Re-thinking Grid Security Architecture

Demchenko, Y.; de Laat, C.; Koeroo, O.; Groep, D.

Published in:
Proceedings: Fourth IEEE International Conference on eScience: eScience 2008

Fourth IEEE International Conference on eScience

Re-thinking Grid Security Architecture

Yuri Demchenko, Cees de Laat
System and Network Engineering Group
University of Amsterdam
Amsterdam, Netherlands
demch@science.uva.nl, delaat@uva.nl

Oscar Koeroo, David Groep
NIKHEF
Amsterdam, Netherlands
okoeroo@nikhef.nl, davidg@nikhef.nl

***Abstract*—The security models used in Grid systems today strongly bear the marks of their diverse origin. Historically retrofitted to the distributed systems they are designed to protect and control, the security model is usually limited in scope and applicability, and its implementation tailored towards a few specific deployment scenarios. A common approach towards even the "basic" elements such as authentication to resources is only now emerging, whereas for more complex issues such as community organization, integration of site access control with operating systems, cross-domain resource provisioning, or overlay community Grids ("late authentication" for pilot job frameworks or community-based virtual machines) there is no single coherent and consistent "security" view. Via this paper we aim to share some observations on current security models and solutions found in Grid architectures and deployments today and identify architectural limitations in solving complex access control and policy enforcement scenarios in distributed resource management. The paper provides a short overview of the OGSA security services and other security solutions used in Grid middleware and operations practice. However, it is becoming clear that further development in Grid requires a fresh look at the concepts, both operationally and security-wise. This paper analyses the security aspects of different types of Grids and a set of use cases that may require extended security functionality, such as dynamic security context management, and management of stateful services. Recent developments in open systems security, and revisiting basic security concepts in networking and computing including the OSI Security Architecture and the concepts used in the Trusted Computing Base provide interesting examples on how some of the conceptual security problems in Grid can be addressed, and on how the shortcomings of current systems and the frequently proposed "ad-hoc" stop-gaps for what are in fact complex security manageability problems may be avoided. This paper is thus intended to initiate and stimulate the wider discussion on the concepts of Grid security, thereby setting the scene for and providing input to a Grid security taxonomy leading to a more consistent Grid Security Architecture.**

Keywords – Grids; Open Grid Security Architecture; Trusted Computing Base; Reference Monitor; Security models; Security Context; Authentication; Authorisation session.

I. INTRODUCTION

In less than a decade Grids have developed from initial research idea to production ready technology and infrastructure. The initial Grid definition in one of the Grid foundational papers the "Anatomy of the Grid" [1] actually described the goal of this new technology at that time: "Grid systems and applications aim to integrate, virtualise, and manage resources and services within distributed, heterogeneous, dynamic "virtual organizations". The more detailed Grid definition developed in later works included such main components as distributed infrastructure, dynamics, virtualisation, and user-defined security – the components that provide a framework for coordinated collaborative resource sharing in dynamic, multi-institutional virtual organizations [1, 2].

The Open Grid Services Architecture v1.5 (OGSA) published by the Open Grid Forum[1] (OGF) in 2006 defines the Grid as "A system that is concerned with the integration, virtualization, and management of services and resources in a distributed, heterogeneous environment that supports collections of users and resources (virtual organizations) across traditional administrative and organizational domains (real organizations)" [3]. In the recently published document GFD.113 the Grid definition is extended to "Scalable, distributed computing across multiple heterogeneous platforms, locations, organisations" [4]. The document defines the following characteristics and goals of Grids in general:

- Dynamic Resource provisioning
- Management of Virtualised Infrastructure
- Resource pooling and sharing
- Self-monitoring and improvement
- Highest quality of service

The following Grid types are identified depending on usage and required common functionality:

Cluster Grids – that have predominantly homogeneous structure and focused on shared use of high performance computing resources.

Collaboration Grids – that are targeted at supporting collaborative distributed group of people over multiple domains and involving heterogeneous resource.

Data Center Grids – are actually adding provider specific aspects in managing resources, users, their associations and supporting whole provisioning life-cycle.

Grid security is identified as one of priority areas but in the recent and current developments at OGF it is mostly focused on the short-term goals to achieve interoperability of

[1] http://www.ogf.org/

978-0-7695-3535-7/08 $25.00 © 2008 IEEE
DOI 10.1109/eScience.2008.53

currently being developed Grid infrastructures, in particular such main security services and mechanisms as Authentication (AuthN), Authorisation (AuthZ) and Web Services Protocol Security [3]. As a fact of accomplishment of this priority goal, the OGF recently published a set of documents: OGSA Security Profile 2.0 (GFD.138 [5], also referred to as "Express Authentication Profile"), Secure Communication Profile 1.0 (GFD.132 [6]), and Secure Addressing Profile 1.0 (GFD.131 [7]).

