MIDDLEWARE PRIVACY PROTECTION IN DATABASES

Praveen Dangi
Assistant Professor
Department of Computer Science
Dronacharya College of Engineering,
Gurgaon (Haryana)
praveenchoudharyv09@gmail.com

Meenu Vijarania
Assistant Professor
Department of Computer Science
Dronacharya College of Engineering,
Gurgaon (Haryana)
meenuhans.83@gmail.com

Sneha Satija
Assistant Professor
Department of Computer Science
Amity University Haryana
snehasatija@gmail.com

ABSTRACT- In this paper, a solution has been developed for mitigating privacy violation in databases, as per legislative requirements. The research work developed a modified P3P to use it as a privacy enforcement language while in parallel, the peer research community developed standard privacy enforcement languages like E-P3P and EPAL. The privacy policy enforcement layer, presented in this paper, is based on EPAL rather than the modified P3P that was originally developed in the absence of an acceptable privacy enforcement language. It further presents a modified SQL to cater to the privacy requirements for managing privacy constrained records. The privacy access layer presented demonstrates the equivalence of ECA and EPAL and therefore proposes an ECA based privacy access layer. However, since EPAL and E-P3P have evolved as a widely acceptable standard for privacy enforcement language. The Privacy Violation Detection and Monitoring layer termed as PRIVDAM is required since it is hypothesized that just as systems are breached despite implementation of information security control, similarly, privacy violations can happen despite privacy middleware being implemented. Such violations can happen due to hacking or due to compromise in the access to the privacy constrained records due to human error like password sharing.

This also presents the findings of applying the solution in the health domain. The health domain was selected, as it is one of the most sensitive areas and has mature laws, such as HIPPA, governing the privacy of records in the medical domain. The overall effectiveness of the solution was found to be very good on the given data. The results showed that the solution provides excellent privacy protection, privacy violation detection with low false positives and acceptable privacy monitoring.

Keyword component: Middleware, EPAL policy, ECA, PRIVDAM Policy.

1. INTRODUCTION

One of the earliest definitions of Privacy as defined by Warren and Brandeis in 1980 has been “the right of an individual to be left alone and to be able to control the flow of information about him or herself”.

However, the area of research on privacy preserving techniques in databases has witnessed an explosive growth in the past few years and many sub areas of research in privacy enhancements technologies have cropped up. This has been fuelled by an exponential growth in centralized digital repositories of personally identifiable information that are highly susceptible to privacy violations and the privacy legislations that are being quickly adopted by governments the world over.

2. PRIVACY PRESERVING TECHNIQUES IN DATABASE

It provides an overview of a broad classification and description of the various techniques and methodologies that have been developed for preserving privacy in databases. There are two categories of privacy preserving techniques:

1) Privacy Policy Based Techniques.
2) PII Hiding Techniques.

The Privacy Policy Based Technique involves privacy preservation that is driven by the stated privacy policies. PII Hiding Techniques focus on the preservation of privacy by the masking of personally identifiable information.

(a) PRIVACY POLICY BASED TECHNIQUES

We define the Privacy Policy Based Techniques as those privacy preserving techniques that are driven by the pre-defined privacy policies. Privacy policies may be defined by the organization for the individuals or may be defined by the individuals themselves.

We have further classified research under privacy policy-based techniques into:

(a) Privacy specification languages.
(b) Privacy middleware.
(c) Privacy protection within databases.

Privacy Specification Languages help in defining the privacy policy whereas Privacy Middleware may use such languages to provide a middleware above the database to provide privacy protection.

![Layered representation of Privacy Policy based Techniques](image-url)
Privacy Specification Languages help in communicating the privacy policy in a machine-readable form. e.g. P3P 1.0 henceforth P3P (Platform for privacy preferences). P3P is a protocol designed to inform Web users of the data-collection practices of Web sites. It provides a way for a Web site to encode its data-collection.

Privacy Specification Languages also help in providing a robust framework for the enforcement of privacy in an organization. E.g. E-P3P (Platform for Enterprise Privacy Practices), EPAL (Enterprise Privacy Authorization Language) and DPAL (Declarative Privacy Authorization Language).

We define the Privacy Middleware as the software layer between the application software and the database that enforces the privacy policies, as well as enables privacy violation detection and privacy preserving access control. We have classified Privacy Middleware in to three sub areas:

(a) Privacy aware access control.
(b) Policy Enforcement.
(c) Privacy violation and detection.

