Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 81 (2015) 31493-31498

ISSN: 2229-712X

Implementation of Efficient Signature Scheme for Leakage Free Dissemination of XML Content using Structure Based Routing

Vivek N. Waghmare and Ravindra C. Thool

S. G. G. S. I. E. & T, Nanded, SRTMUN, Nanded, Maharashtra - 431605, India.

ARTICLE INFO

Article history:

Received: 20 February 2015;

Received in revised form:

15 March 2015;

Accepted: 1 April 2015;

Keywords

Confidentiality,
DOM,
EPON,
Integrity,
Privacy,
SBR.

ABSTRACT

Nowadays security has always been an important issue in the business world to ensure the integrity & confidentiality of content and transactions, to make sure that information is disseminated appropriately. But while dealing with contents encoded in XML, the hierarchical tree structured data have different confidentiality and integrity for different portions of the same content. Thus need dissemination approach specifically tailored to XML that addresses the issues of efficiency and scalability must however be provided by maintaining the security of contents and privacy of the parties acquiring and disseminating contents. The main objective is to provide solution to above problem using structural properties of tree based data model (such as XML document) and Document Object Model (DOM). This proposed signature scheme is based on notion of Encrypted Post Order Numbers (EPON), which facilitates efficient dissemination of selected portion of content. By using structure based routing (SBR) scheme, it prevents information leakage and assures that delivered content to user according to access control policies. The structure based routing framework facilitates the dissemination of contents with varying degrees of confidentiality and integrity in a network.

Introduction

Most of the businesses exchange information and provides service on web using hierarchical (tree) data model (such as XML). Dissemination of hierarchical data model requires different security for different portion of same data. Requirement for such a dissemination approach include the following:

Access Control

To prevent unauthorized users to infer sensitive information through the data they authorized to access.

Data Integrity

Not only the integrity of the data must be verifiable by the user, but also any compromise to the data must be precisely determined.

Data Confidentiality

A user receives only that information that user is allowed to access, according to access control policies and not able to infer any information that user is not authorized to access.

Efficiency and scalability must however be provided by assuring at the same time security of contents and privacy of the parties acquiring and disseminating contents [3]. It is incompetent to provide high bandwidth content distribution systems if integrity of the disseminated contents is not as sure or the property of the contents not protected [11]. Such problems are further complicated while dealing with contents encoded in XML, in that, because of the hierarchical organization of the content, different confidentiality and integrity requirements may exist for different portions of the same content [4] [14]. Therefore need a dissemination approach exactly tailored to XML that, addresses the issues like security, privacy and scalability in a universal manner [2]. The structural properties furthermore contribute towards the efficiency and scalability of the dissemination framework. This solution is based on the simple notion of encrypted Post Order Numbering [1] and its properties. A indispensable feature of this approach that, access control policies specifying which entity can access which portion of the contents, so that contents are disseminated according to these policies.

The resulting dissemination model is a multicast model for XML dissemination [13] that based, on the content structure and access control policies. Furthermore, this approach utilizes the properties of Post Order Numbers towards integrity assurance. This technique endure consumers to verify the integrity of data they receive, and in the case in which data have been tampered with, allows the consumers to determine the affected portions of the data.

Basic Concept

XML (eXtensible Markup Language) [9] is a used as standard for document interchanges languages for the web. It is platform for application integration and management on the Internet. XML document contain information of different sensitivity degrees that must be shared by possible large user communities.

Each tagged elements has zero or more elements, zero or more attributes, and may contain textual information that is data content. Elements can be nested at any depth in the document structure [9]. The relation between parent and child nodes is represented as directed edges, with edges directed from parents to child. There are two types of tags used in XML: the start tag, at the beginning of the element, with the form <tag-name>, and the end tag, at the end tag, at the end of the element, with the form </tag-name>.

