CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The eXtensible Markup Language (XML) is a universal format for structured documents and data over the web. The Hypertext Markup Language (HTML) format describes only the format of the data, and therefore, programs cannot understand the structure of such data. On the other hand, XML provides flexibility in the representation of the structured documents and data.

The World Wide Web Consortium (W3C) introduced XML to complement HTML, for data exchange over the web. Bray et al (2008) explained the most recent version of the XML language recommendation. XML is a powerful tool, meant to deal with the current structure of document information processing. It is a cross-platform Internet environment technology, which is dependent on the content. XML documents (marked-up documents) are self-describing. Therefore, it is possible for programs to interpret the meaning of the data, which are received, filter the document based on its content, and restructure it to suit the application's needs. XML documents provide a platform independent means to describe data, and hence, they can transport data from one platform to another.

XML is a simplified subset of the Standard Generalized Markup Language (SGML), the international standard meta-language for text markup
systems. XML is designed to improve the functionality of the web, by providing more flexible and adaptable information identification. It is called extensible because it allows the user to create his/her own tags that describe the data, unlike HTML that has pre-defined tags to format the data on the web. Although XML takes more space than the binary data, XML is extremely simple and easy to master and use. These features make XML the first choice for the exchange of information between the Simple Object Access Protocol (SOAP) and web services.

Since XML is structured data, it is suitable for a structured data transmission network. Therefore, in recent years, XML applications have been growing. Security has always been of importance to ensure data protection, transactions' integrity, and maintain information privacy and confidentiality. One of the most challenging problems in managing large, distributed, and heterogeneous networked systems, is the ability to specify and enforce security policies between parties, and ensure access to services and resources. The infrastructure over the Internet has turned out to be difficult, due to system heterogeneity and conflicting security requirements.

1.2 SECURITY FOR XML DATA

Nowadays, the main requirement of Internet-wide security standards, is applying the security to the content created, using XML. XML has been adopted widely for a great variety of applications and types of content. XML is used to apply access control on the structure and content of a document. Moreover, XML is the basis of web services protocols that rely on different XML-based languages, such as SOAP, Web Service Definition Language (WSDL), and Universal Description Discovery and Integration (UDDI). Therefore, it is important to enforce the security of XML documents, to realize the exchange of XML documents in anonymous and untrustworthy
environments, and to ensure confidentiality, integrity, authenticity, and non-repudiation of XML documents (Ardagna et al. 2007).

The existing security technologies provide a set of security technologies used in XML security. However, the formats used to satisfy the security requirements are not compatible with the XML security applications. One reason is that, the existing security technologies use binary formats that require specific tools for interpretation and use. Another reason is that, these security technologies are not designed for XML. The XML security system addresses these issues, by reusing the concepts, algorithms, and core technologies of legacy security systems, while introducing changes necessary to support an extensible integration with XML. The XML security standards are as follows (Hirsch 2002):

1. XML Digital Signature for integrity and signing solutions (Bartel et al. 2013).
2. XML Encryption for confidentiality (Imamura et al. 2013).
3. XML Key Management Specification (XKMS) for public key registration, location, and validation (Hallam-Baker & Mysore 2005).
4. Security Assertion Markup Language (SAML) for conveying authentication, authorization and attribute assertions (Ragouzis et al. 2006).
5. EXtensible Access Control Markup Language (XACML) for describing policies and defining access control rules (Rissanen 2013).
Compared to traditional security technologies, there are four major advantages related to XML security.

1. The ability to selectively encrypt and sign portions of a message.
2. The ability to protect data integrity without encrypting it.
3. The ability to construct overlapping digital signatures using different keys.
4. The ability to digitally sign and encrypt data.

The encryption and signature specifications proposed by W3C specifying the format for encrypted XML documents, are important to XML security. However, they cannot allow the programmer to specify how to encrypt and sign the XML documents. Existing XML security specifications have some problems, which need to be improved, such as the efficiency of the authentication process, context-referral integrity for XML data in XML signature, multi-signature generation, and encrypted XML data query.

