International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Privacy Preserving and Information Security Forensics Brokering

Sandhya N
GSS Institute of Technology
Visvesvaraya Technological University
Bangalore, India

Krishna Gudi
GSS Institute of Technology
Visvesvaraya Technological University
Bangalore, India

Sunanda Allur
GSS Institute of Technology
Visvesvaraya Technological University
Bangalore, India

***Abstract*—Information Brokering as the "business of buying and Selling information as a commodity". Information brokering systems (IBSs) connect large-scale loosely federated data sources via a brokering overlay, in which the data brokers make routing decision to direct client queries, based upon the queries details provided by client to the requested data servers. Many existing IBSs assume that brokers are trusted for the service and thus only adopt server-side access control for data confidentiality. Also privacy of data location and data consumer are important and 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. Here a novel approach to preserve privacy of multiple stakeholders involved in the information brokering task and formally defining two privacy attacks, namely attribute-correlation attack and inference attack, and propose two countermeasure schemes automaton segmentation and query segment encryption to securely share the routing in Information Brokering Technology.**

***Keywords*—Attack,Encryption,Segmentation.**

I. INTRODUCTION

An Information broker, also known as an independent information professional, or information consultant, is a person or business that researches information for clients. Here data brokers (information consultants) create a personal profile for each application, individual users in their organization. Common uses for information brokers include market research and patent searches, but can include practically any type of information research. Information (data) broker are nothing but person or firm who locates and resells secondary (already published or otherwise available) information's such as articles, citations, competitor data, research data).

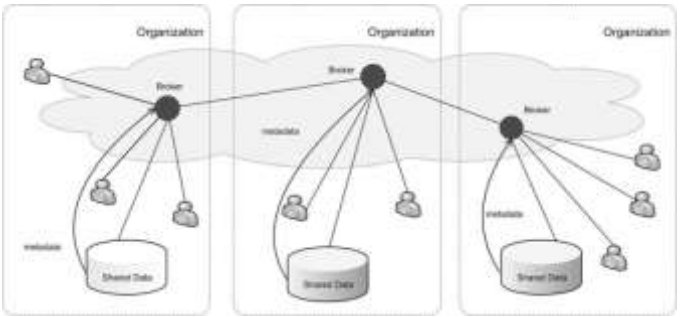

Fig1: The IBS Infrastructure.

Fig1: Consisting of data sources and a set of data brokers that make routing decisions based on the content of the queries.

In distributed system providing data access through a set of brokers know as Information brokering systems (IBSs), Databases of different organizations are connected through a set of brokers, and metadata (e.g., information summary, server locations, data consumer) 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).

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. All the responsibility is undertaken by single broker component for brokering process. Hence broker is the targeted person to attackers or for hacking broker processing.

Data consumers, Data providers and so on need their data that is shared in brokering process sort of privacy over distributed environment.

Threat Model: Existing security mechanisms focusing on confidentiality and integrity cannot preserve privacy effectively. For instance, while data is protected over encrypted communication, external attackers still learn query location and data location from eavesdropping. Combining types of unintentionally disclosed information, the attacker could further infer the privacy of different stakeholders through attribute-correlation attacks and inference attacks.

Attribute Correlation Attack: The 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: Mr.Ami is sent to ER at California Hospital. Doctor Sham queries for her medical records through a medicare IBS. Since Ami has the symptom of cancer, the query contains two predicates: [pName="Ami"], and [symptom="cancer"]. Any malicious broker that has helped routing the query could guess "Ami has leukemia" by correlating the two predicates in the query. Unfortunately, query content including sensitive predicates cannot be simply

encrypted since such information is necessary for content-based query routing. Therefore, we are facing a paradox of the requirement for content-based brokering and the risk of attribute-correlation attacks.

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 Credit card details). 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 leukemia research center, he can tag the queries as "leukemia -related".

II. RELATED WORK

A. *Surveying the RHIO landscape: A description of current {RHIO} models, with a focus on patient identification.*

Regional Health Information Organizations (RHIOs) is a group of organizations and stakeholders that exchanges data electronically to improve the quality, safety, and efficiency of healthcare delivery. A Regional Health Information Organization (RHIO, pronounced rio), also called a Health Information Exchange Organization, is a multi-stakeholder organization created to facilitate a health information exchange (HIE) – the transfer of healthcare information electronically across organizations among stakeholders of that region's healthcare system.

