RDB-XML Security Model based on Masked Relations

Lwin Mar Thin, Nang Saing Moon Kham
University of Computer Studies (Yangon)
lwinmarthin85@gmail.com, moonkhamucsy@gmail.com

Abstract

Relational databases get more and more employed in order to store the contents of a web site. At the same time, XML is a fast emerging technology, which is the standard of data exchange over the web and due to the new requirements emerging from several applications there is a great need for the integration between XML and Database technology. Moreover, Security is becoming one of the most urgent challenges in database research and industry, and the challenge is intensifying due to the enormous popularity of web-based applications. There are a lot of researches about security for XML databases or relational databases, but the researches do not consider the fact that integrated RDB-XML system to publish XML data reside in relational database. This paper explores security techniques for proposed security model in which secure data stored in relational database, access controls for XML data are also specified in relational model, secure key scheme and secure data transfer as a XML document.

1. Introduction

Despite the excitement surrounding XML, it is important to note that most operational business data, even for new web-based applications, continues to be stored in relational database systems. This is unlikely to change in the foreseeable future because of the reliability, scalability and performance associated with relational database systems. Consequently, if XML is to fulfill its potential, some mechanism is needed to publish relational data in the form of XML documents.

Using Relational database to save and manage XML data can bring many advantages for different users. Nowadays relational database is the mainstream database, changing XML data into relational data can not only reserve the specialty of easy to express and independent of platform in the aspect of XML at data application, moreover may fully using the mature data management service of relational database(effective memory, highly effective inquiry, concurrent control, data restore and so on), so making up the obvious shortcoming of XML technology in the aspect of searches, modification, achieving the goal of effectively manages and protects the XML data[4].

Figure 1 shows the overview architecture of the proposed system. The particularity of our integration system is that it employs XML as the user interface format, while all data flows inside the query processor consist of relational tuples and access control rules are stored in relational database. Queries are posed using an SQL query language and the results are formatted as XML documents, making the underlying relational engine transparent to the user. Moreover, this paper shows how to support effectively RDB-XML security models by utilizing security support of relational security models.

- In relational database, sensitive data are encrypted.
- Access control rules are defined by security administrators and stored into relational databases.
- Security check and query evaluation are done by relational database and only valid answers are returned to users in the XML document.

- Encryption key are embedded to this XML document by using steganographic techniques.

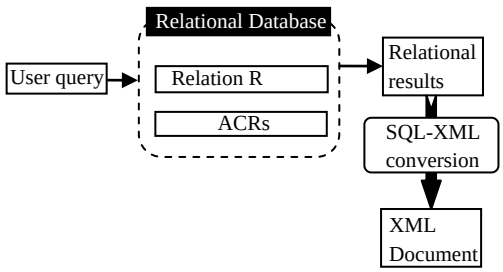

Figure1. Overview Architecture

The rest of the paper is organized as follows: section 2 reviews the related work on relational to XML publishing model, security related research such as encryption techniques, access control rules and steganography. Our motivation is given in Section 3. Section 4 describes about security model and explains its component. Section 5 discusses implementation and experimental results with illustrative examples. Section 6 concludes the paper.

2. Related Work

Research projects such as SilkRoute [5] and XPERANTO [4] have proposed techniques for efficiently publishing relational data as XML. Commercial database products such as SQL Server, Oracle, and DB2 also provide support for publishing relational data as XML. However, this research didn't consider the security over this system.

In the security aspect, a lot of work has already done in the field of database security before, but as we improve the security the attackers also increase the penetration of the attack. ZhaoYong Xia [7] has described a lot of technique for encrypting the data which are used from long time to secure the data. Some of the classical encryption techniques, like DES, triple DES, RSA, AES, are very power full encryption of the data. Xia also described some basic encryption techniques which help to encrypt the data for keeping it safely. These algorithm works in such a way that it balance the cipher text frequency, gives non-weighted code along with no long cycles and no repeated phenomenon.

Access controls are also very important part of the security; they provide security to the data. Lot of work has already been done in the field of access control and different access model have proposed before, which have different policies and can be used in different conditions. Different models proposed for the access control are Mandatory Access Control (MAC), Discretionary Access Control (DAC), Role Based Access Control (RBAC), and Rules Based Access Control. In Rules Based Access Control, access is allowed or denied to resource objects based on a set of rules defined by a system administrator.

Information hiding is field of information security, and it can be applied in XML documents which constitutes towards the field of steganography. Most research of steganography was using cover media such as images, videos and sound. Some of the methods proposed to solve the problem, such as by line shifting, words shifting, up to whitespaces manipulation into the cover text [3]. In this system used the technology of stenography for transfer encryption key into the output XML document as a cover text.

