Secure Multidata Owner K-NN Query Processing on Road Networks using Voronoi Network

T. Akila

Assistant Professor/ Department of CSE, Mahendra College of Engineering, India

Abstract: Mobile devices with geo-positioning capabilities (e.g., GPS) enable users to access information that is relevant to their present location. Users are interested in querying about points of interest (POI) in their physical proximity, such as restaurants, cafes, ongoing events, etc. With the ever-increasing use of smart phones and tablet devices, location-based services (LBSs) have experienced explosive growth in the past few years. To scale up services, there has been a rising trend of outsourcing data management to Cloud service providers, which provides query services to clients on behalf of data owners. However, in this data-outsourcing model, the service provider can be untrustworthy or compromised, thereby returning incorrect or incomplete query results to clients, intentionally or not. In this paper, we propose a family of techniques that allow processing of NN queries in an untrusted outsourced environment, while at the same time protecting both the POI and querying users’ positions. Our techniques rely on mutable order preserving encoding (mOPE), the only secure order-preserving encryption method known to date. We also provide performance optimizations to decrease the computational cost inherent to processing on encrypted data, and we consider the case of incrementally updating datasets.

We present an extensive performance evaluation of our techniques to illustrate their viability in practice. Unlike previous work that verifies k-nearest-neighbor results in the Euclidean space, our approach needs to verify both the distances and the shortest paths from the query point to its kNN results on the road network. We evaluate our approach on real-world road networks with both real and synthetic points of interest datasets. Our experiments on communicate with the service provider through wireless connections. The experiment results show that our approach leads to compact verification objects (VO) and the verification algorithm on mobile devices is efficient, especially for queries with low selectivity.

Keywords: K-NN, Web Search Engine Creation, Voronoi Network

Reference to this paper should be made as follows: T. Akila (2015) ‘Secure Multidata Owner K-NN Query Processing on Road Networks using Voronoi Network’, International Journal of Scientific and Computational Intelligence, Volume 2 Issue 2 2015

1 Introduction

1.1 General Background

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

1.2 The Scope Of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

- Automated prediction of trends and behaviors.
- Automated discovery of previously unknown patterns.

1.3 About The Area

Location-based services (LBS) are a general class of computer program-level services that use location data to control features. As such LBS is an information service and has a number of uses in social networking today as an
entertainment service, which is accessible with mobile devices through the mobile network and which uses information on the geographical position of the mobile device. This has become more and more important with the expansion of the smart phone and tablet markets as well.

LBS are used in a variety of contexts, such as health, indoor object search, entertainment work, personal life, etc. LBS include services to identify a location of a person or object, such as discovering the nearest banking cash machine (a.k.a. ATM) or the whereabouts of a friend or employee. LBS include parcel tracking and vehicle tracking services. LBS can include mobile commerce when taking the form of coupons or advertising directed at customers based on their current location. They include personalized weather services and even location-based games. They are an example of telecommunication convergence.

This concept of location based systems is not compliant with the standardized concept of real-time locating systems (RTLS) and related local services, as noted in ISO/IEC 19762- and ISO/IEC 24730-1. While networked computing devices generally do very well to inform consumers of days old data, the computing devices themselves can also be tracked, even in real-time.

1.4 Architecture Diagram

Query authentication solutions have been proposed for outsourced relational databases. Digital signature aggregation to ensure data integrity and authenticity for outsourced databases. The techniques cannot assure completeness of the result set aggregated signature in order to sign each record with the information from neighboring records by assuming that all the records are sorted in a certain order.

1.5 Web Search Engine Creation

User interface design or user interface engineering is the design of computers, appliances, machines, mobile communication devices, software applications, and websites with the focus on the user's experience and interaction. The goal of user interface design is to make the user's interaction as simple and efficient as possible, in terms of accomplishing user goals—what is often called user-centered design. To run our remote control system we develop a GUI application in J2EE. User can easily execute the project with the help of GUI.

A social networking service is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user (often a profile), his social links, and a variety of additional services.

1.6 Road Network And Skyline Query

A road network can be considered as a graph with positive weights. The nodes represent road junctions and each edge of the graph is associated with a road segment between two junctions. The weight of an edge may correspond to the length of the associated road segment, the time needed to traverse the segment or the cost of traversing the segment. Using directed edges it is also possible to model one-way streets. Such graphs are special in the sense that some edges are more important than others for long distance travel. All of these algorithms work in two phases. In the first phase, the graph is preprocessed without knowing the source or target node. The second phase is the query phase. In this phase, source and target node are known.

