Report from Working Group 3 of the LCG Grid Deployment Board on Registration, Authentication, Authorization and Security

M. Delfino (convenor), T. Kachelhoffer, D. P. Kelsey, R. J. Sobie, V. White

DRAFT 0.1 of LCG-GDB-WG3.

Mods by D Kelsey – structure of document and brief description of issues (25 Nov 02)

1 Scope of the document

Registration, Authentication, Authorization and Security for the LCG-1 prototype.

1.1 Definition of terms

Virtual Organisation (VO) is a set of individuals and/or sites that are defined according to a set of coordinated resource sharing rules.

Authentication is the mechanism used for verifying the identity of a principal, where a principal is a named entity, e.g. a user, a system, or a service.

Authorization is the process whereby a principal is allowed to carry out or is prevented from carrying out a requested action. In the context of Grid computing the computing infrastructure, i.e. systems, resources and services, also require authorization to function as part of a Virtual Organisation.

Registration is one of the components of Authorization. It is the process by which a list of authorized users and infrastructure is created and maintained.

1.2 Grid Security and the GDB working groups

The mandate of WG3 (see Appendix A) includes not only the request for recommendations on common security policies and procedures but also the request to identify technical security issues. The latter topic will naturally overlap with many of the other working groups; with WG1 via the choice of security middleware, with WG2 via the interaction between the management of a VO and resource scheduling and accounting and with WG4 through the security operational procedures.

[DPK: WG3 needs to decide how to tackle the overlaps with other WG's]

Wherever possible WG3 has recommended processes and procedures that are independent of the actual choice of technology.

2 Definition of LCG-1 for the purposes of this report:

LCG-1 is a worldwide **data-processing infrastructure** comprising

DRAFT

- **a dynamically updated set of data processing resources** (computers, disks, tapes, networks, etc.) and
- the **associated** configuration and input, as well as produced, **data files** and
- the **associated** system, middleware, application infrastructure and application **software,**
- which will be **explicitly provided by various institutions** in various countries
- towards the **agreed explicit common goal** of prototyping computing for LHC experiments,
- including tests of the hypothesis that it can present a **single coherent environment from the user's point of view**
- while using a **de-centralized non-homogeneous environment from the provider's point of view,**
- **built and operated by an LCG-1 engineering and operations team** with a continuously updated membership

LCG-1 will be:

- **utilized by** individual users belonging to groups in the LHC experiments, and possibly other projects for test purposes, who will compose **the continuously updated set of LCG-1 authorized users.**

LCG-1 will be built up during 3Q02 to 2Q03 and has a major milestone of **start of operations in 6/03**.

[DPK: Are there 4 separate LCG-1 Grid's (one per experiment) or just one? Do we have an LCG-1 VO as well as 4 experimental VO's?]

3 Domains of Registration, Authorization and Authentication

Two major domains are identified for processes related to Registration, Authorization and Authentication:

- Processes that will define and maintain **the set of LCG-1 authorized users**
- Processes that will define and maintain the **LCG-1 data processing infrastructure**

The rest of the report is organized around these two domains. The Security topic necessarily interweaves the two domains listed above and is treated in a separate chapter.

4 Processes to define and maintain the set of LCG-1 authorized users [DPK: and infrastructure?], i.e. "Registration"

4.1 Registration of Users

4.1.1 Collaborators and the role of Secretariats of the LHC experiments and the CERN EP Division User's Office.

- The number of people associated with the LHC experiments is reported to be around 6000. These will be referred to as **collaborators**.
- The rotation of individual collaborators is estimated at 10% per year (600 people/year). There is an informally maintained asymmetry which results in relative ease in registering collaborators and relative reluctance in de-registering them.
- The list of collaborators is maintained for each experiment by a dedicated individual Secretariat based at CERN and operated by a combination of CERN staff members and contractors.
- Initial registration of a collaborator results in the manual entry by Secretariat staff of:
  - some basic personal information (FIND OUT WHAT THIS IS)

