88	2.671	18	18	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_iuatlas	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	55	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	58	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

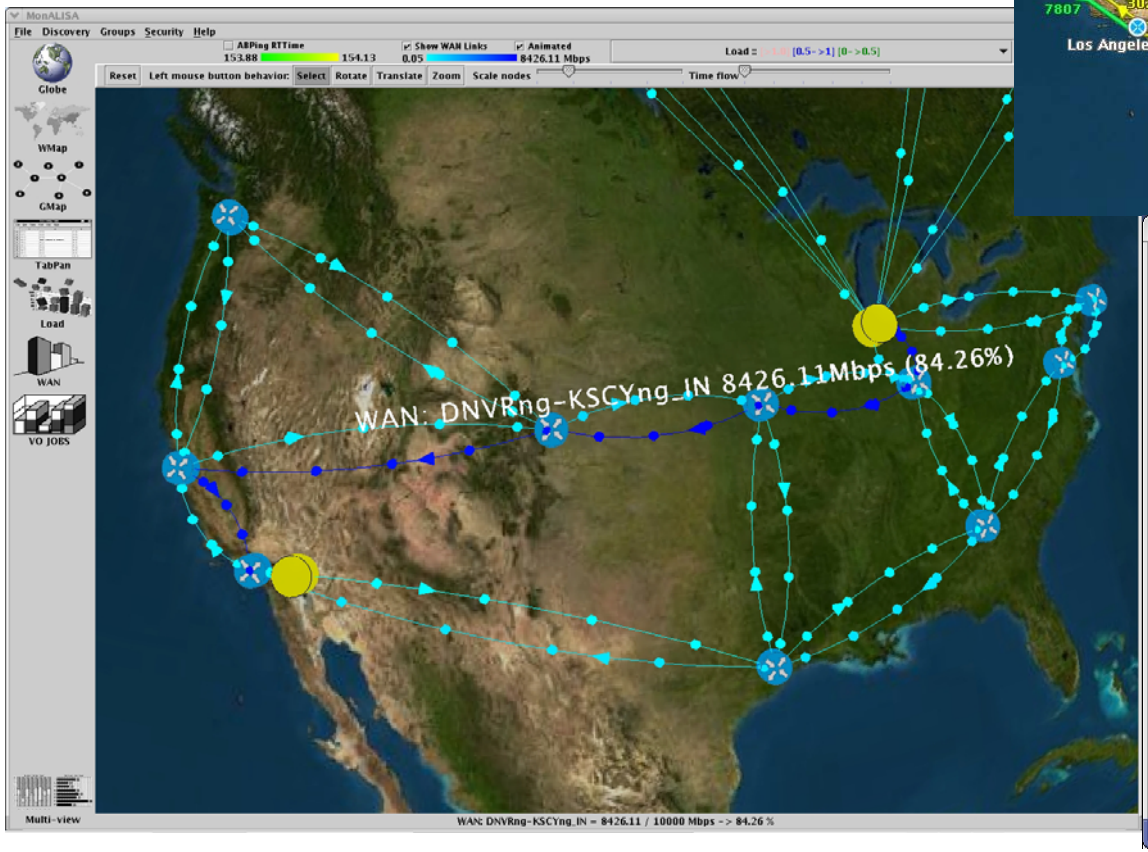
March 2005

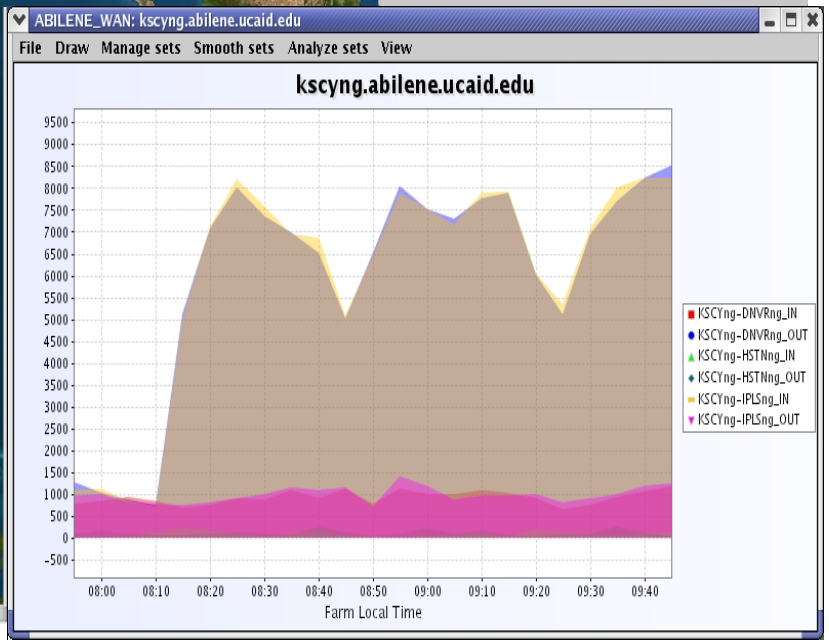
Catalin Cirstoiu

Monitoring ABILENE backbone Network

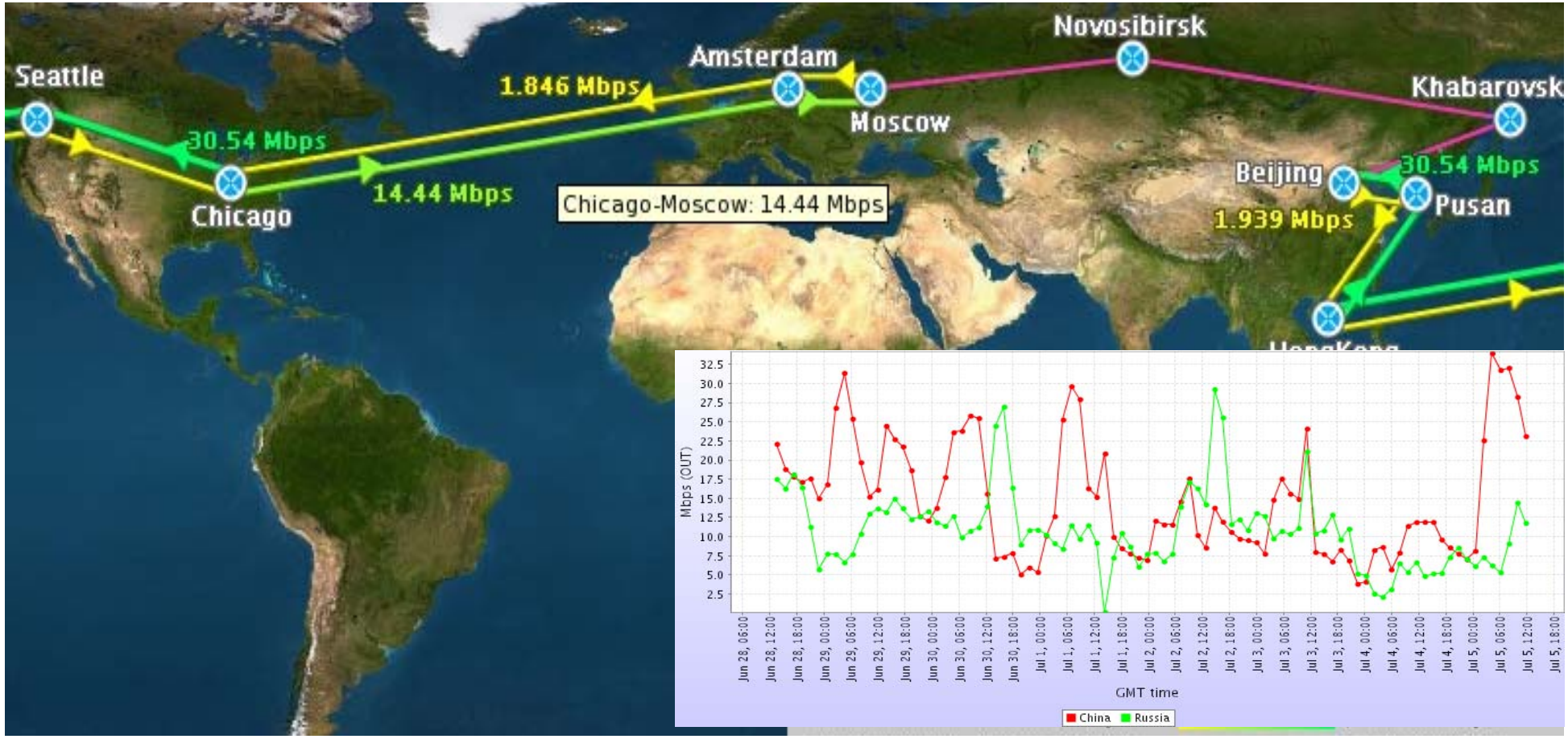
- Test for a Land Speed Record
- ~ 7 Gb/s in a single TCP stream from Geneva to Caltech



