Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.472 – 479

RESEARCH ARTICLE

Effective Concept of Implementing Secure DIBS using Query Segmentation

Mr. Ashutosh Kamble[1], Prof. Deepak Kapgate[2]

[1]Department of CSE, GHRAET, Nagpur (M.S), India

[2]Department of CSE, GHRAET, Nagpur (M.S.), India

[1] kontact.ashu@gmail.com; [2] deepakkapgate32@gmail.com

Abstract— ***Nowadays it is becoming common for enterprises to cooperate in certain areas and compete. To facilitate this requirement enterprise information systems are designed as distributed network systems, where existing information systems and new components are connected together via a middleware. One such widely accepted and commonly used system is Distributed information brokering System (DIBS) which is a peer-to-peer overlay network that comprises diverse data servers and brokering components helping client queries locate the data. Literature on information integration across such systems tacitly assumes that the data can be revealed to the others. However, there is an increasing need for sharing information across autonomous entities in such a way that no information apart from the answer to the query is revealed. We formalize the notion of minimal information sharing across such databases, and develop protocol such that no single entity possess the complete data that can be misused unless it is passed on to the requested entity securely.***

Keywords— ***distributed network; information systems; DIBS; autonomous entities; minimal information sharing***

I. INTRODUCTION

The Information brokering system has a data broker, also called an information broker or information reseller that collects information about consumers and sells that information to other organizations. The information can be collected from a variety of public and non-public sources including courthouse records, website cookies and loyalty card programs. Typically, brokers create profiles of individuals for marketing purposes and sell them to businesses who want to target their advertisements and special offers. At present, there is no legislation that requires a data broker to share the information they have gathered with the consumers they have profiled. Data brokers may refer to themselves as being database marketers or consumer data analytics firms.

In contrast with the situations when the information seeker knows where the needed data is located, a Distributed Information Brokering System (DIBS) needs to help each information seeking query "locate" the corresponding data source(s) or information. Although DIBSs face daunting maintenance challenges, the data source locative capability is highly desired in many important applications, such as emergence health care. And the privacy enhancing measures must be integrated with the query routing operations. In DIBS, data owners collect data independently and manage it with autonomous data servers. While providing data access to legitimate users, data servers have to release certain privacy-sensitive information that needs to be protected.

But all existing systems view or handle access control and query brokering as two orthogonal issues. Access control is a security issue that concerns information confidentiality, while query brokering is a system issue that concerns costs and performance.[11][12] Most of the existing systems work on these two extremes of the spectrum. So there emerges a strong need for an intermediate solution which will be efficient to serve both client queries efficiently in a secure way

with minimum system resource footprint. One such solution is the use of access control over all participating entities using authentication and then passing the client query through various segments from one node to other in various segments such that none of the segment has access to the complete data which otherwise would have been treated as compromised.

II. LITERATURE SURVEY

Information systems are in most cases designed as distributed network systems. A globalisation in a sphere of business creates need for decentralised systems with a correct data distribution, distributed processing, reservation of resources and a reliable communication infrastructure. The distributed information systems are designed as a network of communicating and partially independent components where each component performs its specific task, by itself or with help of other components. In this view, a process in the information system represents a group of components and a scheme of their interaction.[1] Marek Rychl´y and Jaroslav Zendulka proposed a framework for distributed information systems which can be formally represented as asynchronous distributed network models.

Fengjun Li and Bo Luo states that though access control is required in most if not all DIBS. The popular approach[7][8] of XML access control model is proposed where users are members of appropriate roles; and an access control policy consists of a set of role-based 5-tuple access control rules (ACR): R = {subject, object, action, sign, type}, where (1) subject is a role to whom an authorization is granted; (2) object is a set of XML nodes specified by XPath; (3) action is one of \read,"\write," and \update"; (4) sign 2 f+; □g refers to access \granted" or \denied," respectively; and (5) type {LC;RC} refers to either "Local Check" (i.e., authorization is only applied to attributes or textual data of context nodes "self::text() | self::attribute()"), or "Recursive Check" (i.e., authorization is applied to context nodes and propagated to all descendants "descendant-or-self::node()").

