Extending User-Controlled Security Domain with TPM/TCG in Grid-based Virtual Collaborative Environment

Yuri Demchenko, Leon Gommans, Cees de Laat
University of Amsterdam
{demch, lgommans, delaat}@science.uva.nl

ABSTRACT

The paper proposes an integral approach to building multilayer security for Grid based virtual collaborative environment that leverages the general user-controlled complex resource provisioning (CRP-UC) model. The CRP-UC is considered as comprising of three layers: trusted computing platform, secure virtualised workspace, and collaborative/application session. The suggestions on the technology selection are provided for the first two layers: industry adopted Trusted Computing (TCG) platform, and Virtual Workspace Service (VWSS) developed in the framework of the Globus Toolkit. Solutions and implementation are proposed and discussed for the service/application authorisation session and security context management in multidomain applications based on the GAAA Authorisation Framework that can be used with the major service-oriented AuthZ framework. The current implementation of the XML-based authorisation ticket format is discussed and possible extensions to address wider user session management issues are suggested, in particular those related to the TCG-rooted chain of trust and session context negotiation. The paper is based on experiences gained from major Grid based and Grid oriented projects including EGEE, Phosphorus, Globus Toolkit.


1. INTRODUCTION

Grid and Web Services based resources and services virtualisation allows for effective sharing of advanced computing resources and unique equipment via creation of the Virtual Laboratories (VL) or Virtual Organisations (VO) [1].

Available Grid technologies and middleware provide appropriate platform for both creating virtual workspace and collaborative virtual user associations/communities that can be created dynamically on-demand based on the experiment or project agreement, and terminated once the experiment/project has been completed or service/resource delivered or consumed.

Currently typical Grid-based Collaborative Environment (GCE) is created as project or experiment oriented and relies on previously established and offline maintained trust relations between participating members [2]. To become truly open collaborative environment and ensure the same level of security and privacy (to user data protection) as personal or organisation owned systems, the virtual workspace provisioning should rely on the multiplayer security model of the generalised user controlled Complex Resource Provisioning (CRP-UC) [3]. The CRP-UC should address the security from the user point of view at all major components or layers comprising on-demand provisioned collaborative environment or processing environment for user tasks/jobs: computer platform or facility, operating environment or system, virtual workspace, and service or application.

In the current Grid service provisioning model, dynamically created Grid resources rely on the following three components and/or actors: 1) facility and infrastructure providers/operators that are generally interested in protecting their infrastructure and facilities from outside attacks against their infrastructure directly and indirectly via hosting services; 2) service/application providers that are concerned about uninterruptible operation of their services that may be caused by hosting facilities and underlying infrastructure; 3) customers (or content providers in other business model) who are concerned both with the trustworthiness of the (service) provider infrastructure and protection of their working data in running tasks and applications.

Current Grid Security Infrastructure (GSI) and Grid security middleware effectively address service level security but at the same time it (implicitly) relies on the trusted computer platform and trusted workspace organisation what significantly limits the use of virtualised collaborative workspace to mutually trusted collaborating groups of user and/or organisations.

One of recently proposed solutions that attempts to combine services virtualisation, dynamic resource creation
and security is the Virtual Workspace Service (VWSS) being developed in the framework of the Globus Toolkit version 4 (GT4) [4, 5]. In fact, the VWSS security model has been developed from the point of view of Grid services providers and considers the computing platform as trusted. However, for more complex and security concerned use cases like CRP-UC the VWSS should be completed with means to ensure user-centric and user-controlled secure environment.

Business acceptance of virtualised Grid-based services will depend on how successfully the Grid middleware will solve creation of fully user-trusted execution environment for user tasks. The proposed solution is a combination of security solutions at three levels: trusted computing platform, trusted virtualised service environment (workspace), and user-controlled applications/tasks execution.

The goal of this paper is to further advance the development of the security model and infrastructure for building user-controlled secure virtual workspace environments (VWSS-UC) that enable user trusted services and/or execution environments and creates necessary basis for user trusted/controlled GCE. The paper looks into the typical VL organisation in the GCE that requires multidomain CRP and discusses their special requirements with respect to VWSS-UC.

