Using Trusted Computing in Commercial Grids

Po-Wah Yau[1] and Allan Tomlinson[1]

[1] Royal Holloway University of London, Information Security Group
{p.yau, allan.tomlinson}@rhul.ac.uk

Abstract. Trust flows in two directions in a Grid environment. The first is from the Grid user to the Grid resource, that is, the Grid trusts that the user will protect confidential information. The second is from the resource to the user, that is, the Grid will protect the user's Grid job and associated data. This paper comments on how Trusted Computing technology can be used to establish trust in both directions, in three types of Grids that may be interest of to commercial organisations.

1 Introduction

Grid computing is a model of distributed computing to enable the pooling of heterogenous resources, for example, CPU cycles, application software, data and its storage (Foster and Kesselman 2004). These resources are accessed by invoking a Grid service that is defined by the resource host. In general, a two-layer administrative domain hierarchy exists. The top layer is an umbrella administrative domain called a Virtual Organisation (VO). The bottom layer consists of separate autonomous domains that host Grid-enabled resources. This separation of domains can be at many levels, for example, groups in a department, departments in an organisation, organisations in a collaborative project etc. Authorised members of the VO will be able to access VO member resources, even though they reside in different domains.

Grid computing is relatively mature in academia (Doyle 2005; Foster and Kesselman 2004) and is attracting interest from industry. However, there are open security issues that are inhibiting the widespread adoption of Grid computing. Specific applications of Trusted Computing (TC), a recent and industry-backed technology (Mitchell 2005), have been proposed to solve certain Grid security problems (Mao, Martin, Jin, and Zhang 2006). In this paper, we concentrate on trust establishment and how to achieve high levels of assurance that may be required in a commercial setting.

This paper is organised as follows. In Section 2 we briefly describe the TC mechanisms that will be useful in a Grid environment. Section 3 describes how current Grid architectures incorporate trust. The rest of this paper discusses trust establishment and the use of TC in three classes of Grid networks: enterprise Grid in Section 4, fixed VO Grid in Section 5, and dynamic VO Grid in Section 6. Section 7 contains some conclusions for the paper.

2 Trusted Computing

Trusted Computing (TC) is an industry-led initiative to provide a variety of security primitives that make use of a hardware root-of-trust called the Trusted Platform Module (TPM) (Trusted Computing Group 2006), which is a tamper-resistant module which is secured bound to the host platform. The features of TC include protected storage within the TPM (mainly for cryptographic keys), remote platform attestation and data sealing; it is the latter two that we are interested in.

The TPM contains a set of Platform Configuration Registers (PCRs) to securely store an audit log of the host platform's boot process. The log consists of a series of platform 'integrity measurements', essentially the hash digest of some software component on the platform. A trusted platform with a TPM can undergo an authenticated boot process – at each stage of the boot process a measurement is taken of the components required for the subsequent stage, before control is passed to those measured components (Grawrock 2006). The platform can attest to its configuration by presenting the platform measurements, digitally signed by the TPM, to a requesting principal. The 'trust' in TC is the assertion that a platform is in a specific configuration. Whether this configuration is one that is 'trustworthy' for Grid computing is an issue we address later. To avoid confusion, we will state that a platform has the 'required configuration' instead of 'trusted' as used in TC terminology.

Several issues with TC platform attestations have been highlighted, in particular, the management of integrity measurements for the seemingly infinite number of configurations that can occur because of software updates and security patches (Cooper and Martin 2006). This is an issue that we will be commenting on later.

A related technology for creating trusted platforms is virtualisation – providing the ability to run multiple 'virtual machines' on one physical platform. This offers several security properties, e.g. process isolation, which are seen as complementary to TC. An example of the union of the two technologies is Intel's LaGrande virtualisation technology (Grawrock 2006), providing hardware support to create secure (and measurable) compartments for virtual machines to operate in.

3 Grid Computing and Trust

The Grid Security Infrastructure (GSI) (Foster and Kesselman 2004) is the de facto architecture that has been adopted by many Grid implementations. Trust is built upon the asymmetric cryptography based SSL/TLS mutual authentication protocols (Dierks and Allen 1999). These rely on a public-key infrastructure (PKI) of which the trust relationships are well understood (Thompson and Olson et al. 2002) – trust is placed in Certification Authorities (CA), their user registration procedures and their ability to protect their private signature keys. The GSI also includes a delegation capability to extend trust (see Section 6). Following the ethos of Grid computing, other authentication protocols can be used, and the WS-Trust specification (Oasis Standards 2007) contains details of specifying how different credentials can be 'trusted'. The common denominator is that trust is built upon successful entity authentication, i.e. corroborating the identity used.

The issues with relying on identity alone to establish trust in a Grid environment are well documented (Cooper and Martin 2006). Alternative proposals include using attribute certificates (Chadwick, Otenko and Ball 2003) that focus on specifying attributes on which to base access control decisions. We will later see how Trusted Computing can provide another alternative means of establishing trust.

