

Building Trust on the GRID

Trust Issues Underpinning Scalable Virtual Organisations

Brian Matthews, Juan Bicarregui, Theo Dimitrakos

CLRC Rutherford Appleton Laboratory, Oxfordshire, OX11 0QX, UK

b.m.matthews@rl.ac.uk, j.c.bicarregui@rl.ac.uk, t.dimitrakos@rl.ac.uk

Abstract

GRID computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing including distributed very large data-stores and high-performance networking, and innovative applications such as remote control of experiments and distributed group working in virtual organisations. Given the extreme nature of this computing environment, we consider whether existing leading-edge solutions which focus on the security and trust management of open distributed systems extend to the GRID. In this position paper we outline the envisaged GRID architecture and highlight the unique security and trust management challenges which are encountered in such a setting.

Keywords: GRID computing, virtual enterprises, trust requirements, policy specification and deployment, semantic web

1. INTRODUCTION

The concept of the Grid has emerged as a new approach to computing infrastructure within the last five years [7]. Based on the existing Internet, the Grid seeks to extend the scope of distributed computing to encompass large-scale resource sharing including distributed very large data-stores and high-performance networking, and shared use of computational resources, be they supercomputers or large networks of workstations. Currently the applications which are developing this infrastructure are large scale scientific collaborations, such as NASA's Information Power Grid [13], and the European DataGrid project [4] led by the Particle Physics community. Such collaborations have a clear need for the collaborative use of resources (both data and computational) and established communities which can pool their resources for common goals, and thus form an ideal test bed to develop such technology. Tools are appearing to support the Grid concept, notably Globus [10], Condor [2] and Legion [15].

The Grid concept has been generalised to cover any *virtual organisation*, defined as a dynamic collection of individuals and institutions which are required to share resources to achieve certain goals [8]. Thus the Grid will have applications in commerce and industry, supporting distributed collaborative design and engineering, or supporting distributed supply chains. In these contexts, the emerging Web Services technologies are likely to play a key role. In this setting, a key feature in the emerging Grid architecture will be enabling the establishment of trust. Existing approaches to security within distributed systems are stretched by the extreme conditions imposed by the Grid, and significant effort has been undertaken in, for example, Globus to provide support for secure use of resources. However, trust needs to be established at all levels: :

- *Authentication*: the establishment of identity of the user.
- *Policy based management*: the provision and deployment of rules and procedures for governing the choice in the behaviour of a managed system towards users who have been successfully authenticated.

These include, positive/negative authorisation policies and delegation policies to manage access control; refrain policies to define the actions that subjects refrain from performing on target objects even though they may be authorised to do so; and obligation policies to define the actions that must be performed by managers within the system when certain events occur - therefore responding to changing circumstances.

- *Business Rules*: the business and legislative framework determining the nature of policies

In this paper, we discuss the trust requirements imposed by the Grid, and propose an approach based on the publication of trust policies and the automatic negotiation of access rights. A similar discussion on requirements is given in [9] which defines the policy which underlies the Globus security model, which uses X.509 certificates

as its authentication mechanism. However, that paper does not consider the problems of access control and trust management. It also does not consider how to establish trust between entities which do not have previous knowledge of each other in the absence of a fixed global hierarchy of certification authorities, where the entities may need to *discover* or *dynamically form* certification chains in order to enable trust establishment.

2. TRUST REQUIREMENTS ON THE GRID

Consider the following scenario for the use of the Grid.

An engineer within organisation A wants to perform an analysis on a material. By accessing a data portal at site B, she discovers a suitable data set held by a data archive C. The analytical tools are provided at university D within her Virtual Organisation. She initiates the analysis by passing the reference to the data set from B to D, which is then accessed by the analysis tools. D then determines that it does not have enough computational resource available, and determines that a computer is available at different institution E and delegates part of the job there. Finally, D completes the job and return the results to A. D also caches the results of the analysis locally and registers the fact that the precomputed results are available with the portal B and the data provider C.

However, the analysis takes several hours, so the engineer has established a user proxy agent to represent her, collect the results, make payments as appropriate and close down the collaboration.

This scenario highlights several features of the Grid which are relevant to trust.

- Using the Grid requires the collaboration of resources which are controlled by different institutions. Each institution will have their own policies on access control and conditions of use.
- The allocation of resource is dynamic; the computational resource at D may not know until part way through the job that additional resource at E is required.
- The user may have different identities in different domains. For example the engineer may have different Users Ids in her own institution, with the data portal, and at the university.
- Resources need to establish trust between themselves (for example, the archive C and the data portal B, the archive C and the analysis tools D, and the tools D and the compute resource E) on a peer-to-peer basis independently of the trust in the original user.
- Resources may be called upon to participate in the task without previous knowledge of the other participants. For example, E may not know the identity of the engineer, or even the institution she comes from. Trust thus needs to be established on the fly so a mechanism needs to negotiate conditions of use, through the delegation of trust from one party to another.
- Different trust conditions may be applied for different parts of the resource, including restrictions on data. For example, the data on the material lodged at archive C may have particular access conditions set by its originators (who may well be elsewhere than at C). A negotiation with respect to the use of the results needs to take place between the archive and the engineer before the data can be sent to the analysis tools.
- Users and resources may delegate their identity to other parties – for example, the user at A creates a user proxy to act her role as the user in the process. The delegation of trust should follow the delegation of identity.
- Users and resources may be located in different countries under different jurisdictions and thus with as a consequence subject to different legal and business requirements.
- Resource usage tracking and charging may be involved. For example, data portal B and computational resource E may be provided as commercial services. They need to identify the correct party to bill and establish trust in credentials (including possibly contacting third parties such as credit agencies) to be satisfied that bills will be honoured. These credentials need to be propagated by other agencies, for example, the analysis tools provider D, who may not be involved in the monetary transaction.

