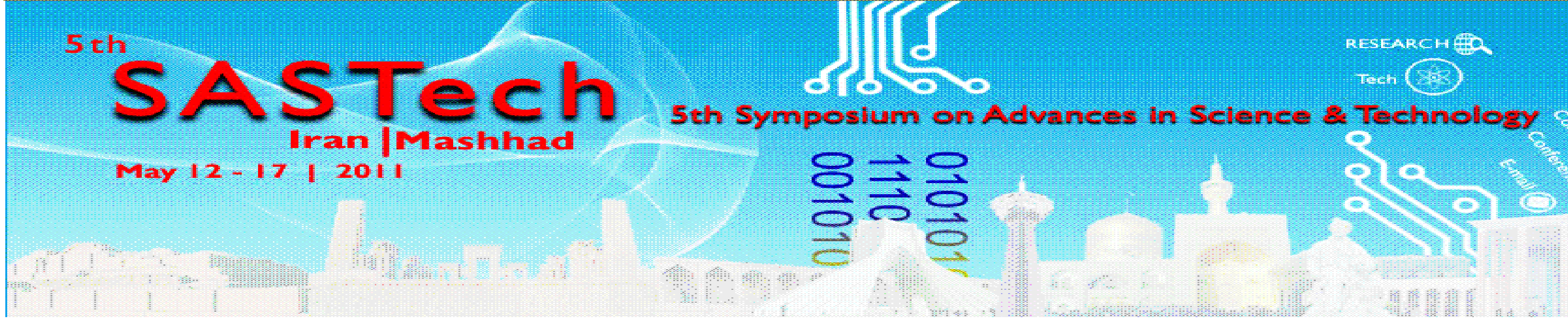

Security in grid computing

Zohre Zare:zohre.zare.it@gmail.com ,
Sakine Maleki:s.maleki1989@gmail.com
Paper Reference Number: 1047
Name of the Presenter: Zohre Zare

Abstract

In this paper, first we survey security problems which exist in grid computing and then we analyze security requirements. At the end we introduce a framework which Erin Cody has posed as a solution for security problems of grid computing. The framework presented classifies security literature into System Solutions, Behavioral Solutions and Hybrid Solutions.

Key words: Grid computing, Resources, Security

1. Introduction

Grid computing is emerging as a viable option for high-performance computing, as the sharing of resources provides improved performance at a lower cost than if each organization were to own its own "closed-box" resources. Grid computing is defined in literature as "systems and applications that integrate and manage resources[1] and services distributed across multiple control domains". According to Foster and Kesselman a grid is a system that conforms to three specific categories: it coordinates resources that are not subject to centralized control, it uses standard, open, general-purpose protocols and interfaces, and it delivers nontrivial quality of service. Kon et al. define grid computing as "coordinated resource sharing and problem solving in dynamic, multi-institution virtual organizations." Fig. 1 depicts a typical grid setup.

In this paper, first we survey security problems which exist in grid computing and then we analyze security requirements. At the end we introduce a framework which Erin Cody has posed as a solution for security problems of grid computing. The classification system group grid security solutions according to system solutions, behavioral solutions, hybrid solutions, as well as technologies related to grid that could be useful in providing grid security. The rest of the paper is organized as follows: Section 2 provides an overview of grid computing and the security issues faced in the environment, Section 3 analyzes system-based solutions, Section 4 discusses behavioral security solutions, and Section 5 discusses hybrid solutions, which embody properties of both behavioral and system solutions. Section 6 discusses related technologies, which are adaptations from other computing areas. Section 7 includes conclusions.

1.1. The Grid Security Problem

We introduce the grid security problem with an example illustrated in Fig. 1. This example, although somewhat contrived, captures important elements of real applications.

