

Privacy Enhanced Data Management for an Electronic Identity System

S. Nimalaprakasan, S. Ramanan, B. A. Malalaseena, K. Shayanthan, C. Gamage and M. S. D. Fernando

Abstract— The electronic identity (eID) is being positioned to be a basic tool for identification, authentication and authorization in application domains ranging from eCommerce in private sector to eGovernance in public sector. A practical and flexible eID should be usable in both a network-connected online setting as well as in conventional offline situations. While improving security of communication and enhancing access control to resources, eID schemes also have the potential to become a serious negative factor on user privacy rights. This paper discusses the specific issue of privacy protection in eID systems and considers a range of solutions that could be implemented in a privacy-enhanced eID system featuring both data access and data management.

I. INTRODUCTION

THE changes been brought about by concepts such as eSociety, eGovernment and eCommerce, which is moving people towards computer and network based services in public and private sectors, require electronic identities as an important means to facilitate interactions between citizens and organizations. The electronic identity (eID) is the electronic counterpart to a legally accepted ID such as national identification card, driving license, passport, etc. A person can present an eID to a computer-based system to prove his or her identity or their right to access information or services. This presentation of an eID may happen to an offline standalone system or to an online networked system. The personal data of eID holders and data belonging to the service providers should be accessed and managed in a secure way. The protection of identity and related data, while providing services in the digital domain, is of the utmost importance to prevent unauthorized or inadvertent disclosure of personal information, identity theft, impersonation and violation of privacy rights.

The notion of privacy is defined by Westin [1] as “the claim of individuals to determine for themselves when, how and to what extent information about them is communicated to others”. The privacy of individuals is a fundamental right and has added importance in this information age [2] of massive information storage and flow.

Manuscript received June 30, 2009.

S. Nimalaprakasan, S. Ramanan, B. A. Malalaseena, K. Shayanthan, C. Gamage and M. S. D. Fernando are with Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. (e-mail: nimal@ieec.org, {ramanan, anuradha, shayanth}@project-eid.org, {chandag.shantha}@uom.lk)

The basic Informational privacy right has two distinct characteristics:

1. The right to be left alone
2. The right to decide what to reveal about oneself.

It is reasonable to assume that people would object to being tracked via the usage of their eIDs for accessing services and being profiled without their consent. If people feel that their privacy rights are being violated or the rights are under threat, it may prevent them from making full use of the electronic based services leading to a failure of the system as a whole [3]. Therefore, individuals should be able to ensure the privacy of their data and activities being accessed or tracked by unauthorized third parties.

In this context, privacy cannot be just another pluggable feature for eID, but must be an integral and central requirement of an eID system. The privacy of information should be ensured while it is transit over networks as well as while the data is stored at various nodes of an information system. Also, the privacy preserving and enhancing mechanisms in an eID system should be enforced in both user side identity management and server side data management.

In this paper, we focus on privacy specific issues and propose policies and mechanisms that could enhance privacy in data management and data access of an eID system. The remainder of this paper is organized as follows. In section 2 we present authorization models in the context of an eID system. This includes privacy-aware access control policies and data protection policies for an eID system. In section 3 we discuss cryptographic schemes and mechanisms that could be used to provide privacy in an eID system based on proposed authorization models. In section 4 we describe user side identity data management and in section 5 we describe server side identity data management. The concluding remarks are given in section 6.

II. AUTHORIZATION MODELS

The success of any identity system depends on the acceptance by its users and privacy issues play a vital role in user acceptance of any electronic identity system. Traditionally, access control is enforced based on policies that specify who can or cannot access/manipulate some data. However in an eID system, data may be accessed in different contexts and users are required to disclose a wide range of distributed information about them including static

identity attributes and dynamic location attributes and transaction attributes. In this scenario, privacy becomes vital as studies have revealed [3] that some users prefer to abort a transaction rather than disclose what they consider to be private information. Also, many users expect a degree of control over secondary use of data to protect their privacy. Protecting user identities by providing anonymity, pseudonymity, unlinkability, and unobservability of users is needed to address privacy issues and these need to be enforced at communication level, system level, and application level. This introduces the need for defining authorization policies and models and the development of new paradigms for access control and in particular paradigms for authorization specification and enforcement for eID systems.

A. Main functional requirements

The following requirements must be considered in an eID system to enhance user data privacy and enforce relevant access controls.