Users carrying location-aware mobile devices are able to query Location-based services (LBSs) for surrounding points of interest (POIs) anywhere and at any time. The location-based skyline queries are to find the nearby spatial object with good services where the distance to the querying user is a spatial attribute and the goodness of the objects. The query returns a set of result that is closer to the client current position.

1.7 Data Provider

To scale up services there has been a rising trend of outsourcing data management to service providers, which provide query services to clients on behalf of data owners. Outsourcing spatial databases to provides an economical and flexible way for data owners to deliver spatial data to users of location-based services. Data as a Service, or DaaS, is a cousin of software as a service. Like all members of the as a Service (aaS) family, DaaS is based on the concept that the product, data in this case, can be provided on demand to the user regardless of geographic or organizational separation of provider and consumer. Additionally, the emergence of service-oriented architecture (SOA) has rendered the actual platform on which the data resides also irrelevant.[3] This development has enabled the recent emergence of the relatively new concept of DaaS.

II. Literature Review

[1]Qian Chen, Haibo Hu, Jianliang Xu —AUTHENTICATING TOPK QUERIES IN LOCATIONBASED SERVICES WITH CONFIDENTIALITY” This paper, we address the more challenging authentication problem on top-k queries, where
private values appear on both sides of a comparison. To start with, we propose two novel cryptographic building blocks, followed by a comprehensive design of authentication schemes for top-k queries based on R-tree and Power Diagram indexes. Optimizations, security analysis, and experimental results consistently show the effectiveness and robustness of the proposed schemes under various system settings and query workloads. Main contributions made in this paper are as follows:

- To the best of our knowledge, this is the first work that addresses privacy-preserving top-k query authentication. Our query definition encompasses most existing location-based top-k and kNN queries.
- We introduce a formal security model and design two cryptographic building blocks (namely PPB and PLB) that can prove to the client the relation of rank values of two objects w.r.t. a query point without disclosing the locations and non-spatial scores.
- We develop a complete set of authentication schemes for both the R-tree and Power Diagram based indexes.
- We propose strategies for both the data owner and the SP to optimize the storage cost of the PLB method.
- We conduct extensive experiments and security analysis to evaluate the performance and robustness of the proposed authentication schemes.

This paper, we studied the problem of privacy-preserving authentication for top-k queries in LBSs. By designing cryptographic building blocks of private ranking-value comparisons, we have presented two authentication schemes based on multi-dimensional R-tree and Power Diagram indexes. The performance and security of our proposed schemes have been verified and analyzed under various system settings.

[2] Haibo Hu, Jianliang Xu, Dik Lun LeelA GENERIC FRAMEWORK FOR MONITORING CONTINUOUS SPATIAL QUERIES OVER MOVING OBJECTS.

This paper proposes a generic framework for monitoring continuous spatial queries over moving objects. The framework distinguishes itself from existing work by being the first to address the location update issue and to provide a common interface for monitoring mixed types of queries. Based on the notion of safe region, the client location update strategy is developed based on the queries being monitored. Thus, it significantly reduces the wireless communication and query reevaluation costs required to maintain the up-to-date query results. We propose algorithms for query evaluation/reevaluation and for safe region computation in this framework. Enhancements are also proposed to take advantage of two practical mobility assumptions: maximum speed and steady movement. The experimental results show that our framework substantially outperforms the traditional periodic monitoring scheme in terms of monitoring accuracy and CPU time while achieving a close-to-optimal wireless communication cost. The framework also can scale up to a large monitoring system and is robust under various object mobility patterns. The proposed framework exhibits the following advantages:

- To our knowledge, this is the first monitoring scheme that addresses the location update issue. By safe regions, the moving clients are query aware and report location updates to the server only when they are very likely to alter the results. Hence, both the wireless communication cost and the query evaluation cost are greatly reduced. Yet the performance gain is not achieved by sacrificing the simplicity at the client side: the new location update logic is as straightforward as before.
- The framework is generic in the sense that it is not designed for a specific query type. Rather, the framework provides a common interface for monitoring various types of spatial queries such as range queries and kNN queries. Moreover, the framework does not presume any mobility pattern on moving objects.
- Since any object movement that might change the query results is captured by the location update, the framework offers us accurate monitoring results at any time, as opposed to deviated results as can be observed in the previous periodic monitoring approach.
- In contrast to periodical reevaluation of queries in most existing studies, query reevaluation in this framework is triggered by location updates only. As these location updates are asynchronous, the workload of the server is evenly distributed over time.

This paper proposes a generic framework for monitoring continuous spatial queries over moving objects. The framework distinguishes itself from existing work by being the first to address the location update issue and to provide a common interface for monitoring mixed types of queries. Based on the notion of safe region, the location updates are query aware and thus the wireless communication and query reevaluation costs are significantly reduced. We provide detailed algorithms for query evaluation/reevaluation and safe region computation in this framework.

