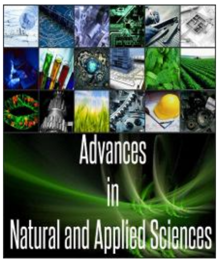
Advances in Natural and Applied Sciences, 9(9) Special 2015, Pages: 37-40

AENSI Journals

Advances in Natural and Applied Sciences

ISSN:1995-0772 EISSN: 1998-1090
Journal home page: www.aensiweb.com/ANAS



Measuring Effective Performance of Secure Labeling Scheme for XML Content Distribution

[1]S. Sankari, [2]S. Bose and [3]Lino Abraham Varghese

[1]Department of Computer Science and Engineering, College of Engineering, Guindy (CEG), Anna University, Chennai – 600 025 Tamil Nadu India
[2]Department of Computer Science and Engineering, College of Engineering, Guindy (CEG), Anna University, Chennai – 600 025 Tamil Nadu. India.
[3]Department of Computer Science and Engineering, College of Engineering, Guindy (CEG), Anna University, Chennai – 600 025 Tamil Nadu. India.

ARTICLE INFO

Article history:
Received on Feb 27th 2015
Accepted on March 20th 2015
Available online 15th June 2015

Keywords:
Publish/Subscribe
Secure labeling scheme
XML content distribution
XML label

ABSTRACT

Distribution of XML content has gained much attention in publish/subscribe model because of its efficiency, security and selectiveness. In publish/subscribe model, producer label the XML document using secure labeling scheme and then distributes the selective content to subscribed consumers through publisher. Several secure labeling schemes to encode the XML documents based on Document Object Model (DOM) have been proposed. In this paper, we consider the two secure labeling schemes namely, Enhanced Dewey Coding (EDC), and Secure Dewey Coding (SDC) for identifying the structural relationships Parent-Child (PC) and Ancestor-Descendant (AD) in XML nodes. We implement EDC and SDC secure labeling scheme and evaluate the performance in identifying the PC and AD relationship for various XML documents. Experiment results measures and validates the time required to identify the PC and AD structural relationships between XML nodes from EDC and SDC labels. Our effort provide performance analysis of efficient XML secure labeling schemes EDC and SDC that can be effective for both producer and consumer in identifying the PC and AD relationship.

To Cite This Article: S. Sankari, S. Bose, and Lino Abraham Varghese., Measuring Effective Performance of Secure Labeling Scheme for XML Content Distribution. **Adv. in Nat. Appl. Sci.,** *9(9): 37-40, 2015*

INTRODUCTION

XML (W3C XML) is becoming a standard for data exchange over the Internet. XML documents are widely used because of its relaxed structure. Content distribution of XML document has received more acceptance in publish/subscribe model. In a publish/subscribe model, producer is the owner of XML documents (Sankari, S and S, Bose, 2013)(Sankari, S and S, Bose., 2014). The XML document contents are securely labeled, encoded, encrypted and distributed by the producer to various consumers through the publisher. A consumer subscribes to a producer and receives accessible XML node labels with their credentials. The publisher is a third-party which stores and forwards the encrypted XML content to the intended consumer by validating their credentials.

XML contents are represented using Document Object Model (DOM), a de facto standard (W3C DOM). In XML DOM, each XML document is represented as a XML tree where each node represents an XML element. Each node must be labeled and the label acts as a structural identifier to find the content of an XML element that is to be distributed to the consumer. The label must be able to describe the relationship between them and also the document order which is more important. But from the label, consumer should not be able to infer additional information about the XML document which becomes a threat from the consumer (insider). So, the label should be secure with less memory size and labeling time.

In this paper, we present an implementation of EDC and SDC secure labeling schemes to estimate the efficiency in identifying the structural relationship like PC and AD between XML nodes. Using the standard XML dataset, we validate EDC and SDC secure labeling scheme based on the computation time in identifying PC and AD relationship. The results show the effectiveness of EDC and SDC labeling scheme in PC and AD relationship identification.