The paper proposes a high-level model that combines the GT4 VWSS with the industry adopted Trusted Computing platform [6] to provide higher trustworthiness of the remote computing platform, and adds the session based user security context management using Authorisation (AuthZ) session management tools being developed in the framework of the GAAA Authorisation Framework (GAAA-AuthZ) [7, 8].

The paper is organized as follows. Section 2 describes typical VL organisation that effectively uses the domain-based resource management and related domain based security model. Section 3 describes general requirements to the Grid security middleware for the general CRP-UC use cases. Sections 4 and 5 provide short overview of the two technologies that are building blocks of the VWSS-UC security infrastructure: Grid VWSS and Trusted Computing platform. Section 6 describes the proposed three-layer security model for the VWSS-UC that incorporates both technologies and includes also applications level security services. Section 7 describes the AuthZ ticket format that can be used for the extended AuthZ session and user security context management.

2. VL ORGANISATION IN GCE

VL concept provides a flexible framework for associating instruments, resources and users into distributed interactive collaborative environment. However, committed to the VL resource still remain in the possession and under direct administration of their original owner enterprises. This will require hierarchical multidomain/multiplayer virtualised resource management and using additional security mechanisms to ensure integrity and trustworthiness of the whole VL workspace and collaborative environment.

The Domain-based resource management model (DM) closer reflects business practice among cooperating organisations contributing their resources (instruments, other facilities and operator personal) to create a Virtual laboratory that can run complex experiments on request from customers. To become consistent the DM should be supported by corresponding organisation of the access control infrastructure that involves all domains and layers [9].

Figure 1 illustrates relations between major components in the hierarchical DM resource management and security model in typical VL organisation. The following administrative and security domains can be defined for users, resources, policy and trust management [9]:

1) Facility that provides administrative/legal platform for all further operational associations; may define what kind of technologies, formats, credentials can be used.

2) VL that can be created on the basis of the VL agreement that defines VL resources, common services (first of all, information/registry and security), administrative structure and VL administrator. Trust relations can be established via PKI and/or VL Certificate population.

3) Experiment/project defined together with the VL resources allocation, members, task/goals, stages, and additionally workflow. It is perceived that experiment related context may change during its lifetime.

4) Experiment session that may include multiple Instrument sessions and Collaborative sessions that involves user security context.
experiment members into interactions.

5) Collaborative session that involves user interaction and interactive instruments control.

Both Experiment and Collaborative sessions are initiated by the user and created based on the positive AuthZ decision and consequently rely on the AuthZ session that in distributed or multidomain scenarios may be supported with the dynamic AuthZ (session) tickets/credentials.

It is essential that Trust Anchors (TA) can be assigned to hierarchical domain related entities to enable security associations and support secure communication. VL TA1 is suggested as minimum required in the DM, Experiment TA2 may be included into the Experiment description. Collaborative session security association can be supported by AuthZ tickets that can also hold session’s dynamic security context.

The Experiment description plays an important role in the DM security infrastructure. It contains all the information required to run the analysis, including assigned users and roles, and a trust/security anchor(s) in the form of the resource and, additionally, the customer’s digital signature(s). The experiment description provides experiment-dependent configuration data for other services to run the experiment and to manage the dynamic security context, in particular for possible experiment specific virtual workspace creation.

3. SECURITY INFRASTRUCTURE
REQUIREMENTS IN VL/CRP-UC

CRP-UC in multi-domain heterogeneous environment requires both a dynamic virtualised workspace environment and a user controlled secure application infrastructure.

For the purpose of current research, we can summarise that typical on-demand resource provisioning includes 2 major stages: resource reservation and the reserved resource access or consumption. In its own turn, the reservation and allocation stage includes 4 basic steps: resource lookup, complex resource composition (including alternatives), reservation of individual resources and their association with the reservation ticket/ID, and finally delivery or allocation. The reservation stage may require execution of complex procedures that may also request individual resources authorisation.

