ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

Private kNN Query Processing in Cloud Enviroments

Monika D.Rokade[1], S.A.Kahate[2], K.S..Kore[3]

M.E. Student, Department of Computer Engineering, SPCOE, Dumberwadi, Pune, Maharashtra, India[1]

Head, Department of Computer Engineering, SPCOE, Dumberwadi, Pune, Maharashtra, India[2]

Department of Computer Engineering, SPCOE, Dumberwadi, Pune, Maharashtra, India[3]

ABSTRACT: Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. In this paper, we propose a holistic and efficient solution that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism. The framework is scalable to large datasets by leveraging an index-based approach. Based on this framework, we devise secure protocols for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index. Moreover, several optimization techniques are presented to improve the efficiency of the query processing protocols. Our solution is verified by both theoretical analysis and performance study.

I. INTRODUCTION

Cloud computing has newly emerged as a new stage for deploying, handling, and provisioning large-scale facilities through an Internet-based organization. Successful examples embrace Amazon EC2, Google App Engine, and Microsoft Azure. As a result, introducing databases in the cloud has converted a promising solution for Database-as-a-Service (DaaS) and Web 2.0 submissions. In the cloud computing classical, the data owner subcontracts both the data and quizzing services to the mist. The data are private properties of the data owner and ought be protected alongside the cloud and enquiring client; on the other indicator, the query might release sensitive information of the customer and should be protected alongside the cloud and data owner. Hence, a vital concern in cloud calculating is to protect both data discretion and query privacy amid the data owner, the customer, and the cloud. The public networking service is one of the divisions that witness such rising alarms. For example, in Fig. 1 operator Cindy wants to exploration an online dating place for friends who share with her comparable backgrounds (e.g., age, instruction, home address). Though the site or the data mist should not disclose to Cindy individual details of any user, specifically those sensitive ones (e.g. home address), Cindy oughtnot disclose the inquiry that involves her own facts to the site or the cloud, whichever. More critical examples happen in business sectors, where inquiries may reveal trusted business cleverness. For example, a retailcorporate plans to open a division in a district. To analyze the target customer ignoble, it needs to query the demographic statistics of that district, which the data proprietor has outsourced to a data cloud. While private details in the demographic data must not be disclosed to the subcontracting cloud or the professional, the district name in that probe should not be released to the cloud or data proprietor, either. It is also renowned that the cloud subtracting model worsens the importance of privacy breaches in the overhead scenarios as a single mist may host querying facilities for many data proprietors. For example, two inquiries from the same operator, one on local clinic manual and another on anti-diabetic medications, together give aadvanced confidence that the user is perhaps suffering from diabetes. All the overhead concerns call for ainquiry processing model that conserves both data secrecy and query privacy between the data owner, the customer, and the cloud. The data proprietor should protect its data secrecy, and does not disclose any information outside what the query outcome can imply. On the additional hand, the client must protect its query secrecy so that the data proprietor and the cloud recognize nothing about the query, and is consequently unable to infer any material about the client. Unfortunately, prevailing privacy-preserving query handling solutions are not satisfactory to solve this new unrulyascending in the cloud classical. Most research work in the nonfiction addresses data discretion or query privacy distinctly. For example, popularization techniques have been

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

predictable to protect data pleasure by hiding quasi-identifier characteristics and avoiding the revelation of sensitive information [3]. Comparable techniques are future for query privacy on both interpersonal data and spatial data [16], [12], [13], Solitary very few, such as the Casper∗agenda [5], consider data and query secrecy as a whole. Furthermore, generalization-based explanations like the Casper∗ still reveal the data or query in a rougher and imprecise form. Not abundant research work speeches the unconditional discretion required for this problem. Though some encryption schemes are projected to protect the data introduced on the outsourcing server [1], they cannot be approved in this problem for numerous reasons. First, precise query processing on encoded data is difficult, if not unbearable at all. Most current encryption schemes only provision some specific questions. For example, space alteration (e.g., a space filling curve) castoff in [13] only supports estimatedkNN queries as it cannot reservation the accurate detachments in the original interplanetary. Second, even though appropriate encryptions are originate for these queries, they developed flawed when useful to our problem, as these encryptions are not intended for mutual privacy defense in the first place. In specific, to evaluate the query on the encoded data, the client must scramble the query by the same arrangement and send it to the subcontracting server, who may then advancing it to the data owner, someplace the query can be decrypted by her encryption limitations. Third, some encryptions or alterations are shown to be susceptible to certain security bouts. For example, distanceconserving space transformations are susceptible to principal constituent analysis [12], [15].

