Innovations for the Grid Security from the Trusted Computing

Wenbo Mao

Hewlett-Packard Laboratories
Filton Road, Stoke Gifford, Bristol BS34 8QZ, UK
wenbo.mao@hp.com

February 22, 2005

Abstract

The Trusted Computing (TC) initiative developed by Trusted Computing Group (TCG) takes a distributed-system-wide approach to the provisions of integrity protection of resources. The TC's notion of trust and security can be described as conformed system behaviours of a platform environment such that the conformation can be attested to a remote challenger. We consider that such a notion of integrity protection of resources naturally suits the security requirements for Grid computing or science collaborations. We identify and discuss in this paper a number of innovations that the TC technology could offer for improving the Grid security.

Keywords Trusted Computing (TC), Trusted Computing Group (TCG), Grid Computing, Grid Security.

1 Introduction

A computational Grid [2, 5, 7] is a distributed computing system comprising a number—possibly large—of physically apart users who may be working on a common task and have similar resource requirements. A Grid system of such collaborators and resource providers is regarded to form a virtual organisation, VO, owing to the fact that these collaborators usually maintain dynamic relations. In the most general setting, a VO of users and resource providers is geographically distributed and in different trust and

management domains. These domains can span governmental, industrial and academic organisations. This implies, even demands, that strong security mechanisms be in place such that the Grid services can be used in a secure and accountable manner.

We assert that the Grid security has, at least, the following two characteristics:

System behaviour conformation Because typical Grid resources—infrastructure, applications, instrument or data— have critically high importance and value, a Grid security strategy should be based mainly on attack prevention. While entity authentication is an important means for controlling access to resources and can also achieve attacker identification after an attack, it does not provide an effective means to attack prevention. Attack prevention is better done with a behaviour conformation mechanism: an entity and its supporting computing system is attested that they have a restricted (and desirable) behaviour which cannot easily lead to any serious damage.

Group-oriented security Resource sharing in a Grid VO is, by definition, a group-oriented activity. Consider the following two simple scenarios: research data are shared by a group of scientists, a large scientific instrument must be operated by a group of users at the same time. Therefore, we argue that a desirable attribute for a Grid security solution is to support group-oriented activities, e.g., in order to secure a conference discussion among a group of entities they need to be served with a shared conference key.

Over past few years effort for increasing computer security has been made by the computing industry. Among the many efforts, we are specifically focusing on the Trusted Computing (TC) initiative by the industrial standard body, the Trusted Computing Group [16]. The purpose of the TCG is to develop, define and promote open, vendor-neutral specifications for trusted computing. It begins with a simple idea: using a tamper-resistant hardware module to enable and manage data and digital identities more securely, protecting them from external software attack and physical theft. The TCG work has so far been developed with sufficient innovations to achieve their goal. These include hardware building block and software interface specifications across multiple platforms and operating environments. The TCG's open specifications (versions 1.1b and 1.2, available at the "Downloads area" of [16]) not only define reasonable notions of trust and security, but also provide concrete mechanisms to achieve protections by means of policy and trusted environment conformance.

We observe that the TCG mechanisms for policy and trusted environment conformation can provide a needed role in Grid security. This is particularly suitable for our two Grid security characteristics. In this paper we propose an innovative approach to Grid security from Trusted Computing effort.

The remainder of this paper is organised as follows. In §2 we list the official Grid security requirements. In §3 we overview the current Grid security solutions and identify their inadequacy with respect to our two characteristics for Grid security. In §4 we overview the Trusted Computing technology. In §5 we consider Trusted Computing technology as the complementary solution to the identified problems in the Grid security. Finally in §6 we provide discussions on issues of the TC implementation and deployment.

2 Grid Security Requirements

The US Department of Energy (DoE) Office of Advanced Scientific Computing Research published a report which provides a good summary of the requirements for Grid security [2]. The Grid requires a security infrastructure with the following properties:

I) Ease of use by users.

II) Conformation with the VO security needs while at the same time working well with site policies of each resource provider site.

III) Provisions for appropriate authentication and encryption of all interactions.

Let's name this set of Grid security requirements "DoE Grid Security Requirements." We hold the view that DoE Grid Security Requirements II and III are compatible to our two characteristics for Grid security. More clarifications will be provided in the remainder of this paper.

3 Current Grid Security Solutions

The Grid Security Infrastructure (GSI) [6] and MyProxy [11] are two primarily important works for the Grid security.

