ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

EFFICIENT TOP-K RANKING SPATIAL KEYWORD QUERY PROCESSING

P.VANMATHI[1], A.DHARSHINIPRIYA[2], S.VIMAL[3], L.ARTHI[4]
UG SCHOLAR[1,2,3,4]
DR. MAHALINGAM COLLEGE OF ENGINEERING AND TECHNOLOGY, POLLACHI

Abstract: The system focuses on outsourcing the relevant result set from the database using efficient Moving Top-k Spatial Keyword (MkSK) queries. An MkSK query, which takes location part in a query, enables a client to be aware of the spatial web objects that best match a query with respect to location and text relevance. The system designs a verification object (VO) for authenticating MkSK queries. The VO will be mailed to the client and it is automatically updated in the webpage. By the VO along with client request, the query will retrieve the related latitude and longitude values from the data owner using MkSK query processing and point out the values in Google map. This system also uses authentication data structures called MIR-tree and MIR*-tree, guarantees that the client has the appropriate top-k result while its spatial location changes continuously.

Keywords: MkSk Queries, Verification object, MIR, MIR*.

1. INTRODUCTION

The increasing mobile use of the internet render it to consider a scenario where spatial keyword search is outsourced to a service provider capable at handling spatial web objects from various sources. The data owner is in the need for outsourcing their spatial data to SP. This provides flexible resources that can effectively reduce the operational and maintenance expenses for data owners. In this paradigm, the data owner delegates the management and maintenance of its database to a SP and the SP is responsible for indexing the data and answering client queries.

This system also aims about location and text search. Taking a location and a set of keywords as arguments, such queries return relevant spatial web objects that match the arguments. Spatial web objects can be points of interest (e.g., restaurants, tourist attractions, hotels, entertainment services) with a web presence, and they have locations as well as textual descriptions. A moving top-k spatial keyword (MkSK) query, which takes into account a continuously moving query location, enables a mobile client to be continuously aware of the top-k spatial web objects that best match a query with respect to location and text relevance, and it has numerous mobile uses.

For example, a mobile client may activate a "restaurant" query in order to be alerted about nearby opportunities to have their lunch. With the MkSk query, a client always has an up-to-date result as the client moves. Upon receiving the query, the SP computes the top-k result and a verification object that encodes the query result. The client gets the VO from SP and then the correctness of the top-k results are verified by the client using VO.

2. Existing System

The existing system authenticate an MkSK query which includes verifying both the top-k result and the accompanying safe zone .The safe zone is calculated based on the objects in the top-k result, so that missing a non-result object may cause a safe zone to fail in the authentication and then the result will fetch the data around the fixed range. So, it will not satisfy the client and the result may also not relevant (i.e. the result which is so away from the user location)

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 query processing and authentication phase, the client first issues an MkSK query. 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 RS and its safe zone results 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.

The Merkle-IR-tree (MIR-tree) is developed based on the IRtree [8], by embedding a series of digests in each node of the tree. In the IR-tree, each entry summarizes the spatial distances and text relevancies of the entries in its child node. This index enables the efficient processing of spatial keyword queries, since it is able to prune the search space according to spatial proximity and text relevance simultaneously. 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. Formally, for a word w, a posting list entry takes the form id, weight 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) is the word digest of e for word w.

The challenge of verifying a top-k result involves the design of a compact VO that achieves low communication cost and short authentication time. The inefficiency of authenticating spatial distances and text relevancies separately using existing techniques can be illustrated by the following example. Imagine a data set of five objects p1, p5 having spatial distances 1, 2, 3, 4, 5 and text relevancies 5, 6, 7, 8, 20, with regard to a query. According to the ranking function, objects p1 and p5 make up the top-2 result with ranking scores 0.2 and 0.25. If authenticating the spatial distance and text relevance separately, one verification set covering the objects with spatial distances smaller than 5 (the maximum spatial distance in the result) and the other verification set covering the objects with text relevancies larger than 5 (the minimum text relevance in the result) are constructed. Therefore, the whole data set is covered by the two verification sets and added to the VO, which incurs very high communication cost. Ideally, only one verification set covering the objects (p1 and p5) with ranking scores smaller than 0.25 (the maximum value of a ranking score of a result object) should be constructed. Thus, only three objects p1, p5 and a summary object representing all the other objects in the data set make up a VO that achieves low communication cost.

The client extracts the top-k result from the VO and authenticates it. It first computes the ranking scores of the objects in the VO ranks them, and obtains the top-k objects. Next, the client re-computes the root spatial and word digests using the str_{rs} and VO and verifies them against the decrypted root spatial and word signatures. The main idea is to re-construct the MIR-tree traversal performed by the server, i.e., guaranteeing the entries in the VO are from the original MIR-tree. Then, it verifies the correctness of the top-k result by checking whether the ranking score computed from Equation (1) of the k^{th} object is no worse than those of all the other entries in the VO.

