A Secured Architecture of Proxy Servers for Information Sharing in Distributed Network

Madhuri D. Dhayarkar
Department of Computer Engineering
Marathwada Mitra Mandal's College of Engineering
Pune, India

Ram B. Joshi
Department of Computer Engineering
Marathwada Mitra Mandal's College of Engineering
Pune, India

***Abstract*— Information is shared via on-demand access in different organizations. Large-scale data sources have been connected through Information Brokering Systems (IBSs). This system includes brokers that are responsible for routing decisions to direct client queries to the requested servers where content is located. Existing IBSs consists of brokers that are trusted and thus only adopt server-side access control for data confidentiality. However the privacy of location of data and information about consumer can still be concluded from metadata (such as query and access control rules) exchanged within the brokering system, thus the protection of the metadata is a major issue in IBS. The proposed scheme presents an overview on information sharing in distributed environment through information brokering system and problems associated with it thus providing scope for healthcare information systems which consists of two domains- personal and public and emergency department. The objective is to overcome two attacks- attribute correlation attack and inference attack providing security enforcement and to provide two countermeasure schemes namely automaton segmentation and query segment encryption scheme. Thus in our scheme, central authority in a system is required for routing decision making responsibility among a selected set of brokering servers by providing access rights. Our scheme also handles risk management issues and provides load balancing, scalability and better privacy.**

***Index Terms*- Access Control, Information sharing, Privacy preservation, Automaton segmentation scheme, Query segment encryption scheme**

I. INTRODUCTION

Today's organizations often operate across organizational boundaries. Strong needs are raised to share the information efficiently and securely to facilitate extensive collaborations among organizations. Previous approaches on sharing of information mainly focus on providing transparency and interoperability among heterogeneous system, fall short of satisfying new requirements of these inter-organizational collaborations. The systems work on two things: (1) servers are autonomous and system-wide communication is not present while responding to the query; so that user creates one-to-one client-server connections for information sharing; (2) in the distributed systems, all the user lost autonomy and are managed by a unified DBMS. There are different types of applications and they need different forms of information sharing. In particular, while some applications uses publish subscribe system and the other applications use the system that provides access to the information on-demand.

As an example, medical data is stored in databases in autonomous enterprises. As a data provider, a participant would not assume free or complete sharing of the data with any unauthorized users as this kind of data is legally private or commercially proprietary. It is required that owner should have full control over the data with the help of access control mechanisms. A feasible solution for storing sensitive data is to construct a data centric overlay including the data sources and a set of brokers helping to locate data sources which will respond to queries. Some mechanisms are used which help to route the queries and thus users can submit queries without knowing data or server's location. Such type of a system provides data access through a set of brokers is referred to as Information Brokering System (IBS). But as the brokers are not trustworthy we propose a novel IBS, named Privacy Preserving Information Brokering (PPIB) to provide secure information sharing.

A. Motivation

Information sharing among inter organizational is an increasing need. Handling heterogeneous data and providing interoperability is a challenge nowadays. In most of the applications enforcement is needed while sharing the secret information which can be shared in a conservative and controlled manner due to business considerations or legal reasons among set of organizations. New technique is provided for healthcare information systems where confidential reports or documents are made available to the doctors, patient (data owner) as well as in emergency department present in the hospitals in case of accidents. Thus our aim is to facilitate access to and retrieval of clinical data across collaborative healthcare providers that include number of hospitals, clinics etc. It is a challenging task as privacy should be maintained while transferring the secret documents. Thus two schemes are used namely automaton segmentation scheme and query segment encryption scheme. These schemes are very helpful in encryption process of the documents so that an unauthorized user or broker or coordinator cannot see the whole content of the document/query. Because of the growing popularity of XML and XML database systems and the need of the privacy, these databases are used in our scheme as they

ISSN: 2278 – 7798

have the ability to hide data from a group of brokers and coordinators and to make the data available to users in an efficient and friendly manner.

The rest of the paper is organized as follows: we discuss literature survey in Section II, and introduce the comparative study of different methods/techniques in Section III. We discuss the drawbacks of existing system in Section IV. We explain our proposed scheme including system architecture and two major schemes in Section V, give implementation details in Section VI including mathematical model, analyze the performance in Section VII, and finally conclude our proposed work in Section VIII.