Napster, Gnutella and Kazza has spurred much attention to peer-to-peer (P2P) computing [3]. Peer-to-peer computing refers to a form of distributed computing that involves large number of autonomous computing nodes (the peers) that cooperate to share resources and services. As opposed to traditional client-server computing, nodes in a P2P systems have equal roles and act as both data providers and data consumers. Furthermore, such systems are highly dynamic in that nodes join or leave the system and change their content constantly. Motivated by the fact that XML has evolved as a standard for publishing and exchanging data in the internet, we assume that the nodes in a P2P system store and share XML documents. Such XML documents may correspond either to native XML documents ot to XML-based descriptions of local services or datasets. Such datasets may be stored in local to each node databases supporting diverse data models and exported by the node as XML data.

III. QUERY SEGMENTATION ALGORITHM

Informative hints may be learnt from the content of the query, so it is critical to protect the query from being intercepted by irrelevant brokering servers. However, it is difficult, if not impossible, to hide the query content from any of the brokers as they are needed to search or match a string present in the metadata or the database, based on which the broker requests coordinator for the data in traditional brokering approaches. Since it is responsible for matching the query with the database index rules which enforce query routing or authorization. In our study, the automaton segmentation scheme provides a new encryption opportunity to encrypt the query in pieces and allow each coordinator to decrypt the piece it is about to process. The query segment scheme consists of the string matching, content validation, and a special secret key based authentication module for processing.

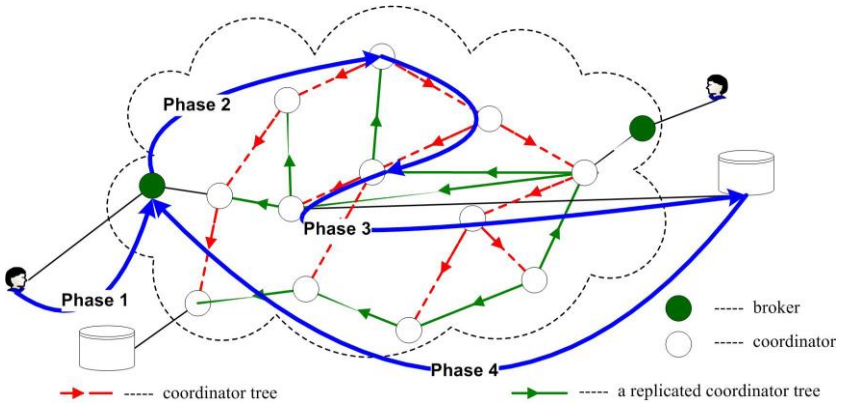

Fig1: The four phases of Query Segmentation

Step 1. User submits a query in form of a string to a broker.
Step 2 . The broker signs this query with his ID and forwards it to the coordinator.
Step 3. The coordinator validates brokers ID and submits this query to the database via metadata if required.
Step 4. The unique secret key present in the database for data relevant to the requested query is fetched and passed from coordinator to user by the broker way.

IV. IMPLEMENTATION DETAILS

The working environment for secure Distributed Information Brokering System with proposed Query Segmentation is implemented using ASP.Net Framework with C# for coding with Visual Studio as frontend and SQL Server 2008 for the database.

The implementation is achieved through modular approach for Regional Health information Organization (RHIO) as a case study. These 4 modules are stated below:

- User Module

The Users are classified into two types they are, Data Users and Data Owner depending on their role and restriction on the data that will be passed to the Co-coordinator. The coordinator passes the details via broker where it will be verified with the secret key and thus will get displayed to the users.

- Broker Module

The broker acts as a mediator between coordinator and data Users. The query submitted by a data user gets verified and thus passed to the co-coordinator.

- Coordinator Module

Coordinator performs the global service between two end users via broker. Once a query is verified by the broker with his ID, he submits it to the coordinator who in turn searches and sends the key to the data users via the broker way

- Admin Module

The admin performs the critical roles like registration of data owners and users, brokers and coordinators. He also registers organization in DIBS and manages the database.

