Secure Information Brokering and Sharing in Distributed Systems

G. Siva Kumar and K. Mahesh Babu

Department of Computer Science Kottam Karunakar Reddy Institute Of Technology.

Summary

To facilitate extensive collaborations, today's organizations raise increasing needs for information sharing via on-demand information access. Information Brokering System (IBS) is a peer-to-peer overlay has been proposed 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. Peer-to-peer (P2P) systems are gaining increasing popularity as a scalable means to share data among a large number of autonomous nodes. However, many existing IBSs adopt server side access control deployment and honest assumptions on brokers, and shed little attention on privacy of data and metadata stored and exchanged with in the IBS. We study the privacy in Privacy- Preserving Information Brokering in Distributed Information Sharing through an innovative automaton segmentation scheme and query segment encryption and data management issues for processing XML data in a p2p setting, namely indexing, replication and query routing and processing. With comprehensive analysis on privacy, end-to-end performance, and scalability, the proposed system can integrate security enforcement and query routing while preserving system-wide privacy with reasonable overhead.

Key words:

Information brokering, privacy, information sharing, access control.

1. Introduction

In recent years, we have observed an explosion of information shared among organizations in many realms ranging from business to government agencies. To facilitate efficient large-scale information sharing, many efforts have been devoted to reconcile data heterogeneity and provide interoperability across geographically distributed data sources. Meanwhile, peer autonomy and system coalition becomes a major trade-off in designing such distributed information sharing systems. Most of the existing systems work on two extremes of the spectrum: (1) in the query-answering model for on-demand information access, peers are fully autonomous but there is no system-wide coordination; so that participants create pair-wise client-server connections for information sharing; (2) in the traditional distributed database systems, all the participates lost autonomy and are managed by a unified DBMS. Unfortunately, neither of them is suitable for many newly emerged applications, such as information sharing for healthcare or law enforcement, in which organizations share information in a conservative and controlled manner, not only from business considerations but also due to legal reasons. As an example, imagine a future where many people have their DNA sequenced. A medical researcher wants to validate a hypothesis connecting a DNA sequence D with a reaction to drug G. People who have taken the drug are partitioned into four groups, based on whether or not they had an adverse reaction and whether or not their DNA contained the specific sequence; the researcher needs the number of people in each group. DNA sequences and medical histories are stored in databases in autonomous enterprises.[9] As a data provider, a participant would not assume free or complete sharing with others, since its data is legally private or commercially proprietary, or both. Instead, it is required to retain full control over the data and access to the data.

In the sensitive data and autonomous data owners, a more practical and adaptable solution is to construct a data centric overlay [3], [4], including the data sources and a set of brokers helping to locate data sources for queries [6], [7]. Mechanisms to route the queries based on their content, which allows users to submit queries without knowing data or server location. In previous study [7], [8], such a distributed system providing data access through a set of brokers is referred to as Information Brokering System (IBS). This system provide scalability and server autonomy. In IBS infrastructure given broker and coordinator, broker are no longer fully trustable. So, system may be abuse by insider or outsider.

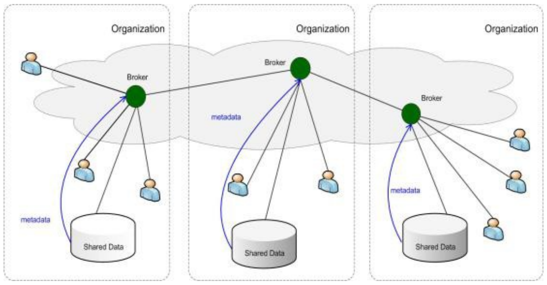

Fig.1. Overview of the IBS infrastructure

Manuscript received November 5, 2015
Manuscript revised November 20, 2015

2. Related work

2.1 Review Stage

Research areas such as information integration, peer-to-peer file sharing systems and publish-subscribe systems provide partial solutions to the problem of large-scale data sharing. Information integration approaches focus on providing an integrated view over a large number of heterogeneous data sources. Peer-to-peer systems are designed to share files and data sets (e.g., in collaborative science applications).

