e-ISSN: 2349-9745
p-ISSN: 2393-8161
Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
www.ijmter.com

Privacy Preserved Distributed Data Sharing with Load Balancing Scheme

Ms. P. Mangayarkarasi[1], Ms. D. Anantha Nayaki[2], Mr. R. Subramanian[3]
[1]*Research Scholar*
[2]*MCA., M.Phil., Assistant Professor, Department of Computer Science,*
[1,2]*Selvamm Arts And Science College (Autonomous), Namakkal, Tamilnadu, India*
[3]*Genius Systems, Erode*

Abstract-Data sharing services are provided under the Peer to Peer (P2P) environment. Federated database technology is used to manage locally stored data with a federated DBMS and provide unified data access. 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.

Distributed data sharing is managed with Privacy Preserved Information Brokering (PPIB) scheme. 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 postencryption modules.

The PPIB scheme is enhanced 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.

I. INTRODUCTION

Peer-to-peer (P2P) computing or networking is a distributed application architecture that partitions tasks or workloads between peers. Peers are equally privileged, equipotent participants in the application. They are said to form a peer-to-peer network of nodes. Peers make a portion of their resources, such as processing power, disk storage or network bandwidth, directly available to other network participants, without the need for central coordination by servers or stable hosts. Peers are both suppliers and consumers of resources, in contrast to the traditional client-server model in which the consumption and supply of resources is divided. Emerging collaborative P2P systems are going beyond the era of peers doing similar things while sharing resources and are looking for diverse peers that can bring in unique resources and capabilities to a virtual community thereby empowering it to engage in

greater tasks beyond those that can be accomplished by individual peers, yet that are beneficial to all the peers.

While P2P systems had been used in many application domains, the architecture was popularized by the file sharing system Napster, originally released in 1999. The concept has inspired new structures and philosophies in many areas of human interaction. In such social contexts, peer-to-peer as a meme refers to the egalitarian social networking that has emerged throughout society, enabled by Internet technologies in general. A peer-to-peer network is designed around the notion of equal peer nodes simultaneously functioning as both "clients" and "servers" to the other nodes on the network. This model of network arrangement differs from the client–server model where communication is usually to and from a central server [1]. A typical example of a file transfer that uses the client-server model is the File Transfer Protocol (FTP) service in which the client and server programs are distinct: the clients initiate the transfer and the servers satisfy these requests.

II. RELATED WORK

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. 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. Distributed hash table technology is adopted to locate replicas based on keyword queries. Such technology has recently been extended to support range queries the coarse granularity cannot meet the expressiveness needs of applications focused in this work. P2P systems often return an incomplete set of answers while we need to locate all relevant data in the IBS.

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, 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 [4]. Various protocols have been proposed for searchable encryption [3], 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 [5] and range queries, supporting queries with complex predicates or structures 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. In this work, we adopt an NFA-based query rewriting access control scheme proposed has a better performance than previous view-based approaches.

III. DATA SHARING IN PEER-TO-PEER (P2P) ENVIRONMENT

III.I. 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 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.

To specify the authorization at the node level, fine-grained access control models are desired. We adopt the 5-tuple access control policy that is widely used in the literature,. The policy consists 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) refers to access "granted" or "denied", respectively; and (5) denotes "local check" or "recursive check". Existing access control enforcement approaches can be classified as *engine-based*, *view-based preprocessing*, and *postprocessing* approaches [6]. In particular, we adopt the *Nondeterministic Finite Automaton* (NFA) based approach as presented, which allows access control to be enforced outside data servers and independent from the data. The NFA-based approach constructs NFA elements for four building blocks of common XPath axes (/x ,//x,/*, and //*) so that XPath expressions, as combinations of these building blocks, can be converted to an NFA, which is used to match and rewrite incoming XPath queries. Please refer to for more details on the QFilter approach.

III.II. Content-Based Query Brokering

Indexing schemes have been proposed for content-based XML retrieval [10], [2]. The index describes the address of the data server that stores a particular data item requested by an user query. Therefore, a contentbased index rule should contain the *content description* and the *address*. 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.

