# Secure and Privacy-Preserving Information in Distributed Information Sharing

**Yamunadevi.M.A[1], Nithya.K[2]**

**[1,2]Computer Science and Engineering, Anna University, SMK Fomra Institute of Technology, Chennai, Tamil Nadu, India**

## Abstract

To facilitate extensive collaborations, today's organizations raise increasing needs for information sharing via on-demand information access. Information Brokering System (IBS) atop a peer-to-peer overlay has been proposed to support information sharing among loosely federated data sources. It consists of diverse data servers and brokering components, which help client queries to locate the data servers. However, many existing IBSs adopt server side access control deployment and honest assumptions on brokers, and shed little attention on privacy of data and metadata stored and exchanged within the IBS. In this article, we study the problem of privacy protection in information brokering process. We first give a formal presentation of the threat models with a focus on two attacks: attribute-correlation attack and inference attack. Then, we propose a broker-coordinator overlay, aa well as two schemes, automaton segmentation scheme and query segment encryption scheme, to share the secure query routing function among a set of brokering servers. With comprehensive analysis on privacy, end to- end performance, and scalability, we show that the proposed system can integrate security enforcement and query routing while preserving system-wide privacy with reasonable overhead.

***Index Terms*** *— Access control, Attack, Query, information sharing, privacy.*

## 1. Introduction

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. As a data provider,a participating organization would not assume free orcomplete 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. However, the centralized DBMS still introduces data heterogeneity, privacy, and trust issues. While being considered a solution between "sharing nothing" and "sharing everything", peer-to-peer information sharing framework essentially need to establish pair wise client-server relationships between each pair of peers, which is not scalable in large scale collaborative sharing. 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 [11], 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" .Experimental results show that PPIB provides comprehensive privacy protection for on-demand information brokering, with insignificant overhead and very good scalability.

## 2. The Problem

### 2.1 Vulnerabilities and the Threat Model

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 meta data 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 in the querying content. For example, a query about AIDS treatment reveals the (possible) disease of the requestor.

Attribute-correlation attack :Predicates of an XML query describe conditions that often carry sensitive and private data 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". 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.

### 2.2 Solution Overview

To address the privacy vulnerabilities in current information brokering infrastructure, we propose a new model, namely Privacy Preserving Information Brokering (PPIB). PPIB has three types of brokering components: brokers, coordinators, and a central authority (CA). The key to preserving privacy is to divide and allocate the functionality to multiple brokering components in a way that no single component can make a meaningful With privacy-preserving considerations, we cannot let a coordinator hold any rule in the complete form. Instead, we propose a novel automaton segmentation scheme to divide (metadata) rules into segments and assign each segment to a coordinator. Coordinators operate collaboratively to enforce secure query routing. A query segment encryption scheme is further proposed to prevent coordinators from seeing sensitive predicates. inference from the information disclosed to it.

### 2.3 Preliminaries

#### 2.3.1 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 [37] 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.

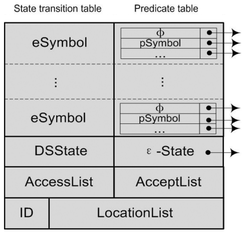

Fig.1. Data structure of an NFIV

## 3. Preserving Query Brokering Schema State

### 3.1. Automaton Segmentation

3.1.1 Segmentation: The atomic unit in the segmentation is an NFA state of the original automaton. Each segment is allowed to hold one.

3.1.2 Deployment: We employ physical brokering servers, called coordinators, to store the logical segments. To reduce the number of needed coordinators, several segments can be deployed on the same coordinator using different port numbers. several NFA states.

3.1.3 Replication: Since all the queries are supposed to be processed first by the root coordinator, it becomes a single point of failure and a performance bottleneck. For robustness, we need to replicate the root coordinator as well as the coordinators at higher levels of the coordinator tree. Replication has been extensively studied in distributed systems.

3.1.4 Handling the Predicates: In the original construction of NFA (similarly as described in QFilter [36] and QBroker [9]), a predicate table is attached to every child state of an NFA stateAs shown in Fig.1.

