Journal of Computer and System Sciences 80 (2014) 1493–1503

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Building access control policy model for privacy preserving and testing policy conflicting problems

Hua Wang[a], Lili Sun[a], Elisa Bertino[b]

[a] Centre for Applied Informatics, Victoria University, Victoria, Australia
[b] Purdue University, West Lafayette, United States

ARTICLE INFO

Article history:
Received 8 January 2013
Received in revised form 20 August 2013
Accepted 10 April 2014
Available online 16 April 2014

Keywords:
Purpose
Privacy
Access Control
Conflicts

ABSTRACT

This paper proposes a purpose-based access control model in distributed computing environment for privacy preserving policies and mechanisms, and describes algorithms for policy conflicting problems. The mechanism enforces access policy to data containing personally identifiable information. The key component is purpose involved access control models for expressing highly complex privacy-related policies with various features. A policy refers to an access right that a subject can have on an object, based on attribute predicates, obligation actions, and system conditions. Policy conflicting problems may arise when new access policies are generated that are possible to be conflicted to existing policies. As a result of the policy conflicts, private information cannot be well protected. The structure of purpose involved access control policy is studied, and efficient conflict-checking algorithms are developed and implemented. Finally a discussion of our work in comparison with other related work such as EPAL is presented.

1. Introduction

Privacy preserving is increasing in its importance since privacy becomes a major concern for both customers and enterprises in today's corporate marketing strategies. This raises challenging questions and problems regarding the use and protection of private messages, especially for context-aware web service [6]. One principle of protecting private information is based on who is allowed to access private information and for what purpose [2]. For example, personal information provided by patients to hospitals may only be used with record purpose, not for advertising purpose. There must be purposes for data collection and data access. The motivations for adopting purpose based approach are 1) the fundamental policies for private information concern with which data object is used for what purposes [20] (for example, customers' age and email address are used for the purpose of marketing analysis), and 2) customers agreed data usage varies from individual to individual. Information technology provides the capability to store various types of users' information required during their business activities. Indeed, Pitofsky [21] showed that 97 percent of web sites were collecting at least one type of identifying information such as name, home address, e-mail address, or postal address of consumers. The fact that the personal information is collected and can be used without any consent or awareness violates privacy for many people. This paper analyzes purpose based methods to secure private information.

Data privacy is defined by policies describing to whom the data may be disclosed and what are the purposes of using the data [1]. For example, a policy may specify that price of an air ticket from an agent may be disclosed, but only with

E-mail addresses: wang@usq.edu.au (H. Wang), sun@usq.edu.au (L. Sun), bertino@cs.purdue.edu (E. Bertino).

http://dx.doi.org/10.1016/j.jcss.2014.04.017

"opted-in" customers, or that the price will be disclosed unless the agent has specifically "opted-out" of this default. While there is recent work on defining languages for specifying privacy policies [22,11], access control mechanisms for enforcing such policies have not been investigated [16]. Ni et al. [19] analyzed a conditional privacy management with role based access control, which supports expressive condition languages and flexible relations among permission assignments for complex privacy policies. But many interested problems remain, for example, developing a formal method to describe and manage purposes and to automatically detect possible conflicts between access policies. As stated by Al-Harbi and Osborn [4] and Adams and Sasse [3]: "Most invasions of privacy are not intentional but due to designers' inability to anticipate how this data could be used, by whom, and how this might affect users"?

Access control is significant when disclosing private information in web service [14]. The importance of privacy has been recognized for a long time, but the concept has not been supported in traditional access models, especially purpose based access control systems. A security officer has to check privacy policies if an access is required. Furthermore, administrators are prone to making mistakes when they generate new access policies to access sensitive data [7]. Such an approach significantly increases the management efforts in distributed environments because of the various privacy requirements and the continuous involvement from security officers. This paper bridges the gap between private information protecting technology and access control models. We start from building a purpose-based access framework and analyzing the conflicts between purposes in access control policies.

The remainder of this paper is organized as follows: Section 2 presents the motivations behind our work in this paper. Section 3 proposes a purpose based access framework which includes detailed information of purposes and access control evaluation. Section 4 provides access control policy structure and authorization models as well as illustrates the impact of generating a new access policy through examples. Section 5 describes conflict problems in access purposes and policies, and develops algorithms for detecting conflicts between purposes. The implementation of the conflicting algorithms is described in Section 6. Section 7 compares the work in this paper and related previous work, the comparisons demonstrate the significance of the work in this paper. Finally, the conclusion of the paper and further work are given in Section 8.

2. Motivations

The important techniques for private information occur in distributed systems specifically tailored to support privacy policies, such as the well known P3P standard [27,11,13]. In particular, Agrawal et al. [2] introduced the concept of Hippocratic databases, incorporating privacy protection in relational database systems. An important feature of their work is that it uses some privacy metadata, consisting of privacy policies and privacy authorizations stored in privacy-policies tables and privacy-authorizations tables respectively. However, they neither discussed the concepts of purpose with hierarchy structure, nor the prohibition of purpose and association of purposes and data elements. LeFevre et al. [15] presented an approach to enforce privacy policy in database systems. They introduced two models of cell level limited disclosure enforcement, namely table semantics and query semantics, but did not consider access control management. Li et al. [16] devised generalization boundary techniques to maximize data usability while, minimizing disclosure of privacy. Inspired by the fact that the permissible generalization level results in a much finer level access control, the authors proposed a privacy-aware access control model in web service environments and also analyzed an access process management through a trust-based decision and ongoing access control policies. However, the concept of purpose was missed. Ni et al. [19] analyzed a role-based access model for purpose-based privacy protection, but their work did not consider usage access management and the conflicts between purposes in policies. The development of access technology entails addressing many challenging issues, ranging from modeling to architectures, and may lead to the next-generation of access management. This paper develops purpose based access technology for privacy violation challenges including complex policy structured models with access control.

Privacy violations may happen when data are released to third parties [2]. Data once released are not any longer under the control of the organizations owning them, and the data owners are not able to control the way data are used. The most common approach to address the privacy of released data is to modify the data by removing all information that can directly link data items with individuals [24]. It is important to note that simply removing identity information, like names or social-security numbers, from the released data may not be enough to anonymize the data. Many examples show that even when such information is removed from the released data, the remaining data combined with other information sources may still link the information to the individuals it refers to [23]. Sweeney [25] proposed approaches based on the notion of k-anonymity as solutions of the problem. Another secure private information techniques such as density-based clustering algorithms happens in the context of data mining [18].

Data mining techniques are today very effective. Thus even though a database is sanitized by removing private information, the use of data mining techniques may allow one to recover the removed information. These techniques are based on modifying or perturbing the data in some way; for example, techniques specialized for privacy preserving mining of association rules modify the data so to reduce the confidence of sensitive association rules [12]. A problem common to those techniques is represented by the quality of the resulting data; if data undergo too many modifications, they may not be useful any longer [10].

Secure private information cannot be easily achieved by traditional access management systems because traditional access management systems focus on which user is performing what action on which data object [28], and privacy policies are concerned with which data object is used for what purpose(s). For example, a common privacy agreement between a data collector and customers is "we use customer information for marketing purposes and to enable help us to resolve problems