ISSN: 2319-6505

Available Online at http://journalijcar.org

International Journal of Current Advanced Research
Vol 4, Issue 11, pp 499-503, November 2015

International Journal of Current Advanced Research

ISSN: 2319 - 6475

RESEARCH ARTICLE

SURVEY PAPER ON QUERY OPTIMIZATION TECHNIQUES IN THE CLOUD

Utsav Raychaudhuri, Chandni Singhal, Neelima Shahi, Shital Jadhav Aditi Kalia, and Vaishali Kolhe

Department Of Computer Engineering, D.Y. Patil College Of Engineering, Akurdi,Pune Savitribai Phule Pune University

A R T I C L E I N F O

Article History:

Received 9thSeptember, 2015
Received in revised form 21st September, 2015
Accepted 10thOctober, 2015
Published online 28th November, 2015

Key words:

Location Privacy, Spatial Databases, Database Outsourcing, Mutable Order Preserving Encoding

A B S T R A C T

Client's interests and their intrigues are distinctive. Gathering substantial measure of information to the client is exceptionally troublesome. At some point data is exceptionally delicate; taking care of the data is extremely troublesome assignment. It is permitted just to remunerate clients. Client send their present area of focuses, they need to think about closest POI's in NN however information Owner does not have that much stockpiling limit so cloud administration is being utilized. Cloud gives power full stockpiling requiring little to no effort however Cloud is not fully trusted. So NN inquiries are being utilized as a part of an untrusted outsourced environment, while at an identical time defensive the POI and questioning the client's area positions. The procedural in view of changeable request safeguarding encoding (mOPE). It is a protected request safeguarding encryption.

INTRODUCTION

There are numerous area based administrations that are accessible, diverse area based administration are (Maps and route, Tracking administration, Information administration, Social networking).The consistent redesigns of remote systems and cell phones .So the advancement has made a difference an exceptional examination in versatile information administrations. Some of these administrations furnish the client with setting mindful data. In particular, area-based administrations and area-ward inquiries have utilizing parcel of hobby. Client needs to handle as to very surprising information in their physical zone. Clients are empowered to get to data about purposes of interest (POI) what's more, are additionally near their area organizes. The majority of the inquiries include area traits is spoken to by closest – neighbor (NN) queries. Clients need to think about k POIs (e.g., eateries, exhibition halls, corner stores) that are closest to the client's present area (kNN). Gathering extensive measure of information to the client is exceptionally troublesome. At some point data is exceptionally touchy. So taking care of the data is exceptionally troublesome undertaking. Data stockpiling likewise enormous issue.so information proprietor does not make information open in all clients. It's permitted just paying clients. Client sends their present area guides and needs toward think about closest POIs in NN however information Owner does not have that much stockpiling limit. So we are utilizing cloud administration. Diverse sort Service like Google furthermore, Amazon are passing to the SaaS (Software as a Administration) business. Cloud gives power full stockpiling requiring little to no effort. But Cloud is not completely trusted so permit question preparing of NN inquiries in an not completely trusted environment, while in the meantime ensuring each the POI and questioning clients' positions. Our procedure is impermanent request safeguarding encoding (mOPE). It is a protected request-pholding encryption. We gives execution enhancements to lessening the computational expense. We present a broad execution assessment of our system is reasonability in hone. We present the two routines. One of the routines is VD- kNN system for secure NN inquiries. This works by handling scrambled. The strategy returns precise one result, yet it is exorbitant for k>1, and it is exceptionally hard to the information proprietor. To handle the requirements of VD-kNN, we present T-kNN, a strategy that works by preparing encrypted Delaunay triangulations, underpins any estimation of k. Also, It support diminishes the heap at the information proprietor. Gives exact question results for k=1, yet when k>1 the outcomes. It returns square measure exclusively rough. Nonetheless, we tend to demonstrate that in take after hone the precision is high. We diagram an instrument for upgrading scrambled Voronoi outlines and Delaunay triangulations that permits us to bargain with effectively, in partner dynamic way, with changing datasets Then Geo-Tagging the closest neighbor inquiry preparing in Untrusted Cloud Environment.