II. RELATED WORK

In this unit, we review present privacy-preserving data outsourcing methods for query dispensation purposes. The communal model is that an untrusted subcontracting server stores and achieves the data on behalf of the data owner, who then invitations trusted operators to query the data. The first group of techniques is based on the oversimplification principle to diminish the disclosure of exact information. For interpersonal data, generalization can guard quasi-identifier attributes and circumvent the disclosure ofsubtle information [3]. For spatial data and query, a comparable technique called place cloaking has been projected to generalize (i.e., blur) the user or thing locations [16], [12], [13], [18]. Though, these techniques still reveal the data or query in a rougher and imprecise form. The second class encrypts or transforms both the query and the data into additional space for assessment, using hashing or interplanetary filling curves. Agrawal et al. planned an order-preserving encryption arrangement(OPES) for 1D numeric standards [1]. SQL statements such as MAX, MIN, COUNT, GROUP BY, and ORDER BY can then be redrafted and processed ended the encrypted data. Though, OPES does not provision SUM or AVG statements, in which the unique data must be decrypted first. Yiu et al. protracted this transformation method to 2D spatial data opinions and proposed ranked space-division (HSD) . To defend mutual privacy as in this paper, Khoshgozaran and Shahabi used interplanetary filling arcs as the transformations for adjacent neighbor search [20]. Though, since distance is not totally preserved in the distorted space, the results are only estimatedkNNs. Another drawback of transformation techniques is the possible disclosure risks. As they are nearly distance-preserving, opponents may utilize this information and recover the original data by lined algebra or principal component examination. To resolve this matter, Wong et al. freshly proposed a new encryption structure that only conserves the scalar product worth of two points, which is satisfactory to answer accurate kNNprobes. Our work falls into this class but distinguishes itself since the others as existence the first work that is devoted to mutual privacy guard. We propose a secure, encryptionintegratedoutline that is suitable for handling complex queries over large-scale, indexed data. It is notable that privacy-preserving exploration on tree-structured data has been deliberate in some current studies [6], [2]; however, these mechanism either consider one-way confidentiality or cannot provide unqualified privacy guarantee. The third group considers a distributed situation where the data are divided and outsourced to a set of self-governing and non-colluding subcontracting servers. The privacy-preserving query handling requires a distributed and protected protocol to evaluate the result deprived of disclosing the data in each subcontracting server. The security basis of such protocols originates from protected multiparty subtraction (SMC), a cryptography problem that computes a safe function from multiple contributors in a distributed network. Privacy-preserving adjacent neighbor queries have been deliberate in this context for data removal. Shaneck et al. presented aanswer [24] for point data on two parties. Qi and Atallahenhanced this solution by smearing a blind-and-permute procedure, together with a secure assortment and a multi-step kNN protocol.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

III. PROBLEM STATEMENT

In the planned research work, we emphasis on solving the k-nearest national (kNN) query problem over encoded database outsourced to a mist: a user issues an encoded query record to the cloud, then the cloud returns the k nearby records to the user. We first current a basic scheme and establish that such a naive result is not secure. To deliver better security, we propose a protectedkNN protocol that defends the confi- dentiality of the data, operator's input query, and data admission patterns. Also, we empirically examine the efficiency of our procedures through various trials. These results indicate that our safe protocol is very effectual on the user end, and this frivolous scheme allows aoperator to use any mobile device to achieve the kNN query

IV. PROPOSED SYSTEM

The organization model comprises of three separate entities: (1) the data proprietor; (2) the outsourced cloud facility provider (for small cloud server, or just server); and (3) the client. The objects are illustrated in Number 1. The data owner has a dataset with n two-dimensional facts of interest, but does not have the compulsory infrastructure to run and uphold a system for processing nearest-neighbor probes from a large number of operators. Therefore, the data proprietor outsources the data stowage and querying services to a cloud breadwinner. As the dataset of opinions of interest is a valuable reserve to the data owner. The waiter receives the dataset of points of attention from the data owner trendyencrypted format, composed with some additional encoded data structures (e.g.,Voronoi diagrams, Delaunay triangulations) wanted for query processing (we willdeliver details about these constructions in Sections 2.2 and 2.3). The server obtainskNN requests from the customers, processes them and revenues the results. Though the cloud provider characteristically possesses powerful computational capitals, processing on encrypted data experiences a significant processing above, so performance thoughtsat the cloud server represent anchiefconcern.The client has a query idea Q and wishes to find the point's adjacentneighbors.The client sends its encoded location query to the server, and obtains k nearestneighbors as aeffect. Note that, due to the detail that the data points are encoded, the client also wants to perform a small part in the query dispensation itself, by supplementarywith certain .