The GSI, which is the security kernel of the Globus Toolkit [9], provides a set of security protocols for achieving mutual entity authentication between a user (actually a user's proxy which is a client-side computing platform) and resource providers. Entity authentication in the GSI protocols involves straightforward applications of the standard SSL Authentication Protocol (SAP) suite [8]. These standard applications can be considered as a "plug-and-play security solution." They achieve quick deployment and ease of use. As a result, the Grid security protocols in the GSI are two-party mutual authentication techniques. Each party has a public-key based cryptographic credential in the formulation of a certificate under the standard public-key authentication infrastructure PKI X.509 [10]. The use of the standard PKI in Grid security is not only suitable for the

VO environment, but also has an important advantage: single sign-on (SSO). The latter means that each user only needs to maintain one cryptographic credential. As always, any security solution must not demand the user to invoke sophisticated operations or tools.

Using PKI requires each user to have obtained a private key as their cryptographic credential. This can be a demanding requirement for many users without a secure computing platform in their locality. MyProxy provides a lightweight solution. It uses an online credential repository which can deliver temporary Grid credentials to the end user. This is achieved via simple user authentication mechanisms such as password. This can be enhanced via a one-time password such as through a SecureID card.

We believe that the combination of the GSI and MyProxy provides a good solution to the DoE Grid Security Requirement I. The two-party authentication protocols of the GSI, however, do not provide an adequate solution to group oriented Grid security applications. For example, consider the DoE Grid Security Requirement III, the GSI cannot easily achieve a common key for a VO wide encrypted communication.

There are other notable Grid security add-ons. These include Akenti [15], Community Authorization Service [12], and Web Services Security [1]. The former two add-ons are focused on important features such as authorisation and accounting. The last, WS-S takes a message level security approach by performing encryption at the Web Services layer, such as the XML messages. These solutions also make use of X.509 PKI. Observe that the services these latter solutions provide are orthogonal to DoE Grid Security Requirements.

Given the above, we can call the current Grid security solutions "plug-and-play PKI" for a conventional client-server environment. It is obvious that two-party protocols based Grid security solutions neither directly nor effectively support a group-oriented security. Additionally, they do not have a inherent means for realising behaviour control for a remote user and its client system environment. For example, consider that WS-Security can achieve message encryption between a resource provider and a user. However, there is no way for a stakeholder in the resource provider to know whether or not the remote client environment is compromised (perhaps by a malicious code) even though it knows that such a compromise is equivalent to the nullification of the channel encryption service.

4 Trusted Computing

In 1999 five companies—Compaq, HP, IBM, Intel and Microsoft—founded the Trusted Computing Platform Alliance (TCPA). In 2003, the TCPA achieved a membership of 190+ companies. TCPA was then succeeded by the Trusted Computing Group (TCG)

[16]. The TCG takes a distributed, system-wide approach to the establishment of trust and security. It defines a concrete concept of Trusted Computing (TC). Consider TC as the desired and conformable system behaviour which is not only established and maintained in a platform environment, but can also be attested to a remote challenger.

The following four notions are at the core of the TC technology:

Trusted Platform Module (TPM): This is a tamper-resistant hardware module for conformed operation and secure storage. It is designed to perform computations which cannot be subverted by the platform owner, including the system administrator. These computations include some public key cryptographic operations (decryption and digital signature generation using a private key in the TPM), platform system status measurement, and secure storage. Each platform has a TPM.

Core Root of Trust for Measurement (CRTM): In the booting time of a platform, the TPM measures the system's data integrity status. The measurement starts from the integrity of BIOS, then that of OS and finally to applications. With CRTM, it is possible to achieve the establishment of a desired platform environment by only loading well behaved systems. This is a strong requirement which is called "secure boot." The TCG also permits a slightly weaker measured boot which is called "authenticated boot." In the latter the TPM will permit loading of codes which do not pass the measurement but will only securely record the status of those which have passed the measurement for attestation purpose (more on this in a moment).

Root of Trust for Storage: The measured integrity of an executable is represented by a cryptographic checksum of the executable. This is then securely stored in a TPM. The TPM component called Platform Configuration Register (PCR) holds this data in an accumulative formulation. TPM has a plural number of PCRs; each of them can be used to accumulate system integrity data for one category of system executables, e.g., one PCR for OS's (a platform can run many copies of OS's, see §5.4) and one PCR for a family of specific applications. The stored platform environment status is maintained until rebooting of the system.

Remote Platform Attestation: That a platform's system has desired and conformed behaviour can be examined by a legitimate remote entity via cryptographic challenge-response mechanisms. Remote platform attestation is the most significant and the most innovation element in the TC technology. With this capability, a remote stakeholder can be assured, with confidence, of the desired and conformed behaviour of a platform.