Tele:

E-mail addresses: comsvivek@gmail.com

Let D be a document and T be a DOM [10] tree representation of D. T (V, E) where V be a set of vertices and E be a set of edges. T is a nonlinear, acyclic data structure. $Content_x$ be content only at x. $Content_x$ contains only the content specific to x and not of other nodes.

The Dissemination of a Document exploits following Structural properties of XML data [1][2];

1. Order preserving XML data i.e. nodes x and y have an order among them in D.
2. Unit of data access is sub-tree representation of a subdocument. The smallest unit is a node.
3. Any element and its corresponding subdocument are accessible by themselves or by a sub-tree rooted at any of their ancestor.

Related Work

Bertino and Ferrari approach supporting access control in both pull and push based distribution of data [3]. Information pull is based on authorization. Consumer sends request to source for XML document. When consumer submits an access request then access control system checks authorization of consumer. Based on this authorization, consumer is returned a view of the requested document that contains all and only those portions. When no authorizations are found then, access is denied. Information push approach is used for distributing documents to users which based on broadcast data to clients. In this case, different users may have privileges to see different, selected portions of the same document. Thus, different views of same document are sent to different consumer. Example, the case of a newsletter sent once a week to all users. Different users have different privilege to access different, selected portion of same document, supporting an information push approach for generating different physical views of the same document and sending them to proper users. The main problem with Information pull and Information push approach is number of views becomes large and such approach cannot be practically applied.

Gladney and Lopspiech proposed solution for above problem which is mainly based on, Multilevel Encryption [3]. In multilevel encryption, different portions of same document are encrypted with different keys and same encrypted copy is broadcast to all subjects.

Issues related to multilevel encryption are as follows,

1. Which and how many keys should be distributed to which subjects?
2. How to securely and efficiently distribute keys?
3. How to encrypt document?

Solutions for these issues are,

1. Encryption of document according to specified access control policies.
2. According to policies apply key, therefore number of policies equal to number of keys.

Merkle proposed a digital signature scheme [1][2][4] based on a secure conventional encryption function over a hierarchy (tree) of document fragments. Buldas and Laur have also found that Merkle trees are binding (integrity-preserving) but not hiding (confidentiality-preserving) the information. The use of commutative hash operations to compute the Merkle hash signature which prevents leakage related to the ordering among the siblings. However it cannot prevent the leakage of signatures of a node and to resolve structural relationships with its descendants or ancestors. Moreover, one-way accumulation is very expensive in comparison to the one-way hash operation. The Merkle hash technique has been widely used in data authentication. Devanbu et al. used the Merkle hash technique for authenticating XML data. Bertino et al. proposed a technique based on the Merkle hash technique for selective dissemination of XML data in a third party distribution framework [4]. But this technique is not scalable and does not remove extraneous data. It is sensitive to data tampering attack and inference attack. Kocher proposed to use Merkle hash trees for distribution to third parties. For secure multicast, Perrig uses static data ordering over symmetric encryption Chatvichienchai and Iwaihara proposed mechanisms for secure updates, without leading to information leakages. However such mechanism does not address the problem of information leakages during verification of integrity of partial XML documents.

Traversal numbers [5] used for querying and navigation of XML data by Zezula et al. However they fail to address any security issues. They use the non-randomized version of traversal numbers, which is unsuitable for security purposes. Traversal numbers have also been used for secure querying of data. Wang et al. have used a notion similar to traversal numbers in defining the structural index in XML databas in order to be able to locate encryption blocks as well as their unencrypted data nodes which, satisfy user query. They use real intervals [0, 1] for root and every child of the root is assigned a sub-interval such as [0.5, 0.6]. The first entry in the interval can be assumed to be referring to the pre-order number and the second one to the post-order number. However they do not derive such an interval from traversal numbers nor do they use traversal numbers for signing trees.

