Towards More Flexible and Increased Security and Privacy in Grids

Willy Weisz

University of Vienna, Institute for Scientific Computing, VCPC
weisz@vcpc.univie.ac.at

Abstract. The development of UNICORE started as a Grid-enabling middleware with a monolithic security policy that restricted Grid activities to a set of users whose credentials (X.509 certificates) are pre-recorded in a UNICORE User Database (UUDB), and to a task distribution completely defined at job-submission time because the sub-jobs have to be signed by the user with his private key. Later on projects aiming at allowing a restricted interoperability with other Grid middleware lead to the adoption of more flexible approaches like the the Explicit Trust Delegation (ETD). ETD involves implicitly a more general concept: That of an attribute or role which is attached to an identified and authenticated entity and which defines the extent of the authorisations granted to that entity by the target resource. Extending this concept to other authorisation-related aspects of Grid computing is today an area of intensive research, that should also be taken up by the UNICORE developers in order to enable the creation of Virtual Organisations (VOs) that are able to take security as seriously as necessary, and to opt for flexibility as much as possible.

1 General Remarks

Virtual Organisations (VOs) that make up the organisational units which use Grid resources have two almost contradictory requirements: (1) Security that is the prime requirement for the establishment of the trust required when allowing the interoperation of resources belonging to different administration domains, and (2) Flexibility that enable VOs to easily adapt to structures in user membership and resources changing during their lifetime.

The initial decision of the UNICORE design was to give security an overriding primacy that resulted in a very strict and rather inflexible Security Model [1], that originally didn't foresee any interoperability with other Grid middleware. Nevertheless the modular design of UNICORE eases the implementation of new UNICORE Security Models that are more suitable to the security and working requirements of VOs as seen as result of the continuing Grid research.

Departing from the traditional OS views on security, and analysing security and authorisation models in real organisations recent projects came up with new approaches to secure and flexible authorisation schemes.

W. Lehner et al. (Eds.): Euro-Par 2006 Workshops, LNCS 4375, pp. 205–214, 2007.

2 Identification and Authorisation in an Organisation

2.1 Identification

In any organisation the security of internal and external operations relies on the identification of the actors and the authorisations granted to them in any possible action scenario, including (manual or automated) information processing. Virtual Organisations on the Grid have the same requirements.

Employees and collaborators as well as resources must be uniquely identifiable in order to allow a well co-ordinated and optimisable running of the operation. In plans and reports their respective tasks, rights and responsibilities are attached to their identifiers; for people this is generally their common name possibly extended by e.g. a function title or an affiliation with a department.

2.2 Attributes of Entities

In bigger organisations a comprehensive list of individually named human and non-human resources may not be practical. In this case functions or roles with their rights, privileges and duties may be defined and attributed in a many-to-many relationship to individual resources. Overall work can thus be defined, planned, carried out and reported upon as a function of these attributes. The details including the assignments can be left to the possibly dislocated departments. These smaller units are also better suited to track promotions (or demotions), changes of responsibilities and privileges of their local personel, and new or modified non-human resources. Only modifications in terms of changed organigrams or roles need to be passed on to the higher company echelons.

2.3 Authorisation

When it comes to empowerments and thus responsibilities, the company policies should be defined as a function of roles and attributes of the entity (e.g. clearance), not of the name of individuals. If a person may assume different roles within the organisation its empowerment should be defined with regard to the role he is actually assuming when performing a certain task. In analogy the security levels of computer systems (including their environment) have to be at the basis of decisions on which applications are handed over to which hosts.

When organisations are co-operating in projects new authorisation challenges arise:

- each partner provides collaborators which have certain roles and capabilities,
- within the project and even the project phases project roles are defined.

For some actions some capabilities for the project as well as some defined within an individual partner organisation may be required. Projects spanning boundaries will thus base their authorisation decisions on the direct product of the authorisation attributes of the individuals in their home organisations and those defined within the projects.

