SPATIAL QUERY AUTHENTICATON

Analysis of Authentication Mechanisms for Outsourced Spatial Databases

[1]J Linita Lyle, [2]Ani Sunny

[1]Department of Computer Science and Engineering,
[1]Mar Athanasius College of Engineering, Kothamangalam, Kerala, India.

Abstract: With the growing popularity of location-based services and the excessive usage of smart phones and GPS enabled devices, the practice of outsourcing spatial data to third party service providers has grown rapidly over the past few years. Meanwhile, the fast arising trend of Cloud storage and Cloud computing services has provided a flexible and cost-effective platform for hosting data from businesses and individuals, further enabling many location-based applications. However, in this database outsourcing paradigm, the authentication of the query results at the client remains a challenging problem. This paper presents an overview and analysis of the Outsourced Spatial Database (OSDB) model in spatial datasets and the security techniques suggested in literature for ensuring the authenticity of the query results obtained by manipulating these outsourced databases.

IndexTerms **– Outsourced Spatial Database, Security, and Authentication.**

I. INTRODUCTION

The embedding of geo-positioning capabilities (e.g., GPS) in mobile devices has sparked off several types of location-based services. Owing to the advancement of such technologies, it is becoming more convenient and motivating for mobile users to share with each other their experience with all kinds of points of interests (POIs) such as cafes, restaurants, tourist spots, worship places etc. In the GPS navigation system, a POI (point of interest) may be defined as a geographically anchored pushpin that someone may find useful or interesting, which is usually annotated with textual information (e.g., descriptions and users' reviews). In recent years various spatial query models and techniques have emerged to enhance user experiences of location based services.

Consider a scenario where a Data Owner (DO) possesses a proprietary spatial dataset, such as a specialized map overlay or a set of points of interest (e.g., local businesses). The data owner may be any non-governmental organization (NGO), or individuals. The dataset can be exploited to a profitable gain. However, the cost of setting up the infrastructure, hiring qualified personnel and advertising an online service may be impractical. Moreover, the dataset will be more valuable if it is combined with the functionality of a general-purpose online map. These reasons provide strong motivation for outsourcing the dataset to a specialized Location-Based Service Provider (LBSP). Due to advancements in cloud storage technology, network technology and improvements in data transmission techniques, database outsourcing has attracted the attention of organizations across the globe.

While analyzing the security of the cloud storage, it can be assumed that the service provider, (LBSP) which maintains the outsourced spatial data, as untrusted. Clients must have the ability to verify the integrity of their own outsourced data. It can be assumed that the data owner is trusted, while the LBSP is untrusted. For example, the LBSP may replace some genuine results with others not among the top results or even not in the data owner's data set, and it may also modify some data records by adding more good reviews and features and deleting bad ones. In addition, an honest LBSP may be compromised by attackers to forge query results.

Query integrity is a line of research that ensures that a query result was indeed generated from the outsourced data (the authenticity requirement) and contains all the data satisfying the request (the correctness requirement). The key idea of the schemes used to ensure query integrity is that the data collector pre-computes and authenticates some auxiliary information (called authenticated hints) about its data set, which will be sold along with its data set to LBSPs. To faithfully answer a spatial query, a LBSP needs to return the correct spatial POI data records as well as proper authenticity and correctness proofs constructed from authenticated hints. The authenticity proof allows the query user to confirm that the query result only consists of authentic data records from the trusted data collector's data set, and the correctness proof enables the user to verify that the returned spatial POIs are the true ones satisfying the query.

The main objective of this paper is analyze the various authentication schemes that have been proposed in literature which includes the data structure used by the source and the directories to maintain data and the protocol for queries and updates. The performance of the authentication mechanism depends on the types of queries issued, the number of queries issued, storage overhead, computation overhead and frequency of updates.

II. BACKGROUND

This section describes the concept of Database as a service (DBaaS), its benefits, the architecture of database outsourcing model and the challenges associated with the same.

2.1 Database As A Service

According to technopedia, DBaaS can be defined as "*A Cloud computing service model that provides users with some form of access to a database without the need for setting up physical hardware, installing software or configuring for performance.*"

By comparison, the DBaaS model is a fee-based subscription service in which the database runs on the service provider's physical infrastructure. Different service levels are usually available. In a classic DBaaS arrangement, the provider maintains the physical infrastructure and database, leaving the data owner to manage the database's contents and operation. Alternatively, a data owner can set up a managed hosting arrangement, in which the provider handles database maintenance and management. This latter option may be especially attractive to small businesses that have database needs, but lack the adequate IT expertise.