For secure publish-subscribe system [1][2][6], several approaches have been proposed to address efficiency issues concerning pub/sub systems. Most approaches e.g. use a spanning tree structure for event routing. In order to reduce the matching that has to be performed by brokers from the root to the leaves, several optimization techniques have been proposed. However, security issues in content-based pub/sub systems have not been investigated.

System Architecture

Generation of EPON Value

In this, take .XML file as a input, then by using Document Object Model (DOM) creates hierarchical tree representation of XML file. Then assign post order number (PON) to each node in tree. Generate sorted random number and combine with post order number (PON). These combined numbers are given as input to order preservation technique which creates encrypted post order number (EPON) that is shown in shown in figure 1. EPON generation module overcome security related flaws of solutions based on the use of PON.

Using EPON value, create structural identifier for each node in tree. Then create integrity identifier using structural identifier and content at that node. Create encoding value for each node using structural identifier and integrity identifier. After encoding, apply encryption on encoded node using symmetric or asymmetric encryption technique that is, shown in Figure 2.

Structure Based Routing

Structure Based Routing involves following Entities:

- Document source is document producer or owner.
- Publisher publishes data to set of subscribers,
- Subscribe subscribes to data and sends request to a router-based publisher.
- Router routes specific portion of data to consumer and other router.
- A router is both a publisher and a subscriber.

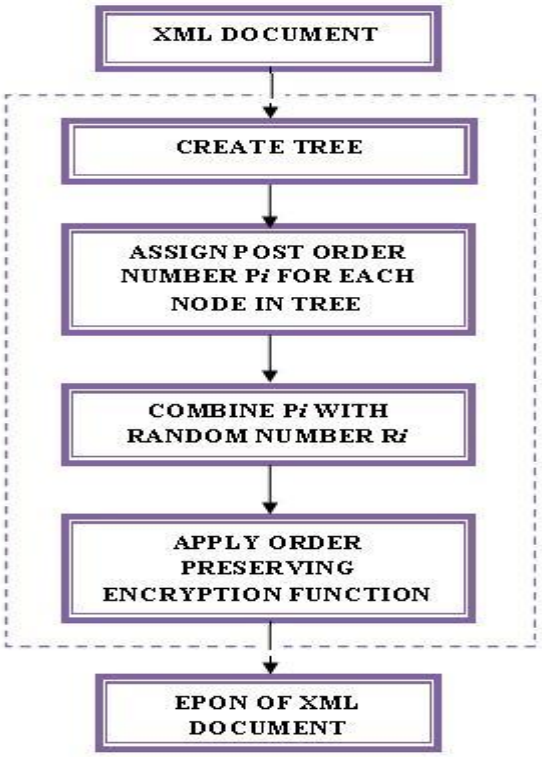

Figure 1. Generation of EPON Value

Document Encoding and Encryption

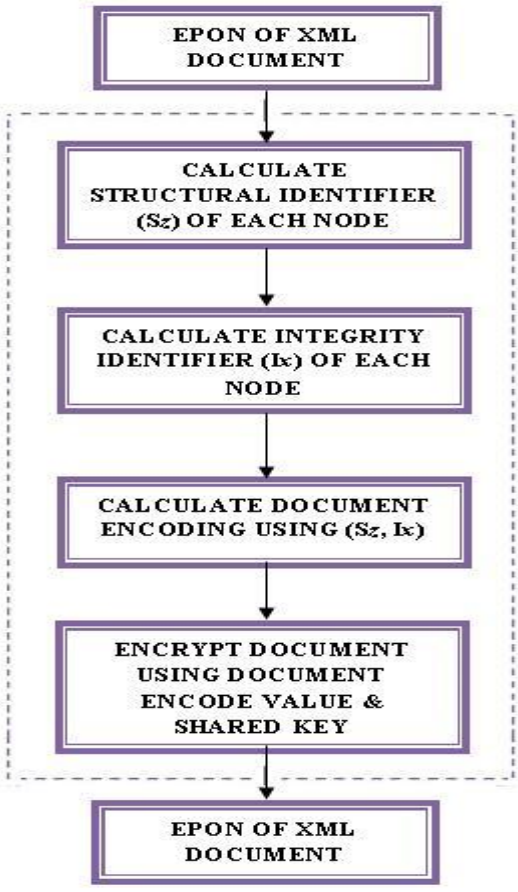

Figure 2. Document Encoding and Encryption

Access Permission

Access permissions on content for a user are expressed on a complete or partial sub-tree of document. Access permissions for a user u on a document D are denoted by Pud is (AllowedSet, NotAllowedSet).