Privacy concerns arise in inter-organizational information brokering since one can no longer assume brokers controlled by other organizations are fully trustable. As the major source that may cause privacy leak is the metadata (i.e., indexing and access control), secure index based search schemes [22], [23] may be adopted to outsource metadata in encrypted form to untrusted brokers. Brokers are assumed to enforce security check and make routing decision without knowing the content of both query and metadata rules. Various protocols have been proposed for searchable encryption. While there are approaches proposed for multidimensional keyword search and range queries, supporting queries with complex predicates (e.g., regular expressions) or structures (e.g., XPath queries) is still a difficult open problem. In terms of privacy-preserving brokering, another related technique is secure computation that allows one party to evaluate various functions on encrypted data without being able to decrypt. Originally designed for privacy information retrieval (PIR) in database systems, such schemes have the same limitation that only keyword-based search is supported.

Research on anonymous communication provides a way to protect information from unauthorized parties. Many protocols have been proposed to enable the sender node dynamically select a set of nodes to relay its requests [30]. These approaches can be incorporated into PPIB to protect location of data requestors and data servers from irrelevant or malicious parties. However, aiming at enforcing access control during query routing, PPIB addresses more privacy concerns other than anonymity, and thus faces more challenges.

Finally, research on distributed access control is also related to our work ([31] gives a good overview on access control in collaborative systems). In summary, earlier approaches implement access control mechanisms at the nodes of XML trees and filter out data nodes that users do not have authorization to access [32]. 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, which causes high maintenance and storage costs. In this work, we adopt an NFA-based query rewriting ac- cess control scheme proposed recently in, which has a better performance than previous view-based approaches.

2.2 Preliminaries

2.2.1 XML Data Model and Access Control

The eXtensible Markup Language (XML) has emerged as the de facto standard for information sharing due to its rich semantics and extensive expressiveness. We assume that all the data sources in PPIB exchange information in XML format, i.e., taking XPath [37] queries and returning XML data. Note that the more powerful XML query language, XQuery, still uses XPath to access XML nodes. In XPath, predicates are used to eliminate unwanted nodes, where test conditions are contained within square brackets “[ ]”. In our study, we mainly focus on value-based predicates. The policy consists of a set of a set of access control rules , where (1) subject is the role to whom the authorization is granted; (2) object is a set of XML nodes specified by an XPath expression; (3) action is operations as “read”, “write”, or “update”; (4) sign ϵ {+, -} refers to access “granted” or “denied”, respectively; and (5) type ϵ {LC, RC} denotes “local check” (i.e., applying authorization only to the attributes or textual data of the context nodes) or “recursive check” (i.e., applying authorization to all the descendants of the context node). A set of example rules are shown below :

Example 2. Example ACRs:

R1 : {role1, / site // person / name, read, + , RC}

R2 : {role1, / site / regions / asia / item, read, +, RC}

R3 : {role2, / site / regions / * / item, [location= ”USA”] / description, read, +, RC}

2.2.2 Content-Based Query Brokering

Indexing schemes have been proposed for content-based XML retrieval [50]–[53]. The index describes the address of the data server that stores a particular data item requested by an user query. Therefore, a content-based index rule should contain the content description and the address. In [9], we presented a content-based indexing model with index rules in the form of , where (1) object is an XPath expression that selects a set of nodes; and (2) location is a list of IP addresses of data servers that hold the content.

Example 3. Example Index Rules:

I1 : { / site / people / person /name, 130.203.189.2 }

I2 : { /site / America / item / name, 135.38.92.1 }

When an user queries the system, the XPath query is matched with the object field of the index rules, and the matched query will be sent to the data server specified by the location field of the rule(s). While other techniques (e.g., bloom filter) can be used to implement content-based indexing, we adopt the model in [9] in our study since it can be directly integrated with the NFA-based access control enforcement scheme. We call the integrated NFA that captures access control rules and index rules content-based query broker (QBroker).

3. Privacy- Preserving Information Brokering

Privacy protection is need for the Information Brokering System (novel IBS), named Privacy Preserving Information Brokering (PPIB). PPIB has two type of brokering Component: (1) brokers and (2) co-ordinators. The brokering are mainly responsible for user authentication and query forwarding, the broker performs the role who can act between the Co-coordinator and the data Users. The request which is all submitted from the data user will be verified and thus it will be passed to the co-coordinator. The coordinators which are linked in a tree structure enforce access control and query routing based on the embedded nondeterministic finite automata also known as query brokering automata. The coordinators, each holding a segment of access control automaton and routing guidelines, are mainly responsible for access control and query routing. [8]

PPIB takes an innovator 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.

To prevent inquisitive or unserviceable coordinators from inferring private information, we design two novel schemes: (a) to segment the query brokering automata, and (b) to encrypt corresponding query segments. System will providing full capability to wage in network access control and to path queries to the right data sources, these two schemes ensure that inquisitive or unserviceable coordinator is not capable to collect sufficient information to guess privacy, like "which data need to be queried, where located and what are the policies to access data". Privacy Preserving Information Brokering (PPIB) enables wide-ranging security and privacy protection for claimed information brokering, with minor overhead and major scalability.