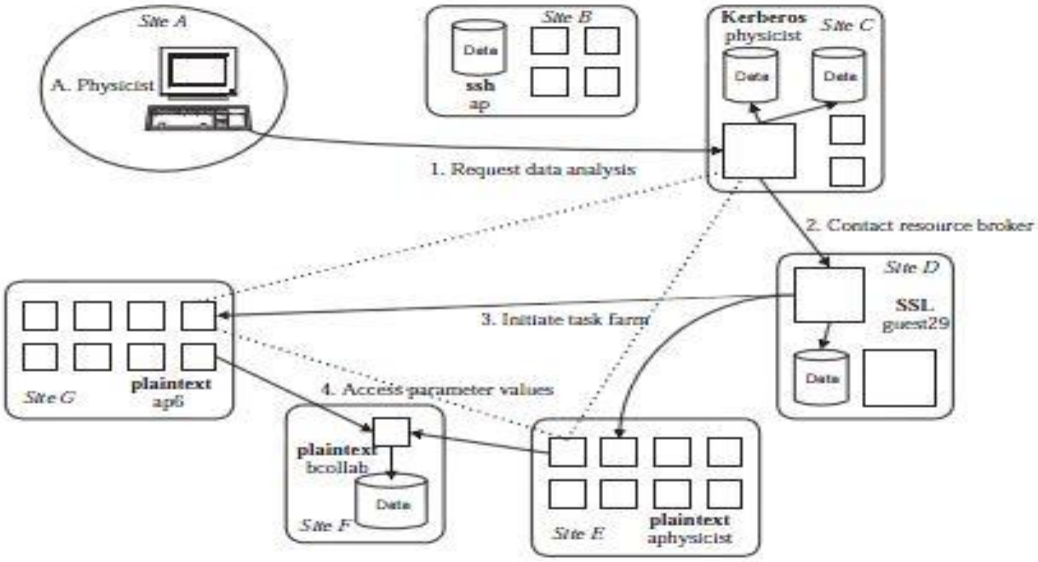

Fig. 1: Example of a large-scale distributed

We imagine a scientist, a member of a multi-institutional scientific collaboration, who receives e-mail from a colleague regarding a new data set. He starts an analysis program, which dispatches code to the remote location where the data is stored (site C). Once started, the analysis program determines that it needs to run a simulation in order to compare the experimental results with predictions. Hence, it contacts a resource broker service maintained by the collaboration (at site D), in order to locate idle resources that can be used for the simulation. The resource broker in turn initiates computation on computers at two sites (E and G). These computers access parameter values stored on a file system at yet another site (F)

[1] Resources refer to computing and management resources such as computer, software applications, etc.

and also communicate among themselves (perhaps using specialized protocols, such as multicast) and with the broker, the original site, and the user.

This example illustrates many of the distinctive characteristics of the grid computing environment:

- The user population is large and dynamic. Participants in such virtual organizations as this scientific collaboration will include members of many institutions and will change frequently.
- The resource pool is large and dynamic. Because individual institutions and users decide whether and when to contribute resources, the quantity and location of available resources can change rapidly.
- A computation (or processes created by a computation) may acquire, start processes on, and release resources dynamically during its execution. Even in our simple example, the computation acquired (and later released) resources at five sites. In other words, throughout its lifetime, a computation is composed of a dynamic group of processes running on different resources and sites.
- The processes constituting a computation may communicate by using a variety of mechanisms, including unicast and multicast. While these processes form a single, fully connected logical entity, low-level communication connections (e.g., TCP/IP sockets) may be created and destroyed dynamically during program execution.
- Resources may require different authentication and authorization mechanisms and policies, which we will have limited ability to change. In Figure 1, we indicate this situation by showing the local access control policies that apply at the different sites. These include Kerberos, plaintext passwords, Secure Socket Library (SSL), and secure shell.
- An individual user will be associated with different local name spaces, credentials, or accounts, at different sites, for the purposes of accounting and access control. At some sites, a user may have a regular account (“ap,” “physicist,” etc.). At others, the user may use a dynamically assigned guest account or simply an account created for the collaboration.
- Resources and users may be located in different countries.

To summarize, the problem we face is providing security solutions that can allow computations, such as the one just described, to coordinate diverse access control policies and to operate securely in heterogeneous environments.

1.2. Security Requirements

Grid systems and applications may require any or all of the standard security functions, including authentication, access control, integrity, privacy, and nonrepudiation. In this section, we focus primarily on issues of authentication and access control. Specifically, we seek to (1) provide authentication solutions that allow a user, the processes that comprise a user's computation, and the resources used by those processes, to verify each other's identity; and (2) allow local access control mechanisms to be applied without change, whenever possible.

In developing a security architecture that meets these requirements, we also choose to satisfy the following constraints derived from the characteristics of the grid environment and grid applications:

Single sign-on: A user should be able to authenticate once (e.g., when starting a computation) and initiate computations that acquire resources, use resources, release resources, and communicate internally, without further authentication of the user.