AllowedSet = $\{e_x \mid$ access is allowed to content in maximal sub-tree rooted at x$\}$.

NotAllowedSet = $\{e_y \mid$ access is not allowed to content in maximal sub-tree rooted at x$\}$.

Content Router

A router in this context is an application level router that is able to route documents based on their structural organization which is represented by its structural identifier. Each router has an associated set of nodes in a document.

Dissemination Network

In document dissemination network, a link is between two content router and intermediate network router. For development of dissemination network uses structural identifier. Subscription process is initiated by a consumer. Use a three way handshake protocol to establish a subscription link between two routers.

Link Setup Protocol

Consumer C sends request for D using {<c-id, c-credentials, document URI, permissions, callback - address>+} to router R. R authenticates C and determines list of signature of content nodes that C has access to.R determines set of subdocuments from set of signatures.If the list of accessible sub-tree include sub-tree with root e_z that is subsumed by content tree. $e^h_{lowest} \leq e^z_{lowest} \leq e_z \leq e_h$ Otherwise, R sends request to some or all neighboring R_s. If access permission includes multiple sub-trees, the subscription request includes each of these sub-trees.Upon receiving a request from R, R_i checks e_x. If correct then send response otherwise repeats link setup procedure.Upon receiving responses R selects a parent R, the router registers C and sends response back to client.

Each R determines if a new node and the existing node can be combined together to form a complete sub-tree of document. Then replace all the nodes stored in database by lowest common ancestor of sub-trees.

Document Distance Based Technique

Document distance based technique is used for router selection. Let ex and e_z be EPON's of x and z nodes. e_z is PEPON at a router and e_x is EPON of the root of sub-tree requested by another router. Then document distance calculated by $(e_x - e_z)$. After this is the router who has minimum document distance they selected as publisher router R_i. This technique reduces the cost of overall dissemination network. It defines how structural and EPONs can be used in establishing multicast paths. The resulting multicast topology is a directed acyclic graph (DAG), which might be in tree case. Every path from document source to a consumer contains a monotonically decreasing sequence of EPON's of document as PEPON's of routers. A parent router never has a less EPON a given router's PEPON.

Content Publishing:

Content published in following ways

Content Delivery to Router: If there is a nonempty set of routers that are subscribes for some nodes in document, then R forward document to these routers based on their requirements is as follows:

- For each router in its subscriber set, a router determines number of registered nodes.
- It identifies and extracts these document nodes from respective sub-trees.
- Then they are encrypt and sent to subscribing router.

Content Delivery to Consume: If R has a nonempty set of consumers for the document, it then forward document to consumer after encrypting it by using encryption technique but based on access control policies.

Content Identification and Extraction Step: Content identification and extraction is carried out at each router that has at least one subscriber. Router keeps list of signatures of the roots of sub-trees that contain PEPONs.

Identification Step: Identification step determines the belongs to relation among each of content roots accessible to each consumer and content sub-trees it receives using simple EPON property, that is, any node y that belongs to sub-tree rooted at node x is such that $e_y < e_x$, where e_x and e_y are EPONs of x and y. e_y is also greater than or equal to e^x_{lowest}, as defined in structural identifier S_x of x. The worst case complexity of identification step is O(mn) where m is number of received content sub-trees and n is number of subscribers at a given router.

