NMI Integration Testbed
NMI-EDIT
Identity and Access Management
Case Study Series

Exploring Technical and Policy Considerations for Inter-Institutional Grids: NMI Testbed Grid

November 2004

NSF Middleware Initiative (NMI) Integration Testbed Case Study Series

Series contact: Mary Fran Yafchak, Southeastern Universities Research Association, maryfran@sura.org.

The NMI Integration Testbed Program provided practical evaluation of NMI components within the context of real projects and application scenarios from June 2002 through November 2004. During that time, NMI Testbed sites collectively submitted over 220 evaluation reports to middleware component developers as direct feedback into the NMI development cycle. Site representatives also actively inspired, promoted and facilitated the integration of middleware throughout their institutions.

The NMI Integration Testbed Case Study Series documents the most significant influences and outcomes of NMI Testbed sites' middleware integration efforts, highlighting intersections with established projects, application contexts and influences, drivers for innovation, decision points and challenges. Through this documentation, the work of these pioneering institutions is captured to provide a breadth of insight and approaches for others to use towards successful middleware development and deployment.

This NMI Integration Testbed Case Study Series is sponsored by the National Science Foundation Middleware Initiative-Enterprise and Desktop Integration Technologies (NMI-EDIT) Consortium of EDUCAUSE, Internet2, and SURA. Additional support was provided by the National Science Foundation Cooperative Agreement NSF 02-028, ANI-0123937.

Executive Summary

The National Science Foundation Middleware Initiative (NMI) supports the development and deployment of grid computing technologies. Managed by SURA (Southeastern Universities Research Association), the NMI Integration Testbed furthers the NMI's overall goal to create and distribute software that allows scientists and educators to share resources and collaborate effectively across the Internet. An important focus area of the NMI Integration Testbed has been the evaluation and integration of middleware to further grid infrastructure development.

In September 2003, SURA and the NMI Integration Testbed sites formalized the building of the NMI Testbed Grid, a grid infrastructure to represent the shared, collaborative and multi-institutional environment characteristic of higher-education research. This grid presents challenges and opportunities for NMI Integration Testbed sites to expand grid middleware services beyond project specific or local solutions commonly found in project-specific grid environments.

NMI component solutions are being investigated in a number of key areas: demonstration of grid-enabled applications, portal Implementation and evaluation, identification and deployment of shared resources, and policy for a shared grid.

Authentication and authorization issues are also being explored through development of a "BridgeCA:" that enables cross-certification between institutional domains.

Facilitating communication between researchers, faculty, and grid technologists is a focal point of the NMI Integration Testbed. Testbed sites are building a robust, user-friendly, and hands-on grid environment for use in outreach efforts to educate scientists and educators about grid computing. To successfully introduce the grid to researchers and to inspire them to try the grid and grid-enable their applications, the project is showcasing real applications benefiting from grid technology.

For more information about the NMI Testbed Grid, including how to join, please contact either of the following individuals:
Mary Fran Yafchak at maryfran@sura.org or Art Vandenberg at avandenberg@gsu.edu.

NMI Components Highlighted in this Case Study

The NMI components discussed in this case study series encompass NMI Releases 1 through 4. Information about NMI Releases can be found at **http://nsf-middleware.org/**.

GridPort

The GRIDS Center's GridPort enables the development of portals and applications on top of underlying distributed and grid computing infrastructure to facilitate computational science.
Home site: http://gridport.net/index.cgi

KX.509/KCA

NMI-EDIT's KX.509 and KCA provide a bridge between a Kerberos and PKI infrastructure. These tools enable the PKI-based security infrastructure of the Globus Toolkit to integrate with Kerberos-based authentication implemented at university campuses.
Home site: http://www.citi.umich.edu/projects/kerb_pki/

MyProxy

The GRIDS Center's MyProxy is a credential repository for the Grid. MyProxy provides a set of flexible authorization mechanisms for controlling access to the repository.
Home site: http://grid.ncsa.uiuc.edu/myproxy/

Exploring Inter-Institutional Grids: NMI Testbed Grid

The work of grid computing technologists, like the work of the users whose research efforts they enhance and support, is based upon growing a body of knowledge. Central to this learning process is applying new knowledge to solve very real problems. Solving one problem, however, often presents researchers and technologists with new challenges – and solving those challenges can further push the envelope of progress. A case in point is the National Science Foundation Middleware Initiative (NMI), established in September 2001. The NMI-Enterprise and Desktop Integration Technologies (NMI-EDIT) Consortium [1] of Internet2, EDUCAUSE and SURA (Southeastern Universities Research Association) were awarded one of the first of two integrator cooperative agreements in 2001. SURA's primary role in this is the development and management of the NMI Integration Testbed program (http://www1.sura.org/3000/NMI-Testbed.html). This program would provide "real life" evaluation and feedback on middleware components, thus furthering the NSF's overall goal of creating and distributing software that allows scientists and educators to share resources and collaborate effectively across the Internet.