4. Security And Privacy Need For Ppib

In information brokering scenario, there are three types of entrepreneur, namely data owners, data providers, and data requestors. Each entrepreneur has its own privacy: (1) the privacy of a data owner (e.g. a patient) is identifiable data and the information keep together by this data (e.g. medical records). Data owners usually sign stiff privacy agreements with data providers to protect their privacy from unauthorized disclosure/user. (2) Data providers store collected data, and create two types of metadata, namely routing metadata and access control metadata. (3) Data requestors divulge identifiable and private information in the querying process. For example, a query process about AIDS or DNA treatment reveals the (possible) disease of the requestor.

Assume that for the brokers, two types of enemy, outside attackers and curious or corrupted brokering components. Outside attackers passively eavesdrop communication channels. Curious or corrupted brokering components follow the protocols be seemingly to accomplish their functions, others' private information from the information disclosed in the querying process.

Data providers push routing and access control metadata to brokers [8], which also strut queries from requestors. Therefore, a curious or corrupted brokering server could: (1) learn query content and query location by impede a local query; (2) learn routing metadata and access control metadata from local data servers and other brokers; (3) learn data location from routing metadata it holds Although attacker may not obtain plaintext data over encrypted data, they can still learn query location and data location from eavesdrop. The attacks into two major classes: (1) the attribute-correlation attack and (2) inference attack.

Attribute-correlation attack: An attacker prevents a query, which typically contains several predicates. Each predicate describes a condition, which sometimes involves sensitive and private data (e.g. name, credit card number, etc.).

Inference attack: Attacker some techniques and result more than one other type of sensitive information so more sever, and further associates to learn explicit and implicit knowledge about entrepreneur

IBS work is designed with user and data privacy. Such privacy protection requirements, therefore a novel IBS, named as Privacy Preserving Information Brokering system (PPIB). As shown in Figure, PPIB contains a broker-coordinator overlay network, in which the brokers are amenable for onus transmission 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.

5. Architecture Of PPIB

PPIB has three types of brokering components: (1) Brokers (2) Coordinators and (3) Central authority (CA). The key to defend privacy is to part the work on more than one components in such a way that more than one node can make a meaningful presumption from the information disclosed to it. Figure 2 shows the architecture of PPIB. Through local brokers (green nodes in Fig) Data servers and requestors from different organizations connect to the system.

Brokers: It is intercommunicating through coordinators (white nodes in Fig). A local broker functions as the “entry” to the system. It’s responsible for authenticates requestors and hides their. It would also permute query sequence to defend against local traffic analysis.

Coordinators: It is responsible for content-based query routing and access control actuation. With privacy-preserving idea, coordinator cannot hold any rule in the complete form. Instead, a novel automaton segmentation scheme to divide (i.e. metadata) rules into segments and assign each segment to a coordinator. Coordinators operate collaboratively to enforce secure query routing.

Coordinator prevents from sensitive predicates, a query segment encryption scheme and automaton segmentation scheme, query divide into segment and encrypt it (each segment).

Central Authority (CA) : It is responsible for key management and metadata maintenance

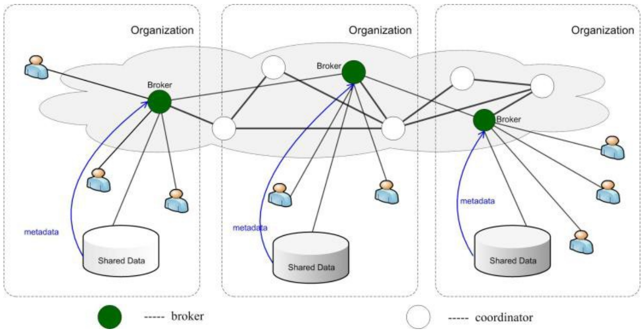

Fig.2. Architecture of PPIB

The Architecture of the privacy preserving brokering system is shown in Fig. 2, where users and data servers of more than one organizations are communicate via a Broker, coordinator overlay component. User requests for data by sending a XML query to the local broker, which further carry the query to the root of the coordinator tree. The query is processed along a path of the multiple organizations coordinator. The brokering process consists of 4 phases:

Phase 1: For join the system, a user needs to authenticate to the local broker. And the user submits encrypted segment an XML query by public level keys, and a unique session key Ks, data servers encrypted with the public key, to return data.