UML diagrams of the Implementation

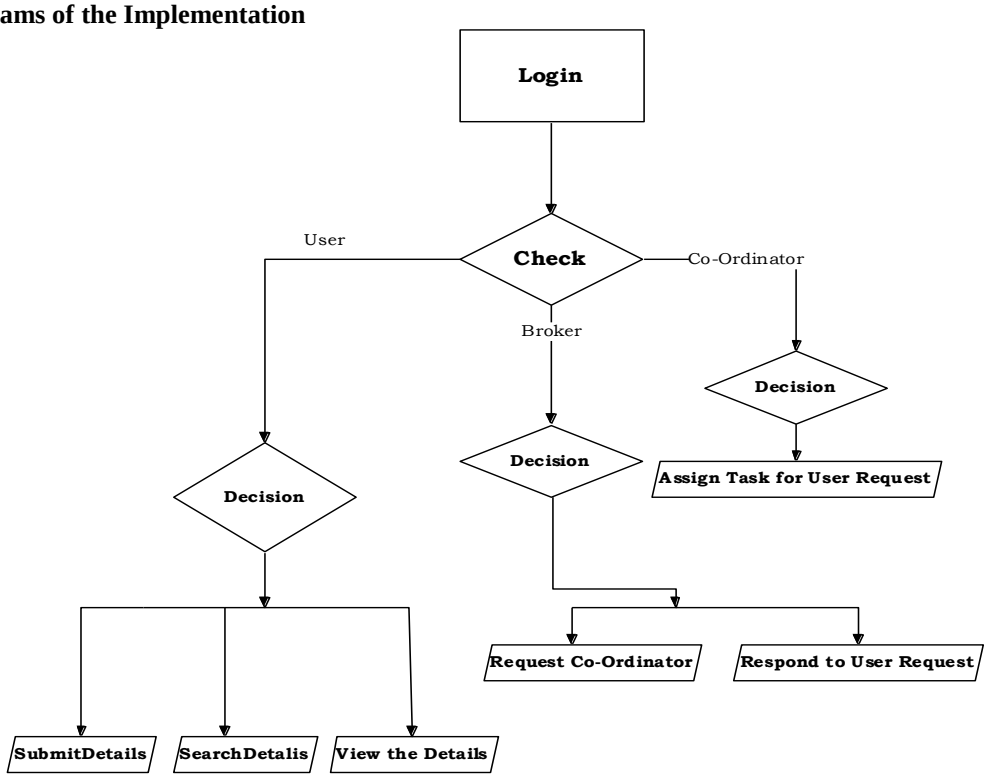

Fig2 : Data Flow Diagram

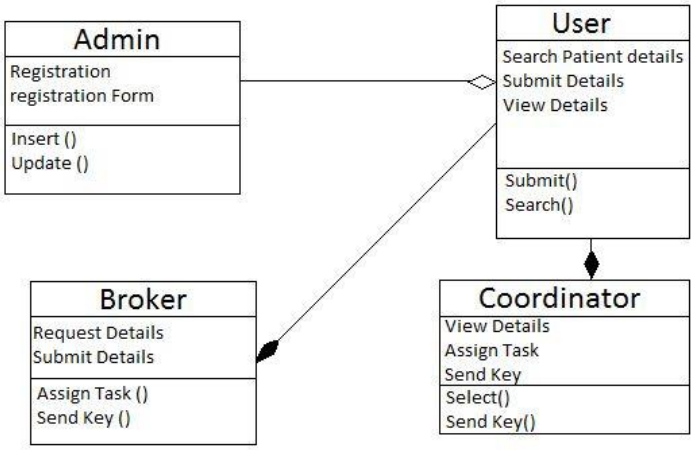

Fig3: Class Diagram

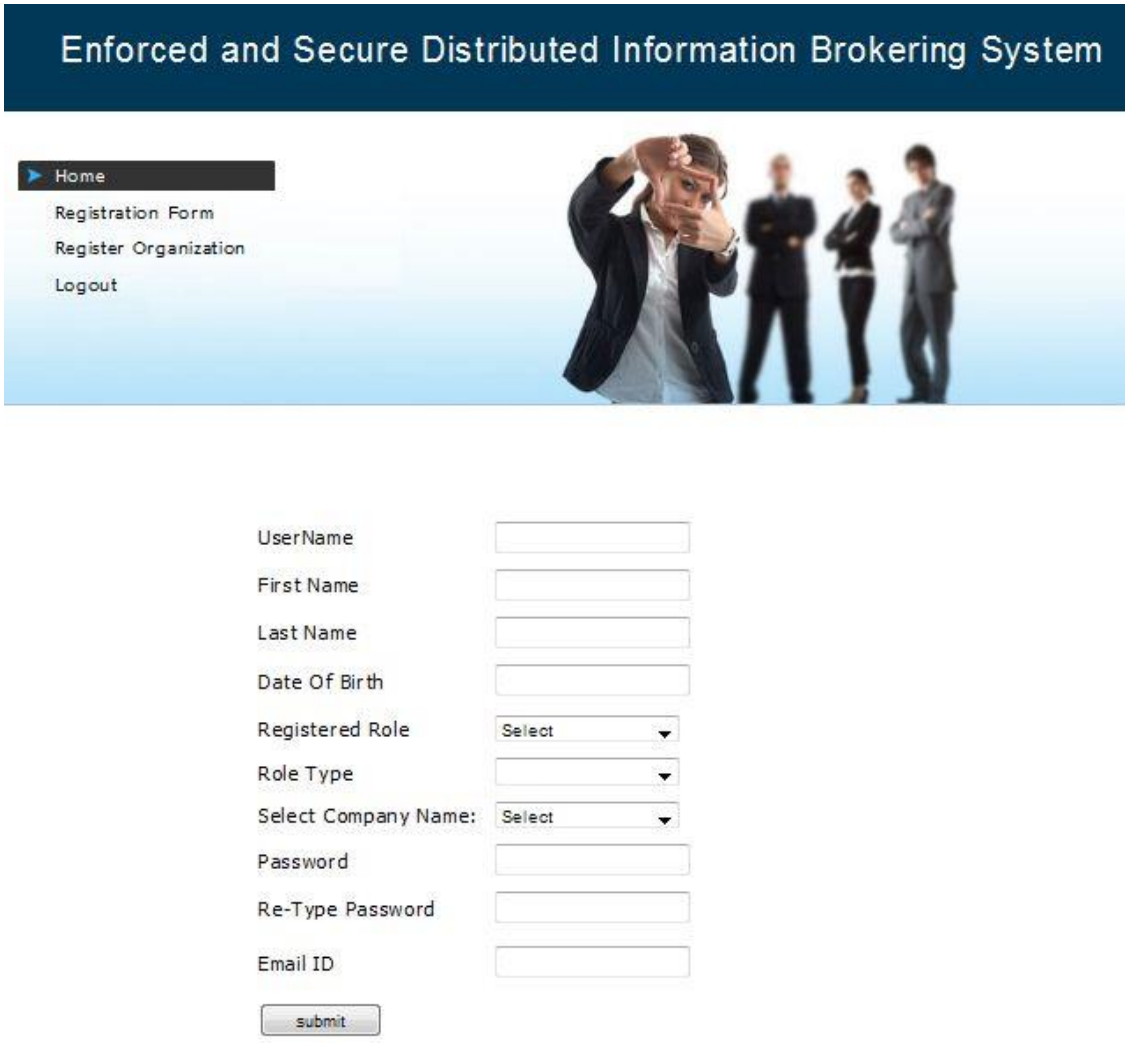

Fig4: Registration Page – A typical role of the Admin

The Above implementation screenshot the shows registration page for a user, broker or a coordinator based on the role that they are eligible to be registered for. Also during registration the admin needs to assign them to an organization. Each role type has a unique ID associated with it which gets assigned to the registrant after successful enrolment.