II. LITERATURE REVIEW

In peer-to-peer (client-server) systems information sharing framework means you have to either share everything or you should not share anything. These systems share files and data sets. But, P2P file-sharing systems may not provide complete set of answers to a request. If the database management system is centralized then it causes privacy and trust issues and is not able to handle heterogeneous data. In IBS, brokers are trusted and thus only adopt server-side access control for data confidentiality.

Peer-to-peer file sharing systems and publish-subscribe systems provide partial solutions to the problem of sharing of data on a large scale. Integrated information provides an integrated view over large numbers of heterogeneous data sources by exploiting the semantic relationship between schemas of different sources.

Distributed hash table technology [12] is adopted to locate replicas based on queries but it is not satisfying the need for privacy while sharing the content. Due to this problem, the XML pub- sub systems is probably the closely related technology to our proposed scheme: where we locate relevant data sources for a given query and route the query to these sources of data which are nothing but servers while the publish subscribe systems are responsible for locating consumers for a given document and route the document to these consumers. They have different concerns: they focus on efficiently delivering the same piece of information to consumers located at different sites while we route large volume but small-size queries to fewer sites.

In [8], pairing-based cryptography and IBE is used to support many-to-many interactions between subscribers and publishers. In our scheme multicasting is not applicable. Thus XML overlay architecture is built that supports query processing and security checking.

The specialized data structures are maintained on nodes to route path queries. In [13], a robust mesh has been built to effectively route XML packets by making the use of self-describing XML tags and the overlay networks. In [14], content-based routing of path queries in peer-to-peer systems is studied to share data among a large number of autonomous nodes. In [4], issues for processing XML data in a peer to peer systems, namely indexing of data, replication of XML data and query routing and processing are studied. The main difference between these approaches and our scheme is that they focus on distributed query routing, while we integrate query routing and access control to preserve relevant private information.

Research has been proposed on distributed access control. Earlier approaches implement access control mechanisms at the nodes of XML trees and filter out data nodes that users do not have authorizations to access [15]. This processing is handled by XML engines. Creation and maintenance of a separate view for each user is handled by view-based access control mechanisms but requires high maintenance and storage cost [16]. In [7], it uses attribute based encryption (ABE) techniques to achieve fine-grained and scalable data access control for PHRs.

III. COMPARATIVE ANALYSIS OF DIFERENT TECHNIQUES

Fig. 1 shows the comparative analysis of different techniques.

Title	Year	Advantage	Disadvantage
Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption[7]	2013	Focus on the multiple data owner scenario, and divide the users in the PHR system into multiple security domains that greatly reduces the key management complexity for owners and users	Usually assume the use of a single trusted authority (TA) in the system creating a load bottleneck
A Survey on Protecting Information Brokerage System against Intruders[10]	2013	Brokers make use of routing protocols that create hard-to-trace communications by using a chain of proxy servers which are untraceable and mainly responsible for user authentication and query forwarding	The broker functionality may be outsourced to third-party providers and thus vulnerable to be abused by insiders or compromised by outsiders
Design and Implement Privacy Protection for Secure Information Brokering Systems[9]	2014	End-to-end query processing performance and System scalability	Several types of attacks possible
Securing Broker-Less Publish/Subscribe Systems Using Identity-Based Encryption[8]	2014	To ensure that a particular subscriber can decrypt an event only if there is match between the credentials associated with the event and the key	A parent can decrypt every event it forwarded to its children
A Novel Approach to Improve the Privacy of Information Brokering in Semantic Web[11]	2014	Enriches the privacy of data shared within Information Brokering System by using Selective encryption, Vigenere Cipher encryption and Selective Reverse Circle Cipher algorithm	Privacy leakage of data requestor privacy, data privacy and metadata privacy

Fig 1: Comparative analysis of different methods

IV. EXISTING SYSTEM

Databases of different organizations are connected through a set of brokers, and metadata is pushed to the local brokers, which further advertise (some of) the metadata to other brokers (green nodes) as shown in Fig. 2 [1]. Local broker receives the query from user, and routes the query based on metadata till it reaches to the database having the expected information to return for the particular query. Information sources in different organizations are loosely federated to provide unified data access, transparency, and on-demand data access.