[3] Haibo Hu, Jianliang Xu, Qian Chen, Ziwei Yang AUTHENTICATING LOCATION-BASED SERVICES WITHOUT COMPROMISING LOCATION PRIVACY. This paper we address this challenging problem by proposing a comprehensive solution that preserves unconditional location privacy when authenticating range queries. Three authentication schemes for R-tree and grid-file index, together with two optimization techniques, are developed. Cost models, security analysis, and experimental results consistently show the effectiveness, reliability and robustness of the proposed schemes under various system settings and query workloads. Contributions in this paper are as follows:
To the best of our knowledge, this is the first work that addresses privacy-preserving query authentication for location-based services. The problem is critical for both mobile value-added service market and database research community.

We develop three authentication schemes for R-tree and grid-file index, which are good for large queries, small queries, and queries on static datasets, respectively. Analytical models of computation and bandwidth costs are developed to justify these schemes. Security analysis shows they are secure by not disclosing any individual location information.

We propose two optimization techniques that are orthogonal to the underlying authentication schemes.

We conduct extensive experiments to evaluate the performance of the proposed schemes, with and without the two optimization techniques in effect. The results coincide with our analysis and further justify the efficiency of our approaches.

This paper, we study the problem of privacy-preserving query authentication for location-based services. With the single-dimension building block of authentication on B+-tree, we propose three authentication schemes for multidimensional indexes, including R-tree and grid-file index. We further enhance the efficiency of the schemes by two optimization techniques, namely, the linear ordering and embedding. The performance of all schemes is evaluated both analytically and empirically, which consistently shows their effectiveness and robustness under various system settings. The security perspective of these schemes is also studied. As for future work, we plan to investigate on more query types for location-based services. In particular, we are interested in privacy-preserving authentication on k-nearest-neighbor queries. As neither the user locations nor their distances to the query point can be disclosed to the client, the authentication is even more challenging than range queries.

This paper, we focus on the Outsourced Spatial Database (OSDB) model and propose an efficient scheme, called VN-Auth, which allows a client to verify the correctness and completeness of the result set. Our approach is based on neighborhood information derived from the Voronoi diagram of the underlying spatial dataset and can handle fundamental spatial query types, such as k nearest neighbor and range queries, as well as more advanced query types like reverse k nearest neighbor, aggregate nearest neighbor, and spatial skyline. We evaluated VN-Auth based on real-world datasets using mobile devices (Google Droid smart phones with Android OS) as query clients. Compared to the current state-of-the-art approaches (i.e., methods based on Merkle hash trees), our experiments show that VN-Auth produces significantly smaller verification objects and is more computationally efficient, especially for queries with low selectivity.

In this approach, the DO generates one signature for every object in the database, which is computed on the hash digest of the concatenation of the binary representation of the object and its Voronoi neighbors. In this way, the client can verify the authenticity of each individual object and its neighborhood. Note that, in this work, we utilize RSA signatures [6] that are typically 128 bytes in size. Alternatively, signatures based on Elliptic Curve Cryptography (ECC) can be significantly shorter, thus reducing the overall communication and storage cost. However, ECC algorithms are computationally intensive and would perform poorly on mobile devices with limited computational capabilities.

The drawback of having one signature per database object is that it may increase considerably the communication cost between the SP and the client. Specifically, the SP has to transmit one 128-byte signature for every object in the result set, so the overhead can be significant for queries with high selectivity (especially for mobile clients). To avoid this cost, we employ a technique called signature aggregation. In particular, given k digests and their corresponding signatures (generated by the same signer), the SP can replace them with a single Condensed-RSA signature. Condensed-RSA has the same size as the original signatures (128 bytes), and it is computed as the modular multiplication of the k signatures. Aggregate signatures are provably secure and can be computed by any party that possesses the individual signatures. This paper, we introduced the VN-Auth query integrity assurance framework for outsourced spatial databases. Our approach separates the authentication information from the spatial index, thus allowing efficient query processing at the service provider. Additionally, since the verification information depends only on the object and its Voronoi neighbors, database updates can be disseminated quickly to their local regions and be performed independently of all other updates in the database. VN-Auth handles not only kNN and range queries, but also more advanced query types, such as reverse kNNs, k aggregate NNs and spatial skylines. More importantly, VN-Auth produces compact verification objects, which enables fast query verification on mobile devices with limited capabilities. Finally, we showed that our approach facilitates progressive result verification, which allows a user to retrieve objects in an incremental fashion until the results are deemed satisfactory.

