The Cancer Biomedical Informatics Grid (caBIG™) Security Infrastructure

Stephen Langella[1], Scott Oster[1], Shannon Hastings[1], Frank Siebenlist[2], David Ervin[1], Tahsin Kurc[1], Justin Pemar[1], Joel Saltz[1]

[1]*Department of Biomedical Informatics*
Ohio State University
Columbus, OH 43210
{langella,oster,hastings,ervin,kurc,jpemar}@bmi.osu.edu
Joel.Saltz@osumc.edu

[2]*Mathematics and Computer Science Division*
Argonne National Laboratory
Argonne, IL 60439
franks@mcs.anl.gov

Abstract

Given the sensitivity of the medically related data and the number of institutions involved, security has quickly become a high priority in the Cancer Biomedical Informatics Grid (caBIG™). In evaluating the security requirements of caBIG several security issues were raised that existing grid security technologies could not address. Considering the scale of caBIG, these issues become challenging. In this paper we will present these issues and the infrastructure developed to address them.

1. Introduction

The Cancer Biomedical Informatics Grid (caBIG™) (1) program is funded by the National Cancer Institute (NCI) and was launched to provide a coordinated approach to the informatics requirements of basic and clinical cancer research and multi-institutional studies. The goal is to accelerate the delivery of innovative approaches for the prevention and treatment of cancer by facilitating sharing, discovery, and integration of distributed information and analytic resources. At the time this paper was written the caBIG™ community was made up of over 800 participants from over 80 organizations working together on over 70 projects. In the future it is expected that the caBIG™ community will grow to hundreds of organizations and many thousands of cancer-research participants from geographically dispersed medical centers, universities, government agencies, and commercial companies.

In this paper we will present the Grid Authentication and Authorization with Reliably Distributed Services (GAARDS) security infrastructure. GAARDS was developed as part of the caGrid (2)(3) infrastructure. caGrid is the core software architecture for caBIG™, caGrid is a service-oriented architecture and implementation that provides core services, toolkits and wizards for the development and deployment of community provided services. In order to articulate the security requirements of caBIG and evaluate existing technologies, a Security Technology Evaluation White Paper (4) was developed. This white paper along with numerous working groups run within the caBIG™ community raised key security issues which became the motivating influences for the GAARDS work. Given the size of caBIG™ one key issue identified was the provisioning of user accounts. caGrid is built on top of the Globus Toolkit (5), the most widely used reference implementation of grid standards. The Globus Toolkit implements support for security via its Grid Security Infrastructure (GSI) (6). GSI utilizes X.509 Identity Certificates for identifying a user. An X.509 Certificate with its corresponding private key forms a unique credential or so-called "grid credential" within the Grid. These grid credentials are used to authenticate both users and services. Although this approach is very effective and secure, it is difficult to manage in a multi-institutional environment. Using the base Globus toolkit, the provisioning of grid credentials is a manual process, which is far too complicated for users. The overall process is further complicated if a user wishes to authenticate from multiple locations, as a copy of their private key and certificate has to be present a every location. Not only is this process complicated, securely distributing private keys is error prone and poses a security risk. Additionally there is a scalability and efficiency problem in vetting user identities, in Grids this it traditionally done by certificate authority administrators. It was determined by the caBIG™ community that this approach would not be adequate. Organizations invest a significant amount of resources into their existing identity management systems and already have processes in place for vetting user identities. It was determine that it would me more efficient to leverage an organization's existing identity management systems to provision grid user accounts. Under this approach users would be able to use their existing credentials assigned to them by their organization to "logon" or obtain grid credentials such that they may access the services of the Grid. Cleary a mechanism is needed to allow users to obtain grid credentials using their existing organization provided

credentials. It is also essential that this mechanism remove the earlier described complications of using and managing grid credentials. The GAARDS infrastructure provides this mechanism through a grid service called Dorian (7); we will provide more details on Dorian later on this paper.

