Scalable Computing: Practice and Experience
Volume 7, Number 3, pp. 95–108. http://www.scpe.org
ISSN 1895-1767
© 2006 SWPS

THE GREAT PLAINS NETWORK (GPN) MIDDLEWARE TEST BED

AMY W. APON[*], GREGORY E. MONACO[†], AND GORDON K. SPRINGER[‡]

Abstract. GPN (Great Plains Network) is a consortium of public universities in seven mid-western states. GPN goals include regional strategic planning and the development of a collaboration environment, middleware services and a regional grid for sharing computational, storage and data resources. A major challenge is to arrive at a common authentication and authorization service, based on the set of heterogeneous identity providers at each institution.

GPN has built a prototype middleware test bed that includes Shibboleth and other NMI-EDIT middleware components. The test bed includes several prototype end-user applications, and is being used to further our research into fine-grained access control for virtual organizations. The GPN prototype applications and namespace form a basis for the design and deployment of a robust and scalable attribute management architecture.

Key words. Middleware, Shibboleth, NMI-EDIT, Signet, Grouper, Grid computing

1. Introduction and Background.

The Great Plains Network (GPN) [1] is a regional consortium of public universities in the states of Arkansas, Kansas, Missouri, Nebraska, Oklahoma, North Dakota, and South Dakota and regional higher education state networks in these states. Member representatives have recognized the strategic importance of sharing resources collaboratively and shared the goals of national efforts, such as the NMI (NSF Middleware Initiative), NMI-EDIT (NSF Middleware Initiative - Enterprise Desktop and Integration Technologies) [2] and Shibboleth [3], devoted to facilitating inter-institutional collaboration. A long-term goal of GPN is to build a regional middleware infrastructure to share existing grid computing resources (computation, storage, data, and applications) across the region and to provide a platform for the development of new tools and technologies.

In June, 2004, GPN was selected to be one of four projects funded by the Extending the Reach (ETR) program. Participating institutions include the University of Arkansas, the University of Missouri, the University of Oklahoma, South Dakota State University, the University of Kansas, the University of Nebraska-Lincoln, the Peter Kiewit Institute, North Dakota State University, the University of South Dakota, and the Great Plains Network Consortium. Community members include volunteers from both academic computing and research groups. The objectives of the ETR project are: 1) strategic planning on a regional level; 2) the development of a knowledge base, including the test bed installation and testing of middleware environments; and 3) educational outreach to participating institutions. One result of this project has been the development and deployment of a unique middleware test bed across GPN institutions that incorporates several NMI-EDIT software components.

1.1. The Challenge.

While national networking and networking-related initiatives such as the National Science Foundations Middleware Initiative (NMI) present new opportunities to improve network capacity, security and reliability for research activities, these initiatives also present challenges to the GPN consortium institutions. The GPN consortium includes campuses that are spread across the central United States, there is no hierarchical organization or a single central authority, and there are several heterogeneous approaches to the implementation of core middleware services across campuses. Unlike campus or state organizations in which policy can be established administratively, GPN is best able to influence regional decision making by calling attention to best practices. At the inception of this project, the proposed participants were at varying stages of planning and implementation for core middleware services. Several campuses were implementing either Lightweight Directory Access Protocol (LDAP) or Microsoft Active Directory systems, and one campus had developed its own identification and authorization system. In addition to technical accomplishments, this middleware project has been an example of successful voluntary collaboration among several institutions with a goal of human capacity-building, training and consulting, and shared resources across campuses and state networks [4, 5]

The remainder of this paper describes the GPN core middleware test bed components, primarily including Shibboleth, prototype resources that have been developed to support the GPN federation, the development of federation infrastructure, synergism with regional research groups and the broader middleware and grid computing community, and conclusions.

[*] University of Arkansas (aapon@uark.edu)
[†] GPN and Kansas State University (greg@greatplains.net)
[‡] University of Missouri-Columbia (springer@cecs.missouri.edu)

2. Shibboleth Middleware Component. Encouraged by the funding of the Extending the Reach Proposal in June, 2004, the GPN began developing a middleware test bed that included Shibboleth, integration with campus identity management systems, the development of prototype resources, and a namespace and attribute architecture to support resource sharing among members of the GPN institutions.

NMI-EDIT's Shibboleth is a project and software package from Internet2/MACE (Internet2's Middleware Architecture Committee for Education). It is a protocol and architecture for sharing attributes among trusted institutions, and is designed to authorize a user to a remote web-based resource through the use of the login and attribute information that is maintained at the home institution of the user. Shibboleth is an ongoing project. The current work by Shibboleth developers includes extensions to non-web-based resources and integration with other middleware tools for grid computing and attribute management [6].

Shibboleth protects resources in the same way that a user ID and password can protect resources. However, Shibboleth protection is based on group membership, or attributes, rather than on the identification of a particular user. Access to resources based on group membership is a common access control mechanism for many types of applications, including defined groups for file access in traditional Unix systems and integrated database applications or student information systems. Shibboleth adds to this capability the distributed management of the user attributes at home institutions, rather than the management of these through a central repository.

An example can illustrate the versatility and a usage scenario of Shibboleth. For example, suppose that the University of Missouri-Columbia (MU) would like to share access to its supercomputing cluster and terascale storage facility to authorized users at the University of Arkansas (Uark). Suppose that the potential users at Uark are members of the biomedical informatics research group and the resource is to be made available for a particular research project. With Shibboleth, users at Uark would login using the LDAP directory service at the University of Arkansas using their institutional login ID and password. The user attribute, member of the biomedical informatics research group would be retrieved from the Uark directory service and passed to the resource at MU. Through the Shibboleth infrastructure, the MU computing resource would acquire the attribute (e.g., group membership) that it needs from Uark. It would test the value of the attribute (e.g., biomedical informatics research group) to verify that the user has been allowed to access the resource, and then the user would have access. In this process the users identity information only needs to be maintained at the home institution. A centralized directory server for the whole GPN biomedical informatics virtual organization is not needed. Only the attributes, such as group membership and institution affiliation, that are required to access the resource need to be passed to the MU resource. Additionally, the Uark user has the option not to allow those attributes to be sent to MU. The administrator of the MU resource may configure the system to grant or deny access based on the access control policy that has been established.

Shibboleth allows the privacy of the user to be maintained if the administrator of the resource allows this. The user could desire this, for example, if the researcher is performing access on a large data resource for AIDS and the researcher does not want to reveal his or her personal identity as an AIDS researcher. Only the group membership information is sent, and the user's privacy is protected.

The architecture can allow members of a group to share a single account on a resource. The architecture also allows for a user identity to be passed to a resource that requires the use of a specific account for each user. For example, access to a computational resource may only be allowed to certain users that have already had accounts established with administrator approval. In this case a user's identity would be passed through the Shibboleth protocol and would be matched to the corresponding account on the resource.

Shibboleth consists of two primary software components, an Identity Provider and a Service Provider. The Identity Provider component of Shibboleth is integrated with the identity management system of a user's home institution. It authenticates a user using local authentication mechanisms, and then allows user attribute information that is maintained in an institutional identity directory to be sent to a requesting remote resource. The Shibboleth Service Provider software component protects access to a resource by remote users, and allows access only by users that meet the attribute requirements for using the resource.

