Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1351 – 1355

SURVEY ARTICLE

SURVEY ON PROTECTING PRIVACY AND SECURITY IN XML INFORMATION BROKERING

Noe Elisa [1], K. Suresh Babu [2]

[1]M.Tech scholar, School of Information Technology, JNTU, Hyderabad, Tanzania

[2]Assistant Professor, School of Information Technology, JNTU, Hyderabad, India

[1] noelisa1983@gmail.com, [2] kare_suresh@yahoo.co.in

Abstract- In a federated information system with diverse participants (from different organizations) such as data producers, data consumers, or both, the need of cross-organizational information sharing naturally arises. However, different types of applications often need different forms of information sharing. In particular, while some applications (e.g., stock price updating) would need a publish-subscribe framework, the on-demand information access is more suitable for other applications. A number of information brokering systems have been developed to provide efficient and secure information sharing. Many existing information brokering systems adopt server side access control deployment and honest assumptions on brokers. However, little attention has been drawn on privacy of data and metadata stored and exchanged within Information Brokering System (IBS).

We proposed an Information Brokering System (IBS) on the top of a peer-to-peer overlay to support information sharing among loosely federated data sources. It consists of diverse data servers and brokering components, which help client queries to locate the data servers.

However, privacy of data location and data consumer can still be inferred from metadata (such as query and access control rules) exchanged within the IBS, but little attention has been put on its protection. We studied the problem of privacy protection in information brokering process. A formal presentation of the threat models with a focus on two attacks: attribute-correlation attack and inference attack is been performed. We propose a flexible and scalable system using a broker-coordinator overlay network. Through an innovative automaton segmentation scheme, distributed access control enforcement, and query segment encryption, proposed system integrates security enforcement and query forwarding while preserving system-wide privacy.

We performed a comprehensive analysis on privacy, end-to-end performance, and scalability, the proposed system integrate security enforcement and query routing while preserving system-wide privacy with reasonable overhead.

Keywords: brokering systems; Access control; information sharing; privacy

I. INTRODUCTION

The Internet enables global sharing of data across organizational boundaries. Distributed file systems facilitate data sharing in the form of remote file access. However, traditional access control mechanisms used in distributed file systems are intended for machines under common administrative control, and rely on maintaining a centralized database of user identities. They fail to scale to a large user base distributed across multiple organizations. The Internet offers the possibility of global data sharing and collaboration. One class of mechanisms commonly used by organizations is shared data access via file sharing, using remote file access in distributed/networked file systems. However, most existing information brokering systems do not offer secure, scalable and dynamic cooperation across organizational boundaries. When users in distinct administrative domains try to share files, either inefficient or cumbersome exchange of information or compromises in security result.

Government entities such as the military are in an analogous situation: they have incentive to share sensitive information about potential military targets, suspicious activities, difficult technical problems, or vulnerabilities with partners at differing levels of trust. Misuse of this information may result in harm. But harm is also possible if information is not shared, as the information could be necessary to prevent loss of life, assets, or advantage.

To better understand such requirements, we overview the unique needs of such interorganization collaboration by considering an example in the healthcare domain. Large-scale health information infrastructures, such as Regional Health Information Organization (RHIO), are being developed to share medical information (e.g., patient records) collected by collaborative health providers (e.g., hospitals) via protected "channels". First, there is no centralized authority to coordinate the data in different hospitals. Each health provider is authorized by its patients to collect medical information independency, and stores it across multiple local data servers. Since the data is private and sensitive, the health providers are responsible for not leaking patient records to irrelevant parties. The health providers desire to share their data to fulfill collaboration, however, they prefer to do it in a restricted and controlled fashion. Data requestors, such as doctors, need to be able to retrieve the medical records with precision and not be distracted by "noisy" data. Finally, the RHIO should be able to maintain a large number of data servers, considering the participant population. In general, such interorganization collaboration application requires an information sharing system that offers full autonomy to underlying databases preserves data security and privacy comprehensively, and provides good scalability.