In an environment where credentials are being issued by multiple authorities, another key security issue is determining which authorities to accept credentials from and at what level of assurance. In order to authenticate users and other peer-services, Grid services need to maintain a list of authorities that they trust as a source for issuing credentials. Grids inherently span multiple institutional administration domains and aim to support the sharing of applications, data, and computational resources in a collaborative environment. In this environment there may exist hundreds of certificate authorities, each issuing hundreds if not thousands of certificates. In such a dynamic multi-institutional environment with tens of thousands of users, credentials will be issued and revoked frequently, and new authorities will be added regularly. Clearly a Grid-wide mechanism is needed for maintaining and provisioning trusted certificate authorities, such that Grid services and users may make authentication and authorizations decisions against the most up-to-date trust information. The GAARDS infrastructure provides a grid-wide trust mechanism called the Grid Trust Service (GTS) (8); more details on the GTS will be provided later on in this paper.

Another key security issued raised was the need for scalable access control enforcement. It was determined that it is critical that access control policy be maintained and enforced locally, giving data providers the ability to grant who has access to their data. At the same time in order to scale it is important that access control policy be based on information managed by the Grid. Since most access control systems base access control policy on membership to groups it was determined that a mechanism for organizing and managing groups spanning organizational boundaries was needed. The GAARDS infrastructure provides a service called Grid Grouper (9) for facilitating this. It is anticipated that existing access control systems could be modified to base their access control policies off of groups managed by Grid Grouper as well as off the local groups that they currently maintain; the Common Security Module (CSM) (10) is an example of such a system. CSM is the access control system used by the caCORE Software Development Kit (11), the toolkit used in developing the majority of caBIG™ applications. CSM has also been adopted into the GAARDS infrastructure as its native access control system.

2. GAARDS

The Grid Authentication and Authorization with Reliably Distributed Services (GAARDS) provides services and tools for the administration and enforcement of security policy in an enterprise Grid. GAARDS was developed on top of the Globus Toolkit and extends the Grid Security Infrastructure (GSI) to provide enterprise services and administrative tools for: 1) grid user management, 2) identity federation, 3) trust management, 4) group/VO management 5) Access Control Policy management and enforcement, and 5) Integration between existing security domains and the grid security domain. GAARDS services can be used individually or grouped together to meet the authentication and authorization needs for Grids. Below is a list of some of the core services provided by GAARDS:

Dorian (7)– A grid service for the provisioning and management of grid users accounts. Dorian provides an integration point between external security domains and the grid, allowing accounts managed in external domains to be federated and managed in the grid. Dorian allows users to use their existing credentials (external to the grid) to authenticate to the grid.

Grid Trust Service (GTS) (8)- The Grid Trust Service (GTS) is a grid-wide mechanism for maintaining and provisioning a federated trust fabric consisting of trusted certificate authorities, such that grid services may make authentication decisions against the most up to date information.

Grid Grouper (9)- Provides a group-based authorization solution for the Grid, wherein grid services and applications enforce authorization policy based on membership to groups defined and managed at the grid level.

Authentication Service - Provides a framework for issuing SAML assertions for existing credential providers such that they may easily integrated with Dorian and other grid credential providers. The authentication service also provides a uniform authentication interface in which applications can be built on.

Common Security Module (CSM) (10) - Provides a centralize approach to managing and enforcing access control policy authorization.