2.2 Benefits of DBaaS

Compared with operating a traditional database on an on-site physical server and storage architecture, a cloud database offers the following distinct advantages:

- Elimination of physical infrastructure. In a cloud database environment, the cloud computing provider of servers, storage and other infrastructure is responsible for maintenance and availability. The organization that owns and operates the database (data owner) is only responsible for supporting and maintaining the database software and its contents. In a DBaaS environment, the service provider is responsible for maintaining and operating the database software, leaving the DBaaS users responsible only for their own data.
- Reduced cost: Through the elimination of a physical infrastructure owned and operated by an IT department, significant savings can be achieved from reduced capital expenditures, less staff, decreased electrical and a smaller amount of needed physical space.
- Instantaneous scalability. Should added database capacity be necessitated by seasonal business peaks or unexpected spikes in demand, a DBaaS provider can quickly offer additional fee-based capacity, throughput and access bandwidth via its own infrastructure. A database operating in a traditional, on-site infrastructure would likely need to wait weeks or months for the procurement and installation of additional server, storage or communications resources.
- Performance guarantees. Through a service level agreement (SLA), a DBaaS provider may be obligated to provide guarantees that typically quantify minimum uptime availability and transaction response times. An SLA specifies monetary and legal remedies if these performance thresholds are not met.
- Specialized expertise. In a corporate IT environment, except for the largest multinational enterprises, finding world-class database experts may be difficult, and keeping them on staff may be cost prohibitive. In a DBaaS environment, the provider may serve thousands of customers; thus, finding, affording and keeping world-class talent is less of a challenge.
- Latest technology. To remain competitive, DBaaS providers work hard to ensure that all database software, server operating systems and other aspects of the overall infrastructure are kept up to date with security and feature updates regularly issued by software vendors.
- Failover support. For a provider of database services to meet performance and availability guarantees, it is incumbent on that provider to ensure uninterrupted operation should the primary data center fail for any reason. Failover support typically encompasses the operation of multiple mirror image server and data storage facilities. Handled properly, failover to a backup data center should be imperceptible to any customer of that service.
- Declining pricing. With advances in technology and an intensely competitive marketplace among major service providers, pricing for a wide range of cloud-computing services undergoes continual recalibration. Declining prices are a major impetus for migrating on-site databases and other IT infrastructure to the cloud.

2.3 Architecture of Outsourced Database Model

Figure 1 depicts the architecture of Outsourced Database Model. It consists of 3 main entities as

- Data Owner
- Service Provider
- Clients

Generally, data owner and end users are considered as trustful entities while service provider is distrustful in context of manipulating data in an unauthorized manner. A data owner uploads his/her data at third party service provider's site using high speed communication link. A data owner can manipulate his/her data which includes functions such as insertion, modification/updation, and deletion. In case of multiple clients, access level permissions can be set for using the data. The service provider stores the data uploaded by the data owner. Data management hardware and software tools are deployed and maintained at the provider's site. The service provider ensures availability of data and provides efficient mechanisms for end users to access or query the underlying data.