From their efforts in implementing and evaluating NMI components, NMI Integration Testbed sites have produced a body of knowledge and experience that informs future real-world middleware deployments. An important focus area of that work has been middleware that furthers grid development. As the NMI Integration Testbed program progressed, the growing importance of grid technology to benefit academic and research applications was evident. Many current grids are established to support a regional, national, or inter-national project. Often, researchers participating in these projects must acquire separate credentials, usually in the form of a digital certificate, to authentication for a specific grid.

SURA and the NMI Integration Testbed sites realized they could help take middleware-enabled infrastructure to the next level by not only exploring key cross-organizational issues in sharing resources among a group of cooperating institutions, but also by leveraging enterprise (and, in the longer term, authorization) credentials to authenticate for grid-aware resources. For the researcher, this would greatly simplify the access and use of grid technologies and help to seamlessly integrate their everyday campus work with their research activities. In September 2003, SURA formally initiated

[1] http://www.nmi-edit.org/

the NMI Testbed Grid (http://www1.sura.org/3000/SURAgrid.html).

NMI Tested Grid participation began with nearly all of the NMI Integration Testbed sites:

- University of Alabama at Birmingham
- University of Alabama in Huntsville
- Georgia State University
- University of Michigan
- University of Southern California
- Texas Advanced Computing Center (TACC), UT Austin
- University of Virginia

The participants identified a need for fuller exploration of grid deployment in a shared, collaborative, multi-institutional environment that is so characteristic of higher education research. Equally important, they sought a robust, user-friendly, and hands-on grid environment to use in outreach efforts to educate scientists and educators about grid computing. This article takes a closer look at how and why the NMI Testbed Grid was formed, its progress to date, and future work.

Lots of Grids, So Why Another Grid?

Grids deployed for "big science" or project specific applications usually have a targeted scope and group of participants. Even though the participants may reside at different institutions, the nature of the project specific grid solutions tends to reflect a single enterprise-like solution, rather than a more complex, inter-enterprise solution. For instance, authentication may be project specific, specially created and defined for project members, regardless of what authentication they may already use in their home institution. Resources for such project specific application grids may be closely held and managed for well-defined needs of the project with less impetus or requirement to address variation. The resulting grid infrastructures are naturally more narrowly focused, since it is the scientific goals and project specific applications that are the driving purposes of the work. Such grids have been pioneering efforts in applying grid technology and concepts, but they may not provide an optimal environment for exploration of deeper issues in authentication and authorization necessary for scaling grids as a general purpose, cross-organizational infrastructure.

NMI Integration Testbed sites are familiar with project specific grid environments, with researchers at several institutions actively involved in national and even international grids of this type. One goal of the NMI Integration Testbed program has been to learn from the experience of these current "real life" deployments and then expand grid middleware services beyond local or project specific solutions. "Before NMI, middleware was in danger of becoming 'balkanized,' with many differing research communities developing independent - and often incompatible - solutions to similar problems of interoperability and resource sharing," according to Kevin Thompson, NSF program director for NMI. "Now, by creating

production-quality middleware using open-source and open-standards approaches, NMI-sponsored projects avoid duplication of effort and provide a common foundation on which varied communities may build their own customized applications." (1)
To support the NMI's interoperability and resource sharing goals, the NMI Integration Testbed team wanted to model a real world grid environment with the creation of the NMI Testbed Grid – a grid that could meet the "need to integrate services across distributed, heterogeneous, dynamic 'virtual organizations' formed from the disparate resources within a single enterprise and/or from external resource sharing and service provider relationships." (2) Several NMI Integration Testbed sites agreed to collaborate on building a grid infrastructure that spans organizations, including clusters of computing resources contributed by different sites, different approaches to local campus authentication, a variety of applications to be grid-enabled, and a range of interested faculty and technology staff represented. By taking this approach and working with the resources at hand, the NMI Testbed Grid is developing a working higher education grid environment in which faculty, researchers and students can form "virtual organizations" and share distributed resources.