Extraction Step: A depth first traversal is used to determine subscribed content root during extraction step. The EPON of root of subscribed content root is compared with EPON of visited node. If these EPONs match, the corresponding sub-tree is extracted. The worst case complexity of extraction step is O(v+e) where v is number of nodes in received sub-tree and e is number of edges in received sub-tree.

Document Verification

Document verification can be done at consumer side. The following steps must be executed for document verification [1] [2]:

- If nodes have been dropped.
- If the order of the nodes has been changed.
- If the content of a node has been compromised.
- If some nodes have been added in an unauthorized manner.
- If the content of one node has been replaced with the content of another node.

Update Management

Update to documents is either content or structure in the context of structure based routing.

Update Content: If any change in content then only the local hash of the node changes. The updates of the changed nodes along with their signatures are forwarded to the routers.

Update Structure: Structural changes have to be reflected in the mapping from user credentials to accessible nodes and their signatures. Therefore, the services that implement the mapping function from user credential to structural identifiers need to be notified accordingly with the new EPONs. If it is a distributed hash table, then the document source updates the hash table. The routers are also notified of the modifications. Removal of a sub-tree is notified to the routers and consumers having the document.

Update Tree: In case of addition of a new sub-tree, the original structure of the document is not affected. Therefore, the update is propagated to all the routers that have consumers with access permission to the new sub-tree. In case of interchanges, the changes need to be propagated to the routers and consumers that are registered for any updated node or an ancestor of that updated node. It is shown in Figure 3.Figure 1. and Figure 2. are used for generation of structural signature scheme.

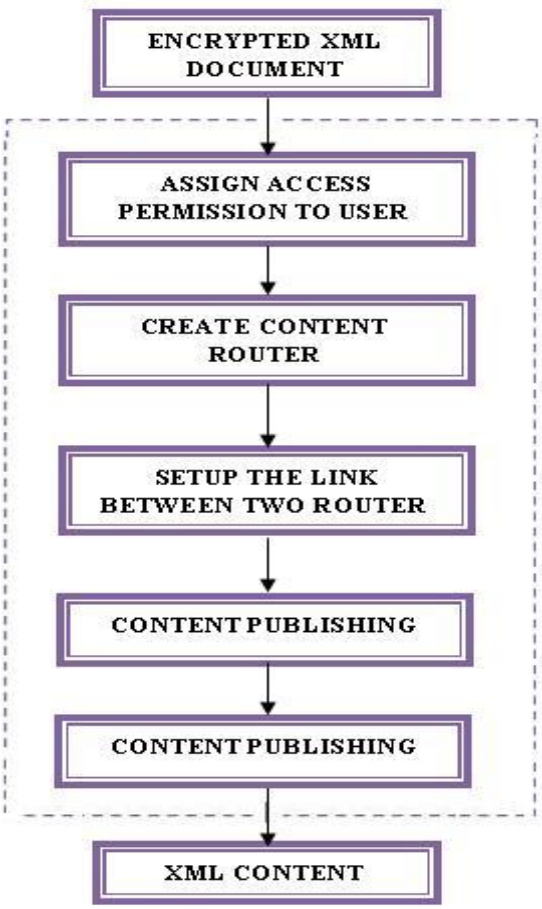

Figure 3. Structure Based Routing

Result Analysis

XML (Extensible Markup Language) is used as standard for document interchanges languages for the web. XML organizes data according to tree structure integrity and confidentiality of XML data is an important requirement for distributed web based application. The XML document is shown in Figure 4.

```
<?xml version="1.0" ?>
<!-- The template
     -->
- <w>
  - <z>
      <x />
      <y />
    </z>
  - <v>
      <t />
      <u />
    </v>
</w>
```

Figure 4. .xml File

Comparison between Merkle Hash Technique and Proposed Signature Scheme:

Merkle proposed a digital signature scheme [1][2][4] based on a secure conventional encryption function over a hierarchy (tree) of document fragments. The use of commutative hash operations to compute the Merkle hash signature prevents leakage related to the ordering among the sibling.