This paper gives the more secure way of securing the data in which more than one security techniques are used one behind the other which provides more security to the RDB-XML system.

3. Motivation

Database systems are well-known for consistent storage, retrieval and manipulation of data. At the same time, XML is generally accepted as data description language for both web-based information systems and electronic data interchange between different organizations. Since database systems form the backbone of essentially any information system, the integration of XML and database systems is a must. Moreover, security is becoming one of the

most urgent challenges in database research and industry. Daily we hear that the cyber crime are increasing it is all because the number of attacker are increasing day by day and the techniques to do an attack are also enhanced. So to secure our data we also have to enhance our security techniques. Attacker can attack the data at two different positions. This means that attacker attack directly to the database and gets the whole database data. On the other hand, the attacker can also attacks the data on the network and gets only the data which is on the network at the particular time. To secure the data it is necessary to secure database as well as the network. The proposed system satisfies both the storage and transmission.

4. Security Model

In most databases used on the web, data is stored in tables in the form they are loaded, mostly in plaintext, which doesn't satisfy high level security and privacy requirements. To ensure data confidentiality, privacy and integrity we should take care of two security issues: secure data storage and secure data transmission.

In this security model the encryption and the access control are used one behind the other, this will give more security to the database. Here only the person who can use the data have the proper access control and know the proper key which will increase the security to the database.

When an attacker break the one kind of security then other security will stop the attacker and if it breaks the second security then the first security will stop the attacker. It will take time for the attacker to break both the security levels. The attack time in the single level security is small in which it is difficult for the administrator to detect the attack but the attacker have to attack at two level when this model is used and the attack time will be large enough for the administrator to detect the attack.

This section describes the concepts of encryption, its access control model relational database and key security scheme. The proposed security model, whose basic architecture is shown in Figure 2. The purpose of such model is to implement encrypted storage and secure transmission which satisfying data confidentiality, privacy and integrity. In this model, when the users query to relations, the system produces masked relations which they can obtain information allowed by access rules. This result relation is need to convert XML document. If the results included the sensitive data, the system embedded the encryption key into this XML document by using steganography. So, we don't need to send this encryption key via other channel.

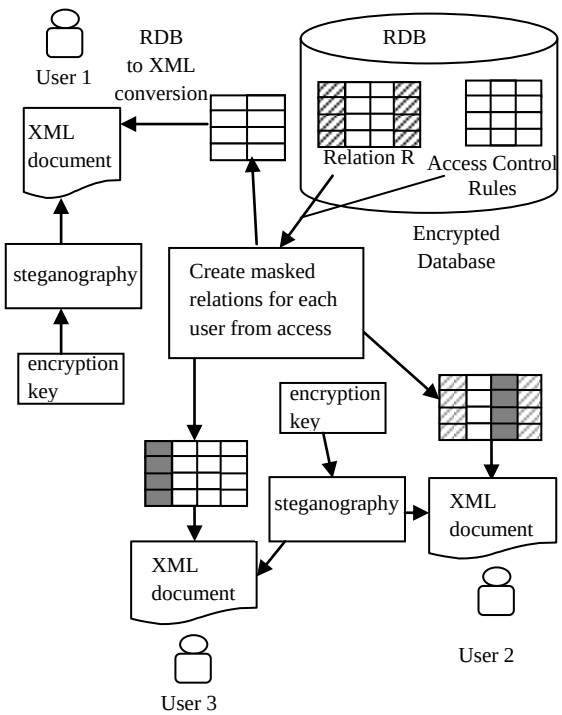

Figure2: Suggested Security Model

4.1. Database Encryption

There are two well-known classes of encryption algorithms: conventional and public-key. Although both can be used to provide data confidentiality, their goals and performance differ widely. Conventional, (also known as symmetric-key) encryption algorithms require the encryptor and decryptor to share the same key. Such algorithms can achieve high bulk

encryption speeds, as high as 100-s of Mbits/sec. Symmetric encryption provides excellent security at a lower hardware cost due to its simplicity. The algorithms are fairly simple and require less computing power to perform data encryption. On the other hand, symmetric encryption keys must be kept secure - you would need to make sure each person i.e, the need to securely deliver the same key to all necessary entities.

Public-key cryptography solves the problem of key distribution by allowing an entity to create its own public/private key-pair. Anyone with the knowledge of an entity's public key can encrypt data for this entity, while only someone in possession of the corresponding private key can decrypt the respective data. While elegant and useful, public key cryptography typically suffers from slow encryption speeds (up to 3orders of magnitude slower than convention algorithms) as well as secure public key distribution and revocation issues. But a suitable infrastructure for exchanging public keys is also required; this is needed to verify the identity of the sender, confirming that the message is not a forgery. In addition, asymmetric encryption produces longer encrypted messages and key generate problem might appear for each user. So, it doesn't suitable for encryption data in database. This asymmetric algorithm is mostly used for authentication.