It can be also mentioned that there is a gap between OGSA Security model/services definition and existing practical Grid implementation in large Grid projects such as LCG/EGEE[2], OSG[3]. These Grid infrastructures use different implementations of Grid middleware and successfully made them working together. Some practical interoperability initiatives came out of these projects and have been brought to OGF, but many others still remain developed outside of the OGF standardisation process. Authors are involved into some of such initiatives and have an intention to bring them to the OGF standardisation process in short term. Meantime we propose this paper as a summary of our ongoing research and development work and gained experience to facilitate early discussions in wider Grid community.

The paper is organized as follows. Section 2 discusses the major use cases for Grids and required security services. Section 3 provides insight into practical Grid security that is to large extent based on authors' practical experience in developing Grid security middleware.

Section 4 provides comparative overview of the two basic security concepts: the OSI/Internet Security architecture that became a common approach and technology for modern networked applications and the Trusted Computing Base (TCB) that is originated from the mainframe technologies and primary focused on the protected computing environment. The section also provides short overview of the Trusted Computing Platform Architecture (TCPA) that develops TCB for modern networking environment.

Section 5 discusses a number of suggested research and development areas that are originated from the practical requirements and may provide missing components build consistent Grid Security Architecture (GSA).

II. Security in Grid Resources and Users Management

The three types of Grids defined in OGF Roadmap document [4] the Cluster Grids, the Collaboration Grids and the Data Center Grids provide a good basis for identifying basic common and specific security functionalities required in each case. It is not a goal of this paper to make detailed specification of all required security functionalities but we simply point on or refer to some differences between required security services/infrastructure operations.

Although the Cluster Grids deal with potentially homogeneous computing environment, the major security challenge/problem here is that the required security solutions need to bridge between open services oriented environment (basically using Web Services or other messaging platform over open Internet or networking environment) and "closed" job execution environment that is typically UNIX based. These two realms use different operational and security models which we discuss later.

The Collaboration Grids need to solve a task of managing distributed multidomain/multi-organisational user and resource associations, which in current Grid practice called Virtual Organisations. Such associations may be static or created dynamically, and Grid resources also may be assigned to VO statically or provisioned dynamically for some experiments. And so, the security infrastructure needs to support inter-domain attributes, policies, and trust management.

The Data Center Grids bring the whole spectrum of the security aspects and problems related to typical provider operation. We can just mention that few of them are related to defining a general Grid resource provisioning model, securing virtual execution environment, and user session management.

It is important to discuss another use case the provisioning of the dedicated high-speed network infrastructure. Although network provisioning tends to use the Grid middleware and consequently manage network as Grid resources, it can bring a new experience and the generic solutions from the multidomain network resource provisioning which can be used for developing common provisioning and security architecture for Grid enabled resources.

Based on their extensive experience in both networking and high-performance Grid computing, authors have a good opportunity to bring together and combine experience from two areas to develop effective and easy manageable security solutions for both Grid and network resources.

III. Practical Grid Security

A. Grid Middleware

Grid infrastructure and applications rely on the Grid middleware that provides a common communication/messaging infrastructure for all resources and services exposed as Grid services, and also allows for a uniform security configuration at the service container or messaging level. This significantly simplifies development of Grid-based applications and allows developers to focus on application-level logic. Recently, Grid middleware being developed in the framework of large international projects and consortia such as EGEE, OSG, Globus Alliance[4] and UNICORE Forum[5] has reached a production level of maturity.

The following describes a commonly accepted practice in the Grid middleware security. Authentication in Grids is based on PKI and can use different types of (user) credentials (PKI, SAML, Kerberos tickets, password, etc.). Delegation (restricted and full) is a necessary mechanism in Grids to

[1] [2] http://www.eu-egee.org/

[2] [3] http://www.opensciencegrid.org/

[3] [4] http://www.globus.org/

[4] [5] http://www.unicore.eu/

manage distributed Grid job submission and staged execution. Delegation is implemented by using X.509 Proxy Certificate. The Proxy certificate is generated by the user client or other entity acting on behalf of a user based on the user master PKC or previous Proxy certificate.

Authorisation is based on the VO attributes assigned to a user by the VO and typically managed by the VO membership service (VOMS). VOMS attributes are provided as VOMS Attribute certificate or VOMS attribute assertion and typically included into the Proxy. These user capabilities will be evaluated by the AuthZ services against the authorisation policy when requesting access to a resource. In fact, Proxy with AC/assertion can be treated as user session credentials and support simple session management functionality.

One of the important functional and structural components of the gLite and Globus middleware is the gLExec module that provides a gateway between open Grid infrastructure environment and protected task execution environment of the Computer Element (CE) or Worker Node (WN) [10].

Trust management is another important component of the Grid security and PKI based authentication and delegation. Trust relations are represented by a certificate chain that include Grid Certification Authority (CA) certificate and may include a number of successively generated Proxies. It is important to notice that global trust relations in Grids are maintained by the International Grid Trust Federation [6] (IGTF).