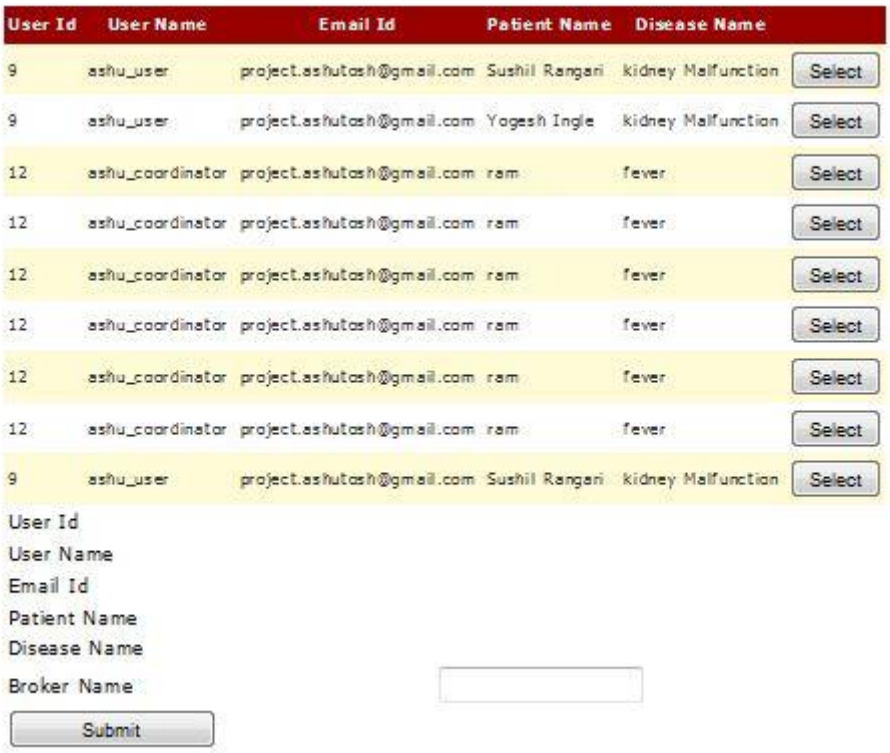

Fig5: Queries submitted by users to the broker as seen in his inbox

Fig6 : Queries forwarded by the broker to the coordinator which are tagged with a secret key and sent to user.

UserId	UserName	EmailId	Patient Name	Requested Disease	Assigned Broker Name	Secret Key
9	ashu_user	project.ashutosh@gmail.com	Sushil Rangari	kidney Malfunction		ukgi74vs
9	ashu_user	project.ashutosh@gmail.com	Sushil Rangari	kidney Malfunction		ukgi74vs
9	ashu_user	project.ashutosh@gmail.com	Sushil Rangari	kidney Malfunction		ukgi74vs
9	ashu_user	project.ashutosh@gmail.com	Sushil Rangari	kidney Malfunction		ukgi74vs
9	ashu_user	project.ashutosh@gmail.com	Sushil Rangari	kidney Malfunction		ukgi74vs

secret key

Validate

Fig7: Data records received by the user in his inbox with a secret key which after authentication displays the data

```
  <PatientName>yuvraj</PatientName>
  <DoctorName>mahesh</DoctorName>
  <Age>35</Age>
  <DiseaseName>cancer</DiseaseName>
  <Email>yuvraj@gmail.com</Email>
  <DiseaseDescription>cancer aaaa...</DiseaseDescription>
  <SecretKey>0mau6422</SecretKey>
</New_Table>
<New_Table>
  <id_no>3</id_no>
  <PatientName>ashutosh</PatientName>
  <DoctorName>sushil</DoctorName>
  <Age>26</Age>
  <DiseaseName>fever123</DiseaseName>
  <Email>project.ashutosh@gmail.com</Email>
  <DiseaseDescription>feswhfv jshfdelkwsehfg lskefjlwv skl
  <SecretKey>ifcwhh8m</SecretKey>
</New_Table>
<New_Table>
  <id_no>4</id_no>
  <PatientName>ram</PatientName>
  <DoctorName>mahesh</DoctorName>
  <Age>44</Age>
  <DiseaseName>fever</DiseaseName>
  <Email>ram@globalinfobase.com</Email>
  <DiseaseDescription>aaaaaaaaaaa</DiseaseDescription>
  <SecretKey>4c0v16t9</SecretKey>
</New_Table>
<New_Table>
  <id_no>5</id_no>
  <PatientName>Sushil Rangari</PatientName>
  <DoctorName>Dr. Warghane</DoctorName>
```

Fig8: XML Code showing stored Secret keys for each user

V. RESULTS CALCULATED

To protect the data privacy, access control policy and data query segmentation, information are divided and distributed among broker-coordinator overlay. As a result, DIBS only require minimal trust (or honesty) in each coordinator, as shown in Figure 9, where \Hide"means \no need to trust". It is clear that whenever the system's level of trust in each brokering component can be lowered without hurting privacy, thus the system's privacy protection capability will be enhanced.

Privacy Type	User Location	Query Content	Data Object Distribution	Access Control Policy	Query Segmentation Policy	Index Information
Broker	Trust	(Partially)Trust	Hide	Hide	NA	Hide
Coordinator	Hide	(Partially)Trust	Hide	(Partially)Trust	(Partially)Trust	Hide
Database	Hide	Trust	Trust	Trust	Trust	Trust

Fig9 : Brokering components showing restricted trust on systems privacy

VI. CONCLUSION

This paper describes the distributed information brokering systems with a strong formal base with a goal to outline some of the critical vulnerabilities of the system. We therefore propose introduction of a new approach to preserve privacy of data in information brokering. Through an innovative Query segmentation scheme, in-network access control, and secret key based authentication, our system integrates security enforcement and query forwarding while providing comprehensive protection. Our analysis shows that it I very resistant to privacy concerns where trust factor is ever changing from system wide brokers. it is also efficient in query processing scalable to fit for small to medium organizations, and considerably light on resource footprint.

