Macrothink Institute™

International Journal of Management Innovation Systems
ISSN 1943-1384
2009, Vol. 1, No. 1: E5

An Novel Approach to Privacy Management System

N. Maheswari (Corresponding Author)

P.G. Department of Computer Science

Kongu Arts and Science College, Erode-638 107, Tamil Nadu, India

E-mail: mahii_14@yahoo.com

K. Duraiswamy

K.S. R. College of Technology, Tiruchengode- 637 209, Tamil Nadu, India

E-mail: kduraiswamy@yahoo.co.in

Abstract

This paper addresses the problem of dealing with privacy management of confidential data stored by enterprises and other organizations. We describe an innovative solution based on an adaptive privacy management system. In this system (arbitrarily complex) data are retrieved from standard data repositories, in such a way that parts of these data are obfuscated and associated with privacy preserving techniques. Confidential data are "first class" objects that can be sent to other parties. Entities that try to access their content can be different from those entities that retrieve these objects. In particular, a Privacy Management System decides what is visible at a given time for each specific request for content. The visibility of (and access to) the obfuscated data is adaptive, depending on the requestor, the context and purpose.

Keywords: Privacy management, Privacy preserving, Entity, Data mining, Data repository

1. Introduction

Enterprises store large amounts of confidential data about their employees, customers and partners. On the one hand, accessing and managing this data is fundamental for their business: confidential information is retrieved, analyzed and exchanged between people (and applications) that have different roles within an organization (or across organizations) to enable the provision of services and transactions. On the other hand, data protection and privacy laws, including (Laurant 2003,QECD 2001,Alliance 2004), dictate increasingly strict constraints about how these data have to be protected, accessed and managed. Failure to comply with such privacy laws can have serious consequences for the reputation and brand of organizations and have negative financial impacts. There is therefore a need to reveal sensitive data but this must be done in a way that is legally compliant.

Privacy management technology can help achieve such a balance: this paper describes the approach to addressing the problem above by providing an adaptive privacy management system for data repositories. Our main objective within this work is to enable adaptive access to confidential information based on the satisfaction of privacy policies with a minimal impact on data repositories in terms of required technological changes. The latter is important in order to aid the practical deployment of the system.

A privacy model is introduced, based on: a Privacy Management System used by people and applications to mediate their interactions with data repositories as dictated by privacy policies, and one or more Privacy Preserving Techniques dealing with the enforcement of privacy policies. The process of disclosing confidential data is adaptive to contextual information

2. Problem Statement

The key problem addressed in this paper is the management of privacy for confidential data stored by enterprises and other organizations.

Privacy management is not just a matter of authentication and authorization. When dealing with confidential (personal) data - among other things - it is necessary to capture the purpose of data, convey the consensus of the data owners (subjects) and make decisions on access requests based on the requestors' intentions.

Privacy preserving techniques can dictate additional terms and conditions under which access to confidential data can be granted: this involves the satisfaction of constraints and obligations which might require the processing of credentials, trust verification and management of contextual information.

In large organisations, people have different roles and skills: business tasks are achieved thanks to collaboration among these people. The rigid enforcement of privacy policies might create disruptions in business practices and introduce unacceptable burdens. For example, confidential data can be stored in a variety of data repositories. Only technical specialists might have the right skills to retrieve these data in a way that is meaningful for business people, marketing departments or strategists. Unfortunately, privacy policy constraints might dictate that these technical people must not access confidential data: in this case they would

not be able to provide a service to the business people. Similar observations apply for applications and services run by different organizations within an enterprise.

Mechanisms are required to address both privacy requirements and business needs. Entities or applications must be enabled to retrieve confidential data by searching data repositories. The process of accessing confidential information has to be flexible and adaptive to contextual information and privacy policies.

An entity should not be prevented from acting on behalf of other people when searching and retrieving data, even if they cover different roles and have different privacy clearances. In such a case different views of this data must be provided, according to predefined privacy policies. In case of non-compliance to specific privacy policies, parts of this data might be removed or simply obfuscated.