Despite its importance, none of existing IBS work is designed with user and data privacy in mind. To satisfy such privacy protection requirements, therefore, a novel IBS, named as Privacy Preserving Information Brokering system (PPIB) is proposed. As shown in Figure 1, PPIB contains a broker-coordinator overlay network, in which the brokers are responsible for forwarding user queries to coordinators concatenated in tree structure while preserving privacy. The coordinators, each holding a segment of access control automaton and routing guidelines, are mainly responsible for access control and query routing. PPIB takes an innovative automaton segmentation approach to privacy protection. In particular, two critical forms of privacy, namely query content privacy and data object distribution privacy (or data location privacy), are enabled by a novel automaton segmentation scheme, with a "little" help from an assisting query segment encryption scheme. This scheme preserves privacy without sacrificing functionality. While providing "full" capability to do in-network access control and to route queries to the right data sources, this scheme ensures the information that a (curious, corrupted or broken) coordinator can gather is far from being enough to infer either "which data is being queried" or "where the data is located". Second, the automaton segmentation scheme can also provide high-quality privacy protection to metadata (e.g., access control policy). Third, user location privacy is protected by multilateral security, a design principle of PPIB.

To the best of this work, (1) PPIB is the first system that uses automaton segmentation to do privacy-preserving in-network access control. (2) PPIB is the first system that integrates automaton segmentation, in-broker access control, and query routing. (3) PPIB provides the most comprehensive privacy protection for information brokering systems, and its performance degradation is insignificant compared with traditional IBS systems (in a practical setting, the performance degradation of PPIB is at milliseconds level). (4) The evaluation results show that PPIB is a scalable privacy solution. Brokers and Coordinator are linked in a peer-to-peer fashion that makes PPIB a scalable system.

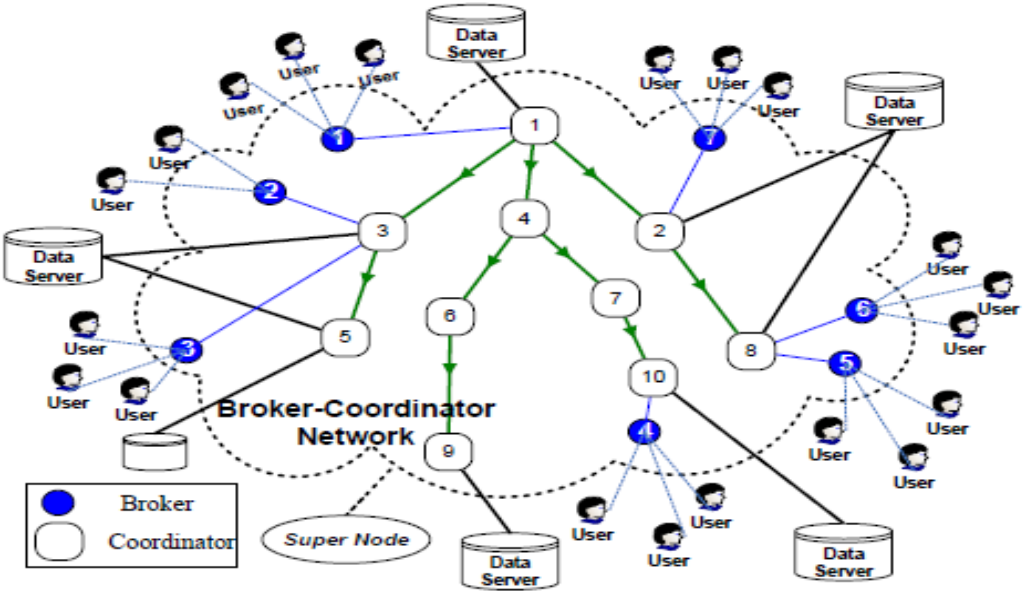

Figure 1: System architecture of an information brokering system

II. PROBLEM STATEMENT

Existing system:

Conceptually, IBS is a peer-to-peer overlay network consisting of data servers, brokering components, and end users. Applications on top IBS always involve some sort of consortium among a set of data owners (or organizations). While expressing a strong need of cross-organizational information sharing, data owners in such a consortium still expect to remain as much autonomous as possible. As a result, data owners collect data independently, and manage it in their local data servers. Data is not poured into some center data warehouse or replicated in distributed databases. Instead, data servers send metadata about their data objects distribution as well as access control rules to the consortium, which will further assign them to brokers to help information brokering. Traditional information sharing approaches always assume the use of trustable servers, such as the central data warehousing server or database servers. However, the honest or semi-honest assumptions (e.g., honest-but-curious assumption as adopted in [2]) may not hold for brokers. In practice, they may either be abused by insiders or compromised by outsiders.