3 Organisations in the Grid

The co-operative use of resources connected by the Internet (or any other network of local networks) and belonging to independent administration domains — *the Grid* — requires the formation of Virtual Organisations which must define a matrix of authorisations based on the policies regulating the authorisations within a domain and those agreed upon in the projects leading to the establishment of a VO.

The communication power of the Grid makes the creation of flexible inter-organisational projects very attractive. The flexibility of the VOs being such an asset also means that frequent changes at short notice may happen, be it on the user side or on the side of the resources. These changes generally bring about modifications of the authorisation matrix.

4 Identification and Authentication on the Grid

The multi-administration structure of the Grid requires that the identity of consumers and resources be stated unambiguously despite the many different organisations responsible for them. This is made possible by the establishment of X.509 Public Key Infrastructures (PKI) [4] where the public key of a cryptographic key pair is embedded in a certificate which i.a. contains a unique identifier for the entity owning the key pair, and is digitally signed by a "trusted third party", a trusted Certification Authority (CA). This certificate, for which the CA declares that its identification item (the Distinguished Subject Name, Subject DN) is uniquely attributed to this single entity, identifies the entity to resource consumers and providers on the Grid. The private key of the pair is used for authentication purposes and for signing digital documents and messages; the public key is used by the communication partners to send the entity encrypted messages.

The world-wide distribution of Grid consumers and resources makes it necessary to also have CAs distributed over the world. The agreement on minimal rules of operation to establish mutual trust has lead in 2005 to the International Grid Trust Federation [5]. Nevertheless organisations or VOs may establish CAs with special trust requirements, e.g. the UNICORE CA. Such a VO-centered CA has the disadvantage that is doesn't scale when the VO expands its user community or resource pool.

Even more than the trust that can be put in the CAs the storage quality of the private key determines the security level of the PKI. If its repository is a computer disk, then the security level is a product of the user protecting the file containing the key and the system administrator providing an overall secure system: Nobody than the owner is allowed to access and be able to use the private key. Much better security is achieved when the key pair is generated in a secure cryptographic token (SmartCard or USB token) and the private key, that is never allowed to leave the token is encrypted by a PIN only known to the owner. Since the private key is only available on the token, it can only be

used by the person who physically owns the token; and even in the case of theft only the person knowing the PIN can activate the key, i.e. make it usable. When the use of such tokens will become the rule, PKIs will reach a really trustable security level.

5 Authorisation Based on Identity

Like the underlying operating systems, the most widely used Grid middlewares, Globus and UNICORE, base their authorisation infrastructure solely on the identity of the user (Discretionary Access Control, DAC). After the requestor has been authenticated his identity is mapped to an OS-based identity and he is welcome to an "almost help-yourself party".

This lack of fine-grain authorisation in most operating systems has led the creators of database systems to define their own access schemes mostly independent of the OS-related identities. They define roles and access rights, and manage them their own way. The lack of authorisation beyond the user identity makes the Grid for the time being unfit for the use of federated databases. Neither the OGSA-DAI project [6] nor the GGF DAIS-WG [7] have tackled the security aspects of a gridified database access. But databases are but one resource that needs fine-grain, role-based access rules, individuals' health records in any kind of container are another example.

This tradition of reducing the authorisation policy to the mere identification and authentication of the entity requesting a resource has determined how secure authorised accesses have been perceived for the Grid. It is clearly insufficient at the level of VOs.

6 Authorisations Based on Properties of the Entities

In high security environments, entities (consumers and resources) are classified according to security clearance levels. The corresponding authorisation scheme (Mandatory Access Control, MAC) allows read access only to objects of the same or a lower clearance level (Read Down) and write objects of the same or a higher clearance level (Write Up). In most organisational environments MAC is too restrictive a scheme.

The Role-Based Access Control (RBAC) has become the preferred authorisation scheme when DAC is too weak and MAC is too restrictive. It allows policies that are more fine-grained than identity-based access rights. Changes of the position in an organisation generally incur changes in roles for the entity. Like with MAC hierarchies of roles can be constructed leading to hierachies of authorisations.

