International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)

International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

Secure Preserving Information Brokering System

Anita Rose J T[1], Shirly Jeny B[2]

[1]*Associate Professor, Dept Of Cse, St.Joseph's College Of Engineering, Chennai, India*
[2]*M.E.Computer Science And Engineering, St.Joseph's College Of Engineering, Chennai, India*

jenyshirly@gmail.com

***Abstract*—Today's organizations raise an increasing need for information sharing via on-demand access. Information brokering systems (IBSs) have been proposed to connect large-scale loosely federated data sources via a brokering overlay, in which the brokers make routing decisions to direct client queries to the requested data servers. Many existing IBSs assume that brokers are trusted and thus only adopt server-side access control for data confidentiality. However, privacy of data location and data consumer can still be inferred from metadata (such as query and access control rules) exchanged within the IBS, but little attention has been put on its protection. In this paper, we propose a novel approach to preserve privacy of multiple stakeholders involved in the information brokering process. We are among the first to formally define two privacy attacks, namely *attribute-correlation attack* and *inference attack*, and propose two countermeasure schemes *automaton segmentation* and *query segment encryption* to securely share the routing decision-making responsibility among a selected set of brokering servers. With comprehensive security analysis and experimental results, we show that our approach seamlessly integrates security enforcement with query routing to provide system-wide security with insignificant overhead.**

***Keywords*—Access control, information sharing, privacy.**

I. INTRODUCTION

Along 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 pair wise 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.

Many existing IBS assume that brokers are trusted and thus only adopt server-side access control for data confidentiality. However, privacy of data location and data consumer can still be inferred from metadata .Encryption to securely share the routing decision making responsibility among a selected set brokering servers. With comprehensive security analysis and experimental results, we show that our approach seamlessly integrates security enforcement with query routing to provide system-wide security with insignificant overhead.

In the context of sensitive data and autonomous data that given by the 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).

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.

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 non deterministic finite automata—the *query brokering automata*. A novel approach is identified to preserve privacy of multiple stakeholders involved in the information brokering process, Automation segmentation, Query segment encryption

II. Problem

In a typical information brokering scenario, there are three types of stakeholders, namely *data owners*, *data providers*, and *data requestors*. Each stakeholder has its own privacy: (1) the privacy of a data owner (e.g., a patient in RHIO) is the identifiable data and sensitive or personal information carried by this data (e.g., medical records). Data owners usually sign strict privacy agreements with data providers to prevent unauthorized use or disclosure. (2) Data providers store the collected data locally and create two types of metadata, namely *routing metadata* and *access control metadata*, for data brokering. Both types of metadata are considered privacy of a data provider. (3) Data requestors may reveal identifiable or private information (e.g., information specifying her interests) in the querying content. For example, a query about AIDS treatment reveals the (possible) disease of the requestor. We adopt the *semi-honest* [12] assumption for the brokers, and assume two types of adversaries, *external attackers* and *curious or corrupted brokering components*. External attackers passively eavesdrop communication channels. Curious or corrupted brokering components, while following the protocols properly to fulfill brokering functions, try their best to infer sensitive or private information from the querying process. Privacy concerns arise when identifiable information is disseminated with no or poor disclosure control.

For example, when data provider pushes routing and access control metadata to the local broker [6], [9], a curious or corrupted broker learns *query content* and *query location* by intercepting a local query, *routing metadata* and *access control metadata* of local data servers and from other brokers, and *data location* from routing metadata it holds. Existing security mechanisms focusing on confidentiality and integrity cannot preserve privacy effectively. For instance, while data is protected over encrypted communication, external attackers still learn *query location* and *data location* from eavesdropping. Combining types of unintentionally disclosed information, the attacker could further infer the privacy of different stakeholders through *attribute-correlation attacks* and *inference attacks*.

Attribute-correlation attack. Predicates of an XML query describe conditions that often carry sensitive and private data (e.g., name, SSN, credit card number, etc.) If an attacker intercepts a query with multiple predicates or composite predicate expressions, the attacker can "correlate" the attributes in the predicates to infer sensitive information about data owner. This is known as the *attribute correlation attack*.

Inference attack. More severe privacy leak occurs when an attacker obtains more than one type of sensitive information and learns explicit or implicit knowledge about the stakeholders through association. By "implicit", we mean the attacker infers the fact by "guessing". For example, an attacker can guess the identity of a requestor from her query location (e.g., IP address). Meanwhile, the identity of the data owner could be explicitly learned from query content (e.g., name or SSN). Attackers can also obtain publicly-available information to help his inference. For example, if an attacker identifies that a data server is located at a cancer research center, he can tag the queries as "cancer-related".

