

Security and Privacy Preserving Data in e-Government Integration

Claudio Biancalana

Department of Computer Science and Automation,
RomaTre University.
Rome, Italy
biancalana@dia.uniroma3.it

Francesco Saverio Profiti

Department of Computer Science and Automation,
RomaTre University.
Rome, Italy
s.profiti@gmail.com

Abstract

In this paper, we present a security infrastructure design to ensure safety in the electronic government system: a combination of well-known security solutions, including Public Key Infrastructure, Shibboleth, Smart cards and Lightweight Directory Access Protocol. In this environment we give an overview in privacy preserving and security for Data Mining processes.

Keywords

Security, Integration, Single Sign On, Data Mining

INTRODUCTION

Member countries of the European Union are speeding into the digitalization of government services, with countries currently offering a surplus of interactive services which are increasing in availability and sophistication. International attempts to develop integrated customer oriented administrative services represent efforts to alleviate the problems of bureaucracy and improve the provision of administrative services. Since the launch of the European Strategy for the development of e-Government, with the "e-Europe 2002" initiative presented in March 2000 at the Lisbon European Council, a change of focus has occurred. The original target to supply services through the internet has evolved into the impact of e-Government programmes in delivering better services to their citizens, more efficient in an inclusive society" which emphasizes on the quality of the services provided and the extent to which online services are meeting user needs. Identified as a major aspect, is the safe access to services European Union wide by establishing secure systems for mutual recognition of national electronic identities for public administration websites and services (European Commission, 2006).

The necessity of an interoperable and scalable security and identity infrastructure has been identified by all implicated parties focusing on the effectiveness of solutions provided.

SECURITY AND ELECTRONIC GOVERNMENT

Electronic Government services are being rapidly deployed throughout Europe. Security is the main concern in this process, creating the need for an interoperable secure infrastructure that will meet all current and future needs. It is a necessity that such an infrastructure will provide a horizontal level of service for the entire system and must be accessible by all applications and sub-systems in the network.

Delivering electronic services will largely depend upon the trust and confidence of citizens. For this aim, means have to be developed to achieve the same quality and trustworthiness of public services as provided by the traditional way [10].

Regarding the level of systems design, some fundamental requirements, as far as security is concerned, have to be met:

- Identification of the sender of a digital message.
- Authenticity of a message and its verification.
- Non-repudiation of a message or a data-processing act.
- Avoiding risks related to the availability and reliability.
- Confidentiality of the existence and content of a message [10].

The best solution makes use of coexisting and complementary technologies which ensure safety throughout all interactions. Such a system provides assurances of its interoperability by using widely recognized standards and open source software. This evolutionary infrastructure design is based on a collaboration of existing cutting edge technologies in a unique manner. Public key infrastructure, Single sign On techniques and LDAP collaborate effectively guaranteeing efficient and secure communications and access to resources.

A Public Key Infrastructure (PKI) based on asymmetric keys and digital certificates, is the fundamental architecture to enable the use of public key cryptography in order to achieve strong authentication of involved entities and secure communication. PKI have reached a stage of relative maturity due to extensive research that has occurred in the area over the past two decades, becoming the necessary trust infrastructure for every e-business (ecommerce, e-banking, e-cryptography).

The main purpose of PKI is to bind a public key to an entity. The binding is performed by a certification authority (CA), which plays the role of a trusted third party. The user identity must be unique for each CA. The CA digitally signs a data structure, which contains the name of the entity and the corresponding public key besides other data.

Such a pervasive security infrastructure has many and varied benefits, such as cost savings, interoperability (inter

and intra enterprise) and consistency of a uniform solution [1].

A PKI smart card is a hardware-based cryptographic device for securely generating and storing private and public keys, digital certificates and performing cryptographic operations.

Implementing digital signatures in combination with advanced cryptographic smart cards minimizes user side complexity while maintaining reliability and security (Only an identity in possession of a smart card, a smart card reader and the Personal Identification Number (PIN) can use the smart card). Smart cards provide the means for performing secure communications with minimal human intervention. In addition smart cards are suitable for electronic identification schemes as they are engineered to be tamper proof.

