Privacy preserving in un-trusted cloud environments for query shortest path

Zhang Lei[1, a], Li Jing[2,b]*(Author for correspondence), Wang Bin[2, c]
[1]Office of Academic Affairs, Jiamusi University, Jiamusi 154007, China
[2]College of Information Science and Electronic Technology, Jiamusi University, Jiamusi 154007,China
[a]8213662@163.com, [b]lijing2483@163.com, [c]jmsuwang@163.com

Keywords: Privacy, database outsourcing, cloud environment

Abstract. The shortest path query is one of the most popular location-based services for navigation. With it, one can determine a best route from a source to a destination in an unfamiliar environment and then save time for travelling. But under this service, some sensitive information may be leaked to others due to its contents, especially when the owners of the routing data do not have the necessary infrastructure to run and maintain a system for processing shortest path, that they outsourcing it to the cloud environments for computing. The cloud server may be un-trusted and has a strong ability of calculation, then it can launch structural pattern attack and the graph reconstruction attack, and with computing the shortest path of the network diagram it will acquire the user's sensitive information. In this paper, we propose a protocol that allows the query of the shortest path in cloud environments and while the same time protecting the querying owner's privacy, so the user can enjoy the convenience of the location service without worried the leak of privacy.

1 Introduction

With the improvement of geo-positioning technologies and the widely application of smart phones, location based service (LBS) technologies have produced a profound effect on people's life. As a form of service, the LBS can let people retrieve the nearest neighbor point of interesting (POI) places or retrieve the location of the destination route gradually becomes indispensable in People's Daily life, and then brought a lot of business opportunities and business interests. Then along with the development of the cloud service technology, with the powerful computing capacity of cloud services, the location based service showed more great application prospect and value.

Nevertheless, use of LBS may pose a privacy threat to whom enjoying the advantages of it. Especially for people query the nearest POI or shorted path with their actual location. Notice that, with those clues the adversary can easily get such as user residence, workplace, religion and politics, which can lead to such as stealing, personal injury, political attack etc. As the existence of potential attack, some parts of user became concerns about their privacy leak and dare not to use. So the problem of privacy prevents the development of LBS to a certain extent. And at the same time, this also affected the expansion of the service industry and the improvement of service quality.

Recently, the privacy problem has drawn extensively interests in researching, and most of the methods are applied to the nearest neighbor query. However, the privacy risk of shortest path query is far higher than that of the nearest neighbor query, because the shortest path query need to protect not only the current location privacy, but also the privacy of the source and destination node and even the path which between them. Lee, et al. in [1] put forward the navigation of the shortest path privacy preserving on obfuscation. It obfuscates the source and destination node with the same query demand nodes or adds dummy nodes and sent the notes set to the LBS. Then the LBS computes the shortest path with the notes set and return the path set to the client. Through the data set the LBS will be confused with the actual query, and then the purpose of preserving the client's privacy is achieved. Mouratidis et al [4] uses a no information leak mechanism. In their mechanism they pre-computation the shortest path of every pair of nodes, and store those edge-weight data in the severs, when the clients query the shortest path, they use the privacy information retrieval (PIR) method to require the data page, and then client find out the shortest path which he need from the data page. Xi et al. [3] based on the calculation of minimum path algorithm of Floyd-Warshall algorithm, create a matrix for the pairs of nodes, and instead of pre-computation every node pairs

they use the PIR methods to compute the edge-weight of every hop and then calculation the shortest path.

The privacy of query shortest path maybe solved through the methods mentioned above. However, along with the cloud outsourcing technology puts forward and development, this technology beside improve the ability of service also take with the privacy problem back again, and the leak of information between the client and server changed to among the client, data owner and cloud server, which makes the methods powerful in conventional LBS becomes unable to adapt the new structure of service, and to preserving the privacy we must to find another methods.

In this paper, we provide a privacy preserving method which in un-trusted cloud environments of outsourcing data for the shortest path query. The method is based on the transformational network graphs, and the obfuscate mechanism this method can preserve the privacy from the data owner and the cloud server. The contributions of this paper are:

- We formalize a general methodology that provably achieves the privacy preserving in outsourcing data for the shortest path query;
- We develop specific schemes that implement this general methodology;
- We evaluate our solutions on road network graphs and assess their trade-offs;
- As far as I know it's the first time to formalize methodology of query privacy preserving of the shortest path in cloud environments.

