ISSN:0975-9646

A review on Security in Distributed Information Sharing

Swati Upadhyay
Dept. of Computer Science
IES College of Technology
Bhopal, India

Prof J p Maurya
Dept. of Computer Science
IES College of Technology
Bhopal, India

***Abstract*— In recent year's privacy preserving data mining has emerged as a very active research area in data mining. Over the last few years this has naturally lead to a growing interest in security or privacy issues in data mining. More precisely, it became clear that discovering knowledge through a combination of different databases raises important security issues. Privacy preserving data mining is one of the most demanding research areas within the data mining community. In many cases, multiple parties may wish to share aggregate private data, without disclosing sensitive information at their end. In this we are discussing some techniques related to security and privacy preservation of information and its sharing.**

Keywords*— Privacy preservation, data security, information sharing, data attacks.

I. INTRODUCTION

The Internet is becoming an increasingly vital tool in our everyday life for professional and personal users and them becoming more numerous. In the current trends business is progressively more conducted over the Internet. Now a day's many organizations are capable to collect large amount of information and wants to securely share it with their valuable customers or clients. Unfortunately, neither model is suitable for many newly emerged applications like law or healthcare enforcement information sharing, in which companies share information in a controlled and conservative manner just because of business considerations or legal reasons [1].

Data provider wants to participating organization would not assume free or complete sharing with others, since this data is originally private or commercially proprietary, or both. As a substitute, it requires retaining full control over the data and the access to the data. Meanwhile, as a consumer, a user provider requesting data from other providers expects to preserve her privacy in the querying process. In such a condition sharing a complete copy of the data with others or uploading data into a centralized repository becomes impractical. To manage this secure and privacy protected system is required. However, the centralized DBMS still introduces data privacy, heterogeneity, and trust issues. While being considered a solution between 'sharing nothing' and 'sharing everything', peer-to-peer information sharing scheme essentially need to establish pair wise client-server relationships between each pair of peers that is not scalable in large scale collaborative sharing [1].

Over the past few years state of the art research in privacy preserving data mining has concentrated itself along two major lines: data which is horizontally distributed and data which is vertically distributed. Horizontally partitioned data is data which is homogeneously distributed. This means that all tuples of data yield over the same item or feature set. Indeed this boils down to numerous data sites collecting the same kind of information over different individuals. For instance, numerous superstores with sensitive sales data may wish to coordinate among themselves in knowing aggregate trends without leaking the trends of their individual stores. This requires secure protocols for sharing the information across the various parties. This data may be scattered multiple places in two ways across different sites: Horizontal partition and Vertical partition. Horizontal partition means, where different sites have different sets of records containing the similar attributes. Vertical partition scheme means, where various sites have different attributes of the same sets of records [2]. Privacy preserving data mining method on the decision tree over horizontally partitioned data using (Un-trusted third party)UTP.

Privacy-preserving information publication attracts nice attention of the community in recent years due to the considerations regarding privacy breaching problems in information publication method and to forestall linking attack, a primary attack in information publication, quite an few Privacy preserving data processing (PPDP) strategies are planned, together with Generalization and randomization [3]. Most of them specialize in static past information set publication and can disclose sensitive info once data is re-published. Privacy is consider as protecting individual data tuples as well as protecting attributes and values of attributes. So each party will reveal as little as possible about its data while still constructing an applicable distributed assessment tree. The only article that is known with reference to the tree by all parties is its structure and which party is responsible for each decision node. More precisely, which party possesses the attribute used to make the decision, except not which characteristic? [4].

Privacy preserving data mining (PPDM) addresses the matter of developing correct models concerning mass knowledge while not access to specific data in individual knowledge record. A wide studied perturbation-based PPDM approach introduces random perturbation to individual values to preserve privacy before knowledge area unit printed. Previous solutions of this approach area unit restricted in their inexplicit assumption of single-level trust on knowledge miners and Multi level trust-Privacy

preserving data mining (MLT-PPDM) permits knowledge homeowners to come up with otherwise discomposed copies of its knowledge for various trust levels. The key problems lies in preventing the information miners from combining copies at completely different trust levels to collectively reconstruct the initial data a lot of correct than what's allowed by the information owner. [5].

a. PPIB

Privacy Preserving Information Brokering (PPIB) is an overlay infrastructure consisting of two types of brokering components, brokers and coordinators. The brokering are mainly responsible for user authentication and query forwarding, the broker performs the role who can act between the coordinator and the data Users. The request i.e. submitted from the data user will be verified and thus it will be passed to the coordinator. The coordinators which 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, each holding a segment of access control automaton and routing guidelines, are mainly responsible for access control and query routing [1].

