Chapter IV

Storage and Access Control Issues for XML Documents

George Pallis, Aristotle University of Thessaloniki, Greece
Konstantina Stoupa, Aristotle University of Thessaloniki, Greece
Athena Vakali, Aristotle University of Thessaloniki, Greece

ABSTRACT

XML documents management is becoming an area of great research value and interest since XML has become a popular standard for data communication and knowledge exchange over the Internet. Therefore, new issues have emerged in terms of storage and access control policies for XML documents. Concerning the storage issues, the majority of proposals rely on the usage of typical database management systems (DBMSs), whereas XML documents can also be stored in other storage environments (such as file systems and LDAP directories). It is important to consider storage and access control together since these issues are essential in implementations for XML documents management. Moreover, the chapter focuses on the recent access control models which guarantee the security of the XML-based data, which are located in a variety of
storage topologies. This chapter’s goal is to survey and classify existing approaches for XML documents storage and access control, and, at the same time, highlight the main differences between them. The most popular XML database software tools are outlined in terms of their storage and access control policies.

INTRODUCTION

Internet is currently the core media for data and knowledge exchange. XML (eXtensible Markup Language)\(^1\), a subset of SGML (Standard Generalized Markup Language), is introduced by the World Wide Web Consortium (W3C) to complement and enhance HTML (Hypertext Markup Language) in electronic data representation and exchange on the Web. XML is becoming widespread and is a text-based markup language (like HTML), but it supports a richer set of features. The main advantage of using XML is that an XML document (differently from an HTML document) can be written once and visualized in a variety of ways. Therefore, XML is currently the most popular standardization effort in web documents representation, and is rapidly becoming a standard for data representation and exchange over the Internet. As a result, large amounts of XML documents are being generated, and their efficient management has become a major necessity. Researchers in both industry and academia have focused on efficiently storing, manipulating and retrieving XML documents.

The main XML-related research issues refer to the XML data accessing, storing, querying and exchanging. Indeed, even if XML lends its power to its ease-of-use and extensibility, it is this structure of XML that results in a controversial fact. From one point of view, this structure characterizes XML as an ideal building block on high-speed applications, whereas, from another point of view, it is this structure that makes XML unsuitable for usage under pre-existing data management environments. Most implementations rely on the usage of typical database management systems (DBMSs), whereas others are based on specific systems (providing ad-hoc functionalities). Moreover, since XML can be used over various application platforms, different management approaches have to be devised, depending on the type of the considered XML documents (structured vs. unstructured), the platform type (DBMS vs. file-based systems), and their main usage. Whatever is the chosen solution, a crucial point in efficiently managing XML documents is devising efficient storage and accessing control techniques. Among data management issues, storage and