In order to get access to the reserved resources the Requestor needs to present the reservation ticket/ID together with own credentials that confirm the Requestors rights to access/use the reserved resource. In the generic policy based access control model, the request is received/intercepted by the service/resource Policy Enforcement Point (PEP) and next evaluated by the Policy Decision Point (PDP) against the resource access control policy. After the positive AuthZ/PDP decision, the AuthZ session can be created and AuthZ ticket generated containing reservation ID and full AuthZ session context. In consequent service/resource requests, if the AuthZ ticket is presented, the request can be evaluated by the PEP (locally, without consulting slower request evaluation with the PDP) and access granted based on matching between the AuthZ ticket and presented Requestor’s credentials [2, 7].

The extended AuthZ ticket functionality can be used to support effective decision making during the reservation stage.

Tools and GSI middleware supporting typical VL/CRP-UC applications should satisfy the following requirements:

- Create user/application workspaces (together with related security services) dynamically since the environment can potentially change from one task or experiment to another,
- (Securely) associate multiple administrative and trust domains (e.g., by means of the Virtual Organisation (VO) or other form of dynamic security association);
- Dynamically create user accounts and handle different/multiple user identities and credentials;
- Negotiate and handle multiple security and access control policies (for both resource provisioning and access stages) that may also include mutual authorisation between VWSS and user client;
- Manage session based user security context.
- Allow for user rights/roles delegation, including delegated hierarchical policies administration;
- Allow for binding the whole chain of trust in dynamic collaborative sessions to the VL facilities/platform root of trust.

4. VIRTUAL WORKSPACE IN GRIDS

The concept of the Virtual Workspace Service (VWSS) where configurable execution environment for running Grid services can be dynamically deployed, was proposed in [10, 11] and has been implemented in the framework of the Globus Toolkit. The technology allows for a finer-grained policy-enforcement by providing user specified workspace deployment with customary configured security services and dynamically created user accounts.

The Virtual Machine (VM) based VWSS is available currently as a technology preview implementation based on the GT4 middleware [4]. It comprises of the Workspace Factory Service (WFS) that allows a Grid client to deploy a VM/Xen-based virtual workspace, and the workspace service that allows a Grid client/user to manage a workspace by starting, stopping, pausing, or destroying it, including the creation and reservation of user accounts for the virtual workspace.

Through plug-ins, the WFS can be configured with security enforcement modules (currently, GT4 AuthZ
service using gridmap) and resource scheduling/management. The WSS itself can use the available Authentication (AuthN) and AuthZ services in the GT4.

To achieve higher trustworthiness at the level of VWSS, it was proposed to create in advance and store preconfigured serialised VM images protected by user credentials that can be deployed on user request [10]. This solution can ensure trusted VWSS environment but remains vulnerable to a VM platform compromise and cannot guarantee user credentials and data protection and further data integrity.

These problems could potentially be solved by incorporating the currently available and industry adopted Trusted Computing platform and by adding user session management functionality.

5. TRUSTED COMPUTING PLATFORM

The Trusted Computing platform (TCG), as promoted by the Trusted Computing Group, provides a foundation for building and managing controlled secure environment for running applications and processing (protected) content [5].

The TCG security model and their trustworthiness definition are a bit controversial. They are considered from the point of view of infrastructure and content providers, or system and network administrators (who may not be the system users). Client platform and users themselves are considered as not trusted or a potential source of security threats, in particular with respect to content and intellectual property right (IPR) violations. Actually, the TCG intends to make a client platform (e.g., PC/laptop) trusted to be a part of protected working or consuming environment.

This focus and the TCG’s initial goal to protect on-line content providers (i.e., video and music) caused a widely discussed concern about user privacy issues [12]. Without discussion the merits of the privacy concerns, we would like to make the observation that the Grid-resource users and the Hollywood content providers share similar concerns as in VWSS-UC, a user is also concerned about remote execution environment trustworthiness, data integrity and data confidentiality.