PPIB takes an innovator automaton segmentation approach to privacy protection. In particular, two critical forms of privacy, are generally known as query content privacy and data object distribution privacy (or data location privacy). These forms are enabled by a novel automaton Segmentation scheme, with a “little" help from an assisting query segment encryption scheme. To prevent inquisitive or unserviceable coordinators from inferring private information, they [6] design two novel schemes: (a) to segment the query brokering automata, and (b) to encrypt corresponding query segments. This system will provide full capability to wage in network access control and to path queries to the right data sources. These schemes ensure that inquisitive or unserviceable coordinator is not capable to collect sufficient information to guess privacy, like “which data need to be queried, where located and what the policies to access data are”. Privacy Preserving Information Brokering (PPIB) enables wide-ranging security and privacy protection for claimed information brokering, with minor overhead and major scalability.

b. INFORMATION BROKERING SYSTEM

Information Brokerage System or IBS is a system provide scalability and server autonomy. In IBS infrastructure given broker and coordinator, the brokers are no longer fully trustable. Thus, system may be abuse by insider or outsider. A distributed system providing data access through a set of brokers is referred to as Information Brokering System (IBS). As shown in Fig. 1, applications atop IBS always involve some sort of consortium (e.g., RHIO) among a set of organizations. Databases of various organizations are connected through a set of brokers, and metadata are “pushed” to the local brokers, which further “advertise” the metadata to other brokers. Queries are sent to the local broker and routed according to the metadata until reaching the right data server(s). In this method a large number of information sources in different organizations are loosely federated to provide a unified, transparent, and on-demand data access [1].

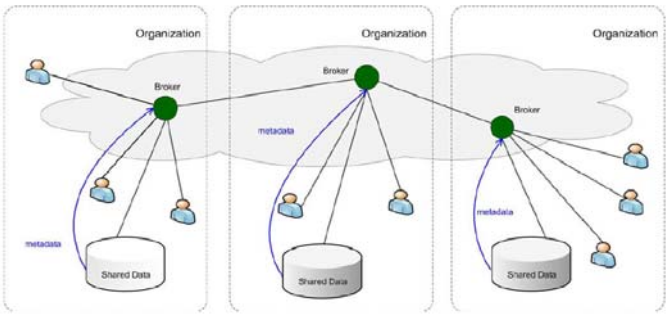

Fig. 1. Overview of the IBS infrastructure [1].

Although the IBS scheme 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.

Privacy type	local eaves-dropper	global eavesdropper	malicious broker	collusive coordinators
User Location	Exposed	Exposed	Exposed	Protected
Query Content	Protected	Exposed	Exposed	Exposed only with compromised root coordinator
AC Policy	Protected	Protected	Protected	Exposed if path coordinators collude
Index Rules	Protected	Protected	Protected	Exposed if path coordinators collude
Data Distribution	Protected	Protected	Protected	Exposed if path coordinators collude
Data Location	Protected	Beyond suspicion	Protected	Exposed with malicious leaf coordinators

Table 1. Privacy Policy Caused by Four Attacks

c. SECURITY AND PRIVACY NEED FOR PPIB

In information brokering scenario, there are three types of entrepreneur called as data owners, data providers, and data requestors. Each entrepreneur has its own privacy:

1) the privacy of a data owner (e.g. a patient) is identifiable data and the information keep together by this data (e.g. medical records). Data owners usually sign stiff privacy agreements with data providers to protect their privacy from unauthorized disclosure/user.

2) Data providers store collected data, and create two types of metadata these are known as routing metadata and access control metadata.

3) Data requestors divulge identifiable and private information in the querying process.