ISSN: 2278 – 7798

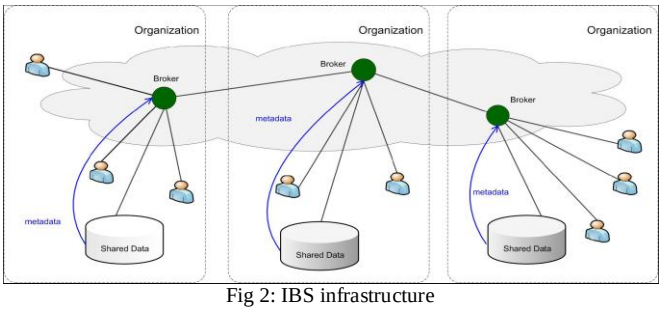


Fig 2: IBS infrastructure

Disadvantages: Though the IBS approach provides scalability and server autonomy, there are privacy issues, as brokers are not fully trustable – they may be abused by insiders or compromised by outsiders.

A. *Privacy Vulnerabilities*

Information Brokering is dependent on the trust of brokers for query forwarding and leads to harm the privacy of user, data and metadata. The user privacy can be described as identity of user, location of user while sending a query and obtain the purpose of the query. User identity can be assumed by authentication process and information about the user location. Location of data and data object distribution privacy is included in data privacy. It describes which type of data is contained in data server. Query indexing that is index rules and ACR are two types included in metadata. It gives us the idea about where the data objects are distributed among data server and provides access to authorized users. Data providers push routing and access control metadata to brokers, which also receives queries from users. Thus, corrupted brokering server could: (1) learn query content and query location of a local query (2) learn routing and access control metadata from local data servers and other brokers (3) learn location of data from routing metadata. In this case, chance for attacker to obtain plaintext data from the data which is encrypted is very less but they are able to learn location of query and data.

B. *Attacks in Existing System*

The attacks are classified as:
(1) Attribute correlation attack: when query is routed through the system, curious broker or attacker from external sources (eavesdropper) may extract the query condition for getting the sensitive information by matching the attributes contained in the query.
(2) Inference attack: By getting more than one type of confidential information, the attacker is able to guess the query location (IP address), query content to identify data owner from it. We prove that PPIB scheme provides privacy protection for on-demand brokering of the information by overcoming these attacks and very good scalability.

V. PROPOSED SYSTEM

To provide privacy protection, we propose new approach for IBS named Privacy Preserving Information Brokering (PPIB). It is an overlay infrastructure consisting of three types of brokering components: (1) brokers and (2) coordinators (3) central authority (CA) as shown in Fig 3 [1]. The idea for privacy preservation is nothing but you have to divide the work among multiple components so that not a single node is able to make inference from the information that is disclosed to that component.

A. *System Architecture*

1. Brokers (main hospital or hospital branches)- are the entry point in the system and are mainly responsible for user authentication and forwarding the query from user who is authorized. The broker forms the middle layer between coordinators and users. The request is submitted from the user is verified and thus it will be passed to the coordinator.

2. Coordinators (departments like blood report, pathology)- are linked in a tree structure enforce access control and query routing based on the embedded nondeterministic finite automata also known as query brokering automata. The coordinators hold a segment of automaton that helps in routing. That segment is nothing but one state in NFA. Each state is attached with dummy state which holds the address of next state.

3. Central authority (data server)- is responsible for key management and maintaining metadata. It also handles leaving/joining of brokers and coordinators to the system and risk management issues in case of failure of any component.