### 3.2. Query Segment Encryption

Informative hints can be learned from query content, so it is critical to hide the query from irrelevant brokering servers.However, in traditional brokering approaches, it is difficult, if not impossible, to do that, since brokering servers need to view query content to fulfill access control and query routing. Fortunately, the automaton segmentation scheme provides new opportunities to encrypt the query in pieces and only allows a coordinator to decrypt the pieces it is supposed to process. The query segment encryption scheme proposed in this work consists of the preencryption and postencryption modules, and a special commutative encryption ule for processing the double-slash ("//") XPath step the query

3.2.1 Level-Based Preencryption: According to the automaton segmentation scheme, query segments are processed by a set of coordinators along a path in the coordinator tree. A straightforward way is to encrypt each query segment with the public key of the coordinator specified by the scheme.

3.2.2 Postencryption: The processed query segments should also be protected from the remaining coordinators in later processing, so postencryption is necessary

3.3.3 Commutative Encryption for "//" Handling: When a query has the descendant-or-self axis (i.e., "//" in XPath expressions), a so-called mismatching problem occurs at the coordinator who takes the "//" XPath step as input. This is because that the "//" XPath step may recursively accepts several tokens until it finds a match. Consequently, the coordinator with the private level key may not be the one that matches the "//" token, and vice versa.

### 3.3. The Overall PPIB Architecture

The architecture of PPIB is shown in Fig. 7, where users and data servers of multiple organizations are connected via a broker-coordinator overlay. 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 KQ.KQ 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 role of the authenticated user to attach to the encrypted query;(2) it creates a unique QID for each query, and attaches QID,<KQ>pkDS and its own address to the query for data servers to return data.

• 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. At the leaf coordinator, all querysegments should be processed and reencrypted by the public key of the data server. • 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 ,KQ to the broker that originates the query.

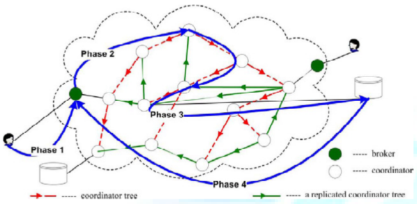

Fig. 2. Query brokering process in four phases

## 4. Maintenance

### 4.1. Key Management

The CA is assumed for offline initiation and maintenance.With the highest level of trust, the CA holds a global view aboutall the rules and plays a critical role in automaton segmentation and key management. There are four types of keys used in the brokering process: query session key KQ ,public/private level keys{pk,sk} , commutative level keys{e,d} , and public/private data server keys {pkDS,skDS}.

### 4.2. Brokering Servers Join/Leave

Brokers and coordinators, contributed by different organizations, are allowed to dynamically join or leave the PPIB system. Besides authentication, a local broker only works as an entrance to the coordinator overly. It stores the address of the root coordinator (and its replica) for forwarding the queries.

### 4.3.Metadata Update

ACR and index rules should be updated to reflect the changes in the access control policy or the data distribution in an organization.

## 5. Privacy and Security Analysis

Three most common types of attackers, local and global eavesdroppers, malicious brokers and malicious coordinators. Collusive Coordinators: Collusive coordinators deviate from the prescribed protocol and disclose sensitive information. Consider a set of collusive (corrupted) coordinators in the coordinator tree framework. Even though each coordinator can observe traffic on a path routed through it, nothing will be exposed to a single coordinator because (1) the sender viewable to it is always a brokering component; (2) the content of the query is incomplete due to query segment encryption; (3) the ACR and indexing information are also incomplete due to automaton segmentation; (4) the receiver viewable to it is likely to be another coordinator. However, privacy vulnerability exists if a coordinator makes reasonable inference from additional knowledge.

## 6. Perform and Analysis

### 6.1.End-to-End Query Processing Time

End-to-end query processing time is defined as the timeelapsed from the point when query arrives at the broker untilto the point when safe answers are returned to the user.