3. Related Work

Relevant work in the area of privacy management for data repositories has been carried out in the area of data encryption. Mechanisms and solutions have been built to encrypt confidential data when it is stored in data repositories. Significant work in this space has been done with Translucent Databases (Wayner 2002). Most of these solutions focus on the "confidentiality" and access control aspects: they have little flexibility in providing policy-driven mechanisms encompassing aspects beyond authentication and authorization i.e. dealing with data purpose, matching the requestors' intentions against this purpose, enforcing obligations, etc.

Seminal work has been done by IBM with their research on Hippocratic Databases (IBM-HIPP 2004), i.e. databases that include mechanisms for preserving the privacy of the data they manage. Their proposed architecture is based on the concept of associating privacy metadata (i.e. privacy policies) to data stored in data repositories, along with mechanisms to enforce privacy. The drawback of this approach is that it might require substantial changes to current data repository architectures, an approach that might take a long time and require substantial investment (of all the involved parties) to succeed. These changes include adding privacy metadata via additional database tables and using modified Java Database Connectivity (JDBC) data adaptors (IBM –HIPP 2004) that deal with these privacy metadata and interact with external privacy engines: this will require customers to buy upgraded versions of databases. In addition, this approach does not take into account that the management of privacy spans across the database boundaries: such management has to be carried out within a broader context as it encompasses aspects such as the management of enterprise-wide privacy policies, obligations and application/service-based privacy policies.

In terms of commercial products, the state of the art in this space is IBM Tivoli Privacy Manager (IBM-TIVOLI 2004, IBM-DOC 2004) This provides mechanisms for defining fine-grained privacy policies and associating them to data. Privacy policies are based on P3P (W3C 2004) but they will evolve towards privacy authorization-based policies - based on the EPAL(IBM-EPAL 2004) specification, i.e. policies containing authorization constraints along with constraints on contextual information and intents. This approach addresses the privacy management problem purely from an access control perspective. It does not include additional aspects relevant for privacy management such as trust management and dealing with ongoing privacy obligations dictated by legislation and enterprise's guidelines.

Our approach differs from the above solutions in that it aims at leveraging current data repository technologies and reducing to the minimum the impact on them, in terms of required changes. In our approach, interactions with data repositories can still happen as usual but with the additional guarantee that confidential data is now protected and contextually released, in a fine-grained way, based on the fulfillment of associated privacy policies.

Additional relevant work has been carried out for privacy management in the area of data mining and statistical databases. In this context, the main goal is to prevent privacy violations when using data mining learning algorithms, data correlations and linking techniques. Current privacy management techniques involve the provision of statistical approaches (i.e. information is not returned as it is but it is statistically modified, for example to reflect average values), data obfuscation and knowledge hiding.

Our approach is complementary to these techniques. We focus on privacy management for traditional data repositories rather than techniques for On-Line Analytical Processing (OLAP) systems and data mining. Our main objective is to ensure that stored data is accessed in a privacy compliant way, in dynamic environments. Some of the technical approaches we introduce in this paper could also be applied in the context of data mining.

4. Solution

This section introduces the model underpinning our privacy management solution, discusses a few relevant scenarios.

4.1Model

The model underpinning our solution consists of three basic components, as shown in Fig. 1.