This paper, we study a novel form of continuous nearest neighbor queries in the presence of obstacles, namely continuous obstructed nearest neighbor (CONN) search. It considers the impact of obstacles on the distance between objects, which is ignored by most of spatial queries. Given a data set P, an obstacle set O, and a query line segment q in a two-dimensional space, a CONN query retrieves the nearest neighbor of each point on q.
according to the obstructed distance, i.e., the shortest path between them without crossing any obstacle. We formulate CONN search, analyze its unique properties, and develop algorithms for exact CONN query processing, assuming that both P and O are indexed by conventional data-partitioning indices (e.g., R-trees). Our methods tackle the CONN retrieval by performing a single query for the entire query segment, and only process the data points and obstacles relevant to the final result, via a novel concept of control points and an efficient quadratic based split point computation algorithm. This paper has made five-fold main contributions, summarized as follows:

1. We formalize CONN search, a new addition to the family of spatial queries in an obstructed space. To the best of our knowledge, this paper is the first attempt on this problem.

2. We introduce the concept of control point that significantly simplifies the computation and comparison of the obstructed distance between two objects.

3. We propose a quadratic-based method to form split points, by solving quadratic inequalities.

4. We develop an efficient algorithm for processing CONN search which can be extended to handle COkNN retrieval.

5. We conduct extensive experiments using both real and synthetic datasets to demonstrate the efficiency and effectiveness of the proposed algorithms.

This paper, for the first time, we identify and solve a novel type of CNN queries, namely continuous obstructed nearest neighbor (CONN) search, which considers the impact of obstacles on the distances between objects. CONN queries are not only interesting from a research point of view, but also useful in many applications such as location-based services, geographic information systems, and spatial data analysis. We carry out a systematic study of CONN retrieval. First, we provide a formal definition of the problem. Then, we present several effective pruning strategies and develop efficient algorithms for CONN query processing. Next, we extend our techniques to handle COkNN search, a natural generalization of CONN query. Finally, we conduct extensive experiments to verify the efficiency and effectiveness of our proposed algorithms using both real and synthetic datasets.

[6]Donald Kossmann Frank Ramsak Steffen Rostl
SHOOTING STARS IN THE SKY: AN ONLINE ALGORITHM FOR SKYLINE QUERIES”

Skyline queries ask for a set of interesting points from a potentially large set of data points. If we are traveling, for instance, a restaurant might be interesting if there is no other restaurant which is nearer, cheaper, and has better food. Skyline queries retrieve all such interesting restaurants so that the user can choose the most promising one. In this paper, we present a new online algorithm that computes the Skyline. Unlike most existing algorithms that compute the Skyline in a batch, this algorithm returns the first results immediately, produces more and more results continuously, and allows the user to give preferences during the running time of the algorithm so that the user can control what kind of results are produced next (e.g., rather cheap or rather near restaurants). Following properties from an online algorithm:

1. The first results should be returned almost instantaneously. It should be possible to give guarantees that constrain the running time to produce, say, the first 100 results.

2. The algorithm should produce more and more results the longer the algorithm runs. Eventually (if given enough time), the algorithm should produce the full Skyline.

3. The algorithm should only return points which are part of the Skyline. In other words, the algorithm should not return good points (e.g., good restaurants) at the beginning and then replace these good restaurants with better restaurants.

4. The algorithm should be fair. In other words, the algorithm should not favor points that are particularly good in one dimension; instead it should continuously compute Skyline points from the whole range.

5. The user should have control over the process. In other words, it should be possible to make preferences while the algorithm is running. Using a graphical user interface, the user should be able to click on the screen and the algorithm will return next points of the Skyline which are near the point that the user has clicked on.

6. The algorithm should be universal with respect to the type of Skyline queries and data sets. It should also be based on standard technology (i.e., indexes), and it should be easy to integrate the algorithm into an existing database system. For a given data set (e.g., hotels or restaurants) one index should be enough to consider all dimensions that a user might find interesting.

In our performance experiments, we could not identify a clear winner. All algorithms have their particular virtues. Our new algorithm, the NN algorithm, gives a big picture of the Skyline very quickly in all situations. However, it is not always the best choice if the full Skyline needs to be computed. In addition to the raw performance, the algorithms differ significantly in other criteria. The NN algorithm is the only algorithm that gives the user control over the process and allows the user to give preferences. The B-tree algorithm gives "extreme" points preference (i.e., points good in one dimension) and returns points which are good in many dimensions very late. The Bitmap algorithm scans the database and uses Bitmaps in order to detect whether a point is part of the Skyline.