Attribute-Based Access Control (ABAC) provides even more flexibility as the attribute relations are less static than the consumer-resource authorisation relations in RBAC.

7 User Database at the Grid Resource Site

The UNICORE and Globus assumption is that secure operations in Grids require comprehensive lists of entities that are allowed to access resources at a local administrative domain (e.g. the UUDB at a UNICORE V-site or the gridmap-file for a Globus host). This of course doesn't scale well and isn't appropriate for Virtual Organisations which may be short-lived and/or allow compositions of users and network-attached hardware varying over the lifetime of the VO.

The Globus approach is more flexible as it allows a remote management of the gridmap-file, e.g. by a VO management system like VOMS [8], whereas the UNICORE User Database (UUDB) can only be maintained from the site where it is located. UNICORE also requires that the X.509 Certificate be stored in the UUDB which must therefore be continuously updated since, for security reasons, the certificates have a limited life time und must thus be regularly renewed. The certificates for Globus are stored at the user's site where they have their first home after they have been issued by the CA.

8 Managing Authorisation for VOs

As has been described in Sect. 2 the authorisation structure in bigger organisations and for inter-organisational projects should move from concentrating on identities and their rights to access resources to policies based on roles and attributes of human and non-human resources. This is also true for Virtual Organisations.

8.1 The Attributes of Requestors in Their Organisation

In his own organisation a user may assume roles or have certain attributes. These attributes are signed by an Attribute Authority (AA) that is legitimised by this organisation and recognised by the resource providers in the VO. Its statements concerning attributes of an entity must bear the proof of its origin and a digital signature verifiable by a recognised certificate.

8.2 The Attributes of Requestors in the VO

Similarly the VO itself may need an Attribute Authority that issues digital documents stating the attributes of the requestor within the VO. These attributes may be functions of the identity of the requestor and/or of his roles and attributes as defined by his home organisation, or be just defined by his role in the VO.

8.3 Privacy — Anonymity

In certain applications it may be important (or even required by law) that the identity of the requestor be anonymised for the time of the resource usage, but nevertheless be traceable at some point in time, e.g. for accounting purposes or for feedback. This can be realised by mapping the identity to a "general user" which is given attributes allowing traceability on a need-to-know basis.

8.4 Attributes of the Resources

Likewise there must be Attribute Authorities that issue information documenting attributes of the resources. They may come from their administration domain or be VO-related, e.g. availability to or costs for the VO.

8.5 Authorisation

In big real organisations a complex set of rules defines who is entitled to take which decisions and who is to implement them. The company policies thus defined are generally expressed as functions of roles and levels in the hierarchy, not of individuals.

Likewise in VOs policies govern the authorisation decisions. The complete set of information on the identities and attributes of consumer and resources triggers a policy decision to grant (or deny) the requestor a set of privileges and access rights that the policy enforcement engine will have to use in order to grant access to resources.

The policy may even require a third party permission: The right to access a person's Electronic Health Record (or identifiable parts of it) that may be distributed over a national Health Grid will require the patient's consent (at least in Austria). This third party will also need to be authenticated and its role or attributes taken into account.

9 Consequences for the UNICORE Development

The need for a more flexible but nevertheless improved security and privacy protection must trigger major changes in the UNICORE security infrastructure. The integration of such developments is facilitated by the modular architecture.

9.1 Authentication

The UUDB is too inflexible for future Grid environments. Without sacrificing security concerns an authentication mechanism is needed that doesn't store all potential users at the resource site, but rather security policies.

The Security Model of UNICORE doesn't allow the use of Proxy Certificates [9]. Without this facility no message-level security is possible. And the extension to allow a limited interoperability with Globus transmits a private key over the communication lines! Even so it is included in an encrypted blob, this is against the proxy concept that the private key corrsponding to a proxy certificate is only used on the system where it has been generated, and never leaves that system.

9.2 A First Use of Attributes: The Explicit Trust Delegation

The requirement to build dynamic Grid jobs for which an agent (e.g. a portal) decides on the distribution of tasks after the end-user has submitted his job

description, require that other instances than the job signer (the end-user) get authority to request actions on behalf of the end-user.

