A Review on: Issues Related to Security on Tree Structure Data

Vivek Waghamre, Dr. Ravindra Thool

**Ph. D. Research Scholar, S. G. G. S. I. E. & T, NANDED, SRTMUN, NANDED
NANDED, MAHARASHATRA -431605, INDIA**

**Professor & HOD of Information Technology , S. G. G. S. I. E. & T, NANDED
NANDED, MAHARASHATRA -431605, INDIA**

Abstract

The problem of tree structured data is that it requires different Integrity and confidentiality for different portion of same content. Integrity assurance technique not only applies integrity to received data by user, but also any compromise to data that must be precisely determined. In tree structures each node contains some content and structural relationships between the nodes. Therefore, it considers structural integrity and content integrity. Confidentiality means a user receives only those nodes as well as organized information to user has privilege according to access control policies, wherever one should not infer others information.

In this paper we reviewed different techniques related to dissemination of tree structured data that exploits structural properties of tree based data model (such as XML document). The approach is based on notion of encrypted post-order numbers which is based on post-order number properties. It facilitates efficient identification, extraction, distribution of selected content portions.

Keywords: *Confidentiality, Integrity, Tree Structured Data.*

1. Introduction

An XML (Extensible Markup Language) has become the standard document interchange language for the web. XML document contain information of different sensitivity degrees that must be shared by possible large user communities. Data sharing among multiple parties require both data integrity & data confidentiality [1]. Data that a consumer is not authorized to access, but belongs to the complete data set is called extraneous data. Flow of extraneous data to a consumer may leak information, even when this data is encrypted.

In particular, extraneous data is prone to offline dictionary attacks even by a legitimate consumer that can exploit contextual knowledge from the data elements it has access to. Therefore, it is important that extraneous data, even if encrypted with keys that the consumer does not have, be removed from the content before its delivery. 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. It is useless 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. Such problems are further complicated when 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. Thus need a dissemination approach specifically tailored to XML that addresses the issues of security, privacy and scalability in a holistic manner [2].

The structural properties also 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. By using such notion, they develop a novel content routing scheme called Structure Based Routing of XML-data. Such routing scheme prevents information leaks and at the same time improves efficiency and scalability of the structure based dissemination model. A key feature of this approach is that it directly takes into account access control policies, that is, policies specifying which entity can access which portion of the contents, so that contents is disseminated according to these policies. The resulting dissemination model is a multicast model for XML dissemination that based, on the content structure and access control policies, builds an overlay topology. Moreover, we exploit the properties of post-order numbers towards integrity assurance. This technique allows 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.

2. Related Work

Many of different design approaches are related to this topic. The models proposed Fernandez [9] and Rabitti [10] are specifically tailored to an object-oriented DBMS storing conventional, structured data. As such, great attention has been devoted to concepts such as versions and composite objects, which are typical of an object-oriented context. Those models support concepts such as positive and negative authorizations, and authorization propagation. This model also supports such concepts, even though it has a larger variety of authorization propagation options.

Three different options are supported by which the SA can specify,

- That an authorization defined at a given level in the hierarchy Propagates to all lower levels.
- That the propagation stops at a specified level down in the hierarchy.
- That no propagation has to be enforced.

By contrast ORION, Rabitti [10] has only one propagation policy, which is equivalent to option (1). Moreover, none of the above mentioned models provide support for secure information push mechanisms. This is the most innovative feature of access control model, which is not found in any access control model previously proposed for object oriented DBMSs. An access control model for WWW documents has been proposed by Samarati [11]. In this model, HTML documents are considered, organized as unstructured pages connected by links. Authorizations can be given either to the whole document or to selected portions within the document. Although [11], the idea of selectively granting access to a document (by authorizing a subject to see only some portions and/or links in the document), this work substantially differs from the proposal. Differences are due to the richer structure of XML documents with respect to HTML documents and to the possibility of attaching a DTD to an XML document, describing its structure. Such aspects require the definition and enforcement of more sophisticated access control policies, than the ones devised for HTML documents. The access control model proposed [11] has great limitations deriving from the fact that it is not based on a language able to semantically structuring the data, as in this model for XML. As such, administering authorizations is very difficult. In particular, if one wants to give access to portions of a document, he/she has to manually split the page into different slots on which different authorizations are given.

This problem is completely overcome by providing semantic information for various document components. Authorizations can thus be based on this semantic information. An access control model for XML documents has been recently proposed by Damiani [11]. This model is very similar to previous models for object-oriented databases and does not actually take into account some peculiarities of XML.

