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

Efficient Data Delivery Mechanism for Distributed Storages

S. Koperundevi[1], R. Anbarasu, ME[2]

II ME (CSE), CSE Department, Selvam College of Technology, Namakkal, Tamilnadu, India[1]

Assistant Professor, CSE Department, Selvam College of Technology, Namakkal, Tamilnadu, India[2]

Abstract: Distributed data servers are used to share data between the users. 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 provided with security and privacy using 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 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.

Site distribution and load balancing schemes are used to enhance the PPIB scheme. 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

Mega data centers have emerged as infrastructures for building online applications, such as the web search, e-mail, and online gaming, as well as infrastructural services, such as GFS and BigTable. Inside a data center, large number of servers is interconnected using a specific data center networking (DCN) structure with design goals. They include the low equipment cost, high network capacity, support of incremental expansion, and robustness. A number of novel DCN network structures are proposed recently and can be roughly divided into two categories. One is switch centric, which organizes switches into structures other than tree and puts the interconnection intelligence on switches [11]. Fat-Tree, VL2 falls into such a category. The other is server centric, which puts the interconnection intelligence on servers and uses switches only as cross bars. DCell, BCube, FiConn, MDCube and uFix fall into the second category. Among others, a server-centric topology has the following advantages. First, in current practice, servers are more programmable than switches, so the deployment of new DCN topology is more feasible. Second, multiple NIC ports in servers can be used to improve the end-to-end throughput as well as the fault-tolerant ability.

Distributed computing is a field of computer science that studies distributed systems. A distributed system consists of multiple autonomous computers that communicate through a computer network. The computers interact with each other in order to achieve a common goal. A computer program that runs in a distributed system is called a distributed program, and distributed programming is the process of writing such programs. Distributed computing also refers to the use of

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

distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers. The word distributed in terms such as "distributed system", "distributed programming", and "distributed algorithm" originally referred to computer networks where individual computers were physically distributed within some geographical area. The terms are nowadays used in a much wider sense, even referring to autonomous processes that run on the same physical computer and interact with each other by message passing.

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. However, although such technology has recently been extended to support range queries 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. 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. To share data among a large number of autonomous nodes, 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 interorganizational 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 [3] 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 [1], 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 [4] and range queries [2], 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 [6] 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

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

concerns other than anonymity, and thus faces more challenges. Finally, research on distributed access control is also related to ourwork. 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. These approaches rely much on the XML engines. View-based access control approaches create and maintain a separate view for each user, which causes highmaintenance and storage costs. In this work, we adopt an NFA-based query rewriting access control scheme proposed recently in [9], which has a better performance than previous view-based approaches.

III. DATA MANAGEMENT FOR DISTRIBUTED STORAGES

Content-Based Query Brokering schemes have been proposed for content-based XML retrieval [7], [5]. 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*. In [9], 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. 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. NFA state in QBroker, where the state transition table stores the child nodes specified by the XPath expression as the child states in . 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. For instance, the accept list of state 5 is [1 0], indicating the state is an accept state for but not for and the access list of state 6 is [1 1], 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.

A long with the explosion of information collected by organizations in many realms ranging from business to government agencies, there is an increasing need for inter organizational information sharing to facilitate extensive collaboration. While many efforts have been devoted to reconcile data heterogeneity and provide interoperability, the problem of balancing peer autonomy and system coalition is still challenging. Most of the existing systems work on two extremes of the spectrum, adopting either the query-answering model to establish pairwise client-server connections for on-demand information access, where peers are fully autonomous but there lacks system wide coordination, or the distributed database model, where all peers with little autonomy are managed by a unified DBMS.

Unfortunately, neither model is suitable for many newly emerged applications, such as healthcare or law enforcement information sharing, in which organizations share information in a conservative and controlled manner due to business considerations or legal reasons. Take healthcare information systems as example. 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

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

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 [10], such a distributed system providing data access through a set of brokers is referred to as *Information Brokering System* (IBS). 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.

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 drawbacks are identified in the existing system.

- Predefined site distribution
- Inefficient load balancing mechanism
- Complex administrator policy model
- Reconfiguration is not supported

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

V. SECURITY AND PRIVACY ENSURED DATA MANAGEMENT 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.

5.1. 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.

5.2. 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.

5.3. 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.

5.4. 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.

5.5. 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.

VI. CONCLUSION

Data servers are placed in distributed manner to provide shared data values. 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. User authentication, query security and data security are provided in the system.

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

REFERENCES

[1] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Proc. CRYPTO'07, 2007.
[2] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proc. TCC'07, 2007.
[3] C. Wang, N. Cao, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. ICDCS'10, 2010.
[4] M. Li, and W. Lou, "Authorized private keyword search over encrypted data in cloud computing," in Proc. ICDCS, MN, 2011.
[5] G. Skobeltsyn, "Query-Driven Indexing in Large-Scale Distributed Systems," Ph.D. Thesis, EPFL, Lausanne, 2009.
[6] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. STOC'09, Bethesda, MD, USA, 2009.
[7] Skyvalidas, and Dimakopoulos, "Replication routing indexes for XML documents," in Proc. DBISP2P Workshop, Vienna, Austria, 2007.
[8] P. Rao and B. Moon, "Locating XML documents in a peer-to-peer network using distributed hash tables," IEEE Trans. Knowl. Data Eng., Dec. 2009.
[9] F. Li, ,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.
[10] 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.
[11] Deke Guo and Guihai Chen, "Expandable and Cost-Effective Network Structures for Data Centers Using Dual-Port Servers", IEEE Transactions On Computers, July 2013.