The NMI Testbed Grid models the more complex, inter-enterprise environment that challenges the functionality of various grid components. For instance, authentication and authorization for the grid may be supported by NMI components, such as NMI-EDIT's KX.509/KCA[2] that integrates campus Kerberos authentication with grid certificates, or the GRIDS Center's[3] MyProxy[4], which provides a credential store for certificates improving access to credentials from any place on the network. These solutions, as well as new ones (cf. BridgeCA: Cross-certification between Testbed Sites) are being investigated in the cross-organizational NMI Testbed Grid. Using solutions such as the Open Grid Computing Environment's (OGCE) (3) portal, creating a portal for the NMI Testbed Grid raises interesting problems in presenting resources from different sites with varying levels of access. For instance, one may need to determine how the portal will handle authentication from campuses that have different enterprise sign-on models, how MyProxy will adapt to such an environment, or how applications can easily access and combine computing resources from multiple organizations. The technical challenge in providing this level of access is multifarious, while the policy issues that the NMI Testbed Grid is exposing present additional investigation opportunities. Each NMI Testbed Grid site brings new resources, new requirements, and its unique perspective and application needs to this grid, creating a distributed and diverse grouping that informs and catalyzes collaborative development.

[2] KX.509/KCA information: http://www.citi.umich.edu/projects/kerb_pki/
[3] GRIDS Center: http://www.grids-center.org/
[4] MyProxy information: http://grid.ncsa.uiuc.edu/myproxy/

Need for Hands-On and Application Focus

How does the researcher envision their application running in a "grid-enabled way" if they are not grid savvy, let alone information technologists? There is no single book one flips open to fully answer questions about the scope of use and value of this cutting edge grid technology. Conversely, how do information technology support staff, who are not application domain experts, customize a grid computing environment to address the very specific, and frequently very complex and demanding, needs of a scientific application? How can researchers and technology staff work together to answer these questions so each may influence the other? It was clear to participants that the NMI Testbed Grid could potentially provide a forum of growing expertise in grid infrastructure, benefiting both academic research and technology infrastructure communities.

The NMI Integration Testbed Program site representatives have seen a distinct growth in interest in grid technology during the Testbed project, fostered in part by outreach activities undertaken as part of the NMI Integration Testbed program. These outreach activities provided valuable lessons to sites building the NMI Testbed Grid. To successfully introduce the grid – inspiring researchers to try the grid and grid-enable their applications, the project needed to showcase real applications benefiting from grid technology. It would not be enough to inform their users that there was a new Beowulf cluster available via the NMI Testbed Grid; they should demonstrate the value of grid-enabling applications within the context of the researcher's *application* domain– how could their application be adapted to run on the grid, how would users of the application gain access (portal?), what kinds of improvements could be expected?

Art Vandenberg, NMI Integration Testbed project lead at Georgia State University, is spearheading the NMI Testbed Grid effort to demonstrate the potential of grids for various applications and illuminate the concept of grids as infrastructure for researchers and faculty. This includes:

- Development of a Catalog of Grid Applications http://art12.gsu.edu:8080/grid_cat/index5.jsp
- Identification and deployment of resources to be shared
- Demonstrations of grid-enabled applications (such as Georgia State's improvements in genome alignment algorithm performance using the grid)
- Implementation, evaluation and (where applicable) merging of portal solutions

NMI components will be integral to the NMI Testbed Grid's ability to create a grid spanning different organizations. Several NMI Testbed Grid sites (see Appendix) are combining resources to implement the initial working grid with a user-friendly "face," to

make grids something researchers will be able to see, in the spirit of "a grid node with a portal is worth a thousand words."

Progress to Date

Through their collaborative efforts, the NMI Testbed Grid has progressed from goal and strategy planning to active grid building in a series of overlapping phases.

Ø Startup

The project began formally in September 2003 with preliminary discussions between the participating sites and the SURA NMI Integration Testbed Program Manager, Mary Fran Yafchak. Consensus was reached that the Testbed Grid would be a means of capturing the growing interest and expertise in grids into deliverables that could make a recognized difference.

Ø Vision & Plan Definition

From team discussions during the period from October 2003 through January 2004, the project leads formulated a vision for the project and developed specific plans and timelines to meet the goals.

Ø Assembling the Grid

By January 2004, the work of building the NMI Testbed Grid was ready to get underway. From January through June 2004, cataloging grid applications and resources was conducted. Project leads also worked with participating site representatives to examine and begin testing the technical and policy issues in inter-institutional authentication and authorization.