The Shibboleth package relies on several underlying software servers and protocols for passing user authentication and attribute information between the Identity Provider and the Service Provider. Figure 2.1 illustrates the flow of the Shibboleth protocol. As shown in Figure 2.1, a user first requests a service, via a web browser, from a Service Provider (SP). The user may want to access a database or to submit a job to a computational resource, for example. The first phase of the Shibboleth protocol is the authentication phase. The SP redirects the request to a Shibboleth "Where Are You From" (WAYF) server, which in turn prompts the user to provide the name of a home institution. This request can be satisfied from a drop-down menu from which the user

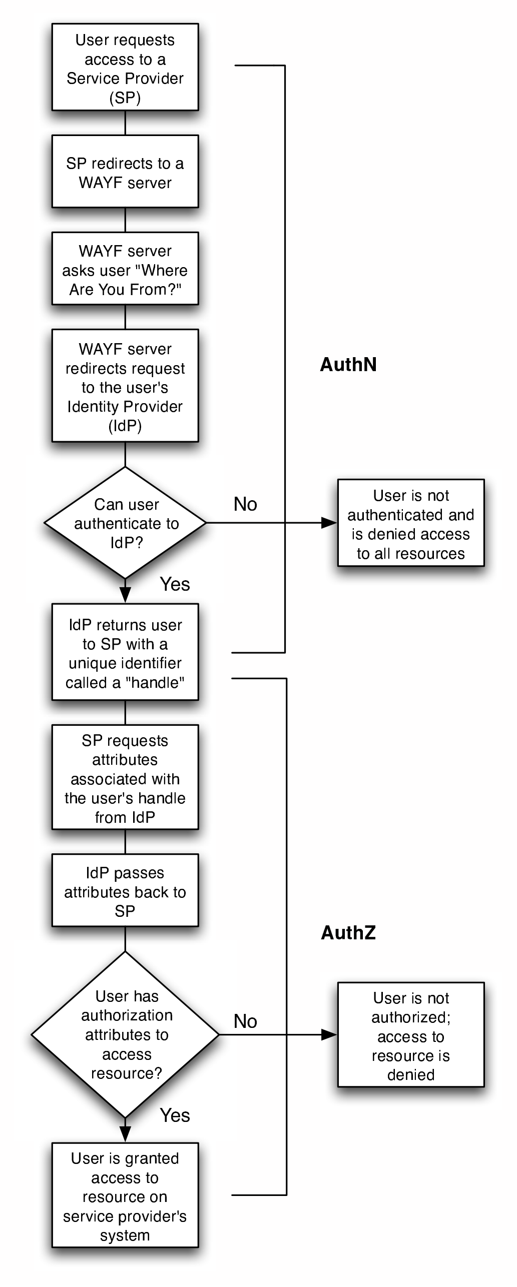

User requests access to a Service Provider (SP)
SP redirects to a WAYF server
WAYF server asks user "Where Are You From?"
WAYF server redirects request to the user's Identity Provider (IdP)
AuthN
Can user authenticate to IdP?
No
User is not authenticated and is denied access to all resources
Yes
IdP returns user to SP with a unique identifier called a "handle"
SP requests attributes associated with the user's handle from IdP
IdP passes attributes back to SP
AuthZ
User has authorization attributes to access resource?
No
User is not authorized; access to resource is denied
Yes
User is granted access to resource on service provider's system

Fig. 2.1. *Shibboleth Protocol*

makes a selection, for example, or software short-cuts can be coded into the system that provide the name of the home institution for certain services. After the home institution of the user has been identified the user's request is redirected to the Identity Provider (IdP) of the home institution. The IdP prompts the user to login using the user's home account and password. After this step, the user is authenticated and the Shibboleth system returns a unique "handle" that is used in the second phase of the protocol.

The second phase of the Shibboleth protocol determines what resources that the user is allowed to access. A distinctive feature of Shibboleth is the separation of the authentication and authorization steps. The Service Provider (SP) requests the necessary attributes from the home institution that are associated with the handle that has been provided in the authentication phase. The user can choose to release or not release attributes to particular requesting services. The IdP passes the attributes back to the SP. As a final step, the SP determines if the attributes that have been presented are sufficient to allow access to the resource that has been requested. The user is granted or denied access based on the attributes that have released from the home institution.

Shibboleth relies strongly on a trust relationship between the home institution of the user (the Identity Provider) and the institution that provides the resource (the Service Provider). The provider components must both own a public key certificate that is signed by a common Certificate Authority. In addition, the providers must also both be configured to accept requests from the other institution. For resource access to be granted to the user, the Identity Provider must accept requests from the Service Provider to release needed user attribute information. The Service Provider must be configured to allow access by users from the Identity Provider institutions that hold appropriate attributes,

The Shibboleth software package is layered on top of standard software environments. The Identity Provider is a Java web application that runs in the Tomcat servlet container over the Apache web server. The Service Provider is somewhat more complex, but uses standard C/C++ and XML based software components.

The Shibboleth Identity Provider is designed to be integrated with local campus directory and identity management systems. This task has been complicated within GPN by the heterogeneous approach to identity management on the various campuses. In general, campuses that have an LDAP-based identity management system and who were successful at building a preliminary installation of a Shibboleth Identity Provider have been successful during the first year of the project. This success includes integrating the Identity Provider with their local campus system.

3. GPN Prototype Resources. Several prototype resources have been established that allow testing of the middleware test bed. These include a biomedical application, a GPN data repository, and a web-based interface to a computational cluster. Each of these resources has one or more associated attributes (or entitlements) that a user must have to be allowed access. Currently the entitlements are configured using the eduPersonEntitlement field of the EduPerson schema [7], as discussed in Section 4.

3.1. Biomedical Application. In order to demonstrate the use of entitlements within the defined framework, a working research application in animal genomics was converted from its original user interface. The original interface used a Java applet to authenticate and authorize access to the research data by members of the research team. Using the Java applet along with a security database identifying users to be granted secure access to the data, members of the research team could access the data via a standard web browser. No access to the web site or the data is allowed except to the researchers. The user interface was custom designed and implemented for the protection of this data for the research team.