1. Privacy Management System
2. Privacy Preserving Techniques
3. Data Repository

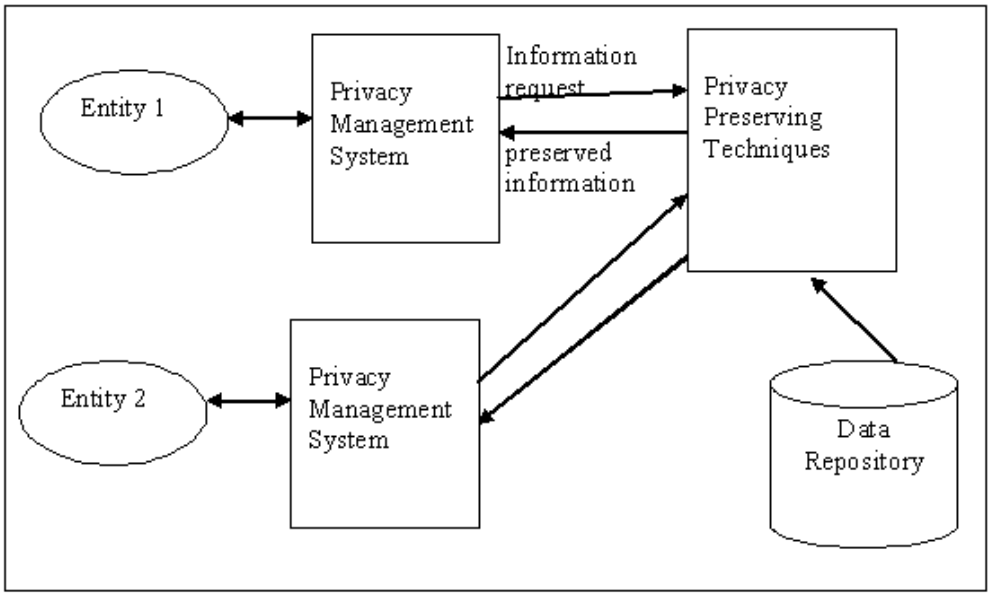

Fig.1 Privacy Management System

The Privacy Management System mediates the interactions between an entity (data requestor), data repositories and the Privacy Preserving Techniques. It allows users to retrieve confidential data from standard data repositories: part of these data can be stored in an obfuscated way and along with the associated privacy preserving techniques.

The Privacy Preserving Techniques decides which confidential information can be accessed by an entity at a specific point in time by taking into account the specific request, relevant privacy policies and requestor's credentials. By privacy-preservation we mean that the techniques must protect individual privacy in the data repositories. This may be crucial to certain knowledge management systems (e.g., the systems coordinating government agencies counter-terror activities). Privacy preserving techniques (Verykios, Bertino, Fovino, Provenza, Saygin, Theodoridis, 2004), are applied for the data. The techniques helps in transferring the information required by the entity and the other confidential information in the database will be hidden.

Figure 1 shows example views of confidential data that are provided by our system. Confidential data can be retrieved by people and applications that have no rights to access its content (i.e. they do not satisfy privacy policies) but are in charge of querying data repositories on behalf of other people: in this case the content is not de-obfuscated. Nevertheless, the obfuscated data can be sent to other entities that can access their content if they satisfy the associated privacy policies.

This basic model can be extended and adapted to a variety of scenarios including enterprise and inter-enterprises contexts. In particular the Privacy Preserving Techniques can be provided by an organization for internal consumption or by one or more external trusted third parties, to enable multi-party interactions and at the same time increase the overall trust and accountability.

4.2 Scenarios

This section briefly describes a few scenarios where our solution adds value in the management of privacy for confidential data:

• *Enterprise Scenario*: an enterprise collects confidential data about employees, customers, partners, etc. People (or applications/services), with different roles and objectives might need to access this confidential information. Roles played by people include IT technicians, researchers, marketing people, project managers and HR people. The kind of confidential information they can access must depend on their role, their declared intent, purpose of the stored data, enterprise policies, legislation and specific customers' (opt-in and opt-out) policies;

• *Federated Identity Management Scenario*: Confidential information can be sent from a service provider *A* to a service provider *B* in the context of multi-party electronic interactions driven by a transaction. Depending on who initiated the transaction (customer, service provider, etc.), the purpose of data and also customers' policies, a subset only of the whole data may be accessed and sent to the other parties, as dictated by privacy policies. For example policy constraints could dictate that specific portions of confidential data cannot be

sent outside an organization for marketing reasons or that it can only be sent to a predefined set of organizations to enable customers' transactions.

• *Healthcare scenario*: it is important to have access control on a patient's medical record. Administrative staff, doctors, nurses, lab technicians, insurance providers, and researchers may have access to some but not necessarily all of a patient's information. Access to information depends on the purpose of data, the intention of the entity trying to access this data and the satisfaction of any specific fine-grained patient's preferences.

By using our approach we are able to associate fine-grained privacy preserving techniques to obfuscated confidential data and force requestors to be compliant to these techniques if they want to access the data. This can be achieved in a flexible way, without *a priori* preventing the various entities from interacting, as dictated by business processes.