We notice that with a platform having the above behaviour, the TC technology has met resistances by being interpreted as providing for monopoly control over the use of software. The TCG considers this a misinterpretation because a TCG platform should be able to execute any software in the "authenticated boot" condition (see CRTM above). At any rate, we avoid this controversial issue here. In the attempted TC application to Grid security there should be much less disagreement since Grid computing either requires behavioural compliances from an individual user as conditions for using remote resource, or implies federation and cooperation among a group of users.

5 Trusted Computing for the Grid Security

We believe that TC technology can offer good solutions to Grid security problems for which current Grid security solutions do not play a role. Specifically, we argue that TC technology addresses particularly well the DoE Grid Security Requirements II and III.

5.1 Secure Storage of Cryptographic Credential

Unattended user authentication is an important feature in the Grid. This means that a user working in a VO is mainly doing so via its proxy. Work within a VO may involve dynamic sessions of resource allocation and hence require user entity authentication without having the user present. In the GSI, and in MyProxy, this is achieved by having a user client platform be issued a proxy certificate. The cryptographic credential of this certificate (i.e., the private key matching the public key in this certificate) is simply stored in the file system of the platform protected under the access control of the operating system. In this way, the client platform does not need to prompt the user for cryptographic operations. The obvious danger of leaving a private key in the file space is mitigated by stipulating a short lifetime for the proxy certificate. The default lifetime of a proxy certificate in the GSI is 12 hours. Upon expiration, a new proxy certificate must be re-issued. We feel this is an unacceptable security exposure.

With a TCP containing a tamper-resistant TPM, it is natural to store a user's cryptographic credentials in the TPM. In TC, each user of a platform can generate many copies of private keys with their matching public keys being certified in the standard X.509 PKI. A TPM can be configured into a "non-migration" mode which will never reveal any private key (up to the tamper-resistance level a TPM is design for). It can also be configured into a "migration" which can export key material upon the owner's consent. Thus, even if a platform is under the control of an attacker, the attacker, though in this situation may be able to misuse the user's credential (still in a conformable manner), cannot retrieve any information stored in the TPM. Thus, in a TC enhanced

Grid security setting, there is no need to use short-lived proxy certificates.

5.2 Distributed Firewall for a VO

In a conventional organisation a firewall plays an effective role in protecting the information assets of the organisation. A conventional firewall relies on the notions of restricted topology and controlled entry points to function. More precisely, a firewall relies on the assumption that every entity on one side of the entry point, the firewall, is to be trusted, and any entity on the other side is, at least potentially, an enemy. Because most attacks are achieved via malicious connections which can be shielded by a firewall, firewalls are a powerful protective mechanism.

A Grid VO is typically composed of multiple physically distinct entities which are in different organisations who usually do not trust each other. There is no longer a notion of a restricted network topology. The current Grid security solution does not utilise the notion of firewall based protection. A user (its proxy) enters a VO without bringing in its own computational resource. Such a VO is in a primitive stage: a user only uses resource "out there," rather than also contributes its own resource as well. In fact, a Grid is valuable precisely because every participant becomes a taker as well as a giver. Imagine the augmented value of a medical research collaboration which combines small databases of some rare illness' information scattered in various hospitals into global database available for access and search.

Bellovin proposed a notion of distributed firewall [3] which exactly suits the situation of a Grid VO. In a distributed firewall, a packet is deemed to be accepted or rejected according to whether it has an acceptable digital signature. The packet's acceptance not only depends on the validity of a signature, but also on the rights granted to the certificate.

At first glance it seems that the current Grid security solutions can achieve a distributed firewall for a VO since these solutions also use public key cryptography and PKI authentication framework which enable the use of digital signatures. The main problem is that the short lifetime of a proxy certificate of any participant makes the packet-level signature verification a performance burden. We repeat that the acceptance of a signature in a distributed firewall application is not only on the validity of the signature in the conventional sense, it should also be judged on the firewall policy granted to a certificate. The short-lived certificate used in the current Grid solutions are mainly limited to the so-called "identity certificate," they are not suitable for distributed firewall use which needs refined policies. We can call a certificate for a distributed firewall use "property certificate."

With TC technology making multiple long-term (property) certificates available to each user of a platform, a Grid VO can readily implement a distributed firewall tech-

nique.

5.3 Attestation of Policy Conformation in a Remote System

A Grid stakeholder has legitimate reasons to worry whether a participating subsystem in a VO conforms to the VO's security policy. For example, consider the need for a remote platform, which is sending in a Grid ftp query for some sensitive information, does indeed run the correct version of the Grid ftp which will flush the downloaded data from the local memory without saving a local copy in the file system after using the data. Likewise, a participating (a giver) client may also have similar concern with respect to a VO.