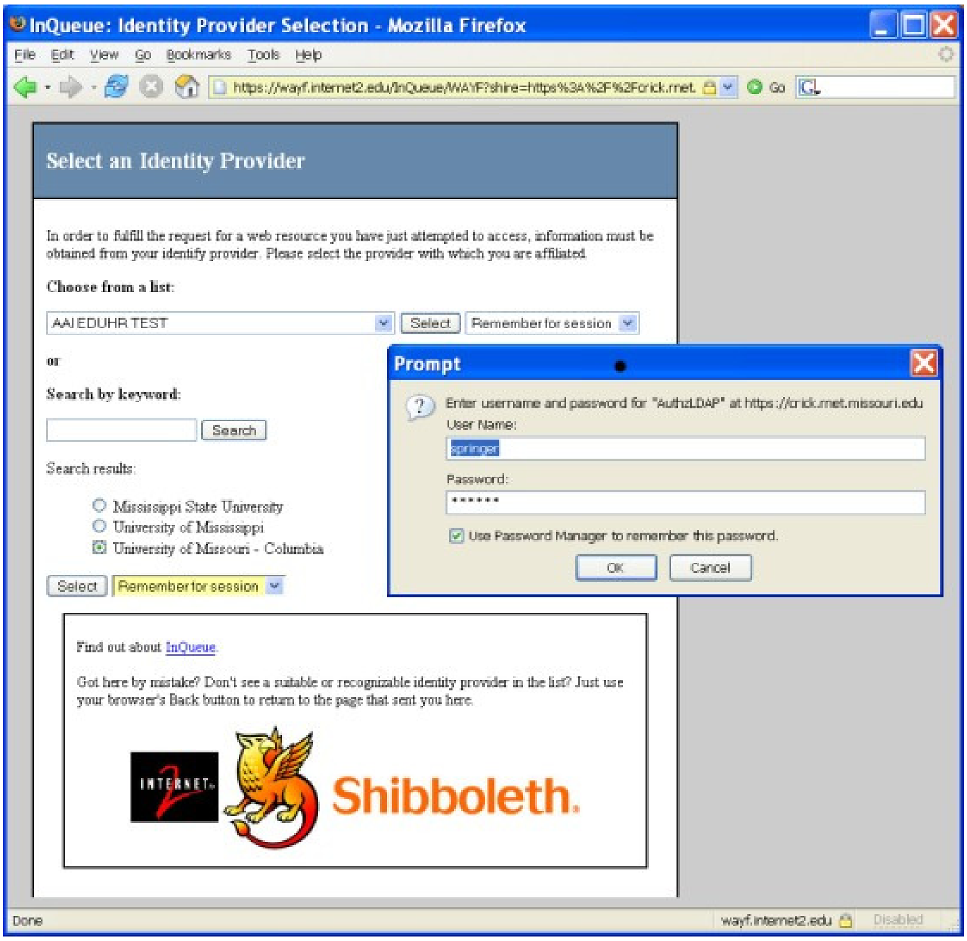
The application was converted to use the Shibboleth protocol for authentication and authorization by replacing the existing Java applet front end. A Shibboleth service web page was created and a link was added to the original application's main web page. This link makes a call to a customized Shiblogin command that internally logs the requesting GPN Shibboleth authenticated user into the protected genomics web site. This is accomplished without the user having to provide a separate login id and password to the application.

Shibboleth login pages for the biomedical application are shown in Figure 3.1. In this figure the user, springer, has selected the home institution, University of Missouri-Columbia, and has entered his password for his home user account. After selecting "OK", the Shibboleth protocol will authenticate user "springer" with the Identity Provider at the University of Missouri-Columbia. The Shibboleth protocol will also obtain this user's attributes (i. e., entitlements) from the Attribute Authority component of the Identity Provider at the University of Missouri-Columbia.

The Shiblogin command receives the Shibboleth generated credentials and entitlements that are verified in the login process. The software ensures that the credentials provided are from an active Shibboleth session and that the required entitlements authorizes the user to access the genomic data. No changes to the original application were required. An additional userid and password was added to the application's security database to grant access to the application by all GPN virtual organization members with the BioSci entitlement.

Within the genomic application, all of the Shibboleth credentials as well as the application's Shiblogin information are available for the application to use as needed. It is possible to provide even more fine-grained control of access. For example, attributes could be defined so that GPN users cannot change any data, but have read-only access to all of the research data.

3.2. GPN Repository. The GPN Repository is a data storage facility that permits members of the collaborating institutions to store and retrieve documents, data and other materials that are to be shared among all of the members. The GPN repository became available to users during the fall of 2004. As with the


InQueue: Identity Provider Selection - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
Go
Search
Select
Remember for session
OK
Cancel
INTERNET2
Shibboleth.
Done
Disabled

https://wayf.internet2.edu/InQueue/WAYF?shire=https%3A%2F%2Fcrick.met.

Select an Identity Provider

In order to fulfill the request for a web resource you have just attempted to access, information must be obtained from your identify provider. Please select the provider with which you are affiliated.

Choose from a list:

AAI EDUHR TEST

or

Search by keyword:

Search results:

○ Mississippi State University
○ University of Mississippi
● University of Missouri - Columbia

Prompt

Enter username and password for "AuthzLDAP" at https://crick.met.missouri.edu

User Name: springer

Password: ******

☑ Use Password Manager to remember this password.

Find out about InQueue.

Got here by mistake? Don't see a suitable or recognizable identity provider in the list? Just use your browser's Back button to return to the page that sent you here.

wayf.internet2.edu

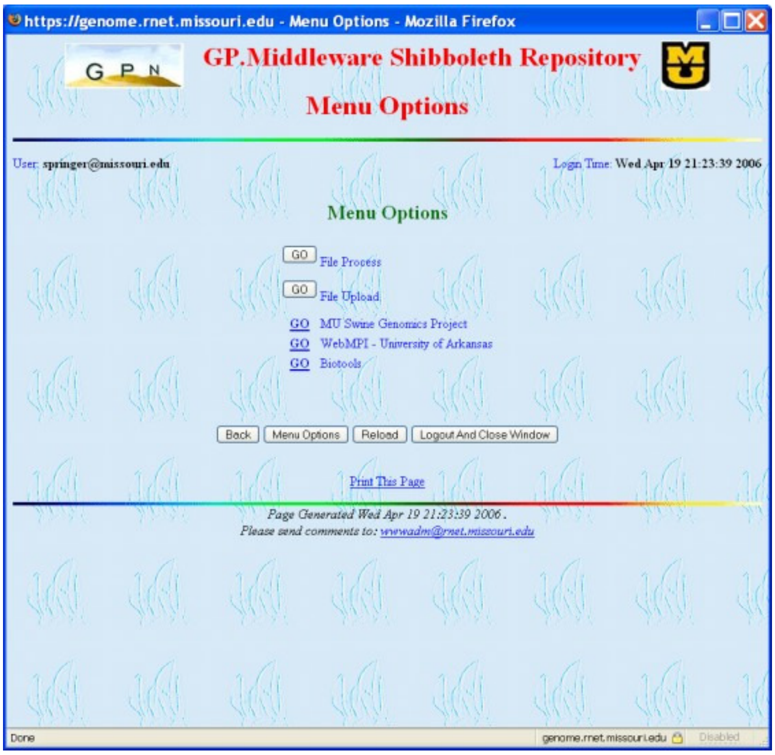
FIG. 3.1. *Shibboleth Login for the User "Springer"*

Biomedical Application, access to the GPN repository is protected by Shibboleth and user authentication to a GPN institution. Currently, access is provided to the GPN repository to any user who is a member of a GPN institution and for whom the Repository entitlement is returned by the campus Identity Provider.

Figure 3.2 shows the front page of the GPN Repository after user "springer" has logged in. It should be noted that the user only needs to log in one time using the Shibboleth protocol, and after that is provided access to several different resources, including the biomedical applications, the GPN repository, and the WebMPI application. Access is granted to the user "springer" for the applications for which his user account has valid attributes. The web page has been configured to allow the user to move easily between the repository and other applications, as shown in the figure. Also note that the network communication is secured via SSL (i. e., the lock on the status line is closed).

The GPN repository is motivated by the need to share large documents within the GPN. Frequently, and many times inappropriately, email (with attachments) is used to share materials among various groups of people. Some of these materials may be quite large and cause many problems such as overflowing mailboxes, causing data to be duplicated in a variety of locations unnecessarily, and can easily become outdated for volatile materials. In general, email is quite inefficient in utilizing resources among collaborators. The GPN repository attempts to overcome some of these inefficiencies by having the materials collected, organized and accessible in one (or more) locations with access restricted to the members of the GPN federation and without a lot of administrative overhead.

