Secure Semantic Service Oriented Information Grid for Cyber Physical System and Applications

Bhavani Thuraisingham, **Principal Investigator**
The University of Texas at Dallas

Collaborators: I-Ling Yen, Latifur Khan, and Murat Kantarcioglu
The University of Texas at Dallas

Sajal Das and Yonghe Liu
The University of Texas at Arlington

Elisa Bertino and Lorenzo Martino
Purdue University

Abstract

This paper describes our approach to designing a secure information grid for cyber physical systems and applications. We discuss access control and accountability for such semantic grid as well as secure infrastructure and storage issues.

1. Introduction

Assured information sharing applications are critical for the Department of Homeland Security (DHS). DHS applications such as border security, emergency preparedness, and immigration and naturalization have to share information between the different organizations. In order to implement assured information sharing systems for such applications, we need computational technologies such as grid services.

While grid computing has received a great deal of attention, little focus has been placed on the security of grid information management. This project represents the first comprehensive multi-university collaborative effort in investigating security issues for the grid, including storage, information management, and collaboration tools for sensitive data. While industry companies (e.g., Oracle) are involved in Grids, their security model is straightforward and often not capable of handling the complexity due to heterogeneous massive data and diverse partners with different data clearance. In the real world, multiple organizations and agencies have to work together and often share vital information. Therefore, comprehensive policies for data sharing, especially for confidentiality, reliability, trust and privacy, as well as accountability is extremely critical. Since security concerns and policy requirements attached to the data naturally emerge from government grid projects, it is important to integrate these into the grid computing infrastructure in a seamless way.

Our research aims at addressing the issues related to the specification, composition, and efficient evaluation of fine-grained access control policies and to design an authorization infrastructure that can easily integrate such access control policies in Information Grids. The underlying architecture for the grid will be based on the service oriented architecture (SOA) paradigm. We are developing a Secure Service Oriented Architecture-based Grid (S-SOAG) that will host the infrastructure services (e.g., scheduling), security services (e.g., attribute based access control and accountability), storage services and information management services.

One may ask the relationship between this research and cyber physical systems. As stated in [1] *a **cyber-physical system** (CPS) is a system featuring a tight combination of, and coordination between, the system's computational and physical elements. Unlike more traditional embedded systems, a full-fledged CPS is typically designed as a network of interacting elements instead of as standalone devices The expectation is that in the coming years ongoing advances in science and engineering will improve the link between computational and physical elements, dramatically increasing the adaptability, autonomy, efficiency, functionality, reliability, safety, and usability of cyber-physical systems. The advances will broaden the potential of cyber-physical systems in several dimensions, including: intervention (e.g., collision avoidance); precision (e.g., robotic surgery and nano-level manufacturing); operation in dangerous or inaccessible environments (e.g., search and rescue, firefighting, and deep-sea exploration); coordination (e.g., air traffic control, war fighting); efficiency (e.g., zero-net energy buildings); and augmentation of human capabilities (e.g., healthcare monitoring and delivery).*

Based on the above definition of CPS, we believe that the secure grid that we are designing and developing is

suitable for cyber phsyical system applications including emergency preparedness, border security and firefighting. The organization of this paper is as follows. In section 2 we will discuss security services for the grid. Section 3 discusses the pervasive infrastructure services. Secure storage services will be discussed in section 4. The paper is concluded in section 5.

2. SECURITY SERVICES

2.1 Accountability

At present there is no grid computing system that addresses accountability as part of its information assurance component. Based on our experience in the context of the TeraGrid system we have identified several crucial requirements for a suitable accountability mechanism for Grids that would be part of the secure infrastructure for the data/information grids being developed for CPS applications.

- *Decentralization.* It implies the need for distributing the accountability tasks across the nodes. Because of the distributed nature of grid systems, accountability cannot be addressed in a single location, but it must involve all the nodes where a job is processed. This requirement also calls for the need for a harmonic and consistent view of the logging information that follows from the job flow across nodes.

- *Scalability.* Scalability in our context has two dimensions: users and nodes. Today, grids have become widely accessible to large user communities because of the availability of web-based portals. Such communities have an impact on the number of job requests that are typically submitted to grids. Additionally the size of grid systems is increasing because more and more organizations are interested in sharing resources. It is important to devise solutions that scale, and thus work properly for grids of almost any size, from the ones consisting of few nodes to large infrastructures with thousands of nodes.

- *Flexibility.* A rich collection of information should be collected and be efficiently stored for later use and analysis, ranging from user authorization data to resource consumption information. The system should be able to combine heterogeneous accountability information as needed. It is however important to identify and select only the data relevant for accountability, as it is not feasible to simply collect all the potentially useful data. The identification of the type of data to collect including information about the users, jobs, and nodes should be specified by using a high-level policy language to simplify administration tasks.