Ø Outreach and Expansion

In July 2004, with the deployment of the NMI Testbed Grid well under way, the project leaders and Testbed manager began reaching out beyond the Grid's current site membership. These outreach efforts were focused on educating and inspiring others through presentations and demonstrations in multiple venues, including North Carolina Grid Initiative Forum (October 2003), GlobusWorld (January 2004), Internet2 Members' meetings (April 2004, September 2004), the Oklahoma Supercomputing Symposium (October 2004) and Supercomputing2004 (November 2004). The project team also put out a call for participation through the SURA IT Committee[5]; several institutions responded and have begun collaborating[6].

Key Aspects of Implementation

Catalog of GRID Applications

The catalog began as a Georgia State survey of its grid potential. Two NSF Research Experience for Undergraduates students (Nicole Geiger, Physics, and Anish Shindore, Computing Information Systems)

[5] The SURA IT Committee is a representative body of the CIO or equivalent from all 62 SURA member institutions. Complete member list at: http://www1.sura.org/1000/1150_Members.html

[6] New institutions (as of October 2004) exploring participation options within the NMI Testbed Grid are: University of Arkansas, Duke University, George Mason University, Georgia Tech, Louisiana State University, University of South Carolina, Texas Tech University, and Tulane University.

reviewed faculty researcher web pages to identify existing computation or data intensive research. Encouraged by the NMI Integration Testbed Program Manager, they eventually extended this survey to include 433 researchers from 17 institutions and 2 national labs (as of August 2004) – and work continues.

The Catalog provides a promising resource of existing grid projects as well as researchers with computing or data needs that may benefit from grid technology. For instance, Tulane University, one of the new sites joining the NMI Testbed Grid, is interested in computational fluid dynamics applications on a grid. A keyword query on "computational fluid dynamics" returns 33 references to researchers doing such research, among whom are Dr. Juan Cabral at George Mason University, another new collaborating site.

Demonstration of Grid-Enabled Applications

One goal of the NMI Testbed Grid is to demonstrate the value of grid-enabled applications. Nova Ahmed, a Computer Science Masters student at Georgia State (now a Ph.D. student), began working with the University of Alabama at Birmingham and later with additional NMI Testbed Grid sites to develop and deploy a genome alignment application using the grid. Ahmed's application quickly brought several high priority issues for the NMI Testbed Grid to the forefront, providing an opportunity to work on authorization and authentication issues through certificate testing, as well as resource policy issues (see following policy discussion). In addition, Ahmed's application work has been a means of inspiring faculty at Georgia State to use grid technology. Ahmed's advisor, Dr. Yi Pan, is working with Ahmed to extend research in using multiple heterogeneous clusters across different institutions in genome alignment.

Portal Implementation and Evaluation

The NMI Testbed Grid's portal work began with a comparison of the portals in development at various sites. Uniting the grouped resources via a single portal or creating a "portal of portals" are options to be considered. Initial work compared the GRIDS Center's GridPort[7] and OGCE, with current efforts focusing on OGCE. Collaboration is important - TACC and Georgia State combined the efforts of their NSF Research Experiences for Undergraduates (REU) and used a series of conference calls to help Georgia State in their portal development. The NMI Testbed Grid members also discuss portal progress on monthly conference calls, each contributing insights and sharing their portals with others. This initial work on individual portals will be important in understanding problems and developing a portal specific to the NMI Testbed Grid environment.

[7] GridPort information: http://gridport.net/index.cgi

Identification and Deployment of Shared Resources

Each NMI Testbed Grid site has contributed hardware, software, and/or labor resources to this project. (See Appendix for listing of resources as of this writing.) Consideration is being made to include this information either via the Catalog of Grid applications or a similar online means.

Policy for a Shared Grid

In their work to deploy NMI Testbed Grid demonstration applications (e.g., the genome alignment described above) the Testbed Grid members encountered policy issues regarding the sharing of various resources between their organizations. The nature of grid infrastructure – the sharing of resources between disparate and seemingly separate organizations – is a revolutionary way for organizations to both support and use applications. Policies have generally not been written to address the new type of sharing that takes place on grids. When approached by Georgia State's Nova Ahmed, seeking to use computing resources, Jim Cotillier, of the University of Southern California, observed, "We've never had a request like this [from non-USC user]." Fortunately, Cotillier is helping Georgia State define the resource requirements and also guiding the application request process for appropriate use of USC's resources. In part because of this experience, NMI Testbed Grid project members are reviewing existing policies in order to understand them better, to develop best practices and offer insights in policies for grid-based resource sharing.