One of the uses of the GPN Repository has been to publish copies of the talks that were made at the annual GPN meeting in June, 2005, in Kansas City. The presenters provided copies of their talks to be made available to the GPN members. One presentation was in excess of 58MB and could not be sent to everyone as an email



https://genome.rnet.missouri.edu - Menu Options - Mozilla Firefox

GP.Middleware Shibboleth Repository

Menu Options

User: springer@missouri.edu

Login Time: Wed Apr 19 21:23:39 2006

Menu Options

Page Generated Wed Apr 19 21:23:39 2006 .

Please send comments to: wwwadm@rnet.missouri.edu

Fig. 3.2. *GPN Repository Page Showing User "springer" Logged In*

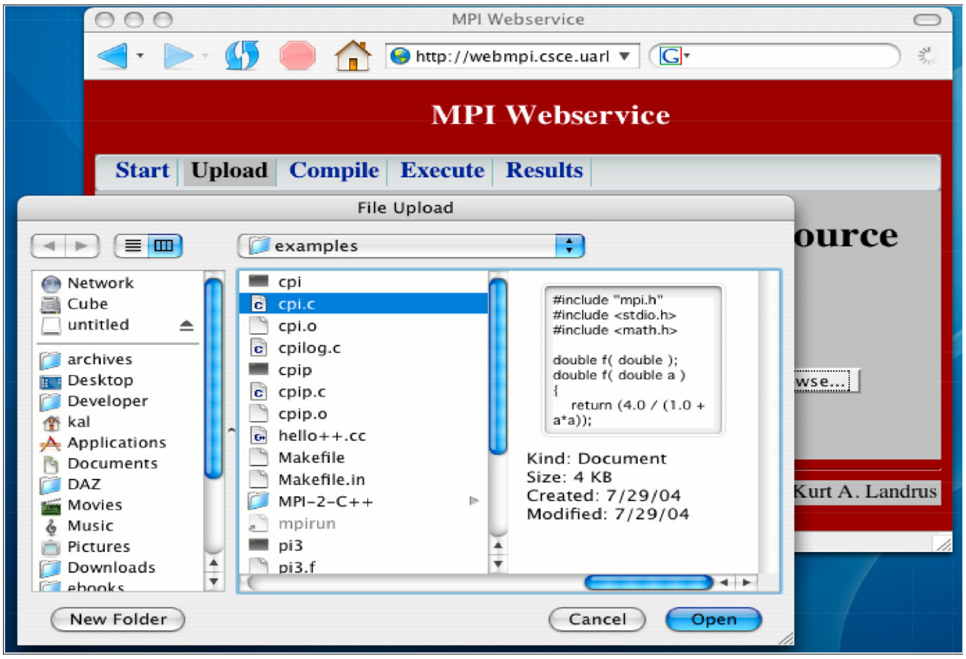
attachment. By placing one copy of the presentation in the repository, all GPN members have access to the presentation whenever they want, it is accessible only to the GPN members and no administrative overhead is necessary to provide this protection such as would be required to publish the document on a protected web site.

While the repository presently is used to house documents, presentations and other GPN materials, in the longer term the repository can also be used to provide high performance access to data used by application programs in Grid computing environments among the VO group servers at several institutions.

An ongoing and complementary project is the development of a web-based Subversion [8] document repository system that is also protected by Shibboleth. Subversion allows document check out and check in, and will allow the protection of document subdirectories based on user attributes.

3.3. WebMPI. WebMPI is a Shibboleth application that enables remote users to access a Linux cluster using a web browser for the purpose of running parallel applications. A typical user may be a student in a university course who is studying parallel programming using the MPI (Message Passing Interface) API. As with other Shibboleth applications, users contact the WebMPI interface, and then are redirected through the Shibboleth protocol to authenticate through the authentication mechanism of their home institutions. User attributes are passed to the WebMPI interface, which then determines if the user is authorized to use the cluster or not.

The current implementation of WebMPI is a prototype. The interface consists of a collection of HTML pages and CGI scripts that perform the user's commands on the underlying cluster resource. The interface consists of five main pages. The first page is an introduction page that provides some brief help information on MPI programming. The second page is a file upload page that allows users to browse to a directory on a local computer, select an MPI source file, and upload the file to the WebMPI cluster. The file upload page is shown in Figure 3.3. The third page is a compile page that allows the user to select a file and compile it using either the C or Fortran compiler with MPI libraries. The fourth page is an execute page that allows the user to select a compiled MPI program and to execute it with additional user parameters. The final page is a results that page allows the user to view the output of the MPI application. The current system is convenient for student programs and other types of small demonstration programs. However, it is limited for large production scientific applications. For example, it does not currently have the capability to manage multiple large input or output files.



```
#include "mpi.h"
#include <stdio.h>
#include <math.h>

double f( double );
double f( double a )
{
   return (4.0 / (1.0 +
a*a));
```

Kind: Document
Size: 4 KB
Created: 7/29/04
Modified: 7/29/04

FIG. 3.3. *WebMPI File Upload Page*

One of the issues encountered in the development of WebMPI that does not come up in the same way for resources such as the GPN Repository is how to handle processing on the underlying cluster. In a normal MPI application, a user logs in to the head node of the Linux cluster using a local account and password. Any new files are created in the user's home directory and all processes that are created run under the user's account. The underlying Linux operating system protects the computer system from a stray process that creates too many files, or that creates files that are too large, or that consumes too much CPU time. In contrast, one of the features of Shibboleth is that it allows the protection of user privacy. While it is possible to pass a user name or account within the Shibboleth architecture when it is desired to do so for a particular application, Shibboleth works by not revealing the user's name or account in general. Access is typically based on group membership, or the possession of a particular attribute. With WebMPI this is a challenge in two ways. First, there needs to be a way to map an incoming user to a subdirectory on the underlying cluster, and secondly, there needs to be a way to map an incoming user to an account in which processes may execute. This is solved in a typical grid application that uses Globus Toolkit [9] by creating an account on the machine and then mapping the user's credentials (i. e., the Distinguished Name) to that account through the grid-mapfile. A Distinguished Name is not passed to the Service Provider by the Shibboleth protocol in general.

One of the attributes that can be maintained by the Identity Management system at the home institution is the PersonID that is a part of the eduPerson schema. The PersonID is an opague identifier that is guaranteed to be unique for a person. WebMPI maintains subdirectory integrity by mapping the PersonID to a subdirectory on the underlying cluster. By using a mapping in this way, a user may authenticate a second time and be mapped to the same user space representing the same subdirectory. For example, a user is able to start a long-running MPI application, and then return at a later time to view results. Note that for a user to be authorized to use WebMPI, the Identity Provider at the home institution of the user must release the PersonID attribute to the WebMPI resource.

User-level MPI processes are handled in WebMPI by executing all MPI applications in a single account named the webmpi account. In the interface, a user requests the execution of an MPI program. Since WebMPI is a web application, the suEXEC feature of the Apache web server can be used to execute the MPI program in the webmpi account on behalf of the user. Results are maintained in the appropriate subdirectory. This technique is secure in that only users who authenticate through the Shibboleth interface are allowed to execute programs; however, it does not allow the logging of usage of individual users. Other solutions to the process mapping problem are being explored, including integration with GridShib [10].