The remainder of this paper is organized as follows. Section 2 reviews the existing related work in privacy preserving of the shortest path. Section 3 details the protocol of processing framework and discusses the main functionalities. Analyze this theory in Section 4. Finally, Section 5 we conclude the paper.

2 Related Works

In this section we discuss the obfuscation and PIR-based methods in the LBS, and along with the work of privacy preserving in outsourcing data for cloud environments.

2.1Obfuscation and PIR-based Methods for Privacy Preserving.

Among the privacy preserving methods with the obfuscate mechanism, k-anonymity is the most famous one, which is imported by Gruteser et al. [7] from the release of the data privacy protection. In [7] anonymizer replaces the coordinates of the originating client u with a region (usually a square or a circle) that includes u and at least k-1 other clients, because there are k clients in the region others unable to identify which the k clients inside the anonymous region is the originator. There exist several methods based on this mechanism. For example, the [8,9,10] is methods for privacy-preserving scheme for continuous query and the [12,13,14,11] is about reducing the anonymous region, however, on the shortest path query we only see one method in literature [1].Its mechanism is: anonymizer chooses the three different kinds of obfuscate methods for shortest path query, namely, independent obfuscated path query, shared obfuscated path query and anti-collusion obfuscated path query. With the degree of privacy from the client sets, anonymizer choose the a method and different amounts of fake nodes, in order to reduce the calculate of redundant data, the fake nodes and real notes must satisfy the amounts of the method need, so the privacy preserve method can't impact the efficiency of the shortest path algorithm. This method has to preserve clients' location privacy from LBS, but the anonymizer sometimes maybe un-trusted, so the method of no information leakage is provided.

PIR method is a type of no information leakage mechanism, and also popular on the k-nearest neighbor query. With the encrypt process of the query information, the PIR is also no information leak to the clients vicinity. Ghinita etc. in [15] first import this computationally PIR into the location privacy preserving, then malicious research is based on this mechanism. Khoshgozaran etc.[16] puts forward the concept of hardware PIR, and establish an index tree method for PIR. In proposes a method using latitude and longitude different decimal in PIR to reduce the high consumption in computing and storage. For the shortest path problem, in literature [4], the LBS pre-computation the shortest path of every pair of nodes, and store those edge-weight data in severs, when the clients query the shortest path, they use the general hardware PIR method to retrieval data

pages they query. In [16] they use the secure co-processor (SCP) for reduce the processing complexity. Xi et al. [3] based on the Floyd-Warshall algorithm, create a matrix for the pairs of nodes, and instead of pre-computation every node pairs they use the PIR methods to compute the edge-weight of every hop and then calculation the shortest path.

2.2Privacy Preserving in Outsourcing Data.

For location privacy preserving in cloud environments of outsourcing data, the two methods for compute the k-nearest neighbor (kNN) are proposed in [6], they all assume cloud environments to be un-trusted, although not, maybe semi-honest, which means processing of kNN must no information leaked. Through methods are powerful in securing kNN queries, is obviously useless for shortest path computation. Therefore [2] proposed a method which transforms the original graphic into outsourcing graphic with 2-HOP delegation, and then cloud server can't reconstruct the original graphic with the outsourcing graphic. Homomorphic encryption method is used in [5] for computation edge-weight in privacy. The same as [2] this method also needs the source and destination for computation, and can't preserve privacy too. Method in [17] can compute the distance between two nodes with no information leak, but this method is suitable for Euclidean distance calculation, and obviously useless for achieving the shortest path in network graphs.

3 The Protocol

In this section we first introduce essential preliminary concepts. We then present the frame of the protocol which can solve the problem of the privacy of shortest path in cloud environments.

3.1Protocol Model.

The system model comprises of three distinct entities: (1) the data owner; (2) the outsourced cloud service provider（for short cloud server）and (3) the client. The entities are illustrated in Fig.1.

The data owner has the routing data set but does not have the infrastructure to run for processing the shortest path, and outsource the data set calculation to a cloud provider. As the data set of the geo-position is calculated in cloud also has the nodes of clients, there must be some processing to transform the network graphs into a type which cloud can't be reconstructed, and send transformation to cloud server.