In the hosting environment of the resource/service provider, an authenticated user identity will be mapped to a local identity and user account, and given limited privileges according to either VO or local policy (or both). The local account is then subject to traditional security protections that protect the host environment from malicious processes. In this case, the service provider 'trusts' the user, but only up to a certain degree. Conversely, the user has no option but to completely trust the service provider with any data, software, scripts sent, and the resulting output. This trust asymmetry may not be acceptable within a commercial setting, and is a concern that has been previously expressed (Cooper and Martin 2006; Mao, Martin et al. 2006).

4 Enterprise Grids

Early commercial adoption of Grid has been using enterprise Grid technology, where the VO consists solely of Grid nodes that reside within a single enterprise (Enterprise Grid Alliance 2004), i.e. a single-domain Grid. Such nodes are likely to have strong trust relationships enforced by internal security policies and services, so the trust asymmetry issue highlighted in Section 3 becomes less relevant. Instead, focus is on the Grid access device (user interface) that is used to obtain commercially sensitive information from the enterprise Grid. Information from an enterprise Grid will have potentially high intellectual property value and require additional protection, especially from insider attacks. Private data/results can be 'sealed' using TC, so that only devices with the required configuration can have access. This is of particular importance if access to the single-enterprise Grid is given to external principals, e.g. third-party suppliers in a coordinated development project, or to remote user devices with limited physical security.

Deciding upon the required configuration in an enterprise Grid is a manageable problem. The enterprise is likely to manage software updates for enterprise devices. High risk devices, where users have the autonomy to perform software updates or installations, would probably not be considered for enterprise Grid access.

However, the management of third-party access devices would be outside the enterprise domain. The relationship between the enterprise and third-parties is likely to be enforced with commercial contracts and non-disclosure agreements. One of the conditions of contract could be that access to the enterprise Grid is only possible using TPM-enabled devices with digital rights management (DRM) software installed. The protocols proposed by (Gallery and Tomlinson 2005) and the Trusted Network Connect (TNC) architecture (Trusted Computing Group 2005) provide a means to measure that this configuration is met before allowing a network connection and access to the enterprise Grid.

5 Fixed VO Grids

In fixed VO Grids, Grid nodes reside in distinct administrative domains, have predetermined trust relationships and the VO membership is relatively static.
Such a Grid can be used for multiple enterprise collaboration, each enterprise contributing, through their single-enterprise Grids, propriety, and often specialist, hardware and software (Foster and Kesselman 2004). In a fixed VO Grid, each enterprise Grid represents one node.

In this type of Grid, the protection of enterprise data has to extend outside the enterprise. Jobs and service requests are likely to be submitted directly to the service provider in question, i.e. the node which is hosting the service. Using TC mechanisms, the required configuration of service provider's Grid gateway can be determined before submitting a job to it. If the gateway receives an incoming job that satisfies policy rules and the submitting user has been authenticated etc., then the service provider will accept the job.

However, the gateway will probably not be the platform that the job will be executed on. Consider a Grid service offering computational time. The gateway will pass the authenticated job to a local job manager (e.g. Condor or Portable Batch System), which could also reside on the gateway host. The local job manager will at some point choose an internal worker node to perform the job. This leads to two main issues with securing enterprise data in the host environment.

Firstly, the owner can ‘seal’ the job and relevant data, so access is only possible if the internal worker nodes meets the required configuration. While it is possible for a worker node to match the requirements of some job owners in order to ‘unseal’ the job for execution, it would be extremely difficult to do so for all potential VO members.

Secondly, the worker node could be compromised as the job is being executed. A Grid node could be a member of several VOs, potentially running jobs/providing services concurrently for rival organisations. Jobs running on the same worker node would have access to each other's memory space – a vulnerability that could be exploited. While the gateway can be measured to ensure that it is in a state to ensure that jobs are submitted to different worker nodes, local users using the enterprise Grid could potentially introduce or exploit vulnerabilities that could comprise the job (Cooper and Martin 2006).

As discussed in Section 2, platform virtualisation can be used to create protected environments. This would address the second issue highlighted above, by segregating jobs on the same worker node. Virtualisation and TC can also be used to address the first issue of determining required configuration. A user could use a Grid service to download a known secure virtual machine image, and ‘seal’ data to that image. The virtual machine could include DRM software to provide enhanced protection. The user can delegate rights to the service provider to use the same Grid service to download the same virtual machine image, in order to ‘unseal’ the user's job for execution. This would also require key migration, from the user's TPM to the worker node's TPM, using a system such as Daonity (Mao, Yan and Chen 2006). Delegation is an issue that we will discuss in the next section.

6 Dynamic VO Grids

There are two issues with fixed VO grids which become even more complicated in dynamic VO Grids: delegation and the architecture needed to support authentication across domains that use different authentication protocols. The vision is that VOs dynamically form to meet the services requested by the Grid user. For example, an enterprise Grid could use a Grid resource broker service to dynamically outsource to other appropriate Grid service providers when additional resources are needed (e.g. for data federation, extra CPU cycles and storage). In order to manage this outsourcing, credentials are delegated using a proxy certificate (Tuecke et al. 2004). This is an X.509 public-key certificate that is generated by a user to extend a trust chain to the entity that the user is delegating rights to. That entity can then make service requests on behalf of the user.