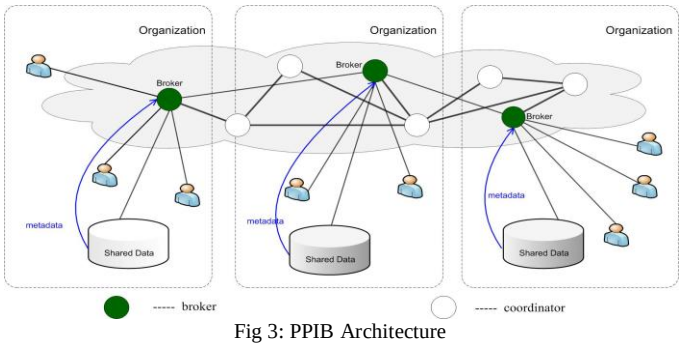


Fig 3: PPIB Architecture

B. *Two Major Schemes in PPIB*

There are two schemes namely (1) automaton segmentation scheme (2) query segment encryption scheme. The first scheme divides metadata into segments and each segment is then assigned to a coordinator and each segment holds one NFA state as shown in Fig. 4. Coordinators enforce secure query routing thus they operate in collaborative manner. Second scheme is query segment encryption scheme to protect query content and it prevents coordinators from seeing sensitive information. In our scheme a query is divided into multiple segments, and each segment is encrypted in a proper way such that no segment apart from the ones needed to enforce secure routing is revealed to the coordinators route.
Thus the provided schemes in the system ensure that corrupted broker or coordinator is not capable to collect enough information to infer privacy.

C. *XML Data and Access Control Model*

Nowadays the extensible Mark-up Language (XML) is an emerging standard for information sharing due to its rich

ISSN: 2278 – 7798

semantics. Information is exchanged by taking XPath queries and returning the data.

The 5-tuple access control policy is used in access control rules.

ACR {subject, object, action, sign, type} where

(1) *Subject* is the role in the system to whom authorization is granted;

(2) *Object* is a set of XML nodes specified in XPath expression;

(3) *Action* is operations as "read", "write", or "update";

(4) *Sign belongs* to {+,-} refers to access "granted" or "denied",

(5) *Type LC* or RC means local check means authorization is applied only to the attributes or textual data of the context nodes or recursive check means authorization is applied to all the descendants of the context node.

Examples of ACRs:

R1:{role1,/site//person/name, read, +, RC}

R2:{role1,/site/regions/asia/item,read,+,RC}

A newly proposed NFA-based query re-writing access control [17] is adopted by PPIB approach and extended that scheme in a decentralized way that can be used by any off the-shelf XML DBs. This NFA approach constructs NFA elements for four building blocks of common XPath axes (/x, //x, /*, and //*). It forms XPath expressions which is combinations of these building blocks, which are converted to an NFA. XPath queries are matched with the NFA and rewritten incoming XPath queries.

D. Content-based Query Brokering

We presented a content-based indexing model with index rules in the form of I={object, location}, where (1) object is an XPath expression that selects a set of nodes (2) location is a list of IP address of data servers that hold the content. When user sends the query to the system, that query is matched with the object field of the index rules, and the query will be sent to the data server specified by the location field of the matched rule(s) to send the response. The NFA that integrates ACR and IR is a content-based query broker (QBroker) as shown in Fig. 4.

Examples of index rules:

I1:{/site/people/person/name,130.203.189.2}

I2:{/site/regions//item[@id>"100"],135.176.4.56}

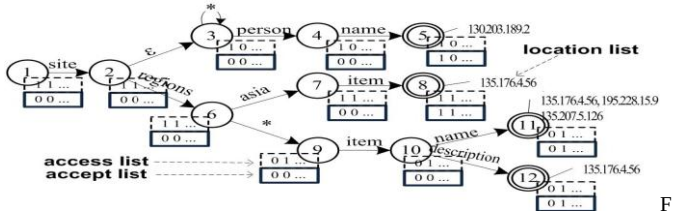


Fig. 4: The state transition graph of the QBroker that integrates index rules with ACRs

In healthcare application, query will be forwarded as

R1:{user1,/userip/hospital/branch/regno/querycontent/dept}

I1: {/ userip/hospital/branch/regno/querycontent/dept, 130.203.189.2}

VI. IMPLEMENTATION DETAILS

A. Mathematical Model

Consider U as the user, B as the broker and C as the coordinator.

