ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 01, January 2018

Moving top-k Spatial Keyword Queries

Ms.Anitha.R[1], Ms.S.Nivedha[2]
Assistant Professor, Department of CSE, Sri Krishna College of Technology, Coimbatore, India[1,2]

Abstract: A Moving top-k Spatial Keyword (MkSK) query takes into account a continuously moving query location and enables a mobile client to be continuously aware of the top-k spatial web objects that best matches a query with respect to the location and text relevance. The increasing use of the web and the proliferation of geo-positioning render it of interest to consider a spatial keyword search is outsourced to a separate service provider capable at handling the voluminous spatial web objects available from various sources. A key challenge is the service provider may return inaccurate or incorrect query results e.g., due to cost considerations or invasion of hackers. Therefore, it is attractive to authenticate the query results at the client side. Present authentication techniques are either inefficient or inapplicable for the kind of query. The MIR-tree (Merkle-IR) and MIR*-tree enables the authentication of MkSK queries at low computation and communication costs in the new authentication data structure system. A verification object for authenticating MkSK queries and an algorithm for constructing verification objects is designed.

Keywords: Moving top-k Spatial Keyword, MIR-tree, Safe zone

I. INTRODUCTION

Data Mining is the computational process of discovering patterns in huge data sets involving methods at the intersection of statistics and database systems.

A spatial keyword query integrates location and text search, taking a location and a set of keywords as arguments. Spatial web objects can be points of interest (e.g., restaurants) with a web presence and have locations as well as textual descriptions.

A MkSK query takes into account a continuously moving query location and enables a mobile client to be continuously aware of the top-k spatial web objects that best matches a query with respect to location and text relevance. For example, a mobile client may activate a "cafe" query in order to be alerted about nearby opportunities for a cup of coffee. With the MkSK query, a client always has an up-to-date result as the client moves. The client can ignore a result and move until an appealing result appears.

A straight forward solution to the MkSK query is to periodically invoke an existing snapshot of spatial keyword query processing technique. However, the approach has the problem that even if snapshot queries are processed very frequently, which is expensive and waste because consecutive results are likely to be very similar and there is no guarantee that the user always has the right, up-to-date result. Another possible solution is to extend a buffering technique for spatial k-Nearest Neighbour (kNN) query processing to top-k spatial keyword querying. Given a kNN query, the technique retrieves k nearest neighbours and uses them to derive a buffer region with the property that stands as long as the user moves inside the region. The kNN result can be derived from the k objects. However, it is not known to extend the technique to MkSK queries, where both text relevance and spatial distance are considered. A safe zone based approach for the processing of MkSK queries returns a safe zone to a client together with the query result. The safe zone is a region containing the users location and in which the top-k result remains unchanged. The safe zone based approach significantly reduces the communication between clients and the Service Provider (SP) as well as computation costs, since the client needs to request new result only when leaving the safe zone [4]. Consider the example MkSK query q, where the query keywords are "vanilla coffee." The text relevance of objects p1 and p2 to q is 2 and 1, respectively, when using the number of matched keywords for defining relevance. The SP returns object p2 as the top-1 result and the gray circle as the safe zone of p2, meaning that as long as the client remains inside the gray circle, object p2 is the top-1 result.

The curved path shows the client's movement. When the client crosses the boundary of the gray circle (at q0), it sends its updated location to the SP that computes and returns a new top-1 result p1 and the white region as the safe zone. Completeness ensures that no valid result object is missed in a query result. Authentication techniques have been

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 01, January 2018

developed for a variety of queries, including relational queries, sliding window queries, spatial queries, text similarity queries, shortest path queries, moving kNN queries, moving range queries, and sub graph search. The contributions include a design data structure for the authentication of MkSK queries and the MIR-tree that enables low computation and communication costs [5]. The MIR-tree is applicable to many variant of the spatial keyword query. A Verification Object (VO) for authenticating the top-k results and safe zones of MkSK queries is designed. Algorithms for constructing VO and verifying the top-k results and safe zones using VO are developed. An enhanced data structure (MIR*-tree) is proposed to further reduce the communication cost. The idea of the MIR*-tree is applicable for any tree-structured ADS, e.g., the MIR-tree, where each node contains multiple entries. A thorough experimental study on real data to evaluate the performance is conducted.