4. GPN Federation Infrastructure. The Shibboleth middleware component and the prototype applications are supported by several infrastructure components and design approaches in the GPN federation. The supporting infrastructure includes federation server support, structured management of attributes, the greatplains.net MACE namespace, and management of entitlements.

4.1. Federation Server Support. Early in the project, GPN institutions joined InQueue [11], a test federation organized by Internet2 for institutions who are learning how to use Shibboleth and the federated trust model. A federation organization supports the GPN federation process by providing two different types of servers and services. First, GPN members of InQueue initially configured their Identity Provider with a public key certificate from the Bossie Certificate Authority recommended by InQueue since this was easy and at no cost. The Bossie CA is not secure or suitable for production use, but does provide a simple and fast way to build a working prototype Shibboleth system. As GPN members have moved toward production use of their Identity Provider, the Bossie certificates have been replaced by certificates from Verisign and other production-quality CA's. InCommon [12] is a formal federation that has been developed to support a production CA and production use of the Shibboleth architecture. In general, in order for the identity and resource providers in GPN to continue to trust each other, they have to be configured with the public key of each CA that is in use and is trusted in a production setting by the GPN members.

A second server that a formal federation organization supports the GPN consortium is the WAYF, or a "Where Are You From?" server. In the Shibboleth protocol, when a user first contacts a Resource Provider through a web page, the Resource Provider must determine the home institution of the user for that person to be authenticated. The Shibboleth protocol redirects a request to the WAYF named by the Resource Provider. The WAYF, in turn, maintains a list of potential Identity Providers, allows the user to select one, and then redirects the user to the authentication process of the Identity Provider selected by the user.

The WAYF in InQueue currently supports dozens of participating organizations, which is distracting when trying to do testing with GPN resources. In addition, the InQueue WAYF does not meet the needs of GPN for the testing and configuration of the middleware test bed. As attributes have been tested, for example, two Identity Providers have been maintained at the University of Arkansas. The first remains integrated with the campus Identity Management system and is visible to all campus users, but the second is configured in a test mode to only allow access by certain test user accounts. GPN has built a test federation by implementing a test WAYF that lists the test Identity Provider. In turn, the resources also being tested are configured to redirect users to the test WAYF. This lightweight test environment provides a mechanism for introducing new resources and new types of attribute access without interfering with production use of Shibboleth on some campuses. As the middleware infrastructure of GPN grows the components in the lightweight test environment will need to be replaced by production versions of these components.

4.2. Structured Management of Attributes. As the resources in the GPN federation have become more complex, the need has arisen for a structured way of managing the attributes that are required for access to these resources. The EDUCAUSE/Internet2 eduPerson Task Force has defined a data structure (object class) that defines attributes that are useful for individuals in higher education. Among the attributes are an individual's institutional affiliation, and their relationship to the institution, such as faculty, student or staff. At a very high level, this provides a coarse-grained means for distinguishing groups within an institution, such as all faculty members or all students enrolled in a particular class in a particular semester. At this level, decisions about authorizations to utilize various services at an institution can be readily made.

The eduPerson Task Force identified the syntax and semantics of these attributes. The development of this object class is now managed by the MACE-Directory Working Group, which is encouraging wide-spread adoption of the attributes among institutions of higher education. While these attributes are common across institutions and as a proposed standard the defined object class is or can be quite useful to enable a wide variety of applications and services both within an institution and external to it.

The GPN group, upon reviewing the attribute fields, has a need for more fine-grained controls over authorizing access to specific, shared, collaborative resources and services that span several institutions among the Great Plains states. The values for attributes for most fields in the eduPerson object class are single-valued and/or predefined (e.g., faculty, staff, student, or member).

To accomplish the goals of the GPN middleware project, an attribute that can take on various values and, in fact, be multi-valued is desired. One field that has the required properties is the eduPersonEntitlement field. Thus, in these first steps of the middleware project the GPN group has used the eduPersonEntitlement field to define the necessary values that facilitate the need for fine-grained decision-making when authorizing access to inter-institutional, collaborative resources. Use of the entitlement attribute in combination with other eduPerson attributes sets the stage for fine-grained control of authorizing access to specific resources and services.

In the current implementation the entitlements are defined as LDAP attributes using the eduPersonEntitlement definition. The entitlement attribute permits multiple values. The entitlement attribute is a semicolon separated string containing MACE registered values that are asserted and verified during the authorization process to grant or deny access to an entitlement dependent resource.

Shibboleth processes these attributes at both the Identity Provider (IdP) and the Service Provider (SP). The simplest way to release the attributes by the IdP is to use the sample attribute release policy which releases specified attributes to all requesters. However, it is possible to release attributes to some, but not all institutions. The SP must define what attributes it will accept, and from whom. In other words, an attribute acceptance policy might allow all GPN institutions to assert eduPersonEntitlement, but no others. Shibboleth by default disallows any institution from asserting attributes scoped to another institution. Also, an IdP may choose to only release certain values of an attribute to a particular SP. For example, a GPN SP might only be able to see GPN-related eduPersonEntitlement values, and not those related to other organizations.

In the Apache configuration for the SP, the required attributes are defined, and if the "ShibRequireAll" directive is specified, all attributes must be present. Otherwise, the default Apache behavior is to authorize access if any one of the attributes matches. These can be matched exactly, or a regular expression may be used to match a range of acceptable values. Note that if regular expressions are used, extreme caution should be paid to ensuring that unwanted matches do not mistakenly get accepted.