The main goals in sharing patient-specific data are to:

1. Improve healthcare delivery by providing immediate, secure, confidential exchange of health information between authorized users.
2. Enable providers and patients to make decisions based on near real-time access to health information.
3. Provide warning and reminders at point of care.
4. Reduce medical errors.
5. Prevent adverse drug reactions.
6. Encourage participation of patients in their own healthcare and chronic disease management.

Accurate patient identification and linking is the foundation of health technology that is implemented in a RHIO or any similar network that shares patient information. Without accurate patient identification, patient safety and quality of care are compromised. When high percentages of duplication or overlaying of records occurs in electronic health record databases, physician trust in the system is lost.

B. *Peer-to-peer management of XML data: Issues and research challenges*

Peer-to-peer (p2p) systems are attracting increasing attention as an efficient means of sharing data among large, diverse and dynamic sets of users. The widespread use of XML as a standard for representing and exchanging data in the Internet suggests using XML for describing data shared in a p2p system. However, sharing XML data imposes new challenges in p2p systems related to supporting advanced querying beyond simple keyword-based retrieval that focuses on data management issues for processing XML data in a p2p setting, namely indexing, replication, clustering and query routing and processing. For each of these topics, we present the issues that arise, survey related research and highlight open research problems. XML has evolved as the new standard for the representation and exchange of semi structured data on the Internet.

Peer-to-peer characteristics

1. Clustering

Data clustering refers to grouping data items together to form clusters (groups) of items with common attributes or properties.

2. Replication

The goals of replication in a p2p system do not differ much from those in a non p2p distributed system. Replication is basically used to improve system performance and to increase data availability in case of peer failures.

3. Query processing

Query processing in traditional distributed systems can be divided into four phases: query decomposition, data localization, global and local query optimization.

III. PROPOSED METHOD

First, to address the need for privacy protection, proposing a novel IBS, namely Privacy Preserving Information Brokering (PPIB). It is an overlay infrastructure consisting of two types of brokering components, brokers and coordinators. Assuming central authority for managing the key and handles metadata maintenance. To privacy vulnerabilities in current information brokering infrastructure, the key to preserving privacy is to divide and allocate the functionality to multiple brokering components in a way that no single component can make a meaningful inference from the information disclosed to it. Fig. 2 shows the architecture of PPIB. Data servers and requestors from different organizations connect to the system through local brokers (i.e., the dark nodes in Fig. 2). Brokers are interconnected through coordinators (i.e., the white nodes).

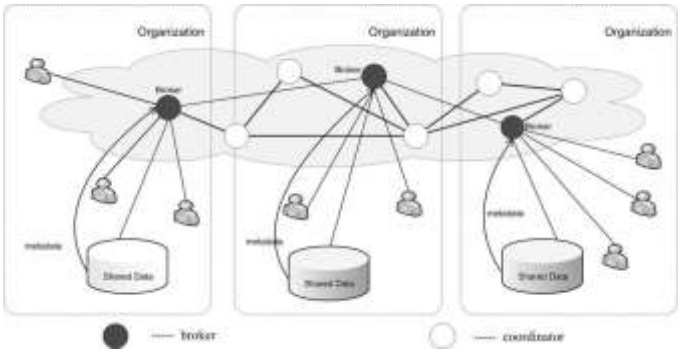

Fig 2: Architecture of Privacy Preserving Information Brokering.

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. Coordinators are responsible for content-based query routing and access control enforcement. With privacy-preserving considerations, we cannot let a coordinator hold any rule in the complete form.

