Chapter 2.5
A Policy-Based Security Framework for Privacy-Enhancing Data Access and Usage Control in Grids

Wolfgang Hommel
Leibniz Supercomputing Centre, Germany

ABSTRACT

IT service providers are obliged to prevent the misuse of their customers' and users' personally identifiable information. However, the preservation of user privacy is a challenging key issue in the management of IT services, especially when organizational borders are crossed. This challenge also exists in Grids, where so far, only few of the advantages in research areas such as privacy enhancing technologies and federated identity management have been adopted.

In this chapter, we first summarize an analysis of the differences between Grids and the previously dominant model of inter-organizational collaboration. Based on requirements derived thereof, we specify a security framework that demonstrates how well-established policy-based privacy management architectures can be extended to provide the required Grid-specific functionality. We also discuss the necessary steps for integration into existing service provider and service access point infrastructures. Special emphasis is put on privacy policies that can be configured by users themselves, and distinguishing between the initial data access phase and the later data usage control phase. We also discuss the challenges of practically applying the required changes to real-world infrastructures, including delegated administration, monitoring, and auditing.

DOI: 10.4018/978-1-61350-323-2.ch2.5

INTRODUCTION

Using compute and storage services starts with selecting an appropriate IT service provider (SP). Within their terms of use and privacy statements, SPs define which information about a customer (and, if the customer is an organization, its users) they require in order to provide the selected service. It also must be specified for which purposes the collected data will be used, and how long it will be retained. Typically, customer and user information is required for accounting and billing purposes as well as for service personalization. Generally, it thus includes personally identifiable information (PII), i.e., data that can be used to uniquely identify a single person.

In order to prevent any misuse of such sensitive data, e.g., selling email addresses to marketing agencies, legislative regulations exist; they restrict how PII may be used on an organizational level and must be mapped to technical solutions, which often have been neglected in the past, resulting in potential vulnerabilities. Although privacy and data protection laws differ between countries and dedicated regulations exist for industrial sectors such as finance and healthcare, one classic and common principle is that data must only be used for purposes which the user has been informed about and agreed to.

As intra-organizational solutions so-called privacy management systems have successfully been implemented and deployed over the past few years. They are tightly coupled with the IT services used by the customers as well as with other management systems, such as billing and invoice management tools. Whenever a user's or customer's data is about to be accessed, rule sets are evaluated to determine whether the current access attempt is in accordance with the privacy policy the user has agreed to. Basically, such systems can be viewed as an extension of traditional access management systems in order to enforce the purpose limitation principle: They also take into consideration *for which specific purpose* someone is trying to access the data; formally specifying such policies requires extensive modeling of the involved roles, the acceptable purposes, and the available PII itself.

In inter-organizational service usage scenarios, such as Grid computing, privacy protection becomes an even more complicated issue, because multiple organizations – typically also located in different countries – are involved and SPs need to retrieve the required user data from the user's home organization in an automated manner.

Instead of a single organization's privacy policy, multiple heterogeneous demands must now be fulfilled regarding PII handling. For example, there usually will be Grid-wide privacy policies, such as those specified by a virtual organization (VO); they must often be adequately combined with SP-specific or user home organization specific policies, as well as policies eventually specified by the users themselves. Combining policies requires the handling of conflicting policy parts in a transparent manner.

In general, privacy management – intentionally with a strong focus on the user – becomes a two-tiered process: First, users must decide which of their data may be submitted to an SP at all, and second they must be able to monitor and control how their data is being used later on.

In the research areas of privacy enhancing technologies (PET) and federated identity management (FIM), various solutions to these issues have been suggested, with many of them already being used in production environments by commercial as well as academic SPs; a short overview will be given in the next section.

However, these solutions were originally not suitable for certain characteristics of Grid environments, such as the concept of VOs, and cover only the PII of the users themselves; thus, they neglect sensitive data submitted along with Grid jobs, such as medical records used as input data for those programs. In this article, we first discuss these differences of Grid environments and