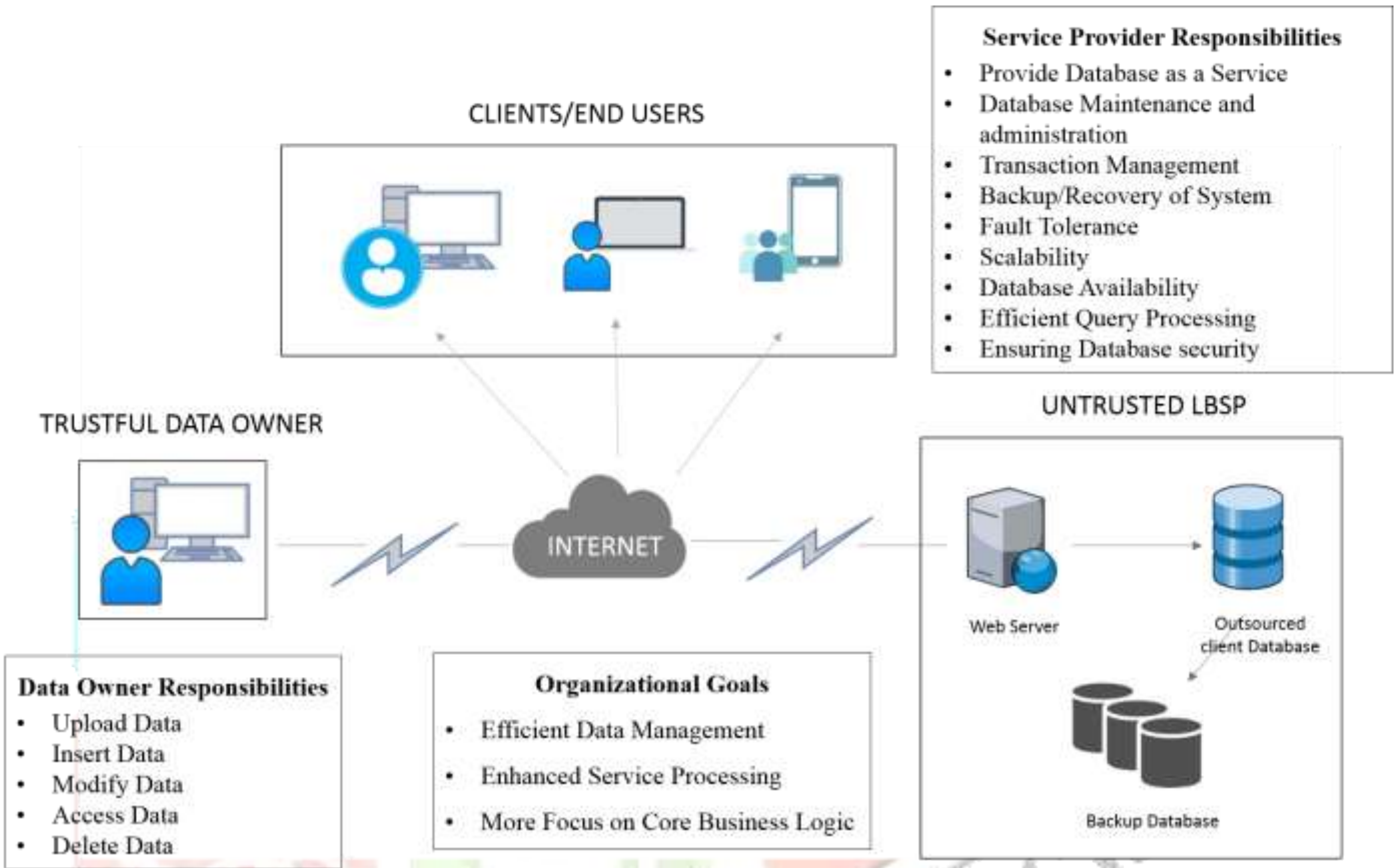

Figure 1: Architecture of Outsourced Database Model

There are 3 types of outsourced database models, categorized on the basis of number of data owners and end users/clients involved. The first model is unified client model in which the database is used by a single entity i.e. the functionalities of the client and the data owner are the same. The data owner does all the operations on the database and the communication link between data owner and client has very high bandwidth. The second type of outsourced database model is the multiple client model where multiple clients are given privileges of read only access. Here, the database can be accessed through mobile devices, laptops, PCs and the communication link has limited bandwidth. The third type of model is the multiple data owner model. In this model, there might be more than one data owner outsourcing his/her database to a third party service provider. Hence, for every group of data owner and client, separate access control and security policies need to be applied. This model is also referred to as multi-authority outsourced database model.

2.4 Challenges in Database Outsourcing

One of the primary challenges in outsourcing database to third party service providers is to ensure security of the outsourced database. Various security issues come into play which include:

- Confidentiality: Confidentiality is a security concern that assures that only authorized users or systems are given permission to access the data. *Data confidentiality* refers to forbidding unauthorized access when data is stored and also when the data is in transit state. Other aspects of confidentiality include *user privacy* and a*ccess privacy*. User *privacy* conceals the identity of the user who fetches or manipulates the data. Access *privacy* refers to keeping the data access patterns private.

- Integrity: Integrity assures that the data stored in a database or in transition state is not modified or manipulated by unauthorized personnel. *Completeness* and *correctness* are two important dimensions of *integrity*. *Completeness* refers to the idea of fetching all results that satisfy a query and *Correctness* refers to fetching the genuine results without any tampering.