TC's notion of remote platform attestation is a ready solution for this sort of service. Each user has "Attestation ID Keys" (AIKs) which can sign PCRs in a TPM. For details, see "Root of Trust for Storage" in §4; a PCR is a securely stored (in the TPM) cryptographic checksum of a specific executable and the secure storage is current for a session since the platform was booted. Therefore, a digitally signed PCR value, which is verifiable upon a challenge by a stakeholder using a public AIK, provides an assertion that the current instantiation of the platform is running the specific executable. Notice that a PCR stores the system integrity data in an accumulative formulation which does not limit the number of executables to be stored, and there are plenty of PCRs (minimum of 24) in a TPM, hence complex site policies can be defined by combining PCRs.

TC innovation in remote platform attestation provides a powerful solution to the integrity protection of resources. Integrity protection of resources is a serious problem which the current Grid security techniques cannot solve.

5.4 Securely Virtualised OS's and Services as "Vaults"

In many enterprise organisations it is typical that many PCs run continuously while not being used for extensive periods of time, e.g., in off hours. Also, in many organisations typical uses of a PC involve word-processing like jobs which require minimal resource utilisation by the prime PC user. According to studies by Microsoft [4] PC utilisation's are between 10 to 20 percent. A similar situation also applies to the servers environment, e.g., [14].

Under the notion of virtual machines, an area of memory in a computing system can be isolated from the rest of the system to provide a simulated computer as if it were a separate computer. One piece of hardware can even enable multiple general-purpose OS's. Relations between these OS's can be configured to satisfy various access control policies. It is thus realistic that large chunks of underutilised platform resources (enterprise PCs and servers) can be organised to provide services for external demanding

users. It is obvious that a stringent security policy conformation is necessary. One basic such policies is that a virtualised OS or service should function like a "vault" which confines its users and processes to certain behaviours which cannot affect the rest of the system. For example, when a buggy code used by a prime PC user is gone dead, the rest of the system services should continue serving uninterrupted.

The TC technology can provide a strong guarantee for a stringent policy conformation. Secure OS and service "vaults" can be loaded with respective PCRs measured and stored in the TPM. IT security administrators in a resource contributing enterprise can challenge a system in the realm from time to time for policy attestation to make sure the proper functioning of the "vaults."

5.5 Group-oriented Security

Combining the distributed firewall technique in §5.2 with the remote platform attestation technique in 5.3, we can imagine a realisation of a group-oriented security for a VO. As in the case of a physical group, in a VO there also needs to be an entity in the position of acting as the group manager or a stakeholder. The group manager is responsible for defining and managing the group security policies. These policies can be setup to each site. The group security policy definition, setting up and management can be achieved using the distributed firewalls technique by letting the manager play the role of a property certification authority who issues property certificates to the group members. The group policy enforcement is then achieved by the group manager challenging and verifying the property attestation with each member of the VO.

For example, upon satisfaction of an attestation according the the VO security policy and the remote site policy, the manager could release a group session key to the attested remote environment and this group session key plays the role of the "security association" (in IPSec language) for that entity to penetrate the distributed firewall (i.e., to secure each packet both in data integrity and in message confidentiality). Thus, conference discussions in this environment can be securely conducted.

6 Trusted Computing Implementation and Deployment Status

The TCG has defined the security subsystems in such a manner so as to allow cryptographic applications to easily evolve from basic hardware protection mechanisms, such as key hardening, to more advanced capabilities, such as platform attestation and key backup and recovery services. The TCG whitepaper "Writing TCG Enabled Trusted

Applications" (at the "Downloads area" of [16]) provides an overview of the strategies that application developers may employ in developing TCG-aware client applications.

The TCG Software Stack (TSS) provides trust services that can be used by enhanced operating systems and applications. The TSS uses cryptographic methods to establish security services and trust relationships, allowing applications to maintain privacy, protect data, perform owner and user authentication, and verify operational capabilities of the platform.

The TCG Crypto Service Providers (CSPs) provide features that are commonly associated with cryptographic functionality. A TCG-enabled platform typically supports both MS Cryptographic API (MS-CAPI) and PKCS#11 [13]. If an application developer has experience writing with MS-CAPI or PKCS#11, it is relatively easy to provide basic TCG enabled capabilities. For most applications, the application developer may harden RSA asymmetric private key operations by simply calling the new CSP that is provided with TPM-enabled platforms. While there may occasionally be a subtle user experience difference based on different vendors' TSS and CSP, the TCG organisation is working to develop common interfaces and actions that may, over time, facilitate a common user experience, independent of the platform.