- U = {u1, u2,...., un}
- B = {B1, B2,….,Bn}
- C = {C1, C2,…., Cn}
- S = {F1, F2, F3, F4}

1. User authentication
   - Input: Login id, password
   - F1 = {uid, Pk, K_Q, E}
     where uid – User id,
     Pk – Public Key,
     K_Q – Session Key,
     E – Encryption
   - Output : E(K_Q, $P_{K,}$ Q)

2. Meta data preparation
   - Input: E(K_Q, $P_{K,}$ Q),
   - F2 = {uid, E, Q_{ID}, $K_{Q,}$ baddr }
     where uid – user id,
     Q_{ID} – query id,
     K_Q – Session key,
     baddr – broker address
   - Output : Encrypted query with baddr

3. NFA= {ST, PT, ID, Locationlist}
   where ST - State transition table
   PT - Predicate table

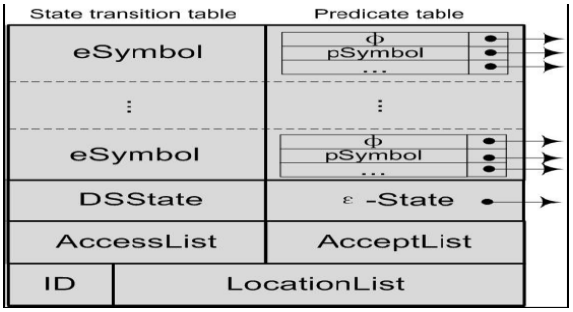


Fig. 5: Data structure of an NFA state

ST = {eSymbol, DSState, AccessList}
where eSymbol – XPath state
DSState – Double slash state
AccessList – determines roles that are allowed to access the state

PT = {Psymbol, condition, AcceptList, Locationlist}
where Psymbol - stores predicates
AcceptList – for which roles the state is an accept state
condition - stores test condition
Locationlist - stores address of index rule

ISSN: 2278 – 7798

4. Role of root coordinator uses two schemes
- Input: E(Q),Q_{ID}
- F3 = {RC, $Q_{ID,}$ Pk}
  where RC – root coordinator,
  Q_{ID} – query id,
  Pk – Public Key of server
- Output: Forwards encrypted query to leaf coordinator

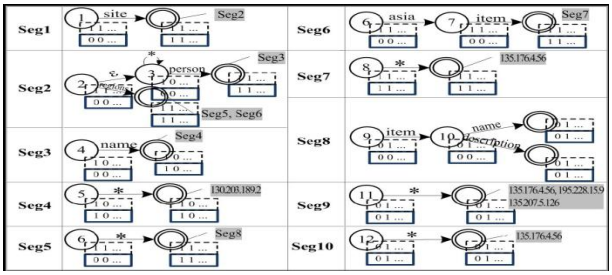


Fig 6: Automaton segmentation scheme

5. Query Segment Encryption
- Input : E(Q)
- F4 = {E, D, K_Q }
- Output: D(E(Q), K_Q)

B. Algorithm

1. User, broker and coordinator registers themselves to the system. Central authority will authenticate broker and coordinator. An authorized user sends encrypted query and a unique session key K_Q which is encrypted with the data server's public key is generated so that data server can return data to the user.
2. Local broker authenticates the user, receives the query and prepares metadata. It creates unique ID for each query and attaches Q_{ID} with its own address. Broker will forward query to the coordinator based on the query contents. As the whole query having query contents is in encrypted form, forwarding of query to the coordinator depends on ACR assigned by central authority. If query is forwarded to the wrong coordinator then user will not receive response showing worst case.
3. Root Coordinator receives the encrypted query. It uses automata segmentation scheme to divide query into segments and assign each segment to a subsequent coordinators and query segment encryption scheme to prevent coordinators from seeing query content. It uses indexing to route the query to leaf coordinator. Failure message is returned to the broker if query is denied access. All the query segments are processed and encrypted with the public key of the data server by the leaf coordinator and the query is then forwarded to the data server. If root coordinator is the leaf coordinator then the user will get response quickly by entering the private key received through email showing the best case.
4. Data server receives encrypted query. That query is decrypted and evaluated by data server and returns data encrypted by K_Q.
5. User decrypts the data using private key received through email and gets the response for the query.
6. In case of emergency scenarios, user will give his/her details to the respective trustworthy doctor and then the doctor will get the person's medical history to start with the treatment. User can change the details afterwards to protect his/her data.