REFERENCES

[1] *Distributed Information System as a System of Asynchronous Concurrent Processes* Marek Rychl´y and Jaroslav Zendulka

[2] Fengjun Li, Bo Luo, Peng Liu Dongwon Lee and Chao-Hsien Chu, "*Enforcing Secure and Privacy-Preserving Information Brokering in Distributed Information Sharing*", IEEE TRANSCATIONS ON INFORMATION FORENSICS AND SECURITY, 2013.

[3]Georgia koloniari and Evaggelia Pitoura. *Content-based Routing of Path Queries in Peer-to-Peer Systems*.

[4] Fengjun Li, Bo Luo, Peng Liu, Dongwon Lee, and Chao-Hsien Chu. "*Automaton Segmentation: A New Approach to Preserve Privacy in XML Information Brokering*", CCS'07, October 29–November 2, 2007, Alexandria, Virginia, USA.

[5] 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.

[6] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, "CoolStreaming/DONet. A data-driven overlay network for efficient live media streaming," in Proceedings of IEEE INFOCOM, 2005.

[7] A. C. Snoeren, K. Conley, and D. K. Gifford, "Mesh-based content routing using XML," in SOSP, pp. 160–173, 2001.

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

[9]E. Damiani, S. Vimercati, S. Paraboschi, and P. Samarati. A fine-grained access control system for XML documents. ACM Trans. Inf. Syst. Secur., 5(2):169-202, 2002

[10]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. [8]

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

[12] M. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces: a new abstraction for information management," SIGMOD Rec., vol. 34, no. 4, pp. 27–33, 2005.

[13] 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.

[14] 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.

[15] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Communications of the ACM, vol. 24, no. 2, 1981.

[16] R. Agrawal, A. Evfimivski, and R. Srikant, "Information sharing across private databases," in Proceedings of the 2003 ACM SIGMOD, 2003.

[17] M. Genesereth, A. Keller, and O. Duschka, "Informaster: An information integration system," in SIGMOD, (Tucson), 1997.

[18] I. Manolescu, D. Florescu, and D. Kossmann, "Answering XML queries on heterogeneous data sources," in VLDB, pp. 241–250, 2001.

[19] J. Kang and J. F. Naughton, "On schema matching with opaque column names and data values," in SIGMOD, pp. 205–216, 2003.

[20] 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 Transactions on Networking, vol. 11 of 1, 2003.
[21] 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 CIDR, pp. 28–43, 2005.
[22] O. Sahin, A. Gupta, D. Agrawal, and A. E. Abbadi, "A peer-to-peer framework for caching range queries," in ICDE, 2004.
[23] A. Carzaniga, M. J. Rutherford, and A. L. Wolf, "A routing scheme for content-based networking," in Proc. Of INFOCOM, 2004.
[24] Y. Diao, S. Rizvi, and M. J. Franklin, "Towards an Internet-scale XML dissemination service," in VLDB Conference, (Toronto), August 2004.
[25] G. Koloniari and E. Pitoura, "Content-based routing of path queries in peer-to-peer systems.," in EDBT, pp. 29–47, 2004.
[26] M. K. Reiter and A. D. Rubin, "Crowds: anonymity for Web transactions," ACM TISS, vol. 1, no. 1, pp. 66–92, 1998.
[27] P. F. Syverson, D. M. Goldschlag, and M. G. Reed, "Anonymous connections and onion routing," in IEEE Symposium on Security and Privacy, (Oakland, California), pp. 44–54, 4–7 1997.
[28] W. Tolone, G.-J. Ahn, T. Pai, and S.-P. Hong, "Access control in collaborative systems," ACM Comput. Surv., vol. 37, no. 1, 2005.
[29] Whiten B. and Taskale G. *An overview of reliable multicast transport protocol* II. IEEE Network 14, 1 (Jan.2000), 37–47.
[30] S. Cho, S. Amer-Yahia, L. V. S. Lakshmanan, and D. Srivastava, "Optimizing the secure evaluation of twig queries.," in VLDB, 2002.
[31] M. Murata, A. Tozawa, and M. Kudo, "XML access control using static analysis," in ACM CCS, 2003.
[32] A. Schmidt, F. Waas, S. Manegold, and M. Kersten. "*The XML Benchmark Project*". Technical report, INS-R0103, CWI, April 2001.