Related Work

Location Cloaking

This techniques replace the exact location of a user with a cloaking region (CR), typically of rectangular shape. To establish result correctness, the CR must enclose the actual user location. Furthermore, CRs must satisfy certain necessity dictated by a privacy paradigm, which expresses the privacy requirements of the user (e.g., spatial k-anonymity (SKA) [1] requires each CR to contain at least k distinct users). Regardless of the method used to develop the CR, query

processing at the LS side is performed with respect to a rectangular region, as opposed to an exact user location. In effect, the result returned by the LS is a super-set of the actual query result.

Private Information Retrieval

(PIR) methods rely on a cryptographic protocol to achieve query privacy [1]. In a pre-processing phase, the LS contructs the POI database into a data structure relevant to the supported type of query, and maps it to an ordered array D[1...n]. At runtime, a query is altered from a context-based (i.e., spatial) query to a query-by-index (i.e., return the ith item), according to the data organization which is known by the users. When a user wishes to get D[i], s/he creates an encoded query object q(i). Using a mathematical transformation, the LS calculates privately (i.e., without studying the value of i) the result r(D,q(i)) and sends it back to the user. PIR protocols ensure that it is computationally tough for the LS to recover the value i from q(i), but at the same time the user can easily re-construct D[i] from r.

Order Preserving Encoding

Order-preserving symmetric encryption (OPE) is an encryption scheme (aka. cipher) whose encoding function preserves numerical ordering of the plaintexts. OPE has a long period of history in the form of one-part codes, which are series of plaintexts and the corresponding ciphertexts, both arranged in alphabetical or numerical order so only a single copy is needed for profitable encoding and decoding. One-part codes were used, for example, during World War I [2]. The reason for new interest in such schemes is that they allow profitable range queries on encoded data. That is, a remote untrusted directory server is able to index the (sensitive) data it receives, in encrypted form, in a data structure that permits efficient range queries (asking the server to return ciphertexts in the directory whose decodings fall within a given range, say [a, b]). By "efficient" it means in time logarithmic (or at least sub-linear) in the size of the directory, as performing linear work on each query is prohibitively slow in practice for large directories. In fact, OPE not only allows convenient range queries but allows indexing and query processing to be done exactly and as conveniently as for unencrypted data, since a query just consists of the encryptions of a and b and the server can locate the desired key in logarithmic-time via standard tree-based data structures. Indeed, subsequent to its publication, [2] has been referenced widely in the database community, and OPE has also been suggested for use in in-network aggregation on encoded data in sensor networks and as a tool for applying signal processing methods to multimedia content protection.

K - Anonymity In Relational Databases

Anonymity was first discussed in relational directories, where advertised data (e.g., census, medical) should not be linked to specific persons. Recent work has focused on K-anonymity as defined in [3]: a relation satisfies K-anonymity if every tuple is identical to at least K-1 other tuples with respect to a set of quasi-identifier traits. Quasi-identifiers are fields (e.g., date of birth, gender, zip code) that can be linked to publicly available data to identify individuals. Records with identical quasi- identifiers form an anonymized group. Two techniques are used to transform a relation to a K-anonymized one: suppression, where some of the tuples are removed and generalized, which involves replacing specific values (e.g., phone number) with more general ones (e.g., only area code). Both methods lead to knowledge loss. It shows that anonymizing a high-dimensional relation leads to unacceptable damage of information due to the dimensionality curse.