Fig. 7 shows that PPIB only requires minimal trust (or honesty) in each component where hide means there is no need to trust. It is clear that whenever the level of trust in each brokering component is decreased, the system's capability to protect the privacy of information will be enhanced.

Privacy Type	User Location	Query Content	Data Server Location	Data Object Distribution	Access Control Policy	Index Information
Broker	Trust	Hide	Hide	Hide	Hide	Hide
Root-Coordinator	Hide	Trust	Hide	Hide	(Partially) Trust	(Partially) Trust
Coordinator	Hide	(Partially) Trust	Hide	Hide	(Partially) Trust	(Partially) Trust
Leaf-Coordinator	Hide	Hide	Trust	Hide	Hide	Trust
Data Server	Hide	Trust	Trust	Trust	Trust	Hide

Fig. 7: Level of trust of each brokering components

VII. EXPERIMENTAL RESULTS

These graphs are based on sign in person i.e role of the person using this system. It shows that response time is changing as the query content changes and depends on role involves in between query forwarding and response i.e how many coordinators involved in response and forwarding. There is small change in response time as flow changes i.e Regular flow, emergency flow and recursive flow.

I] Regular Flow: when system works properly showing the best case and Q1, Q2, Q3 are the three queries.

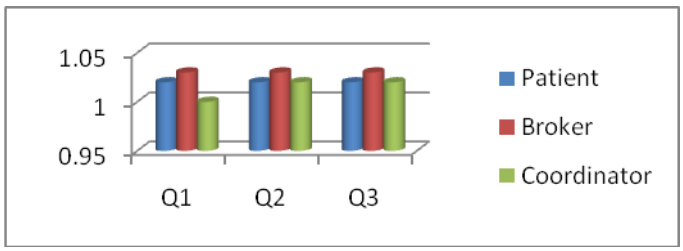


Fig. 8: Regular flow of system

II] Emergency Flow: in case of emergency scenarios

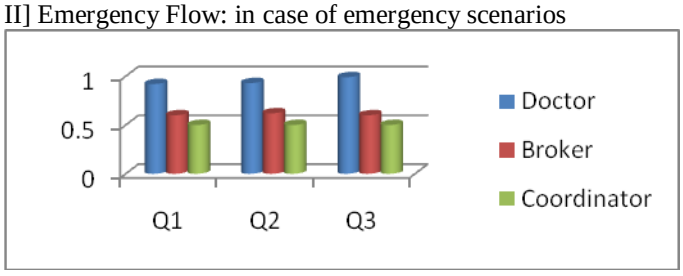


Fig. 9: Emergency flow of system

III] Recursive Flow: when query moves in tree structure of coordinators showing the average or worst case

ISSN: 2278 – 7798

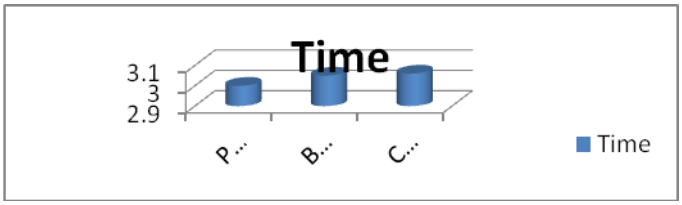
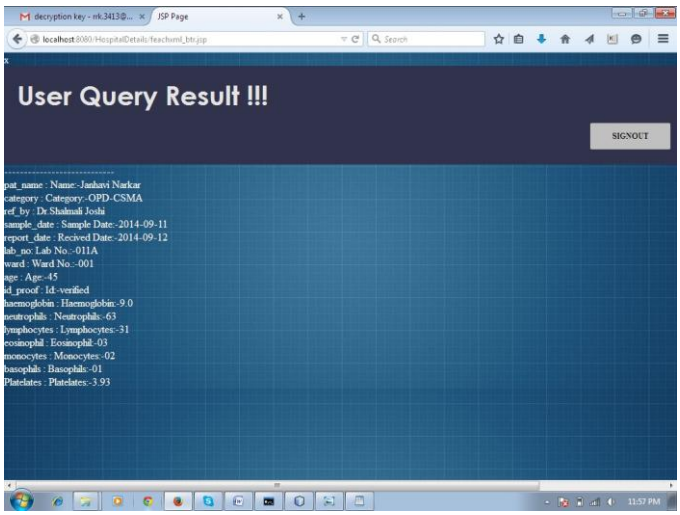