DRAFT

  - a link to the record describing the Institute to which the collaborator is affiliated
  - other collaboration-specific information
- Update of the registration of a collaborator results in the manual entry by Secretariat staff of changes to the information listed above. The update is normally triggered by the collaborator and there is no formal process for regularly verifying the status of collaborators.
- Creation and maintenance of an Institute record is performed by (FIND OUT HOW THIS IS DONE).
- The criteria used by an LHC experiment Secretariat to register a collaborator are not uniform and are not spelled out in any official document.
- The basic personal data of collaborators is stored in a uniform way in a database at CERN (CHECK TO SEE IF THIS IS TRUE) and is subject to privacy protection constraints defined by CERN (as CERN is an Intergovernmental Organization it is not bound to any single member state legal system and it must work by consensus and pragmatic choices).
- There is a link between processes used by the LHC experiment Secretariats and the CERN Users Office. This link is neither well defined nor documented and results in two different entry points into the collaborator registration process which could lead to inconsistencies.
- In order for a person to be an LHC experiment collaborator and a CERN Registered User they must be hosted in CERN EP Division. (This does not apply to CERN staff and fellows, as they are registered through CERN Personnel Procedures distinct from those for Registered Users; registration of CERN staff and fellows from divisions other than EP is unusual and maintenance of the information in these cases is of very poor quality).

> The **collaborator information** maintained by the Secretariats of the LHC experiments and the CERN EP Division User's Office is the result of a very substantial investment of human resources. Re-use of this information for LCG-1 Registration purposes would be very wise.

> The processes in place for creation and maintenance of collaborator information have not been designed with goals relating to Authorization, Authentication and Security of data-processing systems. Therefore, re-use of the collaborator information must be done with great care and must be complemented with additional processes.

> As result of the work on Registration done for LCG-1, future opportunities may arise for streamlining and reducing costs of the maintenance of collaborator information by experiment Secretariats and the User's Office. The LCG Project Management should consider ways to ensure these opportunities are explored and possibly capitalized.

> The LCG-1 Registration Process should be modelled as the **enhancement of collaborator information** (maintained by the LHC experiment Secretariats and the CERN EP Division User's Office) resulting in a maintained **LCG-1 Registered User Information Base** which can support LCG-1 Authorization, Authentication and Security.

4.1.2 Estimate of the size of the necessary LCG-1 authorized user base

- Historical data suggests that the number of active simultaneous users of physics computing services is about 50% of the number of members of the collaboration.

- Taking into account that the LHC experiments are still in construction phase and that LCG-1 will comprise only a part (albeit substantial in capacity) of their computing activities, the target number of 1000 simultaneously authorized users seems reasonable.
- It is in the interest of the LCG-1 activity to have as many users as possible. Therefore, the LCG-1 Registration Process must complete the initial creation of the Registered User Information Base in a time span of a few months and have a response time of about one working day for its maintenance.
- The rotation rate of LCG-1 will likely be higher than that of experimental collaborators as a whole. It is assumed to be about 20% per year.

> The LCG-1 Registration Process must deliver a **peak capacity of about 25 transactions per working day during 2Q03**. It must maintain a capacity of about **1 transaction per working day during the period of LCG-1 operation**.

4.1.3 Other issues – for User registration

The VO database needs to contain the Grid Authentication credentials of the user, e.g. the X.509 certificate, containing the public key, or at least the full "Distinguished Name" (DN). Today EDG just stores the DN in the VO database and not the public key or certificate.

It is extremely important that the registration process confirms that the correct Authentication credential is stored and not just someone who appears to have the right name (Many of our CA's issue certificates to a much larger community than just particle physics). We could consider, for example, some check of proof of possession of the matching private key via encryption or signature of a registration request.

How are requests from users who cannot visit the Secretariat at CERN in person dealt with? This is likely to require the creation of a devolved (national? Per institute?) set of registration authorities who can confirm the correctness of the registration request, e.g. via digitally signed electronic communication.

Other issues with registration are covered in the chapter on Authorization.