2.1Limitations

Existing system has a straightforward solution to the MkSK query which is to periodically invoke an existing spatial keyword query processing technique. However, the queries are processed very frequently, the results are not appropriate because consecutive results are likely to be very similar. There is no guarantee that the user always has the right, up-to-date result and less computation, the client will not be satisfied with these issues.

This system helps the client to retrieve the results only around the safe zone and the results will not be appropriate as per the client request. It displays only the nearest hotel which matches the client request not based on any ranking system. This approach is inefficient since it generates maximum number of objects in the result set.

3. Proposed System

The proposed system is about the authentication of moving top-k spatial keyword queries. The web services are enabled in this system through the intermediate called Service Provider. By using SP, the system may aware of voluminous attacks. The system uses authenticated data structure, the MIR-tree, to efficiently authenticate moving top-k spatial keyword query results. A verification object for authenticating the top-k results and safe zones of MkSK queries is designed. The proposed system has its scope in domains like ranking system. The advantages of this

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

system are using authenticated data structures and construct verification objects.

3.1 Advantages

- The result will be retrieve the results based on ranking.
- The authentication will be provided for the user.

4 Proposed Solution

- Location Tracking
- Authentication Framework & Query Processing
- MIR & Verification Object
- Top K result

4.1 Location Tracking

In this module, track the user's location with the help of latitude and longitude values. The latitude and longitude value varies according to the client's location. Through this, the user can find his/her current location and views all the hotels in map by entering the location.

4.2 Authentication framework & Query processing

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. The client downloads a public key from the key distribution center and the signatures from the SP. Query Processing: In the query processing and authentication phase, the client first issues an MkSK query. Upon receiving the query, the

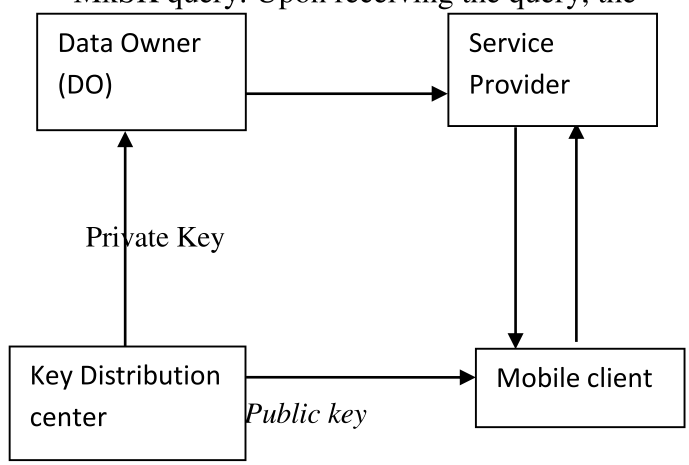


Fig 1 Authentication Framework

SP computes the top-k result. Spatial keyword query processing is also receiving where a range of techniques have been proposed for efficiently processing spatial keyword queries. Let D be a dataset in which each object p€D is a pair <λ,ψ> of a point location p.λ and a text description, or document, p. ψ.

Location= p.λ

Document=p. ψ

A moving top-k spatial keyword(MkSk) query q=< λ,ψ>

ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.3, Special Issue.24, March 2017

4.3MIR & Verification Object

The Merkle-IR-tree (MIR-tree) is developed based on the IR tree, by embedding a series of digests in each node of the tree. In the IR-tree, each entry summarizes the spatial distances and text relevancies of the entries in its child node. This index enables the efficient processing of spatial keyword queries, since it is able to prune the search space according to spatial proximity and text relevance simultaneously. MIR-tree, is used to check text relevancies, a word digest which is stored with each entry in each posting list in the inverted file attached to each non-leaf node.

hw(e)=h((ce1, weight,hw(ce1))|(ce2, weight, hw(ce2))| ..),h()-hash function

Cei(i=1,2,3)- entry in the child node of e.