It is obvious that the brokers become the most vulnerable privacy breach of a IBS, which leads to inevitable security and privacy risks. On one hand, the survival of information brokering depends on the trust of brokers to enforce authentication, access control as well as query forwarding, while on the other hand, failing to provide proper protection of information released in this process may create circumstances that harm the privacy of user, data and the system.

Disadvantages of Existing System

Many existing IBSs (1) assume that brokers are trusted and thus only adopt server-side access control for data confidentiality (2) The centralized DBMS introduces data heterogeneity, privacy, and trust issues (3) In peer-to-peer (client-saver) information sharing framework means "sharing everything or sharing nothing"

Proposed System:

In existing study of IBS, relatively little attention has been drawn to privacy protection. To impose order into the multitude of privacy vulnerabilities in current IBS approaches, taxonomy of privacy in three types is proposed: User Privacy, Data Privacy, and Metadata Privacy. To address the need for privacy protection, a novel IBS is proposed, namely Privacy Preserving Information Brokering (PPIB), it is an overlay infrastructure consisting of two types of brokering components, brokers and coordinators. The brokers make use routing protocols that create hard-to-trace communications by using a chain of proxy servers which is untraceable and mainly responsible for user authentication and query forwarding. The coordinators, concatenated in a tree structure, enforce access control and query routing based on the embedded non-deterministic finite automata, the query brokering automata.

III. SURVEY ON EXISTING IBS

This section provides theoretical background that is required in studying different ways in which the current problem has been deal with in the past. It goes on to give a brief outline of the various protocols that have been used in the existing system, the architectures that are used. Finally deals with ideas which are most related to the proposed research.

1. Broker Access Control for Information Brokerage Systems

An XML brokerage system is a distributed XML database system that comprises data sources and brokers, which, respectively, hold XML documents and document distribution information [8]. However, all existing information brokerage systems view or handle query brokering and access control as two orthogonal issues: query brokering is a system issue that concerns costs and performance, while access control is a security issue that concerns information confidentiality. As a result, access control deployment strategies (in terms of where and when to do access control) and the impact of such strategies on end-to-end system performance are neglected by existing information brokerage systems. In addition, data source side access control deployment is taken for granted as the right thing to do. Challenging this traditional, taken-for-granted access control deployment methodology is done, and argues that query brokering and access control are not two orthogonal issues because access control deployment strategies can have a significant impact on the "whole" system's end-to-end performance. Then propose the first Broker-Coordinator access control deployment strategy where access control is "pushed" from the boundary into the "heart" of the information brokerage system using coordinators.

2. Query Rewriting Techniques for Fine Grained Access Control

Access control is required in most if not all IBS. The popular XML access control model proposed in [1,3,5, 7] is adopted. In this model, users are members of appropriate roles; and an access control policy consists of a set of role based 5-tuple access control rules (ACR): R = {subject, object, action, sign, type}, where (1) subject is a role to whom an authorization is granted; (2) object is a set of XML nodes specified by XPath; (3) action is one of "read,""write," and "update"; (4) sign 2 {+, −} refers to access "granted" or "denied," respectively; and (5) type 2 {LC,RC} refers to either "Local Check" (i.e., authorization is only applied to attributes or textual data of context nodes "self::text()| self::attribute()"), or "Recursive Check" (i.e., authorization is applied to context nodes and propagated to all descendants–"descendant-or-self::node()"). When an XML node does not have either explicit

(via LC rules) or implicit (via RC rules) authorization, it is considered to be "access denied." It is possible for an XML node to have more than one relevant access control rule. If conflict occurs between "+" and "–" rules, "–" rules take precedence. In his IBS, each owner contributes a policy governing the access to her data objects, and the system-wide access control policy is simply the union of all the per-owner policies.