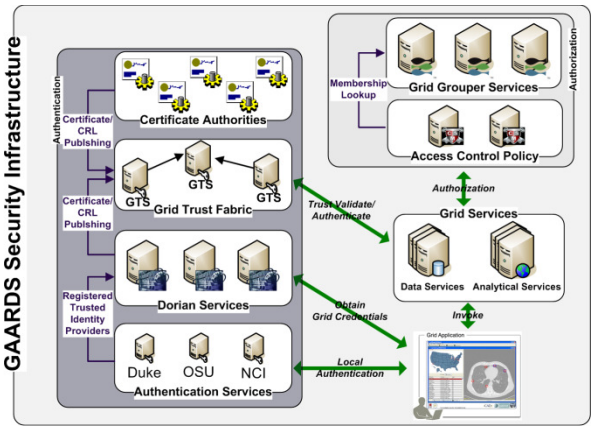

Figure 1 GAARDS Security Infrastructure

Figure 1 illustrates the GAARDS security infrastructure, in order for users/applications to communicate with secure services, they need grid credentials. Obtaining grid credentials requires having a Grid User Account. Dorian provides two methods for registering for a grid user account: 1) Register directly with Dorian 2) By having an existing user account in another security domain. It is anticipated that most users will use their existing locally provided credentials for obtaining grid credentials and only users that are un-affiliated with an existing credential provider should register directly with Dorian. In order to use an existing user account to obtain grid credentials, the existing credential provider must be registered in Dorian as a Trusted Identity Provider. It is anticipated that the majority of grid user accounts will be provisioned based on existing accounts. The advantages to this approach are: 1) users can use their existing credentials to access the grid 2) administrators only need to manage a single account for a given user. To obtain grid credentials, Dorian requires proof or a SAML assertion (12) that proves that the user locally authenticated. The GAARDS Authentication service provides a framework for issuing SAML assertions for existing credential providers such that they may be used to obtain grid credentials from Dorian. The authentication service also provides a uniform authentication interface in which applications can be built on. Figure 1 illustrates the process for obtaining grid credentials, the user/application first authenticates with their local credential provider via the authentication service and obtains a SAML assertion as proof they authenticated. They then use the SAML assertion provided by the authentication service to obtain grid credentials from Dorian. Assuming the local credential provider is registered with Dorian as a trusted identity provider and that the user's account is in good standing, Dorian will issue grid credentials to the user. It should be noted that the use of the authentication service is not required; an alternative mechanism for obtaining the SAML assertion required by Dorian can be used. If as user is registered directly with Dorian and not through an existing credential provider, they may contact Dorian directly for obtaining grid credentials.

Once a user has obtained grid credentials from Dorian they may invoke secure services. Upon receiving grid credentials from a user, a secure service authenticates the user to ensure that the user has presented valid grid credentials. Part of the grid authentication process is verifying that grid credentials presented were issued by a trusted grid credential provider (i.e Dorian, other certificate authorities). The Grid Trust Service (GTS) maintains a federated trust fabric of all the trusted digital signers in the grid. Credential providers such as Dorian and grid certificate authorities are registered as trusted digital signers and regularly publish new information to the GTS. Grid services authenticate grid credentials against the trusted digital signers in a GTS.

Once the user has been authenticated, a secure grid service next determines if a user is authorized to perform what they requested. Grid services have many different options available to them for performing authorization. The GAARDS infrastructure provides two approaches which can each be used independently or can be used together. It is important to note any other authorization approach can be used in conjunction with the GAARDS authentication/trust infrastructure. Grid Grouper service provides a group-based authorization solution for the Grid, where in grid services and applications enforce authorization policy based on membership to groups defined and managed at the grid level. Grid services can use Grid Grouper directly to enforce their internal access control policies. Assuming the authorization policy is based on membership to groups provisioned by Grid Grouper; services can determine whether a caller is authorized by simply asking grid grouper whether the caller is in a given group. The Common Security Module (CSM) is a more centralized approach to authorization. CSM is a tool for managing and enforcing access control policy centrally. Access control policies can be based on membership to groups in Grid Grouper. Grid services that use CSM for authorization simply ask CSM with a user can perform a given action. Based on the access control policy maintained in CSM, CSM decides whether or not a user is authorized. In Figure 1, the grid services defer the authorization to CSM. CSM enforces its group based access control policy by asking Grid Grouper whether the caller is a member of the groups specified in the policy.

3. Dorian

Dorian (7) is a grid user management service that 1) hides the complexities of creating and managing grid credentials from the users and 2) provides a mechanism for users to authenticate using their institution's authentication mechanism, assuming a trust agreement is in place between Dorian and the institution. Dorian provides a complete Grid-enabled solution, based on public key certificates and SAML, for managing and federating user identities in a Grid environment. Grid technologies have adopted the use of X.509 identity certificates to support user authentication. Dorian uses SAML authentication assertions as the enabling mechanism for federating users from local institutions to the grid.