1) Privacy

Access control needs to guarantee the enforcement of the privacy requirements. There are two principal problems that need to be considered: (1) the definition of privacy-preserving access control policies; which requires considering, expressing, and combining protection requirements taking in to account both direct and indirect release of information and (2) information may not be under the control of a single authority; where privacy policies related to information must take in to consideration not only the privacy requirements of the owner, but also the privacy requirements of the collector and relevant privacy laws. It is important that privacy requirements be associated with the data during their movement among different parties in a system and the parties that receive the information follow the privacy rules when managing them. These multiple authorities' scenario should be supported from the administration point of view by providing solutions for modular, large-scale, scalable policy composition and interaction.

2) Anonymity and end-user control:

There are many services that do not need to know the real identity of a user (e.g., a digital library could be accessed by a user that presents a certificate issued by a given association and stating the user's membership in the association). The access control system should allow full end-user control over digital identity to be used. In other words, access control system needs to operate even when interacting parties wish to remain anonymous or to disclose only specific attributes about themselves.

3) Client-side restrictions

In addition to traditional server-side access control rules, users should be able to specify restrictions about the usage of their information released to a third party and their activities that occur under the observation of third parties.

4) Context-aware restrictions

The privacy protection requirements may depend on the evaluation of some conditions (e.g., physical location of user, time-of-day at which activity occurs, etc). Therefore, an access control system should allow the specification of generic constraints not only on subjects and objects but also on contextual information.

5) Flexible and expressive access control rules

The access control rules should be able to express access restrictions based on the typical abstractions used by data/service providers, such as user categorizations and data objects categorizations as well as complex combinations that capture specific scenarios.

B. A privacy-aware access control policy

As it is vital to define access control policy for an eID system that addresses privacy related issues, the following basic elements should be part of such a policy.

1) Subject expression:

A subject expression identifies a set of subjects that satisfy specific properties. It is a set of rules that specify the entities on which access control to be enforced. For example, a subject expression can denote citizens of age 65 and above. For an eID system, subject expressions could be used to categorize citizens into different groups and access controls could be enforced based on the groups.

2) Object expression:

Similar to subjects, each expression identifies a set of objects that satisfy specific properties. This specifies the characterization of the processing to be done by the eID system on them. For example, an object expression can denote an operation to extract certain attributes from the selected object set. In an eID system, the data derived from the system for any given purpose should be controlled so that cross-correlation and information synthesis is prevented to ensure that personal identifiable information (PII) is not revealed.

3) Actions:

A policy makes distinctions about who can perform activities based on the action being performed (e.g., read, writes, and so on) using the eID system. Abstractions can also be defined on actions for specializing a particular action or to group them in sets. For the eID system different access levels should be defined with ability to perform different actions on the system and data. This will also include different levels of service users representing service providers and system users representing the various operating elements of an eID system.

4) Purposes:

Data access requests are made for a specific purpose or purposes. This represents how the data will be used by the recipient. For instance, the data may be used for analytical purposes or for identification purposes. In an eID system data access requests for identification purposes would be more common, but there will be instances for other types of requests, which should be granted only after considering privacy issues.

5) *Conditions:*

Conditions are the system-wide rules affecting how the eID service is operated. These rules may be due to particular laws of a country, due to international agreements, etc. For example, a condition may stipulate that the eID holder consent be obtained before PII is used for a particular purpose.

6) *Obligations:*

A privacy policy may also state that when a certain access is allowed, the parties involved must take some additional steps. An example is that all accesses against a certain type of data for a given purpose must be logged. Another obligation might be that PII must be deleted if its eID holder has not performed a specific action over a specified time period.

C. *Data protection policy*

In an eID system, the different stakeholders interact remotely. This requires the exchange of sensitive information and storage of data on locations that are not under the direct control of the data owner. Such data will have to be remotely accessed by eID system users and the data owners such as eID holders should have some degree of access to remotely manage their data. Therefore, it is essential to implement a joint management of data by the eID service provider, the eID holder and other eID service users. This requirement brings in the following challenges to the system.

1. The development of a powerful access control model that drive the enforcement of policies agreed between the different parties
2. The development of techniques for assessing the protection of data gathered by a party

With respect to the first issue, an approach for selectively encrypting data could be adopted so that users (or groups thereof) can decrypt only the data they are authorized to access [4].