Privacy Protection within Databases modifies databases kernels for enforcing the privacy policies. However this solution is not technology neutral as the privacy middleware is welded into the specific database and hence it does not provide a solution for legacy databases.

In this a large amount of work has been done for protecting privacy within the database. Databases include responsibility for the privacy of data. Introducing the concept of Hippocratic databases, they have proposed architecture of a database system that accurately protects private information. Hippocratic databases use components of secure databases and introduce privacy control within the database itself.

The Hippocratic database uses Privacy Metadata which is defined:

a. External recipients
b. Retention period.
c. Authorized users.

(b). PII Hiding Techniques

PII Hiding Techniques are used in contexts where there are no specific articulated privacy policies. Hence, these techniques generically try to mask all personally identifiable information. The areas of research identified as part of PII Hiding Techniques are:-

- Anonymization techniques.
- Inference Control techniques.
- Privacy Preserving Query Processing techniques.
- Privacy preserving Data Mining techniques.
- Privacy preserving data Integration techniques.

As per legislations, the users have right to define their privacy preferences. We have publishing languages like web publishing languages.

(1) WEB PUBLISHING LANGUAGES:-

For publishing privacy policies over the web in a machine readable format, the W3C developed the platform for privacy preferences languages(P3P).

The Platform for Privacy Preferences Project (P3P) enables Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats).

The P3P specification defines

A standard schema for data a Web site may wish to collect, known as the "P3P base data schema"

- DATA A standard set of uses, recipients, data categories, and other privacy disclosures.
- An XML format for expressing a privacy policy.
- A means of associating privacy policies with Web pages or sites, and cookies.
- A mechanism for transporting P3P policies over HTTP.

P3P policies are describes in XML format as a sequence of STATEMENT elements that have the sub elements.

DATA-GROUP, RECEIPIENT, PURPOSE, CONSEQUENCE AND RETENTION.

The sub elements are defined as follows:

- DATA-GROUP

It provides the list of individual data items that are collected for the stated purposes in the statement.

- RECEIPIENT

It describes the intended users of the collected information. Multiple recipients can be specified in one statement.

- PURPOSE

It describes the purposes for which information is collected. Multiple purposes can be listed in a STATEMENT if all of them have the same values for RECEIPIENT, RETENTION and DATA-GROUP; otherwise they are specified in different STATEMENT elements.

- CONSEQUENCE:
It describes the intended purpose for collecting information in human-readable text.

- **RETENTION:**

It defines the duration for which the collected information can be retained.

Although P3P is a standard for publishing privacy policies over the web, a P3P policy may be semantically inconsistent and interpreted and represented differently by different user agents.

(1) **USER’S PRIVACY PREFERENCE LANGUAGE**

For specifying the privacy preferences of the users, formal languages APPEL have been developed. The languages help the users in programmatically checking their privacy preferences against a P3P policy, which allows them to decide whether to release their data to the website. APPEL is an XML based language that was developed by the creators of P3P for enabling the automatic matching for privacy preferences against P3P policies. In APPEL, a user’s privacy preferences are expressed in a RULESET, which consist of a sequence of RULE’S. A rule consist of two parts

1) Rule behavior (rule head)

This specifies the action to be taken if the rule is executed.

2) Rule body

It provides one or more patterns that are matched against a P3P policy. These patterns are combined using logical operators (i.e. AND, OR, NON-AND, NON-OR)

3. **ENFORCEMENT LANGUAGES**

The Enforcement languages are

(a). E-P3P (Platform for Enterprise Privacy practices).

(b). EPAL (Enterprise Privacy Authorization language.)

(a). E-P3P (Platform for Enterprise Privacy practices).

The Platform for Enterprise Privacy Practices (EP3P) language has been designed to enforce privacy statements expressed in P3P. E-P3P policy manages the personal information internally.

The rules in EP3P policies allow or deny actions on data categories by user –categories for certain purposes under certain conditions while mandating certain obligations. A rule belonging to the rule set consists of

(a) Precedence, (b) Data category, (c) Purpose, (d) data user, (e) Ruling, (f) Action, (g) Obligations, (h) Conditions.