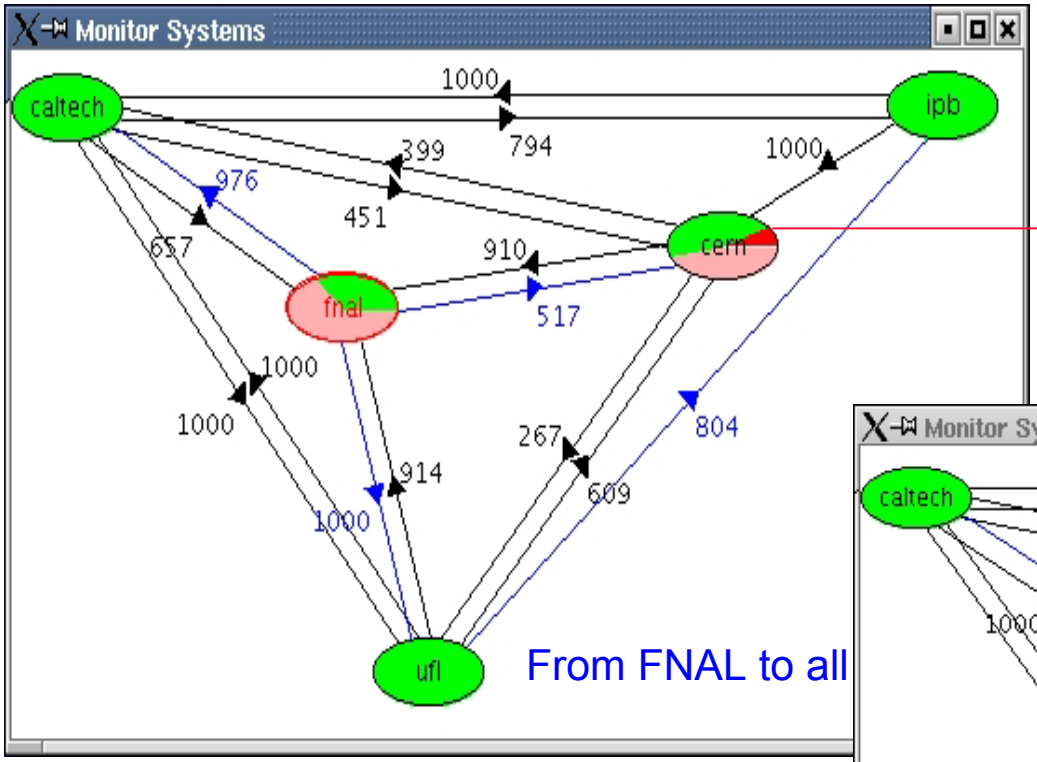




The GLORAID Network

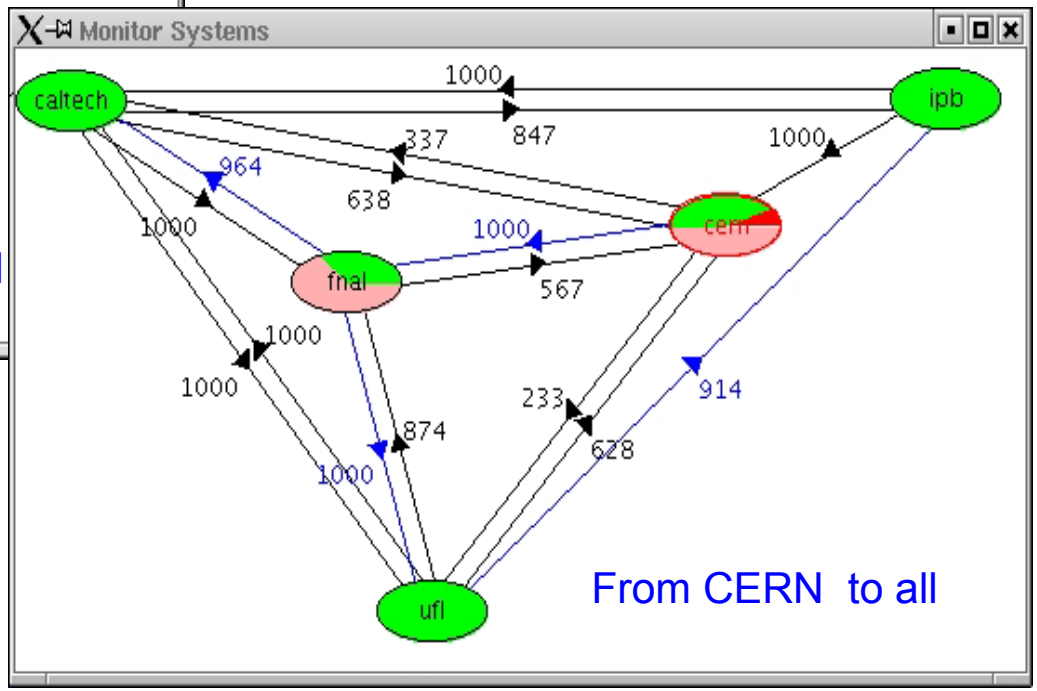


Agents for Global Optimization



Simple “Global Load” filter agent