Corresponding Author: S. Sankari, Department of Computer and Engineering, College of Engineering, Guindy (CEG), Anna University, Chennai – 600 025, Tamil Nadu, India.
E-mail: sankarisivan@yahoo.co.in

1. *Background and Related Works:*

In this section, we will discuss the need for XML secure labeling schemes then the different labeling methods in the related works.

a) Need for XML Secure Labeling Scheme:

Selective contents of XML document are distributed to the consumer based on their subscription. To select certain content from the XML document, Document Object Model (DOM) is followed. In DOM, the whole XML document is considered as an XML tree with root element as root node and the subsequent elements, attributes, content, etc., are considered as node by preserving their structural relationship.

Labeling the nodes of the XML tree allows to uniquely identify the content of an XML document based on the label. Producer uses labeling scheme to label the XML document and identifies the content as labels relevant to a consumer. Producer provides the identified set of labels to the appropriate consumer during subscription. After content distribution through publisher, consumer needs to construct a new XML document with the received content. To create a new XML document, the relationship between the contents must be identified from the labels related to the content. Thus, labeling scheme plays a major role in XML content distribution. Effectiveness in relationship identification of a labeling scheme can be measured by its execution time.

b) Importance of PC and AD relationship:

In XML tree, a node say A is said to be parent of node B, if and only if A exists in previous level to node B through directly connected edge. Otherwise, node B is child to node A, if node B exists in next level of node A connected directly by an edge. Similarly, for example, node A is an ancestor of node B, if and only if node A is reachable by traversing repeatedly to preceding level through parent of node B. Likewise, node B is descendant of node A, when B is reachable by traversing repeatedly to next level through child of node A. Nodes in AD relationship, shares the same path. PC and AD relationship answers twig pattern queries (Li, C and T.W. Ling, 2010). The "/" and "//" in query denotes PC and AD relationship respectively. Also, PC and AD relationship is helpful in informational retrieval from XML documents. Thus, the PC and AD relationship can be identified easily from the EDC and SDC labels itself.

c) Related Works:

Different labeling schemes are used to identify each XML element uniquely. In (Kundu, A and E, Bertino, 2008), the structural properties of the XML DOM are utilized to address the data security issues. The structural identifier called label specified for each node in an XML document is additional information to content. Both (Kundu, A and E, Bertino, 2006) and (Kundu, A and E, Bertino, 2008) uses Post Order Numbering (PON) based on post order traversal (Dietz, P.F, 1982) that requires more memory which is the major drawback. A well-known numbering scheme for trees called Dewey coding is discussed in (Gou, G and R, Chirkova, 2007) and (Tatarinov, I, *et al*., 2002).

The major drawbacks of existing methods are structural information leakage is not prevented and requires more memory and labeling time for secure XML labeling. EDC (Sankari, S and S, Bose, 2013) and SDC (Sankari, S and S, Bose, 2014) are also based on Dewey coding but requires minimum memory space and labeling time. It prevents structural information leak and is used to identify the structural relationship in XML document.

2. *XML Secure Labeling Schemes:*

In this section, we will discuss methodology used in EDC and SDC secure labeling schemes.

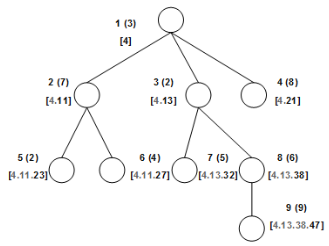
a) Enhanced Dewey Coding (EDC):

EDC (Sankari, S and S, Bose, 2013) is based on level ordering. Level Order Number (LON) is the number uniquely assigned to each node in increasing order while traversing the XML DOM tree in level order. Enhanced Level Order Number (ELON) is based on the general notion of LON with random number added to it. EDC is calculated based on ELON. Fig. 1 shows EDC labeled XML Document with calculated EDC in square bracket for each node.