In summary, we have three reasonable inferences from three distinct combinations of private information: (1) from *query location & data location*, the attacker infers about *who* (i.e., a specific requestor) is interested in *what* (i.e., a specific type of data).

(2) From *query location & query content*, the attacker infers about *where who* is, or *who* is interested in *what* (if predicates describe symptom or medicine, etc.), or *something* about the data owner (if predicate identifies name or address of a personnel), etc. (3) From *query content & data location*, the attacker infers *which* data server has *which* data. Hence, the attacker could continuously create artificial queries or monitor user queries to learn the data distribution of the system, which could be used to conduct further attacks.

III. Background

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(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 , the coarse granularity (e.g., files and documents) cannot meet the expressiveness needs of applications focused in this work. Furthermore,P2P systems often returns an incomplete set of answers while we need to locate all relevant data in the IBS. Addressing a conceptually dual problem, XML publish-subscribe systems (e.g., [19], [20]) are probably the closely relate 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 a top normal IP network.

In particular, specialized data structures are maintained on overlay nodes to route XML queries. In a robust mesh has been built to effectively route XML packets by making use of self-describing XML tags and the overlay networks. To share data among a large number of autonomous nodes, 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 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 (i.e., indexing and access control), secure index based search schemes may be adopted to outsource metadata in encrypted form to un trusted 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 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 and range queries, supporting queries with complex predicates (e.g., regular expressions) or structures (e.g., X Path queries) is still a difficult open problem. In terms of privacy-preserving brokering, another related technique is secure computation 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 give a good overview on access control in collaborative systems.

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 (e.g., a specific portion of XML documents) for each use, which causes high maintenance and storage costs. In this work, we adopt an NFA-based query rewriting access control scheme proposed recently in which has a better performance than previous view-based approaches .

To integrate distributed schemas into a new one and query just that new schema without losing the ability to retrieve data fromthe original schemas. The area in which we try is to federated databases, where the original heterogeneous sources are assumed to be autonomously managed.

Our approach is based on schema matching, the identification of semantic correspondences between elements of different schemas. We propose SASF, the Service-based Approach to Schema Federation, which is composed services: schema translation, schema matching, and schema mapping that are accessed through a user portal.

Our approach exploits both schema-level and instance-level information to discover semantic correspondences, and is able to operate over a full range of matching cardinalities. A demonstration of the usability of SASF in a real-world scenario of federating astronomical databases is presented as a case study. The results, we believe, illustrate the potential of performing data integration through database federation.

We propose a fully homomorphism encryption scheme – i.e., a scheme that allows one to evaluate circuits over encrypted data without being able to decrypt. First, we provide a general result – that ,to construct an encryption scheme that permits evaluation of arbitrary circuits, it suffices to construct an encryption scheme that can evaluate (slightly augmented versions of)its own decryption circuit; we call a scheme that can evaluate its (augmented) decryption circuit boot strippable.

Next, we describe a public key encryption scheme using ideal lattices that is almost boot strippable. Lattice-based cryptosystems typically have decryption algorithms with low circuit complexity, often dominated by an inner product computation that is in NC1. Also, ideal lattices provide both additive and multiplicative homomorphism's (modulo a public-key ideal in a polynomial ring that is represented as a lattice), as needed to evaluate general circuits.

A federated database system (FDBS) is a collection of cooperating database systems that are autonomous and possibly heterogeneous. In this paper, we define a reference architecture for distributed database management systems from system and schema viewpoints and show how various FDBS architectures can be developed. We then define a methodology for developing one of the popular architectures of an FDBS. Finally, we discuss critical issues related to developing and operating an FDBS.

One of the significant aspects of an FDBS is that a component DBS can continue its local operations and at the same time participate in a federation. The integration of component DBSs may be managed either by the users of the federation or by the administrator of the FDBS together with the administrators of the component DBSs.

The amount of integration depends on the needs of federation users and desires of the administrators of the component DBSs to participate in the federation and share their databases.

To identify best practices in emerging and existing regional health information organizations (RHIOs), AHIMA's e-HIM_ work group on patient identification in RHIOs offers HIM professionals a marketplace description of existing RHIO models with a focus on patient identification linkage methods.

This practice brief does not address the issues related to privacy and security reflected in HIPAA regulations ,nor issues such as data quality within the records. It only discusses the quality of linking methods .Organized cross-jurisdictional healthcare data-sharing organizations are referred to as RHIOs throughout this . This is just one of several terms applied to such organizations.

Others include health information exchange (a more generic term used by the e Health Initiative), sub network organization(the term used by the Collaborative Response to the Office of the National Coordinator for Health Information Technology), and health information network (a term used by several organizations). As this is an evolving field, it is likely that other terminology will come into vogue in the future.