The lightweight directory access protocol, or LDAP, is the Internet standard way of accessing directory services that conform to the X.500 data model. LDAP has become the predominant protocol in support of PKIs accessing directory services for certificates and certificate revocation lists (CRLs) and is often used by other (web) services for authentication. A directory is a set of objects with similar attributes organized in a logical and hierarchical manner. An LDAP directory tree often reflects various political, geographic, and/or organizational boundaries, depending on the model chosen. LDAP deployments today tends to use Domain name system (DNS) names for structuring the top-most levels of the hierarchy. The directory contains entries representing people, organizational units, printers, documents, groups of people or anything else which represents a given tree entry (or multiple entries).

Single Sign On (SSO) is a method of access control that enables a user to authenticate once and gain access to the resources of multiple independent software systems. Shibboleth is standards-based, open source middleware software which provides

Web Single Sign On (SSO) across or within organizational boundaries. It allows sites to make informed authorization decisions for individual access of protected online resources in a privacy preserving manner. Shibboleth is a Security Assertion Mark Up Language with a focus on federating research and educational communities.

Key concepts within Shibboleth include:

- **Federated Administration:** The origin campus (home to the browser user) provides attribute assertions about that user to the target site. A trust fabric exists between campuses, allowing each site to identify the other speaker, and assign a trust level. Origin sites are responsible for authenticating their users, but can use any reliable means to do this.

- **Access Control Based On Attributes:** Access control decisions are made using those assertions. The collection of assertions might include identity, but many situations will not require this (e.g. accessing a resource licensed for use by all active members of the campus community or accessing a resource available to students in a particular course).
- **Active Management of Privacy:** The origin site (and the browser user) controls what information is released to the target. A typical default is merely "member of community". Individuals can manage attribute release via a web-based user interface. Users are no longer at the mercy of the target's privacy policy.

A collaboration of independent technologies presented previously leads to an evolutionary horizontal infrastructure.

Introducing federations in e-government, in association with PKI and LDAP technology, will lead to efficient trust relationships between involved entities. A federation is a group of legal entities that share a set of agreed policies and rules for access to online resources (Uk Federation Information Centre, 2007, <http://www.ukfederation.org.uk/>). These policies enable the members to establish trust and shared understanding of language or terminology. A federation provides a structure and a legal framework that enables authentication and authorization across different organizations.

In general the underlying trust relationships' networks of the federation are based on Public Key Infrastructure (PKI) and certificates enable mutual authentication between involved entities. This is performed using SSL/TLS protocol and XML digital signatures using keys contained in X.509 certificates [11][4] obtained from e-school Certification Authorities. An opaque client certificate can contain information about the user's home institution and, optionally, the user's pseudonymous identity. Shibboleth technology relies on a third party to provide the information about a user, named attributes. Attributes are used to refer to the characteristics of a user and not the user straightforward: a set of attributes about a user is what is actually needed rather than a name with respect to giving the user access to a resource.

In the hypothesized architecture, this is performed by the LDAP repository which is also responsible for the association of user attributes.

Additionally LDAP contains a list of all valid certificates and revoked certificates. Digital signatures are used to secure all information in transit between the various sub-systems.

This infrastructure leverages a system of certificate distribution and a mechanism for associating these certificates with

known origin and target sites at each participating server. User side complexity is guaranteed to be minimum without any cutbacks on the overall security and reliability. The model presented in this paper offers the advantages of each single technology used and deals with their deficiencies through their combined implementation:

- Hybrid PKI hierarchical infrastructure delegates the trust to subordinate CAs permitting the creation of trust meshes, under a central CA, between independent organizations. Interoperability is simply addressed.
- PKI supports single sign on with the use of Shibboleth. Shibboleth coordinates with PKI to develop enhanced, complex free, authorization and authentication processes.
- The user becomes part of the designed system using Single Sign On (SSO) technology, that simplifies the access to multiple resources with only one “gain access procedure”. In practice this results in enhancing the security of the whole infrastructure, among other evident technical issues, because a sufficient level of usability is assured. Providing a security infrastructure is not enough, the user must also be able to make use of the security features. Otherwise, the designed service will fail due to the fact that users’ behavior is often the weakest link in a security chain.