This solution requires defining and maintaining, both at the client and server, additional information at the level of metadata needed to enforce selective access. Also, an approach for the implementation of access control based on a hierarchical structure, used for key derivation, reflecting the access control policy defined by the data owner could be used [5]. With respect to the second issue, different strategies for creating indexes that can be used by the data provider to select the data to be returned in response to a query could be adopted, together with quantitative measures to model inference exposure [6].

III. CRYPTOGRAPHIC TECHNIQUES/ SCHEMES

A. *Cryptography for privacy-enhancement*

The methods and techniques based on cryptography can support retaining and protecting a user's privacy in different contexts. These techniques range from simple encryption and digital signature schemes for secured private

transactions over the communication networks to complex multi party protocols for secure function evaluation. There are general privacy enhancing cryptographic schemes such as Group Signatures, Blind Signatures and Ring Signatures. The research presented in this paper is mainly focused on the first two signature schemes because group and blind signatures are closer to our problem domain of privacy protection. We are mainly concerned about improving the state of the art for so called Anonymous Credential Systems (also known as Pseudonym Systems), which are an essential mechanism for privacy enhanced data management in an eID system.

B. *Group signature scheme*

A Group signature scheme allows a member of a group to anonymously sign a message on behalf of the group. The concept was first introduced by David Chaum and Eugene van Heyst in 1991 [7]. For example, a group signature scheme could be used by an employee of a large company where it is sufficient for a verifier to know a message was signed by an employee, but not the particular employee who signed it. Another application is for keycard access to restricted areas where it is inappropriate to track individual employee's movements, but necessary to secure areas to only employees in the group.

An essential entity in a group signature scheme is the group manager, who is in charge of adding group members and has the ability to reveal the original signer in the event of disputes. In some systems, the responsibilities of adding members and revoking signature anonymity are separated and given to a membership manager and revocation manager respectively.

These group signatures are a "generalization" of the credential authentication schemes in which one person proves that he belongs to a certain group. The group signature scheme allows a group member to sign messages anonymously on behalf of the group and signatures can be verified with respect to a single public key of the group and do not reveal the identity of the signer [8].

Soundness and Completeness: Valid signatures by group members always verify correctly, and invalid signatures always fail verification.

C. *Blind signature scheme*

A Blind signature scheme is a form of digital signatures where content of a message is concealed (blinded) before it is signed. The resulting blind signature can be publicly verified against the original message like dealing with regular digital signatures [9]. Blind signature schemes allow a person to get a message signed by another party without revealing any information about the message to the other party. Blind signatures are used to provide unlinkability, which prevents the signer from linking the blinded version of the message and the un-blinded version of the message, if he is given the opportunity to sign both the messages in two different occasions.

In a practical scenario, a blind signature scheme is applied as a cryptographic protocol that involves two parties: a user Alice who wants to obtain a signature on her message m , and a signer Bob who is in possession of his secret signing key. At the end of the protocol Alice obtains a signature on m without Bob learning anything about the message.

D. Anonymous credential systems

Anonymous credential systems allow anonymous yet authenticated and accountable transactions between users and service providers. They provide a powerful mechanism to protect the privacy of the users mainly when engaging in online transactions.

Service providers may require authentication or accountability for users' actions to control access to their resources. For this purpose users need to prove their identity or at least prove the possession of a certificate. Such a certificate may contain a pseudonymous identifier of the user or only the required information needed to access a certain service [10]. However, with the continuous use of this type of certificates, different uses of a particular certificate can be linked. Eventually it will lead to identification of its owner through analysis of a series of transactions made with a specific certificate. Therefore, this is a major concern for privacy of the users.

An anonymous credential system (pseudonym system) can help eliminate linkability between transactions [11]. In such systems, different service providers and credential issuers know the users only by pseudonyms. A user can not be linked with his different pseudonyms. Yet, an organization can issue a credential to a pseudonym, and the corresponding user can prove possession of this credential to another organization (which knows him by a different pseudonym), without revealing anything more than the fact that the user owns such a credential.

Whenever the user needs to provide some personal information, then the system should ensure that no other information other than disclosed is revealed. For an example, if the user has to prove that he/she is of major age, then the system should not prompt to provide their date of birth or name. Moreover, the party who certifies that a user is of age and the party who verifies the statement should not be able to tell whether they communicated with the same user or with different one using any historical data. This can be achieved by the use of anonymous credential systems.