Since the UNICORE Security Model doesn't allow the use of proxies this delegation of rights of the end-user to UNICORE agents is managed by the definition of a trust attribute that the end-user issues for that agent, the Explicit Trust Delegation (ETD) [10].

Even so it is not presented as such, ETD can be seen as the first(?) introduction of a formal policy based on attributes (trust) conferred to a Grid entity (the agent) by an AA (the end-user).

9.3 The Proposed Authorisation Architecture

The authorisation arcitecture to be developed for UNICORE should provide the following functions:

For any subject and target of a request a complete policy or set of policies must be defined by a Source of Authority (SOA); this collection will be used to derive decisions whether to accept or deny requests.

After being authenticated the request for use of resources (including all the identity/role/attribute information provided by the client agent) is handed over to the Policy Enforcing Engine (PEP).

The PEP hands the request over to the Policy Decision Point (PDP) which applies the rules taking into consideration the identity/roles/attributes included in the request, and if needed, requesting further information from Policy Information Points (PIP), like e.g. AAs.

The decision to accept or deny derived by the PDP is then handed over to the PEP which has to enforce it. The PEP should be provided with a default rule (accept or deny) that it must enforce when the PDP is unable to decide (e.g. due to insufficient information from PIPs).

9.4 Attribute Authorities

The collection of attributes of the requestor can be orchestrated by the User Client or it can be initiated by the Policy Decision Point at the resource provider. The former solution seems to be more scalable, at least for the attributes of the requestor in his own organisation.

Since the attributes may be stored in different kinds of databases with differing interfaces, it will be necessary to define a standardised protocol for transporting the attributes over the Grid and an interface for plugins to be developed for the individual underlying databases.

The transport protocols will be based on X.509 Attribute Certificates [11] using the ASN.1 format [12] or the XML-coded Security Assertion MarkupLanguage (SAML) [13].

9.5 Authorisation

Plugins replacing the monolithic UUDB have to be developed that implement the authorisation architecture described in 9.3. The Explicit Trust Delegation will have rules in the policy and will be decided by the PDP.

For the formulation of the policies standard languages will be used, like the eXtended Access Control Markup Language (XACML) [14]. They must be able to describe in easy to learn ways simple policy models as well as complex requirements.

10 Authorisation in the Non-UNICORE World

10.1 VOMS

VOMS manages VOs and their constituency. Users can request to be added to the VO and VOMS managers will accept or deny the request. Users can be assigned attributes and capabilities. At the lowest sophistication level VOMS generates for the Globus middleware on each of the systems available to the VO the gridmap-file which contains the mapping of DNs to user identifications known to the OS.

VOMS performs only PIP functions. The PDP function is left to the Globus Gatekeeper.

10.2 Shibboleth

Shibboleth [15] is a middleware that provides a federated authorisation infrastructure for Web Single SignOn across organisational boundaries. It uses SAML v1.1 for the exchange of attributes.

Shibboleth passes the authorisation information in form of opaque handles which provide anonymity of users without loosing the capability to trace them back, if necessary.

10.3 GridShib

The project GridShib [16] integrates Shibboleth with Grid technology as provided by the Globus Toolkit version 4 (GTK 4). One of the major challenges is the efficient mapping of Shibboleth's opaque handle with the DN of the certificates used in GTK 4.

10.4 PERMIS

PERMIS [17] is a "Privilege Management Infrastructure" that provides a complete policy-based authorisation service. Policies are written in XML to support an RBAC paradigm.

10.5 GridShibPERMIS

The GridShibPERMIS project [18] combines the strengths of Shibboleth as an Identity and Attributes Provider, the Grid Infrastructure of GTK 4 and the PDP provided by PERMIS.

The authentication based on the X.509 certificates is performed by GTK, GridShib provides the PIP, PERMIS provides the policy-based authorisation system with its interface called "GridShibPERMIS Context Handler" acting as the PDP in the GTK authorisation framework.