Let A be a root node, B and C be non-root nodes, assume B is the parent of C. EDC for root and other nodes can be calculated by using equation (1a) and (1b) respectively where "." represents concatenation.

$$EDC_A = ELON_A \quad (1a)$$

$$EDC_C = EDC_B \bullet ELON_C \quad (1b)$$



n – Level Order Number (LON)
(r) – Random Number
[p.e] – Enhanced Dewey Coding (EDC)
p – EDC of parent node
e – Enhanced Level Order Number (ELON) of a node

Fig. 1: Enhanced Dewey Coded (EDC) XML Document.

b) Secure Dewey Coding (SDC):

SDC (Sankari, S and S, Bose, 2014) follows sibling order in each level of nodes. Sibling Order

Number (SON) is a unique number assigned to each child node that ranges from 1 to the total number of child nodes of a parent node while traversing them in sibling order. Secure SON (SSON) is generated by adding a random number to SON. SDC is computed by using SSON. Fig. 2 shows SDC labeled XML document. Let A be a root node, B and C be non-root nodes, assume B is the parent of C. SDC for a root node and non-root nodes are represented in equation (2a) and equation (2b) respectively where "." represents concatenation.

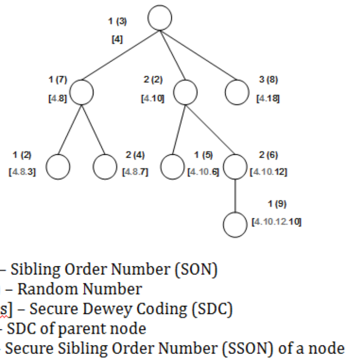


Fig. 2: Secure Dewey Coded (SDC) XML Document.

$$SDC_A = SSON_A \quad (2a)$$

$$SDC_C = SDC_B \bullet SSON_C \quad (2b)$$

c) Relationship Computation:

Structural relationships like Parent-Child (PC) and Ancestor-Descendant (AD) can be computed from secure labeling schemes. The method of computing PC and AD from EDC and SDC labels are identical because both of the labels follows Dewey format.

The following explains how the relationships PC and AD can be calculated from secure labels EDC and SDC.

1) Parent-Child (PC): Node A is parent of node B if only if parent label of node B equals label of node A.
Node A : a1.a2…am
Node B : b1.b2…bn
Node A is parent of node B.
Then, a1 = b1, a2 = b2, … am = bm and m = n-1.

2) Ancestor-Descendant (AD): If label of node A is the prefix of label of node B then node A is an ancestor of node B.
Node A : a1.a2…am
Node B : b1.b2…bn
Node A is an ancestor of node B.
Then, a1 = b1, a2 = b2, … am = bm and m < n.

For example, from Fig. 1 consider two EDC labels 4.11 and 4.11.27. Parent of node 4.11.27 is node 4.11. Also, node 4.13 is an ancestor for node 4.13.38.47. Similarly, from Fig. 2 consider the SDC label 4.8 and 4.8.7. The parent of node 4.8.7 is node 4.8. Likewise, consider two nodes 4.10 and 4.10.12.10. For node 4.10.12.10 the ancestor is 4.10.

3. *Experiments and Evaluation of Results:*

In this section, we present the experimental setup with the XML dataset, the execution results of secure labeling schemes EDC and SDC in identifying PC and AD relationships and the evaluation of results based on computation time.

Table I: XML test dataset .

XML Dataset	File Size (Bytes)	Total no. of nodes
SigmodRecord	467 K	11526
XMark1	1.12 M	17132
XMark2	3.40 M	50266
uwm	2 M	66729

a) Experimental Setup and XML Dataset:

EDC and SDC labeling schemes are implemented using Java in 1.83 GHz Core2 Duo CPU with 3 GB of RAM. For experiments, the labeling schemes EDC and SDC uses XML datasets from (XML Data Repository), (XMark) and (Schmidt, A., *et al*., 2002) which contains various sizes of XML documents. Table 1 shows the XML dataset details like file size and total number of nodes.

b) Evaluation of EDC and SDC results:

From the secure labeling schemes EDC and SDC labels, structural relationships like PC and AD can be computed. The performance of EDC and SDC labeling schemes in identifying PC and AD structural relationship are measured and are evaluated based on their execution time. Fig. 3 shows the PC and AD relationship computation time for the XML documents SigmodRecord, XMark1, XMark2 and uwm. From the graph, it is clear that SDC performs better than EDC in PC and AD structural relationship identification. Thus, SDC secure labeling scheme can be preferred over EDC for XML document labeling.

c) Impact of Node Count in PC and AD Relationship:

Fig. 4 shows the line graph for PC and AD computation time for node count in increasing order with different XML documents. From the graph, we can deduce that for lesser values of node count, the difference in PC and AD computation time from EDC and SDC labels is ignorable. Also, the PC and AD computation time between EDC and SDC labels increases as the total number of nodes increases. Therefore, for PC and AD computation time, SDC labeling scheme is better than EDC when the amount of node is very high.