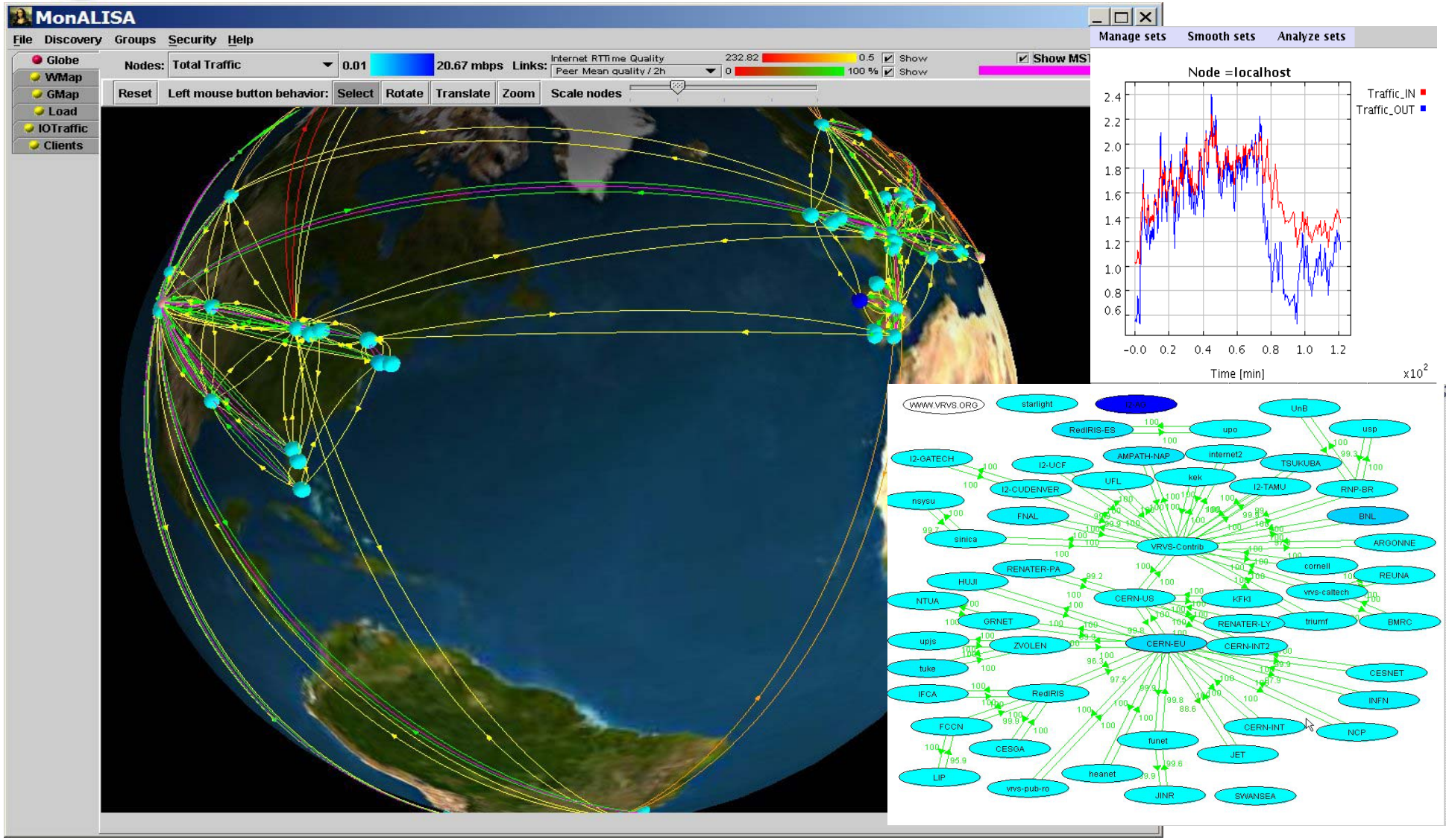
From FNAL to all



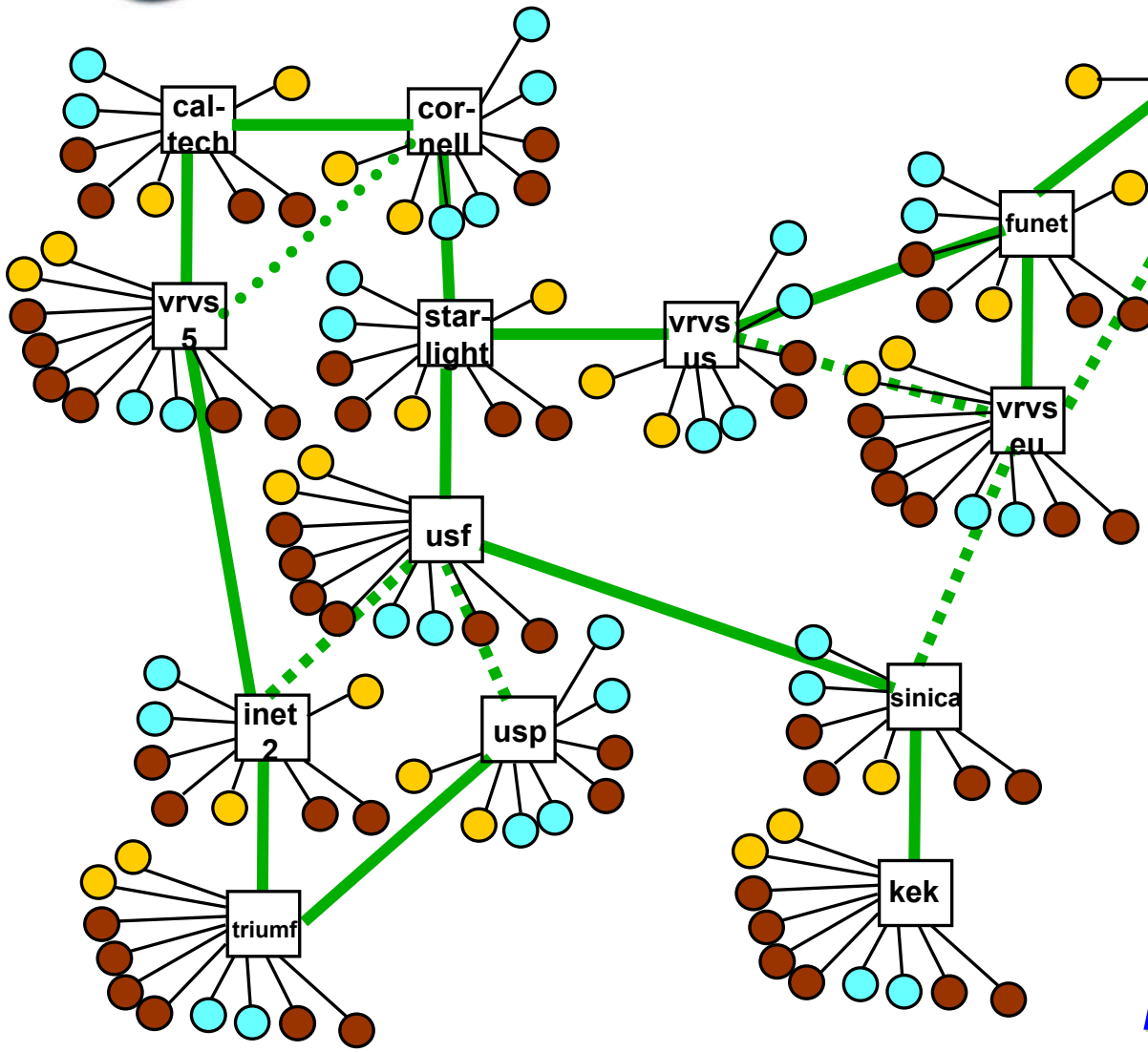
From CERN to all

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

Monitoring VRVS Reflectors