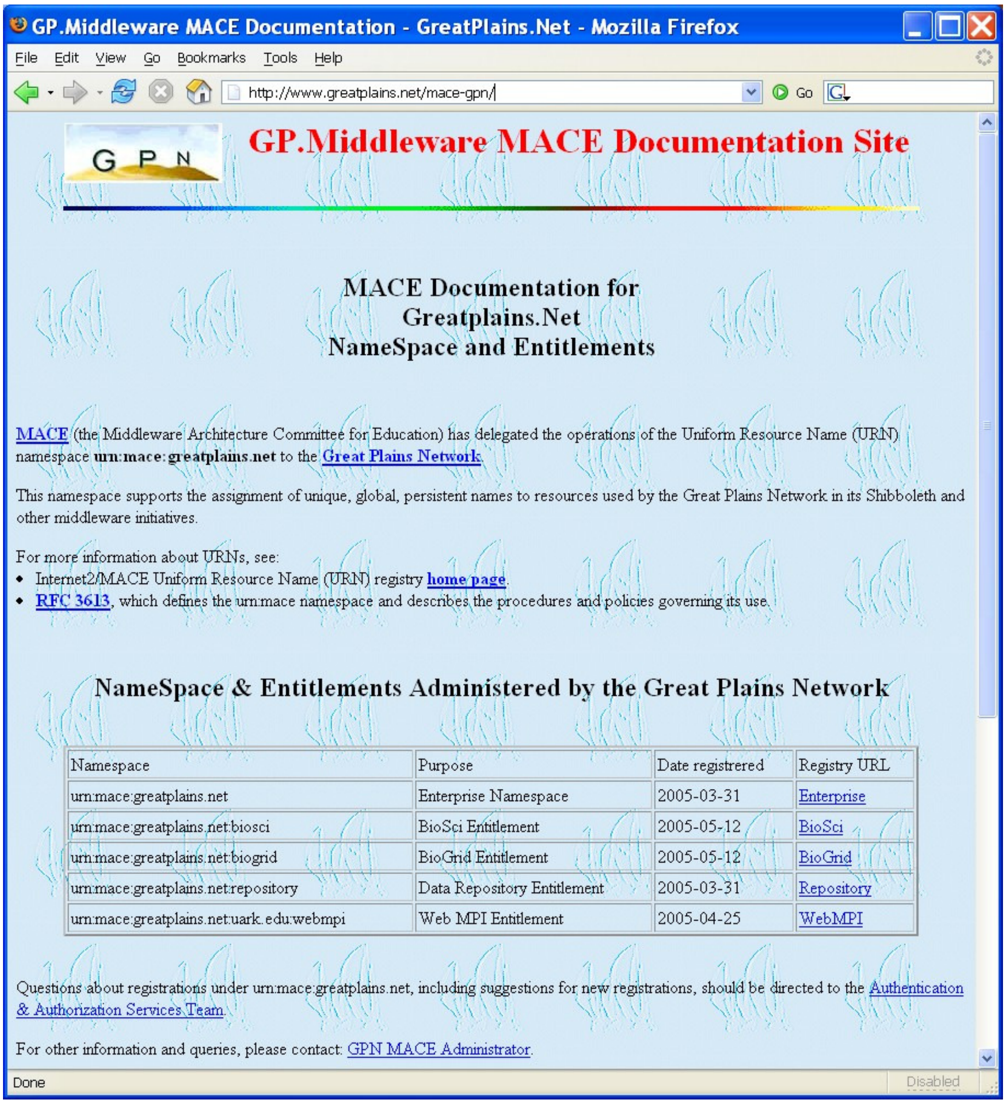
The GPN access to resources can be further mediated by a portal application that determines final access using the combination of entitlements that are asserted. Apache will permit users asserting any GPN MACE registered entitlement. This is achieved using a regular expression. Once the user is at the portal, they may be granted access to one or more resources (or none, as appropriate) depending on what entitlements are asserted. There are currently four resources at two different institutions that require the use of eduPersonEntitlements: the biomedical data and application service, biomedical computing resources, and the GPN Repository at the University of Missouri, and the WebMPI computational resource at the University of Arkansas.

4.3. Registration of the MACE Greatplains.net Namespace. The Great Plains Network (GPN) has registered the name urn:mace:greatplains.net with the Internet2 Middleware Architecture Committee for Education (MACE). This name is the top level of a hierarchical namespace controlled by the GPN for use in its collaboration efforts. In the case of GPN, this namespace includes specific entitlement values that are used to provide fine-grained access control to GPN defined resources.

As part of the MACE namespace registration process, a URL is provided to access online documentation for the registered namespace. In the case of the greatplains.net MACE namespace, the registered URL is: http://www.greatplains.net/mace-gpn. This web site defines the namespace and the entitlement attributes defined for use by the GPN group, as shown in Figure 4.1. The documentation for MACE as well as the corresponding IETF RFC (3613) that MACE is based is provided as links from this web page.

The GPN middleware project is using the MACE namespace to define and specify the eduPersonEntitlement information used to identify resources and authorization information needed to authorize access to specific resources and services in use among the GPN collaborating institutions. Individuals who have authenticated through Shibboleth (at their home institution) and who have eduPersonEntitlements that match the greatplains.net defined entitlements are granted access to the defined resources or services. Individuals without such entitlements are denied access to the corresponding resources and services even though they have been authenticated via Shibboleth from one of the collaborating institutions.

The MACE registered namespace for greatplains.net is critically important to this project since it provides a persistent URN naming convention under control of the GPN collaborating institutions. This mechanism creates a virtual organization consisting of a wide variety of individuals from a collection of institutions that does not fall into the normal classifications for individuals in a typical identity management system within or across multiple institutions. For example, the virtual organization can permit selected students, faculty and staff at various institutions to access services without granting access to all such groups of individuals from all of the institutions. In short, the defined entitlements provide a way to utilize middleware standards, while providing an extensible means for accommodating specific and unique needs of groups of individuals that can be easily tailored for fine-grained discriminating decisions implemented in application programs and systems.


GP.Middleware MACE Documentation - GreatPlains.Net - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.greatplains.net/mace-gpn/
Go
G P N
Done
Disabled

GP.Middleware MACE Documentation Site

MACE Documentation for
Greatplains.Net
NameSpace and Entitlements

MACE (the Middleware Architecture Committee for Education) has delegated the operations of the Uniform Resource Name (URN) namespace **urn:mace:greatplains.net** to the Great Plains Network.

This namespace supports the assignment of unique, global, persistent names to resources used by the Great Plains Network in its Shibboleth and other middleware initiatives.

For more information about URNs, see:
- Internet2/MACE Uniform Resource Name (URN) registry **home page**.
- **RFC 3613**, which defines the urn:mace namespace and describes the procedures and policies governing its use.

NameSpace & Entitlements Administered by the Great Plains Network

Namespace	Purpose	Date registrered	Registry URL
urn:mace:greatplains.net	Enterprise Namespace	2005-03-31	Enterprise
urn:mace:greatplains.net:biosci	BioSci Entitlement	2005-05-12	BioSci
urn:mace:greatplains.net:biogrid	BioGrid Entitlement	2005-05-12	BioGrid
urn:mace:greatplains.net:repository	Data Repository Entitlement	2005-03-31	Repository
urn:mace:greatplains.net:uark.edu:webmpi	Web MPI Entitlement	2005-04-25	WebMPI

Questions about registrations under urn:mace:greatplains.net, including suggestions for new registrations, should be directed to the Authentication & Authorization Services Team.

For other information and queries, please contact: GPN MACE Administrator.

Fig. 4.1. *The GP.Middleware MACE Documentation Web Site*

4.4. Building an Attribute Architecture. The GPN middleware team is defining an attribute architecture to support the Great Plains Network Consortium in its efforts to create a regional collaboration environment among its members. This architecture is predicated on the eduPersonEntitlement attributes defined in the GPN registered MACE NameSpace. The MACE NameSpace: urn:mace:greatplains.net is central to facilitating the development of applications and services to be shared among individuals located at consortium member institutions.

Shibboleth is used for authentication into the virtual organization (i. e., greatplains.net) environment. The urn:mace:greatplains.net NameSpace is used for required authorizations to access the collaboration resources defined within the NameSpace via Shibboleth/EduPerson released entitlements. It is important to note that the enforcement of the entitlement rules for access is enforced in a split responsibility fashion. The first part of the entitlement syntax (urn:mace:greatplains.net) is enforced by the Shibboleth target Apache web server. That is, if the prefix of the entitlement for the GPN MACE NameSpace does not appear in the entitlement list at authentication time, access to the Shibboleth target is denied. The last part of the NameSpace entitlement (everything following the final colon in each entitlement in the list of entitlements asserted during the authentication) is used to enforce the specific authorizations an entity has within the defined entitlement NameSpace.

The presently defined entitlements consist of two groups; the entitlements requested for use by the University of Missouri-Columbia and the entitlements requested for use by the University of Arkansas. Conceptually, these entitlements are hierarchical and interoperable. And, globally these entitlements permit fine-grained control over selected resources or capabilities offered to users located at several institutions located in the Great Plains region.

