

# Building the M-grid

Arto Teräs <arto.teras@csc.fi>

Meeting with Saethor L Jonsson

CSC, Espoo, September 26<sup>th</sup> 2005



# Contents

- **The Finnish Material Sciences Grid (M-grid) overview**
- **Lessons learned in collaborative system administration**
- **Initial user experiences**
- **Grid use**
- **Security challenges**
- **Discussion**



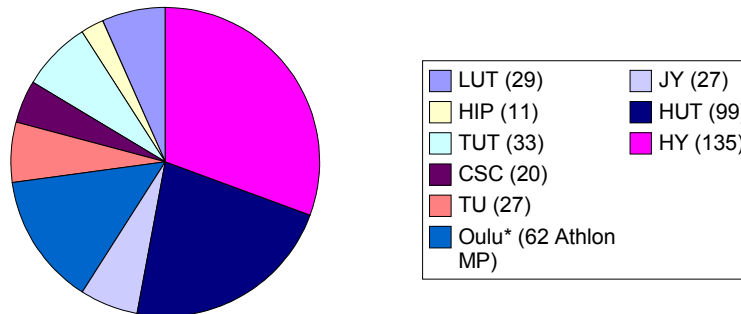
# Finnish Material Sciences Grid (M-grid)

- **Joint project between seven Finnish universities, Helsinki Institute of Physics and CSC, the Finnish IT center for science**
- **Jointly funded by the Academy of Finland and the participating universities**
  - Funding application Nov 2003, deployment Oct 2004
- **First large initiative to put Grid middleware into production use in Finland**
- **Based on Linux clusters, targeted for serial and "pleasantly parallel" applications**
- **Users mainly physicists and chemists**



# Hardware and CPU Distribution

- **Dual AMD Opteron 1.8-2.2 GHz nodes with 2-8 GB memory, 80-320 GB local disk, 1-2 TB shared storage, 2xGbit Ethernet, remote administration hardware**
- **Number of CPUs: 410 (computing nodes only), 1.5 Tflops theoretical computing power**
  - CSC's new cluster `sepeli.csc.fi` will be added soon
- **9 sites, size of sites varies greatly**



# Software Choices

- **NPACI Rocks Cluster Distribution**

- Main developers in the San Diego Supercomputing Center, U.S.A.
- Based on Red Hat Enterprise Linux 3.0 source packages, but customized for clusters
- <http://www.rocksclusters.org>