Fig. 9: Recursive flow of system

A. *Outcome*



VIII. CONCLUSION

In this paper, we propose PPIB scheme, which provides a new approach for privacy preserving in XML information brokering system. We apply our proposed scheme in healthcare information sharing that also supports emergency scenarios. By using automaton segmentation scheme, query segment encryption scheme and ACR, PPIB integrates security enforcement and query forwarding while providing privacy protection. The analysis shows that PPIB system overcomes the two major attacks, provides end-to-end query processing, performance, scalability and supports load balancing. We have made the system self-configurable. Many directions are ahead for future work. Several factors like dynamic site distribution, the workload and trust level at server, and privacy issues between automaton segments can be considered. To provide the service while server is offline is also a challenge.

ACKNOWLEDGMENT

We would like to take this opportunity to thank Mrs. H. K. Khanuja for her supportive guidance and for providing all the facilities, which are necessary in the completion of this paper. We are also thankful to all the respective staff members of the Department of Computer Engineering for their valuable time, support and suggestions. We would also like to thank the institute for providing the required facilities.

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 Transactions On Information Forensics And Security Vol:8 No:6 Year 2013.

[2] N. Koudas, M. Rabinovich, D. Srivastava, and T. Yu, "Routing XML queries," in ICDE '04, p. 844, 2004.

[3] A. Carzaniga, M. J. Rutherford, and A. L. Wolf, "A routing scheme for content-based networking," in Proc. of INFOCOM, 2004.

[4] G. Koloniari and E. Pitoura, "Peer-to-peer management of XML data: issues and research challenges," SIGMOD Rec., vol. 34, no. 2, 2005.

[5] 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, 2006.

[6] 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 ACM CCS '07, pp. 508–518, 2007.

[7] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, Wenjing Lou, "Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption," IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 1, January 2013.

[8] Muhammad Adnan Tariq, Boris Koldehofe, and Kurt Rothermel, "Securing Broker-Less Publish/Subscribe Systems Using Identity-Based Encryption," IEEE Transactions On Parallel And Distributed Systems, VOL. 25, NO. 2,February 2014.

[9] Shaik.Mahaboob Basha, A.Bhaskar, D.V Satish Kaladhar Reddy, "Design and Implement Privacy Protection For Secure Information Brokering Systems," in IJCSMC, Vol. 3, Issue. 8, pg.41–48, August 2014.

[10] Sanchari Saha, Madhusudana H.A, "A Survey on Protecting Information Brokerage System against Intruders," International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 5, November 2013.

[11] Supriya S. Sankpal, Rupali A. Mahajan, "A Novel Approach to Improve the Privacy of Information Brokering in Semantic Web," International Journal of Science and Research (IJSR),Volume 3, Issue 7, July 2014.

[12] 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", in IEEE/ACM Trans. Networking, volume 11 of 1, 2003.

[13] A. C. Snoeren, K. Conley, and D. K. Gifford, "Mesh-based content routing using XML", in Symposium on Operating Systems Principles, pages 160-173, 2001.

[14] G. Koloniari and E. Pitoura. Content-based routing of path queries in peer-to-peer systems. In EDBT, 2004.

[15] M. Murata, A. Tozawa, and M. Kudo. XML access control using static analysis. In ACM CCS, Washington D.C., 2003.

[16] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy. Extending query rewriting techniques for fine-grained access control. In SIGMOD, pages 551{562, Paris, France, 2004.

[17] B. Luo, D. Lee, W.-C. Lee, and P. Liu. QFilter: Fine-grained run-time XML access control via NFA-based query rewriting. In ACM CIKM, Washington D.C., USA, nov 2004.

ISSN: 2278 – 7798