Figure 1: System Module

Privacy Model

As declared previously, the dataset of opinions of interest represents an imperative asset for the data proprietor, and an important foundation of revenue. Therefore, theorganizes of the points should not be recognized to the server.Weshoulder an honest-but-curious cloud facility provider. In this model, the serverimplements correctly the assumed protocol for processing kNNdemands, but will also try toconclude the location of the statistics points. It is thus necessary to encode all informationstored and treated at the server.To allow query assessment, a special type of encryption that permits processing onciphertexts is required. In our case, we usage the mOPE technique from [6].mOPEis a provably protected order-preserving encryption technique, and our techniques receivethe IND-OCPA security assurance against the honest-but-curious server providingbymOPE.Furthermore, we undertake that there is no conspiracy be-tween the customers and server,then the clients will not reveal to the server the encryption solutions.

Secure Range Query Processing Method

ThekNN queries on encrypteddata requires complex operations, but at the core of these operations sits a relatively simple scheme called mutable order-preserving encryption (mOPE) .mOPEallowssecure evaluation of range queries, and is the only provably secure order-preservingencoding system (OPES) known to date. The di_erence between

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

mOPE and previousOPES techniques is that it allows cipher texts to changevalue over time, hence the mutable attribute. Without mutability, it is shown in [6]that secure OPES is not possible.

Algorithms1 :Privacy-Preserving Processing Framework for Distance-based Queries
Input: q: the query at the client I: the index at the data owner E(·): the encryption known to the data owner and cloud.
Output: C: the query result
Step 1: data owner sends shadow index E(I) to the client;
Step 2: data owner sends the decryption E−1 (·) to the cloud;
Step 3: client initializes a set of seeds S with cloud;
Step 4: client initializes the root of E(I) as i, the next node to access;
Step 5: while q is not completed do
Step 6: client retrieves shadow index node E(i), computes and scrambles the local distances from E(q); //
Step 7: server receives the scrambled local distances, decrypts and recodes them; Repeat step2
Step 8: client updates i and C according to the recoded distances; then repeat step 3

Algorithm 2 :Complete Query Processing Protocol
Input: q : the query point at client r: the threshold for distance range query at client k: for kNN query at client root: the root entry of the shadow index at client
Output: C: the set of result objects Procedure:
Step 1: client initializes queue Q and C = ∅;
Step 2: client enqueues root into Q;
Step 3: while Q is not empty do
Step 4: client dequeues entry p from Q;
Step 5: if p is a leaf entry then
Step 6: C = C ∪ p;
Step 7: kNNquery : if |C| == k, return C;
Step 8: else
Step 9: client retrieves shadow index node $\vec{-}e = (e1, e2, ...)$ pointed by p;
Step 10: client gets $D(\vec{-}q, \vec{-}e) = dist\ access(\vec{-}q, \vec{-}e)$;
Step 11: client enqueues (ei, $D(\vec{-}q, ei)$) of qualified ei into Q as below;
Step 12: Range query : those ei whose $D(\vec{-}q, ei)) \leq recode(4r\ 2)$;
Step 13: kNNquery : all ei;