4.2 Registration of infrastructure (systems, resources, services)

There are many issues here:

Not as easy as User registration. The LHC experiments Secretariats do not already do this, nor does the particle physics community have experience with such procedures.

We need a database of and procedure to register and maintain (per VO?)

- Systems
- Resources
- Services
- Applications?

LCG-1 (is this WG2 or us?) (here again, is this our top-level VO?) and the VOs need to define a procedure to register sites, systems, resources, services etc. Each of these principals will require

DRAFT

Authentication credentials, containing a Distinguished Name (an X.509 certificate signed by the local national CA). Given that the resources cannot present themselves in person to the Secretariat at CERN (!) we will need a devolved set of registration authorities. In some cases the RA at a site may be the same person for users and for resources, but this will not always be the case.

The EDG VOMS (Virtual Organisation Membership Service) could be used for BOTH users and infrastructure. Grid proxy certificates, including authorization attributes, could be generated for both.

4.3 Technical issues - registration

The technical issues for registration include:

- Choice of an appropriate database implementation. For users this could be the main CERN user database (Oracle?). What do we do with resources? Register directly in the VOMS database?
- The Grid Authorization system also needs a database (per VO). This is used by the Authorization middleware. In EDG today this is implemented in a VO-LDAP system. The new EDG VOMS is based on MySQL.
- If the master database for users is the CERN Oracle DB, we will then need tools to export (securely) from this into the VO Authorization system (VOMS)
- We need to define access rights to the VO data.
- Is the VO membership data confidential? And if so, how to ensure that it remains private?
- Does the registration DB need to be replicated to all sites (or some sites)? If so, how? And what about issues of confidentiality?
- A **major** issue. Once a user has successfully registered in a VO, how do we ensure that all the necessary account are created at each of the sites? Some sites will be willing to use dynamic or pooled accounts, e.g. atlas037, while others may require more meaningful user names.

5 Authentication

For now, only a very brief list… (needs more work!)

Brief description of technology; PKI, GSI, CA's etc

Establishing trust between LCG-1 and the national CA's. This is based on study of the published policies and operational procedures of the CA's and the identity checks performed by the Registration Authorities (RA's).

Some sites will not allow users to hold long-term private keys on local or network disks. They therefore require that these private keys are either stored in credential repositories (e.g. MyProxy or VSC), or on SmartCards?, or that there is no long-term private key. In the latter case, short-lived (proxy) credentials can be generated online from other authentication systems, e.g. kerberos. The various grid projects are currently discussing all of the issues here in the EDG CA Managers group. It is likely that in the long-term grids will have to support multiple authentication mechanisms. LCG-1 needs to agree whether to support one, two or several authentication mechanisms.

Revocation of credentials

Renewal of authentication credentials for long-lived jobs (EDG uses MyProxy).

DRAFT

6 Authorization

Again, a very brief list which needs much more work…

This is a rather immature, but extremely important, area in Grid security, which will evolve slowly over the lifetime of the LCG project. We will need to consider carefully what can be achieved on the timescale of LCG-1! It is important to realise that Authentication just identifies a user or resource. It is Authorization that actually grants access to the resources, so this is the area needing most careful consideration.

Brief description of the current technologies (grid mapfile, EDG extensions – pooled accounts, VO-LDAP, grid mapfile tools)

Description of soon to be available technologies (Globus CAS, EDG VOMS, LCAS, LCMAPS, SlashGrid, GACL, G-HTTPS etc)

Top-level membership of the VO and the checking that users and resources have the rights to belong to a VO are all covered under the chapter on "Registration".

Here we need to consider more fine-grained Authorization requirements

- Based on Roles within the VO
- Based on membership of Groups (and sub-groups?)
- Based on possession of Capabilities

… and the evaluation of Access Control Lists (ACLs) using the authorization information to grant or deny access to a resource, a service, a file, etc.

We will need to consult with the experiments as to what is required in LCG-1. The SC2 RTAG4 Use Cases does not give enough guidance here.