Protection of credentials: User credentials (passwords, private keys, etc.) must be protected.

Interoperability with local security solutions: While our security solutions may provide interdomain access meechanisms, access to local resources will typically be determined by a local security policy that is enforced by a local security mechanism. It is impractical to modify every local resource to accommodate interdomain access; instead, one or more entities in a domain (e.g., interdomain security servers) must act as agents of remote clients/users for local resources.

Exportability: We require that the code be (a) exportable and (b) executable in multinational testbeds. In short, the exportability issues mean that our security policy cannot directly or indirectly require the use of bulk encryption.

Uniform credentials/certification infrastructure: Interdomain access requires, at a minimum, a common way of expressing the identity of a security principal such as an actual user or a resource. Hence, it is imperative to employ
a standard (such as X.509v3) for encoding credentials for security principals.

Support for secure group communication: A computation can comprise a number of processes that will need to coordinate their activities as a group. The composition of a process group can and will change during the lifetime of a computation. Hence, support is needed for secure (in this context, authenticated) communication for dynamic groups. No current security solution supports this feature; even GSS-API has no provisions for group security contexts.

Support for multiple implementations: The security policy should not dictate a specific implementation technology. Rather, it should be possible to implement the security policy with a range of security technologies, based on both public and shared key cryptography.

2. Overview of grid computing security

This section gives an overview of the current grid computing environment, as well as a brief introduction of the security situations faced in today's distributed computing environment. Formation of a grid involves the organized sharing of a variety of resources with diverse ownerships. The grid gets its superior power and functionality by utilizing synergies resulting from cooperation—i.e. resource owners sharing idle disk space and processor time with users solving complex problems that their own personal resources could not handle. There are three main types of computer grids in use today: computational grids, data grids, and service grids. Each has its own set of vulnerabilities, particularly in the security area, as referenced in Table 1. Since the research papers discussed here deal with security risks that could be faced by any type of grid, it is assumed in this paper that the term “grid computing system” includes each of these three types. Note that while the types of grid listed in Table 1 represent the three common categories of grid computing systems, some grid systems can employ aspects of several or all of the three types, making them “hybrid” grid computing systems. These grids could then face any of the vulnerabilities faced by the grid types they are made up of.
Considering the grid environment's diverse and geographically separated resources and wide variety of users, each with unique needs and goals for the grid system, the issue of managing the security of users and resources becomes an issue. The users of a grid, be it computational,

data, or service oriented, may have conflicting interests with each other, and thus would want some assurance that their grid-based transactions are safe from the eyes of other users.

Without security, a grid setup would is left vulnerable to unauthorized users, malicious processes, and data tampering that could possibly render it useless. A recent survey ranked grid computing as sixth on priority lists for IT spending among industry professionals. With this much interest, security becomes necessary to protect the capital put into the grid system. Designing a secure grid implies taking into account the needs of grid users for secure remote resources that protect the integrity and confidentiality of data; and also the needs of resource owners to ensure that only authorized, trustworthy individuals are using their systems.

According to the classification system for grid computing security research, grid computing security can be classified into the following parts: systems, behavioral, hybrid, and related technologies, as shown in Fig. 2. This classification provides several benefits to the research community. By classifying first under broad ideas (systems vs. behavioral vs. hybrid) and then drilling down into more specific categories, we separate the ideas (securing of grid resources, authentication and authorization, etc.) from the implementations (Entropia, LegionFS, etc.). Further functionality is added by separating the security solutions into behavioral vs. system-based solutions. For example, a grid developer/administrator looking to implement a specific technology on his grid system might turn immediately to the system-based solutions, while a researcher studying behavioral aspects of security would have information needs more suited to the behavior section of the classification system. What follows is a brief introduction of the classification system.

Type of grid computing system	Brief explanation	Most common vulnerabilities
Computational grid	Grid architectures that focus on setting aside resources specifically for computing power; i.e. solving equations and complex mathematical problems; machines participating in this type of grid are usually high-performance servers.	Programs with infinite loops can be used to bring down nodes of this grid, decreasing functionality
Data grid	Grid architecture responsible for storage and providing access to large volumes of data, often across several organizations	Users can overwrite data of other users if they exceed their available space-this corrupts the other users' data
Service grid	A grid which provides services that are not available on a single machine	Users can use the service grid to launch Denial of Service Attack (DOS) against another site

Table 1: Types of grid computing systems

The following section discusses the system solutions from the classification system presented in Fig. 2. It is decomposed into system security for grid resources and intrusion detection systems.