3. Content Based Routing (CBR) of Path Queries in Peer-to-Peer Systems

Peer-to-peer (P2P) systems are gaining increasing popularity as a scalable means to share data among a large number of autonomous nodes [7, 8]. The nodes in a P2P system store XML documents. A fully decentralized approach to the problem of routing path queries among the nodes of a P2P system based on maintaining specialized data structures, called filters this will efficiently summarize the content, Building a hierarchical organization of nodes by clustering together nodes with similar content. Similarity between nodes is related to the similarity between the corresponding filters. The existing CBR System follows hierarchical organization, which is time-consuming job to process the data between the agent and the remote user.

IV. RELATED WORK IN PROPOSED SYSTEM

Research areas such as information integration, Web search, peer-to-peer file sharing systems, and publish-subscribe systems provide partial solutions to the problem of large scale data sharing. Information integration seeks to provide an integrated view over large numbers of heterogeneous data sources by exploiting the semantic relationship between schemas of different sources [7, 6, 9]. It turns out that the PPIB approach will facilitate but is orthogonal to the information integration technology. On the other hand, Web search focuses on locating data sources with high precision and coverage [6, 8]. However, it only supports keyword queries with limited expressiveness. Peer-to-peer systems are designed to share files and data sets (e.g. in collaborative science applications). Distributed hash table technology [15, 8] is adopted to locate replicas based on keyword queries. However, although these technologies have recently been extended to support range queries [11], the coarse granularity (e.g. files and documents) still makes them short of our expressiveness needs. Further, P2P file-sharing systems may not provide complete set of answers to a request while we need to locate all relevant data. Addressing a conceptually dual problem, the XML publish-subscribe systems (e.g. [3, 6]) is probably the closely related technology to our proposed research: while we locate relevant data sources for a given query and route the query to these data sources, the pub/sub systems locate relevant consumers for a given document and route the document to these consumers. However, due to this duality, we have different concerns: they focus on efficiently delivering the same piece of information to a large number of consumers, and we are trying to route large volume but small-size queries to many fewer sites. Accordingly, the multicast solution in pub/sub systems does not scale in our environment and we need to develop new mechanisms.

One idea is to build an XML overlay architecture that supports expressive query processing and security checking on top of normal IP network. In particular, specialized data structures are maintained on nodes of the overlay networks to route path queries. In [14], a robust mesh has been built to effectively route XML packets by making the use of self-describing XML tags and the overlay networks. Kouds et al. [12] describes a decentralized architecture for ad hoc XPath query routing across a collection of XML databases using the open and agreement cooperation models. In [10], content-based routing of path queries in peer-to-peer systems is studied to serve the purpose as sharing data among a large number of autonomous nodes. The main difference between these approaches and proposed is that they focus on distributed query routing, while we seamlessly integrate query routing and security checking (e.g. access control) so as to preserve relevant privacy information. As long as privacy becomes important information that should be protected from unauthorized entities, several approaches have been designed to preserve anonymity in communication. Crowds [9] are a distributed and chained Anonymizer (http://www.anonymizer.com), where users are grouped dynamically and issue request on behalf of a crowd member. In [7], sender anonymity is guaranteed by building up anonymous connections among Onion Routers using Chaum Mix. Since proposed system integrates security checking enroute that involves more privacy concerns other than anonymity, proposed privacy addresses more challenge. As for security check, many researches have been proposed on distributed access control (see [8] for a good overview on access control in collaborative systems). Earlier approaches implement access control mechanisms at the nodes of XML trees and filter out data nodes that users do not have authorizations to access [4, 7]. These approaches rely much on the XML engines. View-based access control approaches create and maintain a separate view (e.g. a specific portion of XML documents) for each user [10, 9]. However, supporting large number of views causes high maintenance and storage cost. Proposed PPIB approach adopts a recently proposed NFA-based query re-writing access control scheme [15, 13] and extends it to a decentralized manner. It has a better performance compared with [7], and any off-the-shelf XML databases can be used due to its query re-writing nature.

CONCLUSION

The existing information brokering system is susceptible from attacks such as user privacy, data privacy, and metadata privacy. So proposed system approach, integrates security enforcement on the query forwarding among the nodes while providing comprehensive privacy protection in XML information brokering and ensuring scalability in network access.