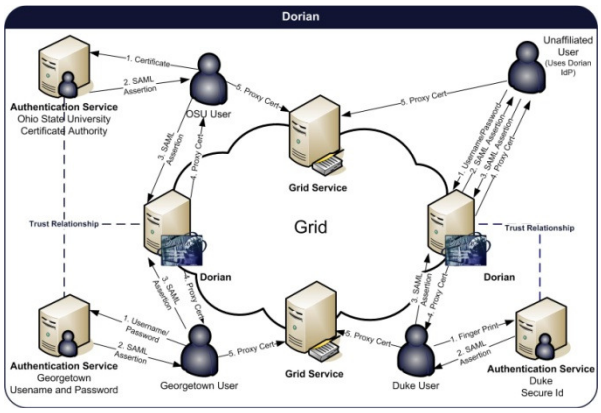

Figure 2 Dorian

Figure 2 illustrates an example usage scenario for Dorian. To obtain grid credentials or a proxy certificate, users authenticate with their institution using the institution's conventional mechanism. Upon successfully authenticating the user, the local institution issues a digitally signed SAML assertion, vouching that the user has authenticated. The user then sends this SAML assertion to Dorian in exchange for grid credentials. Dorian will only issue grid credentials to users that supply a SAML assertion signed by a Trusted Identity Provider. For example, in Figure 2, a Georgetown user wishes to invoke a grid service that requires grid credentials; they first supply the application with their username and password. The application client authenticates the Georgetown user with the Georgetown Authentication Service, receives a signed SAML assertion which it subsequently passes to Dorian in exchange for grid credentials. These credentials can then be used to invoke the grid services. To facilitate smaller groups or institutions without an existing IdP, Dorian also has its own internal IdP. This allows users to authenticate to Dorian directly, thereby enabling them to leverage Dorian to obtain credentials which will allow them to access the grid, this is also illustrated in Figure 2.

4. Grid Trust Service (GTS)

In a Grid environment, there may exist tens or even hundreds of certificate authorities, each issuing hundreds if not thousands of certificates. To make matter worse, in a dynamic multi-institutional environment, the status of identities may be updated frequently. Identities and credentials can be revoked, suspended, reinstated, or new identities can be created. In addition, the list of trusted authorities may change. In such settings, certificate authorities will frequently publish Certificate Revocation Lists (CRL), which specify "black listed" certificates that the authority once issued but no longer accredits. For the security and integrity of the Grid, it is critical to be able to perform authentication and validate a given identity against the most up-to-date information about the list of trusted certificate authorities and their corresponding CRLs. The Grid Trust Service (GTS) (8) is a Web Services Resource Framework compliant federated infrastructure enabling the provisioning and management of a grid trust fabric. The salient features of the GTS can be summarized as follows:

1. It provides a complete Grid enabled federated solution for registering and managing certificate authority certificates and CRLs, facilitating the enforcement of the most recent trust agreements.
2. It allows the definition and management of levels of assurance, such that certificate authorities may be grouped and discovered by the level of assurance that is acceptable to the consumer.
3. The federated nature of the GTS, coupled with its ability to create and manage arbitrary arrangements of authorities into trust levels, allows it to facilitate the curation of numerous independent trust overlays across the same physical Grid.
4. The GTS can also perform validation for a client, allowing a client to submit a certificate and trust requirements in exchange for a validation decision, which allows for centralized certificate verification and validation.

5. Grid Grouper

Grid Grouper (9) is a group/virtual organization management solution for the grid, providing a group based authorization solution for the grid, where grid services and applications enforce authorization policy

based on membership to groups defined and managed at the grid level. Grid Grouper is built on top of Grouper (13) an Internet2 (14) initiative focused on providing tools for group management. Grouper is a java object model which currently supports: basic group management by distributed authorities; subgroups; composite groups (whose membership is determined by the union, intersection, or relative complement of two other groups); custom group types and custom attributes; trace back of indirect membership; delegation. Applications interact with Grouper by embedding the Grouper's java object model within applications. Grouper does not provide a service interface for accessing groups. Grid Grouper is a grid enabled version of Grouper, providing a web service interface to the Grouper object model. Grid Grouper make groups managed by Grouper available and manageable to applications and other services in the grid. Grid Grouper provides an almost identical object model to the Grouper object model on the grid client side. Applications and services can use the Grid Grouper object model much like they would use the Grouper object model to access and manage groups as well as enforce authorization policy based on membership to groups.