Besides OGF activity, there are numerous community driven initiatives to ensure Grid middleware interoperability. They are built around mentioned above Grid projects and consortia. One of such initiatives is the joint OSG-EGEE Authorisation interoperability Working Group that produced the common XACML-Grid attributes and policy profile [11] that is being jointly implemented by partner projects. The profile version 1.0 documented a number of common attributes and policy models for typical Grid applications and formalised use of the policy obligations in Grid what further was developed by authors as the Reference Model for Obligations Handling (OHRM) and being implemented in the GAAA-Toolkit (GAAA-TK) [12].

B. gLExec and Pilot Jobs on WN

The use of pilot job in current VO practice provides an interesting use case that expose limitations of currently used and implemented Grid middleware security model. This situation is becoming even more clear when Grid sites are trying to implement Site Central Authorisation Service (SCAS) what is motivated by needs to make policy based access control enforcement at site consistent and easier manageable.

A pilot job is submitted by the pilot "submitter" on behalf of the real job user. The following scenario is suggested when VO submits a pilot job to the batch system:

- The VO 'pilot job' submitter is responsible for the pilot behavior
- Pilot job obtains the true user job, and presents the user credentials and the job (executable name) to the site (gLExec) to request a decision on a cooperative basis

The pilot job submission AuthZ policy should address the following issues:

- Preventing 'back-manipulation' of the pilot job by user workload
- Protecting project sensitive data in the pilot environment (in particular, not revealing or changing job and user uid)
- Fair resource sharing between user job, in case of multiple user jobs.

The gLExec [10] is used as a gateway to submit both a pilot job to the CE or Grid site and a real user job to the WN. Figure 1 illustrates a case when both gLExec modules are requesting AuthZ decision from the LCAS/LCMAPS service which in their turn request a policy decision from the SCAS. The SCAS allows centralised access control policy management but requires using such policy enforcement mechanism as policy obligations, as they are defined in the XACML policy language [13], to instruct gLExec about required user ID mapping.

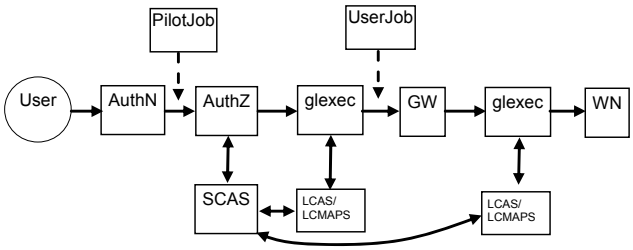


Figure 1. Pilot job submission and involved components.

Current SCAS and XACML-Grid profile implementation allows only simple obligations handling at the time of receiving reply from SCAS without using special mechanism to store AuthZ session context. This could be sufficient for simple AuthZ scenario but in case when an additional AuthZ decision is required when submitting a user job, some security mechanism must be provided to store and communicate the security context of the pilot job AuthZ decision and use it later when evaluating user job submission request.

This can be achieved by using a general AuthZ session context management mechanism such as AuthZ ticket and token which we discuss later.

C. Grid Operational Security Practice

Grid operational security scope includes but not limited to such issues as developing Grid Centers and Grid sites operational security policies, user policies, developing policies and procedures for security incident response, vulnerabilities analysis and security risk assessment. Importance of these tasks was recognised when Grid infrastructures moved to pre-production operation in such projects as EGEE, LCG, OSG. Most of these issues are addressed by the Joint Security Policy Group (JSPG).

[5] [6] http://www.gridpma.org/gridpma.html

The JSPG have developed a set of operational security policy documents that currently used in practical Grid infrastructure operation. At the same time the JSPG identified a number of issues that should be addressed in Grid security. Some issues that required wider cooperation in Grid community have been brought to the Grid Inter-operation Now (GIN) Working Group (GIN-WG) that is focused on urgent solution to ensure interoperability of Grid applications and infrastructure.

It can be suggested from the business IT security practice defined by a number of industry standards, that consistent operational security and risk management should be built on the solid conceptual basis. Some known to authors initial attempts by commercial risk management companies to create an operational and risk evaluation model for Grids has not been resulted in something more than attempts to frame Grid security model into the standard IT security models which are inheritingly built either using Internet/OSI security model or Common Criteria [14] based on the Trusted Computing Base model [15].

In attempt to create a practical model for security risk assessment when analysing Grid vulnerabilities the authors proposed the zone security model for Grid/Web services [16] that defines a number of security zones for the resource or target applications protected by such security measures and services as secure communication channels, secure user login, application container, AuthN, AuthZ, and finally gLExec type of gateway.

We can suggest that developing further this idea will help in creating a better security risk analysis model for vulnerabilities in Grids.