Assume that for the brokers, two types of opponent, outside attackers and curious or corrupted brokering components. Outside attackers passively eavesdrop

communication channels. Curious or corrupted brokering components follow the protocols be seemingly to accomplish their functions, others' private information from the information disclosed in the querying process. Data providers push routing and access control metadata to brokers [8], which also strut queries from requestors. Therefore, a curious or corrupted brokering server could: (1) learn query content and query location by impede a local query; (2) learn routing metadata and access control metadata from local data servers and other brokers; (3) learn data location from routing metadata it holds Although attacker may not obtain plaintext data over encrypted data, they can still learn query location and data location from eavesdrop. The attacks into two major classes: (1) the attribute-correlation attack and (2) inference attack.

Attribute-correlation attack: An attacker prevents a query, which typically contains several predicates. Each predicate describes a condition, which sometimes involves sensitive and private data like name, credit card number, etc..

Inference attack: Attacker some techniques and result more than one other type of sensitive information so more sever, and further associates to learn explicit and implicit knowledge about entrepreneur.

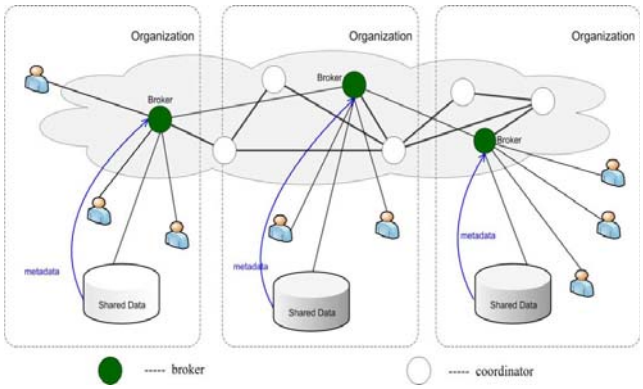

Fig. 2. Architecture of PPIB.

IBS work is designed with user and data privacy. Such privacy protection requirements, therefore a novel IBS, named as Privacy Preserving Information Brokering system (PPIB). As shown in Figure 2, PPIB contains a broker-coordinator overlay network, in which the brokers are amenable for onus transmission user queries to coordinators concatenated in tree structure while preserving privacy. The coordinators, each holding a segment of access control automaton and routing guidelines, are mainly responsible for access control and query routing [6].

Privacy issues of user and data during the design stage is considered and concluded that existing information brokering systems suffer from a spectrum of vulnerabilities associated with user privacy, data privacy, and metadata privacy. The rest of paper is organized as follows. In Section II describes about background. In Section III describes about related work of privacy preserving IBS. Section IV describes about conclusion.

II. BACKGROUND

Privacy concerns arise in inter-organizational information brokering since one can no longer assume brokers controlled by other organizations are fully trustable. Despite the potential benefits, one crucial issue pertaining to the Internet-wide collaborative neural network learning is the protection of data privacy for each participant. In particular, the participants from different trust domains may not want to disclose their private data sets, which may contain privacy or proprietary information, to anybody else. In order to embrace the Internet wide collaborative learning, it is imperative to provide a solution that allows the participants, who lack mutual trust, to conduct neural network learning jointly without disclosing their respective private data sets.

The privacy-preserving public auditing is uniquely integrate the homomorphic non-linear authenticator with random masking technique. In the non-linear blocks in the server's response is masked with randomness produced the server. With random masking, the Third party auditor (TPA) no further has all the necessary information to build up a correct group of non-linear equations and therefore cannot derive the user's data content, no substance how countless linear combinations of the same set of file blocks can be together. Besides the correctness justification of the block authenticator pairs can still be carried out in a new way even with the presence of the randomness [7]. With the establishment of privacy- preserving public auditing, the TPA may at the same time as handle numerous auditing upon different users entrustment. The individual auditing of these assignments for the TPA can be tedious and very inefficient.

III. RELATED WORK

This section describes some related work to privacy preservation and brokering information sharing.