Random Space Perturbation (Rasp) Technique

RASP is one type of multiplicative perturbation, with a novel combination of OPE, scope expansion, random noise injection, and random projection. Let's consider the multifaceted data are numerical and in multifaceted vector space. The directory has k searchable dimensions and n records, which makes a d × n matrix X. The searchable elements can be used in queries and thus should be indexed. Let x represent a d-dimensional record,

x ∈ Rd. Note that in the d-length vector space Rd, the range query conditions are represented as half-space functions and a range uery is rendered to finding the point set in corresponding polyhedron area described by the half spaces [4].

The RASP perturbation utilizess three steps. Its security is based on the existence of random invertible real-value matrix generator and arbitrary real value generator. For each k-dimensional input vector x,

1) An order preserving encryption (OPE) method [1], Eope with keys Kope, is applied to each dimension of x: Eope(x, Kope) ∈ Rd to change the elemental distributions to normal distributions with each dimension's value order still preserved.

2) The vector is then enhanced to d+2 dimensions as G(x) = ((Eopt(x))T ,1,v)T, where the (d + 1) - th dimension is always a 1 and the (d + 2) - th dimension, v, is drawn from a arbitrary real number generator RNG that generates random values from a tailored normal distributions.
3) The (d + 2)-dimensional vector is finally transformed to
F(x, K = {A, Kope, RG}) =A((Eope(x))T, 1, v)T, (1)
where A is a (d+2)×(d+2)arbitrarily generated invertible matrix with aij ∈ R such that there are at least two non-zero values in each line of A and the last column of A is also non-zero.

Kope and A are shared by all vectors in the directory, but v is randomly generated for each individual vector. Since the RASP-perturbed data records are only used for indexing and supporting query processing, there is no need to recover the perturbed data.

Casper

A novel framework that turns traditional geo-positional servers and query processors to provide anonymous service to their customers. In Casper, mobile users can enjoy location-

based services without the need to reveal their private location information. Upon registration with [5] Casper, mobile users specify their convenient level of privacy through a user-specified privacy profile. A user privacy profile includes two arguments k and Amin. k indicates that the mobile user wants to be k-anonymous, i.e., not differentiable among other k users while Amin indicates that the user wants to hide her location data within an area of at least Amin. Large values fork and Amin indicate more strict privacy requirements. Casper mainly consists of two parts, namely, the location anonymizer and the privacy-aware query processor. The location anonymizer is a verified third party that acts as a middle layer between mobile users and the location-based database server in order to:

1. Retrieve the exact location information from mobile users along with a privacy profile of each user,
2. Obscure the exact location information into veiled spatial areas based on each user privacy profile, and
3. Send the veiled spatial areas to the location-based directory server.

The privacy-aware query processor is embedded inside the location-based directory server to tune its functionality to accord with anonymous queries and cloaked spatial areas rather than the exact location information. There are three novel query types that are supported by Casper:

1. Private queries over public information, e.g., “Where is my nearest coffee shop”, in which the individual who issues the query is a private entity while the data (i.e., coffee shops) are public,
2. Public queries over private data, e.g., “How many cars in a certain area”, in which a public entity asks about personal private locations, and
3. Private queries over private data, e.g., “Where is my closest restaurant” in which both the person who issues the query and the requested data are protected.

With this classification in mind, conventional location-based query processors can support only public queries over public data. Due to the absence of the exact location information at the server, the anonymous query processor gives a candidate list of answers instead of a single exact answer. Hence it proves that the candidate list is inclusive, i.e., has the exact answer, and is low, i.e., a high quality answer is given to the users.

Ind-Cpa

Ciphertext identicalness is a property of many encoding schemes. Intuitively, if a cryptosystem possesses the property of indistinguishability, then an adversary will not be able to distinguish pairs of ciphertexts based on the message they encode. The property of identicalness under chosen plaintext attack is considered a basic requirement for most arguably secure public key cryptosystems, though some strategies also provide identicalness under chosen ciphertext attack and robust chosen ciphertext attack. Indistinguishability under chosen plaintext attack is equivalent to the feature of semantic immunity, and many cryptographic proofs use these interpretations interchangeably [6].

