ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 2; Issue 6; Page No. 76-80

Secure Information Brokering In Distributed Information System

Priyanka M. Jamunkar[1], Gayatri M. Bhandari[2],

[1]JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi Pune 412207, India

priyankapurkude@gmail.com

[2]JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi Pune 412207, India

gayatri.bhandari1980@gmail.com

ABSTRACT

To facilitate extensive collaborations, today's organizations raise increasing needs for information sharing via on-demand information access. Information Brokering System (IBS) atop 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. 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 within the IBS. The necessity for the information sharing increases day to day as many organizations use shared data sources to improve their data communication by improving the interoperability among them. To implement the Information sharing, there needs a Broker who acts as the intermediate between the organizations and transfer data between them. In this article, we study the problem of privacy protection in information brokering process. We first give a formal presentation of the threat models with a focus on two attacks: attribute-correlation attack and inference attack. Then, we propose a broker-coordinator overlay, a well as two schemes, automaton segmentation scheme and query segment encryption scheme, to share the secure query routing function among a set of brokering servers. With comprehensive analysis on privacy, end-to-end performance, and scalability, we show that the proposed system can integrate security enforcement and query routing while preserving system-wide privacy with reasonable overhead.

Key Words: information sharing, privacy, access control, and information brokering system

INTRODUCTION:

Information sharing is becoming increasingly important in recent years, not only among organizations with common or complementary interests but also within large organizations and enterprise that are becoming ever more globalized and distributed. Multiple divisions cooperate within large multi- national enterprise as well. For example, in GM, to maintain a proper stock level of parts, people in supply management division need to check the sale information (of car models) gathered and managed by sales people worldwide. To implement the Information sharing, there needs a Broker who acts as the intermediate between the organizations and transfer data between them. This broker must be a trustable person and should not reveal the shared data to others, but it's difficult to trust a third party easily.

The current system lets the Broker to make the routing decisions to direct the client query to the data servers, here the entire query content visible to the broker. In the context of sensitive data and autonomous data providers, a more practical and adaptable solution is to construct a data-centric overlay consisting of data sources and a set of brokers that make routing decisions based on the content of the queries. Such infrastructure builds up semantic-aware index mechanisms to route the queries based on their content, which allows users to submit queries without knowing data or server location. In previous study, such a distributed system providing data access through a set of brokers is referred to as Information Brokering System (IBS).

As shown in Figure 1, applications atop IBS always involve some sort of consortium among a set of organizations. Databases of different organizations are connected through a set of brokers, and metadata (e.g. data summary, server locations) are "pushed" to the local brokers, which further "advertise" (some of) the metadata to other brokers. Queries are sent to the local broker and routed according to the metadata until

reaching the right data server(s). In this way, a large number of information sources in different organizations are loosely federated to provide a unified, transparent, and on-demand data access.

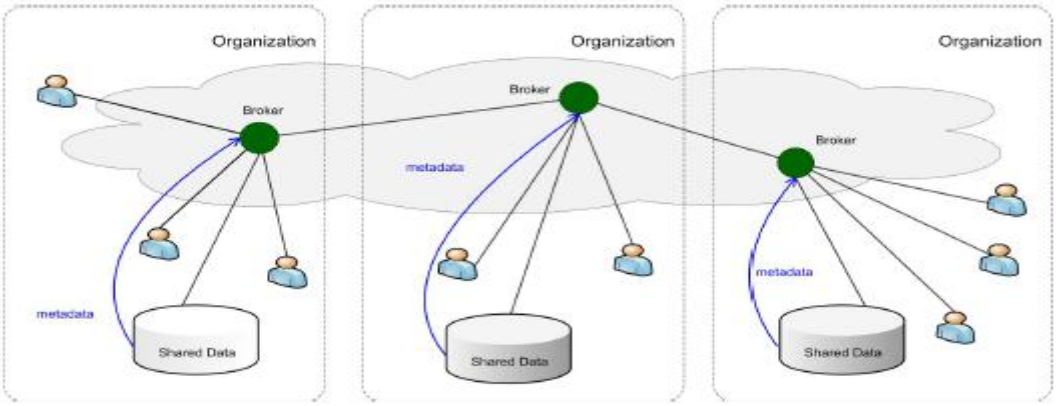

Figure 1: An overview of the IBS Infrastructure.