Sites also need to be in control of local authorisation policy. So, what are the roles of the site and of the VO? Sites will wish to have the ability to ban particular users.

Who is able to define and assign the various roles, groups, capabilities?

How do we revoke membership of a VO and/or possession of the authorization attributes.

7 Security

Operational security issues are handled by GDB/WG4.

The LCG-1 Access policy was addressed by the Computing Resources Review Board (CERN-C-RRB-2002-09 - 9th October 2002).

We need to survey Site security requirements.
Note that the GGF Site-AAA research group and the PPDG-SiteAA project are both active in this area.
http://www.gridforum.org/2_SEC/SAAA.htm
http://www.ppdg.net/pa/ppdg-pa/siteaa/

Our mandate asks us to:

DRAFT

Recommend activities to be developed and deployed by LCG
Recommend policies to be developed and deployed by LCG

There are many other issues related to Grid security: Auditing, accounting, non-repudiation, integrity, confidentiality, firewalls, delegation, etc. How do we deal with these?

7.1 Acceptable Use Policy

It is highly desirable (or do we say absolutely essential?) that each user needs to register just once in a VO. The necessity to register all users at all Grid sites separately, e.g. each time a new site is added to the VO, would be an enormous administrative burden.

To make this possible, we need procedures and tools to propagate this registration, together with the necessary information about the user, out to all sites. We should also aim for a single form to be signed (physically or electronically?) by the user prior to registration. It is normal practice in the existing registration procedures at HEP computing centres to require the user to confirm his knowledge of the existence of various rules, procedures and policies regarding acceptable use of the resources.

Can LCG-1 agree on a single document describing the Acceptable Use Policy of all LCG-1 resources? EDG has such a document (http://marianne.in2p3.fr/datagrid/documentation/EDG-Usage-Rules.html) that also includes implicit reference to the rules of all sites. We should start from here and see what a) sites will agree to and b) what users need to know before using LCG-1.

DRAFT

8 Appendices

*8.1 Appendix A: Reproduction of the instructions given to the Working Group by the GDB (extracted from **put in reference document number**):*

9 GDB Working Groups

9.1 WG3: Security, Registration, Authentication and Authorization

- To report in December 2002

9.1.1 Mandate:

Recommend how common Registration, Authentication, Authorization, Security policies and procedures should be achieved to ensure integrated availability of LCG-assigned infrastructure and resources to valid users. Identify technical issues for Security, Authentication and Authorisation that need to be undertaken by the LCG to meet the deliverables for LCG-1 in July 2003, e.g.

- developing the model for simple user registration, and maintenance of the user database(s)
- deployment of a common model for mapping VO users to local users,
- implementation and operation of the accountability functionality that satisfies the security needs of the sites.

Authentication:

- Survey the existing agreements and trusts for CA/RA between EDG, EDT, DOE, etc. for incompatibilities and determine a procedure for deciding if these are sufficient for the needs of the sites participating in LCG. If so, recommend a process for agreement by the sites. If not, define the procedure to be followed to add the missing sites.
- Identify the constraints from participating sites on authentication and accountability.

Authorization:

- Recommend steps towards acceptable use policies for LCG-1. Ideally the goal should be that a user need only sign once. Define a process for reaching agreement by all participating sites that covers all their requirements for acceptable use of resources.

- Recommend a process and tasks to define and deploy the model for authorization that meets the constraints of the participating sites, and that satisfies the requirements of the experiments. This should include appropriate layers of delegation of authorization.

Security:

- Survey the security requirements of the participating sites and recommend activities and policies to be developed and implemented by the LCG.

9.1.2 Discussion items:

- Is there a need to explicitly define a policy for accepting CA's or is that implied above?
- Evolution of the model – consider how the authentication and authorization mechanisms might change in the transition to OGSA/web services architectures, in particular ws-security. It will be important to ensure that agreements and policies are in place that will permit implementation changes without the need to re-negotiate the agreements.

DRAFT

- Is a model based on VO membership and roles sufficient?
- How should revocation of authorization be done – need both central and local revocation?

DRAFT