In year 2013, Li, Fengjun et al [1] presented a general solution to the privacy-preserving information sharing problem. First, to address the need for privacy protection, we propose a novel IBS, namely Privacy Preserving Information Brokering (PPIB). It is an overlay infrastructure consisting of two types of brokering components, brokers and coordinators. The brokers, acting as mix anonymizer, are mainly responsible for user authentication and query forwarding. The coordinators, concatenated in a tree structure, enforce access control and query routing based on the embedded nondeterministic finite automata—the query brokering automata. To prevent curious or corrupted coordinators from inferring private information, they design two novel schemes to segment the query brokering automata and encrypt corresponding query segments so that routing decision making is decoupled into multiple correlated tasks for a set of collaborative coordinators. While providing integrated in-network access control and content-based query routing, the proposed IBS also ensures that a curious or corrupted coordinator is not capable to collect enough information to infer privacy, such as "which data is being queried", "where certain data is located", or "what are the access control policies", etc. Experimental results show that PPIB provides comprehensive privacy protection for on-demand

information brokering, with insignificant overhead and very good scalability [1].

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. This 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 [1].

In 2012 by Alka Gangrade and Ravindra Patel gives the concept about the two layer protocol uses an Un-trusted Third Party (UTP) and explains how to build privacy preserving two-layer decision tree classifier, where database is horizontally partitioned and communicate their intermediate results to the UTP not their private data. In this protocol, an UTP allows well-designed solutions that meet privacy constraint and achieve suitable performance and finally proposed a new classifier using two-layer architecture that enables Secure multiparty computation (SMC) by hiding the identity of the parties' attractive part in the classification process using UTP. Further they may describe that intermediate result is calculated by every party individually and send only intermediate result to UTP not the input data. During the communication among UTP and all party final result is carried out. It requires less memory space. Also provides fast and easy calculations. Using this protocol, classification will almost secure and privacy of individual will be maintained. Additional development of the protocol is estimated in the sense that for joining multi-party attributes using a trusted third party can be used [2].

Data privacy through optimal k-anonymization was proposed by Bayardo, and Agrawal. They [3] offered a practical method for determining an optimal - anonymization of a given dataset. An optimal anonymization is one which perturbs the input dataset as little as is necessary to achieve -anonymity, where "as little as is necessary" is typically quantified by a given cost metric. he ability to compute optimal anonymizations lets us more definitively investigate the impacts of various coding techniques and problem variations on anonymization quality. It also allows us to better quantify the effectiveness of stochastic or other non-optimal methods. Demonstrate that despite the problems inherent hardness, provably optimal -anonymizations can be obtained for real census data under two representative cost metrics. This anytime quality allows it to be used to obtain good anonymizations even when an optimal anonymization is out of reach [3].

In 2008 by Bart Kuijpers et al. [4] proposed the complexity analysis, in which case the earlier evaluation method is the more efficient and give an algorithm for privacy preserving ID3 over horizontally partitioned data involving more than two parties. For grid partitioned data, here discuss two different evaluation methods for preserving privacy ID3, that is, first merging horizontally and increasing vertically or first merging vertically and next developing horizontally with the help of these concept the complexity analysis of both algorithms shows that it is more efficient to first merge data horizontally and further develop it vertically than the other way around [4].

In 2012 Yaping Li et. [5] presented Enabling Multilevel Trust in Privacy Preserving Data Mining. Privacy preserving data mining (PPDM) addresses the matter of developing correct models concerning mass knowledge while not access to specific data in individual knowledge record. A wide studied perturbation-based PPDM approach introduces random perturbation to individual values to preserve privacy before knowledge area unit printed. Previous solutions of this approach area unit restricted in their inexplicit assumption of single-level trust on knowledge miners and MLT-PPDM permits knowledge homeowners to come up with otherwise discomposed copies of its knowledge for various trust levels. The primary problem lies in preventing the information miners from combining copies at completely different trust levels to collectively reconstruct the initial data a lot of correct than what's allowed by the information owner [5].

All assumption and expand the scope of perturbation-based PPDM to construction Trust (MLT-PPDM) and also the additional trusty an information jack is, the less rattled copy of the info it will access. Preventing such diversity attacks is that the key challenge of providing MLT-PPDM services. Here address this challenge by properly correlating perturbation across copies at totally different trust levels and prove that this resolution is powerful against diversity attacks with regard to privacy goal. That is, for information miners World Health Organization have access to AN impulsive assortment of the rattled copies, this resolution stop them from conjointly reconstructing the first information additional accurately than the most effective effort exploitation a person copy within the assortment. This resolution permits an information owner to come up with rattled copies of its data for impulsive trust levels on demand. This feature offers information house owners most flexibility [5].