In particular, this model has two main shortcomings,

- The first one is that it does not consider the problem of a secure massive distribution of XML documents and thus considers only the information pull mode.
- Second, the model proposed in Damiani [11] does not provide access control modes specific to XML documents, it only provides the read access mode.

By contrast, it provides a number of specialized access modes for browsing and authoring, which allow the SA to authorize a user to read the information in an element and/or to navigate through its links, or to modify/delete the content of an element/attribute. Because of the widespread use of XML and due the relevance of XML security, the Worldwide Web Consortium (W3C) has set up several working groups to address the various security aspects related to XML. For example, The XML Working Groups of the W3C are working on standards for both an XML representation of digital signatures (W3C XML Signature Working Group) and encrypted contents W3CXML Encryption Working Group). The goal of the OASIS Consortium [OASIS Consortium] is the design and development of industry standard specifications for XML-based interoperability. In this framework, the XACML Technical Committee is studying the definition of a standard model for XML based security policies. However, the draft proposal for XACML is based on very simple access control policies, in that notion such as credentials, positive/negative policies, convict management, and dissemination strategies are not taken into account.

Other related work deals with approaches proposing more flexible methods to qualify subjects with respect to traditional identity-based schemes for access control. One of the most relevant research efforts in this area are role-based access control (RBAC) models. In particular, note that the concept of credential has some similarity with that of role [12]. Roles can be seen as a set of actions or responsibilities associated with a particular working activity. Under role-based models, all authorizations needed to perform a certain activity are granted to the role associated with that activity, rather than being granted directly to users. Users are then made members of roles, thereby acquiring the roles authorizations. User access to data is mediated by roles, each user is authorized to play certain roles and, on the basis of the role, he/she can perform accesses on the data. Whenever a user needs to perform a certain activity, the user only needs to be granted the authorization of playing the proper role, rather than being directly assigned the required authorizations. A basic distinction between roles and credentials is that credentials are characterized by a set of attributes, and this allows us to grant access authorizations only to users whose credentials satisfy certain conditions (e.g., access to a document can be granted to all the users with a given age or with a given nationality). This can of course be done also through roles but it requires the creation of a distinct role for each condition to enforce (e.g., enforcing the access control policy of the previous example requires the creation of two distinct roles, one corresponding to the users with the specified age, and the other corresponding to the users with the specified nationality).

This makes the specification and management of authorizations very difficult, given also the large variety of users that typically access XML documents. The concept of subject credential was first presented by Winslett [13], whereas the access control model proposed by Adam [15] provides a formalization of the concept of subject credential by proposing the credential logic-based specification language that we use in this article. Other related work on credential specification for stranger parties. In particular, the work by IBM on Trust Policy Language (TPL), Herzberg and Mass [3] is devoted to the enforcement of an XML-based framework for specifying and managing role based access control in a distributed context. This framework has been extended for mapping subject certificates to a role, based on policies defined by the owner of the resource and on the roles of the issuers of the certificates. In another area, Bertino and Ferrari approach supporting access control in both pull and push based distribution of data [3]. This approach is depending on encrypting different portion of the data with different keys & then distributing the keys to data consumers according to access control policies. 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, the access is denied. Information push approach is used for distributing documents to users which based on broadcast data to clients. Also 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 [5]. Example, the case of a newsletter sent once a week to all users. Different users have different privilege to see 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. Bertino [3] have also investigated the problem of integrity of XML data by using notion of Merkle Hash Tree. Merkle proposed a digital signature scheme [1] [2] [3] based on a secure conventional encryption function over a hierarchy (tree) of document fragments. Merkle trees are binding (integrity-preserving) but not hiding (confidentiality-preserving). The use of commutative hash operations to compute the Merkle hash signature prevents leakage related to the ordering among the siblings. However it cannot prevent the leakage of signatures of a node and the structural relationships with its descendants or ancestors. Moreover, one-way accumulation is very expensive in comparison to the one-way hash operation. Merkle hash trees are a well-known mechanism used in several computer areas for certified query processing. For instance, it has been exploited by Naor and Nissim in [16] to deal with the task of introducing and maintaining efficient authenticated data structures that holds the information of certificates about its validity. More precisely, the paper proposes as data structure a sorted hash tree scheme, such that tree leaves correspond to revoked certificates. Thus, verifying that a certificate is revoked or not implies verifying the existence of certain leaves in the tree.