Instead, we propose novel automaton segmentation scheme to divide (metadata) rules into segments and assign each segment to a coordinator. Coordinators operate collaboratively to enforce secure query routing. A query segment encryption scheme is further proposed to prevent coordinators from seeing sensitive predicates. The scheme divides a query into segments, and encrypts each segment in a way that to each coordinator enroute only the segments that are needed for secure routing is revealed. Section A discusses more on the important scheme and Section B discusses on Privacy and Security Analysis

A Privacy Preserving Query Brokering Scheme

If the Broker is compromised or cannot be fully trusted (e.g., under the honest-but-curious assumption as in our study), the privacy of both requestor and data owner is under risk. To tackle the problem, present the PPIB infrastructure with two core schemes namely automaton segmentation and query segment encryption.

1) Automaton Segmentation: In the context of distributed information brokering, multiple organizations join a consortium and agree to share the data within the consortium. While different organizations may have different schemas, we assume a global schema exists by aligning and merging the local schemas. Thus, 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. 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.

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, define the granularity level to denote the greatest distance between any two NFA states contained in one segment Given a granularity level k, for each segmentation, the next states will be divided into one segment with a probability .Obviously, with a larger granularity level, each segment will contain more NFA states, resulting in less segments and smaller end-to-end over- head in distributed query processing. However, a coarse partition is more likely to increase the privacy risk. The trade-off between the processing complexity and the degree of privacy should be considered in deciding the granularity level. As privacy protection is of the primary concern of this work, we suggest a granularity level≤2.To reserve the logical connection between the segments after segmentation, we 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. To ensure the segments are logically connected, we also make the last states of each segment as "dummy" accept states, with links pointing to the segments holding the child states of the original global automaton.

2) 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 preencryption and post encryption modules, and a special commutative encryption module for processing the double-slash ("//") XPath step in the query.

Level-Based Preencryption: According to the automaton segmentation scheme, query segments are processed by a set of coordinators along a path in the coordinator tree. A straightforward way is to encrypt each query segment with the public key of the coordinator specified by the scheme. Hence, each coordinator only sees a small portion of the query that is not enough for inference, but collaborating together, they can still fulfill the designed function. The key challenges in this approach is that the segment-coordinator association is unknown beforehand in the distributed setting, since no party other than the CA knows how the global automaton is segmented and distributed among the coordinators.

B. Privacy and Security Analysis

There are three types of attackers in the information brokering process eavesdropper, Single Malicious Broker, Collusive Coordinators.

Eavesdropper: A local eavesdropper is an attacker, who can observe all communication to and from the user side.

A global eavesdropper is an attacker who observes the traffic in the entire network.

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. Moreover, since the broker knows the root-coordinator locations, the threat is the disclosure of root-coordinator location and potential DoS attacks.

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. Even though each coordinator can ob- serve traffic on a path routed through it, nothing will be exposed to a single coordinator because (1) the sender viewable to it is always a brokering component;(2) the content of the query is incomplete due to query segment encryption; (3) the ACR and indexing information are also incomplete due to automaton segmentation; (4) the receiver viewable to it is likely to be another coordinator. However, privacy vulnerability exists if a coordinator makes reasonable inference from additional knowledge. For instance, if a leaf-coordinator knows how PPIB mechanism works, it can assure its identity (by checking the automaton it holds) and find out the destinations attached to this automaton are of some data servers. Another example is that one coordinator can compare the segment of ACR it holds with the open schemas and make reasonable inference about its position in the coordinator tree. However, inference made by one coordinator may be vague and even misleading.

CONCLUSION

Information Brokering Technology is one of the approach where a set of brokers, servers and clients will communicate each other, in this communication process different types of attacks like attribute correlation attack and inference attack are found, to overcome this a novel approach called Privacy Preserving Information Brokering found where adopts a important privacy preserving query brokering scheme are automaton segmentation and query segment encryption. Also different privacy and security analysis are done in information brokering. To minimize (or even eliminate) the participation of the administrator node, A main goal is to make PPIB self reconfigurable, Securable.

REFERENCES

[1] 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.

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

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

[4] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, "CoolStreaming/DONet: A data-driven overlay network for efficient live media streaming," in *Proc. IEEE INFOCOM*, Miami, FL, USA, 2005, vol. 3, pp. 2102–2111.

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

[6] G. Koloniari and E. Pitoura, "Peer-to-peer management of XML data: Issues and research challenges," *SIGMOD Rec.*, vol. 34, no. 2, pp.6–17, 2005.