1) Average Query Processing Time at the Coordinator:

2) Average Network Transmission Latency:

3) Average Number of Hops:

4) End-to-End Query Processing Time:

### 6.2. System Scalability

We evaluate the scalability of the PPIB system against complicity of ACR, the number of user queries, and data size (number of data objects and data servers).

1) Complicity of XML Schema and ACR:

2) Number of Queries:

3) Data Size: When data volume increases (e.g., adding more data items into the online auction database), the number of indexing rules also increases.

## 7. Conclusion

In this paper, 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 andquery forwarding while providing comprehensive privacy protection.Our analysis shows that it is very resistant to privacyattacks. End-to-end query processing performance and system scalability are also evaluated and the results show that PPIB is efficient and scalable.Many directions are ahead for future research. First at present, site distribution and load balancing in PPIB are conducted in an ad-hoc manner. Our next step of research is to design an automatic scheme that does dynamic site distribution.Several factors can be considered in the scheme such as the workload at each peer, trust level of each peer, and privacy conflicts between automaton segments. Designing a scheme that can strike a balance among these factors is a challenge. Second, we would like to quantify the level of privacy protection achieved by PPIB. Finally, we plan to minimize (or even eliminate) the participation of the administrator node, who decides such issues as automaton segmentation granularity. A main goal is to make PPIB self-reconfigurable.

## References

[1] I. Stoica, R. Morris, D. Liben-Nowell, D. Karger, M. Kaashoek, F.Dabek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookupprotocol for Internet applications," IEEE/ACM Trans. Netw., vol. 11,no. 1, pp. 17–32, Feb. 2003.

[2] R. Huebsch, B. Chun, J. Hellerstein, B. Loo, P. Maniatis, T. Roscoe,S. Shenker, I. Stoica, and A. Yumerefendi, "The architecture of PIER:An Internet-scale query processor," in Proc. CIDR, 2005, pp. 28–43.

[3] O. Sahin, A. Gupta, D. Agrawal, and A. E. Abbadi, "A peer-to-peerframework for caching range queries," in Proc. ICDE, Boston, MA,USA, 2004, pp. 165–176.

[4] A. Carzaniga, M. J.Rutherford, andA. L.Wolf, "Arouting scheme forcontent-based networking," in Proc. INFOCOM, Hong Kong, 2004,pp. 918–928.

[5] Y. Diao, S. Rizvi, and M. J. Franklin, "Towards an Internet-scale XMLdissemination service," in Proc. VLDB Conf., Toronto, Canada, Aug.2004.

[6] G. Koloniari and E. Pitoura, "Content-based routing of path queries inpeer-to-peer systems," in Proc. EDBT, 2004, pp. 29–47.

[7] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searcheson encrypted data," inProc. IEEE Symposiumon Security and Privacy,2000, pp. 44–55.

[8] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keywordsearch over encrypted cloud data," in Proc. ICDCS'10, Genoa, Italy,pp. 253–262.

[9] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficientlysearchable encryption," in Proc. CRYPTO'07, Santa Barbara,CA, USA, pp. 535–552.

[10] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized private keywordsearch over encrypted data in cloud computing," in Proc. ICDCS,Minneapolis, MN, USA, 2011, pp. 383–392.

[11] D. Boneh and B. Waters, "Conjunctive, subset, and range queries onencrypted data," in Proc. TCC'07, Amsterdam, The Netherlands, pp.535–554.

[12] C. Gentry, "Fully homomorphic encryption using ideal lattices," inProc. STOC'09, Bethesda, MD, USA, pp. 169–178.

[13] M. J. Freedman,Y. Ishai, B. Pinkas, and O. Reingold, "Keyword search and oblivious pseudorandom functions," in Proc. TCC'05, Cambridge, MA, USA, pp. 303–324.

[14] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for web transactions," ACM Trans. Inf. Syst. Security, vol. 1, no. 1, pp. 66–92, 1998.

[15] P. F. Syverson, D. M. Goldschlag, and M. G. Reed, "Anonymous connectionsand onion routing," in Proc. IEEE S&P, 1997, pp. 44–54.