While the IBS approach provides scalability and server autonomy, privacy concerns arise, as brokers are no longer assumed fully trustable the broker functionality may be outsourced to third-party providers and thus vulnerable to be abused by insiders or compromised by outsiders.

1. VULNERABILITIES AND ATTACKS:

Information Brokering System (IBS) over a peer-to-peer connection has been proposed to support information sharing among loosely associated data sources. It Consist of various data sources and brokering component. These components help client queries to locate the data servers. The information broker deal with the IBS systems and are responsible for providing and processing the data between heterogeneous entities, hence they are also called as the Data Broker. In real world, directly or indirectly everybody get affected with such Information Brokering System as the Information Broker collecting information about consumers from a variety of public and non-public sources such as website cookies etc. and sell them to business who want to target their advertisement and special offers.

In information brokering system, there are three types of stakeholders:

A. Data owner

B. Data provider

C. Data requestor

All stakeholders have its own privacy.

A. Data owner: - The privacy of a data owner (e.g. a patient in RHIO) is the identifiable data and sensitive or personal information carried by this data (e.g. medical records). Data owners usually sign strict privacy agreements with data providers to prevent unauthorized use or disclosure.

B. Data provider: - Data providers store the collected data locally and create two types of metadata, namely routing metadata and access control metadata, for data brokering. Both types of metadata are considered privacy of data provider.

C. Data requestor: - Data requestors may reveal identifiable or private information (e.g., information specifying her interests) in the querying content. For example, a query about AIDS treatment reveals the (possible) disease of the requestor.

A. Threats: - The privacy threats arise in he DIBS system is as follows:

1. User Privacy: - User location an easily retrieved by analyzing the IP packet of the query. User identity is a key concern of the privacy, which can be obtained either from authentication process or by associating user location information with other public data. The “what" privacy may not be known directly, but some reasonable inference from the content of the query can be made. Although the user identity, user location, and query content are privacy-sensitive matters, one cannot apply popular privacy preserving techniques directly in the DIBS. This is because a broker needs to learn this privacy-entities information of full query brokering.

2. Data Privacy: - In DIBS, data owner collect data independently and manage it with autonomous data servers. While providing data access to legitimate user, data server has to release certain privacy-entities information that needs to be protected. In general, two questions, “where is the data stored?" and “who stores

what data?", can express privacy concerns of data. The first question concerns data location privacy, and the second question, denoted data object distribution privacy, inquires which type of data is contained in a particularly data server.

3. Metadata Privacy: - Two type of metadata are brokering involves in DIBS, query indexing guideline and access control. The format describes where the data objects are distributed among all the data servers, and the latter assign accessibility to legitimate user according to access control policy provided by data owner. Risk rises when unsecured or dishonest brokering componentry to abuse or leak this privacy-entities information. In existing DIBS approach, a compromised broker can obtain data location information from indexing guideline or access control policy since these information are stored in broker to facilitate routing and access control.

B. Attacks: -

1. Attribute-correlation attack: - Predicates of an XML query describe conditions that often carry sensitive and private data (e.g., name, SSN, credit card number, etc.) If an attacker intercepts a query with multiple predicates or composite predicate expressions, the attacker can "correlate" the attributes in the predicates to infer sensitive information about data owner. This is known as the attribute correlation attack.

Example 1: A tourist Anne is sent to ER at California Hospital. Doctor Bob queries for her medical records through a Medicare IBS. Since Anne has the symptom of leukemia, the query contains two predicates: [pName="Anne"], and [symptom="leukemia"]. Any malicious broker that has helped routing the query could guess, "Anne has a blood cancer" by correlating the two predicates in the query.

2. Inference attack: - More severe privacy leak occurs when an attacker obtains more than one type of sensitive information and learns explicit or implicit knowledge about the stakeholders through association. By "implicit", we mean the attacker infers the fact by "guessing". For example, an attacker can guess the identity of a requestor from her query location (e.g., IP address). Meanwhile, the identity of the data owner could be explicitly learned from query content (e.g., name or SSN). Attackers can also obtain publicly available information to help his inference. For example, if an attacker identifies that a data server is located at a cancer research center, he can tag the queries as "cancer-related".