- *Minimum Impact.* The accountability tools must be light-weight and must not interfere with the ordinary computation and activities performed by the grid nodes.

- *Administration Autonomy.* In the design of the system, non-technical barriers such as the coexistence of multiple administrative domains in the same grid system should be taken into account. Note that this requirement is challenging, because of the difficulty to exactly predict how grid administrators will manage their resources. For instance, it is hard to predict to what extent different administrative domains will trust each other in sharing local information with other sites. A good design should thus preserve the autonomy of grid sites, and limit as much as possible the level of collaboration required for sharing accountability data.

- *Integration with Digital Identity Management and Access Control Systems.* Because actions executed in a grid system ultimately have to be traced back to real users, it is important that the accountability system be integrated with the system in place for managing user identities. In addition, in order to connect all accountability information related to the same job, the accountability system must be aware of how users are identified across different domains. Integration with access control systems is important in order to determine which access control policies and/or which credentials permitted access for a given user, should an unintended access by this user occurs. By analyzing accountability data for access control decisions, the administrators may obtain information for revising the access control policies in place and the credentials required to gain access to the grid resources.

The challenges in the design of accountability mechanisms arise because of the heterogeneous nature of grid software and system components. Currently, few standards are available in the grid realm, and most of the adopted systems differ in the way they operate, especially in terms of tracked information and interoperability. This is especially the case of grid schedulers, access control mechanisms and identity management systems, which play an important role in accountability.

2.2 Access Control

Grid computing represents an important infrastructure that makes it possible for multiple institutions to pool their computing resources and to collaborate in order to efficiently carry on data intensive applications, such as intelligence analysis applications, and computationally intensive tasks. As more organizations and users are becoming interested in using the grid computing systems in a variety of application domains, security becomes a key issue.

Developing security approaches suitable for such a context requires addressing several requirements, such as interoperability of security mechanisms, authentication, authorization and scalability. Grid sites must be able to interoperate while continuing to use their local security solutions. Achieving interoperability is a complex task; it may require interoperability among the security mechanisms as well as the coordination of access control and authorization policies. To this extent, a unified fine-grained access control mechanism should be achieved, based not only on local user identities but also on other qualifying user attributes. Such a requirement is important in order to provide high-level access control policies that can be easily specified and understood. Grid systems dynamically evolve over time. Users and resources can dynamically be added/removed as specific projects are started/completed. Security mechanisms should be designed to reduce the security administrative overhead when dealing with re-configurations of the grid system, the user communities, and the available resources. Finally, it is still an open problem for establishing and managing trust relationship in a grid-computing environment.

Although authorization in distributed systems has been extensively investigated, not much work has been done to address authorization problems facing real large distributed systems such as grids. The current de facto solution, as represented by the GSI component included in the Globus toolkit, adopts a simple low-level approach in authorizing users to use resources at a grid site. This low-level approach relies on an access control list (gridmapfile) that maps a user's global identity (distinguished name, or DN) to a local account. Users whose DN appears on the list is authorized to use the machines, with privileges associated with the local account. This simple approach is in essence the same authorization mechanism used for a single machine, for example the Unix "/etc/passwd" file. In a distributed system like grid, there may be thousands of users and it is not realistic to base authorization decisions on individual users' identities. How can one maintain a grid-mapfile with thousands of entries? Instead, an attribute-based authorization system is desirable. In attribute-based authorization, access control policy does not mention individual user's identity, but rather attributes such as ones describing a person's role in an organization. It has been widely acknowledged in the grid-computing community that attribute-based authorization should be the direction of development for grid security. And there are a number of projects going on in this field, such as the VO Privilege Project, GridShib, and PERMIS. However, there are quite a few decision dimensions when it comes about designing an attribute-based authorization scheme for grid computing. A suitable candidate should address the emerging trends in grid computing. In particular, we observe that recent years have seen grid computing moving towards a more "virtualized" environment. The usage of computational resources is less divided by organizational boundaries, and many users subscribe to virtual communities whose members have similar information interests or computational needs. Such communities often transcend organizational boundaries, and many of them are represented by subscribers to particular websites that serve as a front-end for high-performance computing and data repositories. These trends are also sometimes demonstrated by the terminology such as "virtual organizations". Virtualized environments on one hand enhance security by isolating users and applications. On the other hand, they introduce new challenges in that one needs to assure the security, and in particular integrity, of the virtual machines; accountability may also become more difficult in such a virtualized environment.