The TCG architecture [13] defines five abstract layers: platform, system (including OS), service/application, and user identity. It is built around the functionality of the Trusted Platform Module (TPM) [14] - 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 TCG layers:

- Asymmetric key functions for on-chip key pair generation using hardware random key generation; private key signatures; public key encryption and private key decryption.
- An Endorsement Key (EK) that can be used by a platform owner to establish that identity keys were generated in a TPM, without disclosing its identity.
- Direct Dynamic Attestation (DAA) that securely communicates information about the static or dynamic platform configuration, which is internally stored in TPM in the form of hashed values.
- Protection of communication between two TPM.
- Monotonic counter and the tick counter to enable transaction timing and sequencing.

TPM provides a platform-tied “root of trust” that can be used for secure platform registration and as an initial trusted secure session initiation (or “trusted introduction”).

Other components of the TCG architecture include (in current implementation): a “curtained memory” feature in the CPU; a security kernel in the operating system; a security kernel in each TCG application; and a back-end infrastructure of online security servers maintained by hardware and software vendors [15].

The TCG defines separate specifications for the trusted network infrastructure, client, server storage and mobile devices, and TPM Software Stack (TSS). The TSS defines a set of API’s to major security applications such as Remote Access, Identity Management, PKI, Secure e-mail, and file/folder encryption.

The TCG architecture has been developed with the following philosophy [13, 16]: incremental implementation; available as opt-in functionality; the possibility of anonymous TPM identification through “zero knowledge” cryptography; the possibility to migrate (or backup) TPM keys to another TPM without disclosing them in clear. Trusted platform (TP) lifecycle 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).

TCG Credentials specification [16] defines three types of credentials: already mentioned EK, platform key/credentials (PK), and Attestation Identity Key (AIK). EK and AIK are specified in a form of X.509 Identity Certificate and PK as an X.509 Attribute Certificate.

Pre-deployment EK pair is generated at the TPM manufacturing stage and next used at the deployment stage to generate post-manufacturing EK key pair and credentials when TP is delivered and installed at the user location. PK credentials additionally bind TPM related EK credentials to an instant platform configuration. AIK credentials are generated at the TP registration stage and provide a mechanism to protect privacy sensitive EK during platform registration and operation.

AIK credentials are generated by the platform operated/bound Privacy-CA [13]. However, in some critical
cases revealing Privacy-CA identity (as AIK issuer) is not acceptable due to confidentiality or privacy issue, also assurance level provided by the platform or site locally operated Privacy-CA may not be sufficient for some applications. In such cases the TP can use the TPM supported DAA protocol to access remote DAA service which is supported by the TP deployment/operation infrastructure.

The TCG Trusted Network Connect (TNC) platform [17] is focused on establishing and enforcing security policies before and after endpoints or clients connect to multi-vendor environments. Among other requirements that improve end-points administration, TNC defines end-point configuration measurements against compliance security policies before the connection to the network is allowed. The TNC uses the IETF AAA Authorisation Framework [6] to add TPM based policy enforcement mechanisms to the TCG network infrastructure layer. On other hand, the TNC describes how the TPM functionality can be used to improve security of communications between AAA components in an open multidomain environment, in particularly to support “trusted introduction” of new network devices and reliable key distribution in multidomain network/resource provisioning.

6. BUILDING SECURE VIRTUAL WORKSPACE FOR VL/CRP-UC

Figure 2 depicts the proposed 3-layer VWSS-UC environment for running user tasks and applications that provides integral protection of user tasks/applications at all three layers. It is capable of scaling over multiple administrative and trust domain and allows for running multistage user tasks or complex resource provisioning. The three layers include: a TCG based computing/hosting facility, a Grid based Virtual Workspace Service, and a User Application Environment.

We assume that at the time of requesting access to the service or application, the resource reservation and allocation have been done and the user can reference it with the reservation ID. Reservation ID actually means a kind of contract based on which a service can be deployed on the remote platform in the VWSS container/environment.