V. RESULTS AND DISCUSSION

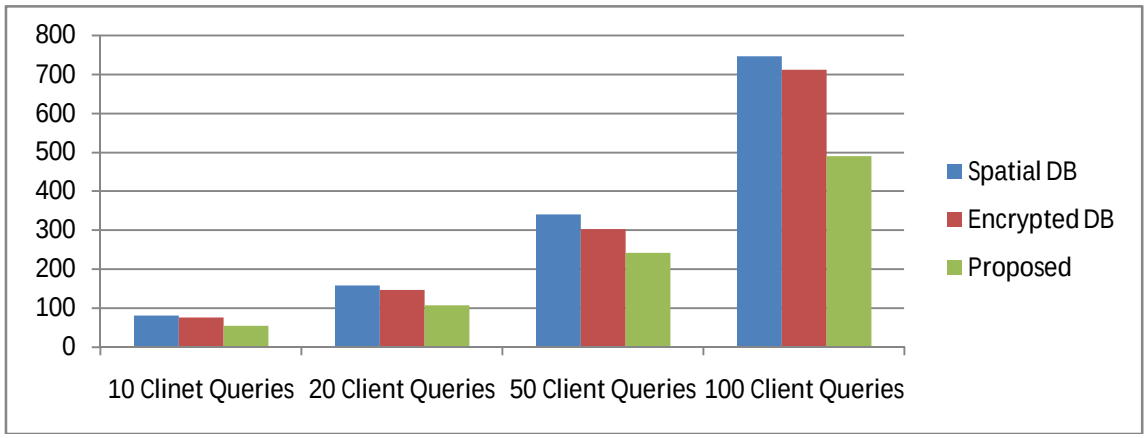

Fig. 2 Data Extraction performance by query with different Databases

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2015

We implemented the proposed protocols in Java, using JDK 1.6. We ran our experiments on a 2.5 GHz Intel Core 2 duo with 3GB RAM running Windows Vista..In fig.2 x axis shows the total number of queries executed by clients and y shows total number of time required in seconds with group query. On the basis of this graph we got all results on satisfactory level.

VI. CONCLUSION

This proposed research work, we study the problem of processing private queries on indexed data for mutual privacy protection in a cloud environment. We present a secure index traversal framework, based on which secure protocols are devised for classic types of queries. Through theoretical proofs and performance evaluation, this approach is shown to be not only feasible, but also efficient and robust under various parameter settings. We believe this work steps towards practical applications of privacy homomorphism to secure query processing on large-scale, structured datasets. As for future work, we plan to extend this work to other query types, including top-k queries, skyline queries and multi-way joins. We also plan to investigate mutual privacy protection for queries on semior unstructured datasets.

REFERENCES

[1] Sunoh Choi et.al., Secure kNN Query Processing in Untrusted Cloud Environments, IEEE TKDE, 2014
[2] Sunoh Choi et.al., Secure Proximity Detection in Untrusted Cloud Environments, submitted to VLDB, 2014
[3] Sunoh Choi et.al., Authenticated Top-K Aggregation in Distributed and Outsourced Databases, IEEE PASSAT, 2012
[4] Sunoh Choi et.al., Secure and Resilient Proximity-based Access Control, ACMDARE, 2013
[5] Sunoh Choi et.al., Secure Sensor Network SUM Aggregation with Detection ofMalicious Nodes, IEEE LCN, 2012.
[6] Raluca Ada Popaet.al., An Ideal-Security Protocol for Order-Preserving Encoding, IEEE SP, 2013
[7] Josep Domingo-Ferrer. A provably secure additive and multiplicative privacy homomorphism.In Proc. 5th International Conference on Information Security, 2002.
[8] W. Du and M. Atallah. Privacy-preserving cooperative statistical analysis.In Proceedings of the 17th Annual Computer Security Applications Conference, 2001.
[9] M. A. Soderstrand et al. Residue Number System Arithmetic: Modern Applications in Digital Signal Processing. New York: IEEE Press, 1986.
[10] R. Gallant, R. Lambert, and S. Vanstone. Faster point multiplication on elliptic curves with effientendomorphisms. In Advances in CryptologyCrypto 2001, LNCS 2139, Springer-Verlag, 2001.
[11] P. Gastaldo, G. Parodi, and R. Zunino. Enhanced montgomery multiplication on dsp architectures for embedded public-key cryptosystems. EURASIP Journal on Embedded Systems, 2008(April), 2008.
[12] BugraGedik and Ling Liu. Location-privacy in mobile systems: A personalized anonymization model. In Proc. ICDCS, 2005.
[13] G. Ghinita, P. Kalnis, and S. Skiadopoulos.Prive: Anonymous locationbased queries in distributed mobile systems. In WWW, 2007.
[14] Gabriel Ghinita, PanosKalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: Anonymizers are not necessary. In Proc. of SIGMOD, 2008.
[15] OdedGoldreich. The Foundations of Cryptography – Volume 2.Cambridge University Press, 2004.
[16] M. Gruteser and D. Grunwald.Anonymous usage of location-based services through spatial and temporal cloaking.In Proc. MobiSys, 2003.
[17] Gisli R. Hjaltason and HananSamet. Distance browsing in spatial databases. ACM TODS, 24(2):265 – 318, 1999.
[18] Haibo Hu and JianliangXu.Non-exposure location anonymity.In Proc. ICDE, 2009.
[19] M. Kantarcioglu and C. Clifton.Privacy preserving k-nnclassifier.In Proc. ICDE, 2005.
[20] A. Khoshgozaran and C. Shahabi.Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy.In Proc. SSTD, 2007.