IV. Two Basic Security Concepts – Historical Overview

Current OGSA/Grid Security services model adopted Web Services security model which in its own turn inherited approach and basic concepts of the OSI Security Architecture and consequently the client/service security model. However, Grid operation generically deals with the managed objects, which are jobs, processes and assigned resources. This creates a gap between inherited limitations of the OSI client/server security model, that may be considered generically stateless, when trying to solve managed objects security problems which in general require stateful services.

The shortage of many proposed and currently used solutions in Grid security motivated the authors to revisit basic security concepts in networking and computing, in particular, the OSI Security Architecture and the security concepts used in Trusted Computing Base (TCB) such as Reference Monitor (RM), Multi-Level Security (MLS), Clark-Wilson integrity and manageability model, which were resulted from mainframe oriented security research in 1970s-80s.

The following provides a short overview of the two security concepts that will provide a necessary context for considering possible deeper research into developing consistent GSA.

A. From OSI/Internet to WSA and OGSA Security

Current Internet infrastructure and networking technologies are built in compliance with the Open Systems Interconnection (OSI) model. The OSI Security Architecture (ISO7498-2/X.800 [17]) provides a common framework and approach for developing secure protocols and applications. The ISO/ITU standards specify the basic security services and mechanisms and their relation to the OSI layers. The standard also suggests relations between security services and security mechanisms. The OSI security architecture is fully applicable to the Internet TCP/IP protocol stack due to their direct mapping at the data, network, and transport layers.

Security services, in the context of the OSI security architecture, are defined as services, provided by a protocol layer of communicating or interacting systems, which ensure adequate security of the systems or data transfers. To ensure openness and interoperability of interacting systems, the services are defined for specific OSI layers and may use one or more security mechanisms. Security policies are used to manage security services and can be a part of an application specific security service implementation.

The philosophy behind OSI security architecture is that security services and mechanisms can be added independently using standard/specified interfaces (as illustrated on Fig. 2). For actual security mechanisms and services matching to the OSI layers refer to the X.800 standard). The following are inherited key features of the OSI/Internet security architecture:

- Internet/OSI model suggests that interconnected systems are managed independently and communicate using protocols specific to each OSI/Internet layer.
- Trust relations between systems are established mutually or via 3rd trusted party, a group of system can create an administrative and trust domains.
- Public Key Infrastructure (PKI) provides a basis for trust management, authentication and key exchange
- Communication and security protocols can use a session related security context.

The same philosophy was inherited by the whole development of the Internet and web based applications and later by the Web Services Architecture (WSA) [8, 9] and consequently by the OGSA services model [3].

The WS-Security services model uses actually the same approach in defining security services interfaces that use the SOAP message header for adding security related information and context. This makes the security services independent from the main service call which is typically placed into the SOAP message body [9]. In this respect WS-Security services can be also considered as orthogonal to main services and in general arbitrary combined. This confirms that current Web Services Security architecture inherited basic principles from the OSI/Internet Security Architecture.

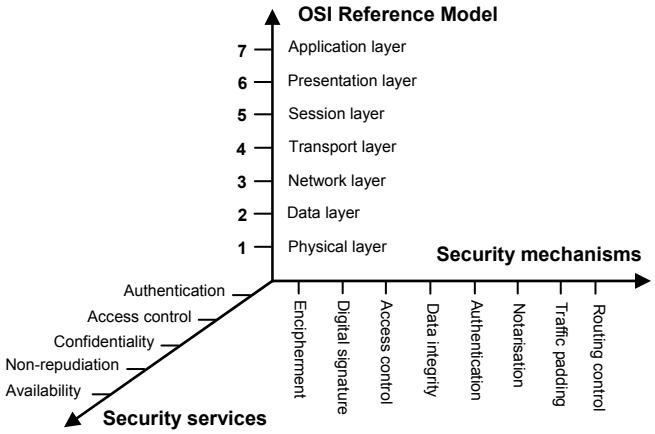


Figure 2. Relation between OSI security services, mechanisms and OSI reference model layers.

We can also make an observation that the introduction of the Web Services Resource Framework (WSRF) [18] and recent developments of the Web Services Resource Transfer (WS-RT) [19] are other attempts to address the problems of managing stateful processes in Grids with generically stateless Web Services.

B. Trusted Computing Base and Reference Monitor Concept

Reference Monitor (RM) concept was proposed by J.P. Anderson in the report "Computer Security Planning Study" (1972) [20] and was used as a basis for developing Trusted Computing Base (TCB) concept and architecture. As originated from the military research, the RM property provides a basis for Multi-Level Security (MLS) that can be abstracted as:

Complete mediation: The security rules are enforced on every access, not just, for example, when a file is opened.

Isolation: The reference monitor and databases must be protected from unauthorized modification.

Verifiability: The reference monitor's correctness must be provable. That is, it must be possible to demonstrate mathematically that the reference monitor enforces the security rules and provides complete mediation and isolation.