(b). **EPAL (Enterprise privacy Authorization Language)**

**EPAL** is a formal language proposed by IBM, which enables enterprises to automate and enforce privacy policies across IT applications and systems. It is an interoperable language for exchanging privacy policies in a structured format between applications. The elements of EPAL rule are as follows:

1). Vocabulary information:

It describe the vocabulary of the rule. The schema for the vocabulary –information is defined by vocabulary id, short description, issuer name, issuer email, issuer address, issuer country, start date, last date, last modified.

2). User category:

It represent a category of individuals that can access data. The user category table has the fields user category, short description, long description.

3). Data category:

It is used in privacy policies. Data category is defined by the data field, table name, short description, long description.

4). Purpose:

It represents the purpose for which a rule authorizes or denies access. It includes purpose, short description, long description.

5). Action:

A privacy policy specifies which actions are allowed and which are denied under particular circumstances. The action is defined by fields action, class name, short description, long description.

4. **EPAL Policy (Enterprise Privacy Authorization Language)**

Enterprises collect personal data while promising fair information practices to their customers. The Enterprise Privacy Authorization Language (EPAL) enables an enterprise to formalize the exact privacy policy that shall be enforced within an enterprise.

IBM’s new XML-based programming language called Enterprise Privacy Authorization Language (EPAL) allows developers to build policy enforcement directly into enterprise applications.

EPAL builds on current privacy specifications, namely the Platform for Privacy Preferences (P3P) that provide privacy controls for information passed between business applications and consumers with browsers.

An EPAL policy is essentially a set of privacy rules. A rule is a statement that includes a ruling, a data user, an action, a data category, and a purpose. A rule may also contain conditions and obligations. Each rule contains a precedence level.

A EPAL policy document consists of three main sections:

1. **Policy Information**: This is used to identify the policy. It consists of information such as Issuer, Version Number,
Start Date, End Date, Replacement Policy Name, Replacement Policy Version.

2. Definitions: This defines all of the possible components that can be used in the following rules. Here is where Data Users, Data Categories, Purposes, Actions, Context Models, Conditions and Obligations are defined.

3. ALLOW or DENY: Rules to define whether Data Users are ALLOWED or DENYED Perform Action on Data Category for Purpose under Conditions.

5. ENTERPRISE PRIVACY POLICY ENFORCER

Enterprise Privacy Policy Enforcer (EPPE) can be loosely coupled with the underlying Privacy Policy Enforcement Layer.

Fig. shows the layer architecture of the Enterprise Privacy Policy Enforcer that includes Policy Definition Interface, Authentication and Authorization Manager, Policy Validator, Policy Translator and a Policy Execution Engine.

A user with low IT skills can define the policy through the Policy Definition Interface, which is user friendly GUI. Policy Translator converts the policy specified by the user through the Policy Definition Interface to a format that can be stored directly in the database and can be executed by the Policy Execution Engine.

The Transaction Notification Layer in the layered Architecture captures the events and sends them to the Policy Execution Engine. The Policy Execution Engine is responsible for detecting the situation and then reacting against the situation that have occurred. The authentication and authorization manager manages the access control of the EPPE framework. The policy validator in EPPE ensures that all conflicts in policy definition are flagged out for the user to resolve. This layer is a critical first level filter for controlling policy conflicts.

6. PRIVACY VIOLATION DETECTION AND MONITORING

Privacy violation detection and monitoring system to minimize the privacy violations that can occur even after proper privacy violation mechanisms have been deployed.

The area of privacy enhancement technologies has seen tremendous growth in the last couple of years. This is mainly due to the enactment of privacy legislations and the widespread use of the internet and its inherent weakness in the protection of the privacy of individuals as well as organizations.

The need of a privacy violation prevention mechanism becomes evident whenever organizations deal with Personal Identifiable Information. With the increase in the amounts of personal data being collected, stored and processed in information systems, the threat of violation of individual privacy and consequent commercial damage to large enterprises is on the increase. Moreover, control over personal information has also decreased as individuals are unaware of which systems stored their information and what all has been stored. Sensitive data such as detailed transaction summaries including social security number, shipping and billing addresses, e-mail id and credit card details are being put to risk on a routine basis.

7. Capability Specification

The credit card company maintains a record of customer’s personal information, credit history, credit report, account balances, payment history, etc. According to its privacy policy, its customer service department is allowed to access the customer’s credit report. This is captured in the modified EPAL Policy.