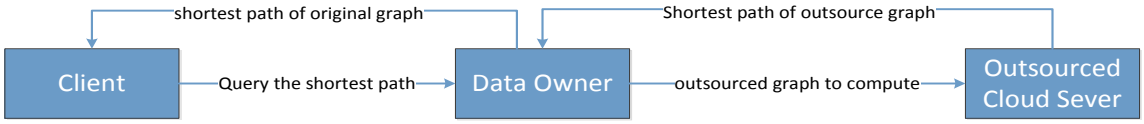


Fig. 1.Protocol model

The cloud server receives the transformational data set from the data owner; it calculates the shortest path. Because there is no query node relative to a transformed one, the cloud server does not know any information about the query of clients. Although the cloud server typically processes powerful computational resources, processing on the shortest path in Dijkstra or Floyd method incurs a significant processing overhead. In this protocol we only use cloud server for shortest computation, and the data set is not encrypted.

The client has a query Q and wishes to find the shortest path to the destination. The client sends a query to data owner, and data owner finds the network graphs which contain the source and destination nodes, then sends transformed graph to cloud server for shortest computation. Note that, in sends query to data owner the location information of the client must to be preserved.

3.2 Preliminary Assumptions and Attacker Model.

In this paper, we formalize the problem of shortest path in outsource data, and the main objective is to preserve clients' privacy from the outsource cloud server. So for tackling it systematically, we now present a series of security and privacy assumptions. We shall assume that:

- The computing capacity of cloud server is powerful, and can efficiently compute the shortest path with the complex road network graphic, even the data sets are very larger.
- There is no collusion between the data owner and the cloud server. This assumption is a reasonable one, because both the data owner and the cloud server's interest are to maximize their business benefit. Collusion between them would result in a loss of trust.

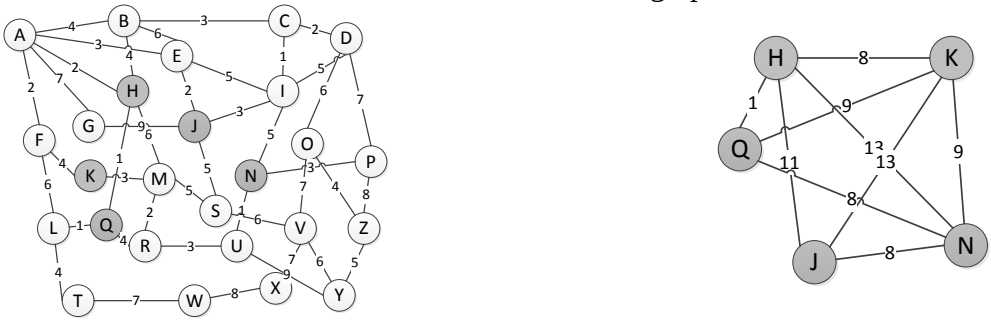
- The data owner is trusted, but the cloud server is un-trust. We regard the methods mentioned above are efficient and in the process of client send information to data owner there is no information leaked. The cloud server is semi-honest. It executes the process of computing shortest but curious about the source and destination nodes of the query without malevolence. With the powerful capacity of computation, the cloud server can reconstruct the network graphs for acquiring the source and destination nodes, and even relevant them with the shortest path.
- When querying, clients maybe in the edges or in nodes, for this protocol we formulate the graphs' nodes as the source nodes, the query is from the nodes not the current position of clients. The original network graph G=(V,E), the source node is s∈V and destination node is d∈V, (s,d)∈E is the edge sets from s to d. The shortest path is the edge sets which weight sum is the smallest.

3.3 Protocol Processing.

Processing of this protocol is divided into two phases, one is the communication between clients and data owner for requesting the road network graphs, the other is the communication between data owner and cloud server for computing shortest distances. We believe that the first phase is trustful, and there is no information leaked, so this protocol is mainly to deal with the second phase. This phase is for disposing of road network graphs. In this phase we use the pattern from [2], transform the original graph into 2-HOP delegation graph. Then we formally define as follows:

Definition 1(Delegation nodes) Let G = (V, E) be an original graph, n is the nodes number, the nodes of original graph $V = \{N_1, N_2, ...N_n\}(1 \le n)$, DV is the nodes random selected in V, and with those nodes we can transform the original graph into a outsourcing delegation graph, $DV \subseteq V, DV = \{N_{D1}, N_{D2}...N_{Di}\}(1 \le i \le k)$, where k is the number sets by client. DV is called as the delegation nodes.