The following can be regarded as the basic security models used in TCB and MLS:

- Bell–LaPadula (BLP) MLS policy model [21] to protect data confidentiality that can be described as "No write down" and "No read up
- Biba model [22] to ensure data integrity that can be described as "No write up" and "No read down". Biba model can be applied to control and management data protection in an open environment.
- Clark-Wilson data integrity policy model [23] that defines both policy enforcement and certification rules that can be shortly summarised as:
  - Authentication of all user accessing system
  - Logging and auditing all modifications
  - Well-formed transactions
  - Separation of duties

The Clark-Wilson model was initially proposed to ensure reliable business operation, it is used in developing internal OS security management policies, and in Grids it can be also applicable for creating Grid Data operational security policies.

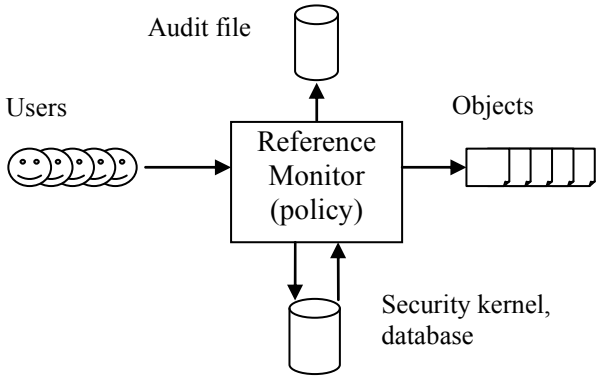


Figure 3. Reference Monitor model (applied in an "orthogonal" way to all system calls).

C. Trusted Computing Platform Architecture (TCPA)

The TCPA [24, 25] provides a basis for building and managing controlled secure environment for running applications and processing (protected) content and can be considered as TCB development for open networking environment.

The TCPA defines the five abstract layers: platform, system (including OS), service/application, and user identity. It is built around the functionality of the Trusted Platform Module (TPM) [25] - a chip built-in into the computer system or a smartcard chip that provides a number of hardware based cryptographic functions to ensure integrity and trust relation between TCPA layers. The following TPM functions are specifically targeted to improve privacy protection in TPM based systems: Endorsement Key (EK) that allows anonymous TPM identification through "zero knowledge" cryptography (without revealing actual identity or secret), the Direct Dynamic Attestation (DAA) that can securely communicate information about the static or dynamic platform configuration. In respect to the trust management, the TPM provides a platform-tied "root of trust" that can be used for secure platform registration and as an initial trusted secure session initiation (also referred to as "trusted introduction").

The TCPA architecture has been developed with the philosophy of covering the whole TP life cycle that includes six phases presumably supported by three types of infrastructures: pre-deployment/provisioning (includes manufacturing, delivery phases), deployment (includes deployment, identity registration, operation phases), and redeployment/retirement (includes recycling and retirement phases). In this respect the TCPA lifecycle model/stages can be naturally integrated with the discussed below the Complex Resource Provisioning model.

The Trusted Network Connect (TNC) [26] is a part of the TCPA that specifies how the network

component/infrastructure can be integrated into TCPA to enforce security policies before and after endpoints or clients connect to multi-vendor environment.

In section 5.3 below we will discuss how the TCPA and TPM can be used to build user-controlled virtual workspace service.

V. Suggested Research Areas for Grid Security Architecture

This section provides suggestions about possible research in the Grid security that may bring to the definition of more consistent GSA. In particular, we are trying to address such issues as defining Complex Resource Provisioning (CRP) model to provide basis for Grid security services integration with the upper layer scientific workflow, user and authorisation session management, defining mechanisms to express and communicate security context between services and domains, user centric and user controlled security services environment, secure invocation of a remote virtualised execution environment.

We also describe some of the suggested solutions that were resulted from our research into different areas of the Grid security. We hope that such information will be helpful for other researchers in this area.

A. Complex Resource Provisioning Model

The whole lifecycle of the Grid resources provisioned on-demand can be abstracted to the common Complex Resource Provisioning (CRP) model. Such abstraction can provide a basis for defining GSA that should answer the major Grid operational models.

A typical on-demand resource provisioning process includes four major stages: (1) resource reservation, (2) deployment (or activation), (3) the reserved resource access/consumption, and additionally (4) resource de-commissioning after it was used. In its own turn, the reservation stage includes three basic steps: (a) resource lookup, (b) complex resource composition (including alternatives), and (c) reservation of individual resources.

It may be also observed that for long running scientific experiments (months and years) there may be a need to define another CRP stage "resource relocation". In the current applications and experiments this problem is reduced to either moving virtual execution environment image or just moving data. However, if a relocation is required for multiple involved machines or more security restriction are applied to data, this case will need more precise definition of the security model and related procedures.

The reservation stage may require the execution of complex procedures that may also request individual resources authorisation. This process can be controlled by the central advance reservation system or meta-scheduling system and driven by the provisioning workflow and related security policy. At the deployment stage the reserved resources are typically bound to the reservation ID, which we will refer to as the Global Reservation Identifier (GRI). The de-commissioning stage is considered as an important stage in the whole resource provisioning workflow from the provider point of view and may include such important actions as global provisioning/access session termination and user/process logout, log information sealing, accounting and billing which are currently considered as separates actions outside of the general provisioning workflow.