The architecture and definitions presently active are accessible at:
`http://www.greatplains.net/mace-gpn`

The entitlements defined include:

- `urn:mace:greatplains.net:biosci` The BioSci attribute allows authenticated identities to access and utilize resources to support research activities in the biological and life sciences among the member institutions.
- `urn:mace:greatplains.net:biogrid` The BioGrid attribute allows authenticated identities to access and utilize region-wide grid computing resources to support research activities requiring grid computing access.
- `urn:mace:greatplains.net:repository` The Repository attribute allows authenticated identities to access and utilize a region-wide data repository for sharing documents, files and data among the participating members.
- `urn:mace:greatplains.net:uark.edu:webmpi` The WebMPI attribute allows authenticated identities to access and utilize the grid computing cluster at the University of Arkansas to develop, test, and run MPI-based parallel programs.

The three entitlements: BioSci, BioGrid, and Repository have a hierarchical relationship. BioSci is the superior (top level) entitlement that incorporates the BioGrid and Repository entitlements. That is, any authenticated entity that incorporates the BioSci entitlement automatically is presumed to also have the BioGrid and Repository entitlements as well. However, an authenticated entity with the BioGrid entitlement does not incorporate the BioSci entitlement nor the Repository entitlement. The same is true for the Repository entitlement. Namely, having the Repository attribute does not incorporate either the BioSci or the BioGrid attribute for authorization purposes.

The WebMPI entitlement has an interoperable relationship with the BioSci entitlement. Namely, at the University of Arkansas, an authenticated entity with the WebMPI attribute is granted access to the grid computing resource for developing MPI programs on a cluster machine. However, another authenticated entity with the BioSci attribute will be authorized for the WebMPI attribute when required or needed. Since our resources are in prototype phase this interoperation capability is not fully functional at this time.

Details of the meaning of the specific entitlements that have been defined can be found by accessing the appropriate links in the attribute table found at: `http://www.greatplains.net/mace-gpn`. These definitions are in active development and are subject to change at any time. The URN attributes are persistent, but the details of the authorizations they carry may change.

At this time, the BioSci, BioGrid, and Repository MACE entitlement attributes are associated with the Shibboleth Target: `https://crick.rnet.missouri.edu/GPN`. The WebMPI MACE entitlement is associated with the Shibboleth Target: `http://webmpi.csce.uark.edu`.

5. Synergism with Other Research Communities. As the GPN continues to build virtual organizations several opportunities have arisen. Synergistic activities among researchers in GPN provide motivation and opportunity to build on the middleware test bed, and our efforts have allowed us to build collaborative research relationships with members of the larger middleware community.

5.1. Regional Collaboration. A group of researchers within the GPN group of institutions has begun discussion and planning to promote biomedical application development and collaboration using the middleware infrastructure being developed within the GPN. This effort is in its infancy, but is due in part to the expanding efforts of the GPN to encourage researchers to become active in projects that enable applications of specific interest to the researchers.

Using the middleware infrastructure, the researchers can concentrate on the deploying of applications and application development of biomedical software for research use. The middleware infrastructure also enables the group to utilize inter-institutional resources, such as Grid technology, to further the individual research efforts. In fact, the GPN MACE entitlements (BioSci and BioGrid), were set up to support the efforts of this group and other groups in the future.

The biomedical sciences group is planning a pilot implementation to be used to prepare for submissions of research proposals to external funding agencies. This planning process includes selecting appropriate applications suited for this activity, develop a prototype pilot implementation and collect data to be used in proposal writing.

A second group of researchers in the area of environmental sciences, water resource modeling and watershed research is also keenly interested in using the developing middleware infrastructure in support of research. The environmental science group has a history of collaboration that precedes the GPN project. The group has met regularly and participated in several proposals to date.

Within the GPN environmental science research community several broad-based water modeling applications have been developed. These applications are used across the region to predict the effects of changes due to weather events, farming practice, and urbanization. Some of the applications use data that is collected across a broad geographic region and shared among the researchers in the area. The needs of this virtual organization to share data, applications, and research results in an authorized manner are helping to drive the development of GPN middleware.

5.2. Integration with Grid Computing Middleware. Grid computing has been one of the driving forces for the GPN middleware group. Once the ability to use Shibboleth for authentication across institutions was accomplished, the ability to access shared resources and applications has become an important part of the group's efforts. While Shibboleth provides an architecture and protocol for authentication and authorization, many other capabilities are required for resource sharing in the region, including service discovery, service monitoring, data management, workflow management, and so on.

Globus Toolkit is the default middleware for providing grid capabilities. Since the formal completion of the ETR project in early 2006 our efforts have shifted somewhat to unifying resources using Globus so that we will have a grid that incorporates these capabilities. We currently have a set of cluster resources at four different institutions that have been installed with Globus Toolkit.

Globus is based on Grid Security Infrastructure (GSI) [13], which relies on X.509 certificates for the authentication and authorization of each user and each resource in the grid. A user certificate contains, among other information, a globally-unique Distinguished Name. The Distinguished Name is used directly in Access Control Lists to determine if a user is allowed to access a particular resource. One or more common Certificate Authorities must sign, directly or indirectly through a proxy certificate, the certificates held by the user and the resource.

GridShib is a software product that allows a middleware infrastructure built using the Globus Toolkit to incorporate the attribute-based authorization protocol of Shibboleth [10]. The software package includes two plug-ins, one that interfaces to Globus Toolkit version 4, and one that interfaces to Shibboleth. The primary purpose of the Globus plug-in is to request attributes about a requesting user from the Shibboleth Attribute Authority (a software component at the Identity Provider site). The Globus plug-in parses the attributes, caches them, makes an access control decision and then allows or disallows access to the resource. The Shibboleth plug-in allows the Attribute Authority to map a Distinguished Name from a GSI X.509 certificate to a local user identifier, and to query the attribute database based on the local user identifier.

GridShib assumes that a user (a grid client) and the grid service both possess an X.509 credential. That is, a user logs in using a Globus login process to acquire a certificate or proxy certificate. GridShib also assumes that the user has an account with a home institution a Shibboleth Identity Provider. The Identity Provider and the grid Service Provider each have a globally unique providerID, and the Service Provider and Identity Provider use a common metadata format and exchange metadata out-of-band.

A beta release of GridShib software became available in fall, 2005. The GPN virtual organization is in the process of incorporating these components into the existing Shibboleth and Globus infrastructure.

5.3. Challenges of Managing Entitlements in Virtual Organizations. Defining entitlements and utilizing them in applications to provide fine-grained discrimination for authorization is challenging. A truly significant aspect of this challenge involves the management of the entitlements in identity management providers (IdP) across a virtual organization. A large number of policy and technical decisions must be developed to allow entitlement data to be incorporated into separate IdPs governed by different business and policy models.

In the current test bed the eduPersonEntitlement field is used to store user attributes in local campus LDAP directories, and the Shibboleth Attributed Authority retrieves these attributes for authorization to a server. The eduPersonEntitlement field was readily available at the time of the development of the GPN middleware test bed, but this approach of storing virtual organization attributes in local campus LDAP directory services is not a long-term scalable solution. How to management entitlements and user attributes is a major topic of discussion in NMI, Shibboleth, Grid, and other efforts. This problem is also the focus of a recent NSF-sponsored workshop on Campus Research Computing Cyberinfrastructure.