- Availability: Availability ensures that data is available to trusted users and systems when they access the database in an authorized manner. It is degree or extent to which database is in operable state and is closely associated with reliability of a database.
- Authenticity: Authenticity implies that contracts, query transactions and communication are genuine and identities of the involved entities (users and system) are verified and known
- Freshness: Freshness of query results forms an important aspect of security when the database is frequently updated by the data owner. In such dynamically changing databases, *freshness* guarantees that the query results are obtained by executing the queries over the most updated database.

Other security issues include accountability, access control and risk management. While outsourcing data to third party service providers, it is vital that efficient security mechanisms are incorporated to ensure that the database is not exploited for malicious intends by the service providers.

III. OVERVIEW OF AUTHENTICATION MECHANISMS FOR OUTSOURCED SPATIAL DATABASES

The idea of outsourcing databases to a third-party service provider was first introduced by Hacigumus et al. [7]. Since then, numerous query authentication solutions have been proposed for auditing query results in outsourced relational databases. In the recent years, with the growing popularity of location based services, spatial databases imposes additional challenges to the authentication process as the spatial distribution of the objects need to be considered as it plays a vital role in the querying process. Authentication Data structure based approaches are used widely and has proven to be more efficient than previously suggested approaches.

Authenticated data structures are a model of computation where untrusted responders answer queries on a data structure on behalf of a trusted source and provide a proof of the validity of the answer to the user. Authentication data structure based approach are usually used for ensuring integrity and authenticity of outsourced databases. The most popular authentication data structure based approaches are MACs (Message Authentication Code), Digital Signature scheme, MHT (Merkle Hash Tree).

MAC function takes a secret key and a variable length data (message) as input and produces the MAC code. It assures integrity and authenticity in unified outsourced database model where data owner and client are the same entities and where bandwidth overhead and computation overhead are also less. MAC is not suitable in multiple clients and multiple data owner model. A deceitful client with secret MAC key can collude with the server and add fraudulent records to the database. Hence, non-repudiation cannot be assured with MACs.

Digital signatures are used to provide authenticity, integrity and non-repudiation and it is used when a large number of entities are involved in outsourcing [12]. A digital signature is a mathematical scheme for demonstrating the authenticity of digital messages or documents. A valid digital signature gives a recipient reason to believe that the message was created by a known sender (authentication), that the sender cannot deny having sent the message (non-repudiation), and that the message was not altered in transit (integrity). Large storage and bandwidth is required for implementing the digital signature schemes.

MHT[11] is used for assuring the integrity, secure verification, authentication of large datasets. It works as follows: Suppose data value is represented as $d_1, d_2 \ldots d_n$. A leaf node N_i ($i = 1, 2 \ldots n$) stores the hashed value of data such that N_1=h (d_1) and so on. The non-leaf node is represented by concatenation of hash values of its children. Figure 2 illustrates the idea of MHT.

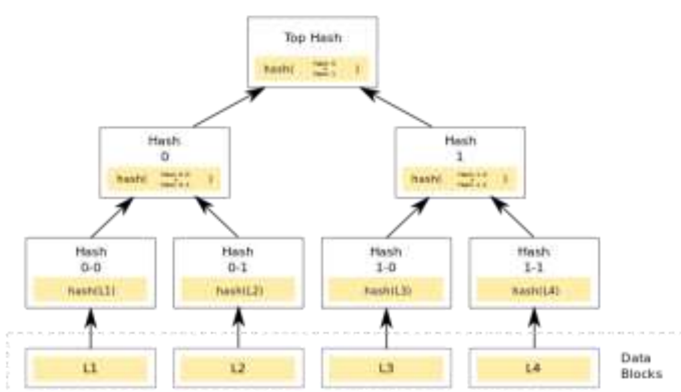

Figure 2: Merkle Hash Tree

The combination of MHT and B-tree is implemented by [10] known as Merkle B-tree to provide integrity (completeness, correctness). In this, the client computes all the hashes of the sub-tree using the verification object in a repeated manner until the root of the tree. Once the hash of the tree is computed, the correctness can be verified using the owner's public key and hashed value of root. As the client is forced to find the hash of whole sub tree, it ensures the criteria of completeness. A dithered B-tree which is combination of original B-tree and its corresponding dithered B-tree is implemented in [5] based on key-value pairs. It prevents a third party from learning whether the key is present in database or not, thus ensuring privacy. As the server stores two kinds of trees, it requires more space. Also communication cost for transferring the page-level data is also more.