Definition 2(Shortest path) with the definition 1, let e_{N_1,N_2} be the edge between N_1 and N_2, $w(N_1, N_2)$ is the weight of edge e_{N_1,N_2} .The shortest distance is the sum of weights between two nodes, we use δ_G define the shortest distance of two delegation nodes which pass without the delegation nodes, if the shortest distance pass the delegation nodes we see there is no edges between them, then use the sets of δ_G and delegation nodes construct the outsource graph. The shortest path is the shortest distance between two nodes of the outsource graph.



(a)Original graph of routing with delegation nodes (b) 2-HOP delegation graph

Fig.2. 2-HOP delegation graph

We illustrate the transformation process of the graph in figure 2.Data owner randomly selects delegation nodes that filled with gray in Fig.2a. Then the data owner pre-computes shortest distance sets of δ_G and constructs the delegation graph. In Fig.2b there is a delegation graph constructs from Fig.2a, we may see there is no edge between nodes Q and J, because the shortest distance passes the node H. With the 2-HOP delegation graph though computing powerful the cloud server is, it can't infer the current routing case through structural pattern attack and the graph reconstruction attack.

We now proceed to describe the protocol. The data owner pre-transform the current road networkgraphic data G into a set of outsourced delegation $G_o = \{G_o^1, G_o^2, ...G_o^{|G|}\}$, and deploy G_o on the cloud server. A link graph G_l is conserved on the data owner, which maintains the relationship between a node in G and a delegation node in G_o .When the client request shortest path to its current

position and destination nodes, it sends the request to data owner. After receiving the request, the data owner check whether its current position is at the edge of the graph, if so, the data owner chooses the nearest node of the edge as its query position. Then the data owner random chooses G_o^c from G_o, where G_o^c contains the source and destination nodes. After those, it sends the chosen G_o^c with the transformed nodes to cloud server. The cloud server computes the shortest distance with the nodes and G_o^c, then sends the shortest distance back to the data owner. The data owner receives the distance and combines with the nodes to transform them with G_l into the original graph. So the shortest path in the original graph is acquired and the data owner sends it to the client. The entire process cloud server only disposes the delegation graph, and can't reconstruct the original graph with through structural pattern attack and the graph reconstruction attack. (The 2-HOP delegation graph can resist the reconstruction attack can be seen in [2])

4 Theoretical analyses

In this section, we analyze the structure of our protocol described in Section 3.On one hand, we analyses the security for the client. On the other, we analyses the processing cost of the protocol.

4.1 Security Analysis.

As the 2-HOP delegation graph can resist structural pattern attack and the graph reconstruction attack. We analyze the security from the probability of the cloud server gain query information infers from the graph which it has conserved. This information can be gained from the relation with client's source and destination nodes and the shortest path.

Now we analyze from the worst case condition. Assume the delegation graph only has three nodes in it. The data owner sends the transformed source and destination nodes to cloud server with two nodes for the shortest path, and the cloud server computes in the delegation graph with three nodes, so the probability of the nodes been inferred by cloud server is 2/3, and edge is 1/3. This means once the cloud server corresponds the delegation graph with the original graph, it has 1/3 probability to know the path of the client. We must sure that the road networkgraphic is very complex, and for security the data owner chooses the delegation must larger than three nodes. If there are four nodes in the delegation graph, the probability is 1/6, when the nodes number is n this may be $1/C_n^2 = 2(n-2)!/n!$. At the same time, delegation nodes have no relation with the original graph, the complexity of the cloud server to infer the query of the client even more difficult.

4.2 Complexity Analysis.

The client sends a query to data owner's temporal cost is depended on service of quality which no relation to this paper, and we assume the cloud server has powerful computation capacity without temporal cost in computing the shortest path. So we only analyze the temporal cost in processing under data owner. The client sends the source and destination nodes to data owner; the data owner first finds the pair of nodes in road networkgraph, once this graph has n nodes in it, the finding temporal cost is O (n). After finds the nodes in road networkgraph, the data owner need to link the pair of nodes to the delegation graph sets with G_l, this link phase need to find the delegation graph which contain the pair of query nodes, Because $G_o = \{G_o^1, G_o^2, ...G_o^{|G|}\}$,to find the related delegation graph from this set needs O (n).So we can see all the temporal cost of this protocol process is O (n)+ O (n)= O (n).