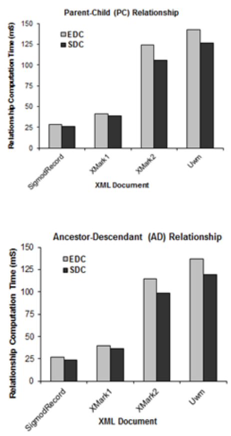


Fig. 3: PC and AD Relationship Computation using EDC and SDC Labels.

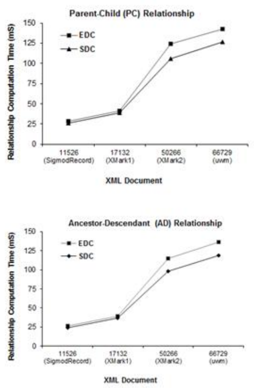


Fig. 4: Impact of Node Count in PC and AD Relationship Computation Time.

4. Conclusion:

In this paper, we carried out a detailed evaluation of two secure labeling schemes called EDC and SDC used for XML content distribution in publish/subscribe model. We explored the basic definitions and the PC and AD relationship computation related to EDC and SDC labels. We implemented EDC and SDC labeling scheme using XML test dataset and the results are measured based on the execution time in identifying the PC and AD relationships. The results shows that SDC secure labeling scheme is fast in computing the relationships like PC and AD compared to the EDC secure labeling scheme and can be used for various XML content distribution applications.

REFERENCES

Dietz, P.F., 1982. Maintaining order in a linked list. Proceedings of the fourteenth annual ACM symposium on Theory of computing, 122-127.

Gou, G. and R. Chirkova, 2007. Efficiently Querying Large XML Data Repositories: A Survey. IEEE Transactions on Knowledge and Data Engineering, 19(10): 1381-1403.

Kundu, A. and E. Bertino, 2006. Secure Dissemination of XML Content Using Structure-based Routing. Proceedings of the tenth IEEE International Enterprise Distributed Object Computing Conference, 153-164.

Kundu, A. and E. Bertino., 2008. A New Model for Secure Dissemination of XML Content. IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, 38(3): 292-301.

Li, C. and T.W. Ling, 2010. "Advanced Applications and Structures in XML Processing: Label Streams, Semantics Utilization and Data Query Technologies", Information Science Reference - Imprint of: IGI Publishing Hershey, PA.

Sankari, S. and S. Bose, 2013. Efficient Encoding of XML Document for Secure Dissemination. Proceedings of the fifth International Conference on Advanced Computing, 565-570.

Sankari, S. and S. Bose, 2014. Secure XML Labeling for Efficient XML Content Dissemination. Proceedings of the sixth International Conference on Advanced Computing (to be published).

Schmidt, A., F. Waas, M. Kersten, M.J. Carey, I. Manolescu and R. Busse, 2002. XMark: a benchmark for XML data management. Proceedings of the 28th international conference on Very Large Data Bases, 974-985.

Tatarinov, I., S.D. Viglas., K. Beyer, J. Shanmugasundaram, E. Shekita and C. Zhang, 2002. Storing and querying ordered XML using a relational database system. Proceedings of the 2002 ACM SIGMOD international conference on Management of data, 204-215.

University of Washington XML Data Repository [Online]. Available: http://www.cs.washington.edu/research/xmldatasets/www/repository.html

W3C Document Object Model (DOM) [Online]. Available: http://www.w3.org/TR/DOM/

W3C Extensible Markup Language (XML) 1.0 [Online]. Available: http://www.w3.org/TR/xml/

XMark - An XML Benchmark Project [Online]. Available: http://www.xml-benchmark.org/downloads.html