In cooperation with the GPN, NMI-EDIT and Internet2/MACE Signet groups, we are looking at the issues and means for virtual organizations to authorize, request and manage authorization data that must be incorporated into separate idPs to be effective. For example, currently the GPN entitlement values must be contained in an identity provider's identity management database for individuals who are authenticated at their home institution and are members of the virtual organization. The virtual organization should be the ones to authorize entitlements in a given idP, but idPs are reluctant to empower anyone but their local identity management team to update anything in their database. Similarly, a single individual may have several identities/roles in an organization and it must be determined which identity is to be granted access in an external virtual organization. The current thinking is that the entitlements must be maintained in a database that is separate from the campus identity management database but must somehow be coordinated with it. These issues are of significant interest to a wide range of institutions, government agencies, and funded research projects. We will be working with the NMI effort to incorporate new technologies as better approaches are developed and to develop or deploy the prototype that will be used in the test bed.

Signet is a developing core middleware component that assists in the management of privileges [14]. Signet provides: 1) a standard user interface for privilege administrators; 2) a consistent, simplified policy definition via roles and including integration with core campus organizational data; 3) improved visibility, understandability and auditability of privilege information; and 4) standard interfaces to other infrastructure services.

Grouper is an open source toolkit for managing groups, and is designed to work complementary with Signet and other middleware components [15]. Grouper is designed to manage multiple sources of group information, such as what may be found among the several GPN institutions and sets of user groups who desire to access the various GPN and other resources. In addition to GridShib, we anticipate that Signet and Grouper will become core components of the developing GPN middleware test bed.

6. Conclusions. The development and the deployment of middleware infrastructure in the GPN is an on-going project. The accomplishments can be described both in terms of the technological accomplishments described in this paper as well as the progress in inter-institutional collaboration. Several lessons have been learned during the task of building a middleware infrastructure in such a heterogeneous and widely-distributed environment, including:

1. Collaboration among individuals is essential. We have supported collaboration with weekly teleconferences and a mailing list. Since the tools are sometimes difficult to install and use it is essential that individuals have community support for asking technical questions. Individuals are much more comfortable contacting colleagues that they already know for help in installing and configuring a new tool than they are in contacting an unknown person. A high level of collaboration allows the project to move along much more quickly.
2. A middleware project among cooperating institutions that lack a central hierarchical administration will be limited by the level of support of the institutions in the region. Specifically, if a university infrastructure resource, such as a campus LDAP server or campus firewall configuration, is required to be a part of the architecture, then staff employees must have time allocated during the week that is specifically devoted to the middleware project. In the case of faculty, the tenure reward system must recognize the contribution to the middleware project as evidence of productive research and teaching. The benefits of middleware infrastructure are long-term, and so an institution must take a long-term view of its contribution to the project and not expect that the individuals efforts will immediately benefit the institution.
3. Communication and constant education about the on-going changes in the middleware community are essential. For example, the release of the GridShib software component for unifying Shibboleth and Globus security models, the development of Signet and Grouper, and the current discussions on Campus Research Computing Cyberinfrastructure are having an impact on the test bed. Some members of the GPN middleware team are participating in a NMI-EDIT Signet and Grouper Early Adopters Workshop and will bring the ideas from those organizations to the larger middleware group.

The GPN ETR project has accomplished the original goal of the deployment of a middleware test bed on a small number of campuses. These prototypes and test components are a building block for further work. The GPN middleware group is in an excellent position to collaborate on new grid and middleware computing projects and have an opportunity to test the evolving mechanisms in a live test bed with multiple participating institutions operating as a virtual organization.

The next goals for GPN include the development of a robust and scalable architecture for attribute management and fine-grained access control, the development of applications and a community of users that can utilize the storage and computing resources that have been developed, and move towards a production quality set of tools in support of research and education in the region.

7. Acknowledgements. We would like to acknowledge the work and contributions of Kathryn Huxtable, Denis Hancock, Larry Sanders, Shafquat Bhuiyan, Arturo Guillen, Abdul Khambati, Kurt Landrus, James McCartney, Linh Ngo, and the support of our colleagues from the Great Plains Network Consortium, without whose help this project would not have been possible.

Funding for this project is provided in part by an Extending the Reach (ETR) grant from Educause on behalf of the NMI-EDIT Consortium of Internet2, Educause, and SURA, and with several statewide university systems and regional networks. See `http://archie.csce.uark.edu/gpn/` for more information about this funded project.

REFERENCES

[1] Great Plains Network Consortium (2006), http://www.greatplains.net/

[2] Enterprise and Desktop Integration Technologies (EDIT) Consortium (2006), http://www.nmi-edit.org/index.cfm

[3] Shibboleth Project, Internet2 Middleware (2006), http://shibboleth.internet2.com

[4] Amy Apon, Greg Monaco, Gordon Springer, and Kathryn Huxtable, *Great Plains Network: Building the Regional Middleware Infrastructure,* NMI-EDIT Case Study Series, January, 2006. http://archie.csce.uark.edu/gpn/publications/GPN-Building-NMI-case-study-final.pdf

[5] Gordon Springer, Shafquat Bhuiyan, and Arturo Guillen, *Great Plains Network: Integrating Shibboleth, Grid and Bioinformatics,* NMI-EDIT Case Study Series, January, 2006. http://beagle.rnet.missouri.edu/GPN/Docs/MUCaseStudy.pdf

[6] Tom Barton, Jim Basney, Tim Freeman, Tom Scavo, Frank Siebenlist, Von Welch, Rachana Ananthakrishnan, Bill Baker, Monte Goode, and Kate Keahey, *Identity Federation and Attribute-based Authorization through the Globus Toolkit, Shibboleth, Gridshib, and MyProxy,* In 5th Annual PKI R & D Workshop, April 2006. http://grid.ncsa.uiuc.edu/papers/gridshib-pki06-final.pdf

[7] EduPerson Object Class Specification, Internet2, April, 2006. http://www.nmi-edit.org/eduPerson/draft-internet2-mace-dir-eduperson-latest.html

[8] Subversion project (2006). http://subversion.tigris.org/

[9] Ian Foster and Carl Kesselman, *Globus: A Metacomputing Infrastructure Toolkit,* Intl J. Supercomputer Applications, 11(2):115-128, 1997.

[10] Welch, Von, Tom Barton, Kate Keahey, Frank Siebenlist, *Attributes, Anonymity, and Access: Shibboleth and Globus Integration to Facilitate Grid Collaboration,* Proceedings of the 4th Annual PKI R & D Workshop, 2005. http://grid.ncsa.uiuc.edu/papers/gridshib-pki05-final.pdf

[11] InQueue Home (2006), http://inqueue.internet2.edu/

[12] Incommon Federation (2006), http://www.incommonfederation.org/

[13] V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, and S. Tuecke, *Security for grid services,* In Twelfth International Symposium on High Performance Distributed Computing (HPDC-12). IEEE Computer Society Press, 2003.

[14] Signet Home Page (2006). http://middleware.internet2.edu/signet/

[15] Grouper Working Group (2006). http://middleware.internet2.edu/dir/groups/grouper/

Edited by: Dana Petcu.
Received: February 08, 2006.
Accepted: May 25, 2006.