Table I. Verification process of merkle hash signature scheme

Node	Nodes used	Leaked Information during Verification of Node in Merkle Hash
X	x	None
Y	y	None
Z	x, y, z	Signature of x, y; y as sibling of x and x, y are child of z
T	t	None
U	u	None
V	t, u, v	Signature of t, u; u as sibling of t and t, u are child of v
W	z, v, w	Signature of z, v; v as sibling of z and v, z are child of w

Table II. Verification process of generated epon value for proposed signature scheme

Node	Nodes used	Leaked Information during Verification of Node in Proposed Signature Scheme
x	(11,11)	Define upper and lower bounds of sub tree
y	(22,22)	,,
z	(43,11)	,,
t	(64,64)	,,
u	(75,75)	,,
v	(96,64)	,,
w	(107,11)	,,

Table III.Time required for generation & verification process for merkle hash scheme

Number of Nodes	Time for Hash Tree Generation in sec	Time for Generation of Merkle Hash Signature Scheme in sec	Time for Verification Process for Merkle Hash Scheme in sec
4000	0.194	0.23	0.422
8000	0.342	0.451	0.623
12000	0.405	0.551	0.82
16000	0.486	0.725	1.201
20000	0.612	0.981	2.01

Table IV. Time required for generation & verification process for proposed structural signature scheme

Number of Nodes	Time for EPON umbering Generation in sec	Time for Generation of Proposed Signature Scheme in sec	Time for Verification Process in sec
4000	0.172	0.25	0.463
8000	0.258	0.55	0.782
12000	0.318	0.779	0.909
16000	0.435	1.085	1.509
20000	0.579	1.345	2.343

In Merkle Hash Technique one way accumulation is very expensive in comparison to the one-way hash operation.

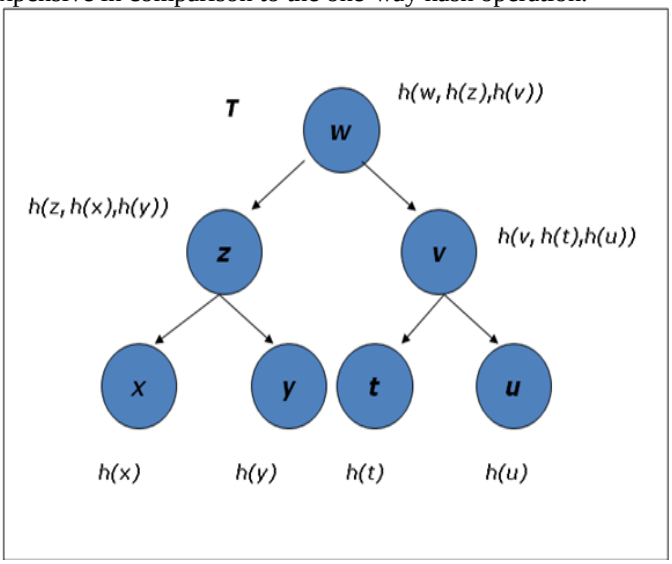

Figure 5. Hash Tree Generation using Merkle Hash Signature Scheme

Generate sorted random number and combine with post order number. These combined numbers are given as input to order preservation technique which creates encrypted post order number for tree as shown in below figure 6 and verification process result for Structural Signature Scheme shown in Table IV.