Communication Topology & Dynamic MST



The VRVS Architecture
http://www.vrvs.org



Reflectors are hosts that interconnect users by permanent IP tunnels.

The active IP tunnels must be selected so that there is no cycle formed.

Tree

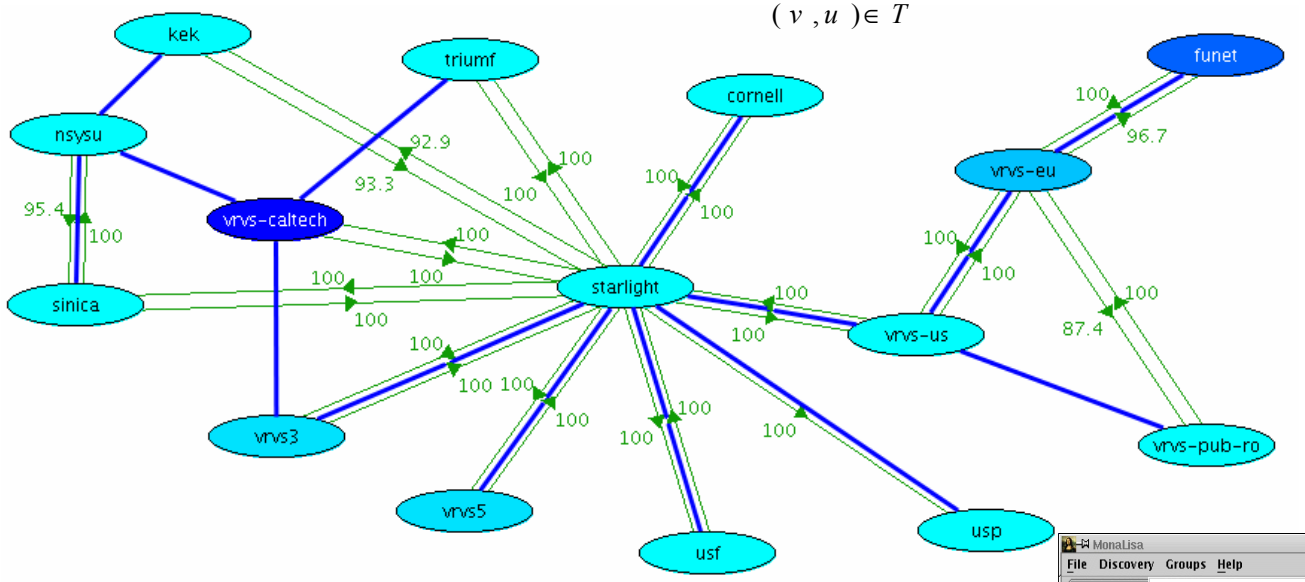
The selection is made according to the **assumed** network links performance.