3. Systems solutions

This section discusses papers that propose system-based solutions to secure grid computing environments. Rather than discussing the implementing of security policies and behavior-based solutions (as discussed in Section 4), this part of the classification deals with solutions whose focus is to manipulate the hardware and software of a grid system directly in order to achieve security. Box-product technologies, topologies and architectures, and intrusion detection systems are addressed in this section.

3.1. System security for grid resources

This section deals with research focused on system-based solutions toward grid security. Proposed solutions falling into this category seek to protect resources on the grid. Access control is a valid method for protecting resources, however it cannot be the only line of defense to ensure that grid nodes, applications, data, and communications are safe from malicious users. This category focuses on protecting the grid resources, which include hardware and computing equipment, applications running on the grid and the data that they contain, as well as communication between grid nodes. Solutions falling into this category address the data and service types of grids (Table 1) and the security situations of Immediate Job Execution and Accessing of Information.

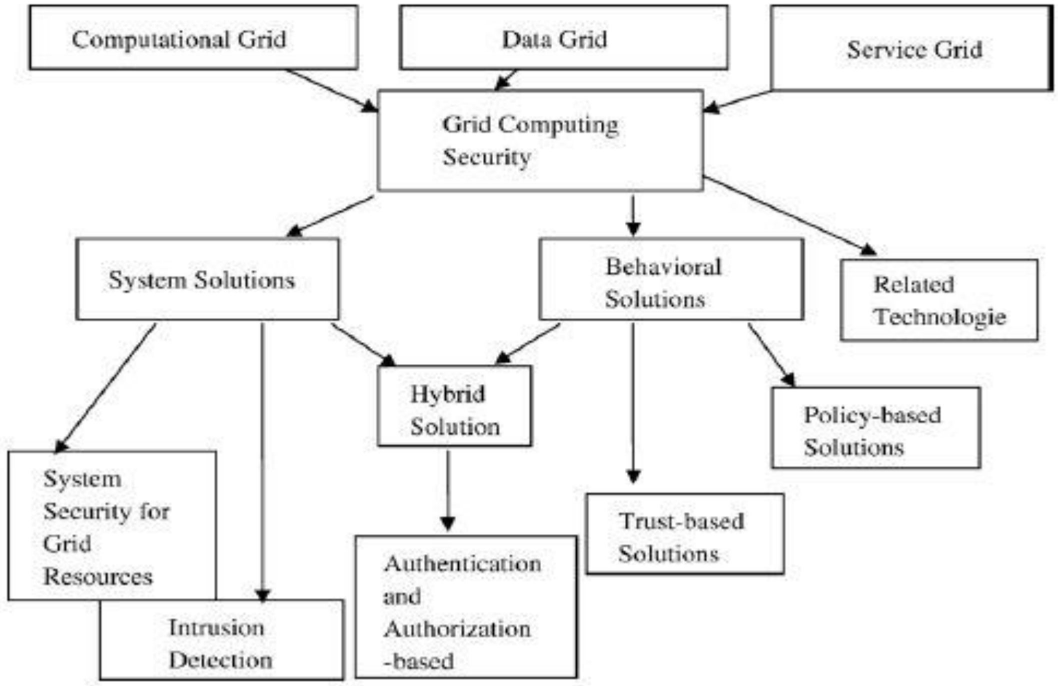

Fig. 2: Classifications for grid computing security.

3.2. Intrusion detection systems (IDS) in grid computing

This section discusses papers that present an intrusion detection system (IDS) model for securing the grid environment. For the context of this article, an “intruder” into the grid is defined as any grid user who intends to harm the grid or its resources, or intends to use the grid for purposes other than what it was designed for. Rather than being a specific software package or brandname box product, intrusion detection is a technological concept which can be implemented using any one of several software and/or hardware methods. IDS grid solutions function in the computational and service grids (Table 1) and address the security solutions of Accessing Grid Information, Setting/Querying Security Parameters, and Auditing Grid Functions.

4. Behavioral solutions

The following section addresses types of solutions that emphasize policy and management controls over hardware/software solutions to maintain a secure grid. Behavioral solutions are intangible and intuitive, rather .than employing a physical technology to maintain security in the grid. Accountability, group management, and trust are all issues that are addressed here.