Due to their security and performance, we use symmetric-key algorithm for encryption of sensitive data stored in relational database by adding secure key scheme.

4.2. Access Control Rules

In this section, access control rules are defined by administrator. For example, consider the relations of EHR (Electronic Health Record) database. The patient relations consists of five attributes p_id, name, address, room, illness, department in which p-id is the primary key and all data are encrypted by using public key encryption algorithm except p_id field. The physician relation consists of eight attributes: p_id, phy_id, name, office, phone no., address, salary, and department. The nurse relation consists of six attributes: n_id, p_id, name, address, salary and department. Now consider the following access policies.

- A patient is allowed to see all of his or her own information from the relations
- Other user can see insensitive data of all information
- Information's about patient having in the corresponding department can accessible by physician and nurse.
- Information about medical staff with exception of their salary and home address is publicly accessible.
- Information about the name and home address, salary, phone. no of medical staff and all information of patients must be accessible to the members of Administrative.

Table1.examples of access authorization

Subject	Object	Table	Action	Sign
public	salary, phone no., address	physician	read	-
public	salary, address	nurse	read	-
public	name, departmen t	physician, nurse	read	+
public	*	patient	read	-
admin	*	patient	read, write	+
admin	salary, phone no., address	physician, nurse	read, write	+
nurse, physician (same department)	*	patient	read	+
patient (p_id)	p_id,*	patient	read	+

The sign attribute '+' or '-' represent positive or negative role of the subject.

4.3. RDB to XML conversion

The output which we get when we fire SQL query is in terms of rows and columns. But client

should get the reply in terms of XML. So we need to convert to the desired result into XML file. This translation is not focused in this paper and we use the VQT algorithm [1][2] already developed by us. The VQT algorithm is superiority to conventional translation algorithm because the VQT algorithm considers not only syntactic aspect but explicit/implicit semantic aspect.

4.4. Proposed Secure Key Scheme

Key management refers to the way cryptographic keys are generated and managed throughout their life. Because cryptography is based on keys that encrypt and decrypt data, the database protection solution is only as good as the protection of the keys to transfer. In this section explores the secure key scheme by using steganographic techniques for embedding encryption key with the help of stego key into output XML document from previous step. The users don't require knowing the knowledge of about encryption algorithms. They need to know only stego key. In this system, the stego key may be the identity of the user. They entered the stego key, the actual output displayed instantly without showing the encryption key. So, they think that stego key may be the real encryption key. In this way, the weakness of symmetric encryption key problem is satisfied. Figure 3 shows the processing steps of stegano key module.

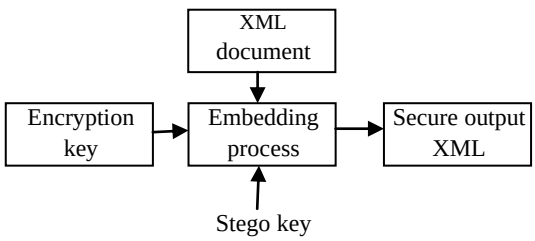

Figure 3: Stegano key module

4.4.1 The proposed Steganographic techniques

These techniques are used to hide the key from generated encryption algorithm.

White space close tag Method

- Data is embedded by inserting a white space before close tag ">", or no white space.
- By inserting or deleting spaces, we can embed and data preserving all meaning of original text.

Line Break insertion Method

- Data is embedded by exploiting Line Break.
- In this method, first tag is taken and a line break is inserted after reading the closing character ">".

Word space Method

- Data is embedded after inserting word space in values.
- At reverse process extra spaces are removed to get back the original file.

5. Implementation and Experimental Results

In this section, we describe the implementation of the rule-based access control for encrypted RDB-XML document. We have implemented on an intel core i5 machine running Windows 7, with 2GB memory and 320GB hard disk. As an experimental platform, we used Microsoft SQL server 2008 and we also used ASP.Net in visual studio 2010 as experimental software program and run under .Net framework.

The basic flow of the RDB-XML security model is as follows. When a user queries to the system, it receives the query and user_id, and extracts tables form the query. Then, the system makes mask relations based on user_id and access rules. Finally, the system returns the results to the user.

Consider the following query Q1 and Q2 as examples of the proposed system.

Q1: select * from patient;

When the inputs are q and p_id 1, the system produces the results in Table 2(a). The results

masked the all patient information except p_id 1. Table 2(b) shows the results of phy_id 3. The department of this physician is cardiology. So, the system masks other department's patient information and all patient information within same department is produced as a results.
Following encrypted relational database table's data convert into output XML file shown in Figure 4 and 5.