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 can be used to implement content-based indexing, we adopt the model in our study since it can be directly integrated with the NFA-based access control enforcement scheme [8]. We call the integrated NFA that captures access control rules and index rules *content-based query broker* (QBroker). QBroker is constructed in a similar way as QFilter. The binary flag indicates that the state is a "double-slash" state. "double-slash" state, whose child state is an -transition state that directly transits to the next state without consuming any input symbol, will recursively accept input symbols. Unlike QFilter that captures ACRs for only one role, QBroker adds two binary arrays to each state to capture rules for multiple roles: determines the roles that are allowed to access this state and indicates for which role(s) the state is an accept state [9]. For instance, the accept list of state 5 is indicating the state is an accept state for but not for and the access list of state 6 is indicating this state is accessible by both roles. A is attached to each accept state. In the brokering process, Qbroker first checks if a query is allowed to access the requested nodes according to the role type and then makes routing decision. If a query can access only a subset of the requested data, it will be rewritten into a "safe" query before forwarding.

IV. PROBLEM STATEMENT

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 postencryption modules. The following problems are identified in the existing system. Predefined site distribution, Inefficient load balancing mechanism, Complex administrator policy model and Reconfiguration is not supported.

V. P2P DATA SHARING WITH SECURITY AND PRIVACY

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, 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 among a set of organizations. Databases of different organizations are connected through a set of brokers and metadata are "pushed" to the *local brokers*, which further "advertise" 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). 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.

V.I. RSA Algorithm

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 $n = p \times q$	
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.II. Secure Hashing Algorithm

The Secure Hash Algorithm is a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) as a U.S. Federal Information Processing Standard (FIPS), and may refer to:-

A 160-bit hash function which resembles the earlier MD5 algorithm. This was designed by the National Security Agency (NSA) to be part of the Digital Signature Algorithm. Cryptographic weaknesses were discovered in SHA-1, and the standard was no longer approved for most cryptographic uses after 2010.

VI. DISTRIBUTED DATA SHARING WITH LOAD BALANCING SCHEME

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.

VI.I. Data Server

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.

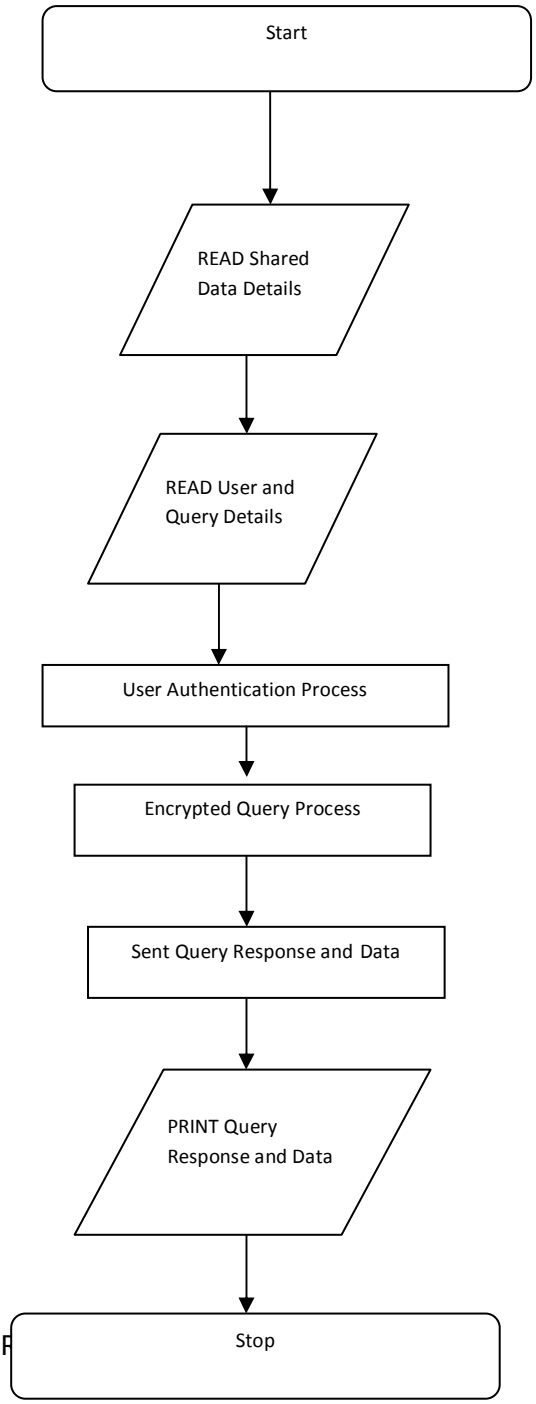
Start
READ Shared Data Details
READ User and Query Details
User Authentication Process
Encrypted Query Process
Sent Query Response and Data
PRINT Query Response and Data
Stop