A similar approach has been proposed by Devanbu et al. [17] to prove the completeness and authenticity of queries on relational data. Similar schemes have also been used for micropayments [18], where Merkle hash trees are used to reduce the number of public key signatures that are required in to provide or authenticating a sequence of certificates. By contrast, the use of such trees for handling XML documents is still a novel aspect, which, to the best of our knowledge, has been so far investigated only by Devanbu et al. in 179. In this work, the authors have developed a scheme, based on Merkle hash trees, allowing clients to validate the answers to certain type of queries against XML sources managed by untrusted publishers. The method developed in [17] is based on the definition of a data structure, called Xtrie, which stores the set of possible paths that can be specified on a given DTD. However, the work presented in [5] has many differences with regard to this proposal. A first difference is the type of XML documents supported by the two approaches. In this approach, they have no limitation on the structure of XML, documents, whereas the approach presented in does not consider attributes and it imposes that data content be only present in leaf nodes. Another important difference is the kinds of queries for which the subject is able to verify authenticity. In our approach, we can certify the authenticity for each possible kind of XPath queries, whereas the approach presented in considers only queries returning whole sub trees. A further distinction is that it also considers completeness with regard to access control rights, besides data authentication, and provides a comprehensive architecture and related mechanisms to support data authentication and completeness services. Bertino proposed a technique based on the Merkle hash technique for selective dissemination of XML data in a third party distribution framework [4]. Gladney and Lopspiech proposed solution for above problem which is based on Multilevel Encryption [3]. In multilevel encryption, different portions of same document are encrypted with different keys and same encrypted copy is broadcasted to all subjects.

Issues related to multilevel encryption are,

- Which and how many keys should be distributed to which subjects?
- How to securely and efficiently distribute keys?
- How to encrypt document?

Solution for these issues:

- Encryption of document according to specified access control policies.
- According to policies apply key.
- Therefore number of policies equal to number of keys.

The Data Dissemination Problem has been studied by a number of projects [19, 2]. However, none of them attempt to reduce costs by automatically merging similar queries. The Query Merging Problem is also related to Client-side caching in client/server configurations [21]. In this approach, data is loaded into each client cache as answers to other queries are broadcast by the server. When a client is ready to make a query, it first checks in its own cache to see if the cache already contains the answer. The difference with this work is that in the client-caching approach, queries are not known, so the server cannot optimize the global cost. This research is also related with the Semantic Query Optimization Problem [23]. The goal of the Semantic Query Optimization problem is to use semantic knowledge (such as integrity constraints) for transforming a query into a form that may be answered more efficiently than the original version. The main difference with this work is that in semantic query optimization only one query at a time is optimized. This limits the opportunities for improvement versus work, where they considered a set of queries.

There are a number of data dissemination products and services in the market [24, 27]. However, as far as know they do not attempt to do any real query merging. Most of these products are very simple, requiring clients to maintain their subscriptions and to "pull" from the server any new information. Servers normally unicast the results to each client, making this approach non-callable and resulting in a very high cost. The Cellular Telephony and Telecommunication research community has also consider the problem of improving the bandwidth use on broadcast channels [25][24]. The difference between this effort and work is the level of abstraction. While the telephony community focus on random memory page requests (and therefore, there is little information available to the optimizer), specially this work focus on queries and query answers which allows to have more sophisticated schemes. The BADD problem [26, 28] has generated a wealth of research in the data dissemination arena. References [33] and [34] have proposed multicast protocol, which can be used as a low level support to this algorithm. Deployment of Internet services through a satellite broadcast channel has been studied in [26] and smart information “push” by [33]. Reference [30] extends the client-side caching by considering caches not only at the client, but also at intermediate locations "close" to the clients. Finally, in [31] the data staging problem is described and heuristics to solve it are presented. The query merging problem in a geographical database is closely related to the polygon covering problem, and to the set covering problem [32]. However, the special characteristics of the Query Merging Problem make it difficult to directly use the well known solutions to those problems.

3. Conclusions

RBAC model supports the formulation of high-level access control policies. Such policies take into account both user characteristics, document contents and structure. In Merkle hash tree, values of sub-trees that are not accessible to consumer also to be forwarded. Also it, avoids sending the hash values of sub-trees that are not accessible to consumer. Query Merging Problem, presented a general framework and cost model for evaluating merging, and also a variety of merging algorithms. It shows that dissemination costs can be significantly decreased by applying a merging algorithm, and that heuristic algorithms work well.

References

[1] A. Kundu and E. Bertino, “A New Model for Secure Dissemination of XML Content,” IEEE Transaction on, May 2008.