The combination of the above mentioned techniques creates strong trust relationships between users and e-Government services, by implementing a “zero-knowledge” procedure of a very strong authorization. Zero-Knowledge is an interactive method for one entity to prove the possession of a secret without actually revealing it, resulting eventually in not revealing anything about the entity’s personal information. The combined techniques mitigate the problem of memorizing many passwords and reduce the vulnerability of using the same password to access many web services. It is essential to distinguish the authentication process from the authorization process. During the authentication process a user is required to navigate to his home site and authenticate himself. During this phase information is exchanged between the user and his home site only; with all information on the wire being encrypted. After the successful authentication of a user, according to the user attributes/credentials, permission to access resources is either granted or rejected. The process in which the user exchanges his attributes with the resource server is the authorization process during which no personal information is leaked and can only be performed after successful authentication.

User Authentication is performed only once when the user identifies himself inside the trust mesh.

Once authenticated inside the trust mesh, users are not required to re-authenticate themselves. When a user navigates to a resource store inside the trust mesh, the authorization process is executed. During this process the service provider requires from the users Identity Provider to present the users access credentials. The Identity provider, after successfully identifying the user and checking if he is previously authenticated, retrieves user credentials for the required resource. If user has not previously been authenticated, the authentication process is initialized. The Shibboleth Identity provider contains four primary components the Attribute Authority (AA), the Handle Service (HS), attribute sources, and the local sign-on system (SSO). Shibboleth interacts with the Ldap infrastructure to retrieve user credentials.

From the Identity Providers point of view, the first contact will be the redirection of a user to the handle service, which will then consult the SSO system to determine whether the user has already been authenticated. If not, then the browser user will be asked to authenticate, and then sent back to the SP URL with a handle bundled in an attribute assertion. Next, a request from the Service Provider's Attribute Requester (AR) will arrive at the AA which will include the previously mentioned handle. The AA then consults the ARP's for the directory entry corresponding to the handle, queries the directory for these attributes, and releases to the AR all attributes the requesting application is entitled to know about that user.

PRIVACY PRESERVING DATA MINING

In large intra-organizational environments, data are usually shared among a number of distributed databases, for security or practicality reasons, or due to the organizational structure of the business. Data can be partitioned either horizontally, where each database contains a subset of complete transactions ([6]; [5]), or vertically, where each database contains shares of each transaction. The role of a data warehouse is to collect and transform the dispersed data to an acceptable format, before they will be forwarded to the Data Mining (DM) subsystem. Such central repository raises privacy concerns, especially if it used in an inter-organizational setting where several entities, mutually untrusted, may desire to mine their private inputs, both securely and accurately. Alternatively, data mining can be performed locally, at each database (or intranet), and then the subresults be combined to extract knowledge, although this will most likely affect the quality of the output.

If a general discussion was to be made about protecting privacy in distributed databases, we would point to the literature for access control and audit policies, authorization and information flow control (e.g., multilevel and multilateral security strategies), security in the application layer (e.g., database views), and Operating Systems security among others. However in this paper we assume that appropriate security and access control exist in the intra-organizational setting, and we mainly focus on the inter-organizational setting where a set of mutually untrusted

entities wish to execute a miner on their private databases. As an alternative layer of protection, original data can be suitably altered (*e.g. randomized*) [2] or *anonymized* before given as an input to a miner, or queries in statistical databases may be. The problem with data perturbation is that in highly distributed environments, preventing the *inference* of unauthorized information by combining authorized information is not an easy problem [3]. Furthermore, in most perturbation techniques lies a *tradeoff* between protecting privacy of the individual records and at the same time establishing accuracy of the DM results [11].

At a high abstraction level, the problem of privacy preserving data mining between mutually untrusted parties can be reduced to the following problem for a two-party protocol: Each party owns some private data and both parties wish to execute a *function F* on the union of their data without sacrificing the privacy of their inputs [9].

In a DM environment, for example, the function *F* could be a classification function that outputs the class of a set of transactions with specific attributes, a function that identifies association rules in partitioned databases, or a function that outputs aggregate results over the union of two statistical databases.

In the above distributed computing scenario, an “ideal” protocol would require a trusted third party who would accept both inputs and announce the output. However, the goal of cryptography is to relax or even destroy the need for trusted parties.

Contrary to other strategies, crypto mechanisms usually do not pose dilemmas between the privacy of the inputs and the accuracy of the output. In the academic literature for privacy preserving data mining, following the line of work that begun with Yao [12], most theoretical results are based on the *Secure Multiparty Computation* (SMC) approach (*e.g.* [6]; [5]). SMC protocols are *interactive protocols*, run in a distributed network by a set of entities with private inputs, who wish to compute a function of their inputs in a privacy preserving manner.