We define a transfer group, which specifies the capability of a user group to transfer his capability to a set of user groups.

In order to specify the transfer group, we add a new element to the EPAL vocabulary, <transfer-group>.

IMPLEMENTATION

For the implementation of the Privacy Policy Enforcement Layer, a modular approach is shown:

1) Authentication Layer

The Authentication layer works on the top of the capability and privacy check layer. The main function of this module is the authentication of user’s capability. While submitting a query, user first logs in this module. After login, the capability certificate of a user is fetched from the capability certificate database.

Since this capability certificate is encrypted with a secure co-processor’s key, it needs to be decrypted.

2) Capability – Privacy compliance check layer

After Authentication of the capability certificates, the second layer performs the important function of capability checking. In our architecture, capability certificate and privacy certificates are specified using an XML language EPAL. This layer serially checks capability of the data requestor to access the privacy constrained data, by matching the capability certificate with the privacy certificate.

3) Encryption – Decryption Layer

This layer ensures that even a database administrator cannot violate a user’s privacy policy. This module encrypts the data while storing in database and decrypts the data while retrieving from database.
8. PRIVACY VIOLATION DETECTION V/S INTRUSION DETECTION SYSTEM

Intrusion Detection System have been extensively studied and they have been classified according to their granularity of data processing, detection method, security, manageability, adaptability and network infrastructure requirements.

Protecting the infrastructure from external unauthorized access is a security issue whereas the protection of individuals information from intentional or unintentional abuse of authorized access is a privacy issue.

However, security and privacy are also co-related. In most cases, privacy requires security ,but sometimes security functions may hinder or actually be a cause of privacy violations e.g. intrusion detection method and logging. This is called the security- privacy paradox.

PRIVDAM borrows some of the learning from Intrusion Detection Systems.

9. INTRUSION DETECTION METHOD

ID’s prepare for and deal with attacks by collecting information from a variety of system and network sources, then analyzing the symptoms of security problems.

ID’s serve three essential security functions:
1. Monitor
2. Detect and
3. Respond to unauthorized activity

An IDS is a process of identifying and responding to malicious activity targeted at computing and networking resources.

CLASSIFICATION OF IDS

ID is concerned with the detection of hostile actions.

This network security tool uses either two main functions:
1. Anomaly Detection
2. Signature Detection

Anomaly Detection which explores issues in Intrusion Detection.

It discriminates between anomaly or attack patterns and Intrusion detection System.

1. Anomaly Detection or Normal behavior Patterns
Normal behavior patterns are useful in predicting both user and system behavior.
In this behavior, everything that does not match the stored profile is considered action. Hence, these systems are characterized by very high detection efficiency.

2. Signature Detection or Misbehavior Signature
The misbehavior signatures fall into two categories:
   a. Attack signatures
      They describe action patterns that may pose a security threat.
   b. Selected text string
      Signatures to match text string which look for suspicious action.

The ID in terms of behavior are
1. Active ID’S
2. Passive ID’S

1. Active ID’S:
They detect and respond to attacks, attempt to patch software holes before getting hacked.
2. Passive ID’S:
Those that simply generated alerts and log network packets.

10. Future Work

From the classification of privacy enhancing technologies, it is evident that some areas of research that are not based on privacy policies need to be based on privacy policy based privacy preservation. There is no known privacy preservation data mining work that is based on privacy policies. it is essential to develop languages that can specify the privacy preserving data mining policies and develops techniques that enforce privacy in data mining that are based on the privacy policies.

This thesis presents the mapping of SQL queries for privacy, further work needs to done to map complex queries for privacy. Complex queries include SQL queries with sub queries, set operations, grouping and aggregates.

In future we also propose some other algorithm in PRIVDAM. The fully developed PRIVDAM system can go a long way if some good visualizations techniques are also provided.

Extending PRIVDAM for real time privacy violation detection prove to be a significant contribution in this area. Considering that information once leaked into the wrong hands it is impossible to be undone, we plan to extend the current PRIVDAM methodology to react in real time.

References
[2] Dynamic Privacy Management: a Plugin Service for the Middleware in Pervasive Computing by Dan Hong Department of Computer science from Google
[6] Protecting Privacy of Health Information using EPAL based privacy by J.Bhattacharyya, from IEEE.