REFERENCES

[1] **Fengjun Li**, Bo Luo, Peng Liu, Anna Squicciarini, Dongwon Lee, and Chao-Hsien Chu. "Defending against Attribute-Correlation Attacks in Privacy-Aware Information Brokering", *Proceedings of the 4th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, November, 2008. Orlando, FL.

[2] A. P. Sheth and J. A. Larson, Federated database systems for managing distributed, heterogeneous, and autonomous databases, ACM Comput. Surveys (CSUR), vol. 22, no. 3, pp. 183–236, and 1990.

[3] A. C. Snoeren, K. Conley, and D. K. Gifford, Mesh-based content routing using XML, in Proc. SOSP, 2001, pp. 160–173.

[4] L. M.Haas, E. T. Lin, and M.A. Roth, Data integration through database federation, IBM Syst. J., vol. 41, no. 4, pp. 578–596, 2002.

[5] A. C. Snoeren, K. Conley, and D. K. Gifford, "Mesh-based content routing using XML," in SOSP, pp. 160–173, 2001.

[6] N. Koudas, M. Rabinovich, D. Srivastava, and T. Yu, Routing XML queries, in Proc. ICDE'04, 2004, p. 844.

[7] G. Koloniari and E. Pitoura, Content-based routing of path queries in peer-to-peer systems, in Proc. EDBT, 2004, pp. 29–47

[8] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy, Extending query rewriting techniques for fine-grained

[9] Access control, in Proc. SIGMOD'04, Paris, France, 2004, pp. 551–562.

[10] F. Li, B. Luo, P. Liu, D. Lee, P. Mitra,W. Lee, and C. Chu, In-broker access control: Towards efficient end-to-end performance of information brokerage systems, in Proc. IEEE SUTC, Taichung, Taiwan, 2006, pp. 252–259.

[11] W. Bartschat, J. Burrington-Brown, S. Carey, J. Chen, S.Deming, and S. Durkin, Surveying the RHIO landscape: A description of current {RHIO} models, with a focus on patient identification, J. AHIMA, vol. 77, pp. 64A–64D, Jan. 2006

[12] F. Li, B. Luo, P. Liu, D. Lee, and C.-H. Chu, Automaton segmentation: A new approach to preserve privacy in XML information brokering, in Proc. ACM CCS'07, 2007, pp. 508–518.

[13] I. Stoica, R. Morris, D. Liben-Nowell, D. Karger, M. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup protocol for Internet applications," in IEEE/ACM Transactions on Networking, vol. 11 of 1, 2003.

[14] R. Huebsch, B. Chun, J. Hellerstein, B. Loo, P. Maniatis, T. Roscoe, S. Shenker, I. Stoica, and A. Yumerefendi, "The architecture of PIER: an Internet-scale query processor," in CIDR, pp. 28–43, 2005.

[15] O. Sahin, A. Gupta, D. Agrawal, and A. E. Abbadi, "A peer-to-peer framework for caching range queries," in ICDE, 2004.

[16] B. Luo, D. Lee, W.-C. Lee, and P. Liu. QFilter: Fine-grained run-time XML access control via NFA-based query rewriting. In ACM CIKM, Washington D.C., USA, nov 2004.

Authors' Profile

Noe Elisa is currently pursuing M.Tech in computer networks and information security from JNTU Hyderabad. He completed his B.SC in telecommunication engineering from university of dar es salaam, Tanzania in a year 2010. His research interests focus on security and privacy issues in distributed information systems, database systems, and communication networks. In particular Secure and privacy-preserving data sharing and data publishing, health informatics and Privacy and security in social networks.

K.SURESH BABU is an Assistant Professor at JNTU, School of IT. He completed his M.Tech.(Computer Science) from Hyderabad Central University(HCU), Hyderabad and presently pursuing his Ph.D. from JNT University Hyderabad in the field of Network Security in MANETs. He has a teaching experience of 12 years. His subjects of interests are Computer Networks, Network Security, Operating Systems, Wireless Networks, mobile Computing, Ethical Hacking and Wireless & Web Security. He has published several papers in both National and international Journals. He also participated and presented papers in International & National conferences and seminars.

