Fact-Sheet GIN @ SC 2006

Context

- Demonstration of initial Interoperation between production Grids by the Grid Interoperation Now (GIN) – Open Grid Forum (OGF) Community Group at the Supercomputing 2006 in Tampa, USA

Grid Interoperation Now (GIN)

- Community Group of the Open Grid Forum, Chairs: Erwin Laure, Stuart Martin
- Goal: Organize and manage a set of interoperation efforts among production Grid projects
- The results of these interoperations may feed back into the interoperability efforts being conducted by the Standards Working Groups
- The group focus on interoperation: What needs to be done to get Grids to work together
- Different from interoperability: Software from multiple communities communicate through standards
- Participants of these groups are members from real production Grids (DEISA, EGEE, TeraGrid, etc.)
- The focus of this group is to plan and implement interoperation in the following specific areas:
    - (1) Job Submission (job description and execution)
    - (2) Data management (data location and movement)
    - (3) Information services (information about site data, including location, admin contacts, etc.)
    - (4) Security discussions (authentication/authorization and identity management)
    - (5) Operations / Initial Applications (applications to test the interoperations)

Area: Job Submission (GIN Jobs)

- Coordinated by Steven Newhouse (OMII-UK & OMII – Europe)
- Interoperation based on Job Submission and Description Language (JSDL) OGF Standard
- A simple JSDL document was defined with a “POSIXApplication” using “/bin/echo hello world”
- The DN’s and CA details of the people that access resources was listed at Websites
- Systems and open ports where listed at Websites, including VOMS information (gridmapfiles, certs)
    - E.g. London e-Science centre: Globus WS-GRAM – ariane.doc.ic.ac.uk:55100
    - Contacts for each Grid were shown on the Website (administrators, responsible persons)
    - Some information about resource types were also shown on Websites (e.g.desktop hosts)
    - Information about different usage policies were written on Websites (e.g. NGS policies)
- Participants that provided resources in that area:
    - TeraGrid, London e-Science centre, UK National Grid Service, EGEE, OSG, NorduGrid/ARC
- Participants that work with specific JSDLs:
    - GridSAM (UK) implementation (no changes to JSDL)
    - NAREGI (JAPAN) implementation
        - To work with the NAREGI job scheduler several changes were needed
        - (1) a different JSDL schema definition and the (2) WallTimeLimit was required
        - To work with the NAREGI workflow tool several differences were identified
        - (1) different JSDL namespace definition and the (2) required Output and the (3) WallTimeLimit
    - CROWNGrid (CHINA) implementation
        - CROWNGrid used GridSAM for execution, by extending its JSDL
        - (1) The JSDL and underlying Web service technologies were changed to support RPC style Web service invocation
        - (2) The JSDL was also extended to make CROWN scheduler support PBS jobs
        - (3) For security, the JSDL was changed to include username, password, etc.

Area: Data management (GIN Data)

- Coordinated by Erwin Laure (EGEE)
- GridFTP interop and tuning (led by Bill Allcock)
- Interop with “SRB-island” and “SRM-island”, but not across these islands
    - Planned: SRB is working on an SRM interface – is this the best way forward?
    - A tool for validating compatibility and interoperability of Storage Resource Managers (SRMs) for heterogeneous storage systems was used
    - Tool checks client-SRM adherence to the spec. and SRM-SRM remote copy capabilities.
    - Tool is being used by the Open Science Grid (OSG) consortium and the SRM-collaboration
- SRB islands (led by Reagon Moore)
    - About 14 data Grids took part in Storage Resource Broker (SRB)
- SRM islands (led by James Casey)
    - About 7 sites took part in Storage Resource Manager (SRM)
- Problems and issues
    - Version compatibility of the software was a huge problem
    - Network challenges (firewalls and tuning)
    - Setting up VO and usage was error-prone (email interactions, error messages not helpful)
    - Establishment of trust relationships between data Grids only done by standard registration mechanisms

Area: Information services (GIN Info)