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 AuthN and 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.

For the dynamic security context management, we need to distinguish between a WSS session in VWSS-UC 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 TCG or TNC mechanisms. We plan to investigate in details what existing service provisioning frameworks and protocols could provide the required functionality and how they can be used in VWSS-UC (at different stages of the VWSS-UC operation).

In the proposed architecture/model, 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: TAn – TA#-TA2-TA1-TA0, where TAn – are trust anchors as shown on picture.

Note that the VWSS trustworthiness may also be increased by using a trusted VM-repository, which stores pre-configured VM-images that are cryptographically bound to a user or to a trusted third party.

7. AUTHORISATION TICKET FORMAT

As discussed in section 6 there are two types of sessions in the proposed VWSS-UC model: WSS session and service or application AuthZ session. Although WSS session may require wider security context support, both of them will have similar AuthZ context and will require similar functionality when considering distributed multi-
domain scenario. In this section we will discuss current implementation of the AuthZ ticket in the GAAA-AuthZ and suggest possible extensions [3].

An AuthZ ticket (AzTicket) is generated as the result of a positive PDP decision. It contains the decision and all necessary information to identify the requested service. When presented to the PEP protecting the resource or service, its validity can be verified and in the case of a positive result, access will be granted without requesting a new PDP decision. Such a specific functionality is provided in the GAAAPI package with the Triage module [3].

Current AzTicket format and its implementation in the GAAA-AuthZ support extended functionality for distributed multidomain hierarchical resources access control, in particular, administrative policy management (as defined in XACML 3.0 Administrative policy profile), capabilities delegation and conditional AuthZ decision assertion (to support XACML policy obligations). The semantics of AzTicket elements is defined in such a way that allows easy mapping to related elements in other XML-based and AuthZ/AuthN related formats, like the Security Assertion Markup Language (SAML) [18] and the eXtensible Access Control Markup Language (XACML) [19, 20].

Figure 3 illustrates the AzTicket data model and shows the top elements. AzTicket that can be used for extended AuthZ session security context management. The listing also contains comments that explain a suggested mapping to SAML2.0 Authorisation assertion elements, which demonstrates that even for basic AuthZ session data, few extension elements are required for extended security context expression.

The AzTicket contains the following major groups of elements:

- **The Decisions/Decision elements** that hold the PDP AuthZ decision(s) bound to the requested resource(s) or service(s) expressed as the Result and ResourceID attributes correspondently.

- **The Actions/Action complex element** contains actions which are permitted for the Subject or its delegates.

- **The Subject complex element** contains all information related to the authenticated Subject who obtained permission to do the actions, including sub-elements: Role (holding subject’s capabilities), SubjectConfirmationData (typically holding AuthN context), and extensible sub-element SubjectContext that may provide additional security or session related information, e.g. Subject’s VO, project, or federation.

- **The Delegation element** allows to delegate the capabilities defined by the AzTicket to another Subjects or community. The attributes define restriction on type and depth of delegation

- The **Conditions** element specifies the validity constrains for the ticket, including validity time and AuthZ session identification and additionally context. The extensible ConditionAuthzSession element provides rich possibilities for AuthZ context expression.

- The **Obligations/Obligation element** can hold obligations that PEP/Resource should perform in conjunction with the current PDP decision.

![Figure 3. The AuthzTicket data model and top elements.](image)

The first three elements the Decision, the Actions/Action, and the Subject have direct mapping to related SAML elements. Other AuthzTicket elements the Delegation, the Conditions, and the Obligations/Obligation can be implemented as extensible elements that can be customary defined in SAML.

The AzTicket is digitally signed and cached by the Resource’s AuthZ service. To reduce communication overhead when using AzTicket for consecutive requests validation, the associated AuthZ token (AzToken) can be generated of the AzTicket. The AzToken may contain just two elements: **TokenID = TicketID** and **TokenValue = SignatureValue**, needed for identification of the cached AzTicket.