A new approach for reliably multicasting time critical data to heterogeneous clients over mesh-based overlay networks. To facilitate intelligent content pruning, data streams are comprised of a sequence of XML packets and forwarded by application-level XML routers.

XML routers perform content based routing of individual XML packets to other routers or clients based upon queries that describe the information needs of downstream nodes. Our PC-based XML router prototype can route an 18Mbit per second XML stream.

Our routers use a novel Diversity Control Protocol (DCP) for router-to-router and router-to-client communication. DCP reassembles on the a received stream of packets from one or more senders using the first copy of a packet to arrive from any sender.

When each node is connected to n parents, the resulting network is resilient to $(n-1)$ router or independent link failures without repair. Associated mesh algorithms permit the system to recover to $(n-1)$resilience after node and/or link failure. We have deployed a distributed network of XML routers that streams real-time air traffic control data. Experimental results show multiple senders improve reliability and latency when compared to tree-based networks.

The work is based upon the assumption that, in certain cases, the value of reliable and timely data delivery may justify increased transport costs if the cost increase allows us to meet a desired reliability goal. Systems often try to avoid the delay penalty by using loss-resistant coding schemes which encode redundant information into the data stream. We extend this redundancy to network delivery paths and senders.

Recent work in overlay networks has shown that multiple, distinct paths often exist hosts on the Internet.

Attempt to leverage these redundant network links. While some may consider this additional bandwidth wasteful, we believe the system described herein presents an interesting and elegant method of utilizing additional network resources to achieve levels of reliability and latency previously difficult to obtain.

Our basic approach is to construct a content distribution *mesh*, where every node is connected to n parents, receiving duplicate packet streams from each of its parents. The value of n is a configuration parameter that is used to select the desired trade-off between latency, reliability, and transport costs.

By maintaining an acyclic mesh, this approach guarantees that the minimum cut of the mesh is n nodes or independent links. Thus, a mesh is resilient to $(n-1)$ node or $(n-1)$ independent link failures

IV. Overall PPIB Architecture

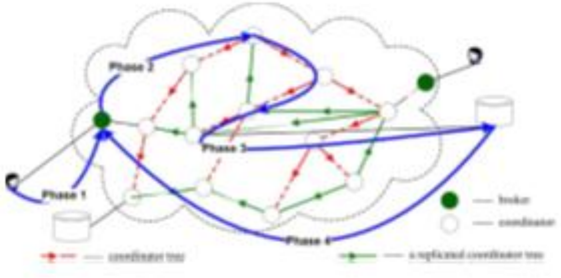
The architecture of PPIB, where users and data servers of multiple organizations are connected via a broker-coordinator overlay. The main aim of this project is to propose a novel approach to preserve privacy of multiple stakeholders involved in the information brokering process. With comprehensive security analysis and experimental results, shows that our approach seamlessly integrates security enforcement with query routing to provide system-wide security with insignificant overhead**.**In particular, the brokering process consists of four phases:

• *Phase 1***:** To join the system, a user needs to authenticate himself to the local broker. After that, the user submits an XML query with each segment encrypted by the corresponding public level keys, and a unique session key . is encrypted with the public key of the data servers to encrypt the reply data.

• Phase 2**:** Besides authentication, the major task of the broker is metadata preparation: (1) it retrieves the of the authenticated user to attach to the encrypted query; (2) it creates a unique for each query, and attaches and its own address to the query for data servers to return data.

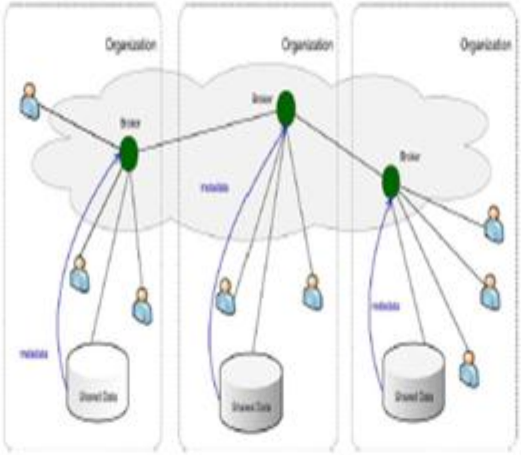
International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)

International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)



4.1 Phases Of Brokering Process

• *Phase 3***:** Upon receiving the encrypted query, the coordinators follow automata segmentation scheme and query segment encryption scheme to perform access control and query routing along the coordinator tree. At the leaf coordinator, all query segments should be processed and encrypted by the public key of the data server. If a query is denied access, a failure message with will be returned to the broker.