3. Secure Pervasive Infrastructures

3.1 Characteristics of Pervasive Infrastructures

Commanding a plethora of distributed, locally controlled computing, storage, and communication infrastructures, the Grid shall not only render its capabilities be available pervasively, but also enforce designated security measures over the computing, storage, and networking services. Furthermore, the increasing mitigation toward wireless mobile devices including sensors within current large-scale data/information infrastructures further magnifies the dynamics and uncertainties involved in managing the Grid. Therefore, by "infrastructure grid" we will mean a "mobile information grid" platform capable of supporting wire-line and wireless networking, pervasive and autonomic computing, thereby also providing grid access to the mobile users carrying cell phones, PDAs, palmtops, sensors and other portable/wearable devices. The goal is to facilitate reliable sharing and exchange of information between various computing, storage and communication devices and networks as well as software/middleware platforms, for use by higher layers such as data/information/knowledge and application layers. However, not much effort has been made in pervasively securing grid infrastructures. Existing research mainly focus on how information/content is received by application users. However, many applications will need to process information, make transactions, choose among multiple sources of information, ensure secured access, and provide/utilize

various services available across heterogeneous dynamic networking environments. Some application tasks may require communications, collaborations, and interactions among heterogeneous devices; other applications may require execution of resource-intensive tasks, while some others may require flexibility in terms of physical location, type of device or network used, and the quality of service (QoS) offered.

We are planning to develop a middleware infrastructure services for supporting pervasive computing, integration of heterogeneous devices, and resource management. We are also planning to develop a middleware layer to interface with the data/information/knowledge layer with a goal to support resource/task discovery and scheduling, coordination, and secure management of the pervasive grid infrastructure. Thus the middleware to be developed will enable continual and secure access to (location-based) services and information in the information grid from anywhere, and reliable discovery of required services and associated resources. The middleware will also help build profiles of users and devices, and capture context-aware information through sensors, then learn and predict security events in the form of anomaly. The components of the secure pervasive infrastructure must work collaboratively in an integrated, pro-active manner.

Our objective is to develop an integrative, multi-layer security framework for high information assurance in the uncertainty characterized, pervasively secure (mobile) information grids. Our fundamental approach will be based on solid mathematical foundation. The expected outcome will be seamless creation, composition, management and scheduling of services (tasks) and heterogeneous resources that provide the desired QoS, availability, authenticated information sharing, reliability, security and privacy.

3.2 Dealing with Heterogeneity immobile Grids

We envision that pervasively secure grid infrastructures will exhibit two dimensions of heterogeneity, namely (i) *device diversity* in the sense that information grids be comprised of heterogeneous devices such as servers, laptops, cellular phones, and sensors which have diverse resource constraints; and (ii) *connection diversity* in that the network connection between different devices may include a multitude of communication interfaces, such as wired and wireless Ethernet, Bluetooth, and cellular phone modem; and network protocols, such as LAN, IEEE 802.11 Wireless LAN, wireless mesh 8, mobile ad-hoc networks, and sensor networks. We aim to exploit both diversities in order to control resource consumptions and information security. Specifically, our research includes the tight coupling of two innovative ideas:

1) *Dynamic Resource Management:* With the help of game theory we are investigating optimal strategies for wireless devices in a pervasive networking platform to satisfy resource and QoS constraints by dynamically configuring multiple wireless interfaces, protocols, and architectures. The generated network topology provides infrastructure support for adaptive control of sensitive information, and will be self-reorganized based on the feedback from it.

2) *Adaptive Control of Sensitive Information:* Using reputation models on the resource-aware topology, we are investigating a novel approach for devices to protect sensitive information by collaboratively and adaptively learning and submitting only the minimum information necessary for the intended data utility purposes. For example, if the intended purpose is to provide location-based services to wireless devices, our objective is to protect location privacy by transmitting only the minimum amount (i.e., maximum granularity) of location information necessary to achieve the desired service quality. The adaptive control will also provide feedback to resource management, such that no resource is wasted while processing sensitive information beyond the minimum necessary.

3.3 A Multi-layer, Integrative Security Framework

Security concern, particularly that due to node compromise or inside attacks, is one of the key reasons hindering the wide deployment of pervasive computing systems (that also involve wireless sensor networks) in mission critical applications, including border and perimeter control and monitoring, airport and harbor security, and environmental monitoring for natural disasters like hurricanes, tsunamis, and floods. We are investigating the development of a multi-layer, integrative security framework for providing high information assurance in the uncertainty characterized (often resource limited) pervasive infrastructure of wireless devices sensor networks to defend against compromised nodes with completely revealed secrets including cryptographic keys. Under such scenario, conventional cryptographic techniques solely relying on encryption/decryption and message authentication code will be essentially ineffective in defending against these attacks. This is because the adversary will be capable of successfully impersonating the compromise node and launching any internal attack, ranging from simple eavesdropping and selective packet forwarding, to much more sophisticated command forging and virus spreading.