This paper work, _rest we done a variety of essential and practical cost metrics associated with ODB systems. Then, we analytically evaluate a number of different approaches, in search for a solution that best leverages all metrics. Most importantly, we look at solutions that can handle dynamic scenarios, where owners periodically update the data residing at the servers. Finally, we discuss query freshness,
a new dimension in data authentication that has not been explored before. A comprehensive experimental evaluation of the proposed and existing approaches is used to validate the analytical models and verify our claims. Our endings exhibit that the proposed solutions improve performance substantially over existing approaches, both for static and dynamic environments. Contributions of the paper

1. Conduct a methodical analysis of existing approaches over all six metrics.
2. Propose a novel authenticated structure that best leverages all metrics.
3. Formulate detailed cost models for all techniques that take into account not only the usual structural maintenance overheads, but the cost of cryptographic operation.
4. Discuss the extensions of the proposed techniques for dynamic environments (where data is frequently updated).
5. Consider possible solutions for guaranteeing query freshness.
6. Implement a fully working prototype and perform a comprehensive experimental evaluation and comparison of all alternatives.

We presented a comprehensive evaluation of authenticated index structures based on a variety of cost metrics and taking into account the cost of cryptographic operations, as well as the cost of index maintenance. We proposed a novel structure that leverages good performance based on all metrics. We extended the work to dynamic environments, which has not been explored in the past. We also formulated the problem of query freshness, a direct outcome of the dynamic case. Finally, we presented a comprehensive experimental evaluation to verify our claims. For future work, we plan to extend our ideas for multidimensional structures, and explore more involved types of queries.

III Implementation

Query authentication solutions have been proposed for outsourced relational databases. Digital signature aggregation to ensure data integrity and authenticity for outsourced databases. The techniques cannot assure completeness of the result set aggregated signature in order to sign each record with the information from neighboring records by assuming that all the records are sorted in a certain order. Their mechanism helps users verify that query results are both complete and authentic. In addition, the challenge token scheme is for a server running outsourced databases to provide a proof of the actual query execution, which is then checked at the client side for integrity verification. Compared to, the scheme also supports more query types without assuming that all the records are sorted. Nonetheless, none of the techniques are specifically designed for spatial databases. Since the query distance is defined by network distance in a road network, the skyline scope defined in existing system no longer works, which calls for new authentication methods. Moreover, we are also interested in studying the authentication problem for dynamic objects, where how to guarantee the freshness of query results is a very challenging issue. In proposed system works with multiple data owners. For example, the POIs and the road networks can come from two different data owners. Hence, how to handle the query verification problem in the presence of multiple data owners is also an interesting direction to explore.

Fig 1.1 data level

User interface design or user interface engineering is the design of computers, appliances, machines, mobile communication devices, software applications, and websites with the focus on the user's experience and interaction. The goal of user interface design is to make the user's interaction as simple and efficient as possible, in terms of accomplishing user goals—what is often called user-centered design. To run our remote control system we develop a GUI application in J2EE. User can easily execute the project with the help of GUI. A social networking service is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user (often a profile), his social links, and a variety of additional services

4.3.2.2 Features

- Query caching
- Sub-SELECTs (i.e. nested SELECTs)
- Replication support (i.e. Master-Master Replication & Master-Slave Replication) with one master per slave, many slaves per master, no automatic support for multiple masters per slave.
- Full-text indexing (Index_(database)) and searching using MyISAM engine
- A broad subset of ANSI SQL 99, as well as extensions
- Cross-platform support
- Stored procedures
- Triggers
- Cursors
- Updatable Views
• True Varchar support

IV. Test Results

V. Conclusion

In this paper, we have studied the problem of authenticating location-based skyline queries in arbitrary subspaces (LASQs). We have proposed a basic MSR-tree authentication method by extending our previous work on skyline query authentication. To enable authentication for large scale datasets and subspaces, we have further proposed a Partial-S4-tree method, in which most of the redundant objects can be easily identified and filtered out from the VO. For authenticating continuous LASQs, we have proposed a perfecting-based solution to avoid frequent query issuances and VO transmissions. Extensive experimental results demonstrate the efficiency of our proposed methods and algorithms under various system settings. In particular, our proposed Partial-S4-tree method outperforms the basic authentication method by up to 69% in terms of the overall query latency and up to 74% in terms of the VO size.

VI. Future Enhancement

In the future work that can explore a few different directions. First of all, how to handle more types of network spatial queries using the general framework and data structure. Secondly, we only considered one data owner party. However, in practice, there might be multiple data owners. For example, the POIs and road networks can come from two different data owners. Hence, how to handle the query verification problem in the presence of multiple data owners is also an interesting direction to explore.

References