Some of the important properties of anonymous credential systems are as follows:

1. *Credential unforgeability*: Must prevent users from showing credentials that have not been issued
2. *Credential non-transferability*: Must prevent users from pooling their credentials

In an anonymous credential system except for the organization which issue and verify credentials of users there is another type of organization that is known as deanonymizing organization. These organizations have the

authority to reveal the details of a user's pseudonym or user's identity depending on the context. This type of anonymity revocation can be applied by the issuing organization to take measures when users misuse their credentials.

E. Database encryption

Database encryption is another method to protect sensitive data by providing data security and ensuring privacy requirements. Database encryption can be performed at different levels of granularity: relation level, attribute level, tuple level, or element level. Both relation level and attribute level imply that the communication to the end user would include the whole relation involved in a query. On the other hand encrypting at element level would require an excessive workload for data owner and clients in encrypting/decrypting data [4]. For balancing the client workload and query execution efficiency, database encryption at tuple level is more preferable.

IV. USER-SIDE IDENTITY DATA MANAGEMENT

In a backdrop of increasing awareness of the importance of identity management, many organizations are quick to declare that they properly secure user privacy. However, in most instances they fail to protect the anonymity and unlinkability aspects of the users. The privacy enhancing policy for the user-side identity data management in an eID System is a primary tool used to achieve anonymity and unlinkability. Identification of users can be categorized into two main types according to the physical functionality of an eID system: visible identification and electronic authentication. The electronic authentication of the cardholder is typically realized using asymmetric cryptography i.e. public key infrastructure (PKI) and allows creating digital signatures with non-repudiation capabilities and legal acceptance.

A. Electronic authentication

Asymmetric cryptography and PKI digital key pairs can be used for various purposes, such as encryption of messages, authentication while consuming services online, placing qualified electronic signatures with legal force, etc. However, it is argued, for security reasons that a key pair used to place an electronic signature with legal force should not also be used for authentication when accessing websites or for encryption purposes [12]. Authentication using an eID in an online environment may become more vulnerable in terms of data privacy management if careful separation of service objectives is not maintained. Accordingly, data privacy policy to be used must take into consideration security of stored data, data with a time-bound value as well as network communication security.

When users disclose their PII, their privacy decreases with respect to the service used. Often, it is not intuitively clear to a user how much the disclosure of certain PII affects their privacy. The goal of our research is to find

measurements which can help the user in estimating the current privacy status. This estimation of the privacy status is meant to help the user in deciding what to do in situations where the user is given different options for disclosure of PII.

B. User Control

Access control for local applications by using an eID would also help to protect their private data, which comes under the Personalization of Applications despite Privacy [13]. Determining which information is indeed needed depends on the particular application and on business and legal requirements.

The eID owners must be given tools to preserve their privacy in the user side. The eID system must also ensure that the communication is secure, anonymous (i.e., does not reveal potentially private information such as the user's IP-address or location to anyone), and correct (the transmitted information is received only by the intended recipient).

Users need assistance to manage their personal information. This is due in part to the volume of the data and the number of different transactions that a user participates in. More importantly, users must painstakingly avoid mistakes since mistakes are never forgotten in the on-line world. Thus, the user's interaction with the systems must be intuitive and easily understandable, with assistance for this management either through automation, if possible, or on-line and contextual help. Many enterprises are not aware of these risks and of the market share they might lose if they violate the trust of their customers. As a consequence enterprises publish privacy statements that promise fair information practices without strong implementations in place.

V. SERVICES-SIDE IDENTITY MANAGEMENT

In the context of services-side identity management, our research was focused on technologies and system solutions for privacy-enhanced data management for eID system. The main technical contributors for privacy control from the services-side are enhanced access control, private information retrieval and policy-based cryptography.

A. Private information retrieval

Private information retrieval (PIR) is a cryptographic technique used to protect privacy of requests to a database, which allows users to retrieve records while hiding the exact query. If a web server is using PIR, any observer, even a malicious administrator of the web server, will be unable to identify the data retrieved by the user [14].

Naive Approaches and their drawbacks:

- *Entire database download*: The entire database transfer (from the server to the client) solves the PIR problem theoretically. But this approach is impractical for real-life databases and applications due to high cost.