Our goal here is to provide a novel paradigm based on a set of rich, powerful, and tractable theoretical and practical design principles, that form a multi-layer defense system that is capable of detecting, revoking, isolating, and purging compromised (sensor) nodes in the information grid in order to secure network operations. The underlying technical approach is based on information theory, epidemic theory, belief and trust model, and digital watermarking techniques. Specifically, our objectives include the following.

To summarize, our framework for **multi-layer integrated defense mechanism** will be capable of detecting, revoking, isolating, and purging compromised nodes in WSNs and providing fused and integrated meaningful information to GRID-P2P users. Our novel methodology will involve fundamental research in information theory, epidemic theory, belief-trust models, and digital watermarking. We have extensive experience and preliminary research and experimental in this area, which demonstrate the high potential and promise of the d framework. We will also provide theoretical and empirical studies on how to increase both data and communication availability for time-critical wireless surveillance networks.

4 Secure Information/Storage Services

Distributed storage systems have been increasingly used to provide highly dependable storage support for critical applications. Security is an important attribute in critical data grid applications. Data stored in data grids could be of sensitive nature. For example, a medical data grid stores medical histories and personal information of patients that needs to be protected. A criminal database may store criminal background information and should be treated in a confidential manner. A commercial data grid may store sensitive business data for restricted retrieval. Many research works in secure data grids focus on access control policies and mechanisms. Though access control is very important in enforcing proper data accesses, it would be rendered useless when some storage nodes are compromised and the sensitive data are exposed to attackers. With the increasing attacks on the Internet, it is imperative to provide mechanisms that ensure secure storage.

Encryption may be used to ensure secure storage in peer-to-peer data grids. When encryption schemes are used, the encryption keys have to be well protected and accesses to them have to be tightly controlled. Therefore, key management is an important issue. Due to the potential problems with key management, secret sharing schemes have been used to protect data in secure storage systems. Most of the existing schemes based on secret sharing based do not consider data management issues, such as where to allocate the data and how to manage the directories. In peer-to-peer data grids, distributed hash table (DHT) based schemes can be used for share distribution and look-up. Each share can be hashed with the data identifier and share identifier in a peer-to-peer network. DHT schemes provide a ready directory management scheme, but they do not bring the benefit of improved access performance that is desired in distributed storage systems. A data item may be stored far away from the clients who access the data frequently since the allocation is purely based on hashing results. This problem becomes more severe in accesses to secret shares since the probability that at least one of the *t* shares is stored far from the client is high.

We are investigating to use a two-level hybrid DHT approach for peer-to-peer secure data storage management. Data shares are stored in a cluster to balance the access cost among shares. A data objects is allocated using a DHT scheme into a cluster and its shares are hashing again into the nodes in the cluster. Data shares for the same data objects are dispersed into different administrative domains to ensure the effectiveness of secret sharing. Thus, each administrative domain is viewed as a single node even if multiple storage nodes are utilized. If a data object is frequently accessed by nodes in a cluster, then its shares are cached in the cluster to achieve better access performance. Caching decisions are made based on the access patterns of various data objects and the storage constraints. Cached data is also hashed within the cluster using DHT techniques and directory for cached objects are maintained using bloom filters.

5. Summary and Conclusion

In this paper we have described the notion of secure grids. We believe that secure grids are important for CPS applications such as emergency preparedness and border security. We have described our approach to developing security services, infrastructure services and storage services for the secure grid. Our goal is to develop a layered framework consisting of plug and play services for secure grid. In the future we envisage integrating information management services. These services will be implemented as semantic web services for the grid. We will demonstrate our technologies utilizing defense and homeland security applications.

6. Acknowledgement

This research is supported by a grant from the Air Force Office of Scientific Research. We thank Dr. Robert Herklotz for his support for this research.

7. Bibliography

Cyber Physical Systems, Wiki Entry

Atkins, D., Cyber Infrastructure, NSF Document, 2002.