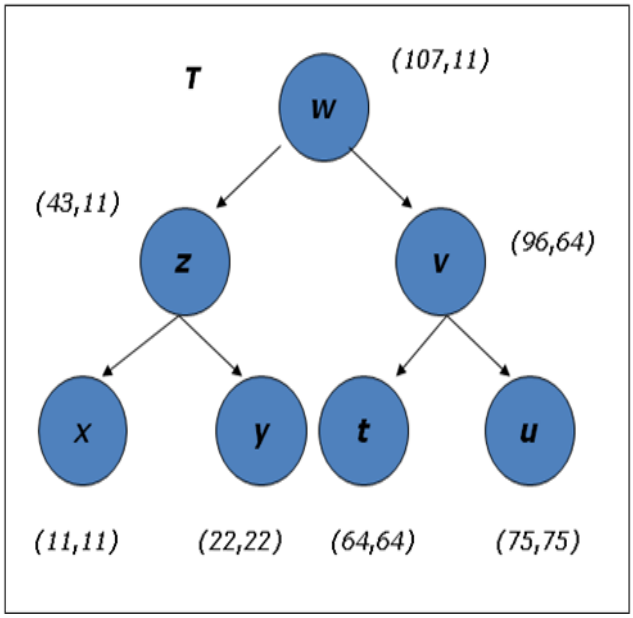

Figure 6. Generation of EPON Value using Proposed Signature Scheme

Table I and II shows comparison of result leaked information during verification of node in both techniques. Merkle hash technique is sensitive to inference attack and data

tampering attack. Merkle has trees are binding (integrity-preserving) but not hiding (confidentiality-preserving). Proposed Signature Scheme, take input as a XML file and create DOM tree for XML file. Then assign post order number to each. The implemented algorithms are tested for different number of nodes, with respective time. TABLE III & IV gives information about time required for verification process for Merkle Hash & Proposed Signature Generation scheme with respect to its number of nodes respectively.

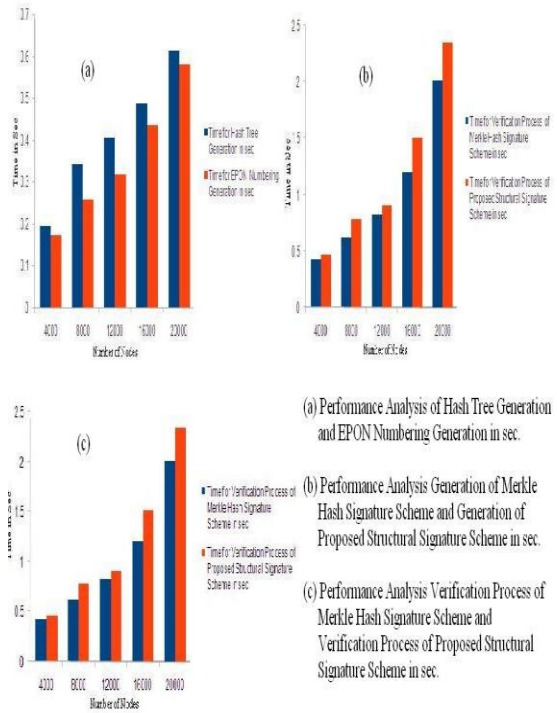

Figure 7. Graphical Representation of Performance Analysis of Merkle Hash & Proposed Signature Scheme

Conclusion

Proposed Signature Scheme solves problem of post order numbering using order preserving encryption technique. In Order preserving encryption scheme, comparison, equality queries can be directly applied to encrypted data. No need to decrypt it. Role based access control policies depend on EPON numbering which ensures that a consumer is delivered only the portion of data that it has access to. Result of query processing over data encrypted using EPON are exact. Proposed Signature Scheme provides stronger security in terms of integrity and confidentiality. It simplifies the transmission of tree based data from a publisher to consumer and improves efficiency of such transmission. A maximal structural block at routers ensures that the routers have access to only that much amount of data that its consumers collectively have access to. This Proposed Signature Scheme is efficient in terms of complexity & cost. Proposed Signature Scheme mainly protects against Inference attack and Data tampering attack.