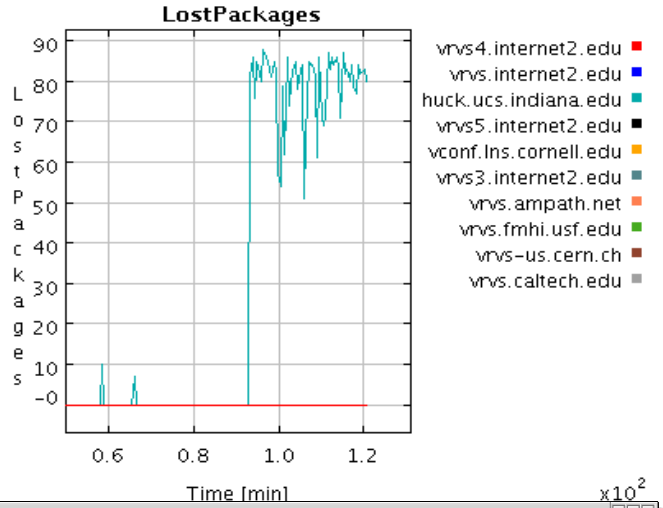
$$w(T) = \sum_{(v,u)\in T} w((v,u))$$

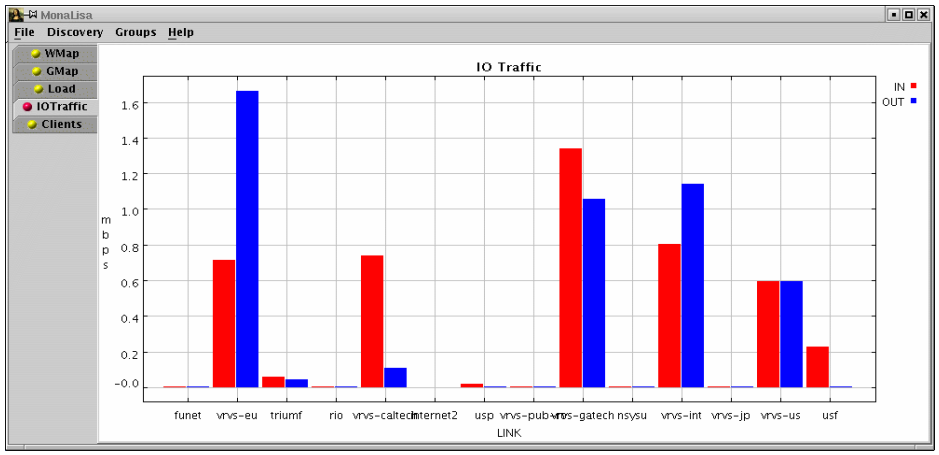
minimum-spanning tree (MST)

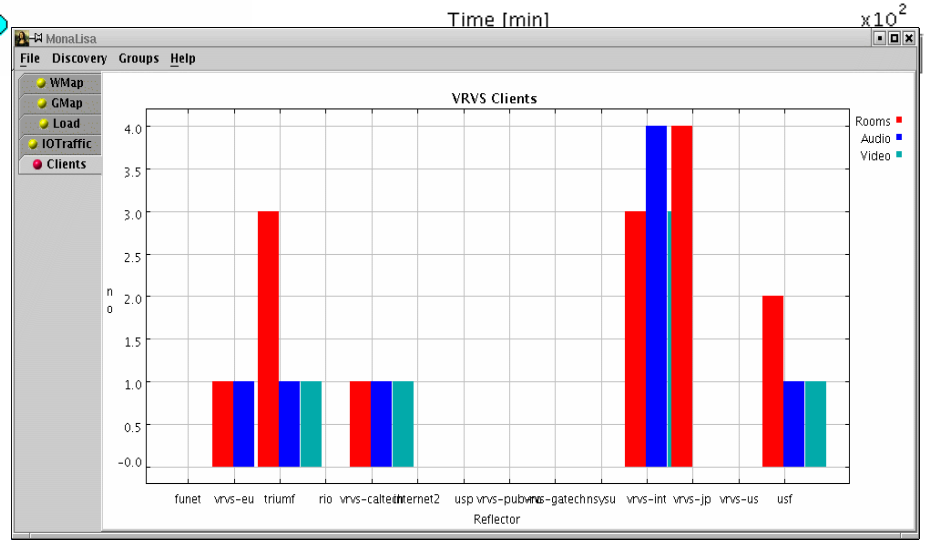
Monitoring VRVS Reflectors ; Agents for Creating a Dynamic Minimum Spanning Tree

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









Current Developments

- GRID Scheduler for STAR and Sphinx
- LISA (host monitoring, end to end network performance measurements and optimization for distributed applications– load balancing; best connectivity)
- LAN topology discovery (using Level 2/3)
- WAN topology discovery and monitoring
- NetFlow measurements
- Snort based DIDS; Development of filters to detect attracts at the site and propagate this information to peer systems
- Lightweight Network Bandwidth measurement tools
- ApMon with background application and system monitoring
- Agents that create on demand optical paths or trees