The rationale behind defining different CRP workflow stages is that they may require and can use different security models for policy enforcement, trust and security context management, but still may need to use common dynamic security context.

The authors have implemented the CRP model in application to the multidomain Network Resource Provisioning authorisation infrastructure (GAAA-NRP) [27] in the framework of the Phosphorus project.

Defining and applying CRP model for Grid specific resource provisioning will aim two goals: building consistent security architecture that will ensure integrity of the whole resource life-cycle, and provide a better formalised framework for Grid services integration into more general e-Science workflow.

B. User and Authorisation Session Management

User and AuthZ session management is considered as an important function when applying access control to managing stateful processes and resources.

The security context and session management are widely used in modern web based applications what can provide a good base for developing similar solutions for Grids that should address such specific requirements as policy-controlled/restricted delegation (currently solved with the Proxy certificate), supporting policy obligations (addressed in XACML-grid profile [12] and discussed below), and others.

The authors addressed this problem in developing AuthZ service for Grid based collaborative applications and for NRP [27, 28] by using AuthZ tickets and token that when used together can address both extended AuthZ context management and performance issues. The proposed and currently implemented in the GAAA-TK solution supports two types of AuthZ tickets: proprietary, and based on the SAML 2.0 Assertions format [29] and SAML 2.0 Profile of XACML [30].

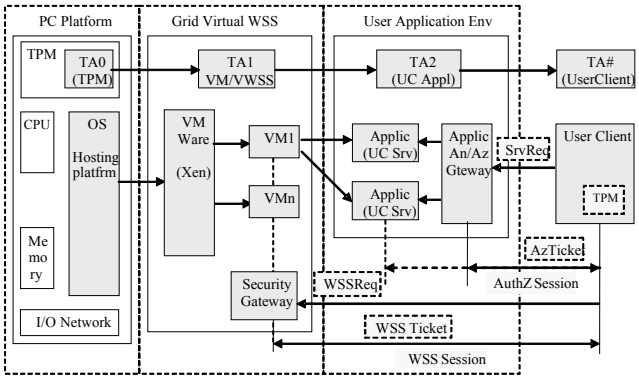
C. Extending User Controlled Security Domain in Virtualised Workspace Service (VWSS)

Modern paradigm of remote distributed services and digital content providing makes security and trust relations between User and Provider more complex. A user and a service provider are two actors concerned with own Data/Content security and each other System/Platform trustworthiness

Figure 4 depicts the proposed in authors work [31] the 3-layer VWSS-UC environment for running user tasks and applications that provides integral protection of user tasks/applications at all three layers. The three layers include: a TCPA/TPM based computing/hosting facility, a Grid based Virtual Workspace Service, and a User Application Environment. The solution extends the original Virtual WorkSpace Service (VWSS) concept [32] and is capable of scaling over multiple administrative and trust domain and

allows for running multistage user tasks or complex resource provisioning.

A virtual workspace is created after a user request is sent to the VWSS security gateway, which checks user credentials and deploys the VM based workspace with characteristics that meet the request's requirements. Such a virtual workspace creates a trusted environment where users can run their tasks or applications. User applications and/or tasks are protected by basic security services to avoid potential data compromise or interruptions. This is first of all achieved by user Authentication (AuhtN) and Authorisation (AuthZ) provided by the Application AuthN/AuthZ Gateway. In the case of complex/multi-component services, their combinations should be secured through the applications level security context management.



- Trust Anchors: T0 (TPM) – TA1 (VM/VWSS) – TA2 (Application) – TA# (User)
- WVSS session and Application AuthZ sessions

Figure 4. Three-layer security model of the VWSS-UC.

For the dynamic security context management, the VWSS-UC distinguishes between a WSS session and an application/service AuthZ session that is related to the user task or application. WSS session may have wider security context but still both of the session types are based on the positive authorisation decision and will require a similar AuthZ context management. WSS sessions that includes VWSS request may also need to incorporate a negotiation stage and possibly want to verify the platform security configuration and/or integrity, which could be achieved through the TPM based mechanisms.

In the proposed architecture, the TPM with its hardware-based secure ID allows for "bootstrapping" a chain of trust to the TMP and hardware platform. This creates a continuous chain of trust from the user to the workspace environment and hosting platform: TA#-TA2-TA1-TA0., where TAn – are trust anchors as shown on the picture.

D. Policy Obligations – Bridging Two Security Concepts

In many Grid applications, policies may specify actions that must be performed either instead of or in addition to the policy decision. In the XACML specification [13], obligations are defined as actions that must be performed in conjunction with policy evaluation on a positive or negative decision. In this way and when using together with gLExec, policy obligations can be used for defining actions that will be performed by the gLExec when submitting Grid jobs to the protected execution environment.