II. METHODOLOGY

A region (range), also called a verification set covers the query result. The data objects that fall into the verification set are sent to the client. In addition, summary objects that are used to derive bounds on the ranking scores of the objects outside the verification set are sent to the client. In other words, the VO contains the objects inside the verification set and a number of summary objects [6]. Using the VO, the client is able to authenticate the result. Authenticating an MkSK query involves verifying both the spatial distances and text relevancies of the result objects. The current state-of-the-art solution for authenticating spatial queries, the MR-tree, and the method for authenticating the query results of text search engines can be used to authenticate the spatial and textual parts of the result objects separately.

However, the approach is inefficient because it generates a large VO based on a large verification set, which causes high communication cost and long authentication time. The MIR-tree is used for efficiently verifying MkSK queries. It is applicable to spatial keyword queries based on other ranking functions. The Authentication algorithms construct a compact VO, i.e., put minimum numbers of objects and MIR-tree entries into the VO. Hence, the communication cost and authentication time are saved. To further reduce the size of a VO, an enhanced AD and the MIR*-tree.

1. Spatial Dataset Processing

Information are collected for few organizations (ATM, Hotels, Petrol bunks, etc.,) from Search Engines & Organization's websites [1]. The information consists of the following attributes: City, Location, Service Provision, Geographical Location, Distance from Main city etc.,

2. Query Authentication thru Key Distribution Center

The mobile client is issuing a query to the SP to get the Top-k results from the Data Server. The authentication of an MkSK query q amounts to guaranteeing that the client always has the correct top-k result while its spatial location changes continuously [7]. The safe zone based approach guarantees that as long as the query does not exit the safe zone, the received top-k result remains valid. In other words, when the query moves across the boundary of a safe zone, it requests an updated top-k result and corresponding safe zone.

The data objects that falls into the verification set are sent to the client. In addition, summary objects that are used to derive bounds on the ranking scores of the objects outside the verification set are sent to the client.

3. Result Authentication using Merkle-IR-Tree (MIR Tree)

The MIR-tree is used for efficiently verifying MkSK queries. In the MIR-tree, to authenticate text relevancies, a word digest is stored with each entry in each posting list in the inverted file attached to each non-leaf node [9]. Formally, for a word w, a posting list entry takes the form (id, weight, $h_w(e)$), where id is the identifier of an entry e in a node in the tree, weight is the weight of w in the pseudo text description of e, and $h_w(e)$ is the word digest of e for word w. Word digests do not occupy any space in nodes. They are stored in the inverted files attached to nodes. The word digests of the root node (e.g., h_c(root) are signed by the Data Owner(DO) and are stored with the MIR-tree.

4. Result Retrieval using Safe Zone Authentication

The challenge of verifying the safe zone is to ensure that the client notices that influence objects are missing if the SP omits some influence objects [2]. The safe zone based approach significantly reduces the communication between clients and the SP as well as computation costs, since the client needs to request new result only when leaving the safe zone.

III. PROPOSED SYSTEM

The framework consists of two phases, i.e., initialization and query processing & authentication. In the initialization phase, the DO first gets a private key from a key distribution center. Next, it signs the ADS constructed on the data set using the private key and transfers the ADS and signatures to the SP. A client downloads a public key from the key distribution center and the signatures from the SP. In the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 01, January 2018

query processing and authentication phase, the client first issues an MkSK query [10]. Upon receiving the query, the SP computes the top-k result, the safe zone, and a verification object that encodes the query result and its safe zone. The client gets the VO from the SP. The top-k Result Set (RS) and its safe zone are obtained from the VO. The correctness of the top-k result and the safe zone can be verified by the client using the VO, the signatures, and the public key. The client needs to send a new request to the SP only when it leaves the safe zone.

IV. VO CONSTRUCTION ALGORITHM

The ADS and the MIR-tree has the desirable property that only a minimum number of objects and MIR-tree entries need to be inserted into the VO. A Verification Set (VS) covering the objects with ranking scores smaller than rank [3]. The smaller the score, the better the ranking. Algorithm 1 shows the pseudo code for constructing the VO for the top-k result. The VO is computed by a depth-first traversal of the MIR-tree using the following conditions: (i) if a non-leaf entry e has a ranking score that is higher than rank k, the VO entry for e is constructed and added to the VO, and its subtree will not be visited (ii) for any visited leaf node, all the objects in it are added to the VO. When constructing the VO, a traversal string str is composed that tracks the search in the MIR-tree. It contains the identifiers of the tree entries and objects added to the VO, as well as the special tokens '½' and '_' used to mark the scope of a node The traversal string is needed in order to avoid duplicate entries in the VO, since the VO for the top-k result and the VO for the safe zone may otherwise have common entries. Finally, the VO and the traversal string str are sent to the client.