• Phase 4: In the final phase, the data server receives a safe query in an encrypted form. After decryption, the data server evaluates the query and returns the data, encrypted by , to the broker that originates the query.

1. Global schema & PPIB components registration

Global organization, having global schema, acts as a consortium, for different organization that belongs to same field, and agree to share their data as global data. Here, a standard schema, known as global schema, is provided for the organization. So, different organization having different schema register with global organization and shares the global schema.

PPIB components such as brokers, coordinators register with Information Brokering System (IBS). Requestor registers with corresponding brokers, who acts as the entrance to the IBS. Coordinator send request to Central Authority to join the system.

2. Central Authority:

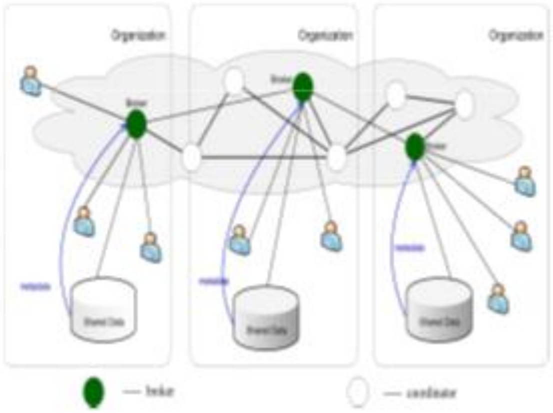
The CA is assumed for off-line initiation and maintenance. With the highest level of trust, the CA holds a global view about all the rules and plays a critical role in automaton segmentation and key management except the query session keys created by the user; the other keys are generated and maintained by the CA. The data servers are treated as a unique party and share a pair of public and private keys, while each of the coordinators has its own pairs of level key and commutative level key. Along with the automaton segmentation and deployment process, the CA creates key pairs for coordinators at each level and assigns the private keys with the segments. The level keys need to be revoked in a batch once a certificate expires or when a coordinator at the same level quits the system.

3. Query routing:

Data servers and requestors from different organizations connect to the system through local brokers .Brokers are interconnected through coordinators. A local broker functions as the “entrance” to the system .It authenticates the requestor and hides his identity from other PPIB components. It forwards the requestor query to root coordinator. The coordinator process the query against its automaton segment assigned to it. After successful processing, it sends the query to the child coordinators for further processing. If denied, it sends the failure message the corresponding broker.

4. Query processing:

Finally, the data server receives the processed query in an encrypted form .After decryption, the data server evaluates the query and returns the data, encrypted by KQ, to the broker that originates the query.



4.3 Proposed System Of Brokering

V. CONCLUSION

With attention drawn on privacy of user, data, and metadata during the design stage, existing information brokering systems suffer from a spectrum of vulnerabilities associated with user privacy, data privacy, and metadata privacy. In this project, we propose PPIB, a new approach to preserve privacy in XML information brokering. Through an innovative automaton segmentation scheme, in-network access control, and query segment encryption, PPIB integrates security enforcement and query forwarding while providing comprehensive privacy protection .Our analysis shows that it is very resistant to privacy attacks. End-to-end query processing performance and system scalability are also evaluated and the results show that PPIB is efficient and scalable.

REFERENCES

[1] W. Bartschat, J. Burrington-Brown, S. Carey, J. Chen, S.Deming, and S. Durkin, “Surveying the RHIO landscape: A description of current{RHIO} models, with a focus on patient identification,” J. AHIMA, vol. 77, pp. 64A–64D, Jan. 2006.

[2] A. P. Sheth and J. A. Larson, “Federated database systems for managing distributed, heterogeneous, and autonomous databases,” ACM Comput. Surveys (CSUR), vol. 22, no. 3, pp. 183–236, 1990.

[3] L. M.Haas, E. T. Lin, and M.A. Roth, “Data integration through database federation,” IBM Syst. J., vol. 41, no. 4, pp. 578–596, 2002.

[4] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, “CoolStreaming /DONet: A data-driven overlay network for efficient live media streaming,” in Proc. IEEE INFOCOM, Miami, FL, USA, 2005, vol. 3, pp. 2102–2111.

[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] G. Koloniari and E. Pitoura, “Peer-to-peer management of XML data: Issues and research challenges,” SIGMOD Rec., vol. 34, no. 2, pp. 6–17, 2005.

[8] M. Franklin, A. Halevy, and D. Maier, “From databases to data spaces: A new abstraction for information management,” SIGMOD Rec., vol. 34, no. 4, pp. 27–33, 2005.

[9] 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.

[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, pp. 508–518.

[11] D. L. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” Commun. ACM, vol. 24, no. 2, pp. 84–90, 1981.