Nested Merkle B_+ Tree which is an index structure is implemented by [15] to provide query assurance for dynamic outsourced XML databases. In this, all paths in XML document are listed out by tree traveler algorithm. The result contains leaf entries pointing to records, data records and not-in-result (co-path) entries. When the client fires query, server returns the root with signature and timestamp. Actual result and co-path are assembled in verification object (VO). The client recalculates hast of root and verifies it with the signature for assuring completeness and correctness.

In spatial databases, MHTs are suitable only for range queries and top K queries as the server returns only two full paths (Minimum value path and maximum value path) to the client.

MR Trees based on the idea of MHT and R* Trees is suggested in [15]. Leaf nodes are identical to those of the R* tree. Each entry R_i corresponds to a data object. Every leaf node of the tree stores a digest that is computed on the concatenation of the binary representation of all objects in the node. Internal nodes are assigned a digest that summarizes the child nodes' MBRs (minimum bounding rectangles) and digests. Digests are computed in bottom-up fashion and the single digest at the root is signed by the DO. The resulting *VO* contains all the objects in every leaf node visited, and the MBRs and digests of all the pruned nodes. With this information, the client can reconstruct the root digest and compare it against the one that was signed by the data owner. In addition, the client also examines the spatial relations between the query and each object/MBR included in the *VO*, in order to verify the correctness of the result.

An enhanced version of MR Trees, MIR Tree is suggested in [3, 6]. MIR is a combination of MHT and IR tree [18].The IR-tree is essentially an R-tree, each node of which is enriched with a reference to an inverted file for the objects contained in the subtree rooted at the node. In the IR-tree, a leaf node contains a number of entries of the form (O, O.r), where O refers to an object in database and O.r is the bounding rectangle of object O. A leaf node also contains a pointer to an inverted file for the text documents of the objects being indexed. The inverted file is stored separately from the R-tree, for two reasons: First, it is more efficient to store each inverted file contiguously, rather than as a sequence of blocks or pages that are scattered across a disk [18]. Second, the inverted file can be distributed across several machines, while this is not easily possible for the R-tree [14].

Various drawbacks have been identified in the structure of MR trees and in the verification process. First, the authentication information (hash digests) embedded in the MR-tree reduces the node fan out which results in increased I/O accesses during query processing. Second, in the presence of updates, all the digests on the path from an affected leaf node to the root have to be recomputed. When updates are frequent, query performance is degraded [8]. Finally, the overhead of the *VO* can be significant, especially for queries that return only a few objects. This is due to the fact that SP has to return all objects that lie inside the leaf nodes that are visited during query processing. An extension of the MR-tree, called MR*-tree [16], mitigates this last drawback by ordering the entries of each node and constructing hierarchical relationships of the digests. Nevertheless, it does not eliminate the VO overhead entirely, and it increases the verification cost at the client.

An approach based on voronoi diagrams was suggested in [9]. Voronoi diagram is a partitioning of a plane into regions based on distance to points in a specific subset of the plane. The approached referred to as VN-Auth separates the authentication information from the spatial index, thus allowing efficient query processing at the service provider. Additionally, the verification information depends only on the object and its voronoi neighbors, and hence database updates can be disseminated quickly to their local regions and be performed independently of all other updates in the database. VN-Auth is proven to handle not only kNN and range queries, but also more advanced query types, such as reverse kNNs, k aggregate NNs and spatial skylines.VN-Auth produces compact verification objects, which enables fast query verification on mobile devices with limited capabilities.

IV. ANALYSIS OF AUTHENTICATION MECHANISMS FOR OUTSOURCED SPATIAL DATABASES

The performance overhead due to authentication-related computations in an authenticated data structure based on a hashing scheme, also known as the authentication overhead consists of the following cost parameters.

- Time Overhead: Includes any cost that adds to the time performance of the data structure. The *rehashing overhead* is the time spend by the source or the directory to rehash over the data in order to recompute the digest of the data structure after an update in the data set. The *query answering overhead* is the extra time consumption for the directory to create the answer authentication information. The *verification time* is the time spent by the user in order to verify the answer given by the directory.

- Communication Overhead: Corresponds to the communication cost introduced by the model, i.e., the size of the update authentication information for the source-to-directory communication and the size of the answer authentication information for the directory-to user communication. We are particularly interested in the size of the proof that is given to the user which is part of the answer authentication information.

- Storage Overhead: The total number of hash values used by the authentication scheme

Even with the most efficient implementations, the time for computing a hash function is an order of magnitude larger than the time for a comparison of basic number types (e.g., integers or floating-point numbers). Thus, the rehashing overhead dominates the update time and the practical performance of an authenticated data structure is characterized by the authentication overhead. For efficient authenticated data structures the above cost parameters need to be minimized.