Srinivas, D. propose a privacy-preserving public auditing system for data storage security in Cloud Computing. He tries to utilize the homomorphic non-linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process that not only reduces the burden of cloud user from the tedious and possibly pricey auditing task, but also alleviates the users' terror of their outsourced data security. Taking into account TPA may concurrently handle multiple audit sessions from dissimilar users for their outsourced data files, he further extend this privacy-preserving public auditing protocol into a multi-user scenario, where the TPA can perform multiple auditing tasks in a batch manner for better effectiveness. Extensive examination shows that this schemes are almost certainly secure and highly efficient [7].

In year 2006, In-broker access control: Towards efficient end-to-end performance of information brokerage systems was proposed by F. Li et al [8]. Moreover, their in-broker brokerage system provides a full replication of access control and location information among all the brokers, which brings higher robustness to the whole

system. In traditional information brokerage systems, attackers could block a portion of data sources by DoS attacks. Since the security check is at the DBMS end, the attackers could exhaust the network access and the system resource of the target data server by sending a huge number of identical (or similar) queries which have no access right to the quested data [8].

According to in-broker access control approach, not only the DoS attacking data cannot reach the data server but also the broker can easily recovery with the help of other brokers. However, compared with databases (relational tables or XML trees), the size of access control rules is minimum. In this in-network access control system, it is practically applicable to maintain a full version of access control rules at each broker, i.e. access control function components are fully replicated at each broker. In this way, attackers are not able to block-out a portion of data, since their fake queries are mostly closed-out at the brokers. The brokers endure the incoming attacks, while the brokerage network and data sources are successfully protected. To turn down the system, attackers need to successfully DoS all the brokers. This is practically impossible considering the number of brokers in the system. Since the broker only holds the access control and location information, replication at the broker level is not as expensive as the data level replication in other two architectures [8].

One idea is to build an XML overlay architecture that supports expressive query processing and security checking atop normal IP network. In particular, specialized data structures are maintained on overlay nodes to route XML queries. In [9], a robust mesh has been built to effectively route XML packets by making use of self-describing XML tags and the overlay networks. Kouds et al. also proposed a decentralized architecture for ad hoc XPath query routing across a collection of XML databases [10]. To sharing data among a large number of autonomous nodes, [11] 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.

Saha and Madhusudana presented a study on Protecting Information Brokerage System against Intruders. For the sake of massive scalability of a large messaging fabric they typically want to allow many brokers to be connected together into a network so that many clients can be logically connected together. In Distributed, information sharing the data entities is shared among several inter-communicating computers, wherein at least one of said computers is a server computer, and wherein each computer, which is not a server computer, is a client computer. To address the need for autonomy, federated database technology has been proposed to manage locally stored data with a federated DBMS and provide unified data access. A local broker functions as the ―entrance‖ to the system. It authenticates the requestor and hides his identity from other PPIB components. It would also permute query sequence to defend against local traffic analysis. Coordinators are responsible for content-based query routing and access control enforcement [12].

In year 2011, Kuehn et al [13] presented Interoperability and information brokers in public safety: an approach toward seamless emergency communications. They consider the notion of interoperability as a necessary prerequisite to achieving seamless communication among public safety units and the public for collaborative and coordinative purposes. Under the term ‘emergency communications’, they understand all the bi-directional communication that takes place in an emergency among three groups: (1) public safety agencies, such as police agencies, fire departments, and emergency medical services (EMS) that are in the thick of the disaster; (2) affiliated organizations, such as hospitals, municipal services, telecommunications, logistics, and utilities (e.g., electricity, gas, water) as well as media or schools and other public or private institutions or companies; and (3) the concerned and/or affected public who have been injured, are in imminent danger of injury, or are able to provide critical information about particular situations. The underlying assumption is that these three groups would benefit from greater information sharing in an open communication system [13].

