ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Decentralized Data Allocation with Load Balancing Mechanism

A. Rajalakshmi, Dr. A. M. J. Md Zubair Rahman, ME, MS, Ph.D.,

II ME (CSE), Al-Ameen Engineering College, Erode, Tamilnadu, India

Principal, Al-Ameen Engineering College, Erode, Tamilnadu, India

Abstract: Mutual data sharing is provided in distributed file sharing environment. Information brokering systems (IBSs) are used to connect large-scale loosely federated data sources via a brokering overlay. Information brokers redirect the client queries to the requested data servers. Privacy preserving methods are used to protect the data location and data consumer. Brokers are trusted to adopt server-side access control for data confidentiality. Query and access control rules are maintained with shared data details under metadata. A Semantic-aware index mechanism is applied to route the queries based on their content and allow users to submit queries without data or server information. Privacy Preserved Information Brokering (PPIB) scheme is used to preserve privacy for distributed data sharing process. Attribute-correlation attack and inference attacks are handled by the PPIB. PPIB overlay infrastructure consisting of two types of brokering components, brokers and coordinators. The broker's acts as mix anonymizer are responsible for user authentication and query forwarding. The coordinators concatenated in a tree structure, enforce access control and query routing based on the automata. Automata segmentation and query segment encryption schemes are used in the Privacy-preserving Query Brokering (QBroker). Automaton segmentation scheme is used to logically divide the global automaton into multiple independent segments. The query segment encryption scheme consists of the preencryption and post encryption modules. The PPIB scheme is improved to support dynamic site distribution and load balancing mechanism. Peer workloads and trust level of each peer are integrated with the site distribution process. The PPIB is improved to adopt self reconfigurable mechanism. Automated decision support system for administrators is included in the PPIB. The system reduces the data distribution overhead and response time for the clients.

I. 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 multinational 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 gathered and managed by sales people world-wide. In such information sharing systems, the data gathered by a specific division are typically stored and maintained in a database *local* to the division, but the needs to access the data may potentially come from any *remote* division. Although the Internet and various virtual private networks provide good data communication links, there are major challenges in (a) achieving scalable, agile and secure remote access of distributed data; (b) handling the heterogeneity among data management systems and data formats which are not always structured and may be incompatible with each other; (c) handling the dynamics of modern business applications; and (d) location discovery.

To tackle these challenges, mediation and federation based information brokering technologies have been proposed. In particular, recent eXtensible Markup Language (XML) has become a promising solution integrating incompatible data while preserving semantics. An XML based information brokerage system comprises data sources and brokers which, respectively, hold XML documents and document distribution information. In such systems, databases can be queried through brokers with no schema relevant or geographical difference being noticed. However, from the security, especially access control, point of view, existing information brokerage systems have a fundamental misconception. That is, they 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 data confidentiality [11]. As a result, access control deployment strategies and the impact of such strategies on end-to-end system performance are neglected by existing systems. In addition, data source side access control deployment is taken-for-granted as the "right" thing to do. In this paper, we challenge this traditional, taken-for-granted access control deployment methodology, and show that query brokering and access control are not two orthogonal issues because access control deployment strategies can have significant impact on the "whole" system's end-to-end performance.

II. LITERATURE REVIEW

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 by exploiting the semantic relationship between schemas of different sources [1]. The PPIB study assumes that a global schema exists within the consortium, therefore, information integration is out of our scope. Peer-to-peer systems are designed to share files and data sets (e.g., in collaborative science applications). Distributed hash table technology is adopted to locate replicas based on keyword queries. However, although such technology has recently been extended to support range queries [8], the coarse granularity cannot meet the expressiveness needs of applications focused in this work. Furthermore, P2P systems often return 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 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 consumers. However, due to this duality, we have different concerns. The pub/sub systems focus more on efficiently delivering the same piece of information to a large number of consumers, while we are trying to route a large volume but small-sized queries to 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 atop normal IP network. In particular, specialized data structures are maintained on overlay nodes to route XML queries. In [5], a robust mesh has been built to effectively route XML packets by making use of self-describing XML tags and the overlay networks. Kouds *et al.* also proposed a decentralized architecture for ad hoc XPath query routing across a collection of XML databases [6]. To share data among a large number of autonomous nodes, [2] 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. Privacy concerns arise in interior ganizational 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, secure index based search schemes 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 [9], however, to the best of our knowledge, all the schemes presented so far only support keyword search based on exact matching. While there are approaches proposed for multidimensional keyword search [10] and range queries, supporting queries with complex predicates or structures (e.g., XPath queries) is still a difficult open problem. In terms of privacy-preserving brokering, another related technique is secure computation [7] 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. 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. 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 [3]. 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 [4], which causes high maintenance and storage costs. In this work, we adopt an NFA-based query rewriting access control scheme proposed, which has a better performance than previous view-based approaches [3].