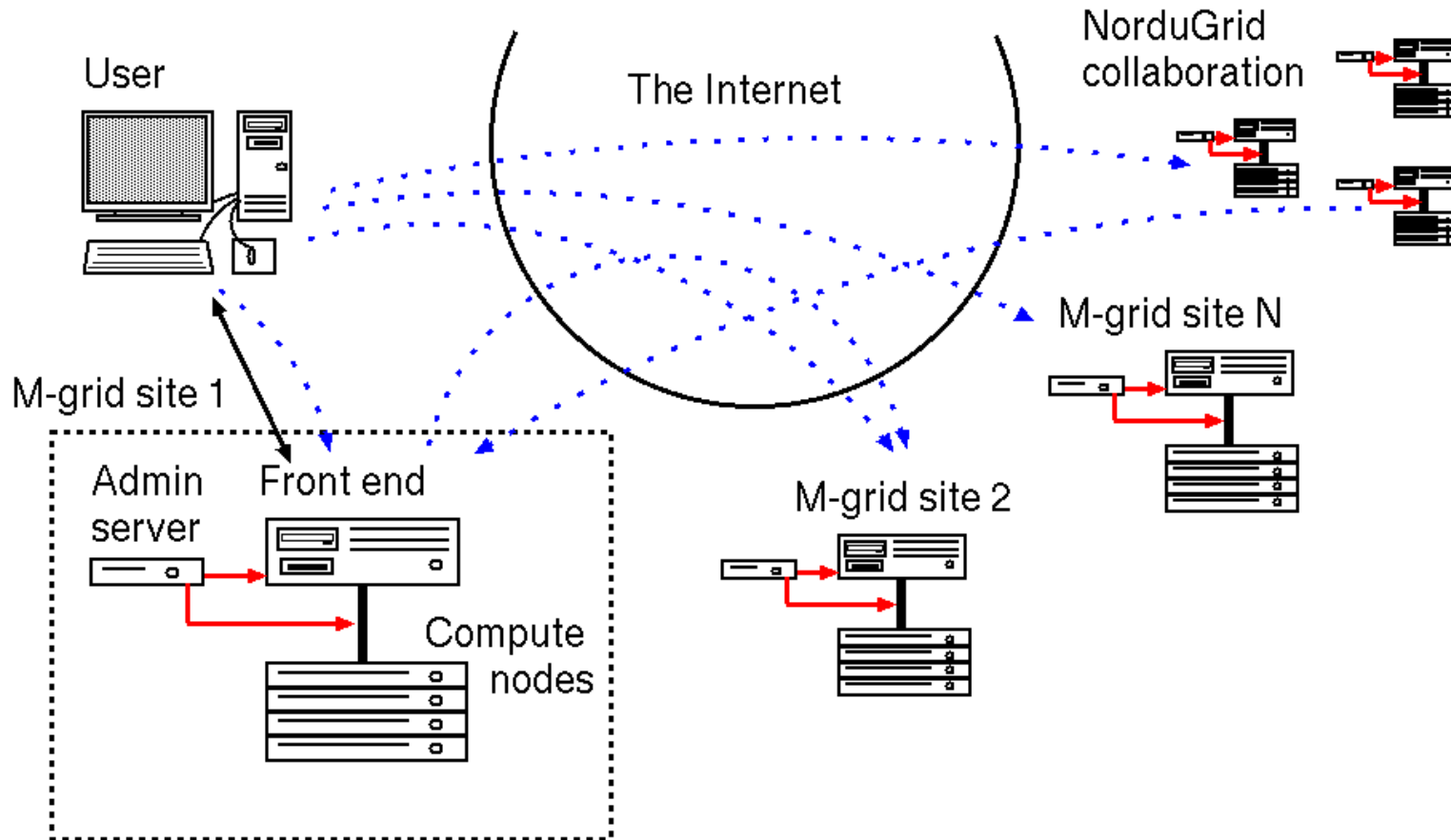


- **NorduGrid ARC Grid middleware**

- The most popular middleware in Nordic countries, one of the few suitable for production use
- <http://www.nordugrid.org>



# Grid — The Whole Picture



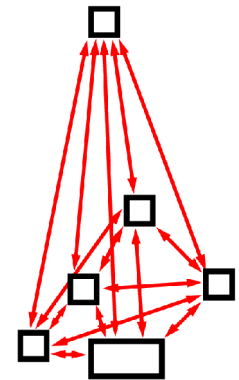
# System Administration in M-grid

- **Tasks divided between CSC and site administrators**
- **CSC administrators**
  - Maintain (remotely) the OS, batch queue system, Grid middleware and certain libraries for all sites except Oulu
  - Separate small test cluster for testing new software releases
- **Site administrators**
  - Local applications and libraries, system monitoring, user support
- **Regular meetings of administrators & support network**



# Goals of Shared System Administration

- **Centrally administered foundation while maintaining local control**
  - A new paradigm — traditionally in Finland HPC resources have been centralized at CSC
- **Easier for universities than setting up their own system from scratch**
  - However, needs a significant amount of work both from CSC and the local sysadmins
- **Take advantage of the local sysadmin expertise in software used by the local researchers**
  - Faster and better user support



**36 pairs for collaboration!**



# Positive Experiences

- **Local sysadmins have found CSC support valuable**
  - Having also local control (root) is important psychologically
- **Participants have used their expertise to pick up suitable tasks, fruitful discussion on the mailing list**
- **Collaboration has strengthened relationships between groups also in their research**
- **Systems are closer to the user**
  - Easier to talk to the own group sysadmin, less support requests to CSC
- **Local sysadmins are often also users**  
**=> direct usability feedback to CSC**



# Negative Experiences

- **Configuring the batch queue system took time**
  - Initially used Sun Grid Engine v. 5.3 which wasn't very well suited for clusters, v 6.0 was capable but complicated
- **Wiki-based FAQ hasn't really taken off**
- **User documentation became scattered**
  - Mainly due to lack of human resources (people assigned to other projects before finishing the docs)
  - Compiling the documentation needs central coordination
- **Some users found support poor**
  - Clearly divided between sites: on some sites users are very happy



# User Experiences

- **Users got started relatively quickly: the average total load of the M-grid soon rose above 50%, currently close to 100% in most clusters**
- **Performance has been quite satisfactory**
- **Problems centered around Fortran compiler and MPI runs**
  - MPI works, but killed jobs can leave unfreed resources behind
  - One compiler isn't satisfactory for everybody: currently Pathscale, Portland Group and GNU compilers used



# Grid Use

- **Policy: Users may submit jobs both locally and through Grid interface**
  - Grid jobs have higher priority than local jobs in 20% of each system, and may fill all available free nodes
- **NorduGrid ARC middleware was installed in August 2005**
  - Problems with the 64 bit environment and Sun Grid Engine support took time (mostly solved now)
  - Little grid use yet, time will show how users adopt the Grid environment; our collaborative network will hopefully be helpful



# Security Challenges in the Grid

- **Grid crosses organizational boundaries**  
**=> Collaboration and mutual trust needed!**
- **Some new risks and all the old ones with wider impact area**
  - Compromised user account most probable method of intrusion
- **Definitions of responsibilities necessary to build trust**
  - Risk analysis
  - Acceptable use policy
  - Incident response



# Security Challenges in the Grid (cont.)

- **Getting all the relevant parties involved**
  - Computing centers, university IT departments, local admins, CERTs and also users
  - International collaboration
- **Distributed systems with hundreds of users are always vulnerable**
  - Focus on detecting break-ins quickly
  - Clear procedures how to act when a system is compromised
- **Data protection and privacy issues**
  - Implementation details vary between sites



# Conclusion

- **Sharing system administration tasks can work**
  - Partners need to know each other — face to face meetings are very useful in avoiding flame wars
- **User support in a distributed system potentially very good but needs special attention**
- **Grid projects can strengthen ties between groups even before actual Grid use**
- **Rocks Cluster Distribution is one good choice for a Linux cluster operating system**
- **International collaboration on security and policies needed**



# More Information

- M-grid home page: <http://www.csc.fi/proj/mgrid/>
- Rocks home page: <http://www.rocksclusters.org>
- NorduGrid home page: <http://www.nordugrid.org>
- Contact people:
  - Arto Teräs <[arto.teras@csc.fi](mailto:arto.teras@csc.fi)>
  - Kai Nordlund <[kai.nordlund@helsinki.fi](mailto:kai.nordlund@helsinki.fi)>
  - Olli-Pekka Lehto <[oplehto@csc.fi](mailto:oplehto@csc.fi)> (Rocks)
  - Urpo Kaila <[urpo.kaila@csc.fi](mailto:urpo.kaila@csc.fi)> (security)
- Thank you! Questions?