This delegation model could be cumbersome for a dynamic VO Grid as a trust relationship between the user and a service provider may not exist. The service provider may have to 'pull' this relationship using either identity federation or credential translation services, and then make a policy decision to accept the service request or not. Identity federation introduces its own problems, e.g. the need to protect user personal identifiable information. Establishing trust in this way may not provide enough assurances in a commercial setting (see Section 3).

TC can be used to make the delegation model more efficient, by allowing a node to 'push' requirements to delegating Grid services. These requirements could be made in the form of dynamically negotiated Service Level Agreements (SLAs) (Foster and Kesselman 2004). For example, a Grid service may mandate that a requesting entity installs DRM software as part of the SLAs it negotiates. Therefore, the relatively heavyweight mechanisms required for identity federation are not needed because the Grid service has assurance that the data it provides will not be forwarded.

6 Conclusions

A symmetric trust relationship should be established when using Grid computing. In one direction, a Grid service must be able to establish trust in the user and the access device being used. There are many issues with the current methods of establishing trust based on identity. In the opposite direction, the user must be able to trust that the Grid service will not compromise the user's job data. Trusted Computing provides a set of security primitives to achieve this.

Sensitive commercial data can be sealed to a certain platform configuration. Trust is then built on an entity's ability to attest that its platform is in the required configuration to unseal data. In an enterprise Grid, this means determining that an access device, which could have Digital Rights Management software installed, can attest to a state that indicates that it has not been compromised. In fixed VO Grids, a Grid user can seal data on a virtual machine, and delegate rights to a Grid service to retrieve the same virtual machine image (from another Grid service) to unseal data. The delegation mechanism in dynamic VO grids can use platform attestation in addition, or instead of, relying on identity-based credentials. Grid technology will be-

come more feasible for commercial use by incorporating Trusted Computing technology, which allows trust to be established in a manner that matches commercial requirements.

Acknowledgements

This work is being funded by the Engineering and Physical Sciences Research Council (EPSRC) UK e-Science programme of research (EP/D053269). For more details of this project please refer to www.distributedtrust.org.

References

D. Chadwick, A. Otenko, and E. Ball. Role-based access control with x.509 attribute certificates. *IEEE Internet Computing*, 7(2):62–69, March 2003.

A. Cooper and A. Martin. Towards a secure, tamper-proof grid platform. In *Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid, Singapore, May 2006*, pages 373–380. IEEE Press, May 2006.

T. Dierks and C. Allen. The TLS Protocol Version 1.0. RFC 2246, Internet and Engineering Task Force, January 1999.

T. Doyle. Meeting the particle physics computing challenge. *PSCA International Public Service Review: Trade and Industry*, (8):88–89, Autumn 2005.

Enterprise Grid Alliance. Accelerating the adoption of Grid solutions in the enterprise. White paper, Enterprise Grid Alliance, Dec 2004.

I. Foster and C. Kesselman. *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers, San Francisco, 2nd edition, 2004.

E. M. Gallery and A. Tomlinson. *Secure delivery of conditional access applications to mobile receivers*, volume 6 of *IEE Professional Applications of Computing*, chapter 7, pages 195–237. IEE Press, London, 1st edition, 2005.

D. Grawrock. *The Intel safer computing initiative: Building blocks for Trusted Computing*. Intel Press, 2006.

W. Mao, A. Martin, H. Jin, and H. Zhang. Innovations for grid security from trusted computing – protocol solutions to sharing of security resource. In *Proceedings of the 14th International Workshop on Security Protocols, Cambridge, UK, March 2006, to appear*. Springer-Verlag LNCS, March 2006.

W. Mao, F. Yan, and C. Chen. Daonity – Grid security with behaviour conformity from trusted computing. In *Proceedings of the first ACM workshop on Scalable Trusted Computing, Alexandria, Virginia, US*, pages 43–46. ACM Press, Nov 2006.

C. J. Mitchell. *Trusted Computing*, volume 6 of *IEE Professional Applications of Computing*. IEE Press, London, 1st edition, 2005.

Oasis Standards. *WS-Trust 1.3*, March 2007.

Thompson and Olson et al. *CA-based Trust Model for Grid Authentication and Identity Delegation*. Open Grid Forum, Oct 2002.

Trusted Computing Group. *TCG Trusted Network Connect TNC Architecture for Interoperability Specification Version 1.1 Revision 2*, May 2005.

Trusted Computing Group. *TPM Main Part 1 Design Principles Specification Version 1.2 Revision 94*, March 2006.

S. Tuecke et al. Internet X.509 public key infrastructure (PKI) proxy certificate profile. RFC 3820, Internet and Engineering Task Force, June 2004.