11 Conclusion

UNICORE provides a solid framework for Grid computing that has already started to inter-operate with other Grid middleware like Globus, has a solid security infrastructure for a rather small, not too mobile user and resource community without the need to leave the UNICORE environment. When it comes to communications with other security schemes the isolation of the approach precludes really secure connections and information transmissions.

Since the inception of UNICORE the understanding of security on the Grid has evolved towards more flexibility while providing more control over integrity and privacy of information and usage of resources. UNICORE/GS, the follow-up to the UNICORE framework used today, must provide a completly overhauled security infrastructure. A look into developments in and surrounding the Globus Toolkit provides guidelines and ideas for the development of a new UNICORE Security Infrastructure Model, based on policies that take into consideration the identity as well as attributes of users and resources.

The existence of standards for the expression and communication of attributes and rules will make the inter-operation with other Grid middleware easier than in the past. Even the problem of delegation of trust, which is the big hurdle for a bi-directional UNICORE-Globus inter-operation should be solvable.

References

1. Goss-Walter, T., Letz, R., Kentemich, T., Hoppe, H.-C. and Wieder, P.: An Analysis of the UNICORE Security Model, Global Grid Forum, Grid Forum Document - Informational 18 (GFD-I 18), 2003, http://www.gridforum.org/documents/GFD.18.pdf
2. Erwin, D. (ed.): UNICORE Plus Final Report, 2003, ISBN-3-00-011592-7, http://www.unicore.org/documents/UNICOREPlus-Final-Report.pdf
3. Grimm, Ch. and Pattloch, M. (coord.): Analyse von AA-Infrastrukturen in Grid-Middleware, Version 1.1, March 2006 http://www.d-grid.de/fileadmin/user_upload/documents/DGI-FG3-4/Analyse-AAI_v1_1.pdf
4. Housley, R., Polk, W., Ford, W. and Solo, D.: Internet X.509 Public Key Infrastructure — Certificate and Certificate Revocation List (CRL) Profile, IETF RFC 3280, April 2002, http://www.ietf.org/rfc/rfc3280.txt
5. http://www.gridpma.org

6. http://www.ogsadai.org.uk
7. http://forge.gridforum.org/projects/dais-wg/
8. Alfieri, R. et al.: From gridmap-file to VOMS: managing authorization in a GRID environment, April 2004, http://infnforge.cnaf.infn.it/docman/view.php/7/61/voms-FCGS.pdf,
9. Tuecke, S., Welch, V., Engert, D., Pearlman, L. and Thompson, M.: Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile, June 2004, IETF RFC 3820, http://www.ietf.org/rfc/rfc3820.txt
10. Snelling, D., van den Berghe, S. and Li, V. Q.: Explicit Trust Delegation: Security for Dynamic Grids, Fujitsu Sci. Tech. J., 40,2,pp.282-294, December 2004, http://www.fujitsu.com/downloads/MAG/vol40-2/paper12.pdf
11. Farrell, S. and Housley, R.: An Internet Attribute Certificate Profile for Authorization, April 2002, IETF RFC 3281 http://www.ietf.org/rfc/rfc3281.txt
12. CCITT Recommendation X.208: Specification of Abstract Syntax Notation One (ASN.1), 1988
13. Security Assertion Markup Language (SAML) v2.0, OASIS Standard, 2005, http://docs.oasis-open.org/security/saml/v2.0/saml-2.0-os.zip
14. eXtensible Access Control Markup Language (XACML) 21.0, OASIS Standard, 2005 http://docs.oasis-open.org/xacml/2.0/access_contrpl-xacml-2.0-core-spec-os.pdf
15. http://shibboleth.internet2.edu
16. http://gridshib.globus.org
17. http://www.permis.org
18. Chadwick, D.W., Novikov, A. and Otenko, O.: GridShib and PERMIS Integration, Terena Networking Conference 2006, 15-16 May 2006, Catania (Sicily), Italy http://www.terena.nl/events/tnc2006/core/getfile.php?file_id=753