State of the art solutions can provide censored responses where private information is stripped out. Our solution can do this as well; the main competitive advantage of our approach consists of the fact that data could be retrieved by people who might not be entitled to access confidential parts of it but are authorised to collect and organise this information on behalf of other people (who might have the right to access it). These data are obfuscated and their deobfuscation is subject to the fulfilment of privacy techniques. As a consequence, incremental disclosure of confidential data can be obtained by requestors by providing the right credentials and satisfying privacy constraints.

Compared with traditional "views" on data (for example views on database tables), our approach reduces the need for defining a broad set views to accommodate multiple different cases, depending on requestors' capabilities and clearance: access and privacy constraints are direclty associated to data and dictate what can be seen at any point in time.

5. Discussion

It is the case that our Privacy Management System can potentially be bypassed as requestors could try to access data by applying data mining techniques. However, in this case, any obfuscated data is going to be unintelligible. This forces the requestor to interact with the Privacy Preserving Techniques as dictated by the associated privacy policies.

A more problematic issue arises because once confidential data have been disclosed to a legitimate requestor (that satisifed the associated privacy policies), it may not be possible to prevent this entity from misusing these data. At this stage also the association of sticky policies to data can be broken. Unfortunately, this is a common problem for systems that must enforce privacy and at the same time must release confidential data. With our approach we ensure that sticky privacy policies are strongly associated to data at least until the first disclosure happens.

We need to fully understand how applications and services will deal with the association of privacy policies to data. This is definitely work in progress.

Another important aspect that needs to be explored further is the overall lifecycle management of privacy policies associated to confidential data, including their renewal and

modification. The management of keys is strictly related to the management of policies as decryption keys will be issued based on policy fulfillment. By changing a policy, our system can automatically change the associated encryption key. Revocation of keys and one-time usage of keys have to be addressed in this context. Related to these aspects, we are currently looking at ways to change encryption keys based on successful disclosures of data. This could be done via a combined interaction between the Privacy Management System and the Privacy Preserving Techniques.

6. Conclusion

This paper describes an innovative approach to deal with an adaptive management of privacy for confidential data. The discussed solutions, based on a Privacy Management System and the Privacy Preserving Techniques, allow an incremental disclosure of confidential data depending on the satisfaction of privacy policies, with minimal disruption to common business interactions. Confidential information can be retrieved and transmitted between people that potentially have the right to access only parts of it: different views (in the sense of visible data) of this information are provided, depending on the requestors' credentials, the context and privacy policies. This is the main advantage of our approach if compared with current solutions.

References

IBM-HIPP. (2004). *HippocraticDatatabases*. http://www.almaden.ibm.com/software/Quest/Projects/hippod

IBM-TIVOLI. (2004). *IBM Tivoli Privacy Manager*. http://www 306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/

IBM-DOC. (2004). *IBM Tivoli Privacy Manager*. online technical documentation, http://publib.boulder.ibm.com/tividd/td/PrivacyManagerfore-business1.1.html

IBM-EPAL. (2004). *The Enterprise Privacy Authorization Language (EPAL)*. EPAL specification, http://www.zurich.ibm.com/security/enterprise-privacy/epal/

Laurant, C. (2003). *Privacy International - Privacy and Human Rights 2003: an International Survey of Privacy Laws and Developments*, Electronic Privacy Information Center (EPIC), Privacy International. http://www.privacyinternational.org/survey/phr2003/

OECD. (2001). *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. http://www1.oecd.org/publications/e-book/9302011E.PDF

Online Privacy Alliance. (2004). *Guidelines for Online Privacy Policies*. Online Privacy Alliance, http://www.privacyalliance.org/

Verykios, V, Bertino, E, Fovino, I.N, Provenza, L.P, Saygin, Y, Theodoridis, Y. (2004). *State- of-the-art in privacy preserving data mining*. ACM SIGMOD Record, 3(1), 50-57.

Wayner, P. (2002). *Translucent Databases*, Flyzone Press

W3C. (2004). *P3P specification.*, http://www.w3.org.P3P/brochure.html