Phase 2: The major task of the broker is metadata preparation: (1) it extracts the role of the user authenticated and attaches it to the encrypted XML query; (2) it make a unique ID for each query, and attaches QID with its own address (as well as < Ks > pkDS) to the query so that the data server can directly return the data.

Phase 3: When the root of the coordinator tree receives the query and its metadata from a local broker, it follows schemes i.e. the automata segmentation scheme for segment the XML query and the query segment encryption scheme to perform access control and to route the query within the coordinator tree, until it reaches a leaf coordinator, which forwards the query to the related data servers.

Phase 4: In the final phase, the data server gets a safe query in an encrypted form. The data server evaluates the query and returns the data after decryption, encrypted by Ks, to the broker of the query.

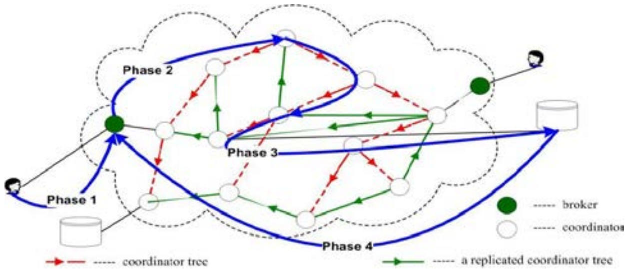

Fig. 3. Query brokering process in 4 phases.

6. APPLICATIONS

Information (Data) Brokers collect data and provide data mining services for various organizations, for example in the FBI, Credit Monitoring Services, DoD, etc. The companies are a high value target for social engineers as they contain huge amounts of information that could be used to further elevate. Because of relaxed regulations and federal laws much of our personal information is collected by government agencies and stored or managed by these Information Broker Companies.

Information brokering is suitable for many newly emerged applications, such as information sharing for healthcare or law enforcement, in which organizations share information in ailliberal and controlled manner, not only from business considerations but also due to legal reasons.

Healthcare information systems, such as Regional Health Information Organization (RHIO) [1], to facilitate retrieval of clinical data thereon collaborative health providers.

Law enforcement, for example young police officers, police academics, researchers agencies use information brokering technologies to share on demand data with other agencies and the public.

7. CONCLUSION

Privacy issues of user and data during the design stage is considered and concluded that existing information brokering systems suffer from a spectrum of vulnerabilities associated with user privacy, data privacy, and metadata privacy. In this paper, PPIB proposed architecture is discussed, a new approach to preserve privacy in XML information brokering. By using automaton segmentation scheme, within network access control and query segment encryption, PPIB put together security enforcement and query forwarding at the same time as providing comprehensive privacy protection. We claim that our analysis is very resistant to privacy attacks. Node-to-node query processing performance and system scalability are also evaluated and the results show that PPIB is efficient and scalable.

ACKNOWLEDGEMENT

The authors would like to thank A. Fengjun Li, Bo Luo, PengLiu, Dongwon Lee, and Chao-Hsien Chu for valuable discussions.

References

[1] A. Fengjun Li, Bo Luo, Peng Liu, Dongwon Lee, and Chao-Hsien Chu “Enforcing Secure and Privacy-Preserving Information Brokering in Distributed Information Sharing” IEEE Trans. Information Forensics and Security, Vol. 8, No. 6, June 2013.

[2] 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 ACM CCS ’07, pp. 508–518, 2007.

[3] 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,” Journal of AHIMA 77, pp. 64A–D, January 2006.

[4] D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symposium on Security and Privacy, 2000, pp. 44–55.

[5] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data,” in Proc. ICDCS’10, Genoa, Italy, pp. 253–262. P. F. Syverson, D. M. Goldschlag, and M. G. Reed, “Anonymous con- nections and onion routing,” in Proc. IEEE S&P, 1997, pp. 44–54.

[6] W. Tolone, G.-J. Ahn, T. Pai, and S.-P. Hong, “Access control in col- laborative systems,” ACM Comput. Surv., vol. 37, no. 1, pp. 29–41,2005.

[7] S. Cho, S. Amer-Yahia, L. V. S. Lakshmanan, and D. Srivastava, “Op- timizing the secure evaluation of twig queries,” in Proc. VLDB, 2002, pp. 490–501.

[8] 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 infor- mation brokerage systems,” in Proc. IEEE SUTC, Taichung, Taiwan, 2006, pp. 252–259.