3. TRUST AS INFORMATION

Another more abstract view on the Grid is that it requires three layers to work effectively [11]:

- firstly a data and computational layer, which provides the raw data sources, processors, and connecting protocols;
- secondly an information layer over this providing information and location of the underlying resources; and
- thirdly a knowledge layer to allow reasoning over that information to give intelligent assistance to the users.

The traditional Grid infrastructure, such as the Grid Security Infrastructure (GSI) from Globus [1], has depended on interfaces at the protocol level to provide the security infrastructure. However, this approach has concentrated on authentication and not provided a sufficient infrastructure for the rest of the trust hierarchy exemplified above, especially with respect to access control (i.e. authorisation, filtering, delegation, etc.) and the specification and deployment of policies. Further, this mechanism does not appear to have considered the case where the collaborating resources have no prior knowledge of each other (or their certifying authorities).

We propose to supplement this infrastructure by raising the level of the trust analysis into the information and knowledge grids of the layered Grid architecture. This would build upon the established literature in trust analysis, such as [12],[5] which provides a framework for analysing how trust should be transmitted between agents in distributed systems, especially dealing with how to propagate trust between agents without prior knowledge of each other.

The basis of this infrastructure would be the explicit declaration and publication of trust policies by participating resources on the Grid using a policy specification language, such as Ponder [3]. Agents wishing to utilise resources would then be able to present their credentials, policies and requirements to the participating resources and an automated process would verify the credentials, possibly referring to trusted third parties, to establish identity, deduce authorisation based upon supplier and consumer policies and to authorise (or not) access under the specified the conditions of use. The nature of this task, where statements made of the properties of systems (i.e. policies and requirements from agents) are transmitted across the network, and automatic reasoning is used to determine action from those statements, is in line with proposals under the *Semantic Web* activity of the World-Wide Web Consortium [16]. Consequently, we propose stating the policies and trust statements in using the Resource Description Framework [14], and use the emerging tools of the Semantic Web to support the reasoning. An initial outline of this approach is given in [6].

REFERENCES

- [1] R. Butler, D. Engert, I. Foster, C. Kesselman, S. Tuecke, J. Volmer, V. Welch. *A National-Scale Authentication Infrastructure*. IEEE Computer, 33(12):60-66, 2000.
- [2] The Condor Project home page, University of Wisconsin at Madison <http://www.cs.wisc.edu/condor/>
- [3] N. Damianou, N. Dulay, E. Lupu, M Sloman, *The Ponder Specification Language*. Workshop on Policies for Distributed Systems and Networks (Policy2001), HP Labs Bristol, 29-31 Jan 2001.
- [4] The European DataGrid Project home page <http://www.eu-datagrid.org/>
- [5] Theo Dimitrakos, *System Models, e-Risks and e-Trust. Towards bridging the gap? Towards the E-Society: E-Business, E-Commerce, and E-Government*. 1st IFIP Conference on e-Commerce, e-Business, e-Government, Kluwer Academic Publishers, Zurich, Switzerland, (to appear October 2001)
- [6] Theo Dimitrakos, Brian Matthews, Juan Bicarregui, *Towards encoding security and trust management policies in RDF*. 2001 (submitted).
- [7] I. Foster and C. Kesselman (eds), *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Pub.,1998
- [8] I. Foster, C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organization*. (To appear in Intl. J. Supercomputer Applications, 2001).
- [9] I. Foster, C. Kesselman, G. Tsudik, S. Tuecke, *A Security Architecture for Computational Grids*. Proc. 5th ACM Conference on Computer and Communications Security Conference, pg. 83-92, 1998.
- [10] The Globus Project home page <http://www.globus.org>
- [11] Keith Jeffery, *Knowledge, Information and Data*, CLRC Information Technology Department, September 1999 <http://www.escience.clrc.ac.uk/Publications/9/KnowledgeInformationData.doc>
- [12] A. Jøsang. *The right type of trust for distributed systems*. Proc. New Security Paradigms Workshop, ACM, 1996.
- [13] NASA's Information Power Grid (IPG) home page <http://www.ipg.nasa.gov/>
- [14] Ora Lassila , & Ralph R. Swick (eds), *Resource Description Framework (RDF) Model and Syntax Specification*.W3C Recommendation 22 February 1999
- [15] The Legion Project home page, University of Virginia <http://www.cs.virginia.edu/~legion/>
- [16] *Semantic Web Activity Statement*, World Wide Web Consortium, 2001 <http://www.w3.org/2001/sw/Activity>