Obligations are included into the policy definition and returned by PDP to PEP which in its turn should take actions as prescribed in the obligation instructions or statements. In the context of the GSA, obligations provide an important mechanism for policy decision enforcement in the provisioned Grid resources, in particular, mapping global user ID/account to local accounts or groups, assigning quotas, usage limits, etc.

The proposed obligations handling model is described in details in [33] and allows two types of obligations execution: at the time of receiving obligations from the PDP and at the later time when accessing a resource or performing an authorised action. The latter can be achieved by using AuthZ tickets or SAML assertions that hold obligations together with AuthZ decisions.

E. Using Identity Based Cryptography for building Dynamic Security Associations

Trust management is an important issue and a problem in Grid security. It would not be a complete overview of possible research areas in developing a consistent GSA if we not mention the Identity Based Cryptography (IBC) [34]. The IBC allows using recipient's public credentials to generate the encryption key when sending a message to the recipient, and the user can request the local IBC Key Generation Server (KGS) to obtain own private key.

IBC in application to Grid has been a topic for many research projects and papers in the academic community (see for example [35]) but it is still less known for Grid practitioners. We expect that IBC can provide a simple way of building dynamic interdomain trust relations or distributing security context between domains that doesn't have direct trust relations. Such an approach will use pre-configured IBC KGS to distribute security information between domains, and in this way "exchange" the IBC based intra-domain trust infrastructure for simpler trust and key management in dynamic multidomain applications.

VI. SUMMARY

In this paper we describe a set of research areas where prevalent Grid security solutions and today's architectures may no longer be able to in themselves provide a single consistent security architecture: provisioning of ensembles of resources across multiple domains whilst maintaining consistent life cycle management, authorization session management, user-controlled security domains (also refrred to as virtual overlays), enforcement of policy where resource usage is subject to additional policy obligations that may be resource-specific and depend on the resource state, and the use of identity-based cryptography for dynamic security associations.

By analyzing several currently deployed Grid security systems and architectures in this paper we have attempted to indicate the limits of their applicability. Although by no means exhaustive in itself, we expect that tackling the new research areas in a consistent and comprehensive way will lead to developing a security architecture that can encompass

both the existing set of scenarios as well as being able to deal with the more complex scenarios, without having to resort to 'one-off' and ad-hoc solutions.

The concepts proposed in this paper may provide the basis for the discussion of a comprehensive Grid security taxonomy, and for the development of a consistent Grid Security Architecture amongst the Grid community at large.

ACKNOWLEDGMENT

This work is supported by the FP6 EU funded Integrated project PHOSPHORUS (Lambda User Controlled Infrastructure for European Research) IST-034115.

REFERENCES

[1] Foster, I., Kesselman, C. and Tuecke, S. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," International Journal of Supercomputer Applications, 15 (3). 200-222. 2001.

[2] Foster, I., Kesselman, C., Nick, J. and Tuecke, S. "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration," Globus Project, 2002. [Online]. Available: http://www.globus.org/research/papers/ogsa.pdf.

[3] GFD.80 "The Open Grid Services Architecture, Version 1.5," I. Foster, H. Kishimoto, A. Savva, D. Berry, A. Grimshaw, B. Horn, F. Maciel, F. Siebenlist, R. Subramaniam, J. Treadwell, J. Von Reich. Open Grid Forum, Sept. 5, 2006.

[4] GFD.113 "Technical Strategy for the Open Grid Forum 2007-2010," D. Snelling, C. Kantarjiev. Open Grid Forum, Aug. 7, 2007.

[5] GFD.131 "Secure Addressing Profile 1.0," D. Merrill. Open Grid Forum, June 13, 2008.

[6] GFD.132 "Secure Communication Profile 1.0," D. Merrill. Open Grid Forum, June 13, 2008.

[7] GFD.138 "OGSA Basic Security Profile 2.0," D. Snelling, D. Merrill, A. Savva, Open Grid Forum, July 28, 2008.

[8] Web Services Architecture. W3C Working Group Note 11 February 2004. [Online]. Available: http://www.w3.org/TR/ws-arch/

[9] Web Services Security Roadmap (2002). [Online]. Available: http://www.ibm.com/developerworks/library/specification/ws-secmap/

[10] Groep, D., O. Koeroo, G. Venekamp, "gLExec: gluing grid computing to the Unix world," Journal of Physics: Conference Series 119 (2008) 062032

[11] An XACML Attribute and Obligation Profile for Authorization Interoperability in Grids. [Online] Available: https://edms.cern.ch/document/929867/1

[12] Pluggable GAAA-TK library. [Online]. Available: http://staff.science.uva.nl/~demch/projects/aaauthreach/index.html#aaauthreach