We believe that research for privacy preserving DM could borrow knowledge from the vast body of literature on secure *e-auction* [8] and *e-voting* systems [7]. These systems are not strictly related to data mining but, they exemplify some of the difficulties of the multiparty case (this has been pointed out first by [9] but it only concerned e-auctions, while we extend it to include e-voting systems as well). Such systems also tend to balance well the efficiency and security criteria, in order to be implementable in medium to large scale environments. Furthermore, such systems fall within our distributed computing scenario and have similar architecture and security requirements, at least at our abstraction level.

In a sealed bid e-auction for example, the function *F*, represented by an auctioneer, receives several encrypted bids and declares the winning bid. In a secure auction, there is a need to protect the privacy of the losing bidders, while establishing accuracy of the auction outcome and verifiability for all participants. Or, in an Internet election, the function *F*, represented by an election authority, receives several encrypted votes and declares the winning candidate. Here the goal is to protect the privacy of the voters (*i.e.*, unlinkability between the identity of the voter and the vote that has been cast), while also establishing eligibility of the voters and verifiability for the election result.

During the last decade, a few cryptographic schemes for conducting online e-auctions and e-elections have been proposed in the literature.

Research has shown that it is possible to provide both privacy and accuracy assurances in a distributed computing scenario, where all participants may be mutually untrusted, without the presence of an unconditionally trusted third party.

CONCLUSIONS

Internationally numerous governments are becoming available online every day. As unattached efforts of addressing electronic government are implemented globally, the need for an interoperable horizontal security infrastructure is stressed.

The effective security infrastructure design presented in this paper is a solution which makes use of coexisting and complementary open source technologies and standards. Provides secure and effective communication supported by ease of use for the end user. Scalability and interoperability is an advantage of this design suitable to meet the needs of electronic government.

In this environment we studied the context of DM security; of course, further research is needed to choose and then adapt the specific cryptographic techniques to the DM environment, taking into account the kind of databases to work with, the kind of knowledge to be mined, as well as the kind of specific DM technique to be used.

REFERENCES

- [1] Carlisle Adams, Sharon Boeyen: UDDI and WSDL extensions for Web service: a security framework. XML Security 2002: 30-35
- [2] Agrawal, Ramakrishnan Srikant: Privacy-Preserving Data Mining. SIGMOD Conference 2000: 439-450
- [3] Domingo-Ferrer, Antoni Martínez-Ballesté, Francesc Sebé: MICROCAST: Smart Card Based (Micro)Pay-per-View for Multicast Services. CARDIS 2002: 125-134
- [4] Pho Duc Giang, Le Xuan Hung, Sungyoung Lee, Young-Koo Lee, Heejo Lee: A Flexible Trust-Based Access Control Mechanism for Security and Privacy Enhancement in Ubiquitous Systems. MUE 2007: 698-703
- [5] Murat Kantarcioglu, Chris Clifton: Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. DMKD 2002
- [6] Yehuda Lindell, Benny Pinkas: Privacy Preserving Data Mining. J. Cryptology 15(3): 177-206 (2002)
- [7] L. Mitrou, Dimitris Gritzalis, Sokratis K. Katsikas: Revisiting Legal and Regulatory Requirements for Secure E-Voting. SEC 2002: 469-480M. Naor and B. Pinkas, Computationally Secure Oblivious Transfer, Advances in Cryptology: Proceedings of Crypto 1999.
- [8] M. Naor and B. Pinkas, Computationally Secure Oblivious Transfer, Advances in Cryptology: Proceedings of Crypto 1999.
- [9] Benny Pinkas: Cryptographic Techniques for Privacy-Preserving Data Mining. SIGKDD Explorations 4(2): 12-19 (2002)
- [10] Roland Traummüller: Electronic Government, Second International Conference, EGOV 2003, Prague, Czech Republic, September 1-5, 2003, Proceedings Springer 2003.
- [11] Wei Wang, Guosun Zeng, Mingjun Sun, Huanan Gu, Quan Zhang: Factoid Mining Based Content Trust Model for Information Retrieval. PAKDD Workshops 2007: 492-499
- [12] Andrew Chi-Chih Yao: How to Generate and Exchange Secrets (Extended Abstract) FOCS 1986: 162-167