[2] A. Kundu and E. Bertino, “Secure Dissemination of XML Content Using Structure Based Routing,” in Proc. 10th IEEE Int. Enterprise Distrib.Object Computing. Conf. (EDOC06), 2006, pp. 153164.

[3] E. Bertino and E. Ferrari, "Secure and Selective Dissemination of XML Documents," ACM Trans. Inf. Syst. Secure, vol. 5, no. 3, pp. 290331,2002.

[4] E. 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. 12631278,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] A. K. Datta, M. Gradinariu, M. Raynal, and G. Simon, "Anonymous Publishsubscribe in p2p Networks," presented at the Int. Parallel Distrib.Process. Symp., Nice, France, 2003.

[8] F. Cao and J. Singh, "Efficient Event Routing in Contentbased Publish Subscribe Service Networks," in Proc. of IEEE INFOCOM 2004, pp. 929-940.

[9] Fernandez, E., Gudes, E., & Song, H. "A Model for Evaluation and Administration of Security in Object-oriented Databases," IEEE Trans. Knowl. Data Eng,1994, 275292.

[10] Rabitti, F., Bertino, E.,Kim, W., Ndwoelk, D. “A Model of Authorization for Next-Generation Database Systems,” ACM rans. Datab. Syst. 1991, 16, 1, 88131.

[11] Damiani, E., De Captitani Di Vimercati, S. Paraboschi, S., & Samarati, P. "Securing XML Documents," In Proceedings of the 6th International Conference on Extending Database Technology (Konstanz, Germany), 2000 pp. 121135.

[12] Osborn, S. Ed. Proceedings of the 5th ACM Workshop on “Role-Based Access Control,” (Berlin, Germany) ACM, New York, 2000.

[13] Winslett, M., Ching, N., Jones, V., & Slepchin, I. "Using Digital Credentials on the World Wide Web," J. Comput. Secu.197, 7.

[14] Adam, N., Atluri, V., Bertino, E., & Ferrari, E. “A Content-Based Authorization Model for Digital Libraries,” IEEE Trans. Knowl. Data Eng.2002, 14, 2,296315.

[15] Herzberg, A., Mass, Y., & Mihaeli, J. “Access Control Meets Public key Infrastructure or Assigning Roles to strangers,” In Proceedings of the IEEE Symposium on Security and Privacy (Oakland, Calif.). IEEE Computer Society Press, Los Alamitos, Calif. 2000.

[16] M. Naor and K. Nissim, “Certificate Revocation and Certificate Update,” Proc. Seventh USENIX Security Symp., 1998.

[17] 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.

[18] S. Charanjit and M. Yung, Paytree: “Amortized Signature for Flexible Micropayments,” Proc. Second Usenix Workshop Electronic Commerce, 1996.

[19] D. Giord. Polychannel, “Systems for Mass Digital Communication,” CACM, February 1990.

[20] T. Imielinski and B. Badrinath, "Mobile Wireless Computing," CACM, 37(10):18-28, Oct 1994.

[21] A. M. Keller and J. Basu, "A Predicate-Based Caching Scheme for Client-Server Database Architectures," The VLDB Journal, 5(1), January 1996.

[22] T. Bowen, G. Gopal, G. Herman, T. Hickey, K. Lee, J. Raitz, and A. Weinrib. “The Data cycle Architecture,” CACM, 35(12), December 1992.

[23] J. J. King. Quist: “A System for Semantic Query Optimization in Relational Databases,” In Very Large Data Bases, 7th International Conference, September 9-11, 1981,Cannes, France, Proceedings, pages 510- 517. IEEE Computer Society Press, 1981.

[24] A. Chan. Transactional publish/subscribe: “The Proactive Multicast of Database Changes,” In ACM SIGMOD, June 1998.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

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

Vivek N. Waghmare completed his B. Tech., M. Tech. from SGGS Institute of Engineering & Technology, Nanded, and Walchand College of Engineering, Sangli, India in 2008 and 2010 respectively. He is currently pursuing his Ph. D. at SGGS Institute of Engineering & Technology, Nanded under SRTMUN, Nanded, India. His research area includes HPC and Network Security.

Dr. Ravindra C. Thool received his BE, ME and Ph.D. in Electronics from SGGS Institute of Engineering & Technology, Nanded, India, in 1986, 1991 and 2003 respectively. He is currently working as professor and head with Information Technology department in the same organization. His research area includes Computer Vision, Image processing and multimedia information systems. He has published several research papers in refereed journals and professional conference proceedings. He is member of IEEE, Life member of ISTE.