- *Anonymization techniques*: Sending queries anonymously to a server and anonymously receiving the answers is also possible. However, in this scheme, servers can collect general statistics (e.g., highest accessed record). Also most anonymization techniques depend on a trusted third party. The client has to trust the third party instead of the server.

B. Policy-based cryptography and applications

In open computing environments like the Internet, many transactions may occur between entities without pre-existing trust relationships. The concept of policy-based cryptography makes it possible to perform policy enforcement in large-scale open environments. According to the data minimization principle only strictly necessary information should be collected for a given purpose. A policy specifies the constraints which a specific action can be performed on certain information.

The mechanism of policy-based cryptography will perform policy enforcement while respecting the data minimization principle. Privacy-aware policy enforcement is enabled by policy-based encryption and policy-based signature [15].

Policy-based encryption allows encrypting data according to a policy, so that only entities satisfying the policy are able to decrypt and retrieve data. Also policy-based signature allows generate a digital signature on data according to a policy, so that only entities satisfying the policy are able to generate/verify a valid signature.

VI. CONCLUSIONS

The privacy of user and system data is of crucial importance for an eID system that meets user expectations and acceptance. In this paper, we presented an authorization model and accompanying basic security and control mechanisms that could be used to enforce privacy protection in an eID system. An eID system built on these models and mechanisms can use different policies for privacy-aware access control and data management.

This research studied the use of cryptographic mechanism such as group signing and database encryption to achieve stronger data level privacy and discussed the need to address both user-side and server-side of privacy protection.

In this research we have focused on techniques for privacy enhancements for an eID system. The identified techniques have been implemented in the open source software project titled *Project eID* (www.project-eid.org) developed by this research team.

REFERENCES

- [1] Alan F. Westin, *Privacy and Freedom*, The Bodley Head Ltd, New York, 1967
- [2] David Banisar, *Privacy & Human Rights*, EPIC, 2000
- [3] G.W. van Blarckom, J.J. Borking J.G.E. Olk, *Handbook of Privacy and Privacy-Enhancing Technologies: The case of Intelligent Software Agents*

- [4] Ernesto Damiani, Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. Metadata management in outsourced encrypted databases. In Proc. of the 2nd VLDB Workshop on Secure Data Management (SDM'05), Trondheim, Norway, September 2005.
- [5] Ernesto Damiani, Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, and Pierangela Samarati. Key management for multiuser encrypted databases. In Proc. of the International Workshop on Storage Security and Survivability, Fairfax, Virginia, USA, November 2005.
- [6] Ernesto Damiani, Sabrina De Capitani di Vimercati, Sara Foresti, Pierangela Samarati, and Marco Viviani. Measuring inference exposure in outsourced encrypted databases. In Proc. of the First Workshop on Quality of Protection, Milan, Italy, September 2005. (short paper).
- [7] David Chaum, Eugene van Hevst. Group Signatures. Kruislaan 413, 1098 SJ Amsterdam, The Netherlands : CWI Centre for Mathematics and Computer Science, 1998.
- [8] Jan Camenischy, Markus Michels. A Group Signature Scheme Based on an RSA Variant. 1998.
- [9] What is a Blind Signature Scheme. RSA Laboratories. [Online] [Cited: August 22, 2008.] [http://www.rsa.com/rsalabs/node.asp?id=2339\](http://www.rsa.com/rsalabs/node.asp?id=2339).
- [10] Herreweghen, Jan Camenisch and Els Van. Design and Implementation of the idemix Anonymous. s.l. : IBM Research, Zurich Research Laboratory.
- [11] A. Pfitzmann and M. Kohntopp. Anonymity, unobservability and pseudonymity: a proposal for terminology. In Designing Privacy Enhancing Technologies: Proceedings of the International Workshop on the Design Issues in Anonymity and Observability, LNCS 2009, pages 1-9. Springer-Verlag, 2000.
- [12] Dealing with Privacy Obligations in Enterprises, Marco Casassa Mont, Trusted Systems Laboratory, HP Laboratories Bristol, June 30, 2004
- [13] Privacy-Enhancing Access Control Enforcement, Yves Deswarte Matthieu Roy, September 13, 2006
- [14] Private Information Retrieval - An overview and current trends, Dmitri Asonov, Humboldt-Universitat zu Berlin
- [15] Policy-Based Cryptography and Applications, Walid Bagga , Refik Molva, Institut Eur'ecom, Corporate Communications