Fig. No: 6.1. Distributed Data Sharing with Load Balancing Scheme

VI.II. Information Broker

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.

VI.III. Coordinator

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.

VI.IV. Query Processing

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.

VI.V. Load Balancing Process

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.

VII. PERFORMANCE ANALYSIS

The privacy preserved data sharing scheme is designed for the Peer to Peer network environment. The data server is used to provide the shared data files to the users. The information broker is used to manage the user and meta data values. User authentication and query parsing operations are carried out under the information broker environment. The coordinator application is used to perform query redirection process. The load balancing operations are carried out under the coordinator application environment. The RSA algorithm is used for the data security process. The Secure Hashing Algorithm (SHA) is used for the data integrity verification process. The system performs the encrypted query processing for data sharing process.

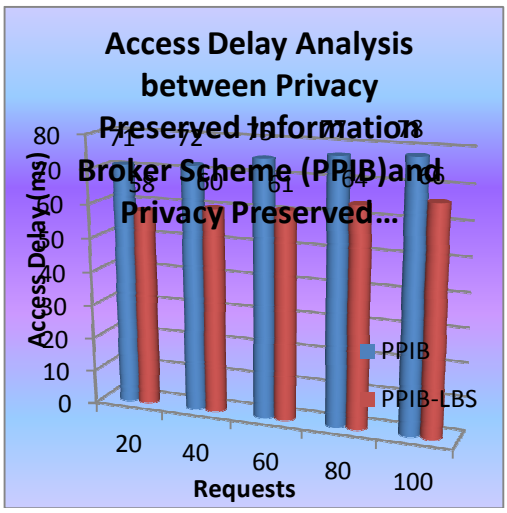

Figure No.7.1. Access Delay Analysis between Privacy Preserved Information Broker Scheme (PPIB) and Privacy Preserved Information Broker with Load Balancing Scheme (PPIB-LBS)

The distributed data sharing system is tested with Privacy Preserved Information Broker Scheme (PPIB) and Privacy Preserved Information Broker with Load Balancing Scheme (PPIB-LBS). The Privacy Preserved Information Broker Scheme (PPIB) is used to perform the data sharing with security features. The data server load level is analyzed in the Privacy Preserved Information Broker with Load Balancing Scheme (PPIB-LBS). The system is tested with two performance measures to evaluate the quality levels. They are access delay and load level measures. The access delay is estimated for the query response time. Access delay is calculated with query submitted time and data delivery time values. The access delay is verified with different data request levels. Figure 7.1. shows the access delay analysis between the PPIBS and PPIB-LBS schemes. The analysis result shows that Access delay in PPIB-LBS is 20% reduced than the PPIB scheme.

