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. 5, May 2014, pg.978 – 982

RESEARCH ARTICLE

A New Approach to Preserve Privacy in Distributed Information Sharing Using Privacy Preserving Information Brokering

Vijaymahanthesh K.H[1], T.R. Muhibur Rahman[2]

[1]M.Tech scholar, Department of Computer Science and Engineering, Ballari Institute of Technology and Management, Bellary, Karnataka, India

[2]Associate Professor, Department of Computer Science & Engineering, Ballari Institute of Technology and Management, Bellary, Karnataka, India

[1] mahanthesh.mlk@gmail.com; [2] muhibr19@gmail.com

Abstract—Information sharing among organizations has been increased these days. To connect large scale loosely- federated data sources, information brokering system (IBSs) have been introduced. In this system, the brokers make routing decisions to direct the client queries to the requested data servers. Many existing IBSs assume that brokers are trusted and thus only adopt server-side access control for data confidentiality. Yet, 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 concentration has been set on its protection. This paper presents two countermeasure schemes automaton segmentation and query segment encryption schemes to preserve the privacy of multiple stakeholders involved in the information brokering process.

Keywords —Information Sharing; Privacy; Automaton Segmentation; Query Segment Encryption; Data Confidentiality

I. INTRODUCTION

In recent years information sharing is becoming increasingly among Organizations in various areas ranging from business to government agencies, there is an increasing need for inter-organizational information sharing to facilitate extensive collaboration. A number of information systems have been developed to provide efficient and secure information sharing [1].The difficulty of balancing peer autonomy and system coalition is still challenging.

Most of the existing system work on two extremes of the spectrum, 1) adopting either the query-answering model to establish pair wise client-server connections for on-demand information access, where peers are fully autonomous but there lacks system wide coordination, or 2) the distributed database model, where all peers with little autonomy are managed by a unified DBMS.

Unfortunately, model is suitable for many newly emerged applications, such as healthcare or law enforcement information sharing. Take healthcare information systems as example. Regional Health Information Organization (RHIO) [2] aims to facilitate access to and retrieval of clinical data across collaborative healthcare providers that include a number of regional hospitals, outpatient clinics, payers, etc. As a data provider, a participating organization would not assume free or complete sharing with others, since its data is legally private or commercially proprietary, or both. Instead, it requires to retain full control over the data and the access to the

data. Meanwhile, as a consumer, a healthcare provider requesting data from other providers expects to preserve her privacy (e.g., identity or interests) in the querying process.

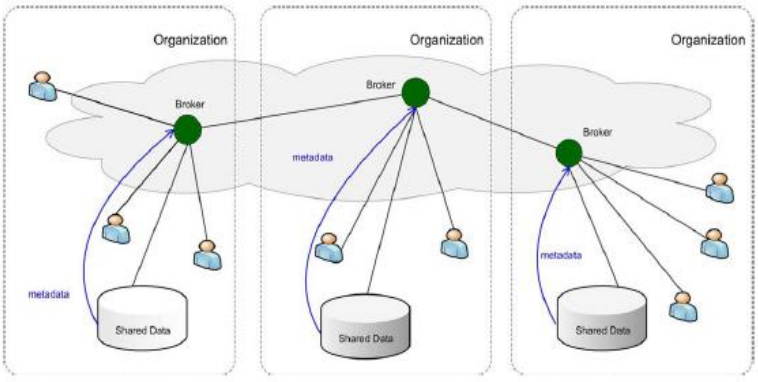

Fig.1. Overview of the IBS infrastructure.

A distributed system providing data access through a set of brokers is referred to as Information Brokering System (IBS). As shown in Fig. 1,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.

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.

II. PRIVACY ATTACKS

The attacker could further infer the privacy of different stakeholders through attribute-correlation attacks and inference attacks.

A. 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.

B. 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 centre, he can tag the queries as "cancer-related".

III. RELATED WORK

Information integration, peer-to-peer file sharing systems and publish-subscribe systems provide partial solutions to the problem of large-scale data sharing. Peer-to-peer systems are designed to share files and data sets (e.g., in collaborative science applications). Distributed hash table technology [3], [4] is adopted to locate replicas based on keyword queries. However, although such technology has recently been extended to support range queries [5], the coarse granularity (e.g., files and documents) cannot meet the expressiveness needs of applications focused in this work. Furthermore, a P2P system often returns an incomplete set of answers while we need to locate all relevant data in the IBS.

Addressing a conceptually dual problem, XML publish-subscribe systems (e.g., [6], [7]) are probably the closely related technology to the proposed research problem: while PPIB aims to locate relevant data sources for a given query and route the query to these data sources, the pub/sub systems locate relevant consumers of a given document and route the document to these consumer.

One idea is to build an XML overlay architecture that supports expressive query processing and security checking atop normal IP network. In particular, specialized data structures are maintained on overlay nodes to route XML queries. In [8], a robust mesh has been built to effectively route XML packets by making use of self-describing XML tags and the overlay networks. To share data among a large number of autonomous nodes, [9] studied content-based routing for path queries in peer-to-peer systems. Different from these approaches, PPIB seamlessly integrates query routing with security and privacy protection.