In order to utilise the enhanced capabilities of the TCG enabled platforms, the application developer must use the SDKs provided by the TPM manufacturer or OEM to expose the advanced trustworthy capabilities. An application developer may take advantage of a trusted platform's attestation capabilities by modifying their applications to require and verify the proper credentials provided by an attestation server. Eventually, most of the TPM and platform vendors will support the necessary credentials for attestation to function properly. Interoperability and compliance testing is being put in place and all the platform vendors have committed to support this mandatory aspect of the TCG specifications. Attestation servers are available from multiple vendors, including Verisign and Wave Systems, and some of these server products can assist in bridging the capability requirements of the platform's current limitations.

TCG-enabled PC platforms with TPM version 1.1b, both in desktop and notebook machines are now widely available from several computing systems manufactures. These include Dell, Fujitsu, HP, IBM and Intel (TCG "Fact Sheet", available at the "Downloads area" of [16]). These commercial-off-the-shelf products offer key storage for securing users' cryptographic credentials.

7 Concluding Remarks

As Grid security is becoming a more and more important topic, a number of problems remains untackled by the current Grid security solutions. We identify that group-oriented

security and distributed system behaviour conformance are among the essential requirements for Grid security while being indifferent from the current Grid security solutions. We argue that trusted computing technology, thanks to its inherent properties of group-oriented security and system behaviour conformation, can provide suitable solutions to the identified Grid security problems.

As we are still in an early stage of problem identification and solution search, the suggested approaches should be considered as initial input to substantial further investigations, which should include not only their plausibility, but also their alignment with the current Grid security solutions. Nevertheless, as hardware and software support for TC is gradually becoming available, it is timely to consider how such tools can be used to maximum effect in enhancing trust and security in Grid environments.

Acknowledgements

Greg Astfalk reviewed an early draft of this paper and provided insightful comments and suggestions. Graeme Proudler provided help with the TC technology background. Paul Vickers provided help with the Grid computing background. Nigel Edwards and Dirk Kuhlmann provided insights in the notion of secure virtualisation.

References

[1] B. Atkinson, et. al. Specification: Web Services Security (WS-Security), Version 1.0, 05 April 2002.

[2] R. Bair (editor), D. Agarwal, et. al. (contributors). National Collaboratories Horizons, Report of the August 10-12, 2004, National Collaboratories Program Meeting, the U.S. Department of Energy Office of Science.

[3] S. Bellovin. Distributed Firewalls, *;login:*, November 1999, pp 39-47.

[4] W.J. Bolosky, J.R. Douceur, D. Ely and M. Theimer. Feasibility of a service distributed file system deployed on an existing set of desktop PCs. In Proceedings of International Conference on Measurement and Modeling of Computer Systems, 2000, pages 34-43.

[5] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*, chapter 2: computational Grids, pages 15–51. Morgan Kaufmann, San Francisco, 1999.

[6] I. Foster, C. Kesselman, G. Tsudik and S. Tuecke. A security architecture for Computational Grids, 5th ACM Conference on Computer and Communications Security, pages 83–92, 1998.

[7] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the Grid: Enabling scalable virtual organizations. *International Journal of High Performance Computing Applications*, 15(3):200–222, 2001.

[8] A.O. Freier, P. Karlton, and P.C. Kocher. The SSL Protocol, Version 3.0. INTERNET-DRAFT, draft-freier-ssl-version3-02.txt, November 1996.

[9] Globus Toolkit. Available at `www-unix.globus.org/toolkit/`.

[10] ITU-T. Rec. X.509 (revised) the Directory — Authentication Framework, 1993. International Telecommunication Union, Geneva, Switzerland (equivalent to ISO/IEC 9594-8:1995.).

[11] J. Novotny, S. Teucke and V. Welch. An Online Credential Repository for the Grid: MyProxy, Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10), IEEE Press, August 2001.

[12] L. Pearlman, V. Welch, I. Foster, C. Kesselman and S. Tuecke. A Community Authorization Service for Group Collaboration, Proceedings of the 3rd International Workshop on Policies for Distributed Systems and Networks, p 50, 2002.

[13] RSA Security. PKCS#11 v2.20: Cryptographic Token Interface Standard. 28 June 2004. Available at `www.rsasecurity.com/pub/pkcs/pkcs-11/v2-20/pkcs-11v2-20.pdf`.

[14] `www.serverwatch.com/`.

[15] M. Thompson, A. Essiari and S. Mudumbai. Certificate-based Authorization Policy in a PKI Environment, ACM Transactions on Information and System Security (TISSEC), 6(4), 566-588 (2003).

[16] `www.trustedcomputinggroup.org`.