M. Behrens, P. Ziu, Command and Control Grid Systems, http://www.r2ad.com/papers/Grid-Study-R2AD.pdf. white paper. Defense Information Systems Agency 2005

A. R. Butt, S. Adabala, N. H. Kapadia, R. J. O. Figueiredo, J. A. B. Fortes: Grid-computing portals and security issues, Journal of Parallel and Distributed Computing, Vol. 63, Issue 10 , October 2003

B. N. Chun and A. C. Bavier, Decentralized trust management and accountability in federated systems. 37th Hawaii International Conference on System Sciences, January 2004

Condor-g: http://www.cs.wisc.edu/condor

R. Corin, S. Etalle, J.D. Hartog, G. Lenzini, and I.Staicu, A Logic for Auditing Accountability in Decentralized Systems, In IEEE Symposium on Security and Privacy, 2006

R. Kailar, Accountability in Electronic Commerce Protocols, Proceedings of the IEEE Symposium on Security and Privacy, in Oakland, CA, May 1995

L. Khan, M. Awad and B. Thuraisingham, "A New Intrusion Detection System using Support Vector Machines and Hierarchical Clustering," to appear The VLDB Journal: The International Journal on Very Large Databases, ACM/Springer-Verlag. *The VLDB Journal* 16, 4 (Oct. 2007)

M. Huang, A. Bavier, L. Peterson, PlanetFlow : Maintaining Accountability for Network Services, Operating Systems Review, January 2006

M. Humprey, M.R. Thompson, and K.R. Jackson. Security for Grids. In *Proceedings of the IEEE*, 93(3), March 2005

M. Humphrey, M. R. Thompson, Security Implications of Typical Grid Computing Usage Scenarios, High Performance Distributed Computing, 2001

I. Foster, C. Kesselman, S. Tuecke, The Anatomy of the Grid, Intl J. Supercomputing Applications, 2001

I. Foster, The Grid: A New Infrastructure for 21st Century Science, Physics Today, Vol. 55 #2, 2002

Globus Toolkit: http://www.globus.org

Teragrid: http://www.teragrid.org

D. Chadwick. Authorisation in Grid Computing. *Information Security Technical Report*, 10(1) 2005.

P. De, Y. Liu, and S. K. Das, "An Epidemic Theoretic Framework for Vulnerability Analysis of Broadcast Protocols in Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, 2008.

A. Gupta, B. Liskov, R. Rodrigues. One hop lookups for peer-to-peer overlays. USENIX Conference on Hot Topics in Operating Systems, Vol. 9, 2003.

H. M. Ammari and S. K. Das, "Promoting Heterogeneity, Mobility and Energy-Aware Voronoi Diagram in Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, 2008

N. Zhang and W. Zhao, Privacy Protection Against Malicious Adversaries in Distributed Information Sharing Systems, *IEEE Transactions on Knowledge and Data Engineering*, 2008.

D. Ellard, J. Megquier. DISP: Practical, efficient, secure and fault-tolerant distributed data storage. ACM Transactions on Storage, Vol. 1. No. 1, Dec. 2004

W. Freeman and E. Miller. Design for a decentralized security system for network-attached storage. IEEE Symposium on Mass Storage Systems and Technologies, Maryland, March 2000.

J. Garay, R. Gennaro, C. Jutla, and T. Rabin. Secure distributed storage and retrieval. Int'l. Workshop on Distributed Algorithms, Springer-Verlag LNCS, No. 1320, 1997.

M. Kallahalla, E. Riedel, R. Swaminathan, Qian Wang, and Kevin Fu. PLUTUS: Scalable secure file sharing on untrusted storage. Conference on File and Storage Technology, California, Mar-Apr, 2003.

M.O. Rabin. Efficient dispersal of information for security, load balancing and fault tolerance. Journal of the ACM, Vol. 36, No. 2, Feb. 1989, pp. 335-348.

M. Satyanarayanan. Integrating security in a large distributed system. ACM Transactions on Computer Systems, Vol. 7, No. 3, March 1989, pp. 247--280.

A. Shamir. How to share a secret. *Communications of the ACM*, Vol. 22, No. 1, Jan. 1979, pp.

Prof. Bhavani Thuraisingham joined The University of Texas at Dallas (UTD) in October 2004 as a Professor of Computer Science and Director of the Cyber Security Research Center in the Erik Jonsson School of Engineering and Computer Science. She is an elected Fellow of three professional organizations: the IEEE (Institute for Electrical and Electronics Engineers), the AAAS (American Association for the Advancement of Science) and the BCS (British Computer Society) for her work in data security. She received the IEEE Computer Society's prestigious 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management." Prior to joining UTD, Dr. Thuraisingham worked for the MITRE Corporation for 16 years which included an IPA (Intergovernmental Personnel Act) at the National Science Foundation as Program Director for Data and Applications Security. Her work in information security and information management has resulted in over 80 journal articles, over 200 refereed conference papers, over 70 keynote addresses and three US patents. She is the author of nine books in data management, data mining and data security.