[13] Godik, S. et al, "eXtensible Access Control Markup Language (XACML) Version 2.0," OASIS Working Draft 04, 6 December 2004. - http://docs.oasis-open.org/ xacml/access_control-xacml-2_0-core-spec-cd-04.pdf

[14] ISO 15408:2005 Information technology - Security techniques - Evaluation criteria for IT security, 2005

[15] Abrams, M.D., M.V. Joyce, "Trusted Computing Update," Computers & Security, Vol. 14 No.1 pp. 57-68, 1995. - http://www.acsac.org/secshelf/papers/trusted_computing_update.pdf

[16] Demchenko, Y., et al. "Web Services and Grid Security Vulnerabilities and Threats Analysis and Model," Proc. 6th IEEE/ACM International Workshop on Grid Computing, November 13-14, 2005. Seattle, Washington

[17] ITU-T Rec. X.800 Security Architecture for Open Systems Interconnection for CCITT applications. ITU-T (CCITT) Recommendation, 1991.

[18] Web Services Resource Framework (WSRF), Primer v1.2, Committee Draft 02, 23 May 2006. [Online] Available: http://docs.oasis-open.org/wsrf/wsrf-primer-1.2-primer-cd-02.pdf

[19] Web Services Resource Transfer (WS-RT) Version 1.0, August 2006. http://www.ibm.com/developerworks/webservices/library/specification/ws-wsrt/

[20] Anderson, J., "Computer Security Technology Planning Study," ESD-TR-73-51, ESD/AFSC, Hanscom AFB, Bedford, MA 01731 (Oct. 1972) [NTIS AD-758 206]. [Online] Available: http://csrc.nist.gov/publications/history/ande72.pdf

[21] David E. Bell and Leonard La Padula, "Secure Computer System: Unified Exposition and Multics Interpretation, » ESD-TR-75-306, ESD/AFSC, Hanscom AFB, Bedford, MA 01731 (1975) [DTIC AD-A023588]. http://csrc.nist.gov/publications/history/bell76.pdf

[22] Biba, K. J. "Integrity Considerations for Secure Computer Systems," MTR-3153, The Mitre Corporation, April 1977.

[23] Anderson, R., F. Stajano, J. Lee, "Security Policies." [Online]. http://www.cl.cam.ac.uk/~rja14/Papers/security-policies.pdf

[24] Trusted Computing Group (TCG). [Online]. Available: https://www.trustedcomputinggroup.org/home

[25] TCG Infrastructure Working Group Reference Architecture for Interoperability. Specification Ver. 1.0. 16 Jun. 2005. http://www.trustedcomputinggroup.org/specs/IWG/IWG_Architecture_v1_0_r1.pdf

[26] TNC Architecture for Interoperability. Specification Version 1.1, 1 May 2006. [Online]. Available: https://www.trustedcomputinggroup.org/specs/TNC/TNCArchitecture_v1_1_r2.pdf

[27] Demchenko Y, A. Wan, M. Cristea, C. de Laat, "Authorisation Infrastructure for On-Demand Network Resource Provisioning," Proc. 9th IEEE/ACM International Conference on Grid Computing (Grid 2008), Tsukuba, Japan, Sept. 29 - Oct. 1, 2008. pp. 95-103. ISBN 978-1-4244-2579-2.

[28] Demchenko Y, L. Gommans, C. de Laat, A. Wan, O. Mulmo, "Dynamic security context management in Grid-based applications", Future Generation Computer Systems (2007), ref: doi:10.1016/j.future.2007.07.015

[29] Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard, 15 March 2005. [Online]. http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

[30] SAML 2.0 Profile of XACML 2.0, Version 2. Working Draft 2, 26 June 2006. [Online]. http://docs.oasis-open.org/xacml/2.0/xacml-2.0-profile-saml2.0-v2.zip

[31] Demchenko Y., Frank Siebenlist, Leon Gommans, Cees de Laat, David Groep, Oscar Koeroo, "Security and Dynamics in Customer Controlled Virtual Workspace Organisation," Proc. HPDC2007 Conference, Monterey Bay California, June 27-29, 2007.

[32] Virtual Workspaces. [Online]. Available: http://workspace.globus.org/index.html

[33] Demchenko, Y., C. de Laat, O. Koeroo, H. Sagehaug, "Extending XACML Authorisation Model to Support Policy Obligations Handling in Distributed Applications," Proc. 6th International Workshop on Middleware for Grid Computing (MGC2008), December 1, 2008, Leuven, Belgium. In press.

[34] A. Shamir. "Identity-based cryptosystems and signature schemes," In G.R. Blakley and D. Chaum, editors, Advances in Cryptology, Proc. of CRYPTO'84, pages 47-53. Springer-Verlag LNCS 196, 1985.

[35] Exploiting Hierarchical Identity-Based Encryption for Access Control to Pervasive Computing Information. [Online]. Available:- http://stinet.dtic.mil/cgi-bin/GetTRDoc?AD=ADA457869&Location=U2&doc=GetTRDoc.pdf