References

1. A. Kundu, E. Bertino, “A new model for secure dissemination of XML Content,” IEEE Transaction, May 2008.

2. Kundu and E. Bertino “Secure dissemination of XML content using structure based routing,” in Proc. 10th IEEEInt. EnterpriseDistrib.Object Comput. Conf. (EDOC’06), 2006, pp.153-164.

3. E. Bertino and E. Ferrari, “Secure and selective dissemination of XML document,” ACM Trans. Inf. Syst. Secur., Vol.5, no.3, pp.290-331, 2002.

4. Bertino, B. Carminati, E. Ferrari, B. M. Thuraisingham, and A. Gupta, “Selective and authentic third party distribution of XML documents,” IEEE Trans. Knowl. Data Eng., Vol.16, no.10, pp.1263-1278, oct.2004.

5. Su Cheng Haw, and G. S. V. Radha Krishna Rao, “Query optimization techniques for XML Databases,” 2005.

6. C. Wang, A. Carzaniga, D. Evans, “Security issues and requirements for internet-scale publish subscribe systems,”2002.

7. Extensible Markup Language [Online]. Available: http://www.w3.org/XML/.

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

9. R. Agrwal, J. Kiernan, R. Srikant, “Order Preserving Encryption for numeric data,” in proc.2004 ACM SIGMOD Int.Conf.Mana. Data, pp. 563-574.

10. A. K. Datta, M. Gradinariu, M. Raynal, and G. Simon, "Anonymous publish subscribe in p2p networks", presented at the Int. Parallel Distrib.Process. Symp., Nice, France, 2003.

11. F. Cao and J. Singh, "Efficient event routing in content-based publish subscribe service Networks," in Proc. of IEEE INFOCOM 2004, pp. 929-940.

12. Q. Hu, D. L. Lee, and W. C. Lee, “Optimal Channel Allocation for Data Dissemination in Mobile Computing Environments”, In 18th International Conference on Distributed Computing Systems, May 1998.

13. M. Lazaro and P. Sage, “Any Information, Anywhere, Anytime for the Warghte”, In Proceedings of the SPIE, volume 3080, pages 35-42, 1997.

14. P. Devanbu, M. Gertz, C. Martel, and S. Stubblebine, “Authentic Third-Party Data Publication”, Proc. 14th Ann. IFIP WG 11.3 Working Conf. Database Security, Aug.2000.

15. FERNANDEZ, E., GUDES, E., AND SONG, H. "A model for evaluation and administration of security in object-oriented databases", IEEE Trans. Knowl. Data Eng, 1994, 275292.

16. R. Douglass, J. Mork, and B. Suresh. Battleeld “Awareness and Data Dissemination (badd) for the Warghter,” In Proceedings of the SPIE, volume 3080, pages 18-24, 1997.

17. M. Tan, M. D. Theys, H. J. Siegel, N. B. Beck, and M. Jurczyk. “A Mathematical Model, Heuristic, and Simulation Study for a Basic Data Staging Problem in a Heterogeneous Networking Environment,” In Proceedings of the 7th International Computing Workshop (HCW'98). IEEE, 1998.

18. R. Lindell, J. Bannister, C. DeMatteis, M. O'Brien, J. Stepanek, M. Campbell, and F. Bauer. “Deploying Internet Services Over a Direct Broadcast Satellite Network: Challenges and Opportunities in the Global Broadcast Service,” In MILCOM. IEEE, 1997.

19. T. Stephenson, B. DeCleene, G. Speckert, and H. Voorhees. Badd phase ii. Dds “Information Management Architecture,” In roceedings of the SPIE, volume 3080, pages 49-58, 1997.

20. H. Salkin & J. Saha. “Set Covering: Algorithms, Results and Codes,” In Bulletin of the Operations Research Society of America, volume 20, suppl.2, Nov 1972.

21. J. Dukes-Schlossberg, Y. Lee, and N. Lehrer. “Iids: Intelligent Information Dissemination Server,” In MILCOM 97 Proceedings, volume 2, pages 635-639. IEEE, 1997.