If considered for extended user session management and binding with the VWSS and TCG layers the following functionality should be added to the current AzTicket format:
• TPM AIK (used as the platform trust anchor) and VWSS public keys or other security credentials;
• WSS session information element similar to \(<\text{ConditionAuthZSession}\) that should contain the reference or description of VM instance;
• Elements or attributes that can support mutual AuthZ or session negotiation what is desirable to have even if the negotiation protocol will have own messages format, because the User/AuthZ session credentials have to be bound to requestor/subject credentials and their AuthN context.
• Supporting consumable resource attributes (e.g., usage time, data transferred, number of access), and additionally collecting accounting data.

To provide described above functionality of the dynamic VWSS and application authorisation sessions security context handling, special features should be added to existing Grid oriented AuthZ frameworks such as Globus Toolkit 4.0 AuthZ Framework (GT4-AuthZ) [21] or gLite Java Authorisation Framework (gJAF) [22, 23]. They are currently being developed as pluggable GAAAPI modules of the GAAA-AuthZ Toolkit to support all the necessary security context processing and communication between a PEP and a PDP [24]. They can be added as external plugins to GT4-AuthZ and gJAF frameworks as they can be called in a standard way from either PEP or PDP.

8. SUMMARY AND FUTURE RESEARCH

This paper is partly based on the results of the ongoing research and development of the security infrastructure for the CRP-UC that targets the two major use cases/application areas of collaborative resource sharing and on-demand networking resources provisioning. This work is being conducted by the authors in the framework of different international, European, Dutch nationally and industry funded projects, including EGEE, Globus Toolkit, Phosphorus, and GigaPort Research on Network.

The paper proposes a three-layer security model for organising VWSS-UC that incorporates currently available and emerging technologies such as GT4’s Virtual Workspace Service (VWSS), Trusted Computing Platform, and the GAAA-AuthZ Authorisation session management tools (currently also being implemented as a gJAF extension). Incorporating TPM’s hardware-based trust anchors/credentials into the VWSS-UC security model allows for creating dynamically a continuous chain of trust from the user client to the virtual workspace environment and hosting platform. This chain of trust will provide a basis for managing dynamic VL/VWSS or service/application authorisation sessions security context.

To achieve scalability in the security context management the two types of sessions are defined: service or application AuthZ session, and VWSS session that should allow for wider security context management and secure session negotiation. It should be also investigated what other standards, solutions or framework can be used for extended security context management during both resource provisioning/reservation and access stages, including WS-Agreement and other network oriented protocols such as TNC [17] and COPS (Common Open Policy Service) [25].

The paper proposes a format of the AuthZ ticket that allows for performance optimisation and for extended AuthZ session context management to support complex resource provisioning in distributed multi-domain environment.

More research and modeling will be required to identify additional functionality to address wider security context management in the user controlled VL/VWSS session and to incorporate mutual AuthZ and security policy negotiation into AuthZ session management and AuthZ ticket format.

Suggested further development will include more detailed investigation how the TCG and TPM can be practically incorporated into the proposed VWSS-UC architecture and integrated with the current middleware, in particular, how the support for the virtual TPM available in Xen v3.0 [26] can be used for this. This work will also rely on recent developments in the Daonity [27] and OpenTC [28] projects that provide practical examples of using TCG for improving security of user credentials and security context management in Grid applications, including fine-grained client-side VM policies management.

Additionally, it is expected that using TCM-based TCG technologies can solve known problem of protecting and storing user secure credential used for user authentication and single-sign-on (SSO), as identified in [29]. The TPM can provide the same functionality as prospective smartcard based solutions but already integrated into advanced trusted infrastructure management.

The authors believe that the proposed security model for user-controlled virtual workspace organisation in VL/GCE can also be used for other use cases that require distributed, dynamically created and/or mobile services. It intends to contribute to the development of the SOA security model in connection with services virtualisation [30].

The authors believe that the proposed approach and solutions will provide a good basis for further discussion among Grid and application security specialists and will be of interest to the industry.
9. REFERENCES


