An access control framework for data integration

Mohand-Sad Hacid
Universit de Lyon, Universit Lyon 1
LIRIS CNRS
Lyon, France
Mohand-Said.Hacid@liris.cnrs.fr

Mehdi Haddad
Universit Paris-Est Crteil
LACL
Paris, France
mehdi.haddad@u-pec.fr

***Abstract*—Nowadays, database systems are essential to any business managing information. For example, even web sites usually make use of databases as data repositories behind the interface. A database is necessary to provide the requested service. In addition to web applications, databases play an important role for many corporations. Indeed, databases are used to manage all the needed data in such corporations in order to achieve their business. Today database systems are concerned with the management of health data, insurance data, scientic data, legislation data, military data, human communications data, emails and tweets among other kinds of data. This demonstrates the importance databases play in our everyday life. This information should also provide services able to control access to information they handle. Indeed, even though these systems oer great benets, users are reluctant to use such systems if their privacy is not preserved. The importance of database systems and their security at economical, scientic and societal levels has led governmental organizations to recognize the need for information security at both technological and societal levels. In this talk, we are going to explore the role of formal approaches to database security and mainly in the context of data integration.**

***Index Terms*—data integration, access control, inference problem**

I. Introduction

In this talk, we discuss problems and approaches to controlling the access to a data integration system. In a data integration system, a mediator is defined. This mediator aims at providing a unique entry point to several heterogeneous sources. In this kind of architecture security aspects and access control in particular represent a major challenge. Indeed, every source, designed independently of the others, defines its own access control policy. The problem is then: "How to define a representative policy at the mediator level that preserves sources policies?" Preserving the sources policies means that a prohibited access at the source level should also be prohibited at the mediator level. Also, the policy of the mediator needs to protect data against indirect accesses. An indirect access occurs when one could synthesize sensitive information from the combination of non-sensitive information and semantic constraints. Detecting all indirect accesses in a given system is referred to as the inference problem.

Securing data integration systems requires understanding the relationships that could arise between data. These relationships could influence the way access control and policy composition have to be performed. Data semantics (e.g., functional dependencies) could allow users to indirectly access prohibited data. We will discuss a methodology that allows controlling the access to a data integration system. This methodology allows dealing with both direct and indirect accesses. The different phases of this methodology are the following:

- ***Propagating and combining source policies:*** This phase aims at transferring the source authorization rules to the mediator. This propagation preserves policy composition principle. Indeed, it ensures that every sources authorization is preserved at the mediator level. This phase preserves data sources from direct access.
- ***Detection phase:*** This phase characterizes the role that data semantics plays in inferring sensitive information. In particular we focus on semantic constraints expressed by functional dependencies. We also introduce in this phase a graph-based approach that allows identifying flaws that could remain after the propagation phase is completed. Indeed, the propagation phase considers only direct accesses. The detection phase identifies, a priori, a set of transactions that could allow malicious users have access to sensitive information.
- ***Reconfiguration phase:*** This phase proposes solutions to deal with policy violations. Violating transactions are first identified at design time. In this phase, we propose solutions for avoiding policy violations that could happen if inference mechanisms are used.

Acknowledgment

This work is supported by Thomson Reuters in the framework of the Partner University Fund project : Cybersecurity Collaboratory: Cyberspace Threat Identification, Analysis and Proactive Response". The Partner University Fund is a program of the French Embassy in the United States and the FACE Foundation and is supported by American donors and the French government.