XML has a number of advantages over a byte stream for multicast delivery. First, XML permits the network to interpret client data needs in terms of well-defined XML queries. Second, XML packets suggest what logical units of data will be processed together by' a client and thus can aid network scheduling. Third, many tools and standards exist for XML making it easy for both the data originator and receiver to build robust applications. Finally, our approach allows applications and databases to push part of their processing into the network fabric. We expect that query languages such as XQuery will become standardized, allowing a single language to be used to describe data requirements. This standardization will permit applications to program our network fabric to deliver the data they need in a simple, consistent fashion.

IV. PRIVACY PRESERVING INFORMATION BROKERING SYSTEM

To address the privacy vulnerabilities in current information brokering infrastructure, a new model, namely privacy preserving information brokering (PPIB) has been proposed.

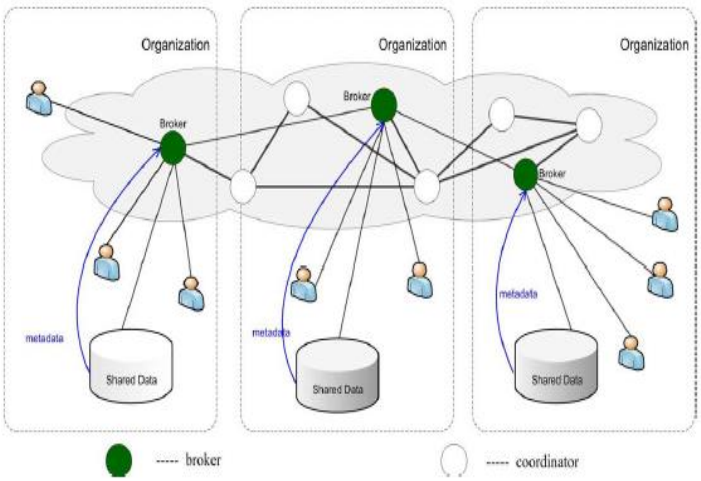

Fig.2. Architecture of PPIB

Fig. 2 shows the architecture of PPIB. It is an overlay infrastructure consisting of two types of brokering components, brokers and coordinators.Brokers are interconnected through coordinators.The brokers, are mainly responsible for user authentication and query forwarding.

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 a 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 en route only the segments that are needed for secure routing are revealed.

A. Automaton Segmentation

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.

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. 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 a 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 not 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.

Algorithm 1: The automaton segmentation algorithm:

deploySegment ()

Input: Automaton State S

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=getCoordinator ()

10:Coordinator.assignSegmen t(Seg)

11: return Coordinator. Address

2) *Deployment:* We employ 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. Therefore, the tuple {coordinator, port} uniquely identifies a segment. After the deployment, the coordinators can be linked together according to the relative position of the segments they store, and thus form a tree structure. 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. Queries are processed along the paths of the coordinator tree in a similar way as they are processed by the global automaton: starting from the root coordinator, the first XPath step (token) of the query is compared with the tokens in the root coordinator. If matched, the query will be sent to the next coordinator, and so on so forth, until it is accepted by a leaf coordinator and then forwarded to the data server specified by the outpointing link of the leaf coordinator. At any coordinator, if the input XPath step does not match the stored tokens, the query will be denied and dropped immediately.

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. For robustness, we need to replicate the root coordinator as well as the coordinators at higher levels of the coordinator tree.

B. Query Segment Encryption

A query segment encryption scheme is 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 are revealed.

V. CONCLUSIONS

Information brokering system was introduced to meet the increasing need for information sharing. To protect the privacy of user, data and metadata, we propose privacy preserving information brokering system (PPIB), through the automaton segmentation scheme and query segment encryption scheme. PPIB integrates security enforcement and query forwarding while providing comprehensive privacy protection.

REFERENCES

[1] Fengjun Li, Bo Luo, Peng Liu Dongwon Lee and Chao-Hsien Chu, "Enforcing Secure and Privacy-Preserving Information Brokering in Distributed Information Sharing", IEEE TRANSCATIONS ON INFORMATION FORENSICS AND SECURITY, 2013.

[2] W.Bartschat,J. BUrrigton-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.

[3] 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," IEEE/ACM Trans. Netw., vol. 11, no. 1, pp. 17–32, Feb. 2003.

[4] 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 Proc. CIDR, 2005, pp. 28–43.

[5] O.Sahin, A.Gupta, D. Agrawal, and A. E. Abbadi, "A peer-to-peer framework for caching range queries," in Proc. ICDE, Boston, MA, USA, 2004, pp. 165–176.

[6] A.Carzaniga, M.J.Rutherford, and A. L.Wolf, "A routing scheme for content-based networking," in Proc. INFOCOM, Hong Kong, 2004, pp. 918–928.

[7] Y.Diao, S. Rizvi, and M. J. Franklin, "Towards an Internet-scale XML dissemination service," in Proc. VLDB Conf., Toronto, Canada, Aug. 2004.

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

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