4.1. Comprehensive policy controls

This portion of the review deals with papers that achieve security through policy definition. The papers suggesting trust as a security solution could be viewed as a subset of this portion. However, those papers were specific to trust, while the following papers cover several types of policies in their solutions, thus it seems more appropriate to group the trust-based papers separately. Research falling into the policy controls category, consequently, discusses policy sets governing a wide range of grid computing actions, rather than focusing on one area of activity while participating in a grid. These policies address all areas of grid computing, including authorized user selection, sign-on procedures and access control, and local vs. global security settings. Comprehensive Policy Controls function best in computational grids (Table 1) and address all security situations identified in [7].

As policy controls primarily affect the human component of the grid, comprehensive policy sets designed to manage groups of users are a logical extension of this method of grid security.

5. Hybrid solutions

A thorough review of the literature concerning grid computing security issues revealed that the particular concept of authentication and authorization of grid users could be addressed equally by system-based solutions and behavior-based solutions alike. Thus, it is more appropriate to create a Hybrid Solution sub-category to address this issue, since it falls equally under System and Behavioral grid security solutions.

6. Related technology solutions

The solutions discussed here come from technology areas other than grid computing specifically. However, these technology areas, and their associated security solutions, bear much similarity to those required by grid computing, so their contribution to security issues in grid computing should not be overlooked.

7. Conclusions

The purpose of this review was to provide an extensive literature survey of current research in the area of security in grid computing, and to identify areas of grid computing security in which more extensive research is needed. More importantly, this paper contributes to the overall body of research concerning security in grid computing through the creation of a comprehensive framework for classification of grid security research. The proposed classification system breaks grid research into four main categories, which are system solutions, behavioral solutions, hybrid solutions, and related technologies.

System solutions were further divided into system security for grid resources, which involves securing the storage and CPU power donated to the grid by resource owners; and intrusion detection system solutions, which rely on currently existing IDSs or the creation of new IDS to secure the grid. Behavioral solutions were subdivided into comprehensive policy-based solutions, which rely on policy over technology to provide security and whose policies cover a wide range of topics, and trust-based solutions, which define and quantify trust as a variable to be used for grid security. The hybrid solutions in Section 5 contain the authentication and authorization solutions for grid computing security, which include both system-based and behavior-based topics. Section 6 described related-technology solutions, which borrow from similar technology areas the security solutions that could be ported to grid computing. In the analysis of grid security literature, no one solution is perfect for every grid. Analysis of the individual situation is needed to find that situation's ideal solution.

The final, but equally important, area of future research addressed here is that of commercialization of grid resources. Currently, there are several new initiatives designed at making pre-existing grid infrastructures available for use by corporations, such as the SunGrid. This grid leasing system allows corporations to buy time on the grid for $1 per hour. As this was a corporate pricing decision reached by Sun Microsystems, information does not exist as to the reasons behind their use of this “rental” system, rather than allowing grids for sale, or as part of more structured lease agreements.

References

[1] A.S. Grimshaw, A.S. Humphrey, A. Natrajan, A philosophical and technical comparison of Legion and Globus, IBM J. Res. Develop. 48 (2) (March 2004).
[2] B. Jacob, How grid infrastructure affects application design, http://www-106.ibm.com/developerworks/library/gr-infra.html,(Document view: March 28 2006).
[3] E. Cody, R. Sharman, Raghav H. Rao, Sh. Upadhyaya, Security in grid computing: A review and synthesis, http://www.siencedirect.com (Document view: October 12 2010).
[4] H. Cassanova, Distributed computing research issues in grid computing, ACM SIGACT News 33 (3) (2002) 50–70.
[5] I. Foster and C. Kesselman, editors. *Computational Grids: The Future of High Performance Distributed Computing*. Morgan Kaufmann, 1998.
[6] I. Foster, K. Kesselman, The Grid: Blueprint for a Future Computing Infrastructure (Morgan Kaufmann in Computer Architecture and Design), 1999.
[7] L. Ramakrishnan, Securing Next-Generation Grids, IEEE IT Pro, March/April 2004.
[8] P. Shread, Survey finds grid becoming strategic IT investment, http://www.gridcomputingplanet.com/news/article.php/3323731 (Document view: March 28 2006).