Efficient Identification of Structural Relationships for XML Queries using Secure Labeling Schemes

S. Sankari, Anna University, Chennai, India
S. Bose, Anna University, Chennai, India

ABSTRACT

XML emerged as a de-facto standard for data representation and information exchange over the World Wide Web. By utilizing document object model (DOM), XML document can be viewed as XML DOM tree. Nodes of an XML tree are labeled to uniquely identify every node by following a labeling scheme. This paper proposes a method to efficiently identify the two structural relationships namely document order (DO) and sibling relationship that exist between the XML nodes using two secure labeling schemes specifically enhanced Dewey coding (EDC) and secure Dewey coding (SDC). These structural relationships influence the performance of XML queries so they need to be identified in efficient time. This paper implements the method to identify DO and sibling relationship using EDC and SDC labels for various real-time XML documents. Experiment results show the identification of DO and sibling relationship using SDC labels performs better than EDC labels for processing XML queries.

KEYWORDS

1. INTRODUCTION

XML (World Wide Web Consortium [W3C], 2008) is the widely used open standard for content representation and dissemination over the Internet. XML is the mostly followed data format to exchange data or information between incompatible systems. XML is very famous for its flexible and adaptable structure. This leads to the user to define their own structure based on the requirement. XML document is extensively used in various real-world application areas (Ashish, & Maluf, 2009; Maluf, Bell, Ashish, Knight, & Tran, 2005; Singh, 2007) like e-commerce, social networks, research communities, healthcare, enterprises, government and private organizations, etc., and the file size ranges from KB (KiloBytes) to EB (ExaBytes). XML big data is a recent trending research area. XML document plays a predominant role in publish/subscribe system (Datta, Gradinariu, Raynal, & Simon, 2003; Kundu, & Bertino, 2008) that efficiently disseminates the selective XML content to different subscribed users. Publish/subscribe system basically follows service oriented architecture (SOA). SOA is an architecture that provide services and mostly adapted in large scale enterprises (Alwadain, Fielt, Korthaus, & Rosemann, 2013; Beydoun, Xu, & Sugumaran, 2013). SOA can be implemented using web services.

DOI: 10.4018/IJIIT.2016100104

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Publish/subscribe system (Sankari, & Bose, 2014) involves three entities namely producer, consumer and publisher. Producer is considered as the owner of the XML document. Producer disseminates the selective content of an XML document to the subscribed consumers through the publisher. To assure the confidentiality and integrity of the XML document content to be disseminated, producer has to securely label, encode and encrypt the XML document. Consumers who are authorized for an XML document subscribes to the producer. Subscribed consumers receive secure labels for their accessible XML content and credentials for authentication and decryption from the producer. Producer disseminates the securely labeled, encoded and encrypted selective XML content to the publisher. A third-party usually called publisher or message broker receives the encrypted XML content with the subscribed consumers details from the producer. Publisher disseminates the encrypted the XML content to the respective subscribed consumers based on their request. Subscribed consumer receives the encrypted XML content and decrypts the XML content.

Producer needs to label the XML document to uniquely identify the XML content. To accomplish this, an XML document can be viewed as an XML tree by following a standard called document object model (DOM). DOM (W3C, 2015) allows to visualize the XML document as XML tree by exploiting the hierarchical structural relationship existing in the XML document. Elements, attributes, content, etc., present in an XML document are usually represented as nodes in the XML tree. An XML labeling scheme is followed to uniquely label every node of an XML tree. Hence, an XML label uniquely identifies every node in the XML tree. The basic requirements of an XML labeling scheme are minimum label size, efficient labeling time and every label has to preserve the structural relationships existing among the nodes of an XML tree. Document order (DO), sibling, parent-child (PC), ancestor-descendant (AD) and lowest common ancestor (LCA) are the structural relationships that normally exist between the nodes of an XML tree. XML label acts as a key to distinguish the content of an XML document. Producer sends the XML labels that are additional information apart from the actual XML content to the subscribed consumers. From these labels, consumer can deduce structural information of an XML document that the producer has. Therefore, any consumer can become an internal attacker and can exploit this information leakage to perform any security attack. Hence, producer prefers a secure XML labeling scheme that retains basic features of a labeling scheme and prevents information leak by assuring secure content dissemination. Enhanced Dewey coding (Sankari, & Bose, 2013) and secure Dewey coding (Sankari, & Bose, 2014) are the two recent secure XML labeling schemes proposed for publish/subscribe system.

XML queries are generally expressed in XPath (W3C, 2014a) and XQuery (W3C, 2014b) format. Document order and sibling are the essential structural relationships that play a vital part in XPath and XQuery processing. Document order (W3C, 2014c) represents the order of nodes in an XML tree based on the order that actually exists in the original XML document. Identification of document order in an ordered XML document from the secure labels is very important for answering XML queries. Identifying document order in efficient time is imperative to improve the performance of XML queries. Sibling relationship (W3C, 2014c) denotes the set of nodes that belongs to same parent node in an XML tree. All the child nodes of a parent node show a sibling relationship among themselves. The nodes exhibiting sibling relationship with other nodes implies that the sibling nodes belongs to same level which is considered as the main notable feature of sibling nodes. Identification of sibling relationship also plays an important role in answering XML queries that involves sibling relationship. Therefore, the time required to identify sibling relationship influences the performance of XML queries. Hence, the secure XML labeling scheme that supports identification of document order and sibling relationship in efficient time is preferred. The objectives of this paper include:

1. Efficient identification of document order and sibling relationship among the XML tree nodes using EDC and SDC labels;
2. Implementation and evaluation of experimental results of various real-time XML dataset for identifying document order and sibling relationship using EDC and SDC labels.
A Fuzzy-Based Recommender System for Electronic Products Selection using Users’ Requirements and Other Users’ Opinion
[www.igi-global.com/article/a-fuzzy-based-recommender-system-for-electronic-products-selection-using-users-requirements-and-other-users-opinion/126200?camid=4v1a](www.igi-global.com/article/a-fuzzy-based-recommender-system-for-electronic-products-selection-using-users-requirements-and-other-users-opinion/126200?camid=4v1a)

Rethinking “Oil Nationalism”: The Case of Anglo Iranian Oil Company (AIOC)
[www.igi-global.com/article/rethinking-oil-nationalism/142499?camid=4v1a](www.igi-global.com/article/rethinking-oil-nationalism/142499?camid=4v1a)