BridgeCA: Cross-Certification Between Testbed Sites

Marty Humphrey & Jim Jokl at the University of Virginia are focused on a critical next step to move from "a grid" to "the Grid'" – cross-certification to insure secure access to resources between organizations, enabling large-scale deployment. Cross-certification has been implemented in a series of steps that exercise functionality while illuminating new needs and directions. This began with the establishment of the Testbed Bridge CA, a prototype bridge CA, at UVa in February 2004. This was followed by the simulation of two campuses within UVa to test viability, specifically with regard to current NMI component levels. Upon verification the next month, UVa removed the simulated environment and went "live" by cross-certifying with the University of Southern California.

With the completion of that successful "real life" trial, UVa began pursuing cross-certification among additional Testbed Grid sites, seeking to increase both the amount and complexity of participating institutions. To assist in this process, in March 2004 UVa established a web page that provides Testbed Grid sites with a documented process for cross-certification. In April 2004,Testbed Bridge CA cross-certification was successfully completed with UAB and TACC. Further tests to exercise the

BridgeCA and run it elsewhere, using local CA certificates (and possibly commercial certificates) are planned.

Future Expansion

The NMI Testbed Grid will continue as part of the NSF-funded NMI Integration Testbed Program through February 2005. The intent beyond that is to expand the infrastructure, applications and collaborators as an evolving and self-sustaining activity.

Offering a collaborative forum and demonstration mechanism for others who are considering grid technology is a primary goal of the NMI Testbed Grid. Project members are planning and conducting outreach through various means. Members participated in strategic regional planning meetings with SURA as early as January 2004, leading to the spin-off of several SURA working groups in key aspects of grid development. The July 2004 call for participation continued this regional influence, with a number of new institutions joining the Testbed Grid thereafter. In addition, the Testbed Grid manager has received inquires from various other organizations interested in collaborating on development of this type.

With the growing number of inquires to join, the project team is considering a more formal "intake process" – one that would add to the growing application catalog, establish a policy for shared resources and facilitate immediate access to peer support for grid and application deployment. The process would also provide input on best practices for how to help organizations understand grid computing and provide a roadmap for "getting on the grid".

More Information

For more information about the NMI Testbed Grid and joining it, please contact either of the following individuals:

Mary Fran Yafchak, NMI Integration Testbed Manager, SURA, at maryfran@sura.org (315.593.0718) or

Art Vandenberg, NMI Integration Testbed Project Lead, Georgia State University, at avandenberg@gsu.edu (404.463.9685).

References

1) *About the NSF Middleware Initiative*, NMI Middleware Initiative. http://www.nsf-middleware.org/about_NMI/

2) *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*, I. Foster, C. Kesselman, J. Nick, S. Tuecke, Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002. http://www.globus.org/research/papers/ogsa.pdf

3) OGCE: Open Grid Computing Environment, OGCE home page, The Open Grid Computing Environments Collaboratory. http://www.collab-ogce.org/nmi/index.jsp

Appendix

Table 1 Current Participants and Components (Planned and Existing) as of October 2004
(*Italics* indicate new sites joining since the July 26, 2004 "Invitation to Participate" from SURA)

SITE	"On-Grid" Resources	GT v.	Apps	CA	Portal	KX.509	Shib	REU
UAB	-Beowulf Cluster -HPC Cluster -SMP -Condor Pool	GT 2.4.3 GT 3.0.2 NMI R4	BLAST	Bgrid CA & BridgeCA	Y OGCE	N	Y	N
UAH	---	---	---	---	---	N	N	N
UARK	Beowulf Cluster	---	---	---	---	N	Y	N
DUKE	---	---	---	---	---	N	N	N
GMU	CPU Cluster Weather data	---	Geospatial	---	---	N	N	N
GSU	Data Cluster	3.0/2.4	Genome alignment	GeoTrust certs	Y OGCE	N	Y	Y
GT		---	I2 End-to-End Performance	---	---	Y	N	N
LSU		3.2.1	Cactus Apps	CA	Y	---	N	---
UMICH	CPU Clusters: • PBS Queue • Condor Pool	3.0/2.4	• Grid3 • ATLAS G4Sims	KX.509 DoE Grid	Y	---	N	Y
SC	CPU Cluster Viral genomes	---	---	---	Y	N	N	N
USC	Linux Cluster	---	N	CA & BridgeCA	---	Y	Y	N
TACC	CPU Cluster	GT3.2 & GT3.0.2	Computa-tional Fluid Dynamics	CA & BridgeCA	Y OGCE	N	N	Y
TTU	Cluster	---	Math & Statistics	---	---	N	N	N
TULANE	Cluster	---	---	---	---	N	N	N
UVA	Cluster	---	---	CA & BridgeCA	---	Y	N	Y