Moreover, although XACML is an integral policy description language, the structure of the existing XACML policy is very complex and hard. Hence, it is necessary for the users to understand XACML well, in order to write all the securing policy specifications. On the other hand, the Query languages of a RDBMS are easy and simple to use by all the users. Moreover, SQL-like query languages overcome the difficulties of XACML, by storing the XACML policies and rules in relations. Therefore, it is easy for the users to use and understand the XACML policies and rules. Moreover, storing the data and rules in tables provides more flexibility. Hence, it is necessary to provide the table based storage features of web databases.
Most of the researches are aimed towards representation and querying, authorization, and access control policies, for safeguarding the XML information system stored at the server. Therefore, it is necessary to provide a new algorithm for providing security for temporal XML based databases, not only for the effective representation of the temporal XML document, but also for providing facilities for encryption and decryption.

### 1.3 OBJECTIVES OF THE RESEARCH WORK

The main objectives of this research work are as follows:

1. To propose new algorithms for overcoming the complex structure of the XACML policies and rules by secure mapping of the XACML policies and rules into temporal relations.

2. To develop a secure temporal database based on XML.

3. To propose a new XML based storage structure for databases.

4. To provide the capability for decrypting the encrypted XML documents partially.

### 1.4 THESIS CONTRIBUTIONS

This research work aims to overcome the complex structure of the XACML policies and rules by secure mapping of them into temporal relations. Moreover, it aims to provide the capability for decrypting the encrypted XML documents partially. To fulfil these objectives, the thesis provides the following contributions:
1. Proposes new algorithms for mapping and storing the XACML policies and rules in the form of rules in temporal relations.

2. Proposes a new time-stamp based algorithm, for decrypting the encrypted XML documents partially.

The first contribution is proposing new algorithms for mapping the XACML policies and rules into relational rules, and storing them in the form of rules in temporal relations, to ease the access control to the XML documents. This proposed work relieves the users from the effort of learning and understanding the XACML policies and rules; hence, it saves the users' time and effort. It controls the access to the XML documents stored in either native or relational databases, using the XACML policies. Finally, it applies the constraints of rules, and obligations and provides the response to an access request effectively.

The second contribution of this research work is proposing a new time-stamp based algorithm, for decrypting the encrypted XML documents partially. This proposed algorithm utilizes a new temporal model for XML representation, based on a combination of valid time and transaction time, through the tuple time stamping to decrypt parts of the encrypted XML documents. Moreover, one attribute of the key components of the encrypted elements is selected to distinguish among them. The time-stamp and the selected attribute guide the receiver to decrypt the required parts of the encrypted XML documents, instead of decrypting all the parts of them. The proposed algorithm not only reduces the decryption time, but also protects the security of the data.

The new temporal model is proposed not only for the effective representation of the temporal XML documents, but also for providing
facilities for encryption and decryption; i.e., the receiver uses the time-stamp and the selected attribute to retrieve the parts that must be decrypted, to contribute to the query results on the encrypted XML documents. Moreover, this model uses the valid time and transaction time for time-stamping, and hence, can maintain the history data efficiently. This model can be used to develop secure temporal applications.

1.5 OUTLINE OF THE THESIS

The outline of this thesis is as follows:

1. **Chapter 2** presents a survey of the recent research that described and implemented the XML security standards and access control to the XML documents.

2. **Chapter 3** shows the background information about the XML signature and encryption.

3. **Chapter 4** depicts the standard XACML and the Role-Based Access Control (RBAC) profile of XACML.

4. **Chapter 5** explains the new algorithms for mapping the XACML policies and rules into a RDBMS.

5. **Chapter 6** discusses the time-stamp based algorithm, for decrypting the encrypted XML documents partially.

6. **Chapter 7** gives the conclusions of this work and suggests the future work.