A Data Sharing Agreement Framework

Vipin Swarup, Len Seligman, and Arnon Rosenthal

The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102
{swarup,seligman,arnie}@mitre.org

Abstract. When consumers build value-added services on top of data resources they do not control, they need to manage their information supply chains to ensure that their data suppliers produce and supply required data as needed. Producers also need to manage their information supply chains to ensure that their data is disseminated and protected appropriately. In this paper, we present a framework for data sharing agreements (DSA) that supports a wide variety of data sharing policies. A DSA is modeled as a set of obligation constraints expressed over a dataflow graph whose nodes are principals with local stores and whose edges are (typed) channels along which data flows. We present a specification language for DSAs in which obligations are expressed as distributed temporal logic (DTL) predicates over data resources, dataflow events, and datastore events. We illustrate the use of our framework via a case study based on a real-world data sharing agreement and discuss issues related to the analysis and compliance of agreements.

1 Introduction

Data sharing refers to the act of letting another party use data. It has become prevalent with the spread of Internet technologies such as email, websites, and P2P systems. Computer users routinely share ideas, documents, multimedia files, etc. while organizational entities share large-scale data sets, e.g., scientific datasets, counter-terrorism information, etc. Four critical capabilities that are essential to any data sharing system are: a means for providers and consumers to discover each other; a means for consumers to access data; a means for consumers to understand data; and a means to manage sharing policies. In this paper, we will focus on issues regarding cross-boundary sharing policies.

In organizations, sharing provides value by improving the ability of information recipients to achieve mission objectives. However, sharing requires information to cross various kinds of boundaries and this carries risks, e.g., risks of disclosure of the information to adversaries, and risks of compromise of the sources and methods used to collect the information. In addition to these mission benefits and risks of sharing, information providers also have incentives and disincentives to share. Possible incentives include prestige, satisfying legal requirements or organizational directives, rewards, and altruism, while disincentives include fear that consumers might misuse or fail to protect the data or not give due credit to the data’s creator or provider.
The decision to share involves a trade-off between the risks of sharing and the risks of not sharing information. These risks can change dramatically when providers and consumers have confidence that certain obligations will be met by their sharing partners. This is greatly facilitated by policy management, which provides a mechanism to express and realize desired behavior of a data sharing arrangement (e.g., access controls, constraints on usage and subsequent sharing, guarantees about data quality and availability, compensation, penalties for violating agreements).

At the sharing transaction level, sharing policies have typically focused on meeting protection goals, e.g., usage control policies that are enforced using techniques such as digital rights management (DRM) engines and trusted computing platforms. At the organizational level, data sharing policies are expressed in documents called Memoranda of Understanding (MOUs) or Memoranda of Agreement (MOAs) that document the obligations and expectations of parties to the agreement. Current practice is to use textual memoranda, an approach with several disadvantages. First, there is little help for the writers, so important sharing issues are often omitted. Second, the documents are typically filed away in a drawer and seldom used thereafter. Third, it is hard to provide automated support for reasoning about the contents of textual memoranda.

Service Level Agreements (SLAs) have emerged as the paradigm used by organizations to express policies regarding service-level parameters. For instance, for network provider services, SLAs describe obligations over parameters such as availability, latency, throughput, packet loss, etc. Similarly, for data sharing services that focus on the sharing of information, SLAs can describe a variety of obligations over a similar set of parameters. However, there are several key aspects of data sharing services that are not addressed by traditional SLAs that focus on functional business services. First, data sharing obligations may require a provider to actively engage in actions that result in wider sharing of its data. These obligations may include parameters such as data freshness and quality, regular update dissemination, etc. Second, data sharing obligations may require a data recipient to engage in certain actions, e.g., further share the data, share derivatives of the data, share audit records of actions that it invoked on the data, notify the provider when the data is shared with a third party, etc. Third, obligations may restrict what the recipient may do with the data, e.g., whether the recipient can print a document. Finally, data objects subsume other data objects and this property can be exploited—e.g., an agreement about European Union persons’ data is relevant to a demand for German persons’ data.

Obligation policies are a key component of SLAs. An obligation is a constraint that a principal commits to satisfy in the future. Obligations are distinct from provisions (i.e., actions that a principal takes in order to gain access to a resource) and expectations (i.e., actions that one principal expects another to satisfy). Unlike access control rules which must always be satisfied, obligations may be violated. Hence, obligations are typically associated with penalty clauses that describe the consequences of violating the obligations. Obligation platforms provide support for obligation policies, maintain representations of obligations,