III. DISTRIBUTED DATA SHARING SCHEMES

Regional Health Information Organization (RHIO) 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 in the querying process. In such a scenario, sharing a complete copy of the data with others or "pouring" data into a centralized repository becomes impractical. To address the need for autonomy, federated database technology has been proposed to manage locally stored data with a federated DBMS and provide unified data access. However, the centralized DBMS still introduces data heterogeneity, privacy, and trust issues. While being considered a solution between "sharing nothing" and "sharing everything", peer-to-peer information sharing framework essentially need to establish pairwise client-server relationships between each pair of peers, which is not scalable in large scale collaborative sharing.

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 our previous study, such a distributed system providing data access through a set of brokers is referred to as *Information Brokering System* (IBS). Applications atop IBS always involve some sort of consortium (e.g., RHIO) 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 an unified, transparent, and on-demand data access. 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.

In this article, we present a general solution to the privacy-preserving information sharing problem. First, to address the need for privacy protection, we propose a novel IBS, namely *Privacy Preserving Information Brokering* (PPIB). It is an overlay infrastructure consisting of two types of brokering components, *brokers* and *coordinators*. The brokers, acting as mix anonymizer, are 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 nondeterministic finite automata—the *query brokering automata*. To prevent curious or corrupted coordinators from inferring private information, we design two novel schemes to segment the query brokering automata and encrypt corresponding query segments so that routing decision making is decoupled into multiple correlated tasks for a set of collaborative coordinators. while providing integrated in-network access control and content-based query routing, the proposed IBS also ensures that a curious or corrupted coordinator is not capable to collect enough information to infer privacy, such as "which data is being queried", "where certain data is located", or "what are the access control policies", etc. Experimental results show that PPIB provides comprehensive privacy protection for on-demand information brokering, with insignificant overhead and very good scalability.

Privacy Preserved Information Brokering (PPIB) scheme is used to preserve privacy for distributed data sharing process. Attribute-correlation attack and inference attacks are handled by the PPIB. PPIB overlay infrastructure consisting of two types of brokering components, brokers and coordinators. The brokers acts as mix anonymizer are responsible for user authentication and query forwarding. The coordinators concatenated in a tree structure, enforce access control and query routing based on the automata. Automata segmentation and query segment encryption schemes are used in the Privacy-preserving Query Brokering (QBroker). Automaton segmentation scheme is used to logically divide the global automaton into multiple independent segments. The query segment encryption scheme consists of the preencryption and post encryption modules. The following drawbacks are identified in the existing system. They are predefined site distribution, inefficient load balancing mechanism, complex administrator policy model and reconfiguration is not supported.

IV. RSA ALGORITHM

The RSA algorithm is used to secure the user information and query values. The domain name service sensitive attributes are secured using the RSA algorithm. The Rivert, Shamir, Adelman (RSA) scheme is a block cipher in which the Plaintext and cipher text are integers between 0 and n-1 for some n. A typical size for n is 1024 bits or 309 decimal digits.

Key Generation

Select p,q	p and q both prime , $p \neq q$
Calculate n = p x q	
Calculate $\phi(n)=(p-1)(q-1)$	
Select integer e	$\gcd(\phi(n),e) = 1;\ 1 < e < \phi(n)$
Calculate d	$d = e^{-1} \bmod \phi(n)$
Public key	KU = {e, n}
Private key	KR = {d, n}

Encryption

Plaintext	$M < n$
Cipher text	$C = M^e \pmod n$

Decryption

Cipher text	C
Plaintext	$M = C^d \pmod n$

V. INFORMATION BROKERING WITH SITE DISTRIBUTION AND LOAD BALANCING

The PPIB scheme is improved to support dynamic site distribution and load balancing mechanism. Peer workloads and trust level of each peer are integrated with the site distribution process. The PPIB is improved to adopt self reconfigurable mechanism. Automated decision support system for administrators is included in the PPIB. The privacy preserved information brokering system is designed to perform data access under multiple data provider environment. The system performs data querying process using encrypted query model. Server selection and data access management operations are controlled by the brokers and coordinators. The system is divided into five major modules. They are data server, information broker, coordinator, query processing and load balancing process. The data server is designed to maintain the shared data files. Information broker is designed to manage Meta data and user information. The coordinator module is designed to handle data access and query processing. The query processing module is designed to manage user data requests. Load balancing module is designed to distribute data delivery loads.