In Grouper/Grid Grouper groups are organized into namespaces called stems. Each stem can have a set of child stems and set of child groups with exception to the root stem which cannot have any child groups. For example let's take a university which is compromised of many departments each of which has Faculty, Staff, and Students. In terms of organizing the university in Grid Grouper, a stem could be created for each department, each department stem would contain three groups one for each Faculty, Staff, and Students.

6. Conclusion

The GAARDS infrastructure serves as the core security infrastructure for caGrid and provides services and tools for the administration and enforcement of security policy in an enterprise Grid. The GAARDS infrastructure is under active development with plans on adding additional components as well as maintaining and enhancing existing components based on community demand. Currently we are in the process of developing a module for Grid FTP allowing Grid FTP to leverage the GAARDS infrastructure for enforcing access control. The GAARDS/Grid FTP security module along with the caGrid BDT framework will allow users to query grid services and move the results of the query out of band more efficiently via Grid FTP. In the future we also plan on adding support for auditing, provisioning of service/host credentials, directory services, and support for GAARDS in the caGrid workflow infrastructure.

Recently the GAARDS infrastructure has been accepted as a Globus Incubator Project, setting it on a path for adoption into the Globus Toolkit. This will make GAARDS available to a wider community which helps drive the direction and development of the future GAARDS infrastructure.

7. Works Cited

1. *Cancer Biomedical Informatics Grid (caBIG).* [Online] https://cabig.nci.nih.gov.

2. *caGrid: Design and Implementation of the Core Architecture of the Cancer Biomedical Informatics Grid.* **Joel H. Saltz, Scott Oster, Shannon L. Hastings, Stephen Langella, William Sanchez, Manav Kher, Peter A. Covitz.** No. 15, s.l. : Bioinformatics, 2006, Vol. Vol. 22.

3. *caGrid.* [Online] http://www.cagrid.org.

4. **Daemer, Ken Lin and Gary.** *caBIG™ Security Technology Evaluation White Paper.* 2006.

5. *The Globus Toolkit.* [Online] http://www.globus.org.

6. *Security for Grid Services.* **V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, and S. Tuecke.** s.l. : IEEE Press, 2003. Twelfth International Symposium on High Performance Distributed Computing (HPDC-12).

7. *Dorian: Grid Service Infrastructure for Identity Management and Federation.* **Stephen Langella, Scott Oster, Shannon Hastings, Frank Siebenlist, Tahsin Kurc, Joel Saltz.** Salt Lake City, Utah : The 19th IEEE Symposium on Computer-Based Medical Systems, 2006.

8. *Enabling the Provisioning and Management of a Federated Grid Trust Fabric.* **Stephen Langella, Scott Oster, Shannon Hastings, Frank Siebenlist, Tahsin Kurc, Joel Saltz.** Gaithersburg : 6th Annual PKI R&D Workshop, 2007.

9. **Langella, Stephen.** *Grid Grouper.* [Online] http://www.cagrid.org/mwiki/index.php?title=GridGrouper:Main.

10. *Common Security Module (CSM).* [Online] http://ncicb.nci.nih.gov/infrastructure/cacore_overview/csm.

11. *caCore.* [Online] http://ncicb.nci.nih.gov/infrastructure/cacore_overview.

12. *OASIS Security Services (SAML) TC.* [Online] http://www.oasisopen.org/committees/tc_home.php?wg_abbrev=security.

13. *Grouper* [Online] http://middleware.internet2.edu/dir/groups/grouper/.

14. *Internet 2.* [Online] http://www.internet2.edu/.