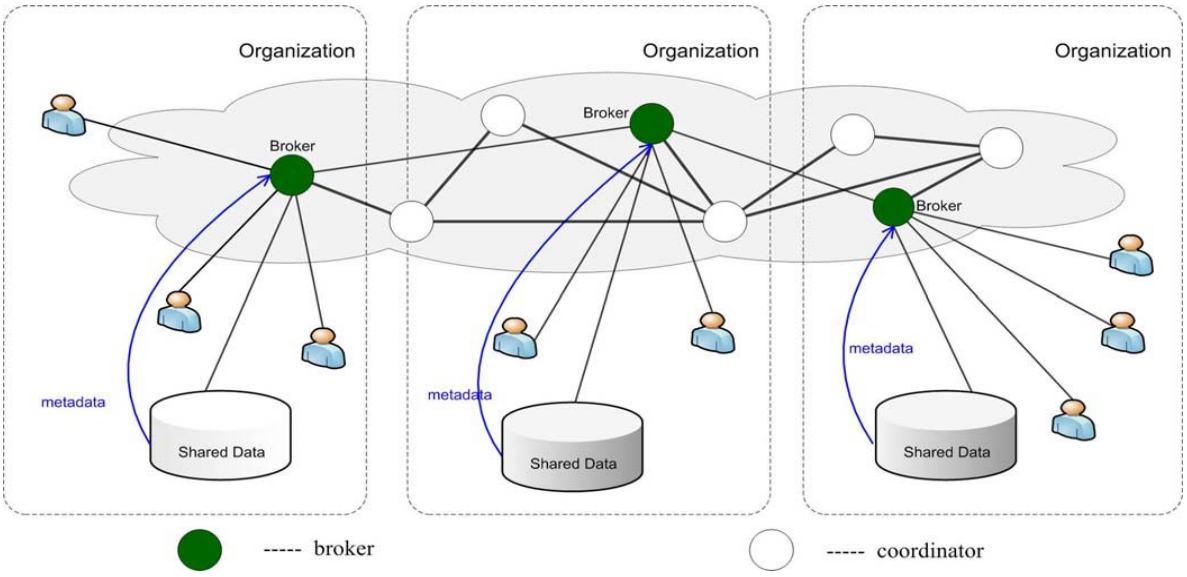

Figure 2: Architecture of PPIB

Figure 2 shows the architecture of PPIB i.e. Privacy Preserving Information Brokering. Data servers and requestors from different organizations connect to the system through local brokers (i.e., the green nodes in Fig. 2). Brokers are interconnected through coordinators (i.e., the white nodes). A local broker functions as the "entrance" to the system. It authenticates the requestor and hides his identity from other PPIB components. It would also permute query sequence to defend against local traffic analysis.

2. PRIVACY-PRESERVING QUERY BROKERING SCHEME:

The QBroker approach has severe privacy vulnerability as discussed. If the Broker is compromised or cannot be fully trusted, the privacy of both requestor and data owner is

under risk. To solve the problem, present the PPIB infrastructure with two core schemes are prescribed. In privacy–preserving query brokering scheme, first explains the details of automata segmentation and query segment encryption schemes

A. Automaton Segmentation: -

In distributed information brokering, multiple organizations join a consortium and agree to share the data within the consortium. Different organizations may have different schemas; we assume a global schema exists by aligning and merging the local schemas. The key idea of automaton segmentation scheme is to logically divide the global automaton into multiple independent yet connected segments, and physically distribute the segments onto different brokering components, known as coordinators. The access control rules and index rules for all the organizations can be crafted following the same shared schema and captured by a global automaton.

Algorithm 1 the automaton segmentation algorithm:

deploySegment()

Input: Automaton State

Output: Segment Address: *addr*

1: **for each** symbol *K* in *S.StateTransTable* **do**
2: *addr = deploySegment (S.StateTransTable(k)).nextState)*
3: *DS = createDummyAcceptState();*
4: *DS.nextState ←addr*
5: *S.StateTransTable(k).nextState ←DS*
6: **end for**
7: *Seg=createSegment()*
8: *Seg.addSegment(S)*
9: *Coordinator=get Coordinator()*
10: *Coordinator.assignSegment(Seg)*
11: **return** *Coordinator.address*

1. Segmentation: - The atomic unit in the segmentation is an NFA state of the original automaton. Each segment is allowed to hold one or several NFA states. Then further define the granularity level to denote the greatest distance between any two NFA states contained in one segment. Reserve the logical connection between the segments after segmentation, and define the following heuristic segmentation rules: (1) NFA states in the same segment should be connected via parent-child links; (2) sibling NFA states should not be put in the same segment without their parent state; and (3) the "accept state" of the original global automaton should be put in separate segments.

