ISGC - Academia Sinica

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

The agent-based dynamic systems are able to carry out a wide range of monitoring tasks in the LHC Data Grid (and other Grids).

**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 provide 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

**Dynamic Thread Pool**
- SNMP get & walk
- `rsh | ssh` remote scripts
- End-To-End measurements

**Farm Monitor**
- Configuration Control
- Other tools (Ganglia, MRT…)

**Trap Listener**
- snmp trap

**WEB Server**
- Dynamic loading of modules or agents

**Trap Agent**
- (ucd – snmp) perl

**PULL**

**PUSH**
Service Monitor UNIT & Data Handling

- Data Stores
- Data Cache
- Service & DB
- Client (other service) Web client
- Client (other service) Java
- WSDL SOAP
- McKoi DB
- MySQL
- Other tools
- User defined loadable Modules to write /sent data
- Registration
- Discovery
- Predicates & Agents
- Configuration Control (SSL)
- Lookup Service
Secure – Automatic Update Mechanism for Services, Clients & Embedded Applications

- All running services are updated using the discovery mechanism.
- At startup, each service checks if 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

MonaLisa Service
MySQL
IDB

Discovery
Lookup Service

Filter Agents / Data

Lookup Service

TOMCAT
JSP/servlets

Pseudo Client
MySQL

Filter Agents / Data

WEB
WAP
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, very soon NetFlow

End to End Network performance

- IPERF, Pipes, Abing, ABping …

Batch Queuing Systems

- LSF, PBS, Condor, NQS, Grid Job Manager

Applications

- ORCA, GridFTP, TMDB, Proof, VRVS, Apache, RRD …
MonALISA DEMO
Monitoring And Controlling Optical Switches

Optical Path

External Application

Application Monitoring

Distributed Agent Proxies

MonALISA

Agents

Real-time monitoring

Agents Control

Optical Switch

Connection Map

MonALISA

Optical Path
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
- D0 SAR
- ABILENE backbone
- GLORIAD
- STAR
- ALICE
- VRVS System
- RoEduNET backbone
- INTERNET2 PIPES

It has been used for Demonstrations at:

- SC2003
- Telecom 2003
- WSIS 2003
SUMMARY

- MonALISA is able to dynamically discover all the “Service Units" used by a community and through the remote event notification mechanism keeps an update state for the entire system.
- Automatic & secure code update (services, embedded applications and clients).
- Dynamic configuration for services. Secure Admin interface.
- Access to aggregate farm values 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 ~150 sites

http://monalisa.caltech.edu
Global Client for HEP Grid Sites
SC03 Grid3 setup

@ CALTECH
Grid03: Monitoring Farms, Jobs, Network traffic

[Map and charts showing network traffic and statistics]
Real-time Data for Large Systems
“Ixshare” cluster at cern ~ 600 ndoes
Monitoring CMS applications: Summary Plots for Probability Density for Several Parameters
Historical Plots for large events generated by the CMS – ORCA production framework
Mobile Agents and Filters

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

From FNAL to all

From CERN to all
Monitoring ABILENE backbone Network

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

Finding a tree \( T \) that contains all the vertices of a graph \( G \) spanning tree and has the least total weight over all such trees minimum-spanning tree (MST)

**Input:** A weighted connected graph \( G = (V,E) \) with \( n \) vertices and \( m \) edges

**Output:** A minimum- spanning tree \( T \)

\[
w(T) = \sum_{(v,u) \in T} w((v,u))
\]
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”
- 1~5, 2~3
- 3 {Delay3, AS3, Net3, Descr3}
- 4 {Delay4, AS4, Net4, Descr4}
- 5 {Delay5, AS5, Net5, Descr5}
- d {DelayD, ASd, NetD, DescrD}
Monitoring the WAN Topology, the Latency and Routes