- Coordinated by Laura Perlman (ISI) and Satoshi Matsuoka (TIT)
- In order to identify appropriate resources for users within a Grid infrastructure there must be some form of resource information, provided through schemas and with standard query mechanisms.
- There are three different schemas used in production Grid projects today
    - (1) Glue Schema, (2) ARC schema and the (3) CIM schema with NAREGI extensions.
- (1)The Glue schema is used by OSG, EGEE and Teragrid
    - It has been mapped to LDAP, XML and the relational model.
    - The GIN-BDII is a top-level BDII with information from all Grids in accordance to Glue
- (2) The NDGF are using the ARC schema
    - The ARC-BDII is a top-level BDII with information from all Grids in accordance to ARC
- (3) NAREGI is using the CIM schema with a vendor extension
    - The GIN Cell Domain is a NAREGI cell domain with information from from all Grids in accordance to CIM and the NAREGI vendor extensions.
- Identified a subset of information items that can be used as a common minimum set
- Works have been done on translation of these information items between these schemas
- Not all attributes can be found in all Grids, this will lead to missing data when using them
    - The implementation may have errors, sites may not be publishing the information correctly
- Interoperation was demonstrated by using Google Earth, showing all participating Grid sites
- Participants with information: Grid, Contact, Schema, Datamodel, Interface, Client
    - EGEE, Laurence Field, Glue, LDIF, LDAP, ldapsearch
    - DEISA, Thomas Soddemann, Glue, XML, WS, WS
    - OSG, Wang Shaowen, Glue, LDIF, LDAP, ldapsearch
    - NDGF, Balazs Konya, Nordugrid, LDIF, LDAP, ldapsearch
    - NAREGI, Yuji Saeki, Naregi, ??, ??, Naregi API
    - TeraGrid, Laura Pearlman, Glue, XML, WS, wsrf-query
    - PRAGMA, Sugree Phatanapherom, Glue, LDIF, HTTP, wget
    - NGS, Matt Viljoen, MDS, LDIF, LDAP, ldapsearch
    - APAC, Gerson Galang, Glue, LDIF, LDAP, ldapsearch

Area: Security discussions (GIN Auth)

- Coordinated by Dane Skow
- Developed GIN policies and experimented with cross-domain authorization using VOMS
- GIN VOMS service provided by EGEE
    - GIN participants have registered themselves to the VOMS service
- An Excel spreadsheet has shown an inventory of which CAs are recognized by which Grid
- It was planned to publish that information within an information service for each Grid
- Configuration information for the VOMS service usage was published on Websites
- A current list of DNs was generated daily and updated (with an RSS feed!)
- The following forms of authentication were performed
    - GSI compliant x.509 proxy certificates for authentication to pre-WS services
    - OGSA Basic profile authentication
    - Transport of supported authorization attributes via VOMS extensions

Area: Operations / Applications (GIN Ops)

- Coordinated by Cindy Zheng (PRAGMA)
- This area tried to used an application to help the technical other groups to verify interoperation
- Based on the early experiences additional applications will be added
- Additional applications are required to validate the technical approaches
- Furthermore they identify needs and opportunities for improved interoperation
- Specific issues from pilot-applications to know where to start would be helpful (GIN-DATA)
- After initial application testing, GIN began to structure interoperation work through a set of matrices showing interoperation by Grid projects
    - E.g. showing GridFTP connectivity and performance between participating Grid facilities

GIN Resource Testing Portal (no special area)

- GIN Resource Testing Portal that provides access to and monitors the availability of GIN resources
- A GIN resource testing portal was demonstrated (based on GridSphere)
- During the demonstrations CHARMM (Chemistry at HARvard Molecular Mechanics) and MadCity traffic simulation workflows have been executed and monitored in the GIN Resource Testing Portal.
- Jobs of the workflows are utilising GRAM4 through the GEMLCA architecture on the TeraGrid, GRAM2 on the TeraGrid, the OSG and the UK NGS, and broker-based submission to EGEE sites
- The demonstrations also cover the Grid Resource Monitoring capabilities of the portal showing the actual state of the GIN VO resources

Observation from the author

- About 20 Grids were participating, mainly coordinated by well known people of the Grid community
- Decentralized and uncoordinated documentation of what was done at Supercomputing 2006
- OMII – Europe may contribute to this group (as a non-production Grid) with documentation/analysis

Future

- The GIN efforts were known to be useful and therefore the most people working in that group would like to continue these efforts, bringing new areas of application into the efforts
- The real definition of the future of GIN will be discussed at OGF19 in North Carolina

Links

- http://forge.ggf.org/sf/projects/gin