For a probabilistic asymmetric key encoding algorithm, indistinguishability under chosen plaintext attack [7] (IND-CPA) is defined by the following game between an attacker and a challenger. For schemes based on computational security, the attacker is formed by a probabilistic polynomial time and it must finish the game and yield a guess within a polynomial number of time steps. In this interpretation E(PK, M) represents the encoding of a message M under the key PK:

1. The contender generates a key pair PK, SK based on some security value k (e.g., a key size in bits), and publishes PK to the attacker. The contender retains SK.
2. The attacker may perform a polynomially bounded number of encryptions or other operations.
3. Eventually, the attacker outputs two speific chosen plaintexts M0, M1 to the challenger.
4. The contender selects a bit b € {0, 1} uniformly at random, and sends the challenge ciphertext C = E(PK, Mb) back to the attacker.
5. The attacker is free to perform any number of additional computations or encryptions. Finally, it outputs a guess for the value of b.

A cryptosystem is indistinguishable under chosen plaintext attack if every probabilistic polynomial time adversary has only a negligible "advantage" over random guessing. An adversary is said to have a negligible "advantage" if it wins the above game with probability, where €(k) is a negligible function in the security parameter k, that is for every (nonzero) polynomial function there exists k0 such that for all k>k0.

Although the adversary knows M0, M1 and PK, the probabilistic nature of E means that the encryption of Mb will be only one of many valid ciphertexts, and therefore encrypting M, M1 and comparing the resulting ciphertexts with the challenge ciphertext does not afford any non-negligible advantage to the adversary.

While the above definition is specific to an asymmetric key cryptosystem, it can be adapted to the symmetric case by replacing the public key encryption function with an "encryption oracle", which retains the secret encryption key and encrypts arbitrary plaintexts at the adversary's request.

System Model

Consider a data management system hosting data service, as illustrated in Figure 1, in which three different entities are involved: data owner, data user and a storage server. The data owner has a collection of data files. Data owners are encouraged to outsource their data from local systems to global space for great flexibility. For protecting data files, they are encrypted before uploading into such global space. Thus enabling search and retrieval over such encrypted data is of paramount importance. The data owner has a collection of n files say, C = {f1, f2, ...fn } which may be of extension .txt, . doc and .pdf. For protecting the file from the unauthorized person we need to apply different types of privacy homomorphism algorithms

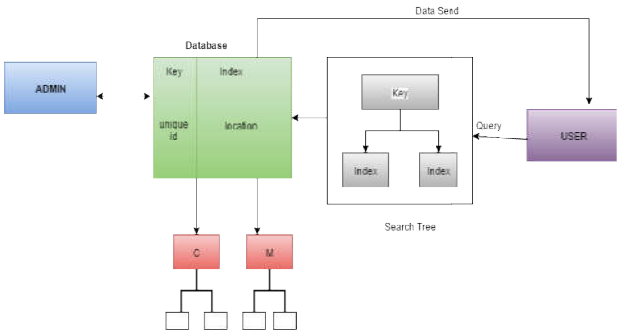

Figure1 Data Exchange Module

Homomorphic encryption has been used to provide a strong privacy protection for the sensitive data. Homomorphic encryption allows addition and multiplication without the need for decryption to be directly performed on cipher texts and that too without loss of generality. We use the popular Paillier's homomorphic encryption.

Literature Survey

A Hybrid Technique for Private Location-Based Queries with Database Protection

Cell phones with geo positioning abilities permit clients to recover points of interest (POI) in their vicinity. To secure client protection, it is essential not to uncover precise client directions to un-trusted vendors that give area based administrations. Right now, there are two fundamental ways to deal with securing the area protection of clients

1. Concealing areas inside shrouding locales (CRs)
2. Scrambling area information utilizing private data recovery (PIR) conventions.