V. CONCLUSION

Database outsourcing is a popular data management technology in the recent era which is accepted widely due to its inherent profitable features. In this paper, the concept of DBaaS, its architecture, its benefits and the challenges associated are discussed. Security mechanisms based on authentication data structures for ensuring authenticity of query results in spatial databases are reviewed. With the advancement in geo positioning technologies and Geographic Information Systems (GIS), the area of spatial database has gained widespread attention in the current scenario. Hence, ensuring efficient and secure systems and methods to manipulate and process the underlying data remains an open challenge.

REFERENCES

[1]. B. Hore, S. Mehrotra, and G. Tsudik, “A Privacy-Preserving Index for Range Queries,” Proc. 30th Int’l Conf. Very Large Data Bases (VLDB’04), pp. 720-731, Aug. 2004.

[2]. Chung-Min Chen, Andrzej Cichocki, Allen McIntosh, Euthimios Panagos, “Privacy-Protecting Index for Outsourced Databases”, In Proc. of ICDE Workshops 2013, pp. 83-87.

[3]. D. Wu, G. Cong, and C. S. Jensen. A framework for efficient spatial web object retrieval. VLDB J., 21(6):797–822, 2012.

[4]. Dongxi Liu, Shenlu Wang, “Programmable Order-Preserving Secure Index for Encrypted Database Query” In Proceedings of 2012 IEEE Fifth International Conference on Cloud Computing, pp. 502-509, 2012.

[5]. E. Mykletun, M. Narasimha, and G. Tsudik, “Authentication and *integrity* in outsourced databases,” In Proc. of ACM Trans. On Storage, vol. 2, 2006, pp. 107-138.

[**6].** G. Cong, C. S. Jensen, and D. Wu. Efficient retrieval of the top-k most relevant spatial web objects. In VLDB, pages 337–348, 2009.

[7]. H. Hacigumus, S. Mehrotra, and B. Iyer, “Providing Database as a Service,” Proc. IEEE 18th Int’l Conf. Data Eng. (ICDE), Feb. 2002.

[8]. H. Pang, J. Zhang, and K. Mouratidis. Scalable Verification for Outsourced Dynamic Databases. PVLDB, 2(1):802–813, 2009.

[9]. Hu, L., Ku, W., Bakiras, S., Shahabi, C.: Spatial query integrity with voronoi neighbors. IEEE Trans. Knowl. Data Eng. 25(4), 863–876 (2013)

[10]. Li Feifei, Marios H, George K, “Dynamic Authenticated Index Structures for Outsourced Database” In Proc. of ACM SIGMOD’06. Chicago, Illinois, 2006, pp. 121-132

[11]. R. C. Merkle. A certified digital signature. In *CRYPTO*, pages 218–238, 1989.

[12]. R. J. Morteza Noferesti, Mohammad Ali Hadavi, "A Signature-based Approach of Correctness Assurance in Data Outsourcing Scenarios," ICISS 2011, India, pp. 374-378.

[13]. R. Zhang, J. Sun, Y. Zhang and C. Zhang, "Secure Spatial Top-k Query Processing via Untrusted Location-Based Service Providers," in IEEE Transactions on Dependable and Secure Computing, vol. 12, no. 1, pp. 111-124, Jan.-Feb. 2015.

[14]. Schnitzer, B., Leutenegger, S.: Master-client R-trees: a new parallel R-tree architecture. In: SSDBM, pp. 68–77 (1999)

[15]. Viet Hung Nguyen, Tran Khanh Dang, Nguyen Thanh Son , Josef Küng, “Query Assurance Verification for Dynamic Outsourced XML Databases”, Second International Conference on Availability, Reliability and Security (ARES'07), 2007, pp. 689 - 696.

[16]. Y. Yang, S. Papadopoulos, D. Papadias, and G. Kollios. Authenticated Indexing for Outsourced Spatial Databases. VLDB J., 18(3):631–648, 2009.

[17]. Y. Yang, S. Papadopoulos, D. Papadias, and G. Kollios. Spatial Outsourcing for Location-based Services. In ICDE, pages 1082–1091, 2008.

[18]. Zobel, J., Moffat, A.: Inverted files for text search engines. ACM Comput. Surv. 38(2), 1–56 (2006)