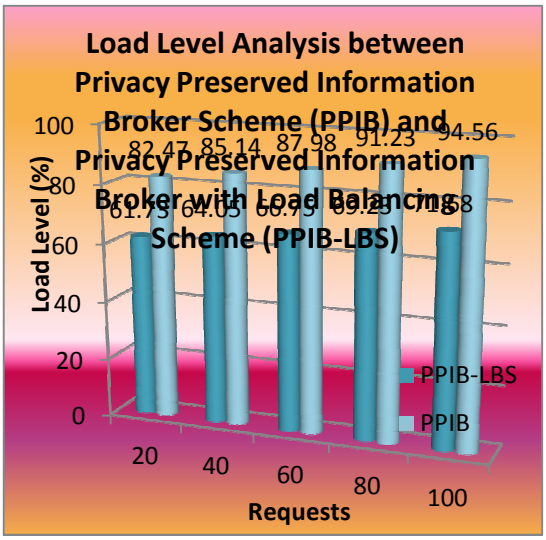

Figure No.7.2. Load Level Analysis between Privacy Preserved Information Broker Scheme (PPIB) and Privacy Preserved Information Broker with Load Balancing Scheme (PPIB-LBS)

The load level for the data sharing schemes is analyzed with different request count levels. The load level is measured for the data servers. The request redirection is performed in normal order under the PPIB scheme. In the PPIB-LBS scheme the request redirection is performed with the data server

load details. Figure 7.2. shows the load level analysis between the PPIB and PPIB-LBS schemes. The analysis result shows that Load level in PPIB-LBS is 30% reduced than the PPIB scheme. The PPIB-LBS scheme produces better results than the PPIB scheme in both access delay and load level measures.

VIII. CONCLUSION AND FUTURE WORK

The privacy preserved distributed data sharing scheme is developed to support data sharing under Peer-to-Peer network environment. The data values are maintained in encrypted form. The data values are collected by the user using the query value. The user submits the encrypted query value. The query is processed by the information broker and the coordinator applications. The coordinator application selects the relevant data server. The data server sends the response to the user. Request count based load balancing scheme is used in the P2P resource sharing environment. The system enhances the response size based load balancing process for the data server selection process. The system can be enhanced with the following features.

- The system can be enhanced to handle malicious and anonymous request based attacks.
- The system can be improved to share data using wireless data servers.
- The data caching scheme can be included in the system to improve the data delivery process.
- The system can be upgraded with bandwidth scheduling scheme.

REFERENCES

[1] Bugra Gedik, Martin Hirzel and Kun-Lung Wu, "Elastic Scaling for Data Stream Processing", IEEE Transactions On Parallel And Distributed Systems, June 2014
[2] G. Skobeltsyn, "Query-Driven Indexing in Large-Scale Distributed Systems," Ph.D. Thesis, EPFL, Lausanne, 2009.
[3] C. Wang, K. Ren and W. Lou, "Secure ranked keyword search over encrypted cloud data," in *Proc. ICDCS'10*, Genoa, 2010.
[4] Emmanuelle Anceaume and Yann Busnel, "A Distributed Information Divergence Estimation over Data Streams", IEEE Transactions On Parallel And Distributed Systems, February 2014.
[5] M. Li and W. Lou, "Authorized private keyword search over encrypted data in cloud computing," in *Proc. ICDCS*, Minneapolis, MN, USA, 2011.
[6] Michele Bristow, Liping Fang and Keith W. Hipel, "Agent-Based Modeling of Competitive and Cooperative Behavior Under Conflict", IEEE Transactions On Systems, Man and Cybernetics: July 2014.
[7] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proc. STOC'09*, Bethesda, MD, USA, 2009.
[8] Fengjun Li and Chao-Hsien Chu "Enforcing Secure and Privacy-Preserving Information Brokering in Distributed Information Sharing", IEEE Transactions On Information Forensics and Security, Vol. 8, No. 6, June 2013.
[9] Qingquan Zhang and Tian He, "Collaborative Scheduling in Dynamic Environments Using Error Inference", IEEE Transactions On Parallel And Distributed Systems, March 2014.
[10] P. Rao and B. Moon, "Locating XML documents in a peer-to-peer network using distributed hash tables," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 12, Dec. 2009.