2. Deployment: - This employs physical brokering servers, called coordinators, to store the logical segments. To reduce the number of needed coordinators, several segments can be deployed on the same coordinator using different port numbers. The coordinator holding the root state of the global automaton is the root of the coordinator tree and the coordinators holding the accept states are the leaf nodes.

3. Replication: - Since all the queries are supposed to be processed first by the root coordinator, it becomes a single point of failure and a performance bottleneck. It adopt the passive path replication strategy to create the replicas for the coordinators along the paths in the coordinator tree, and let the centralized authority to create or revoke the replicas.

4. Handling the predicates: - To handle the predicates, either from the query or from the ACR, the original strategy is lookup-and-attach. That is, if an XPath step in the query matches a child state in the state transition table (i.e., an eSymbol), predicate carried in that particular XPath step or predicate stored in the predicate table will be attached to the corresponding XPath step in the safe query.

B. Query Segment Encryption: -

Informative hints can be learned from query content, so it is critical to hide the query from irrelevant brokering servers. However, in traditional brokering approaches, it is difficult, if not impossible, to do that, since brokering servers need to view query content to fulfill access control and query routing. Fortunately, the automaton segmentation scheme provides new opportunities to encrypt the query in pieces and only allows a coordinator to decrypt the pieces it is supposed to process. The query segment encryption scheme proposed in this work consists of the pre-encryption and post-encryption modules.

To address the privacy vulnerabilities in current information brokering infrastructure, we propose a new model, namely Privacy Preserving Information Brokering (PPIB). PPIB has three types of brokering components: broker coordinators, and a central authority (CA). Fig. 2 shows the architecture of PPIB. Data servers and requestors from different organizations connect to the system through local brokers (i.e., the green nodes in Fig. 2). Brokers are interconnected through coordinators (i.e., the white nodes).

3. PRIVACY AND SECURITY ANALYSIS:

In information brokering system, there are different types of attackers. According to their roles, we have eavesdroppers and active attackers that can compromise any brokering server from their cooperation mode. In this section, we consider three most basic types of attackers, local and global eavesdroppers, malicious brokers and malicious coordinators.

1. Eavesdroppers: - A global eavesdropper is an attacker who observes the traffic in the entire network. It watches brokers and coordinators gossip, so it is capable to infer the locations of local brokers and root-coordinators. This is because the assurance of the connections between user and broker, and between broker and root-coordinator.

2. Single Malicious Broker: - A malicious broker deviates from the prescribed protocol and discloses sensitive information. It is obvious that a corrupted broker endangers user location privacy but not the privacy of query content.

3. Collusive Coordinators: -Collusive coordinators deviate from the prescribed protocol and disclose sensitive information. Consider a set of collusive (corrupted) coordinators in the coordinator tree framework.

4. CONCLUSION:

In existing information brokering systems privacy of user data and metadata which existing system fails to do so. In this paper, we proposed PPIB to protect location of data requestors and data servers from irrelevant or malicious parties. The PPIB is working efficiently against attacks and threats that are prevalent to its functioning with the methods like automaton segmentation and query segment encryption. The PPIB is new approach to preserve privacy XML information brokering.

5. REFERENCES:

1. 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.
2. W. Bartchat, 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.
3. Fengjun Li, Bo Luo, Peng Liu Dongwon Lee and Chao-Hsien Chu, "Enforcing Secure and Privacy-Preserving Information Brokering in Distributed Information Sharing" IEEE 2013.
4. P. Sheth and J. A. Larson, "Federated database systems for managing distributed, heterogeneous, and autonomous databases,"ACM Computing Surveys (CSUR), vol. 22, no. 3, pp. 183–236,1990.
5. M. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces: A new information management," SIGMOD Rec., vol. 34, no. 4, pp. 27–33, 2005.
6. 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.