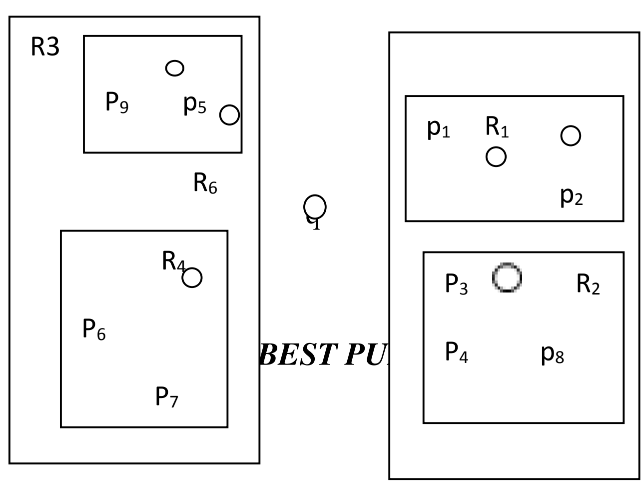
The MIR-tree, has the desirable property that only a minimum number of objects and MIR-tree entries need to be inserted into the VO.

The VO size is critical to minimize the client/server communication cost and reduce the client-side result verification time. We proceed to propose an enhanced ADS, the MIR*-tree, that enables a reduction of the VO size. Specifically, the digests in the MIR*-tree are computed differently than in the MIR-tree. The verification object has been sent at the time of user searching the location.

4.4Top K result

The challenge of verifying a top-k result involves the design of a compact VO that achieves low communication cost and short authentication time. The client extracts the top-k result from the VO.

Retrieve the top-k objects with the highest ranking scores, measured by the combination of their distance to the query location and the relevance of their text description to the query keywords chicken, mutton and veg .



○

Figure 2: Top-K Result

Figure 2 depicts the top-k spatial web objects. Let q be an MkSK query with query keyword 'c', where k=1. In this figure, the p_1,p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9 are the web objects whereas R_1, R_2, R_3, R_4 are the appropriate results. Its top-1 result is p1 by using traversal string str_{rs} is $[[[p_1p_2][p_3p_4p_8]][R_3R_4]]$ and then the verification set is constructed and sent to the client. The client now retrieve the top-k results.

5.Conclusion And Future Work

The authentication of MkSK queries, aiming for a solution that (i) verifies the soundness and completeness of the top-k result and the safe zone of a given MkSK query and (ii) optimizes the server-side computational cost and the client/server communication cost, i.e., the VO size.

The future work is planned to find the current location of client by using Wi-Fi and the dataset will be updated automatically to the client.

REFERENCES

[1] Berlin Heidelberg "Efficient and Scalable Method for Processing Top-k Spatial Boolean Queries" in Proc. 33rd Int. Conf. Very Large Databases,pp. 147–158, 2007.

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

[3] 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.

[4] Dingming Wu · Gao Cong · Christian S. Jensen Springer "A framework for efficient spatial web object retrieval - VLDB J., vol. 21, no. 6, pp. 797–822, 2012

[5] I.D.Felipe, V.Hristidis, and N.Rishe "Keyword Search on Spatial Databases" in Proc.Int.Conf.Data Eng., pp. 656-665, 2008

[6] Gao Cone Dingming Wu, Lyon, "Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects",Proc.VLDB Endowment,vol. 2,no 1 France August 24-28,2010

[7] H. Hu, J.Xu, Q.Chen and Z.Yang, "Collective spatial keyword querying" in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2012, pp. 301–312.New York, NY, USA © 2011

[8] D.L.Huff, "Efficient continuously moving top-k spatial keyword query processing" IEEE Regional Sci.,vol. 7, no. 3, pp. 323–329, 1973. Washington, DC, USA ©2011

[9] R.C,Merkle, "Efficient Safe-Region Construction for Moving Top-K Spatial Keyword Queries" in Proc. Adv. Cryptology,1989, pp. 218–238.October 29–November 2, 2012

[10] E. Mykletun, M. Narasimha, and G. Tsudik, "Authentication and integrity in outsourced databases," ACM Trans. Storage, vol. 2,pp. 107–138, 2006.

[11] H. Pang, A. Jain, K. Ramamritham, and K.-L. Tan "Verifying Completeness of Relational Query Results" in Data Publishing.. In SIGMOD, 2005.

[12] H. Pang and K. Mouratidis. "Authenticating the query results of text search engines", PVLDB, 1(1):126–137, 2008.

[13] D. Wu, M. L. Yiu, C. S. Jensen TODS, "Moving spatial keyword queries: formulation, methods, and analysis 38(1):7, 2013

[14] D. Yung, E. Lo, and M. L. Yiu, "Authentication of moving range queries", 21st ACM Int. Conf. Inf. Knowl. Manage.,pp. 1372–1381, 2012

[15] W. Zeng and K.-L. Tan, Spatial datamining-Datamining Concepts and techniques, 2 edition Hankamber in Proc, Int. Conf. Data Eng, pp. 560–571. 2004.

[16] J. Zhang, M. Zhu, D. Papadias, Y. Tao, and D. L. Lee. "Location-based Spatial Queries", In SIGMOD, 2003.