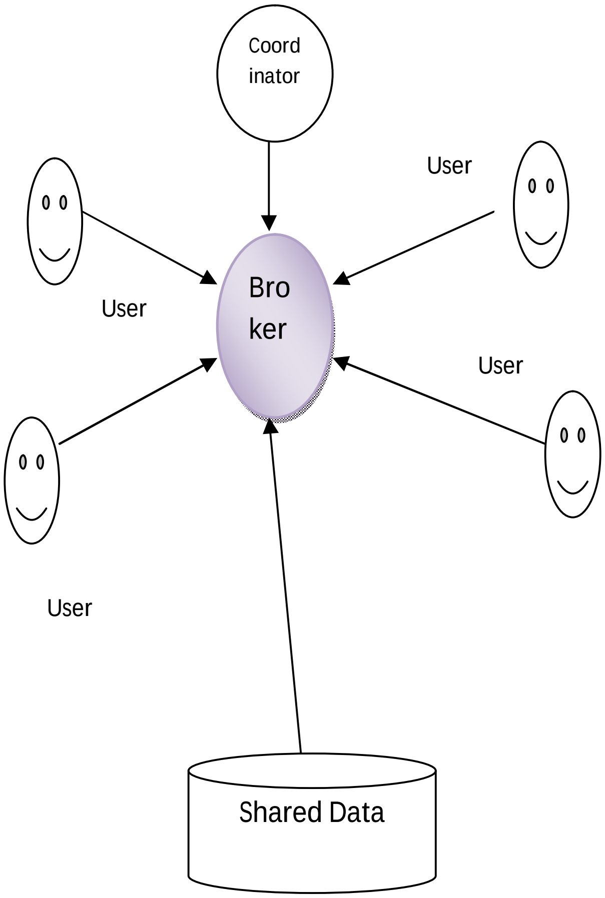

Fig. No: 5.1. Information Brokering with Site Distribution and Load Balancing

The data server provides the shared data to the users. The data values are maintained in encrypted form. Data providers are connected into the information brokers. Data response is prepared by the providers and redirected to the users. Information broker manages the user information and meta data for shared data values. Shared data details are maintained under the meta data environment. User authentication is performed to validate the user requests. Query values are forwarded to the coordinators for site selection process. The coordinator is connected with the broker to perform query processing. Data access control for the user is managed by the coordinator. Encrypted query values are processed under the coordinator to identify the relevant data provider. Query routing is performed with reference to the automata.

The query values are submitted by the users in encrypted format to the information broker. The broker redirects the query values to the coordinator. The data providers are selected by the coordinator and provider information is redirected to the users. Query responses are redirected to the users from the associated providers. The site distribution process is used to manage the request redirection process. Requests are redirected with reference to the server request load and count values. The response load is equally distributed to the servers. Access control verification is carried out for the data providers.

VI. CONCLUSION

Distributed data sharing is performed to share data between organizations. Privacy Preserved Information Brokering (PPIB) scheme is used to provide security and privacy for data access in distributed networks. Load balancing and site distribution schemes are dynamically managed by the system. Self tuning security mechanism is used in the system. The system supports distributed data sharing process. Security and privacy is ensured in the data storage and query process. Data provider load is efficiently handled by the system. The system organization is managed with administrator and historical access details.

REFERENCES

[1] J. Kang and J. F. Naughton, “On schemamatching with opaque column names and data values,” in Proc. SIGMOD, 2003, pp. 205–216.
[2] G. Koloniari and E. Pitoura, “Content-based routing of path queries in peer-to-peer systems,” in Proc. EDBT, 2004, pp. 29–47.
[3] M.Murata, A. Tozawa, andM. Kudo, “XML access control using static analysis,” in Proc. ACM CCS, 2003, pp. 73–84.
[4] T. Yu, D. Srivastava, L. V. S. Lakshmanan, and H. V. Jagadish, “Compressed accessibility map: Efficient access control for XML,” in Proc. VLDB, China, 2002, pp. 478–489.
[5] A. C. Snoeren, K. Conley, and D. K. Gifford, “Mesh-based content routing using XML,” in Proc. SOSP, 2001, pp. 160–173.
[6] N. Koudas, M. Rabinovich, D. Srivastava, and T. Yu, “Routing XML queries,” in Proc. ICDE'04, 2004, p. 844.
[7] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in Proc. STOC'09, Bethesda, MD, USA, pp. 169–178.
[8] 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.
[9] M. Bellare, A. Boldyreva, and A. O'Neill, “Deterministic and efficiently searchable encryption,” in Proc. CRYPTO'07, Santa Barbara, CA, USA, pp. 535–552.
[10] M. Li, S. Yu, N. Cao, and W. Lou, “Authorized private keyword search over encrypted data in cloud computing,” in Proc. ICDCS, Minneapolis, MN, USA, 2011, pp. 383–392.
[11] 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.