Equations

- Soundness of a Top-k result

$$\forall p \in RS(p \in D)$$

Every object in the top-k result must be present in Data Object's (DO) dataset. Specifically, both the location and the text description of the object are not tampered with and the ranking score $rank_q(p)$ of p is computed correctly.

- Completeness of a top-k result

$$\forall p \in RS(\forall p' \in D - RS\left(rank_q(p) \leq rank_q(p')\right)$$

All objects not in the top-k result have ranking scores that are no better than any of those of the result objects.

- Soundness of a safe zone

$$\forall p \in I(p \in D)$$

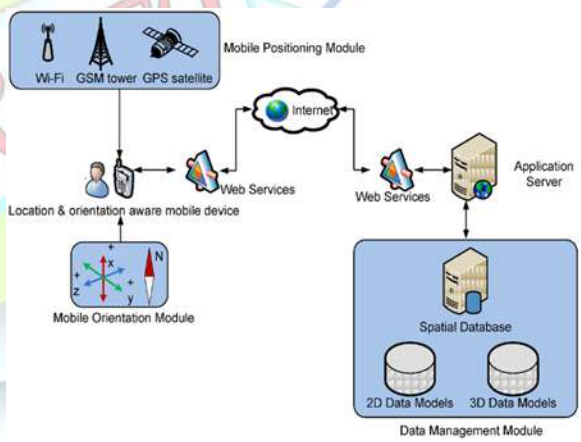
Since a safe zone is computed from the result objects and the influence objects, the facts that the locations and the text descriptions of the influence objects are not tampered with, the text relevancies of influence objects are computed correctly, and sub-condition guarantee the soundness of a safe zone

- Completeness of a safe zone

$$\forall \lambda' \in \Upsilon^k(RS)(Q(\lambda', \Psi, k) = Q(\lambda, \Psi, k) \wedge \forall \lambda' \notin \Upsilon^k(RS) Q(\lambda', \Psi, k) \neq Q(\lambda, \Psi, k)$$

For all locations λ' in the safe zone, the result of the query with parameters λ', the keywords of the original query Ψ, and k on the original data set is the same as the result of the original query $Q(\lambda, \Psi, k)$. For all λ' not in the safe zone, the two results are different.

Figures and Tables



V. CONCLUSION

Top-k spatial queries are a useful tool for location-based applications. A new approach for processing top-k spatial queries is found. The top-k results and safe zones of MkSK queries are designed. Algorithms for constructing and using verification objects for verifying the top-k results and safe zones are developed. An enhancement of the MIR-tree, the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 01, January 2018

MIR*-tree is proposed to reduce the communication cost. Extensive empirical studies on real data sets demonstrate that the approaches are capable of outperforming two baseline algorithms that utilize existing techniques by orders of magnitude.

REFERENCES

[1]. NIST, "Proposed federal information processing standard for digital signature standard (dss)," Federal Register, vol. 57, no. 21, pp. 3747–3749, 1992.

[2]. "Secure hashing algorithm," Nat. Inst. Sci. Technol., FIPS 180-2, 2001.

[3]. F. Aurenhammer and H. Edelsbrunner, "An optimal algorithm for constructing the weighted Voronoi diagram in the plane," Pattern Recognit., vol. 17, no. 2, pp. 51–57, 1984.

[4]. D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "Aggregate and verifiably encrypted signatures from bilinear maps," in Proc. 22nd Int. Conf. Theory Appl. Cryptographic Techn., 2003, pp. 416–432.

[5]. T. Brinkhoff, "A framework for generating network-based moving objects," Geoinformatica, vol. 6, no. 2, pp. 153–180, 2002.

[6]. Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient query processing in geographic web search engines," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2006, pp. 277–288.

[7]. W. Cheng and K.-L. Tan, "Query assurance verification for outsourced multi-dimensional databases," J. Comput. Security, vol. 17, no. 1, pp. 101–126, 2009.

[8]. G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," Proc. VLDB Endowment, vol. 2, no. 1, pp. 337–348, 2009.

[9]. P. T. Devanbu, M. Gertz, C. U. Martel, and S. G. Stubblebine, "Authentic data publication over the internet," J.Comput. Security, vol. 11, no. 3, pp. 291–314, 2003.

[10]. R. Fagin, A. Lotem, and M. Naor, "Optimal aggregation algorithms for middleware," J. Comput. Syst. Sci., vol. 66, no. 4, pp. 614–656, 2003.