A newly system based on economic incentives to share data or discourage data leakage and a hybrid of code-splitting and secure multi-party computation to provide various assurances of secrecy was offered by Mardziel et al [14]. They also present study on how to incorporate these mechanisms into practical applications, including online social networks, a recommendation system based on users’ qualifications rather than identities, and a “personal information broker” that monitors data leakage over time. They aim to design mechanisms and protocols toward building applications that aim to balance these competing principles of need to know with responsibility to share. They would like both sites and users to be disincentivized or prevented from sharing information with eavesdroppers [14].

IV. CONCLUSION

In recent years privacy preserving data mining has emerged as a very active research area in data mining. There are various types of attackers in the information brokering process. With little 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 paper we are presenting a study of privacy preservation based information brokering data sharing.

REFERENCES

[1] Li, Fengjun, 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, pp. 888 – 900,June 2013.

[2] Alka Gangrade, Ravindra Patel “Privacy Preserving Two-Layer Decision Tree Classifier for Multiparty Databases”, International Journal of Computer and Information Technology, ISSN No: 2277 – 0764, Volume 01, Issue 01, September 2012.

[3] Bayardo, Roberto J., and Rakesh Agrawal "Data privacy through optimal k-anonymization", In IEEE Proceedings of 21st

International Conference on Data Engineering, 2005. ICDE 2005., pp. 217-228., 2005.

[4] Bart Kuijpers, Vanessa Lemmens, Bart Moelans, "Privacy Preserving ID3 over Horizontally, Vertically and Grid Partitioned Data", avrxi-0803.155v1 [cs.db], 11 march 2008. Online: http://arxiv.org/pdf/0803.1555.pdf

[5] Yaping Li, Minghua Chen, Qiwei Li, And Wei Zhang "Enabling Multilevel Trust In Privacy Preserving Data Mining" , IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 9, September 2012.

[6] Kawatghare, Mr Mukesh, and Ms Pradnya Kamble. "Review on Enforcing Secure And Privacy Preserving Information Brokering In Distributed Information Sharing", International Conference on Advances in Engineering & Technology (ICAET), ISSN: 2278 - 0661, pp. 45-40, 2014.

[7] Srinivas, D. "Privacy-Preserving Public Auditing In Cloud Storage Security." International Journal of computer science and Information Technologies,ISSN: 0975-9646, vol. 2, no. 6, pp. 2691-2693, 2011.

[8] Li, Fenjun, Bo Luo, Peng Liu, Dongwon Lee, Prasenjit Mitra, Wang-Chien Lee, and Chao-Hsien Chu "In-broker access control: Towards efficient end-to-end performance of information brokerage systems", In IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, vol. 1, pp. 1-8-, 2006.

[9] Snoeren, Alex C., Kenneth Conley, and David K. Gifford. "Mesh-based content routing using XML." ACM SIGOPS Operating Systems Review 35, no. 5, pp. 160-173, 2001.

[10] Koudas, Nick, Michael Rabinovich, Divesh Srivastava, and Ting Yu. "Routing XML queries." In Proceedings of IEEE 20th International Conference on Data Engineering, pp. 844, 2004.

[11] Koloniari, Georgia, and Evaggelia Pitoura. "Content-based routing of path queries in peer-to-peer systems." In proceedings of Advances in Database Technology-EDBT-2004, pp. 29-47, Springer Berlin Heidelberg, 2004.

[12] Saha, Sanchari, and Madhusudana HA. "A Survey on Protecting Information Brokerage System against Intruders." International Journal of Engineering and Innovative Technology (IJEIT), Volume 3, Issue 5, pp. 105 – 108, November 2013.

[13] Kuehn, Andreas, Michael Kaschewsky, Andreas Kappeler, Andreas Spichiger, and Reinhard Riedl. "Interoperability and information brokers in public safety: an approach toward seamless emergency communications." Journal of theoretical and applied electronic commerce research, vol. 6, no. 1, pp. 43 – 60, 2011.

[14] Mardziel, Piotr, Adam Bender, Michael Hicks, Dave Levin, Mudhakar Srivatsa, and Jonathan Katz. "Secure sharing in distributed information management applications: problems and directions." In Proceedings of the Annual Conference of the International Technology Alliance (ACITA). 2010.