Past work concentrated on finding great payoffs in the middle of security and execution of client assurance methods, however dismissed the essential issue of ensuring the POI dataset D. Case in point, area shrouding obliges expensive measured CRs, prompting inordinate exposure of POIs (O(| D |) in the most pessimistic scenario). PIR, then again, decreases this bound to O(p| D |), yet to the detriment of high preparation and correspondence overhead. A half breed, two-stage way to deal with private area based queries, which gives assurance to both the clients and the database. In the first step, client areas are summed up to coarse-grained CRs which give solid security. Next, a PIR protocol is applied with respect to the obtained query CR. To ensure exorbitant divulgence of POI areas, a cryptographic convention has been devised that secretly assesses whether a point is encased inside a rectangular area. Likewise a calculation to efficiently bolster PIR on elements of POI sub-sets has been acquainted [1]. The system uncovers O(1) POI, requests of greatness less than CR-or PIR based procedures. Experimental results demonstrate that the cross breed methodology is versatile, and plainly beats the immaculate PIR approach regarding computational and correspondence.

An Ideal-Security Protocol for Order-Preserving Encoding

Order preserving encryptionan encryption plan where the sort request of cipher texts matches the sort request of the comparing plaintexts permits databases and different applications to process queries, including request over encoded information effectively. The perfect security measure for request protecting encryption set forth in the writing is for the cipher texts to uncover no data about the plaintexts other than request. Despite the fact that more than twelve plans were proposed, every one of these plans release more data than request. It introduces the first request protecting plan that accomplishes perfect security. The primary system is impermanent cipher texts, implying that after some time, the cipher texts for a little number of plaintext qualities change, and demonstrate that alterable cipher texts are required for perfect security. The subsequent convention is intuitive, with a little number of collaborations. The plan was executed and assessed on micro benchmarks furthermore, in the connection of a scrambled MySQL database application. The demonstration that notwithstanding giving perfect security, the plan accomplishes 12 requests of greatness higher execution than the best in class request safeguarding encryption plan, which is less secure than the scheme [2].

Blind Evaluation of Nearest Neighbor Queries Using Space Transformation to Preserve Location Privacy

A central way to deal with performance of the class of Nearest Neighbor (NN) queries, the central class of queries utilized as a part of a large number of the area based administrations, without un- covering the query keeping in mind the end goal to protect the security of this data. The thought behind the methodology is to use one-route changes to delineate space of all static and element items to another space and to determine the query indiscriminately in the changed space. On the other hand, keeping in mind the end goal to turn it into a practical methodology, the change utilized must be capable to determine NN questions in the changed space precisely and all the more should significantly forestall malevolent utilization of changed information by untrusted substances. Conventional encryption based systems bring about lavish O(n) calculation cost (where n is the aggregate number of focuses in space) and conceivably logarithmic correspondence cost for determining a kNN inquiry. This is on account of such methodologies which regards focuses as vectors in space and don't misuse their spatial properties. Conversely, utilization of Hilbert bends as effective restricted changes and configuration calculations to assess a kNN query in the Hilbert changed space. Subsequently, diminish the complexity of processing a kNN question to O(KN) and exchanging the outcomes to the customer in O(K), individually, where N, the Hilbert bend degree, is a little inconsistent. The outcomes demonstrate that firmly summarized, the outcome sets produced from performing kNN in queries in the first space while upholding the new area protection measurements termed u-obscurity and a-secrecy, which are more grounded furthermore, more summed up security measures than the usually utilized K-obscurity furthermore, shrouded area size measures [3].

Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation

With the wide organization of open distributed computing databases, utilizing clouds to host information question administrations has turned into an engaging answer for the

focal points on versatility and expense sparing. On the other hand, some information may be delicate that the information proprietor would not like to move to the cloud unless the information confidentiality and security are ensured. On the other hand, a secured question administration ought to still give efficient inquiry preparing and significantly lessen the in-house workload to completely understand the benefits of distributed computing. The RASP information perturbation system has been proposed to give secure and efficient reach question and kNN inquiry administrations for ensured information in the cloud. The RASP in- formation bother strategy joins request saving encryption, dimensionality extension, arbitrary clamor infusion, and irregular projection, to give solid flexibility to assaults on the irritated information and inquiries. It additionally safeguards multidimensional extents, which permits existing indexing methods to be connected to speedup reach question preparing. The kNN-R calculation is intended to work with the RASP reach question calculation to handle the kNN inquiries. The assaults on information have been painstakingly investigated and inquiries under an unequivocally dened danger model and practical security suspicions. Broad trials have been directed to demonstrate the benefits of this methodology on efficiency and security [4].

Secure kNN Computation on Encrypted Databases

Information suppliers like Google and Amazon are moving into the SaaS (Software as a Service) business. They turn their tremendous foundation into a distributed computing environment and force- fully select organizations to run applications on their stages. To authorize security and protection on such an administration model, the information running on the stage has to be ensured. Tragically, conventional encryption strategies that go for giving unbreakable" insurance are frequently not sufficient since they don't bolster the execution of utilizations for example, database inquiries on the scrambled information. The general issue of secure calculation has been examined on an encoded database and propose a SCONEDB (Secure Computation ON an Encrypted Data Base) model, which catches the execution and security prerequisites. Another lopsided scalar-item safeguarding encryption (ASPE) that jam an exceptional kind of scalar item has been created. APSE has been utilized to develop two secure plans that back kNN calculation on encoded information; each of these plans is appeared to oppose reasonable assaults of an alternate foundation learning level, at an alternate overhead cost [5].

Order-Preserving Symmetric Encryption

The cryptographic investigation has been started for request safeguarding symmetric encryption (OPE), a method proposed in the database group for permitting proficient extent queries on scrambled information. Interestingly, a demonstration has been given of a clear unwinding of standard security ideas for encryption, for example, vagary against picked plaintext assault (IND-CPA) is unachievable by a pragmatic OPE plan. Rather, a security idea has been proposed in the soul of pseudo random capacities (PRFs) and related primitives asking that an OPE plan look as arbitrary as it could reason- ably be expected subject to the request saving limitation. An outline of a productive OPE has been conspired and demonstrate its security under the idea in light of pseudo randomness of a basic block cipher. The development depends on a characteristic connection that has been revealed between an arbitrary request safeguarding capacity and the hyper geometric likelihood circulation. Specifically, it makes discovery utilization of an effective examining calculation for the last [6].

CONCLUSION

A study is conducted on processing problems of private queries on indexed data in a cloud. A secure traversal framework in indexed environment is given to secure protocols for such classic queries. The assumptions and approached mentioned in this paper are thoroughly useful, efficient to perform and effectively used under settings of different parameters. It has been summarized that the process mentioned here, on privacy homomorphism, is used to protect processing queries on cloud is high scalable.

References

1. Gabriel Ghinita, Panos Kalnis, Murat Kantarcioglu, and Elisa Bertino, A Hybrid Technique for Private LocationBased Queries with Database Protection, SSTD09
2. Raluca Ada Popa, Frank H. Li, and Nickolai Zeldovich, An Ideal-Security Protocol for Order-Preserving Encoding, IEEE S&P13
3. Gabriel Ghinita, Panos Kalnis, Murat Kantarcioglu, and Elisa Bertino, Approximate and exact hybrid algorithms for private nearest-neighbor queries with database protection, Geoinformatica11
4. Huiqi Xu, Shumin Guo, and Keke Chen, Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation, TKDE12
5. W. K. Wong, David W. Cheung, Ben Kao, and Nikos Mamoulis, Secure kNN Computation on Encrypted Databases, SIGMOD09
6. Alexandra Boldyera, Nathan Chenette, Younholee and Adam O'Neil, Order Preserving Symmetric Encryption, GA USA

****