5 Conclusions and Future Works

In this paper, we study how to preserve privacy for shortest path query in un-trusted cloud environments. We define a protocol which uses the 2-HOP delegation graph to resist the structural pattern attack and the graph reconstruction attack. But in this protocol we assumes the data owner as an honest one, in the real world it maybe sometimes doesn't true, so in the future we will do research in this phase.

Acknowledgments

This work was supported by Jiamusi university of science and technology research key project(project number：Lz2013-010, Lz2013-011, Lz2014-005), Jiamusi university of science and technology research youth fund projects(project number：Lq2013-033), Jiamusi university of Science and technology research projects(project number：12541788), The youth program of national natural science fund(project number：61203052).

References

[1] Lee K C K, Lee W C, Leong H V, et al. Navigational path privacy protection: navigational path privacy protection[C]//Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 691-700.

[2] Gao J, Yu J X, Jin R, et al. Outsourcing shortest distance computing with privacy protection[J]. The VLDB Journal—The International Journal on Very Large Data Bases, 2013, 22(4): 543-559.

[3] Xi Y, Schwiebert L, Shi W. Privacy preserving shortest path routing with an application to navigation[J]. Pervasive and Mobile Computing, 2013.

[4] Mouratidis K, Yiu M L. Shortest path computation with no information leakage[J]. Proceedings of the VLDB Endowment, 2012, 5(8): 692-703.

[5] Zhang Yingguang, Su Sen, Chen Weifeng, et al. Privacy-preserving shortest distance computing in cloud environment. J.Huazhong Univ.of Sci.&Tech. (Natural Science Edition), 2013.

[6] Elmehdwi Y, Samanthula B K, Jiang W. Secure k-nearest neighbor query over encrypted data in outsourced environments[C]//Data Engineering (ICDE), 2014 IEEE 30th International Conference on. IEEE, 2014: 664-675.

[7] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proceedings of the 1st international conference on Mobile systems, applications and services. ACM, 2003: 31-42.

[8] Wang E K, Ye Y. A New Privacy-Preserving Scheme for Continuous Query in Location-Based Social Networking Services[J]. International Journal of Distributed Sensor Networks, 2014, 2014.

[9] Pan X, Xu J, Meng X. Protecting location privacy against location-dependent attacks in mobile services[J]. Knowledge and Data Engineering, IEEE Transactions on, 2012, 24(8): 1506-1519.

[10] Hashem T, Kulik L, Zhang R. Countering overlapping rectangle privacy attack for moving< i> k</i> NN queries[J]. Information Systems, 2013, 38(3): 430-453.

[11] Jia J, Zhang F. Nonexposure Accurate Location-Anonymity Algorithm in LBS[J]. The Scientific World Journal, 2014, 2014.

[12] ZHAO Ze-mao, LI Lin, ZHANG Fan, ZHANG Pin, ZHOU Jian-qin, WANG Jia-bo. The location privacy protection method with dispersed sub cloaking region. Journal of Shandong University (Natural Science), 2013, 48(7): 56-61.

[13] Zhong C, Liu L, Zhao J. Privacy-Preserving Location-Based Query Using Location Indexes and Parallel Searching in Distributed Networks[J]. The Scientific World Journal, 2014, 2014.

[14] Talukder N, Ahamed S I. Preventing multi-query attack in location-based services[C]//Proceedings of the third ACM conference on Wireless network security. ACM, 2010: 25-36.

[15] Ghinita G, Kalnis P, Khoshgozaran A, et al. Private queries in location based services: anonymizers are not necessary[C]//Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008: 121-132.

[16] Khoshgozaran A, Shirani-Mehr H, Shahabi C. Spiral: A scalable private information retrieval approach to location privacy[C]//Mobile Data Management Workshops, 2008. MDMW 2008. Ninth International Conference on. IEEE, 2008: 55-62.

[17] Šedšnka J, Gasti P. Privacy-preserving distance computation and proximity testing on earth, done right[C]//Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, 2014: 99-110.