Table2. (a) p_id:1

p_id	name	address	room	illness	department
1	Kafkel lk56okalfd;	Urweipi iapjp46asoif	aloilll	Affioaf4af	djkah;fkoe n24 nm

Figure4. Output XML document for p_id 1

p_id	name	address	Room	illness	department
1	Kafkel lk56okalfd;	Urweipi iapjp46asoif	Aloilll	Affioaf4af	djkah;fkoe n24 nm
3	Joie; ioe;rJfa; Diervv34	Fu355 ia;i343hii;	Ieoi[o	Fii 343 oifa	djkah;fkoe n24 nm
7	Hfaj jlakf iei459 k fj;a	Df;kl keok kafkf	Djhe 334fi`	Ertkjd; k34k	djkah;fkoe n24 nm

Table2. (b) phy_id: 3

Figure5. Output XML document for phy_id 3

Q2: select name from patient
where illness = "heparin";

In this example, the query depended on the predicate illness= "heparin" which data are encrypted in relational table. So, we need to encrypt the data "heparin" for search directly with this encrypted value in patient relation to get a desired result. Table 3 shows the query results of admin who has authority to see all patient information. So, the system no needs to mask the result.

Table3. Admin

name
Wapoei Ifjuap; Ioe
Ldjlaf jf;af Kjf;ae
tkdfhk;34 Fkja;fk'a
bi333w Kja;34; Dki

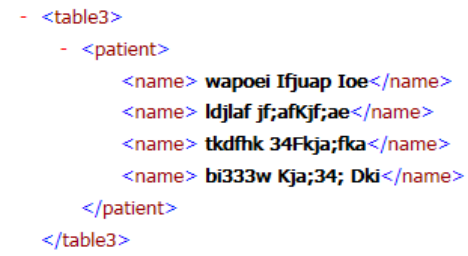

Figure6. Output XML document for Admin

To get the original data record, the user needs to decrypt this output XML documents with key.

6. Conclusion

XML is rapidly emerging as the dominant standard for exchanging data on the World Wide Web, making the ability to publish data as XML increasingly important. We have suggested and implemented the access control model for encrypted RDB-XML considering data levels. This system suggested in this paper can have the following contributions.

Practicality: the RDB-XML system can support more practical access control processing by using relational database.

Performance: especially in query processing, when a user queries to relations, the system produces mask data which is prohibited to access. So, we do not need to publish all of relational data as a XML document.

Security: the encryption and the access control are used one behind the other; this will give more security to the database and network level. Moreover, secure key scheme are also added to this system.

7. Discussion

Security and performance is tradeoff whenever security needed applications. Strong security application can lessen performance due to overhead of using security techniques .The proposed security model is considered efficient because it provides maximum security to the databases and network, whilst the added time cost for encryption, decryption is very minimal. To reduce the time spent on encryption and decryption, the scheme divides the data sensitive and non-sensitive data. This makes the time cost for their encryption and decryption to have less significant on the overall performance of the scheme. Moreover, this system defines access control rules over data levels which is not allowed or permitted. So, we don't need to load all data to publish. The only downside of the scheme is that if query predicate values are encrypted in the database, we need to firstly encrypt this value to search in relational database.

References

[1] J. Kim et al., Formal Verification and Quantitative Evaluation of QP-T Algorithm, Dynamics of Continuous, Discrete and Impulsive Systems (DCDIS) Series B, Vol. 3, pp. 1369-1373, June 2007.

[2] J. Kim et al., VQT: Value Cardinality and Query Pattern-based R-Schema to XML Schema Translation with Implicit Referential Integrity, Journal of Zhejiang University-Science A (JZUS-A), Vol. 9, No. 10, November 2008.

[3] L.Y.Por and B.Delina, "Information Hiding: A New Approach in Text Steganography", In Proceedings of the 7th International Conference on Applied Computer & Applied Computational Science (ACACOS'o8), Hangzhou, China, 2008

[4] M. Carey, K. Jerry, S. Jayavel, S. Eugene, and S. Subramanian, "XPERANTO: Middleware for publishing object-relational data as XML documents," in *Proceedings of International Conference on Very Large Data Bases*, 2000, pp. 646-648.

[5] M. Fernandez, W. Tan, and D. Suciu, "SilkRoute: Trading between relations and XML," *Computer Networks*, Vol. 33, 2000, pp. 723-745.

[6] Z. Ao-Ying, X. Zheng-Chuan, G. Zhi-Mao,etat., Adaptation of XML Storage Schema in VXMLR[J]. Chinese Journal of Computes, 2004, Vol.27 No.4: 433-441.

[7] ZhaoYong-Xia Information Engineering School Wuhan University of Sciences and technology Zhongnan Branch, Wuhan ,China," The Technology of Database Encryption" 2010 Second International Conference on MultiMedia and Information Technology.